Spatial Patterns of Snow Distribution in the sub-Arctic

Katrina E. Bennett[1], Greta Miller[1], Robert Busey[2], Min Chen[1], Emma R. Lathrop[1], Julian B. Dann[1], Mara Nutt[1], Ryan Crumley[1], Shannon L. Dillard[1,5], Baptiste Dafflon[3], Jitendra Kumar[4], W. Robert Bolton[2], Cathy J. Wilson[1], Colleen Iverson[4], and Stan Wullschleger[4]

[1] Los Alamos National Laboratory, Earth and Environmental Sciences, Los Alamos, NM, USA
[2] University of Alaska Fairbanks, International Arctic Research Center, Fairbanks, AK, USA
[3] Lawrence Berkeley National Laboratory, Earth and Environmental Sciences, Berkeley, CA, USA
[4] Oak Ridge National Laboratory, Environmental Sciences Division and Climate Change Science Institute, Oak Ridge, TN, USA
[5] University of Wisconsin–Madison, Department of Geography, Madison, WI, USA

Correspondence to: Katrina E. Bennett (kbennett@lanl.gov)

Abstract. The spatial distribution of snow plays a vital role in sub-Arctic and Arctic climate, hydrology, and ecology due to its fundamental influence on the water balance, thermal regimes, vegetation, and carbon flux. However, the spatial distribution of snow is not well understood, and therefore, it is not well modeled, which can lead to substantial uncertainties in snow cover representations. To capture key hydro-ecological controls on snow spatial distribution, we carried out intensive field studies over multiple years for two small (2017-2019, ~2.5 km^2) sub-Arctic study sites located on the Seward Peninsula of Alaska. Using an intensive suite of field observations (>22,000 data points), we developed simple models of spatial distribution of snow water equivalent (SWE) using factors such as topographic characteristics, vegetation characteristics based on greenness (normalized different vegetation index, NDVI), and a simple metric for approximating winds. The most successful model was the random forest using both study sites and all years, which was able to accurately capture the complexity and variability of snow characteristics across the sites. Approximately 86% of the SWE distribution could be accounted for, on average, by the random forest model at the study sites. Factors that impacted year-to-year snow distribution included NDVI, elevation, and a metric to represent coarse microtopography (topographic position index, or TPI), while slope, wind, and fine microtopography factors were less important. The characterization of the SWE spatial distribution patterns will be used to validate and improve snow distribution modelling in the Department of Energy's earth system model, and for improved understanding of hydrology, topography, and vegetation dynamics in the sub-Arctic and Arctic regions of the globe.

Keywords: Snow Distribution, Random Forests, Spatial Patterns, Topography, Vegetation, Alaska, sub-Arctic

Check TWP, precip, check site names, read through.

1 Introduction

Covering the land for more than six months each year, snow plays a vital role in the climate, hydrology, and ecosystems of the Arctic and sub-Arctic. Snow directly impacts climate through modulation of atmospheric circulation patterns via the snow-albedo feedback mechanism (Fletcher et al., 2009) and atmospheric moisture budgets through its control on the amount of water available for evaporation (Callaghan et al., 2011). Thus, snow controls water availability, soil moisture, and temperature, affecting all components of ecosystem including vegetation (Evans et al., 1989; Schaefer and Messier, 1995; Scott and

Rouse, 1995) animal populations (Forchhammer et al., 2008; Manning and Garton, 2012), microbial decomposition, and carbon flux (Mauritz et al., 2017; Zona et al., 2016). The distribution of snow and timing of its melt is key to understanding how changes in hydrology, soil thermal regimes, and vegetation interact across Arctic and sub-Arctic landscapes (Jafarov et al., 2018). Snow distribution, however, is a particularly elusive and difficult feature to characterize, leading to challenges in how to quantify snow properties over space and understand how snow changes over time, especially in remote, under-monitored watersheds of the Arctic and sub-Arctic.

Within the Arctic and sub-Arctic hydrologic cycle, spring snowmelt is the single most important event contributing to the annual water budget of high-latitude watersheds (Ford and Bedford, 1987; Stuefer et al., 2013). The importance of understanding end-of-winter snow distribution and its impact on snowmelt and importance in snow modeling has been shown in the previous studies on this topic, including historical analyses (Homan and Kane, 2015), modeling, and analyses to characterize the spatial pattern of snow distribution and investigate how different factors affect the spatial distribution (Mendoza et al., 2020; Freudiger et al., 2017; Mott et al., 2018; Revuelto et al., 2014; Trujillo et al., 2007).

There are two main approaches to identify and quantify the influence of different factors on snow distribution: physically-based (dynamical) and statistical (empirical) models (Tarboton et al., 2000; Grünewald et al., 2013). Physically-based dynamical models including SnowTran-3D (Liston and Sturm, 1998; Liston et al., 2007; Hirashima et al., 2004), the distributed blowing-snow model (DBSM) (Pomeroy et al., 2007; Essery and Pomeroy, 2004), and the 3D snowdrift model (Jaedicke and Sandvik, 2002) have been successfully applied to the Arctic. These physically-based models account for both mass and energy exchanges and allow for detailed representation of different snow processes such as deposition, accumulation, redistribution, sublimation and melting (Tarboton et al., 2000; Grünewald et al., 2013). However, high-quality meteorology, topography, and vegetation parameterizations are generally required as input, and the computational cost can be high for dynamical models (Liston, 2004; Grünewald et al., 2013).

In contrast, empirically-based statistical models use the relationships between snow depth or snow water equivalent (SWE) and topography, vegetation, and wind to predict the snow distribution. Statistical models have parsimonious model structures, so they are computationally inexpensive and easy to use, but their drawbacks are that they are site-specific and require substantial data for model calibration (Tarboton et al., 2000). Decision trees (König and Sturm, 1998) and multiple linear or non-linear regression models (Wainwright et al., 2017; Dvornikov et al., 2015) are examples of statistical snow distribution models that have been applied in the Arctic and sub-Arctic regions.

More recently, machine-learning approaches have been used to quantify snow distribution using a variety of different algorithms and remote-sensing methods. Broxton et al. (2019) applied artificial neural networks to estimate snow density, which was then combined with aerial lidar snow depth to predict SWE. Revuelto et al. (2020) used random forests to predict lidar snow depth distribution from several topographic predictors. King et al. (2020) used random forests for bias correction of a SWE data assimilation product. Other studies have applied machine-learning algorithms using remotely sensed observations as predictors, including brightness temperature, fractional snow-covered area, or the normalized difference snow index (Liu et al., 2020; Bair et al., 2018). Meloche et al. (2022) used random forests to predict snow depth from topography, an up-wind slope index, and ecotypes in an area of Nunavut. While machine-learning approaches have shown to be an effective method for predicting snow distribution, few studies have incorporated vegetation characteristics into the models, and validated models with intensive field observations of SWE (Anderton et al., 2004).

A particular interest of this paper is to investigate the spatial pattern of snow distribution based on intensive field snow sampling surveys to identify what factors have control on the snow distribution at the local scale for two study sites in the southwestern and central Seward Peninsula, Alaska. Our main focus is on identifying secondary factors of snow distribution as opposed to the primary variables of temperature and precipitation. Specifically, our goal is to build a statistical model to 1) characterize the spatial pattern of the end-of-winter snow distribution, 2) identify the key factors controlling the spatial distribution, and 3) predict the snow distribution for the local study sites. We expect our analysis will be useful for the validation of the physically-based permafrost hydrology models such as the Advanced Terrestrial Simulator (ATS), which has been developed at fine spatial scales to understand permafrost dynamics for the region (Atchley et al., 2016, 2015; Painter et al., 2016). Further, the statistical snow distribution model will be used to validate and improved snow redistribution in Department of Energy (DOE)'s Energy Exascale Earth System model (E3SM) land surface model (ELM) and to eventually improve our understanding of changing hydrology, topography, and vegetation dynamics in the Arctic and sub-Arctic.

