Water Resources Research

RESEARCH ARTICLE
10.1029/2022WR032204

Moving Land Models Toward More Actionable Science: A Novel Application of the Community Terrestrial Systems Model Across Alaska and the Yukon River Basin

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

[1]National Center for Atmospheric Research, Boulder, CO, USA, [2]Institute of Arctic and Alpine Research, University of Colorado, Boulder, CO, USA, [3]Earthmover, New York, NY, USA

Key Points:

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

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

Y. Cheng,
yifanc@ucar.edu

Citation:

Cheng, Y., Musselman, K. N., Swenson, S., Lawrence, D., Hamman, J., Dagon, K., et al. (2023). Moving land models toward more actionable science: A novel application of the Community Terrestrial Systems Model across Alaska and the Yukon River Basin. *Water Resources Research*, *59*, e2022WR032204. https://doi.org/10.1029/2022WR032204

Received 15 FEB 2022
Accepted 18 DEC 2022

Author Contributions:

Conceptualization: Yifan Cheng, Joseph Hamman
Data curation: Yifan Cheng, Sean Swenson
Formal analysis: Yifan Cheng
Funding acquisition: Keith N. Musselman, Joseph Hamman, Andrew J. Newman
Investigation: Yifan Cheng, Sean Swenson

Abstract 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. In this study, we aim to lower the barrier of using complex land models in regional applications by developing a generalizable optimization methodology and workflow for the Community Terrestrial Systems Model (CTSM), to move them toward a more Actionable Science paradigm. Further end-user engagement is required to make science such as this "fully actionable." We applied CTSM across Alaska and the Yukon River Basin at 4-km spatial resolution. We highlighted 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 10 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, 13 had improved flow simulations after optimization and the mean Kling-Gupta Efficiency of daily flow increased from 0.43 to 0.63 in a 30-year evaluation. 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 & Slater, 2005; Lawrence et al., 2012; 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 (Knoll et al., 2019; Pavelsky & Zarnetske, 2017; Sharma et al., 2019).

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

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

Methodology: Yifan Cheng, Sean Swenson, Joseph Hamman, Katherine Dagon, Daniel Kennedy, Andrew J. Newman
Project Administration: Andrew J. Newman
Resources: Sean Swenson, David Lawrence, Joseph Hamman, Andrew J. Newman
Supervision: Keith N. Musselman, Andrew J. Newman
Validation: Yifan Cheng
Visualization: Yifan Cheng
Writing – original draft: Yifan Cheng
Writing – review & editing: Yifan Cheng, Keith N. Musselman, Sean Swenson, David Lawrence, Joseph Hamman, Katherine Dagon, Daniel Kennedy, Andrew J. Newman

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, Clark, Barlage, et al., 2015), allow for parameter sensitivity and application-oriented optimization studies of these advanced LMs.

In this study, we 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 improve 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 the physical representation like convection or orographic impacts (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 (Jafarov et al., 2012; 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).

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

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

The sophisticated land process representations in CTSM with high spatial resolution make it a potentially robust tool in projecting climate impacts on hydrology, yet its complexity undermines its useability for real-world

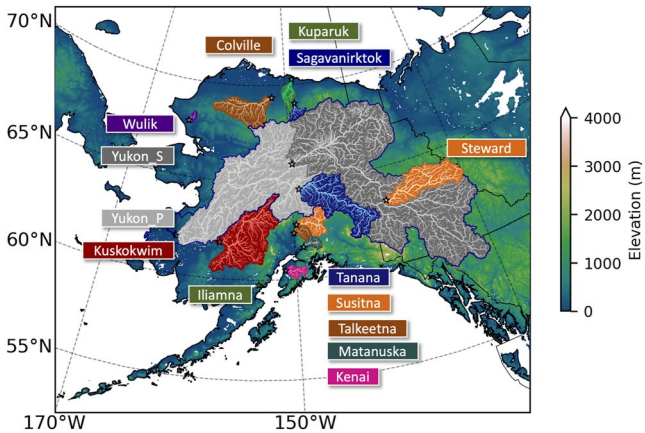

Figure 1. Study domain. The dark blue line denotes the boundary of the Yukon River Basin and black stars denote the outlets of the highlighted river basins.

applications. Therefore, in this study, we provide: (a) a methodology for efficient optimization of CTSM to lower the barrier of using CTSM in real-world applications and enhance the actionability of CTSM; (b) a high-resolution Arctic CTSM configuration focused on improved hydrologic simulation fidelity; (c) an evaluation of the performance of the regional CTSM configuration and its actionability; (d) tools available to the scientific community to apply our methodology to other regions and applications. We evaluated the model performance, or actionability, by using hydrology-related metrics that are related to events of concern from our IAC and the climate information survey available in Section 4.5. Additionally, this study lays the foundation for knowledge co-production research with Indigenous communities on a range of topics, including improving our understanding of climate-induced impacts on the rivers and fishes, and communities necessary to inform adaptation efforts in our study domain.

Truly actionable science involves a high-level of stakeholder engagement throughout the development, application, and evaluation of results to determine specific thresholds of performance for a specific use case. This study does not produce actionable science in that sense, but we aim to lower the barriers of using CTSM and other complex, process-rich land models in regional applications by developing an optimization workflow and providing an example evaluation for a specific application. Therefore we are moving models like CTSM toward a more actionable Earth Science paradigm (Findlater et al., 2021) with this exemplar 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, that is, one at the Pilot station and one near Stevens Village denoted as Yukon_P and Yukon_S in Figure 1.

3. Baseline CTSM Configuration

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

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° (~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, for example, orographic precipitation, altitudinal temperature gradients (Monaghan et al., 2018; Rasmussen et al., 2011). Therefore, we performed a simple downscaling to add

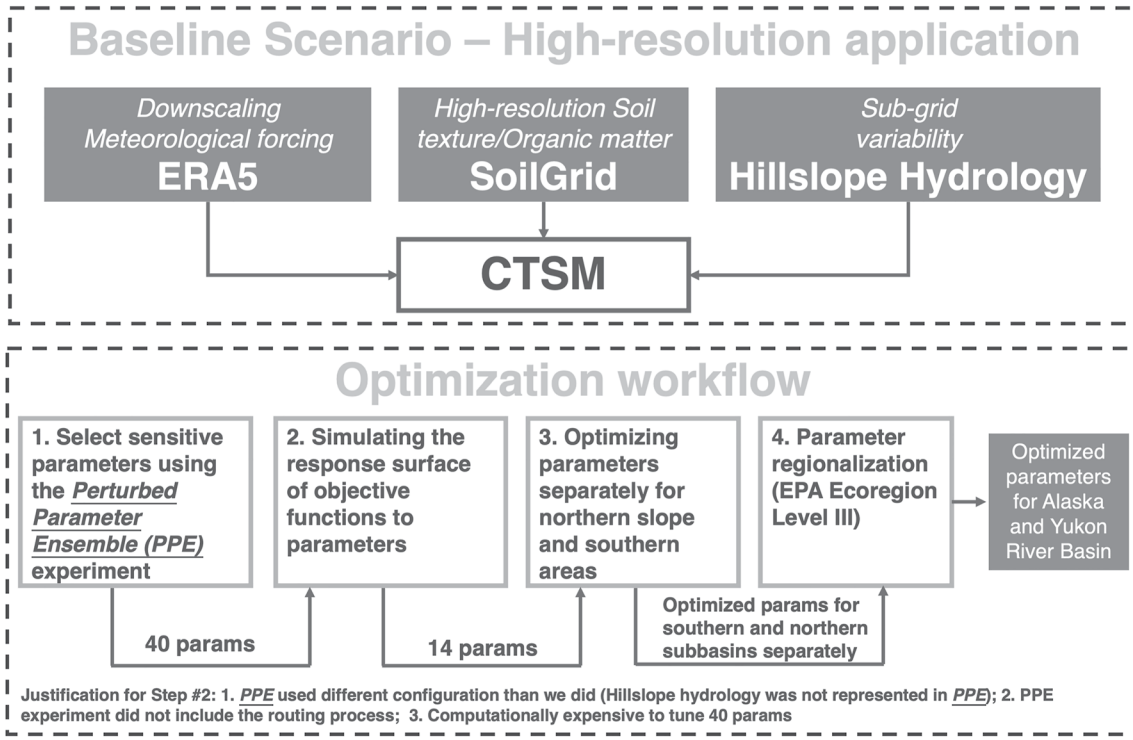

Figure 2. Community Terrestrial Systems Model baseline scenario and workflow for optimization.

