Title: Ecohydrological modeling in deciduous boreal forest: Model evaluation for application in non-stationary climates

Authors: Adrienne M. Marshall[1*], Timothy E. Link[1], Gerald N. Flerchinger[2], Dmitry J. Nicolsky[3], Melissa S. Lucash[4]

1. Department of Forest, Rangeland, and Fire Sciences, University of Idaho, 875 Perimeter Drive, Moscow, ID 83844, USA
2. USDA Agricultural Research Service Northwest Watershed Research Center, 251 E. Front Street, Suite 400, Boise, ID 83702, USA
3. Geophysical Institute, University of Alaska Fairbanks, 2156 Koyukuk Drive, Fairbanks, AK 99775, USA
4. Department of Geography, University of Oregon, 1251 University Oregon, Eugene, OR 97403, USA

* Corresponding author contact: adriennemarshall@uidaho.edu

Keywords:
Ecohydrology, boreal, soil moisture, soil temperature, Generalized Likelihood Uncertainty Estimation, global uncertainty analysis, hydrologic model calibration, model diagnostics

Abstract

Soil moisture is an important driver of growth in boreal Alaska, but estimating soil hydraulic parameters can be challenging in this data-sparse region. To better identify soil hydraulic parameters and quantify energy and water balance and soil moisture dynamics, we applied the physically-based, one-dimensional ecohydrological Simultaneous Heat and Water (SHAW) model, loosely coupled with the Geophysical Institute of Permafrost Laboratory (GIPL) model, to an upland deciduous forest stand in interior Alaska over a 13-year period. Using a Generalized Likelihood Uncertainty Estimation (GLUE) parameterization, SHAW reproduced interannual and vertical spatial variability of soil moisture during a five-year validation period quite well, with root mean squared error (RMSE) of volumetric water content at 0.5 m as low as 0.020. Many parameter sets reproduced reasonable soil moisture dynamics, suggesting considerable equifinality. Model performance generally declined in the eight-year validation period, indicating some overfitting and demonstrating the importance of interannual variability in model evaluation. We compared the performance of parameter sets selected based on traditional performance measures (RMSE) that minimize error in soil moisture simulation, with those that were designed to minimize the dependence of model performance on interannual climate variability. The latter case moderately decreases traditional model performance but is likely more suitable for climate change applications, for which it is important that model error is independent from climate variability. These findings illustrate (1) that the SHAW model, coupled with GIPL, can adequately simulate soil moisture dynamics in this boreal deciduous region, (2) the importance of interannual variability in model parameterization, and (3) a novel objective function for parameter selection to improve applicability in non-stationary climates.

1. Introduction

1.1 Soil moisture in deciduous boreal forests

Boreal regions are changing rapidly, with alterations to permafrost extent and active layer thickness (Meredith et al., 2019), nonlinear changes in forest composition and productivity (Beck et al., 2011; Scheffer et al., 2012), increased fire frequency and severity (Kasischke et al., 2010), and projected rapid changes to soil moisture dynamics following permafrost thaw (Teufel and Sushama, 2019). In boreal North America, we do not know whether future conditions will be drier or wetter: while warming is clearly established, changes in precipitation are generally increasing but relatively uncertain and will likely be as important as temperature for future wetting/drying trends (Fischer et al., 2017; Meredith et al., 2019). Increasing precipitation may increase soil moisture. However, decreases in snow and permafrost and increases in wildfire frequency and intensity due to warming may lead to soil drying (Meredith et al., 2019). This underscores the lack of understanding of regional variability in wetting and drying, and exemplifies why IPCC denotes soil moisture as a major source of uncertainty in this region (Meredith et al., 2019).

Soil moisture is particularly important in upland deciduous forests. Experimental manipulations and dendrochronological studies identify moisture limitation as an important control on summer productivity for many upland forest types in interior Alaska (Cahoon et al., 2018; Yarie & van Cleve, 2010). Early snowmelt in northern latitudes is negatively correlated with vegetation greenness, suggesting that productivity may be limited by summer soil moisture, even at relatively high latitudes in interior Alaska (Barichivich et al., 2014). These deciduous forest types are projected to become increasingly prevalent as wildfire frequency increases (Hansen et al., 2020; Johnstone et al., 2010; Mann et al., 2012) and therefore a quantitative understanding of energy and water balances in these forest types is critical to understanding consequences of climate change at high latitudes.

Physically-based ecohydrological models are an essential tool for understanding and predicting the interacting effects of changing vegetation and climate on soil moisture dynamics. These models have a rich history in interior Alaska: for example, Bonan (1989) constructed a physically-based model including climate, soil moisture and temperature, and forest fires at a monthly timescale. Other approaches have modeled climate and vegetation controls on energy exchanges in boreal forests, but without a focus on soil moisture (Baldocchi et al., 2000). The Geophysical Institute Permafrost Laboratory (GIPL) model has been used for investigations of feedbacks between vegetation, snow, and permafrost dynamics over large scales, though it assumes saturated soils (Jorgenson et al., 2010). Most hydrological studies have generally emphasized streamflow, rather than soil moisture dynamics, as the process of interest (e.g., Endalamaw et al., 2017). Despite these relevant efforts, the magnitude of projected changes in soil moisture in Alaska varies widely across models; most project drying due to deep drainage following permafrost thaw, though a few project wetting trends (Andresen et al., 2020). There is therefore an ongoing need to improve simulations of soil moisture dynamics in deciduous upland forests of interior Alaska via careful parameterization and uncertainty quantification of ecohydrologic models. Long-term meteorologic, hydrologic, and ecological data are relatively sparse in this remote environment, so model evaluation is both critical and challenging.

1.2.GLUE parameter estimation and uncertainty analysis

Hydrologic models can be calibrated with manual or automated methods (e.g., Flerchinger et al., 2012). Automated methods have gained popularity in recent decades due to their reproducibility across multiple model applications, ability to minimize error, potential for identification of equifinality, and relative independence from an individual modeller's understanding of parameter values, importance, and interactions (Liu & Gupta, 2007; Pianosi et al., 2016). Generalized Likelihood Uncertainty Estimation (GLUE) methods are a popular Bayesian approach for parameter estimation that were designed to address many of these issues (Beven & Binley, 1992; 2014). GLUE methods require specification of a prior probability distribution function for each parameter. Many parameter sets are then defined by sampling from the prior distribution. A likelihood or objective function that determines how well model outputs fit calibration data is selected. The modeler defines a lower model performance limit that determines whether a particular parameter set is considered acceptable, or "behavioral" in the GLUE parlance. The model is run with each parameter set, and those that produce acceptable results are used to define uncertainty bounds of predictions. These methods have gained broad acceptance for their relative ease of use, ability to delineate (often informal) uncertainty bounds, Bayesian capacity to integrate prior knowledge about parameter distributions with new data, and consistency with the idea of equifinality (Beven, 2006).

Equifinality may be theoretically appropriate in many cases, but in practice, one optimal parameter set is often needed. Pareto optimization is commonly used to identify the set of solutions that are each not dominated by any other solution with respect to the objective functions used. Pareto optimization is particularly attractive relative to index-based approaches because it does not require any arbitrary weighting decision for the relative importance of different objective functions. As the number of objectives included in a Pareto frontier increases, the set of Pareto-optimal solutions grows rapidly. Selecting a solution from a large set of Pareto-optimal solutions can be accomplished via preference selection (Das, 1999; Khu & Madsen, 2005). Briefly, when m objective functions are considered, the ideal solutions are included in the Pareto front for all possible combinations of m-1 objective functions. Failing the existence of such solutions, individual solutions can be defined as efficient in order k with degree p, where k is the number of objective functions included ($1 \leq k \leq m$), and p is the number of k-dimensional subspaces in which the given solution set is Pareto-optimal. These sets are then denoted as $[k, p]$.