This paper is organized as follows: in data and methodology, we introduce the two study sites used for our analysis, describe the field observations of snow, and we outline the site characteristics used as factors in the study, including topography, vegetation, and winds. We present the methodology for each modelling approach, the results, a discussion of findings, and our next steps in the research. We then summarize the main conclusions for the study.

2 Data and Methodology

2.1 Study Sites

The Teller watershed (2.3 km^2) and Kougarok Hillslope study site (2.5 km^2) are located in the southwest part of the Seward Peninsula, Alaska (Figure 1). The climate of the Seward Peninsula is characterized by cool continental conditions, typified by long, cold winters and short, cool summers, and high precipitation (Peel et al., 2007). The mean annual air temperature at the Nome Municipal Airport (1980-2018, located approximately 35 km from Teller and 78 km from Kougarok) is -2.3 °C, with a mean January temperature of -14.4 °C and mean July temperature of 11.2 °C. Annual precipitation is 430 mm with 45 % falling as snow (National Centers for Environmental Information, National Oceanic and Atmospheric Administration, https://www.ncdc.noaa.gov/). At the Nome Municipal Airport (National Centers for Environmental Information, National Oceanic and Atmospheric Administration, https://www.ncdc.noaa.gov/) average historical (1981-2010) precipitation falling from October to March is 161 mm, while total annual precipitation (rain and snow) is 425 mm. Snow covers the ground, generally, from approximately October through May, dependent on the year.

The Teller watershed, with elevations ranging from 50 m to 300 m, is located nearby the coast and underlain by discontinuous permafrost, with near-surface permafrost adjacent to areas with no permafrost or deep permafrost table locations overlain by a perennially thawed layer (i.e., talik) (Jorgenson et al., 2008; Busey et al., 2008; Uhlemann et al., 2021; Léger et al., 2019). Topographic features at the sites include terrace and risers, and stream bed (Figure 1). The streams in the Teller watershed connect to the Sinuk River, about 1 km to the south of the study site (#2, Figure 1). Along the streams, willow shrubs grow, while sedge-willow-dryas tundra and mixed shrub-sedge tussock tundra-bog dominate the rest of the watershed (U.S. Fish and Wildlife Service, 2015; Konduri and Kumar, 2021).

The Kougarok Hillslope study site is located inland on the leeward side of the Kigluaik Mountains along a minor drainage of the Kuzitrin River, which flows into the Imuruk Basin. The Kougarok Hillslope (referred to as Kougarok from here on in) is situated on a gently rising upland dome that tops out at an elevation of ~110 m (#3, Figure 1). The site is overlain by an active soil layer containing organic peat and mineral horizons (McCaully et al., 2021), vegetated by alder shrubland, tussock tundra, alder savanna, and rocky areas dominated by dwarf shrubs and lichens (Iversen et al., 2019; Salmon et al., 2016). The site is underlain by permafrost approximately 15–50 m thick, with average active layer thickness of 56 cm (Hinzman et al., 2003).

To obtain spatially consistent and replicable (for different study sites and years) estimates of precipitation falling in each separate year of study, the total winter precipitation (TWP) for each year was estimated based on the ERA5-Land hourly reanalysis product (Muñoz 2019), accessed via Google Earth Engine. For each winter, we summed the winter (October to March) precipitation and averaged these totals across the study areas.

2.2 Field Observations of Snow

From March 22nd to March 31st, 2017 (Teller watershed), March 26th to April 5th, 2018 (Teller watershed, Kougarok Hillslope), and March 31st to April 7th, 2019 (Teller watershed), end-of-winter snow surveys were carried out at the study sites to collect snow depth and snow density to calculate SWE. SWE measures the amount of water contained within the snowpack and characterizes the hydrological and thermal impacts of snow cover better than snow depth (Jonas et al., 2009; Sturm et al., 2010; Liston and Elder, 2006), which is why we focus on SWE.

We measured snow characteristics in several different ways. Snow depth was captured every 1-15 m using a Snow-Hydro™ GPS Snow Depth Probe (Sturm and Holmgren, 2018). When the snowpack was greater than 130 cm in depth, the length of the snow depth probe, an avalanche probe was used to manually measure the snow depth. Snow bulk density was measured with a SWE Coring Tube, also manufactured by Snow-Hydro™ (reported error of -9% to 11%, Dixon and Boon, 2012; Young et al., 2018; López-Moreno et al., 2020). Approximately three to five bulk density samples were taken for each sampling location, with an interval of ~200-300 meters between measurements. The snow tube was pressed into the snow until the ground was hit and the depth of the snowpack in cm ($Snow\ Depth$) was recorded. After the tube was removed from the snowpack, the snow in the tube was bagged and weighed in grams ($Snow\ Weight$). The cross-sectional area of the snow coring tube ($Coring\ Tube\ Area$) was 30 cm^3. Then $Snow\ Density$ (g/cm^2) was calculated from Equation (1):

$$Snow\ Density\ = \frac{Snow\ Weight}{Snow\ Depth \times Coring\ Tube\ Area}\,. \tag{1}$$

Inverse distance weighting (IDW) was used to assign snow density for each of the observed snow depth locations, and is a common method for interpolating environmental variables (Franke, 1982; Zimmerman et al., 1999). The IDW method added uncertainty to our SWE estimates since we only collected sparse density measurements. These sparse density measurements ranged over years (Teller) between 0.27 to 038 kg/m^3, with a standard deviation of 0.03 to 0.04 kg/m^3. SWE (cm) was then calculated from the average of the snow bulk density measurements using Equation (2), where $Water\ Density$ was 0.997 g/cm^3:

$$SWE = Snow\ Depth\ \times \frac{Snow\ Density}{Water\ Density}. \tag{2}$$

SWE was then used as the response variable in our statistical models.

2.3 Site Characteristics

To model snow redistribution, various landscape factors were estimated for topographic, vegetation, and wind characteristics, as described below. An overview of data sources, factors, and descriptions of the factors is given in Table 1.

2.3.1 Topography

To evaluate the effects of topography on SWE distributions, a digital elevation model (DEM) was analyzed to estimate elevation, aspect, and slope for each of the study locations. The Teller watershed and Kougarok Hillslope DEMs were derived at a 5 m resolution from Interferometric Synthetic Aperture Radar (IfSAR) data available from http://ifsar.gina.alaska.edu/.

To consider the effects of topography at different spatial scales on SWE distributions, we included a fine-scale “microtopography” and a topographic position index (TPI) to capture the range of terrain variability across spatial scales. Values are calculated by the difference between the cell and the average elevation of all cells in a surrounding square window of 15 m and 155 m width for microtopography and TPI, respectively.

Microtopography in the Arctic can range from sub-meter (e.g., tussocks) to 1-10 m (e.g., hummocks) in scale (Sturm and Holmgren, 1994). The 15 m microtopography is the finest scale that can be derived using a DEM at 5 m resolution, and is equal to the curvature of the terrain (Jenness, 2006). We also consider coarse-scale topographic features defined using TPI, such as terraces and risers, and the stream channel (10s to 100s of m). To determine the scale of TPI, we applied a smoothing-average window at a range of widths, and then picked the optimal width with the best random forest model performance and highest feature importance (see Section 4.4, Figure A1). We derived both of the above described products from the DEM following Lopez-Moreno et al. (2009) and Weiss (2001).