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).

$$\Pi^{P}_{\mathrm{M,H},g} = \frac{\overline{P^{\mathrm{WRF}}_{\mathrm{M,H},g}}}{\overline{P^{\mathrm{ERA5}}_{\mathrm{M,H},g}}} \tag{1}$$

$$P^{ds}_{m,h,g} = P^{\mathrm{ERA5}}_{m,h,g} \times \Pi^{P}_{\mathrm{M,H},g} \tag{2}$$

where P denotes precipitation. Π denotes the multiplicative correction factor, which has three dimensions, that is, 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. Lowercase 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.

$$\Sigma^{v}_{\mathrm{M,H},g} = \overline{v^{\mathrm{WRF}}_{\mathrm{M,H},g}} - \overline{v^{\mathrm{ERA5}}_{\mathrm{M,H},g}} \tag{3}$$

$$v^{\mathrm{d}s}_{m,h,g} = v^{\mathrm{ERA5}}_{m,h,g} + \Sigma^{v}_{\mathrm{M,H},g} \tag{4}$$

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

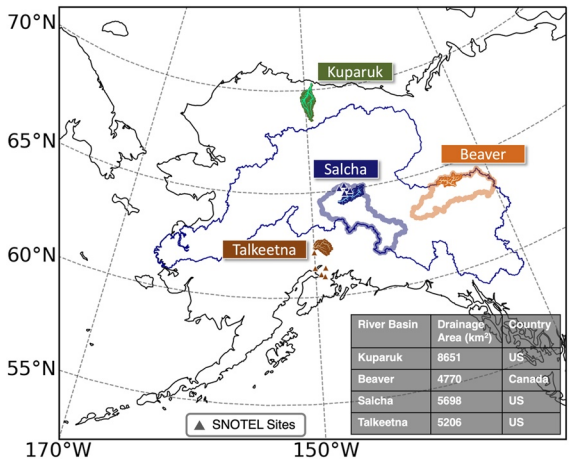

Figure 3. Selected representative medium-sized basins for parameter estimation. Triangles denote SNOTEL sites with snow observations. Salcha River Basin is a subbasin of the Tanana River Basin (thick blue line) and Beaver River Basin is a subbasin of the Steward River Basin (thick orange line).

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.

$$\mathrm{PCT}_{\mathrm{SAND}} = \frac{\gamma_{\mathrm{sand}}}{\gamma_{\mathrm{sand}} + \gamma_{\mathrm{silt}} + \gamma_{\mathrm{clay}}} \times 100\% \tag{5}$$

$$\mathrm{PCT}_{\mathrm{CLAY}} = \frac{\gamma_{\mathrm{clay}}}{\gamma_{\mathrm{sand}} + \gamma_{\mathrm{silt}} + \gamma_{\mathrm{clay}}} \times 100\% \tag{6}$$

$$\rho_{OM} = \frac{\gamma_C \cdot \rho_{\mathrm{bulk}} \cdot 10^{-3}}{0.58} \tag{7}$$

Percentages of sand and clay, $\mathrm{PCT}_{\mathrm{SAND}}$ and $\mathrm{PCT}_{\mathrm{CLAY}}$, 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.

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, that is, 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.

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

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

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 31 August, which is close to the date with the lowest annual SWE. If it is not perennial snow, the melting rate is calculated based upon the annual peak SWE and the first day when SWE falls below 0.1 mm. The snow objective function is the quadratic mean of the three relative error terms,

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

The aggregated snow metric O_S is unitless.

4.2. Parameter Sensitivity

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

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

Table 1
Summary of 14 Parameters Selected for Optimization, Their Categories, Relevant Physical Processes, Ranking Based on 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	72,000	[20,000, 250,000]	235,175	155,394
Hydrology	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
	liq_canopy_storage_scalar	Maximum storage of liquid water on leaf surface	11	0.1	[0.025, 4]	3.695	2.372
Plant hydraulics	krmax[a]	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	232.078
Stomatal resistance and photosynthesis	medlynintercept[a]	Medlynintercept of conductance-photosynthesis relationship	6	100	[120,000]	7.326×10^{3}	1.200×10^{4}

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

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.

- *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 Latin Hypercube Sampling (LHS, McKay et al., 2000) 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 in Supporting Information S1). The amplitude of the response curve was used to evaluate each parameter's sensitivity. For the basins where we conducted multi-objective optimization, we aggregated the amplitudes of the response surface across both objectives to make sure that the selected parameters should be generally sensitive for both objectives.

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, that is, 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, 19 parameters were in the top 10 most sensitive across all basins. We selected all parameters with a total score higher than one, meaning they were at least one of the six most sensitive parameters in any one basin, or somewhat sensitive in multiple basins. This resulted in 14 parameters being selected for full optimization (Table 1). It is possible that multiple parameters shared the same scores and therefore the same ranks, for example, *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 LHS 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 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, that is, 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 parameter values. We then ran CTSM using the N sets of parameter values and calculated their objective functions.

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

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

4.4. Parameter Regionalization

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

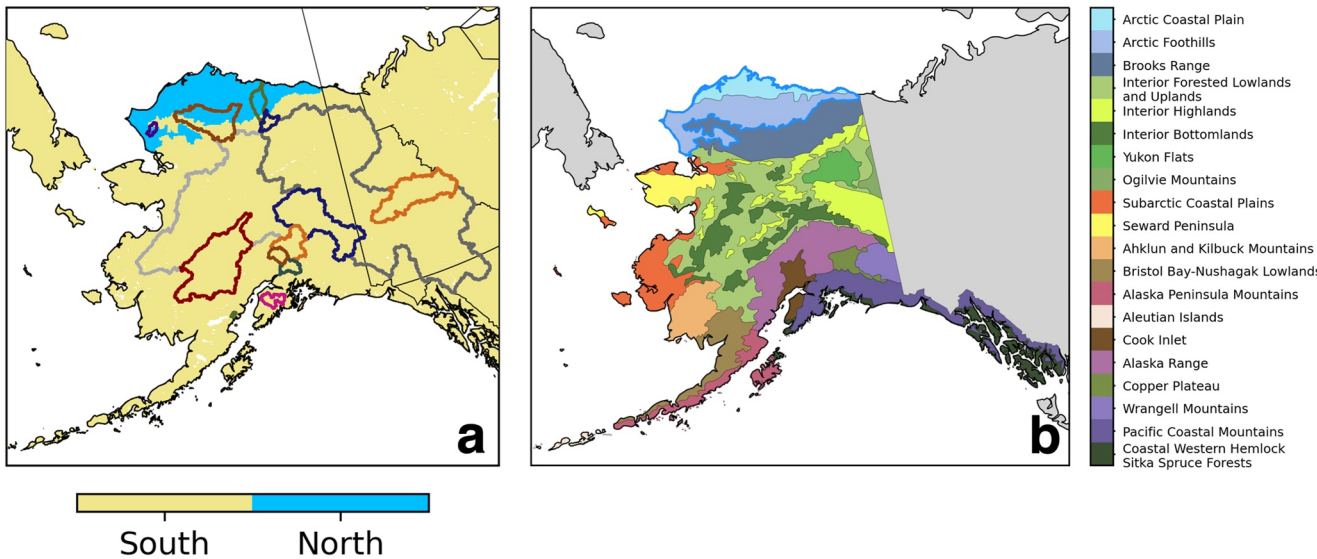

Figure 4. Parameter regionalization based on ecohydrology region classification. In this figure panel (a), background colors denote the selection of optimized parameters, and river basins are highlighted using solid lines with colors corresponding to Figure 1. In this figure panel (b), regions using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

flow objectives across the three basins and snow objectives across 10 SNOTEL sites, which serve as the representative objectives across the entire southern region so the selected SNOTEL sites and basins do not necessarily need to overlap (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; Liu et al., 2020). The EPA ecohydrology region classification helps determine the boundary between the southern and northern regions. Because the Kuparuk mostly resides north of the arctic circle, optimized parameters for the Kuparuk 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).