In hydrologic modeling, the objective functions used for determination of GLUE behavioral parameters, identifying Pareto-optimal solutions, and preference ordering can incorporate (1) multi-site data, (2) multi-variable data, such as multiple states or fluxes, and (3) multi-response models, such as model efficiency at high and low flows (Efstratiadis & Koutsoyiannis, 2010). Here, we use both multi-variable data and multi-response models to identify optimal parameter sets for soil moisture and temperature simulation, following identification of behavioral parameter sets using a single response and variable. Our multi-response models are designed to address the commonly identified problem of model overfitting. For example, Coron et al., (2014) found greatly diminished model performance in a validation relative to calibration period, identifying this as a major source of uncertainty for non-stationary climates. Here, we propose that interannual variability in model performance should be an important criterion for assessing parameter suitability for non-stationary climates. Specifically, the sensitivity of model error to interannual variations in temperature and precipitation can be used in conjunction with traditional model performance metrics such as root mean squared error to identify optimal parameter sets that are more appropriate for future climate projections than those identified with traditional performance measures alone.

1.3 Study objectives

Here, we implement a physically-based ecohydrological model, loosely coupled with a permafrost model, to simulate soil moisture dynamics in a deciduous boreal forest. Specifically, we address the following research questions: (1) how well can SHAW-GIPL reproduce soil moisture dynamics at multiple depths on daily-to-interannual timescales? (2) what is the relationship between model performance in the calibration and validation periods? and (3) can parameter sets be selected to minimize model error dependence on interannual climate variability? These findings will benefit future studies of how climate change might affect boreal forests. Importantly, we introduce a novel objective function designed to optimize model suitability for future climate scenarios.

2. Methods

2.1 Study site

In this study, we used data from the Bonanza Creek Long Term Ecological Research site (BNZ LTER). This upland site (UP1A, 64.7355 °N, 148.3027 °W) in the discontinuous permafrost zone was historically dominated by white spruce, but burned in 1983. By 1997, the site was classified as a paper birch (*Betula papyrifera*)/shrub/herbaceous woodland (van Cleve et al., 2015). The site has no near-surface permafrost and is underlain by silt loam. Annual average air temperature was -1.1 °C and average annual precipitation was 337 mm over the years included in this study (2003-2015).

2.2 Meteorological data

Hydrometeorological data were obtained from the BNZ LTER database for UP1A and the primary upland weather station for BNZ LTER, located approximately 1 km to the north and 90 m higher than the study site. We obtained hourly air temperature, relative humidity, and precipitation from the UP1A site; wind speed was available from LTER1 only (van Cleve et al., 2017a, b; 2018a, b). Shortwave radiation data was obtained from the US-Uaf AmeriFlux site at the University of Alaska campus in Fairbanks, AK (Ueyama et al., 2002), located approximately 25 km northeast of the UP1A site. Precipitation data from UP1A was located under the forest canopy, so data from LTER1 were used instead.

Over 2003-2015, 0.65% of air temperature, 6.1% of wind speed, 0.65% of relative humidity, 0.03% of shortwave radiation, and 31% of precipitation data were missing. For air temperature, relative humidity, wind speed, and shortwave radiation, missing values for gaps ≤4 hours were first infilled with a linear gap-filling procedure. For longer gaps, linear regressions were performed between UP1A, LTER1, and three other nearby upland stations with meteorological data (UP2A-UP4A), using data from each station in turn based on correlation coefficients between each station and UP1A. For variables with a zero-limited minimum (precipitation, relative humidity, windspeed), the regressions were constrained at zero. For air temperature, a monthly interaction term was included to account for seasonally varying lapse rates. Precipitation data were gap-filled from daily data collected at the Fairbanks airport (Menne et al., 2012; 23 km away). We calculated that on wet days at LTER1, the median number of wet hours is 3. For each day of airport data with precipitation > 0 mm, we selected 3 random hours and distributed the precipitation for that day evenly over those hours. The missing LTER1 data were then directly substituted with data from the airport (Figure 1).

Precipitation undercatch is a known issue, particularly in cold conditions (e.g., Sevruk, 1996). Comparing modeled and observed snow water equivalent is a useful way to validate precipitation data, but observed snow water equivalent was not available at UP1A. Instead, we used snow depth data from LTER1, assuming that SHAW's snow density simulations are reasonably accurate. Using weather inputs for the open LTER1 site, we ran the SHAW model without vegetation and compared modeled and observed snow depth at that site (Figure 2; van Cleve et al., 2018c). Snow depth RMSE for 2004-2009 was 10.6 cm, Nash-Sutcliffe efficiency was 0.34, and mean bias was 8.9 cm. We considered this fit reasonable given the length of the time series, missing precipitation data, and uncertainty in modeled density.

2.3 Vegetation data

Vegetation has been regularly surveyed since 1990 at UP1A, with diameter at breast height recorded for each individual tree (van Cleve et al., 2015). We used the allometric equations from Yarie et al. (2007) to estimate height and biomass for each individual tree, then calculated the average of individual heights and sum of biomass for each species and sampling date. For years in which vegetation survey data were not available, we linearly interpolated height and biomass if the missing data was within the total time domain of sampling for that species, and filled in with constant values for years modeled before or after the first or last vegetation surveys. The results of this analysis indicated that paper birch was by far the dominant tree species, so we modeled only paper birch. As the canopy was still developing post-fire during the period modeled, there was relatively limited understory vegetation.

Leaf phenology can affect model results via impacts on transpiration and precipitation interception. To develop reasonable estimates of leaf phenology, we conducted each model run twice: in the first run, LAI was set to zero from day 274 (October 1) to day 91 (April 1). In the spring, LAI was linearly increased to its maximum value between day 91 and day 121 (May 1), and linearly decreased between day 274 and day 305 (November 1). The date of snowmelt was recorded based on this initial run, and the model was rerun, using snow disappearance date from the prior run to estimate leaf-on timing. We based this on a generalized additive model of annual leaf-on timing at nearby Chena Ridge as a function of snow disappearance date at a snow pillow at LTER 1 ($R^2 = 0.84$, N = 21, Figure 3b; Anderson et al., 2020).

Several parameters for deciduous hardwood species in interior Alaska were obtained based on literature values. Minimum stomatal resistance was estimated as 150 s/m (Endalamaw et al., 2017). Critical leaf water potential at which stomatal resistance doubles was estimated as -173 m (Federer, 1977). Root and leaf resistance were estimated as 66×10^4 and 33×10^4 $m^3 s/kg$ (Flerchinger, 2017). Rooting depth was set to 0.6 m (Safford et al., 1990). The characteristic dimension for birch leaves was set to 3.6 cm, based on the assumption that characteristic dimension is equal to 72% of the width of two intersecting parabolas (Campbell & Norman, 1998). Maximum LAI was estimated as 2.8 m^2/m^2 (Bonan, 1991).

2.4 Soil data

Soil texture at UP1A is silty loam throughout the profile, based on four soil pits (Yarie, 1998). We fit generalized additive models to estimate how sand, silt, clay, organic matter, and bulk density varied continuously with depth. Fitted sand, silt, and clay were rescaled to sum to 100%. Saturated volumetric water content (θ_{sat}) for each soil depth was estimated as the maximum observed soil water content over the 13 years included in the study.

Soil water retention curve and hydraulic conductivity parameters were obtained from the National Cooperative Soil Survey Soil Characterization Database for sites throughout interior Alaska (Beaudette et al., 2020; Burt, 2009). To approximately match the soils at UP1A, we selected observations with greater than 60% silt and less than 15% organic matter. This resulted in data from 57 sites, with 158 total observations. The central 95th percentile values of θ_{res}, α, and K_{sat} were used to define parameter ranges in the GLUE analysis. K_{sat} was additionally allowed to vary up to 100 cm/hr, rather than the 17 cm/hr indicated by the NCSS data because modeled K_{sat} is often higher than lab-determined K_{sat}, due to the presence of macropores (e.g., Blain & Milly, 1991; Chappell et al., 1998; Grayson et al., 1992). The pore-connectivity parameter is often set to 0.5 for mineral soils (Mualem et al., 1976); we used the mean value from the NCSS soils database of 0.46.