2.3.2 Vegetation

Owing to the importance of shrubs for trapping drifting snow in the Arctic (Sturm et al., 2001a; Essery and Pomeroy, 2004; Dvornikov et al., 2015; McFadden et al., 2001), we considered different vegetation indicators to reflect types and distributions: vegetation type and vegetation greenness.

Vegetation type was extracted from an updated vegetation map for both study sites (Konduri and Kumar, 2021) and used as a continuous feature ranked for each year and site according to IDW-interpolated SWE. The continuous ranking was applied so that feature importance could be compared across all model features on a consistent basis. The relative rankings from each year and site were then averaged to produce a final ranking used in the models (Table A1).

As high-resolution vegetation distribution information is not widely available for most of the Arctic, Normalized Difference Vegetation Index (NDVI) was used to approximate vegetation characteristics. NDVI is indicative of the abundance of photosynthetically active vegetation (Rouse et al., 1974) and is useful to capture the branch abundance, deciduous canopy cover and maximum height of shrubs in the Arctic tundra landscape (Boelman et al., 2011). NDVI was derived from the 8-band WorldView-2 images obtained on July 27, 2011 at 1.5m resolution for the Teller watershed, and July 14, 2017 for the Kougarok Hillslope. Both images were downloaded from the DigitalGlobe website (https://www.digitalglobe.com/). To better understand how vegetation type was related to NDVI, we binned vegetation types that were present on more than 1% of the total watershed area which resulted in six representative categories that encompass the range of wetland species, short shrub, and tall shrub vegetation types. We computed statistics and significance to differentiate the NDVI values for those bins using ANOVA and Tukey's HSD test (Table A1).

2.3.3 Wind

Previous studies showed that snow is usually accumulated on leeward slopes and blown away from windward slopes (Evans et al., 1989; Liston and Sturm, 1998; Winstral et al., 2002; Mott et al., 2011), so we explored whether the prevailing wind would have an impact on the snow distribution. To understand wind patterns, weather stations within the study locations and from the nearby Nome weather station were analyzed (Table A2).

Because faster wind speeds have a greater impact on the redistribution of snow, we considered winter (October to March) wind speeds greater than 5 m/s (Table A2, Liston and Sturm, 1998; Berg, 1986; Sturm and Wagner, 2010; Sturm and Stuefer, 2013). At the weather station at the top of the Teller watershed, the prevailing wind direction (wind speeds > 5 m/s) was the average of data from the 2016-2017, 2017-2018, and 2018-2019 winters, and at the Kougarok weather station, the prevailing wind direction (wind speeds > 5 m/s) was based on data from the 2018-2019 winter only due to instrument issues in previous years. The prevailing wind direction (wind speeds > 5 m/s) for each study site was used to represent the exposure of a particular location to wind as a function of aspect (Dvornikov et al., 2015). We divided the prevailing winds into the eight cardinal and ordinal directions (N, NE, E, SE, S, SW, W, NW) using 45-degree bins, and then derived a unique wind and aspect factor equation for each directional bin. For the Teller watershed, the prevailing wind direction was 102° (E), so the wind and aspect factor (Wf) was calculated using Equation 3:

$$Wf = -\sin(A), \tag{3}$$

where A is the aspect (Dvornikov et al., 2015; Evans et al., 1989; Liston and Sturm, 1998). For Kougarok, the prevailing wind direction was 45° (NE), so the Wf was calculated using Equation 4:

$$Wf = -\cos(A) - \sin(A). \tag{4}$$

The wind and aspect factor gives positive values for leeward slopes and negative values for windward slopes, since snow is known to blow away from windward slopes and accumulate on leeward slopes (Dvornikov et al., 2015). More details on the wind factors as applied in this work are included in the Appendix and Figure A2.

3 Methodology

3.1 Modelling

We fit three different types of models, linear regression, general additive, and random forests, to quantify the impacts of different factors on snow distribution and characterize the spatial pattern of the snow distribution.

3.1.1 Linear Regression Model

The linear regression model is shown in Equation (5),

$$y = \beta_o + \beta_1 x_1 + \cdots \beta_p x_p + \epsilon, \tag{5}$$

where y denotes SWE, and β are the coefficient of factors expressed by the weighted sum of its p features with an error term, ϵ. Linear models are useful in that they produce simple relationships between the response variable and factors, and the factor rankings are easy to interpret. However, relationships between snow distribution and factors may not be linear, and thus linear models may not provide a good result, additionally, linear models are susceptible to correlations between parameters. The linear models were implemented using the Linear Regression class of the Scikit-learn package in Python (Pedregosa et al., 2011).

3.1.2 Generalized Additive Model

A generalized additive model, or GAM, is a class of statistical models in which the usual linear relationship between the response and predictors are replaced by several nonlinear smooth functions to model and capture the non-linearities in the data, i.e.,

$$g(E_Y(y|x)) = \beta_o + f_1(x_1) + f_2(x_2) + \cdots + f_p(x_p), \tag{6}$$

where g denotes SWE as a link function that links the expected value to the predictor variables $x_1, \ldots x_p$, which denote smooth, nonparametric functions. This type of model is useful because it allows for non-linear relationships between SWE and the factors. GAMs are generally easy to interpret, but while they are non-linear, GAMs still require a fit to a distribution or shape. The GAMs were implemented using the LinearGAM class of the pygam package in Python (Servén et al., 2018).

3.1.3 Random Forest Model

Random forests are based on decision trees, a series of yes/no questions asked about our data eventually leading to a predicted class (or continuous value in the case of regression, Breiman, 2001). A random forest model is defined by a large number of individual decision trees that operate as an ensemble. Each individual tree in the random forest generates a vote for classes (classification) or mean prediction (regression) of the individual trees, and the one with the most votes is used for the final prediction (Liaw and Wiener, 2002). Random forests are useful in that they are not impacted by correlations in the data, generally can protect against over-fitting, and have built-in feature randomization. The drawback of these models is that they can be difficult to control, if used in black box mode. The random forests were implemented using the RandomForestRegressor class of the Scikit-learn package in Python (Pedregosa et al., 2011).

3.1.4 Model Implementation

For all three model (linear, GAM, random forest) iterations, SWE, the response variable, was square root transformed to ensure its distribution was normal. Model inputs were also normalized to have a mean of 0 and a standard deviation of 1. We utilized a split sample approach for all models, with a train and test set. Performance metrics for all three statistical models were the coefficient of determination (R^2) and root mean squared error (RMSE) on the test set, which were evaluated on the untransformed (squared) model output. Due to the decorrelation effects involved in bootstrapping, the predictive accuracy of random forests is generally robust to collinearity across features (Dormann et al., 2013). However, feature collinearity can still be an issue for determining feature importance (Gregorutti et al., 2017). Prior to modelling, we used variance inflation factors as well as pairwise correlation coefficients to assess collinearity among features and ensure collinearity was not a significant issue.

We implemented the random forest model using various subsets of the input factors. One implementation, labeled the “final model”, was trained on data from all years at the Teller watershed and the Kougarok Hillslope combined, with the TWP for each year included as a feature in the model. We also trained the model separately for only the Teller watershed using all years combined (2017-2019), with the TWP included as a feature. In addition, we implemented individual random forest model runs for each year and each site, without TWP. The model runs on individual sites and years were also implemented for linear regression models and GAMs so that the performance of the three different statistical models could be compared.