4.5. Actionability of CTSM

The definition of actionable science above suggests the necessity to set a performance benchmark that a model has to hit before it could be used to inform decision-making. We selected the following two hydrologic metrics, percent bias in flow duration curve (FDC) high-segment volume (%BiasFHV, Yilmaz et al., 2008) given the interest of the IAC and the results of the climate information survey focused on flooding, and percent bias in FDC low-segment volume during summer season (%BiasFLV$_{summer}$, May–October) because summer low flows have significant impacts on the juvenile production for salmon (Ohlberger et al., 2018). No universally accepted benchmarks exist for these metrics so we used the uncalibrated model performance as our benchmark. In addition, we selected an additional hydrologic metric, the Nash Sutcliffe Efficiency (NSE) for daily flow, because it is a widely used metric by water management authorities in the United States. We adopted a benchmark of daily NSE of 0.5 (Moriasi et al., 2015).

Complementary to solely evaluating flow simulations against observations, we conducted a climate sensitivity analysis to assess whether CTSM captures the response of flow simulations to changes in precipitation and air temperature. We followed the technique developed in Wood et al. (2004). For each river basin, regionally averaged flow ($\overline{\overline{Q}}$), air temperature ($\overline{\overline{T}}$), and precipitation ($\overline{\overline{Pr}}$) were calculated for each hydrologic year and for the obser-

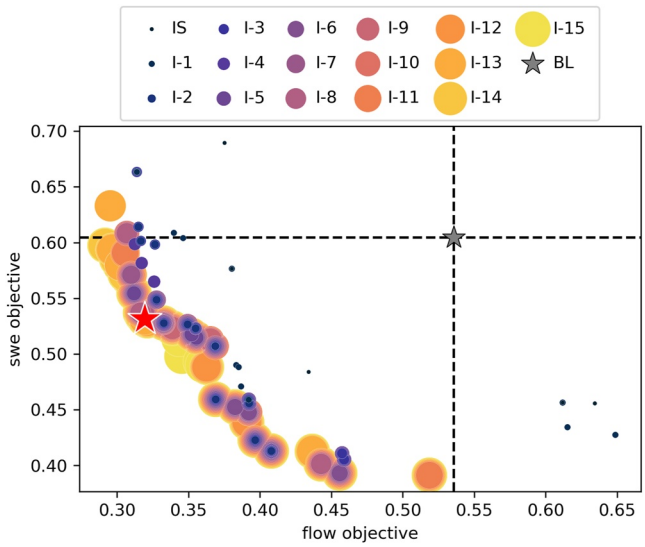

Figure 5. Simulated Pareto front of optimization for southern basins. Each colored dot corresponds to a Pareto optimal set of parameters. *IS* denotes initial sampling, *I*-1 denotes the first iteration, and so on, and *BL* denotes the baseline configuration. The red star denotes the selected optimized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

vations, baseline, and optimized simulations. To quantify the uncertainties in flow responses to climate variables, we conducted bootstrapping 300 times with each bootstrapping sample generating a new series of precipitation, air temperature, and streamflow by resampling the available hydrological years with replacement ($n = 5{,}000$ times, Brunner et al., 2020). For each new series, we fitted a simple linear regression between the flow and climate variables and the slope denotes the corresponding responses. By assessing the baseline and optimized responses to the observed response, we can evaluate whether the optimization improves estimated climate sensitivities.

4.6. 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; Cheng et al., 2022; Yu et al., 2014). We performed the analysis on the 14 optimized parameters for the southern and northern regions separately. The change in the objective function is calculated as

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

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

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

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

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

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

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

5. Results

5.1. Optimization

For the southern basins, flow simulation is improved substantially while snow simulation only sees minor improvements (Figure 5). Dots with the same color in Figure 5 constitute the simulated Pareto front for a given optimization iteration. A Pareto front consists of simulated Pareto optimal, that is, if none of the objective functions can be improved in value without degrading some of the other objective values. In general, the simulated Pareto front shifts toward the origin, signifying improved model performance. Overlapping dots indicate the new iteration failed to improve the Pareto front at that point. The two-dimensional Pareto front serves as the basis for choosing our optimal parameter set. Future work could explore using an ensemble of optimal parameter sets along

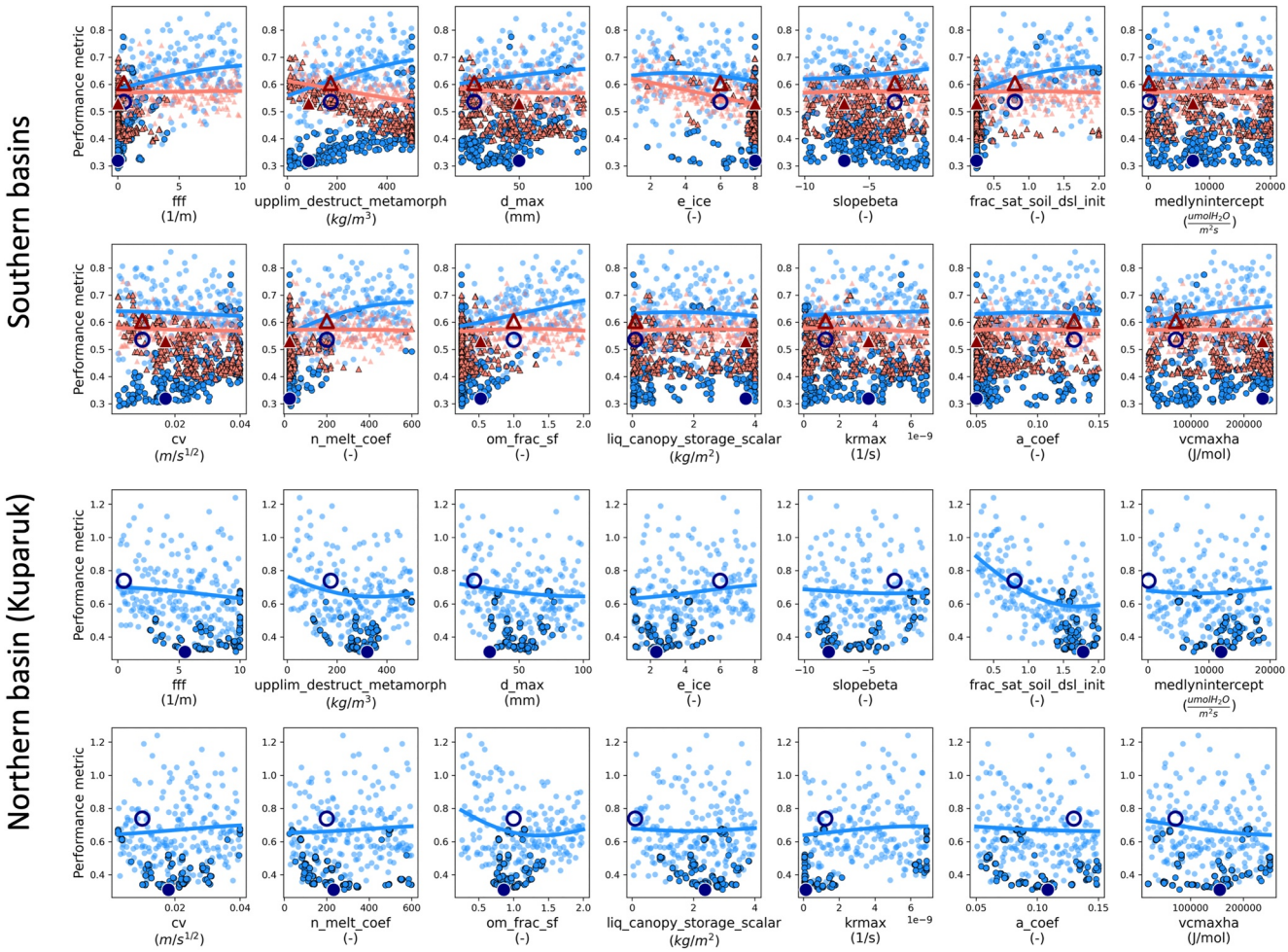

Figure 6. Mean response curve of flow (blue dots) and snow (red triangles) objectives to parameters. Transparent dots and triangles denote initial samples, solid dots, and triangles with black edges denote samples during optimization iterations, the large hallow dot and triangle with darker colors denote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

the Pareto front, but that is outside the scope of this initial investigation. We choose the set of optimized parameters that correspond to the minimum averaged flow and snow objectives, highlighted using a red star in Figure 5. For this parameter set, the corresponding flow and snow objective functions for the southern basins are 0.319 (0.681 KGE) and 0.531 respectively, while the baseline flow and snow objectives are 0.536 (0.464 KGE) and 0.604 respectively. Flow simulation in the Kuparuk is also significantly improved through optimization. Since we conducted a single-objective optimization for Kuparuk, we simply selected the set of parameters resulting in the best flow simulation. The optimized flow objective is 0.311 (0.689 KGE) while the default flow objective is 0.739 (0.261 KGE).