Initial soil temperatures were set to the average values on the initial day of the simulation, moisture content was set to saturation, and the model was run with one spin-up year to minimize the impacts of these initial conditions. Hourly soil moisture and temperature data were obtained from the BNZ LTER for model calibration (Chapin & Ruess, 2018; vanCleve et al., 2018d). We calculated daily averages for comparison with daily model output.

2.5 SHAW model

The Simultaneous Heat and Water (SHAW) model is a one-dimensional, physically based model that solves the energy and water balance throughout a soil-plant-atmosphere continuum (Flerchinger & Saxton; 1989; Flerchinger & Cooley, 2000; Flerchinger et al., 2016). In this study, we used Version 3.0.2 running at an hourly time step. SHAW was originally developed to simulate soil freezing and thawing processes in saturated and unsaturated soils (Flerchinger & Saxton, 1989). The model simulates energy and water transfer throughout a multi-species plant canopy, multi-layer snowpack, surface litter, and multi-layer soil profile. Precipitation partitioning in SHAW is based on a user-defined air temperature (0 °C in this case). Interception of solid and liquid precipitation is modeled as a function of LAI and estimated maximum precipitation intercepted per unit of LAI. Snow accumulation and ablation are modeled via a complete mass and energy balance. SHAW uses Richards' equation for unsaturated infiltration and here we used the van Genuchten-Mualem model (Mualem, 1976) for the soil water retention function. Saturated hydraulic conductivity and heat capacity are dynamically altered based on liquid water and ice content. SHAW has successfully been applied in cold regions, including coupling with geomorphically based hydrologic models (Zhang et al., 2013), permafrost models (Langford et al., 2019), and ecosystem biogeochemistry models (Wang et al., 2014).

2.6 GIPL model

For the sake of computational efficiency, SHAW simulates a relatively shallow soil profile (6 m in this study) and requires a lower thermal boundary condition. In temperate environments this value is often set to the average annual air temperature (Flerchinger, 2017), but the complex interactions of snow, vegetation, and climate conditions in discontinuous permafrost regions make this an inappropriate assumption in this context (Jorgenson et al., 2010). We therefore used the GIPL 2.0 model (Marchenko et al., 2008; Nicolsky et al., 2009) that solves a nonlinear 1-D heat equation with phase change and has been widely verified in the region (Jafarov et al., 2003, Nicolsky et al., 2017). The GIPL model is calibrated (Nicolsky et al., 2009) using the available ground temperature observations at the UP1A site and is used to compute temperature dynamics in a 70-m deep soil column. Deeper soil layers are used to improve the temperature dynamics at the shallower layers (Alexeev et al., 2007). Temperature at the 6m depth is used as a lower boundary condition in SHAW. Both SHAW and GIPL use the same hydrometeorological and soil data.

2.7 GLUE parameterization and uncertainty analysis

We focused our GLUE analysis on soil hydraulic parameters and the lower boundary condition for temperature because these terms are particularly important for simulating soil moisture and temperature dynamics in cold regions (O'Connor et al., 2020). We defined three soil layers ranging from 0-10 cm (Layer 1), 11-40 cm (Layer 2), and 41 cm-6m (Layer 3). The prior soil parameter distributions were determined based on the NCSS soils data described above and were allowed to vary independently for each layer (Table 1). The lower boundary condition prior was based on results of the GIPL model and the fact that this site is known to be near-surface permafrost-free, and varied to allow for potential errors in the GIPL results. We sampled 50,000 parameter sets from these uniform priors using latin hypercube sampling (Carnell, 2019).

To identify behavioral parameter sets, we required the root mean squared error (RMSE) of volumetric water content at 50 cm (θ_{50}) to be less than 0.03 and Nash-Sutcliffe Efficiency (NSE) greater than 0.5. Values at 50 cm were selected because this is an approximate rooting depth for birch (Burns & Honkala, 1990), and therefore important for plant growth. These fit criteria were calculated over May to September, to avoid periods when soil was likely frozen and soil moisture measurements less reliable (Chandler et al., 2017). Outputs were analyzed at the daily scale to reduce the impact of minor temporal variations. The water years 2003-2007 were used for model calibration, and 2008-2015 for validation.

2.8 Analysis of GLUE results

All analyses were conducted in the R programming language (version 3.6.0; R Core Team, 2019). To determine how narrowly the value of each parameter was constrained by the GLUE method, we tested whether the distribution of each parameter in the accepted parameter sets (the posterior) was significantly different from the prior distribution using two-sided Kolmogorov-Smirnov (K-S) tests (Birnbaum & Tingey, 1951). We also tested the robustness of the GLUE-derived uncertainty bounds by calculating the fraction of observations that fell within the uncertainty bounds. These fit statistics were also calculated on soil temperatures at multiple depths. To determine whether the best-performing parameter sets in the calibration period also performed well in the validation period, we ranked each accepted parameter set and assessed the correlation between the rankings for each period.

2.9 Model diagnostics to reduce overfitting

We developed two novel objective functions based on model performance in warmer or wetter conditions that were used to determine the Pareto frontier. For each parameter set, we used the mgcv package in R to fit a generalized additive mixed model (GAMM) to data from the calibration years as follows:

$$g(\theta\, RMSE_{i,j}) = \alpha_o + \beta_j + f_1(T_{mean,i}) + f_2(P_i) + \epsilon_{i,j}$$

where θ $RMSE_{i,j}$ is the RMSE of θ for the *i*th water year and *j*th depth, *g* indicates a gaussian family of model, α_o is an intercept, β_j is a random effect for depth, f_1 and f_2 are thin plate splines, T_{mean} is mean air temperature, P is total annual precipitation, and $\varepsilon_{i,j}$ is a normally distributed error term (Wood, 2017). We then used the fitted GAMM to estimate the difference in θ_{50} RMSE between a hypothetical year with average temperature and precipitation, and one in which T_{mean} is increased by 2 °C or P is increased by 60%. The change in θ_{50} RMSE estimated by the GAMM based on variability in air temperature and precipitation are hereafter referred to, respectively, as ΔRMSE (T+2) and ΔRMSE (1.6P). These values were selected to reflect the projected direction of change for this region (Lander et al., 2016) while remaining within the observed range of variability of our dataset, relative to mean values.

2.10 Pareto optimization and preference ordering

Pareto optimization and preference ordering were used to identify three optimal parameter sets, using the R package "rPref" (Roocks, 2016) and our novel objective criteria. The first parameter set, called Case 1, was simply designed to optimize model performance against observed soil moisture by selecting the solution set with lowest θ_{50} RMSE. Case 2 was designed to minimize the dependence of model performance on interannual climate variability while maintaining performance in the calibration period. In this case, we identified a four-objective Pareto front based on θ_{50} RMSE, soil temperature RMSE at 1 m, and the absolute values of ΔRMSE (T+2) and ΔRMSE (1.6P). In order to select Case 2 from the solutions included in the four-objective Pareto front, we identified the parameter sets that were present in all four possible combinations of *m*-1 (three) objective functions, following Khu & Madsen (2005). We then selected the solution with the minimum |ΔRMSE (T+2)|. Finally, Case 3 was selected exclusively to minimize the |ΔRMSE (T+2)| and |ΔRMSE (1.6P)| by selecting from the parameter sets in the two-objective Pareto front using these two variables and choosing the solution with minimum |ΔRMSE (T+2)|.