Since the random forest performed the best of the three models and has the most comprehensive feature importance metrics, all testing for model features and hyperparameters was completed with the random forest model. The same features were then applied in the linear model and GAM predictions. Model hyperparameters (tree density, max depth, max features, min samples split, and min samples leaf) were selected using Bayesian Search (using the BayesSearchCV function from the Scikit-Optimize package in Python) with 8-fold cross validation on the training set. The training set was a randomly selected 80% of full dataset, and the remaining 20% of the dataset was used for validation. A complete list of the hyperparameters used for the final model and the separate the Teller watershed and the Kougarok Hillslope model runs are given in Table A3.

Using the random forest model, we measured the contribution of each input feature in predicting SWE distribution with both impurity feature importance and permutation feature importance (Louppe et al., 2013). Impurity importance, also known as Mean Decrease in Impurity (MDI) or Gini importance, is proportional to the total number of splits that each feature divides across all trees in the random forest, where features with more splits are more important (Breiman, 2001). While impurity importance is computationally cheap, it is biased towards features with many possible split points, and suffers from overfitting to the training set. Permutation importance, also known as Mean Decrease in Accuracy (MDA), is based on the decrease in model performance when a single feature is randomly shuffled, and more important features result in larger decreases in performance when permuted (Breiman, 2001). Permutation importance is more computationally expensive. We measured permutation importance on the test set. For this study, we analyzed both importance metrics (MDI and MDA) to ensure that the relative feature importance rankings generally agreed.

4 Results

4.1 Meteorological Conditions

Meteorological stations located in the study sites (Figure 1) recorded temperatures of -7.1°C (the Teller watershed, 2017-2019, including the top and bottom meteorological stations) / -8.1°C (the Kougarok Hillslope, 2018) during the winter months of October to March, while the Nome airport reported average 2017-2019 winter temperatures of -8.0°C. From 2017-2019, the Nome Airport reported winter temperatures that were slightly warmer than the recent (1981-2010) climatological period (-10.5°C). Precipitation recorded at the Nome Airport climate station indicated that from October to March, precipitation was 13.3 cm, 28.0 cm, and 25.0 cm, while precipitation from December to March at the Nome airport was 7.2 cm, 15.2 cm, and 16.4 cm for 2017, 2018, and 2019, respectively. ERA5 average winter (October to March) total precipitation values are 26.3 cm, 40.3 cm, and 44.9 cm for both the Teller watershed and the Kougarok Hillslope for 2017-2019, respectively. From October to March, the prevailing wind speed and direction for the Teller watershed (top station) were predominantly East-Northeast (2017-2019), while Kougarok was predominantly Northeast in 2019 (Table A2, Busey et al., 2017). The Kougarok Hillslope meteorological station experienced issues with its wind sensor in 2018, thus those values are not reported herein. The wind directions experienced at the study sites are similar to that experienced at the Nome Airport (East-Northeast). Wind speeds ranged from 4.9 to 6.7 m/s, much lower than wind speeds reported at the Nome (10.1 to 12.4 m/s) in 2017-2019 (Table A2).

4.2 Snow Depth, Density and SWE

Snow depth was collected at thousands of locations (> 22,000 points) while snow bulk density was measured at hundreds of locations (Table 2, Figure 2). The number of observations varied from year to year, with the greatest number of snow depth observations being collected in 2017, followed by 2019, and with 2018 having a similar number of observations to 2019 in the Teller watershed. In 2017, the survey was not planned for Kougarok, and in 2019, weather concerns prohibited safe travel to the study site. Note that while the Teller watershed's 2017 survey collected the most points, the 2019 survey was the most spatially extensive (Figure A3). No snowfall occurred during the any of the 2017-2019 end-of-winter snow surveys. The average snow depth at the Teller watershed was observed to be lowest in 2017, and similar depths were noted in 2018 and 2019 (109.0 cm, 106.4 cm, respectively). The average snow depth at the Kougarok Hillslope in 2018 was 75.3 cm. SWE was estimated to be lowest in 2017, lower at the Kougarok Hillslope versus Teller in 2018, and the highest SWE values were recorded in 2019 for Teller (Table 2). High spatial variability in snow depth was measured (Table 2, Figure 2). Snow depth ranged from 2 to 300 cm (mean 89 cm, standard deviation (SD) 40 cm, coefficient of variation (CV) 0.45). Snow density measurements ranged from 0.146 to 0.433 g cm^{-3} (mean density of 0.300 g cm^{-3}). SWE was strongly correlated with snow depth, with correlation coefficients ranging from 0.95 to 0.97 (Figure 3a). Snow density positively correlated with snow depth with correlation coefficients from 0.51 to 0.59 (Figure 3b) with higher correlations deeper than 60 cm. SWE was calculated from snow depth and snow density as described previously. SWE ranged from 3.3 to 90.5 cm (with mean 27.4 cm, SD 14.6 cm).

Deleted: 7
Deleted: 2
Deleted: 275
Deleted: 6
Deleted: 37
Deleted: 3
Deleted: 16
Deleted: 51
Deleted: 288
Deleted: and CV of
Deleted: 0.4
Deleted: 86.
Deleted: 7
Deleted: 5
Deleted: 2
Deleted: 4

4.3 Topographic, Vegetation, and Wind Features

Topographic features in the study sites illustrate the variation across the landscape (Figure 4a, b). Elevational gradients are strongest at the Teller watershed, topping out at 300 m, while Kougarok's dome-like feature is approximately 100 m. Slopes at the Teller watershed are steeper in the middle of the basin and along the stream banks, while at Kougarok Hillslope slopes are shallower on the west and steeper to the east, with overall more gentle elevational gradients than Teller. TPI highlights dominant features such as the terraces and risers, the stream bank in the Teller watershed, and the top of the dome at the Kougarok Hillslope.

Wind and aspect factors illustrate that the Teller watershed's aspects are largely unidirectional (south, south east facing), while the Kougarok Hillslope's west hillslope is predominantly S facing, with a north facing slope on the east side over the crest of the dome (Figure 4a, b). NDVI patterns in July and August are reflective of taller stature shrub patches located across the study sites (Figure 1, Figure 5, Boelman et al., 2011). Vegetation maps illustrate the differences in the two study sites. The Teller watershed contains low-to-mid-slope willow-birch and willow shrub complexes, with Ericaceous dwarf shrub tundra and wet meadows located in the upper slopes. At the Teller watershed, an ANOVA test of NDVI versus vegetation type (those with >1% of total land cover) indicate significant differences ($p < 0.0001$), with the highest NDVI values occurring in willow shrubs (Table A1). A Tukey's HSD test showed significant differences among all vegetation types ($p < 0.0001$, Figure 5). The Kougarok Hillslope's slopes are largely mixed shrub-sedge tussock tundra, with patches of alder-willow shrubs, dryas-lichen dwarf shrub tundra, birch-Ericaceous-lichen shrub tundra, and willow-birch shrub on the east slope where no snow measurements were taken in 2018 (Konduri and Kumar, 2021).