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, for example, *frac_sat_soil_dsl_init* and *e_ice*, 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 the specified 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.

5.2. A 30-Year 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 from 1991 to 2020 water year (WY). The daily KGE improves at 13 out of 15 basins and the mean KGE across the 15 basins increases from 0.43 to 0.63 after optimization. Furthermore, even though we only conducted optimizations for four medium-sized 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 −0.36 to 0.70 and the daily KGE for the Yukon River at Stevens Village (502,458 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.

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 or barely worse than that of any single optimized parameter set (Figure 8a). In addition, their default flow simulations are similar to the ones using spatially distributed parameters. Compared to the baseline, optimized flows in Wulik and Colville both show a smaller wet bias in spring, a larger dry bias in summer, and a smaller dry bias in fall (Figure 7). 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, that is, 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

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

Since we used the absolute value of KGE difference as the denominator, regardless of $\mathrm{KGE}_{\mathrm{optz}}$ being higher or lower than $\mathrm{KGE}_{\mathrm{base}}$, 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 eight, three, and two river basins respectively (Figure 8b). Poorly simulated flow volume and correlation mostly contribute to the poorer flow simulation in Colville and Iliamna, respectively.

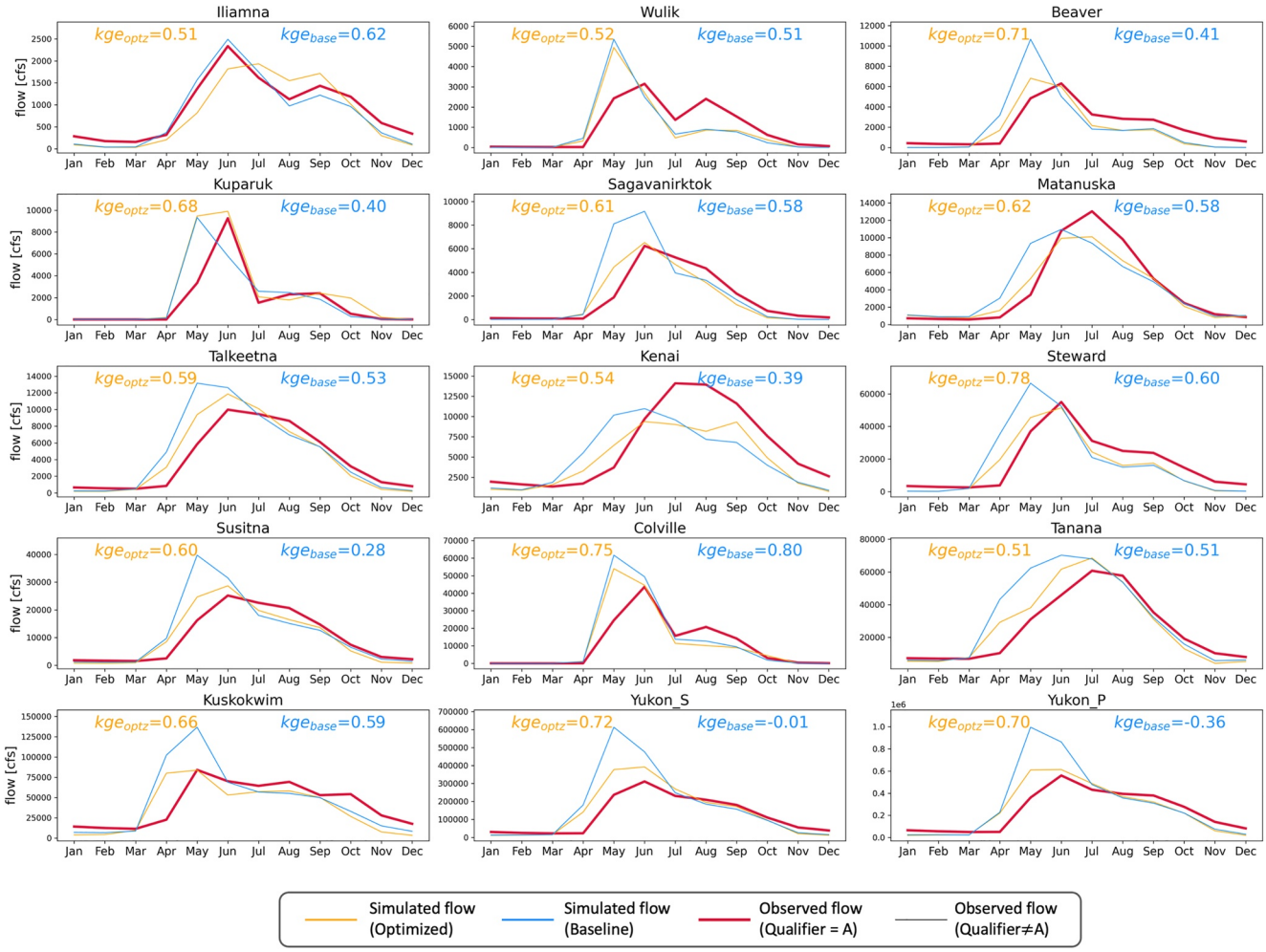

Figure 7. Model evaluation for mean monthly flow time series (from 1991 to 2020 water year). Kling-Gupta Efficiency is based on qualified observed flow at a daily time step.