3. Results

3.1 Soil moisture simulations

Of the 50,000 parameter sets tested in the GLUE analysis, 3196 (6.4%) runs met the criteria of RMSE < 0.03 and NSE > 0.5 at 50 cm. Modeled soil moisture generally reproduced the interannual variability and differences between depths in the calibration data (Figure 4). In particular, at the 5 and 10 cm depths, observations were within the range of values modeled by behavioral parameter sets 97-98% of the time, capturing both spring wetting due to snowmelt and soil thaw and summer rainfall events quite well (Table 2). At the 50 cm depth, observations were within the uncertainty range 80% of the time, but were only within this range 59% of the time at the 20 cm depth. Model performance at 20 cm was relatively poor; the model underestimated the spring wetting pulse in most cases and remained too dry throughout the summer. Despite uncertainties in the input precipitation data, the model appears to capture most precipitation events in the shallow layers fairly well. At the deepest layer, the model underestimated the magnitude of the spring moisture pulse in some cases and overestimated it in others. Notably, the summer of 2014 was unusually wet, and the model captured these unusual weather conditions fairly well.

3.2 Soil temperature simulations

The parameter sets that captured soil moisture dynamics well also simulated soil temperature dynamics reasonably well (Figure 5). At 0 cm, there is minimal variability within the behavioral parameter sets, suggesting that near-surface soil temperature is minimally affected by the parameters included in this analysis. At 50 cm and deeper, simulated temperatures generally warm too early in the spring and are generally too cold in the winter. For example, at 1 m, the coolest simulations thaw 34 days too early, on average, whereas the warmest thaw 44 days too early. Model fit statistics indicate that modeled temperature was fairly accurate, particularly in the deeper layers, perhaps due to reduced variability with intraseasonal temperature fluctuations (Table 3). Overall, 59% of observations at 50 cm and 93% at 2 m were within the range of uncertainty bounded by the behavioral parameter sets. Both the observations and simulations indicate apparent warming in the latter part of the temperature record, particularly with respect to winter temperatures. At the 2 m depth, this is captured well, though in the 50-100 cm depths, the model tends to simulate colder winter temperatures than were observed.

3.3 Selected parameter sets

The Pareto fronts used for parameter set selection suggest some tradeoffs between the objective criteria used (Figure 6). On average across all behavioral parameter sets, ΔRMSE (T+2) increased by 0.028 (±0.011 S.D.), and was positive in all cases, and statistically significant ($p < 0.05$) in 16% of cases. In contrast, ΔRMSE (1.6P) was generally negative, suggesting better model performance in wetter conditions; the average value was -0.05 (±0.013 S.D.) and this term was statistically significant in 19% of cases. The three cases selected highlight the tradeoffs between overall performance in the calibration period and sensitivity of performance to interannual variability. Case 1, which minimized θ_{50} RMSE, had moderate RMSE of temperature at 1 m (Figure 6a) and |ΔRMSE (T+2)| (Figure 6b) relative to other behavioral parameter sets, and relatively high |ΔRMSE (1.6P)| (Figure 6c). In contrast, Case 3 had low |ΔRMSE (T+2)| and |ΔRMSE (1.6P)| (Figure 6f) at the expense of relatively high θ_{50} RMSE, though Case 3 RMSE of soil temperature was similar to that of Case 1 (Figure 6a). Case 2, a compromise between these two, had moderate performance on each measure, with the exception of relatively low RMSE of soil temperature at 1 m. While the Pareto optimization used only RMSE of θ_{50} and temperature at 1 m, these tradeoffs are also reflected across depth. Case 1 generally had the smallest discrepancies in soil moisture at all depths, followed by Case 3, with the largest differences in Case 2 (Table 2). In contrast, with respect to soil temperature, Case 2 had the lowest error at all depths except for the surface; Case 1 and Case 3 had very similar errors in soil temperature throughout the soil profile (Table 3). Despite the differences in effect sizes, the effects of air temperature and precipitation on model error were not statistically significant in any of the three cases selected.

Figure 7 illustrates the dependence of model performance on temperature and precipitation in each case. In Case 1, performance appears to depend relatively strongly on temperature, with results suggesting that errors are lower at relatively high and low values of precipitation, though the terms are not statistically significant ($R^2 = 0.25$; $p > 0.05$). In contrast, in Case 3, average θ_{50} RMSE is higher (Table 2) but little dependence on temperature is apparent ($R^2 = 0.01$; $p > 0.05$). Again, Case 2 represents a compromise solution, with some dependence on both air temperature and precipitation evidenced by a moderate R^2, though neither term is significant ($R^2 = 0.17$, $p > 0.05$).

Parameter values identified in each of the three cases are in Table 4. K_{sat} decreases consistently throughout the soil profile in a realistic way in Case 2, but not in Case 1 or Case 3. K_{sat} is of similar order of magnitude across cases in Layer 1, but not in the other two layers. The increase in θ_{res} with depth in Case 2 is also more physically plausible than the decrease in Case 1 or the variability in Case 3. The effects of α depend on parameter interactions with θ_{res}, and the direction of change in α with depth varies across cases.

The prior and posterior parameter distributions indicate the distribution of accepted values for each parameter; the widths of these distributions illustrates the relative importance of each parameter for model performance on the objective criteria we defined (Figure 8). As expected, parameters at the 50 cm depth used for our objective criteria had the narrowest posterior distribution of values in the behavioral parameter sets, as well as those that were Pareto optimal across our four criteria (RMSE of θ_{50} and soil temperature at 1 m, ΔRMSE (T+2), and ΔRMSE (1.6P)). At 50 cm, K_{sat}, α, and θ_{res} all had significantly narrower posteriors than the prior distribution ($p < 0.0001$). All three parameters at the first and second layer, as well as the lower boundary condition, had fairly wide posterior distributions, suggesting a wide range of values may be acceptable or that parameter interactions are more important than individual values. Despite the fact that these distributions were fairly wide, these were statistically significantly different from the prior distributions in all cases except for that of θ_{res} in the first layer ($p < 0.05$). There was no significant difference between the lower boundary condition prior and posterior distributions ($p > 0.05$). The distributions of parameter values in Pareto-optimal sets were significantly different from the behavioral parameter sets from the calibration period for α in all three layers, K_{sat} in the third layer, and θ_{res} in the shallowest layer (Figure 8). In many of these cases, the Pareto-optimal distribution was narrower but with similar central tendencies, suggesting increased parameter identifiability.

3.4 Performance during the validation period

In all three cases, model performance generally declined in the validation period (Table 2). However, this decline was generally most pronounced in Case 1. For example, θ_{50} RMSE declined by 0.013 in Case 1, but only by 0.003 in Case 3. This is a natural consequence of minimizing |ΔRMSE (T+2)| and |ΔRMSE (1.6P)| in Case 3, and illustrates the validity of this approach. In terms of model performance throughout both the calibration and validation period, Case 2 appears to be a particularly appealing compromise: validation RMSE changes fairly minimally at all depths. In all three cases, model performance at 20 cm is quite poor, and is notably worse in the validation than calibration periods.

Across all behavioral parameter sets, model performance of the behavioral parameter sets was significantly different during the validation than calibration period at 50 cm and 20 cm for both NSE and RMSE (two-sided $p < 0.0001$; $n = 3196$; Figure 9). Surprisingly, RMSE at the 5 cm layer was lower in the validation than the calibration period, with an average difference of 0.009. NSE was lower in the validation than calibration period in all cases, though the differences were greatest at the 20 and 50 cm depths. As in the three cases selected, model performance across all behavioral sets was particularly variable and poor at the 20 cm depth.