4.4 Model Optimization and Testing

Several different tests were undertaken to determine the optimal features used in the final model; we tested this optimization on both sites and for all years. We tested the random forest model to determine the optimal scale of TPI, with a smoothing-average window from 55 m to 505 m (Figure A1). The TPI at a scale of 155 m corresponded to the best model performance. Further, the random forest model was used to determine which vegetation features to use in the final model. We tested the model using NDVI, vegetation type as twelve one-hot-encoded categorical features, and vegetation type as a continuous feature ranked by observed SWE, as well as combinations of these three features (Figure A4). We found that NDVI and the continuous ranking of vegetation type performed the best so these two features were included in the final model. However, while the spatial pattern of vegetation type improved model performance, the relative order of the vegetation type ranking did not appear to be important, since other randomized rankings performed similarly well (Figure A4). It should be noted that while NDVI and continuous vegetation performed the best, the differences across all tests were similar (R^2 values between ~0.86-0.87).
The variance inflation factors of the input features are shown in Figure 6. Since the variance inflation factors are all well below the accepted threshold of 5 (Karimi et al., 2019) and the largest correlation coefficient between two input variables (0.49 for microtopography and TPI) is well below the accepted threshold of 0.70, collinearity is not expected to severely distort model estimation and predictions (Dormann et al., 2013).

4.5 SWE Prediction

The SWE prediction model results for the linear regression, GAM, and random forest for the individual sites and years are shown in Table 3. In general, linear regression performed the worst (R^2 ranging from 0.29 to 0.44), random forest performed the best (R^2 ranging from 0.72 to 0.92), and GAM performed in between the other two models (R^2 ranging from 0.45 to 0.72). The spatial maps of predicted SWE for the three different models are shown in Figure A5 for the Teller watershed and Figure A6 for the Kougarok Hillslope. Because of the success in simulating SWE for the study sites and years using random forest, we focus the remainder of our study on the random forest results to discuss the implications and the driving factors that are ranked to be most important for prediction of SWE in the study sites.
The random forest model results for training and testing data are given in Table 4. The final random forest model which includes data from all years and sites captures approximately 86 % of the variance in SWE and has an RMSE of 5.81 cm on the test set. The scatter plot of predicted and actual SWE measurements of the test set from the final model (Figure 7) shows a linear trend. In comparison to the $y = x$ line, the linear fit in Figure 7 shows the model slightly overestimates low SWE measurements and underestimates high

SWE measurements, which could be due to the tendency of random forests to decrease the variance by averaging across many trees.

We also considered the random forest model developed using the individual study sites and years as a feature to predict SWE. The iterations of the random forest model where we considered only the Teller watershed and all years performed slightly better than the final model, with an R^2 of 0.86 and an RMSE of 5.78 cm for the Teller watershed (Table 4). The random forest models trained on individual years and sites ranged in model performance, as shown in Table 4.

In Figure 8, we illustrate the spatially predicted SWE from the final random forest model for the Teller watershed (2017-2019) and the Kougarok Hillslope (2018). SWE values across the basin reflect the year-to-year variability in the amount of precipitation that fell on the study sites. However, we observe that SWE is variable across the Teller watershed and the Kougarok Hillslope around major landscape features, such as the stream bed and terraces and risers within the Teller watershed. Another location where SWE appears to be higher is just below the dome on the west hillslope in the Kougarok Hillslope. The SWE patterns in SWE also reflect areas of higher NDVI values, where shrubs are identified as darker patches in both the Teller watershed and the Kougarok Hillslope (see Figure 4).

The spatial error, calculated as the predicted minus the observed for the upper and lower quartiles of error in the random forest SWE prediction is shown in Figure A7 for the Teller watershed for each year, and Figure A8 for the Kougarok Hillslope for 2018. In these figures we observe that spatial error varies from year to year, but is not spatially systematic. Errors are higher in the years where there was lower SWE in the basin, such as in 2018 compared to the lower SWE year of 2017 in the Teller watershed, when the survey resolution was similar (Figure A3). In 2019, the survey captured higher SWE values but similar CV compared to 2018, while 2019 survey extent was broader with a finer spatial resolution (Figure A3). A more detailed consideration of the spatial error in these data sets is beyond the scope of this work.

Deleted: 6

Deleted: 7

Deleted: higher

Deleted: In

Deleted: a finer spatial resolution and thus we expect greater spatial variability and higher error

4.6 Feature Importance

Figure 9 shows the impurity and permutation feature importance results for the final random forest model. Both of the importance metrics provide similar results, and the same variable ranking with the exception of permutation importance metric ranking elevation as higher than NDVI, and impurity ranking NDVI higher than elevation. TWP is the key primary factor driving SWE distribution. Overall, NDVI, elevation, and TPI are the most important secondary features for predicting SWE distribution at our study sites. Features such as slope, vegetation type, wind/aspect factor, and microtopography are ranked as the least important in SWE prediction.

4.7 SWE Correlations between Years

Heatmaps that illustrate significant SWE correlations between years 2017-2019 for the Teller watershed are shown in Figure 10, based on random forest modeled SWE trained with data from all years and sites. The weakest correlations are for the years with low (2017) and high (2019) SWE (Pearson correlation coefficient r = 0.67), with stronger correlations between the two higher SWE years (2018 and 2019, r = 0.89). These correlations tend highlight the consistency in SWE values for the study site across highly variable climate conditions.

5 Discussion

Changes in snowpack characteristics have important implications for a changing Arctic and are anticipated to be a major driver of ecosystem shifts (Bjerke et al., 2015; Cooper, 2014), water and energy balances (AMAP, 2019; Pulliainen et al., 2020), and biodiversity changes (Niittynen et al., 2018; Riseth et al., 2011). Changes in snow have implications for Arctic communities where snow may impact many resources (Huntington et al., 2004), and for global climate change (Overland et al., 2019) and carbon cycles (Rogers et al., 2011; Arndt et al., 2020). Thus, understanding how to better model snow distribution and the important features involved in snow distribution is fundamental to improving how we interpret, and plan for, changing Arctic snow in the future (Zhu et al., 2021; Kouki et al., 2021; Mudryk et al., 2020).

5.1 Snow Depth, SWE, and Density Observations

Snow depth and snow density observations collected from two small study sites located on the Seward Peninsula of Alaska comprise an extensive dataset that provides a reliable and representative estimate of the variation in SWE in this region. Snow depth showed high variability at both study sites and years, with a medium level of variability compared to the variability range reported for other Arctic regions (Bruland, Sand, and Killingtveit 2001; Hannula et al. 2016; Dvornikov et al. 2015; Stuefer, Kane, and Liston 2013; Homan and Kane 2015; Sturm et al. 2010). Compared to snow depth, snow density showed relatively low variability, similar to the findings for other non-forest Arctic areas by Homan and Kane (2015) and Hannula et al. (2016).

Consistent with previous studies (Homan and Kane 2015; Assini and Young 2012; Dvornikov et al. 2015; Sturm et al. 2010), there was a high correlation between snow depth and SWE in our study, confirming that SWE is more closely linked to snow depth than to snow density. Sturm et al. (2010) suggested a nonlinear relationship between snow depth and density for a large region of the Northern Hemisphere, while Homan and Kane (2016) did not find any relationship between snow depth and density for a 200 by 240 km region of Alaska's Central Arctic Slope. Our study showed an overall positive linear correlation between snow depth and density, indicating that snow depth has some control on the snow density for the study sites. However, snow depth and density showed no relationship for shallow snow (<60 cm), while the linear relationship for deeper snow (>60 cm) was strongest at most sites and years (with the exception of 2018 at Kougarok), consistent with what has been found for a study region consisting of a variety of landscapes in Saskatchewan, Canada (Shook, 1997). The nonlinear relationship between snow depth and density found in the aforementioned work may be due to the fact that the snow surveys documented were conducted for different climate and landscape classes where snow density was largely controlled by climate (wind and temperature) and landscape classes (such as taiga, tundra, mountain, coast). Whereas, a linear relationship between snow depth and density was observed in this study because the climate and landscape in the small study sites are more homogenous.