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

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

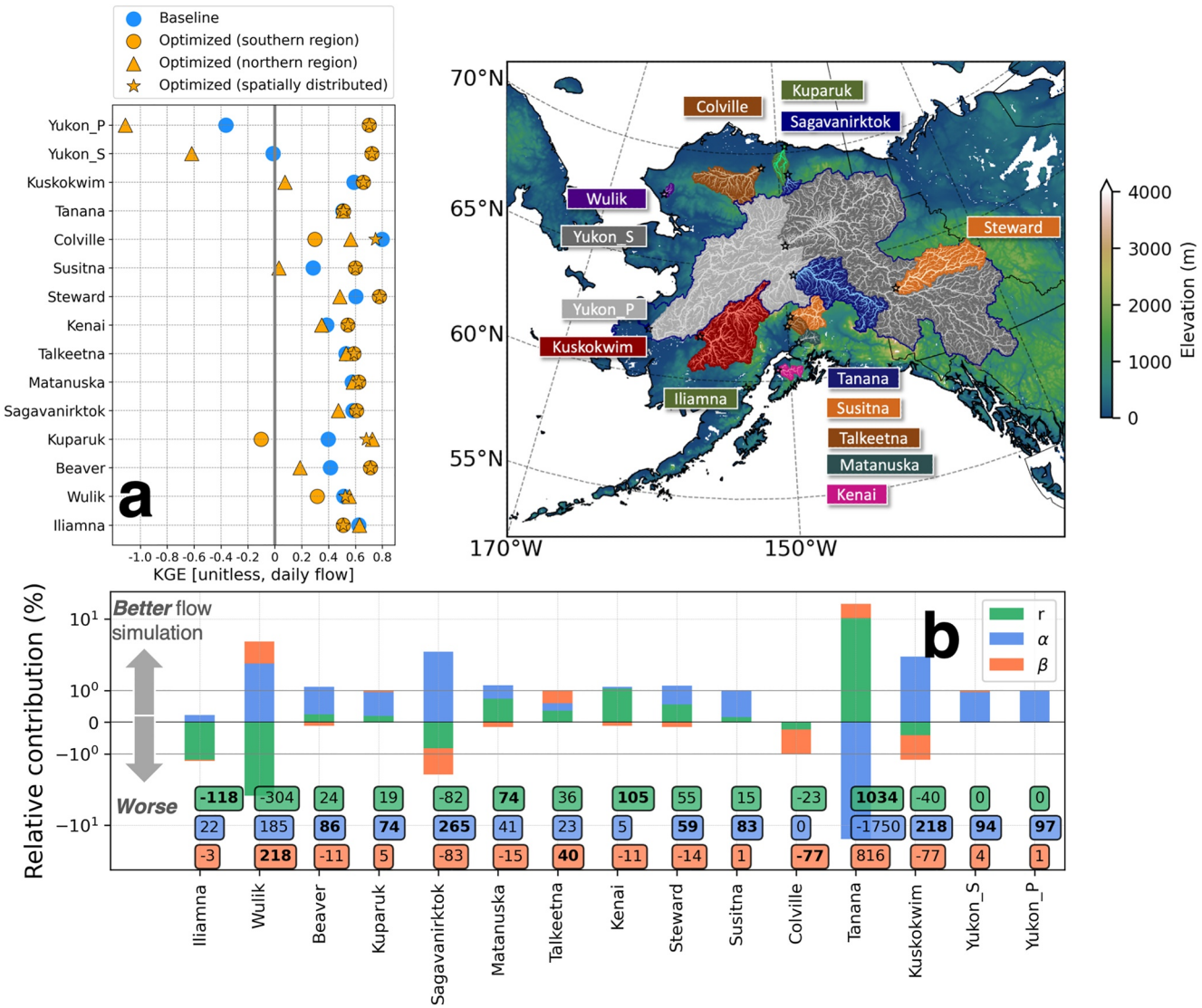

Figure 8. A 30-year evaluation from 1991 to 2020 water year (a) default model performance (blue dot), and model performance using southern optimized parameters (orange dot), northern optimized parameters (orange triangles), and spatially distributed parameters by parameter regionalization (orange stars)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

5.3. Evaluation of CTSM Actionability

As discussed in Section 5.2, the optimized model has improved out of sample overall flow simulation. Specifically using NSE, our actionability flow metric, the mean NSE of daily flow increases from 0.09 to 0.45 after optimization (Table 2). Nine out of 15 basins meet the 0.5 daily NSE benchmark of flow performance after optimization while only two met the benchmark before optimization. In addition, 13 out of 15 basins have improved high flow (a proxy for flooding) magnitudes after optimization. The mean %BiasFHV across all basins is 0% and 19.5% for optimized and baseline simulations respectively. However, the dry bias increases for low-flow events. After optimization, the mean $\%BiasFLV_{summer}$ across all basins decreases from −108.0% to −136.2% and only 7 out of 15 basins show better performance in capturing low flow events (Table 2).

Streamflow climate sensitivities for large river basins are improved while climate sensitivities for small river basins are slightly degraded from the baseline simulation. Subplots in Figure 11 are organized based on the confluence area, with the smallest basin on the top left and the largest basin on the bottom right. When optimization improves the sensitivity to precipitation change, we marked "**P**" in the upper right corner. Similarly, we

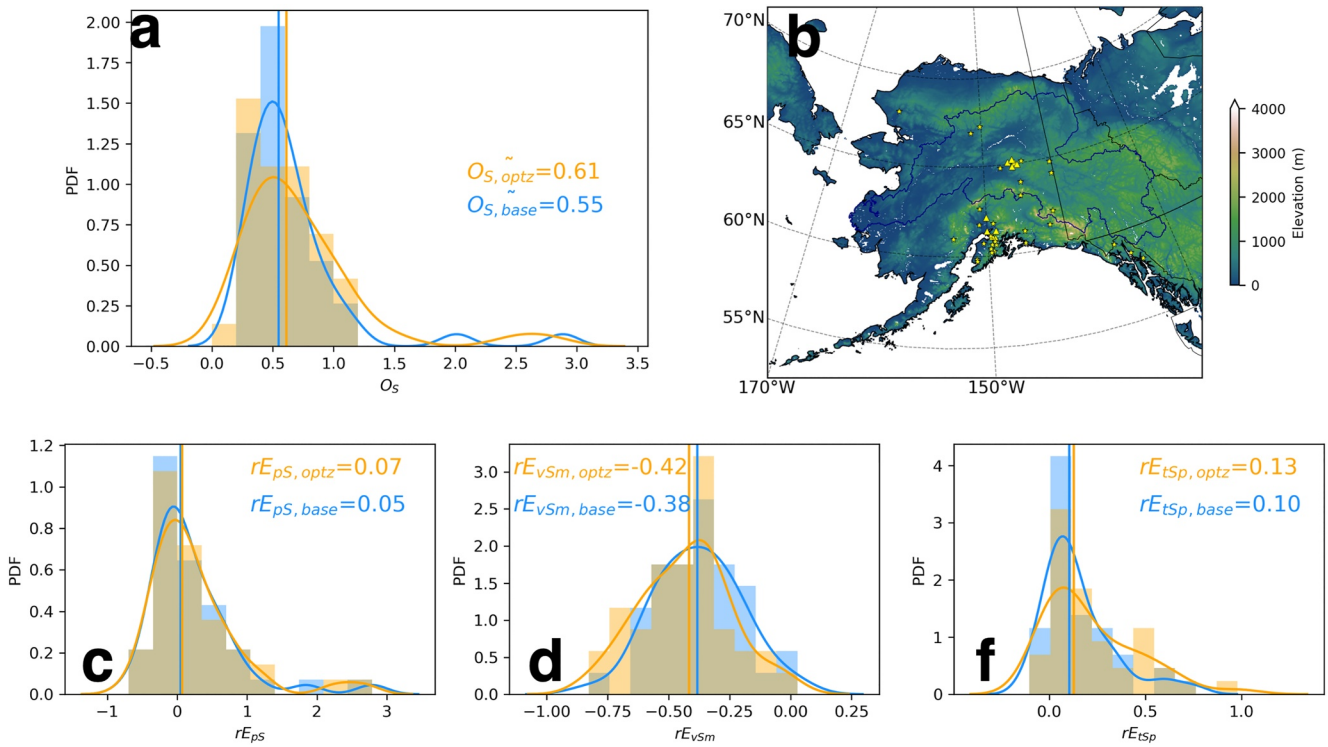

Figure 9. A 30-year evaluation of model performance on snow simulations. Panel (a) shows the distribution of aggregated snow metrics (O_S) across all SNOTEL sites in Alaska, 10 of them used in optimization are highlighted in yellow triangles and the remaining 30 are highlighted in yellow stars in Panel (b). Panels (c, d, f) summarize the distribution of individual snow metrics, including relative errors in annual peak SWE (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 parameter values and blue corresponds with model runs using default parameter values.

marked "**T**" when optimization improves the sensitivity to air temperature change. For the eight largest basins, at least one climate sensitivity is improved and both climate sensitivities are improved in four of them. However, the climate sensitivity for smaller river basins is generally slightly worse in the optimized simulation (Figure 11).

Interestingly, this implies that a better flow simulation does not necessarily lead to a better streamflow climate sensitivity. For example, the KGE for Kuparuk increases from 0.40 to 0.68 after optimization (Figure 8a) while both climate sensitivities are worse in the optimized simulation (Figure 11). Additionally, a worse flow simulation does not necessarily lead to worse climate sensitivity. For example, Colville has slightly lower daily streamflow performance after optimization with KGE decreasing from 0.80 to 0.75 (Figure 8b) but its climate sensitivities to precipitation and air temperature are both improved (Figure 11). This feature may be related to the fact that our optimization metrics are focused on daily flow and daily snowpack, leaving other components of the water budget that could impact climate sensitivities less constrained. It has been shown that metrics not closely related to the optimization metrics result in larger variations in optimized model performance (Mendoza, Clark, Mizukami, et al., 2015). This suggests that metrics more directly related to climate sensitivity should be included in optimization objective functions.