While model performance generally declined during the validation period, comparing the ranks of parameter sets in the calibration to validation period suggests that the parameter sets that performed best in the calibration period tend to be preferable in the validation period as well, though there is considerable variability (Figure 10). At the 50 cm depth used to identify acceptability of parameter sets, there was a modest relationship between the rank of parameter sets during the two periods, suggesting that some parameter sets that performed very well in the calibration period may not perform as well in the validation period, and vice versa. There were stronger relationships between calibration and validation performance at the other depths, with a weaker relationship at 10 cm. This suggests that the GLUE method resulted in overfitting for the variable used for calibration (50 cm soil moisture), but not for other variables.

4. Discussion

4.1 Modeled soil moisture and temperature dynamics

With all three parameter sets, SHAW generally captured the variations of soil moisture across depths, water years, and seasons reasonably well, though there were some considerable errors, particularly in the timing and magnitude of the snowmelt peak each year. Other studies using parameter inversion with SHAW to model soil moisture obtained lower RMSE values than those found in this study, with values ranging as low as 0.008, and up to 0.023 (Flerchinger et al., 2012; Gribb et al., 2009; Hymer et al., 2000). However, all of these studies used less than one year time periods; very low RMSEs are more achievable with short duration data. These sites were also in temperate regions with more complete co-located hydrometeorological data.

SHAW consistently underestimated soil moisture at the 20 cm depth. The higher observed moisture at this depth suggests either an unusual soil layer or a problem with sensor installation. Common problems with soil moisture sensors include contact with air pockets or rocks (Robinson et al., 2008); this would tend to reduce apparent soil moisture and is therefore an unlikely explanation in this case. The more likely explanation in this case is that the soil layer containing the 20 cm sensor or subjacent layers had high clay content that led to greater moisture retention.

Errors in modeled soil temperature at the 1 and 2 m depths suggest that SHAW is lacking some insulative layers that are present at the site. Given the importance of snow depth and thermal conductivity in this region for modulating winter temperature impacts on soil temperatures (e.g., Yi et al., 2015; Zhang et al., 2005), errors in snow depth or density may be the most likely explanation. Lacking snow depth or snow water equivalent data at this site or nearby with similar vegetation conditions, this hypothesis is difficult to confirm. Another possible explanation is errors in modeled soil thermal conductivity, due to misspecification of soil properties or the fact the SHAW assigns the thermal conductivity of quartz (8.8 W/m/°C) to all sand and rock content, whereas conductivity of the minerals that make up the coarse fraction could be much lower (perhaps 2.0 W/m/°C; He et al., 2021). While soil texture was well constrained by observations, rock content was roughly estimated at 30% and could contribute to this error. Observed soil moisture was not available at this depth. However, at the 50 cm depth, modeled soil moisture tended to dry slightly too early in the year. This would tend to reduce soil thermal conductivity, which would have the opposite effect of that seen in the soil temperature results. A final possible explanation is a surface litter layer with greater insulating properties than those represented by the 5 cm layer that we included in the model.

4.2 Parameter values and importance

The results of the GLUE parameter estimation provide calibration-constrained parameter estimates (Figure 8). The prior densities of parameter values were constrained by those from the NCSS soils database and literature values. However, the relationships between parameters were not constrained. The relationships between parameters at depth aligned with theoretical expectations more consistently in Case 2 than in Case 1 or 3. Calibrated K_{sat} was an order of magnitude greater than the lab-based values recorded in the NCSS database, consistent with other modeling studies (e.g., Blain & Milly, 1991; Chappell et al., 1998; Grayson et al., 1992). Interestingly, van Genuchten α was the only parameter that was relatively identifiable throughout the profile, even though only soil moisture at 50 cm was used in the calibration; other soil hydraulic parameters were primarily only important at the calibration depth. The α parameter was calibrated to be in the lower region of the prior range in the first layer, higher in the second layer, and values in the third layer depend on whether the Pareto-optimal sets or all behavioral sets are considered.

The strong relationship we observed between α and *n* in the NCSS database may be worthy of further consideration. Several studies have recently identified relationships between bulk density and van Genuchten parameters, though these were in peat soils (Liu & Lennartz, 2019; O'Connor et al., 2020). While those relationships are more useful for estimating hydraulic parameters based on easily measurable properties, the relationship observed here is useful for constraining parameter space in Monte Carlo approaches. Another important implication is that if this relationship was not accounted for, selecting the mean values of α and *n* based on the database or other sources would result in unrealistic parameter combinations.

4.3 Parameter selection based on calibration, validation, and Pareto optimality

The simulated soil moisture dynamics contained greater errors in the validation years than calibration, based on model efficiency measures and visual inspection of simulation results. These differences were greatest at the 50 cm depth, which was used for model calibration. This suggests likely overfitting by the GLUE method, though we also note that the longer validation than calibration period used in this study may lead to additional interannual variability beyond that captured by our simple analysis of climate space, and therefore greater errors. However, this longer validation period is appropriate for the case when hydrologic models calibrated on several years of data are used for long-term analysis of future climate scenarios. The parameter distribution of Pareto-optimal sets was generally similar to, but narrower than those included in the behavioral parameter sets. This suggests that identifying the Pareto front increased parameter identifiability but did not generally suggest dramatically different values than those identified by the GLUE method alone.

The loss of model performance in the validation period for most parameter sets suggests that the GLUE method used here resulted in overfit models. Moreover, this overfitting appears to relate to simple climatic characteristics of individual water years. Despite the fact that the calibration years spanned the range of interannual precipitation and mean temperature values fairly well, model performance was generally worst in the driest and warmest years. This result is concerning for climate change impacts studies; while physically-based models such as SHAW are widely used for climate change impact studies, this result suggests that errors increase with warming and precipitation change even within the historical range of variability. Given this observation, we suggest that uncertainty estimates should be viewed as the absolute minimum range of uncertainty due to parameter estimation in non-stationary climates. Here, we identified one simple method of reducing overfitting when one parameter set is needed for practical purposes. This parameter set considerably reduced the impact of interannual climatic variability on model performance.

4.4 Assumptions and limitations

Hydrologic modeling errors can originate from errors in the input or calibration data, model structure, and/or parameter estimations (Liu & Gupta, 2007). In this analysis, we focus on understanding, quantifying, and minimizing error due to parameter uncertainty. The loose coupling of SHAW and GIPL provides some insight to model structural error by identifying that GIPL appears to produce a useful and realistic temperature boundary condition for SHAW. However, other sources of model structural error were not addressed. For example, preferential flow through macropores is probably important in boreal soils and may be effectively addressed by a bimodal van Genuchten-Mualem approach (e.g., Coppola et al., 2009); this is not currently possible in the SHAW model. SHAW is also a one-dimensional model, so it does not directly model lateral hydrologic flow. This is likely an acceptable simplification in a relatively xeric upland site like the site studied here, but would be of concern in more mesic lowland sites.

We also did not attempt to quantify uncertainty due to errors in the input or calibration data, though some of our methods aim to account for the most likely sources of error in these data. Specifically, we assessed the potential impacts of precipitation undercatch (Sevruk, 1996), limited our moisture calibration to the generally non-freezing period (e.g., Chandler et al., 2017), and calibrated on daily, rather than hourly data to avoid potential short-timescale errors in soil moisture observations, such as diurnal temperature-based fluctuations (Seyfried & Grant, 2007). However, over a thirteen-year period, there are likely uncertainties in the input climate data, particularly given the large fraction of missing precipitation data and substitution from a somewhat distal meteorological station.

5. Conclusions

In this study, we used a GLUE approach to parameterize soil hydraulic parameters for a physically-based hydrologic model and characterize the uncertainty of modeled soil moisture dynamics in a deciduous, upland boreal forest. The model reproduced soil moisture and temperature dynamics well, capturing interannual variability and variability across soil layers. The model and observations were both very responsive to summer rainfall at shallow depths, as the site is free of near-surface permaforst. This suggests that summer rainfall is important for maintaining soil moisture at this, and similar sites (e.g., Hinzman et al., 2002). Given the uncertainty of summer precipitation in climate models, this implies an uncertain future for hydrologic conditions at these sites.