Finally, when we considered SWE between three study years at the Teller watershed site, we found correlations across the years. This is consistent with other research on snow repeat patterns (Sturm and Wagner, 2010; Liston, 2004; Kirnbauer and Blöschl, 1994; Deems et al., 2008; Homan and Kane, 2015; Woo and Young, 2004; König and Sturm, 1998; Rees et al., 2014; Dozier et al., 2016; Erickson et al., 2005; Winstral and Marks, 2014), indicating that there are driving factors that influence snow distribution consistent across the year-to-year snow variability (i.e., high and lows).

5.2 SWE Modelling and Prediction

The patterns of our SWE maps illustrate the power of utilizing random forest tools over linear methods of estimating SWE distributions (e.g., Broxton et al, 2019, Revuelto et al. 2020, King et al. 2020). When compared to linear and GAM models, we found that random forests significantly outperformed those models. This is mostly likely due to the fact that SWE distributions are controlled by highly non-linear interactions between topography and vegetation characteristics, thus the flexibility offered by the random forest model can more accurately account for these interactions. While GAM models have been applied successfully to estimate non-linear relationships in snow depth (López-Moreno and Nogués-Bravo, 2005) in the Spanish Pyrenees, they were not as successful at predicting SWE in our study compared to the random forests model. Random forest also allowed us to test different hypotheses of configurations for the model, determining clearly the success of those configuration and features combinations.

5.3 Features Impacting Snow Distribution

Two different techniques measuring importance in the random forest model gave similar results, which were that the greatest controls on SWE were TWP (i.e., precipitation representing climate variability), followed by the key secondary factors NDVI, elevation, and TPI. These results in general were consistent with those in previous studies in terms of how those factors affected snow distribution in the Arctic, i.e., more snow was accumulated in the areas with tall shrubs (Sturm et al., 2001b, a; Sturm and Wagner, 2010; Sturm et al., 2005), at higher elevations (Dozier et al., 2016; Homan and Kane, 2015), and within features such as the stream bed and at the bottom of terraces and risers (Gisnås et al., 2014; Grünberg et al., 2020). NDVI in our study tend to reflect the taller shrub patterns present in the landscape. In this work, we applied the NDVI estimate from July, near the end of the vegetation growth period, which has been corroborated in previous work to be the peak NDVI (Boelman et al., 2011). Although vegetation type in our work was slightly less important than NDVI in our modeling overall, we found taller stature shrub types had higher NDVI values. Vegetation type has been noted to play a primary role in end-of-winter snow depth patterning and is also strongly related to variability in winter and spring soil temperatures in the Arctic (Grünberg et al., 2020). Because of the strong feedbacks between snow, shrubs, and climate (Boike et al., 2019; Sturm et al., 2001a, b), this finding indicates the importance of understanding vegetation dynamics in sub-Arctic regions.

Interestingly, in a study examining snow depth distribution using random forests across a low relief high Arctic site located in Nunavut, Canada, Meloche et al. (2022) indicate that vegetation ecotype was the least important factor in snow depth distributions. This was largely explained by the fact that ecotypes are correlated with topographic parameters (such as TPI), and this correlation led to low feature importance in the models. When the authors' removed the TPI parameters in their modeling, ecotype became more important, while model skill was reduced. Our study differs from this work in that it was undertaken using SWE, was carried out in a sub-arctic region with moderate relief, and we used a vegetation index (NDVI) to directly represent shrub influence on the snow, rather than utilizing an ecotype approach. Hence, the ecotype method applied in conjunction with the random forests approach by Meloche et al. 2022 reduced the influence vegetation had on snow distribution, which made vegetation appear unimportant as a factor in the snow depth distributions. Rather, as our research indicates, vegetation characteristics are an important secondary driving variable in snow distributions at Arctic sites.

Our results showed that elevation effects are a dominant factor driving snow distributions at our study sites. We observe an increase in SWE at higher elevations at the Teller watershed due to a slight orographic effect, consistent with the study for a small high-arctic glacier Svenbreen (Małecki 2015). However, an apparent orographic effect also occurs because wind blows snow from outside the catchment into the upper wetland meadow. Homan and Kane (2015) discussed a relationship between snow and elevation below specific elevation bands, above which snow is controlled by moisture availability, however the Teller watershed's maximum elevation (300 m) is above the value, suggesting that the threshold may change due to other local or regional factors. However, we also observed a decrease in SWE at the top of the Kougarok Hillslope, where snow is removed completely from the upper windswept top (Assini and Young 2012; Shook and Gray 1996; Homan and Kane 2015).

TPI in our model was found to be the third most important variable, indicating the importance of coarse-scale features in the sub-Arctic landscapes of the Seward Peninsula. These coarse-scale features including stream banks and terrace risers are areas of topographic variability where shrubs grow and snow accumulates. Thus, they act as hydrology focal points in the basins where higher enhanced soil moistures and soil warming, and associated increased ecological productivity, can occur (Westergaard-Nielsen et al., 2017). These are also features that act to entrain snow distributed by wind. Indeed, recent research into snowdrift landscape patterns in the Arctic have found that wind transports snow into coarse-scale features called drift traps, including stream beds, lake features, outcrop features and more. These drift traps contain as much as 40% of SWE found on the landscape and play a significant role in the distribution of snow in the Arctic (Parr et al., 2020). Meloche et al. (2022) also found that topographic parameters of TPI and up-wind slope index were the two most important features in random forest modeling of snow depth distribution for a low-relief high arctic basin.

In our study, microtopography was found to be one of the least important factors driving snow. However, microtopographic patterns were highly correlated with curvature. In research from Dvornikov et al. (2015), curvature was found to have a dominant control on the snow depth at a shrub tundra area in Central Yamal and there was a positive correlation between shrub heights and snow depth was observed for convex slopes. However, research in fine-scale polygonal tundra sites of the high Arctic, microtopography was found to be an important control on snow (Wainwright et al. 2017). Our study utilized DEMs with a 5 m resolution, with the caveat that microtopography features dominant in this landscape (e.g., drainage paths, terraces) range from centimeters to meters in scale. Thus, we require finer scale DEM sources to investigate this in more detail (Adams et al., 2018; Harder et al., 2020; Revuelto et al., 2021).

5.4 Future Work

The ability to model and predict SWE is important on a number of fronts. First of all, we intend to utilize these findings to compare them with physics-based modeling efforts, as well as for future machine-learning efforts. Our models are being used for investigation of sub-grid SWE variability in E3SM's ELM (Caldwell et al., 2019; Bisht et al., 2018), along with the investigation of ecosystem-type constructs for upscaling of SWE.

Several of our findings require further investigation to clearly understand their importance for snow distribution. For example, NDVI was an important parameter in our model predicting snow distribution, but we do not know which vegetation characteristics (e.g., shrub height, density, allometry, or wetness) that NDVI represents, making it harder to extrapolate the importance of vegetation characteristics for snow distribution at other sites across the Arctic and sub-Arctic. Further, there are likely relationships between TPI and NDVI that should be investigated, including at smaller spatial scales. We are also considering the application of more advanced wind functions (Winstral et al., 2002) in our models, and implementing a physically-based wind model for comparison and testing against our statistical models (Crumley et al., 2021; Liston, 2004).

We used multiple study sites with varying characteristics across multiple climate years to develop our snow distribution estimates. Because of this unique data set, we were able to develop robust machine-learning-based models that we hypothesize are representative of broader SWE patterns across time and space. However, these theories require additional validation at more locations. This hypothesis testing will be incorporated into current and future work that will be carried out using broader observations of SWE and snow depths that can be compared with other remote sensed SWE data products.