5.4. Shapley Decomposition

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

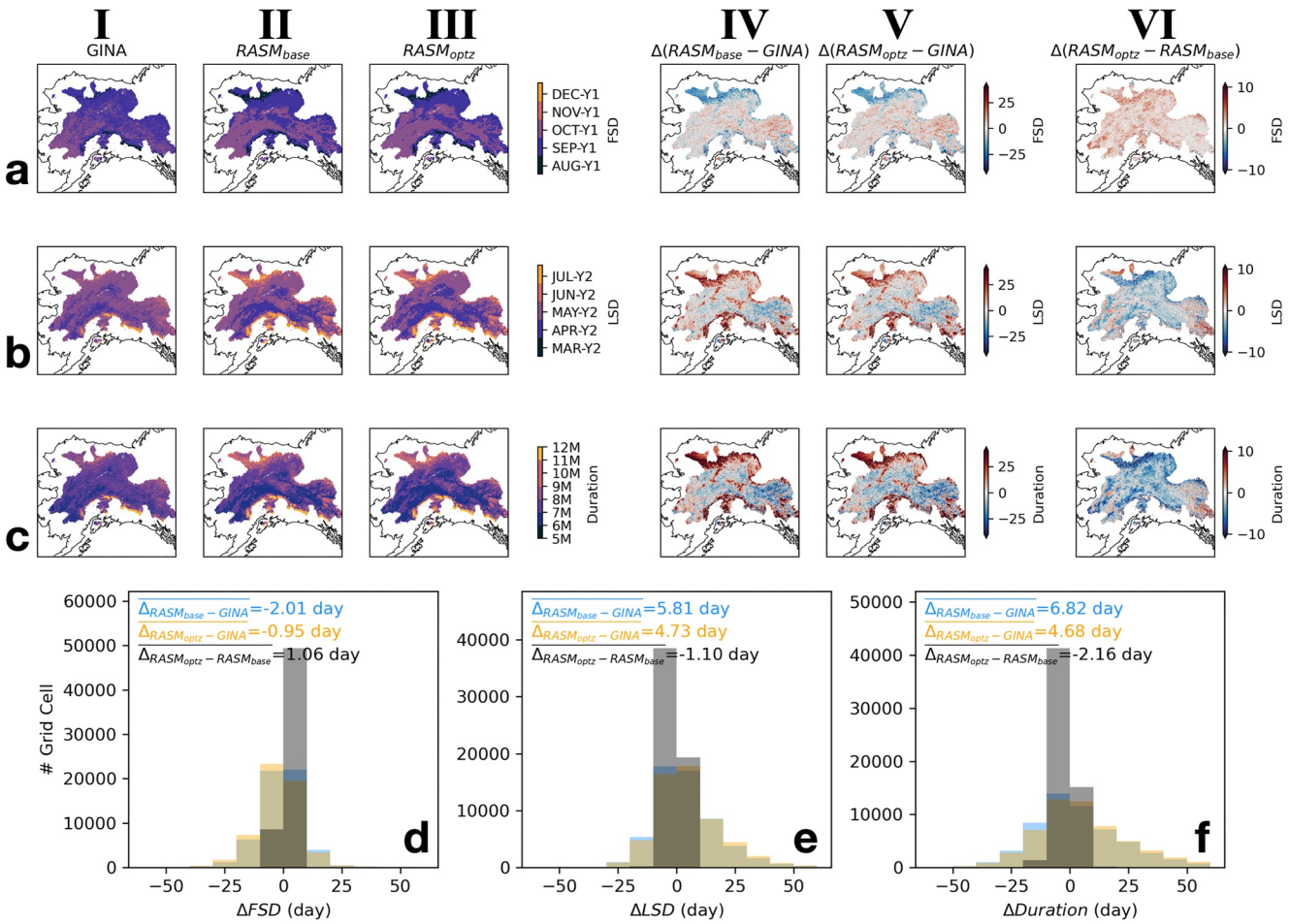

Figure 10. Evaluation of the first snow date (FSD, row (a), last snow date (LSD, row (b), and snow duration (row c) using the Geographic Information Network of Alaska (GINA) data set (2001–2020). Columns *I, II, III* denote the GINA data set, baseline simulation, and optimized simulations, columns *IV, V* denote the biases in baseline and optimized simulations compared with the GINA data set, and column *VI* denotes the difference between optimized and baseline simulations. Figures (d–f) show the histogram of biases across all grid cells for FSD, LSD, and duration, respectively.

In the southern region, the variation of parameter contributions is much smaller than that in the northern region. The perturbation of *n_melt_coef* contributes the most to the flow improvement in the southern region, which only accounts for 30% of the KGE increment (Figure 12). Other than *n_melt_coef*, the top 6 parameters that contribute to the improved flow simulation include *frac_sat_soil_dsl_init, liq_canopy_storage_scalar, upplim_destruct_metamorph, om_frac_sf, e_ice*. In addition, the same parameter perturbation leads to opposite contributions in our flow and snow simulations. The perturbation of *upplim_destrct_metamorph, frac_sat_soil_dsl_init, and liq_canopy_storage_scalar* improves flow simulation while degrading snow simulation, while the perturbation of *d_max and vcmaxha* worsens flow simulation while improving snow simulation.

6. Discussion and Conclusions

We have developed the first high-resolution application and optimization of CTSM for Arctic hydrology. 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 on model parameter choices (Günther et al., 2019; Raleigh et al., 2015). The mean NSE of daily flow increases from 0.09 to 0.45 across 15 river basins for a 30-year evaluation. 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 −0.55 to 0.50. 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

Table 2
Nash-Sutcliffe Efficiency, Percent Bias in Flow Duration Curve (FDC) High-Segment Volume (%BiasFHV), and Percent Bias in FDC Low-Segment Volume During Summer Season (%BiasFLV$_{summer}$) for All River Basins (a 30-Year Evaluation From 1991 to 2020 Water Year)

		NSE optz [−]	NSE base [−]	%BiasFHV optz [%]	%BiasFHV base [%]	%BiasFLV$_{summer}$ optz [%]	%BiasFLV$_{summer}$ base [%]
Basin name	Nation	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

Note. If models outperform benchmarks, the metric is highlighted in bold font.

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 in this study generally improves the actionability of CTSM as defined using our application specific metrics. The overall daily flow simulation using NSE (or KGE), high flow, and streamflow climate sensitivities for large basins are all improved, but there is a degradation for low flow simulation and climate sensitivities for small basins. Optimizing for KGE has been shown to possibly negatively impact low flow simulation (Althoff & Rodrigues, 2021; Gupta et al., 2009). In addition, low flow events during the warm season are largely affected by subsurface flow regimes and thermal characteristics of frozen soils and permafrost. Sensitive parameters to these physical processes might not be identified in our existing workflow, and CTSM does not include many inter-grid cell lateral flow processes, which is an area of needed future research. In addition, we acknowledge that the model results have not yet been used to inform decision-making but this will be further investigated as part of the Arctic Rivers Project.