Our comparisons of parameter selection based on traditional performance measures alone, as well as those focused on interannual variability in model performance, suggested that the latter should be considered to apply models for non-stationary climates. Some trade-offs were required between these performance measures, but our experience suggests that compromise solutions can be identified. Both the hydrologic and modeling conclusions here were supported by long-term data that captured a wide range of interannually varying conditions. Our results suggest that this is critical for understanding the appropriate parameters and uncertainty of a hydrologic model for this region. The LTER data used here is a vital resource for not only ecological, but also hydrologic understanding. These findings support the potential to use SHAW for further investigations in boreal regions and demonstrate a global parameter estimation and sensitivity analysis using open source tools. Finally, the novel objective function illustrates a simple method to select parameter sets appropriate for modeling in non-stationary climates. We suggest that work assessing hydrologic change in future climate conditions should explicitly assess and aim to minimize the impacts of interannual climate variability on model error.

Acknowledgements

This research was funded by NSF Grant #1737413. The authors are extremely grateful to the decades of data collection at the BNZ LTER and to the NRCS for the soil water retention curve data used to guide prior parameter distributions.

Data availability

Data products used in this study are available in the public domain and cited in the references. SHAW model inputs and outputs for the three cases are stored at the NSF Arctic Data Center at: https://doi.org/10.18739/A2M61BQ93.

References

Andresen, C. G., Lawrence, D. M., Wilson, C. J., McGuire, A. D., Koven, C., Schaefer, K., Jafarov, E., Peng, S., Chen, X., Gouttevin, I., Burke, E., Chadburn, S., Ji, D., Chen, G., Hayes, D., & Zhang, W. (2020). Soil moisture and hydrology projections of the permafrost region – a model intercomparison. *The Cryosphere (Online)*, *14*(2), Article PNNL-SA-151044. https://doi.org/10.5194/tc-14-445-2020

Anderson, J.; Elsner, C.; Fathauer, T.; Euskirchen, E.S. 2020. *Greenup values for interior Alaska 1976 - Present*, Bonanza Creek LTER - University of Alaska Fairbanks. BNZ:300, http://www.lter.uaf.edu/data/data-detail/id/300. doi:10.6073/pasta/91782ca741d448eff8ebe63fa387b351

Baldocchi, D., Kelliher, F. M., Black, T. A., & Jarvis, P. (2008). Climate and vegetation controls on boreal zone energy exchange. *Global Change Biology*, *6*(S1), 69–83. https://doi.org/10.1046/j.1365-2486.2000.06014.x

Barichivich, J., Briffa, K. R., Myneni, R., Schrier, G. V. der, Dorigo, W., Tucker, C. J., Osborn, T. J., & Melvin, T. M. (2014). Temperature and Snow-Mediated Moisture Controls of Summer Photosynthetic Activity in Northern Terrestrial Ecosystems between 1982 and 2011. *Remote Sensing*, *6*(2), 1390–1431. https://doi.org/10.3390/rs6021390

Beaudette D., Skovlin, J. and Roecker, S. (2020). soilDB: Soil Database Interface. R package version 2.5. https://CRAN.R-project.org/package=soilDB

Beck, P. S. A., Juday, G. P., Alix, C., Barber, V. A., Winslow, S. E., Sousa, E. E., Heiser, P., Herriges, J. D., & Goetz, S. J. (2011). Changes in forest productivity across Alaska consistent with biome shift: Changes in forest productivity across Alaska. *Ecology Letters*, *14*(4), 373–379. https://doi.org/10.1111/j.1461-0248.2011.01598.x

Beven, K. (2006). A manifesto for the equifinality thesis. *Journal of Hydrology*, *320*(1), 18–36. https://doi.org/10.1016/j.jhydrol.2005.07.007

Beven, K. J., Smith, P. J., & Freer, J. E. (2008). So just why would a modeller choose to be incoherent? *Journal of Hydrology*, *354*(1), 15–32. https://doi.org/10.1016/j.jhydrol.2008.02.007

Beven, K., & Binley, A. (1992). The future of distributed models: Model calibration and uncertainty prediction. *Hydrological Processes*, *6*(3), 279–298. https://doi.org/10.1002/hyp.3360060305

Beven, K., & Binley, A. (2014). GLUE: 20 years on. *Hydrological Processes*, *28*(24), 5897–5918. https://doi.org/10.1002/hyp.10082

Birnbaum, Z.W. and F. H. Tingey (1951). One-sided confidence contours for probability distribution functions. The Annals of Mathematical Statistics, 22/4, 592–596. doi: 10.1214/aoms/1177729550.

Blain, C. A., & Milly, P. C. D. (1991). Development and application of a hillslope hydrologic model. *Advances in Water Resources*, *14*(4), 168–174. https://doi.org/10.1016/0309-1708(91)90012-D

Bonan, G. B. (1989). A computer model of the solar radiation, soil moisture, and soil thermal regimes in boreal forests. *Ecological Modelling*, *45*(4), 275–306. https://doi.org/10.1016/0304-3800(89)90076-8

Bonan, G. B. (1991). A biophysical surface energy budget analysis of soil temperature in the boreal forests of interior Alaska. *Water Resources Research*, *27*(5), 767–781. https://doi.org/10.1029/91WR00143

Burns, R. M., & Honkala, B. H. (1990). *Silvics of North America. Volume 2: Hardwoods* (Vol. 2). U.S. Department of Agriculture, Forest Service. https://www.srs.fs.usda.gov/pubs/misc/ag_654/table_of_contents.htm

Burt, R. 2009. Soil survey field and laboratory methods manual. Soil Surv. Investigations Rep. 51, version 1.0. USDA-NRCS, National Soil Survey Center, Lincoln, NE.

Cahoon, S. M. P., Sullivan, P. F., Brownlee, A. H., Pattison, R. R., Andersen, H.-E., Legner, K., & Hollingsworth, T. N. (2018). Contrasting drivers and trends of coniferous and deciduous tree growth in interior Alaska. *Ecology*. https://doi.org/10.1002/ecy.2223

Carnell, R. (2019). lhs: Latin Hypercube Samples. R package version 1.0.1. https://CRAN.R-project.org/package=lhs

Chandler, D., Seyfried, M., Mcnamara, J., & Hwang, K. (2017). Inference of Soil Hydrologic Parameters from Electronic Soil Moisture Records. *Frontiers in Earth Science*, *5*. https://doi.org/10.3389/feart.2017.00025

Chapin, F. S.; Ruess, R. W. 2018. *Bonanza Creek LTER: Hourly Soil Moisture (VWC) at Various Depths from 2002 to Present in the Bonanza Creek Experimental Forest near Fairbanks, Alaska*, Bonanza Creek LTER - University of Alaska Fairbanks. BNZ:5, http://www.lter.uaf.edu/data/data-detail/id/5. doi:10.6073/pasta/0c60c37f60d7f77f8646018d821e6481

Chappell, N. A., Franks, S. W., & Larenus, J. (1998). Multi-scale permeability estimation for a tropical catchment. *Hydrological Processes*, *12*(9), 1507–1523. https://doi.org/10.1002/(SICI)1099-1085(199807)12:9<1507::AID-HYP653>3.0.CO;2-J

Coppola, A., Basile, A., Comegna, A., & Lamaddalena, N. (2009). Monte Carlo analysis of field water flow comparing uni- and bimodal effective hydraulic parameters for structured soil. J*ournal of Contaminant Hydrology*, 104(1), 153–165. https://doi.org/10.1016/j.jconhyd.2008.09.007