6 Conclusion

The extensive snow depth and density dataset from this study is of high value for calibrating and validating physically-based models of snow distribution. As the patterns of snow distribution for a given location are similar from year to year, the spatial patterns of snow distribution characterized in this study can be used to represent the typical patterns of snow distribution and model the relative spatial patterns of snow distribution for other years for the study sites and across the region. Linear relationships between snow depths greater than 60 cm and snow density revealed homogeneity in the study sites. Snow depth, on the other hand, varied considerably with strong linear relationship to SWE.

We found that random forest models could simulate the SWE distribution most accurately when compared to linear and GAM model approaches, and we were able to simulate the distribution of SWE across the landscape of these small sub-Arctic study sites. The results of the statistical model are useful for understanding the surface water hydrology during spring snowmelt and explaining differences in permafrost distribution and active layer depth, which have an impact on groundwater hydrology. Using the

machine-learning-based model random forest, we were able to determine which factors—in addition to precipitation—were most important at these sub-Arctic study sites for SWE distribution. These factors were NDVI, followed by two topography indexes, elevation and TPI.
These factors were NDVI, followed by two topography indexes, elevation, and TPI.

7 Data Availability

Snow observations collected for this study are available on a publicly available repository, the NGEE Arctic data portal accessible at https://ngee-arctic.ornl.gov/ (Wilson et al., 2020b; Bennett et al., 2020; Wilson et al., 2020a). Data used as inputs to simulate SWE distribution is available (ifSAR, http://ifsar.gina.alaska.edu/), while the vegetation data are also available on the NGEE Arctic portal (Konduri and Kumar, 2021). Python modeling codes for developing linear, GAM, and random forest models will be posted on the NGEE Arctic portal.

8 Author Contribution

KEB conceptualized the random forest modeling, wrote original text, editing, review response, and is the PI of the NGEE Arctic project in 2021-present; GM completed all modeling and analysis, original writing, and figure development; RB was instrumental in field logistics, and observation data collection; MC worked on early modeling iterations, literature review, and original writing; ERL, JBD, JK, and MN provided data collection, and data management; RC, edited and contributed to composition; SD completed additional NDVI analyses to address reviewer comments; BD, SW, CI, and JK provided feedback and edited composition, WRB provided logistical support and observational data collection, along with project management of all collaborative efforts between LANL and UAF; and CJW edited, designed and implemented overall project logistics for field survey, conceptualized earlier implementations of the study design for the composition, oversaw all work, and was PI of the project during the implementation of this work.

9 Competing Interests

The authors declare that they have no conflict of interest.

10 Acknowledgements

Funding for this research was provided by the Department of Energy Office of Science, Office of Biological and Environmental Research through Next Generation Ecosystem Experiment (NGEE) Arctic project. We gratefully acknowledge Mary's Igloo, Sitnasuak, and Council Native Corporation guidance and for allowing us to conduct our research on their traditional lands**.** The authors gratefully acknowledge the contributions of Haruko Wainwright from Berkeley National Lab in earlier versions of this work. We would also like to thank two anonymous reviewers for their thoughtful and detailed comments on the first iteration of the paper.

11 References

Adams, M. S., Bühler, Y., and Fromm, R.: Multitemporal accuracy and precision assessment of unmanned aerial system photogrammetry for slope-scale snow depth maps in Alpine terrain, Pure Appl. Geophys., 175, 3303–3324, 2018.

AMAP: An Update to Key Findings of Snow, Water, Ice and Permafrost in the Arctic (SWIPA) 2017, Arct. Monit. Assess. Programme AMAP Oslo Nor., 1–12, 2019.

Anderton, S. P., White, S. M., and Alvera, B.: Evaluation of spatial variability in snow water equivalent for a high mountain catchment, Hydrol. Process., 18, 435–453, https://doi.org/10.1002/hyp.1319, 2004.

Arndt, K. A., Lipson, D. A., Hashemi, J., Oechel, W. C., and Zona, D.: Snow melt stimulates ecosystem respiration in Arctic ecosystems, Glob. Change Biol., 26, 5042–5051, https://doi.org/10.1111/gcb.15193, 2020.

Atchley, A. L., Painter, S. L., Harp, D. R., Coon, E. T., Wilson, C. J., Liljedahl, A. K., and Romanovsky, V.: Using field observations to inform thermal hydrology models of permafrost dynamics with ATS (v0. 83), Geosci. Model Dev., 8, 2701, 2015.

Atchley, A. L., Coon, E. T., Painter, S. L., Harp, D. R., and Wilson, C. J.: Influences and interactions of inundation, peat, and snow on active layer thickness, Geophys. Res. Lett., 43, 5116–5123, 2016.

Bair, E. H., Abreu Calfa, A., Rittger, K., and Dozier, J.: Using machine learning for real-time estimates of snow water equivalent in the watersheds of Afghanistan, The Cryosphere, 12, 1579–1594, https://doi.org/10.5194/tc-12-1579-2018, 2018.

Bennett, K., Bolton, R., Lathrop, E., Dann, J., Miller, G., Nutt, M., and Wilson, C.: End-of-Winter Snow Depth, Temperature, Density, and SWE Measurements at Teller Road Site, Seward Peninsula, Alaska, 2019, https://doi.org/10.5440/1798170, 2020.

Berg, N. H.: Blowing snow at a Colorado alpine site: measurements and implications, Arct. Alp. Res., 18, 147–161, 1986.

Bisht, G., Riley, W. J., Hammond, G. E., and Lorenzetti, D. M.: Development and evaluation of a variably saturated flow model in the global E3SM Land Model (ELM) version 1.0, Geosci. Model Dev., 11, 4085–4102, 2018.

Bjerke, J. W., Tømmervik, H., Zielke, M., and Jørgensen, M.: Impacts of snow season on ground-ice accumulation, soil frost and primary productivity in a grassland of sub-Arctic Norway, Environ. Res. Lett., 10, 095007, 2015.

Boelman, N. T., Gough, L., McLaren, J. R., and Greaves, H.: Does NDVI reflect variation in the structural attributes associated with increasing shrub dominance in arctic tundra?, Environ. Res. Lett., 6, 1–12, 2011.

Boike, J., Nitzbon, J., Anders, K., Grigoriev, M., Bolshiyanov, D., Langer, M., Lange, S., Bornemann, N., Morgenstern, A., Schreiber, P., Wille, C., Chadburn, S., Gouttevin, I., Burke, E., and Kutzbach, L.: A 16-year record (2002–2017) of permafrost, active-layer, and meteorological conditions at the Samoylov Island Arctic permafrost research site, Lena River delta, northern Siberia: an opportunity to validate remote-sensing data and land surface, snow, and permafrost models, Earth Syst Sci Data, 11, 261–299, https://doi.org/10.5194/essd-11-261-2019, 2019.

Breiman, L.: Random forests, Mach. Learn., 45, 5–32, 2001.

Broxton, P. D., Van Leeuwen, W. J., and Biederman, J. A.: Improving snow water equivalent maps with machine learning of snow survey and lidar measurements, Water Resour. Res., 55, 3739–3757, 2019.

Busey, R. C., Hinzman, L. D., Cassano, J., and Cassano, E.: Permafrost distributions on the Seward Peninsula: past, present, and future, Ninth International Conference on Permafrost, Fairbanks, AK, 215–220, 2008.

Busey, R. C., Bolton, W. R., Wilson, C. J., and Cohen, L.: Surface meteorology at Teller site stations, Seward Peninsula, Alaska, ongoing from 2016, https://doi.org/10.5440/1437633, 2017.