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

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

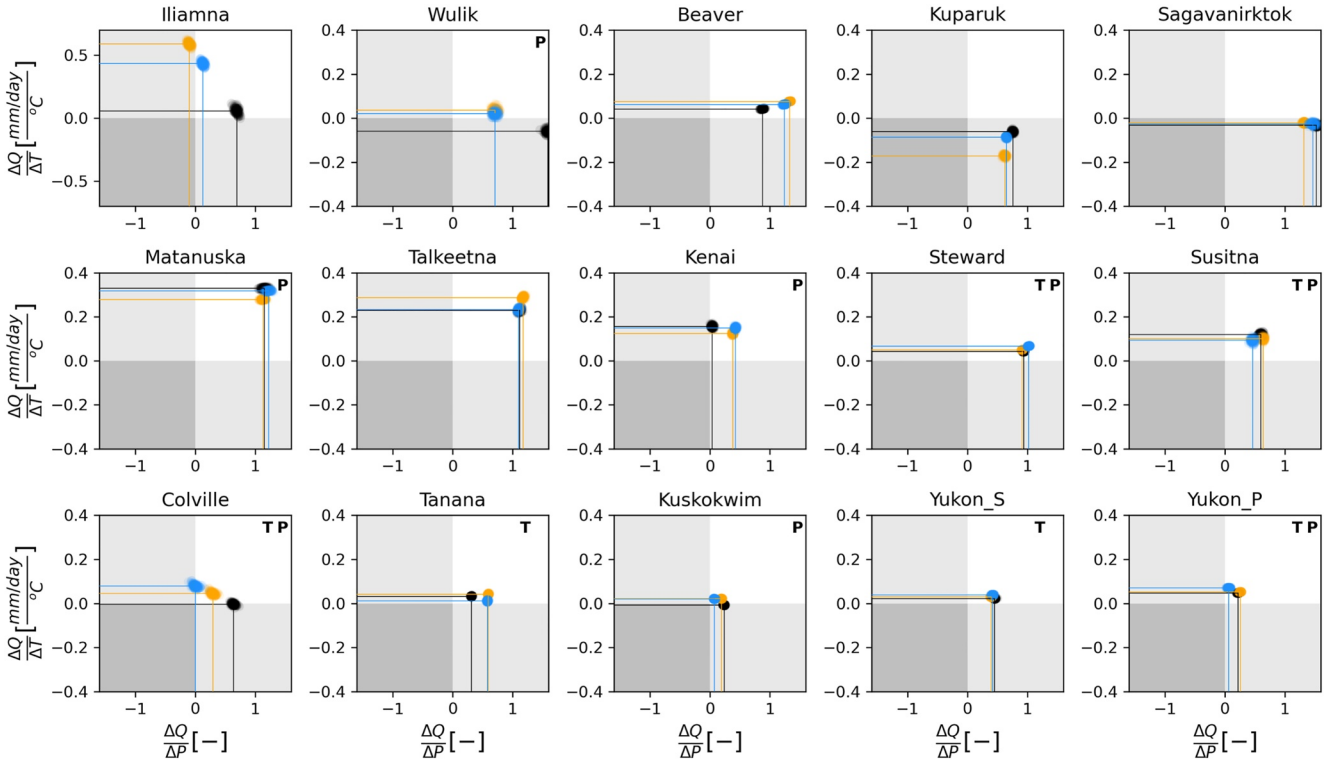

Figure 11. A climate sensitivity analysis. The *x*-axis denotes the rate of basin-averaged discharge change with precipitation change, and the *y*-axis denotes the rate of basin-averaged discharge change with air temperature change. Black, orange, and blue denote observation, optimized simulation, and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

availability of the data sets used in this study, that is, the ERA5 meteorological forcing data (European Centre for Medium-Range Weather Forecasts, 2019), soil texture from SoilGrids (Hengl et al., 2017), MERIT Hydro vector-based river network (Yamazaki et al., 2019), and especially the CTSM PPE global parameter sensitivity analysis. There also exist high-resolution WRF runs available for CONUS (Liu et al., 2017; Rasmussen & Liu, 2017) and ongoing efforts for global coverage, which can be used for downscaling meteorological forcing data outside Alaska. Correctly selecting sensitive parameters provides the foundation for the success of parameter optimization. The PPE experiment can be extensively used in selecting sensitive CTSM parameters not only to runoff and SWE but to any other variable simulated by CTSM. Finally, Arctic Alaska and the Yukon is one of the most challenging regions in hydrological modeling due to the complicated land surface processes that are important in this region. The improvements in hydrological simulation achieved by our optimization framework in the challenging Arctic environment may motivate the exploration of the effectiveness of this optimization framework for regions outside the Arctic.

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

The Shapley decomposition analysis showed the different parameter contributions across regions, reflecting the spatial heterogeneity of parameter sensitivities. The heterogeneity manifests primarily in two ways. First, similar parameter perturbations lead to the opposite direction of effects across regions. For example, an increase in activation energy for $V_{c,max}$ in photosynthesis *(vcmaxha)*, that is, from the default value to the upper limit, contributes 2.0% of flow KGE increment in the northern region but −8.2% in the south. Second, the opposite parameter

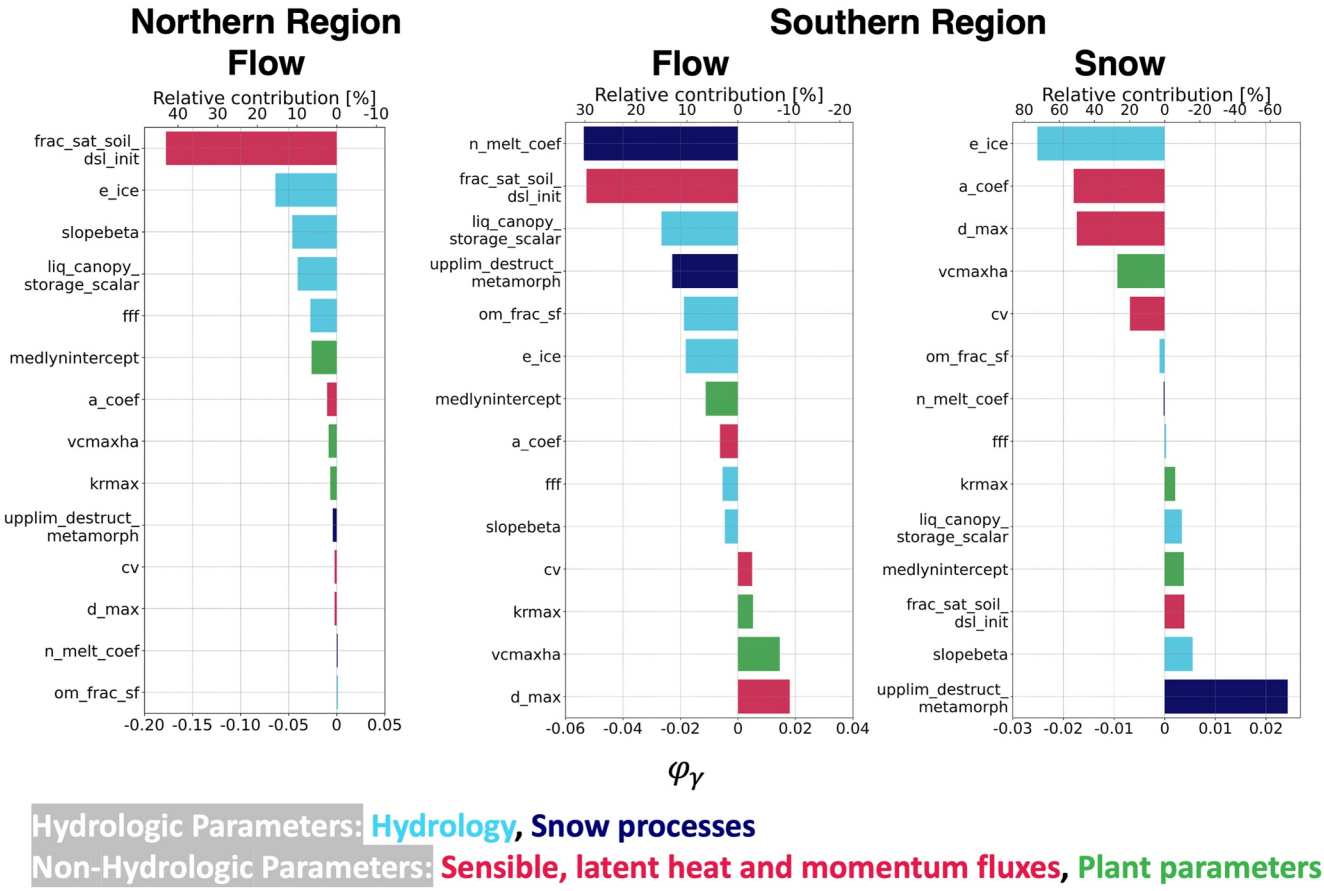

Figure 12. Contribution of each parameter to the changes in objective function using Shapley decomposition.

perturbation leads to the same direction of effects. A decrease of *frac_sat_soil_dsl_init* in the southern region, that is, 0.8 to 0.25, and an increase of *frac_sat_soil_dsl_init* in the northern region, that is, 0.8 to 1.78, both contribute positively to their flow KGE increments, with the RC value of 29.7% and 43.0% 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 provides an important example of utilizing spatially distributed parameters, which should be informative to future CTSM development for allowing this feature.