Coron, L., Andréassian, V., Perrin, C., Bourqui, M., & Hendrickx, F. (2014). On the lack of robustness of hydrologic models regarding water balance simulation: A diagnostic approach applied to three models of increasing complexity on 20 mountainous catchments. Hydrology and Earth System Sciences, 18(2), 727–746. https://doi.org/10.5194/hess-18-727-2014

Das, I. (1999). A preference ordering among various Pareto optimal alternatives. Structural Optimization, 18(1), 30–35. https://doi.org/10.1007/BF01210689

Efstratiadis, A., & Koutsoyiannis, D. (2010). One decade of multi-objective calibration approaches in hydrological modelling: A review. *Hydrological Sciences Journal*, *55*(1), 58–78. https://doi.org/10.1080/02626660903526292

Endalamaw, A., Bolton, W. R., Young-Robertson, J. M., Morton, D., Hinzman, L., & Nijssen, B. (2017). Towards improved parameterization of a macroscale hydrologic model in a discontinuous permafrost boreal forest ecosystem. *Hydrol. Earth Syst. Sci.*, *21*(9), 4663–4680. https://doi.org/10.5194/hess-21-4663-2017

Federer, C. A. (1977). Leaf Resistance and Xylem Potential Differ Among Broadleaved Species. *Forest Science*, *23*(4), 411–419. https://doi.org/10.1093/forestscience/23.4.411

Fischer, R., Walsh, J. E., Euskirchen, E. S., & Bieniek, P. A. (2017). Regional Climate Model Simulation of Surface Moisture Flux Variations in Northern Terrestrial Regions. *Atmospheric and Climate Sciences*, *8*(1), 29–54. https://doi.org/10.4236/acs.2018.81003

Flerchinger, G. N. (2017). *The Simultaneous Heat and Water (SHAW) Model: Technical Documentation*. Technical Report NWRC 2017-02. https://pdfs.semanticscholar.org/4906/2b4ebd161d0df6179b113dd9b1e5abf7813f.pdf

Flerchinger, G. N., Caldwell, T. G., Cho, J., & Hardegree, S. P. (2012). Simultaneous Heat and Water (SHAW) Model: Model Use, Calibration, and Validation. *Transactions of the ASABE*, *55*(4), 1395–1411. https://doi.org/10.13031/2013.42250

Grayson, R. B., Moore, I. D., & McMahon, T. A. (1992). Physically based hydrologic modeling: 1. A terrain-based model for investigative purposes. *Water Resources Research*, *28*(10), 2639–2658. https://doi.org/10.1029/92WR01258

Gribb, M. M., Forkutsa, I., Hansen, A., Chandler, D. G., & McNamara, J. P. (2009). The Effect of Various Soil Hydraulic Property Estimates on Soil Moisture Simulations. *Vadose Zone Journal*, *8*(2), 321–331. https://doi.org/10.2136/vzj2008.0088

Hansen, W. D., Fitzsimmons, R., Olnes, J., & Williams, A. P. (n.d.). An alternate vegetation type proves resilient and persists for decades following forest conversion in the North American boreal biome. *Journal of Ecology*, *n/a*(n/a). https://doi.org/10.1111/1365-2745.13446

He, H., Flerchinger, G. N., Kojima, Y., Dyck, M., & Lv, J. (2021). A review and evaluation of 39 thermal conductivity models for frozen soils. *Geoderma*, *382*, 114694. https://doi.org/10.1016/j.geoderma.2020.114694

Hinzman, L. D., Ishikawa, N., Yoshikawa, K., Bolton, W. R., & Petrone, K. C. (2002). *Hydrologic Studies in Caribou-Poker Creeks Research Watershed in Support of Long Term Ecological Research*. *5*(2), 67–71.

Hymer, D. C., Moran, M. S., & Keefer, T. O. (2000). Soil Water Evaluation Using a Hydrologic Model and Calibrated Sensor Network. *Soil Science Society of America Journal*, *64*(1), 319–326. https://doi.org/10.2136/sssaj2000.641319x

Kasischke, E. S., Verbyla, D. L., Rupp, T. S., McGuire, A. D., Murphy, K. A., Jandt, R., Barnes, J. L., Hoy, E. E., Duffy, P. A., Calef, M., & Turetsky, M. R. (2010). Alaska's changing fire regime—Implications for the vulnerability of its boreal forests. *Canadian Journal of Forest Research. 40: 1313-1324*, *40*, 1313–1324.

Khu, S. T., & Madsen, H. (2005). Multiobjective calibration with Pareto preference ordering: An application to rainfall-runoff model calibration. *Water Resources Research*, *41*(3). https://doi.org/10.1029/2004WR003041

Johnstone, J. F., Hollingsworth, T. N., Chapin, F. S., & Mack, M. C. (2010). Changes in fire regime break the legacy lock on successional trajectories in Alaskan boreal forest. *Global Change Biology*, *16*(4), 1281–1295. https://doi.org/10.1111/j.1365-2486.2009.02051.x

Jorgenson, M. T., Romanovsky, V., Harden, J., Shur, Y., O'Donnell, J., Schuur, E. A. G., Kanevskiy, M., & Marchenko, S. (2010). Resilience and vulnerability of permafrost to climate change. *Canadian Journal of Forest Research*, *40*(7), 1219–1236. https://doi.org/10.1139/X10-060

Liu, Y., & Gupta, H. V. (2007). Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework. *Water Resources Research*, *43*(7). https://doi.org/10.1029/2006WR005756

Mann, D. H., Rupp, T. S., Olson, M. A., & Duffy, P. A. (2012). Is Alaska's Boreal Forest Now Crossing a Major Ecological Threshold? *Arctic, Antarctic, and Alpine Research*, *44*(3), 319–331. https://doi.org/10.1657/1938-4246-44.3.319

Marchenko, S., Romanovsky, V., & Tipenko, G. (2008). Numerical Modeling of Spatial Permafrost Dynamics in Alaska. *Ninth International Conference on Permafrost, 7.*

Menne, M.J., I. Durre, B. Korzeniewski, S. McNeal, K. Thomas, X. Yin, S. Anthony, R. Ray, R.S. Vose, B.E.Gleason, and T.G. Houston, 2012: Global Historical Climatology Network - Daily (GHCN-Daily), Version 3.26. NOAA National Climatic Data Center. http://doi.org/10.7289/V5D21VHZ [September 4, 2020].

Meredith, M., M. Sommerkorn, S. Cassotta, C. Dersken, A. Ekaykin, A. Hollowed, G. Kofinas, A. Mackintosh, J. Melbourne-Thomas, M.M.C. Muelbert, G. Ottersen, H. Pritchard, & E.A.G. Schuur. (2019). Polar Regions. In *IPCC Special Report on the Ocean and Cryosphere in a Changing Climate* [H.-O. Pörtner, D.C. Roberts, V. Masson-Delmotte, P. Zhai, M. Tignor, E. Poloczanska, K. Mintenbeck, A. Alegría, M. Nicolai, A. Okem, J. Petzold, B. Rama, N.M. Weyer (eds.)].

Mualem, Y. (1976). A new model for predicting the hydraulic conductivity of unsaturated porous media. *Water Resources Research*, *12*(3), 513–522. https://doi.org/10.1029/WR012i003p00513

Nicolsky, D. J., Romanovsky, V. E., & Panteleev, G. G. (2009). Estimation of soil thermal properties using in-situ temperature measurements in the active layer and permafrost. *Cold Regions Science and Technology. 55: 120-129*. https://doi.org/10.1016/j.coldregions.2008.03.003

O'Connor, M. T., Cardenas, M. B., Ferencz, S. B., Wu, Y., Neilson, B. T., Chen, J., & Kling, G. W. (2020). Empirical Models for Predicting Water and Heat Flow Properties of Permafrost Soils. *Geophysical Research Letters*, *47*(11), e2020GL087646. https://doi.org/10.1029/2020GL087646

Pianosi, F., Beven, K., Freer, J., Hall, J. W., Rougier, J., Stephenson, D. B., & Wagener, T. (2016). Sensitivity analysis of environmental models: A systematic review with practical workflow. *Environmental Modelling & Software*, *79*, 214–232. https://doi.org/10.1016/j.envsoft.2016.02.008

Robinson, D. A., Campbell, C. S., Hopmans, J. W., Hornbuckle, B. K., Jones, S. B., Knight, R., Ogden, F., Selker, J., & Wendroth, O. (2008). Soil Moisture Measurement for Ecological and Hydrological Watershed-Scale Observatories: A Review. *Vadose Zone Journal*, *7*(1), 358–389. https://doi.org/10.2136/vzj2007.0143

Roocks, P. Computing Pareto Frontiers and Database Preferences with the rPref Package. *The R-Journal*, 8(2):393-404, Dec. 2016. https://doi.org/10.32614/RJ-2016-054

Sazonova, T. S., & Romanovsky, V. E. (2003). A model for regional-scale estimation of temporal and spatial variability of active layer thickness and mean annual ground temperatures. *Permafrost and Periglacial Processes*, *14*(2), 125–139. https://doi.org/10.1002/ppp.449

Scheffer, M., Hirota, M., Holmgren, M., Nes, E. H. V., & Chapin, F. S. (2012). Thresholds for boreal biome transitions. *Proceedings of the National Academy of Sciences*, *109*(52), 21384–21389. https://doi.org/10.1073/pnas.1219844110

Sevruk, B. (1996). Adjustment of tipping-bucket precipitation gauge measurements. *Atmospheric Research*, *42*(1), 237–246. https://doi.org/10.1016/0169-8095(95)00066-6

Seyfried, M. S., & Grant, L. E. (2007). Temperature Effects on Soil Dielectric Properties Measured at 50 MHz. *Vadose Zone Journal*, *6*(4), 759–765. https://doi.org/10.2136/vzj2006.0188

Stedinger, J. R., Vogel, R. M., Lee, S. U., & Batchelder, R. (2008). Appraisal of the generalized likelihood uncertainty estimation (GLUE) method. *Water Resources Research*, *44*(12). https://doi.org/10.1029/2008WR006822

Teufel, B., & Sushama, L. (2019). Abrupt changes across the Arctic permafrost region endanger northern development. *Nature Climate Change*, *9*(11), 858–862. https://doi.org/10.1038/s41558-019-0614-6

Ueyama, M., Iwata, H., Harazono, Y. (2002-) AmeriFlux US-Uaf University of Alaska, Fairbanks, Dataset. https://doi.org/10.17190/AMF/1480322

van Cleve, K.; Chapin, F.S.; Ruess, R. W. 2015. *Bonanza Creek LTER: Tree Inventory Data from 1989 to present at Core research sites in Interior Alaska*, Bonanza Creek LTER - University of Alaska Fairbanks. BNZ:320, http://www.lter.uaf.edu/data/data-detail/id/320. doi:10.6073/pasta/8366b043fbb4dfc220196425284d90a7

Van Cleve, K.; Chapin, F. Stuart; Ruess, Roger W. 2017a. *Bonanza Creek LTER: Hourly Air Temperature Measurements (sample, min, max) at 50 cm and 150 cm from 1988 to Present in the Bonanza Creek Experimental Forest near Fairbanks, Alaska*, Bonanza Creek LTER - University of Alaska Fairbanks. BNZ:1, http://www.lter.uaf.edu/data/data-detail/id/1. doi:10.6073/pasta/006bae44c88f7d8b6fabe8cfebee86ff

Van Cleve, Keith; Chapin, F. Stuart; Ruess, Roger W. 2017b. *Bonanza Creek LTER: Hourly Wind Speed and Direction at 3m and 10 m from 1988 to Present in the Bonanza Creek Experimental Forest near Fairbanks, Alaska*, Bonanza Creek LTER - University of Alaska Fairbanks. BNZ:2, http://www.lter.uaf.edu/data/data-detail/id/2. doi:10.6073/pasta/a01890214ed4a1e6e449181689b7f604

Van Cleve, Keith; Chapin, F. Stuart; Ruess, Roger W. 2018a. *Bonanza Creek LTER: Hourly Precipitation Measurements from 1988 to Present in the Bonanza Creek Experimental Forest near Fairbanks, Alaska*, Bonanza Creek LTER - University of Alaska Fairbanks. BNZ:4, http:// www.lter.uaf.edu/data/data-detail/id/4. doi:10.6073/pasta/b8258ab0d1ee0707d3d6fd0ee460545c

Van Cleve, Keith; Chapin, F. Stuart; Ruess, Roger W. 2018b. *Bonanza Creek LTER: Hourly Relative Humidity Measurements (mean, min, max) at 50 cm and 150 cm from 1988 to Present in the Bonanza Creek Experimental Forest near Fairbanks, Alaska*, Bonanza Creek LTER - University of Alaska Fairbanks. BNZ:241, http://www.lter.uaf.edu/data/data-detail/id/241. doi:10.6073/pasta/a3d132eae808914131dc3f2ddcb5e403

Van Cleve, Keith; Chapin, F. Stuart; Ruess, Roger W. 2018c. *Bonanza Creek LTER: Hourly Snow Depth Measurements from 1988 to Present in the Bonanza Creek Experimental Forest near Fairbanks, Alaska*, Bonanza Creek LTER - University of Alaska Fairbanks. BNZ:161, http://www.lter.uaf.edu/data/data-detail/id/161. doi:10.6073/pasta/fc1f61d73a5f4cb73b9b4757d5e29be6

Van Cleve, Keith; Chapin, F. Stuart; Ruess, Roger W. 2018d. *Bonanza Creek LTER: Hourly Soil Temperature Measurements at Various Depths from 1988 to Present in the Bonanza Creek Experimental Forest near Fairbanks, Alaska*, Bonanza Creek LTER - University of Alaska Fairbanks. BNZ:3, http://www.lter.uaf.edu/data/data-detail/id/3. doi:10.6073/pasta/937dadb2ba822a7a9987f394160ef4f1

Wood, S.N. (2017) Generalized Additive Models: An Introduction with R (2nd edition). Chapman and Hall/CRC.

Yarie, John. 1998. *Soil physical and chemical properties based on genetic horizon from 4 replicate pits placed around the replicate LTER control plots sampled in 1988 and 1989*, Bonanza Creek LTER - University of Alaska Fairbanks. BNZ:134, http://www.lter.uaf.edu/data/ data-detail/id/134. doi:10.6073/pasta/475a1825dfa264822ed53ca3574bb8e6

Yarie, J., & van Cleve, K. (2010). Long-term monitoring of climatic and nutritional effects on tree growth in interior Alaska. *Canadian Journal of Forest Research. 40: 1325-1335*, *40*, 1325–1335.

Yi, Y., Kimball, J. S., Rawlins, M. A., Moghaddam, M., & Euskirchen, E. S. (2015). The role of snow cover affecting boreal-arctic soil freeze–thaw and carbon dynamics. *Biogeosciences*, *12*(19), 5811–5829. https://doi.org/10.5194/bg-12-5811-2015

Zhang, T. (2005). Influence of the seasonal snow cover on the ground thermal regime: An overview. *Reviews of Geophysics*, *43*(4). https://doi.org/10.1029/2004RG000157

Zhang, Y., Cheng, G., Li, X., Han, X., Wang, L., Li, H., Chang, X., & Flerchinger, G. N. (2013). Coupling of a simultaneous heat and water model with a distributed hydrological model and evaluation of the combined model in a cold region watershed. *Hydrological Processes*, *27*(25), 3762–3776. https://doi.org/10.1002/hyp.9514