The parameter regionalization in this study is simple and effective, yet can still be improved. Spatially distributed parameters in Colville and Wulik, that is, basins overlapping both southern and northern regions, generated better flow simulations than the parameters optimized for either region. However, for Iliamna, a southern basin, its flow simulation using northern optimized parameters is better than the one using southern optimized parameters, with daily flow KGE of 0.63 and 0.51, respectively (Figure 8a). The similarities between Iliamna 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 Tanana and Matanuska (Figures 7 and 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 RMSE of the simulated flow and snow objectives are 0.04 and 0.03 respectively, and the RMSE of the simulated flow objective is 0.09 for the northern region. In addition, the Shapley decomposition analysis is based upon the surrogate model, so the contribution of each parameter perturbation reflects the simulated response surface. However, it is infeasible to disentangle each parameter's contribution without a surrogate model. We would need to run CTSM 16,384

(2^{14}) times for the Shapley decomposition while in this study we only ran CTSM 500 times. Additional benefits from using surrogate models might arise by incorporating other observational constraints, for example, Active Layer Thickness, snow depth, or evapotranspiration.

Finally, the development of the optimization framework and example application specific evaluation lowers the barrier of applying complex land models in regional applications and therefore enhances the actionability of the robust scientific tools. The authors hope this work lays the foundation for a process-focused, stakeholder-useful, high-resolution coupled land and atmospheric modeling for cold regions both historically and under future projections to quantify climate change impacts on inland freshwater systems. The authors cannot highlight enough the necessity to include end-users to make science fully actionable and call for more end-user involvement across the range of model development and global, regional, and basin studies using CTSM and other models to move to even higher levels of actionability.

Data Availability Statement

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). For the full CTSM simulation results, please contact Yifan Cheng. The following sources were used to obtain the historical data sets used for evaluation in this research: discharge data for the US State of Alaska from the United States Geological Survey (USGS) National Water Information System, available at https://waterdata.usgs.gov/nwis; discharge data for the upstream Yukon River Basin from the Environment and Natural Resources in Canada, available at https://wateroffice.ec.gc.ca/search/historical_e.html; snow water equivalent data for the SNOTEL sites in the US State of Alaska from the USDA Natural Resources Conservation Service, accessed through a Python software package, *ulmo*, available at https://ulmo.readthedocs.io/en/latest/; satellite derived snow data from the Geographic Information Network of Alaska (GINA) data sets, available at https://gina.alaska.edu/. The high-resolution soil texture and organic matter fraction data are from the *SoilGrids* global gridded soil information, accessed through a Python software package, *WebCoverageService* in *owslib*, and the instructions for downloading SoilGrids data is available at https://www.isric.org/web-coverage-services-wcs. The ecohydrology region classification level III for the US State of Alaska is available at EPA website (https://www.epa.gov/eco-research/ecoregion-download-files-state-region-10#pane-01).

Acknowledgments
This project was funded by NSF Navigating the New Arctic Grant 1928078 and supported by the National Center for Atmospheric Research, which is a major facility sponsored by the National Science Foundation under Cooperative Agreement No. 1852977. We would like to acknowledge high-performance computing support from Cheyenne (https://doi.org/10.5065/D6RX99HX) provided by NCAR's Computational and Information Systems Laboratory, sponsored by the National Science Foundation. We thank our Indigenous Advisory Council and numerous Tribal and First Nation decision-makers for providing important insights to help inform the CTSM configuration design. We also thank the Arctic Rivers Summit investigator team including Nicole Herman-Mercer, Karen Cozzetto, and the Institute for Tribal Environmental Professionals for facilitating the Indigenous Advisory Council and survey efforts. We thank Ryan Toohey for 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).

References

Althoff, D., & Rodrigues, L. N. (2021). Goodness-of-fit criteria for hydrological models: Model calibration and performance assessment. *Journal of Hydrology*, *600*, 126674. https://doi.org/10.1016/J.JHYDROL.2021.126674

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., et al. (2015). Hyper-resolution global hydrological modelling: What is next? *Hydrological Processes*, *29*(2), 310–320. https://doi.org/10.1002/HYP.10391

Brunner, M. I., Melsen, L. A., Newman, A. J., Wood, A. W., & Clark, M. P. (2020). Future streamflow regime changes in the United States: Assessment using functional classification. *Hydrology and Earth System Sciences*, *24*(8), 3951–3966. https://doi.org/10.5194/HESS-24-3951-2020

Cheng, Y., Nijssen, B., Holtgrieve, G. W., & Olden, J. D. (2022). Modeling the freshwater ecological response to changes in flow and thermal regimes influenced by reservoir dynamics. *Journal of Hydrology*, *608*, 127591. https://doi.org/10.1016/j.jhydrol.2022.127591

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., et al. (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, 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., et al. (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., et al. (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*, *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., et al. (2021). Ocean, cryosphere and sea level change. In V. Masson-Delmotte, P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, et al. (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., et al. (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., et al. (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., de Jesus, J. M., Heuvelink, G. B. M., Gonzalez, M. R., Kilibarda, M., Blagotić, A., et al. (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. Retrieved from 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), 10. 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

Jafarov, E. E., Marchenko, S. S., & Romanovsky, V. E. (2012). Numerical modeling of permafrost dynamics in Alaska using a high spatial resolution dataset. *The Cryosphere*, *6*(3), 613–624. https://doi.org/10.5194/TC-6-613-2012

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., den Bossche, J. V., Fleischmann, M., Wasserman, J., McBride, J., Gerard, J., et al. (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), L02503. https://doi.org/10.1029/2005GL024960

Knoll, L. B., Sharma, S., Denfeld, B. A., Flaim, G., Hori, Y., Magnuson, J. J., et al. (2019). Consequences of lake and river ice loss on cultural ecosystem services. *Limnology and Oceanography Letters*, *4*(5), 119–131. https://doi.org/10.1002/LOL2.10116

Lawrence, D. M., Fisher, R. A., Koven, C. D., Oleson, K. W., Swenson, S. C., Bonan, G., et al. (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*, *113*(F2), F02011. 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*, *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., et al. (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

Liu, Y., Rafieeinasab, A., Feng, X., Wu, W., Cosgrove, B. A., Dugger, A. L., & Gochis, D. (2020). Toward improving parameter regionalization for the national water model using the CAMELS data set. In *AGU Fall Meeting Abstracts* (Vol. 2020, pp. H111–0005).

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., et al. (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., et al. (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., et al. (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., et al. (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

Moriasi, D. N., Gitau, M. W., Pai, N., & Daggupati, P. (2015). Hydrologic and water quality models: Performance measures and evaluation criteria. *Transactions of the ASABE*, *58*(6), 1763–1785. https://doi.org/10.13031/TRANS.58.10715

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

Ohlberger, J., Buehrens, T. W., Brenkman, S. J., Crain, P., Quinn, T. P., & Hilborn, R. (2018). Effects of past and projected river discharge variability on freshwater production in an anadromous fish. *Freshwater Biology*, *63*(4), 331–340. https://doi.org/10.1111/FWB.13070

Oleson, K. W., Lawrence, D. M., Gordon, B., Flanner, M. G., Kluzek, E., Peter, J., et al. (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*(1), 17–37. Retrieved from 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., et al. (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., et al. (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. (2017). *High resolution WRF simulations of the current and future climate of North America*. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory. https://doi.org/10.5065/D6V40SXP

Rasmussen, R., Liu, C., Ikeda, K., Gochis, D., Yates, D., Chen, F., et al. (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

Sharma, S., Blagrave, K., Magnuson, J. J., O'Reilly, C. M., Oliver, S., Batt, R. D., et al. (2019). Widespread loss of lake ice around the Northern Hemisphere in a warming world. *Nature Climate Change*, *9*(3), 227–231. https://doi.org/10.1038/s41558-018-0393-5

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*, *107*(D10), 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 & 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

Wood, A. W., Leung, L. R., Sridhar, V., & Lettenmaier, D. P. (2004). Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs. *Climatic Change*, *62*(1), 189–216. https://doi.org/10.1023/B:CLIM.0000013685.99609.9E

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.

Yilmaz, K. K., Gupta, H. V., & Wagener, T. (2008). A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model. *Water Resources Research*, *44*(9), 9417. https://doi.org/10.1029/2007WR006716

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*, *57*(12), e2021WR029879. https://doi.org/10.1029/2021WR029879