Caldwell, P. M., Mametjanov, A., Tang, Q., Van Roekel, L. P., Golaz, J., Lin, W., Bader, D. C., Keen, N. D., Feng, Y., and Jacob, R.: The DOE E3SM coupled model version 1: Description and results at high resolution, J. Adv. Model. Earth Syst., 11, 4095–4146, 2019.

Callaghan, T. V., Johansson, M., Brown, R. D., Groisman, P. Ya., Labba, N., Radionov, V., Bradley, R. S., Blangy, S., Bulygina, O. N., Colman, J. E., Essery, R. L. H., Forbes, B. C., Forchhammer, M. C., Golubev, V. N., Honrath, R. E., Juday, G. P., Meshcherskaya, A. V., Phoenix, G. K., Pomeroy, J., Rautio, A., Robinson, D. A., Schmidt, N. M., Serreze, M. C., Shevchenko, V. P., Shiklomanov, A. I., Shmakin, A. B., Sköld, P., Sturm, M., Woo, M.-K., and Wood, E. F.: Multiple effects of changes in arctic snow cover, Ambio, 40, 32–45, https://doi.org/10.1007/s13280-011-0213-x, 2011.

Cooper, E. J.: Warmer shorter winters disrupt Arctic terrestrial ecosystems, Annu. Rev. Ecol. Evol. Syst., 45, 271–295, 2014.

Crumley, R. L., Hill, D. F., Wikstrom Jones, K., Wolken, G. J., Arendt, A. A., Aragon, C. M., Cosgrove, C., and Community Snow Observations Participants: Assimilation of citizen science data in snowpack modeling using a new snow data set: Community Snow Observations, Hydrol. Earth Syst. Sci., 25, 4651–4680, https://doi.org/10.5194/hess-25-4651-2021, 2021.

Deems, J. S., Fassnacht, S. R., and Elder, K. J.: Interannual consistency in fractal snow depth patterns at two Colorado mountain sites, J. Hydrometeorol., 9, 977–988, 2008.

Dixon, D. and Boon, S.: Comparison of the SnowHydro snow sampler with existing snow tube designs, Hydrol. Process., 26, 2555–2562, https://doi.org/10.1002/hyp.9317, 2012.

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., Gruber, B., Lafourcade, B., and Leitão, P. J.: Collinearity: a review of methods to deal with it and a simulation study evaluating their performance, Ecography, 36, 27–46, 2013.

Dozier, J., Bair, E. H., and Davis, R. E.: Estimating the spatial distribution of snow water equivalent in the world's mountains, Wiley Interdiscip. Rev. Water, 3, 461–474, 2016.

Dvornikov, Y., Khomutov, A., Mullanurov, D., Ermokhina, K., Gubarkov, A., and Leibman, M.: GIS and field data based modelling of snow water equivalent in shrub tundra, Fennia, 193, 53–65, https://doi.org/10.11143/46363, 2015.

Erickson, T. A., Williams, M. W., and Winstral, A.: Persistence of topographic controls on the spatial distribution of snow in rugged mountain terrain, Colorado, United States, Water Resour. Res., 41, 2005.

Essery, R. and Pomeroy, J.: Vegetation and topographic control of wind-blown snow distribution in distributed and aggregated simulations for an Arctic tundra basin, J. Hydrometeorol., 5, 735–744, 2004.

Evans, B. M., Walder, D. A., Benson, C. S., A., N. E., and Petersen, G. W.: Spatial interrelationships between terrain, snow distribution and vegetation patterns at an arctic foothills site in Alaska, Holarct. Ecol., 12, 270–278, 1989.

Fletcher, C. G., Kushner, P. J., Hall, A., and Qu, X.: Circulation responses to snow albedo feedback in climate change, Geophys. Res. Lett., 36, https://doi.org/10.1029/2009GL038011, 2009.

Forchhammer, M. C., Schmidt, N. M., Høye, T. T., Berg, T. B., Hendrichsen, D. K., and Post, E.: Population dynamical responses to climate change, Adv. Ecol. Res., 40, https://doi.org/10.1016/S0065-2504(08)00017-7, 2008.

Ford, J. and Bedford, B. L.: Hydrology of Alaskan wetlands, U.S.A, Arct. Alp. Res., 19, 209–229, 1987.

Franke, R., 1982. Scattered data interpolation: tests of some methods. Mathematics of computation, 38(157), pp.181-200.

Freudiger, D., Kohn, I., Seibert, J., Stahl, K., and Weiler, M.: Snow redistribution for the hydrological modeling of alpine catchments, WIREs Water, 4, e1232, https://doi.org/10.1002/wat2.1232, 2017.

Gisnås, K., Westermann, S., Schuler, T. V., Litherland, T., Isaksen, K., Boike, J., and Etzelmüller, B.: A statistical approach to represent small-scale variability of permafrost temperatures due to snow cover, The Cryosphere, 8, 2063–2074, 2014.

Gregorutti, B., Michel, B., and Saint-Pierre, P.: Correlation and variable importance in random forests, Stat. Comput., 27, 659–678, https://doi.org/10.1007/s11222-016-9646-1, 2017.

Grünberg, I., Wilcox, E. J., Zwieback, S., Marsh, P., and Boike, J.: Linking tundra vegetation, snow, soil temperature, and permafrost, Biogeosciences, 17, 4261–4279, 2020.

Grünberg, I., Wilcox, E. J., Zwieback, S., Marsh, P., and Boike, J.: Linking tundra vegetation, snow, soil temperature, and permafrost, 30, n.d.

Grünewald, T., Stötter, J., Pomeroy, J. W., Dadic, R., Moreno Baños, I., Marturià, J., Spross, M., Hopkinson, C., Burlando, P., and Lehning, M.: Statistical modelling of the snow depth distribution in open alpine terrain, Hydrol Earth Syst Sci, 17, 3005–3021, https://doi.org/10.5194/hess-17-3005-2013, 2013.

Harder, P., Pomeroy, J. W., and Helgason, W. D.: Improving sub-canopy snow depth mapping with unmanned aerial vehicles: lidar versus structure-from-motion techniques, The Cryosphere, 14, 1919–1935, 2020.

Hinzman, L., Kane, D., Yoshikawa, K., Carr, A., Bolton, W., and Fraver, M.: Hydrological variations among watersheds with varying degrees of permafrost, 2003.

Hirashima, H., Ohata, T., Kodama, Y., Yabuki, H., Sato, N., and Georgiadi, A.: Nonuniform distribution of tundra snow cover in Eastern Siberia, J. Hydrometeorol., 5, 373–389, 2004.

Homan, J. W. and Kane, D. L.: Arctic snow distribution patterns at the watershed scale, Hydrol. Res., 46, 507–520, https://doi.org/10.2166/nh.2014.024, 2015.

Huntington, H., Callaghan, T., Fox, S., and Krupnik, I.: Matching Traditional and Scientific Observations to Detect Environmental Change: A Discussion on Arctic Terrestrial Ecosystems, AMBIO J. Hum. Environ., 33, 18–23, https://doi.org/10.1007/0044-7447-33.sp13.18, 2004.

Iversen, C., Breen, A., Salmon, V., VanderStel, H., and Wullschleger, S.: NGEE Arctic Plant Traits: Vegetation Plot Locations, Ecotypes, and Photos, Kougarok Road Mile Marker 64, Seward Peninsula, Alaska, 2016, Next Generation Ecosystems Experiment - Arctic, Oak Ridge National Laboratory (ORNL), Oak Ridge, TN (US); NGEE Arctic, Oak Ridge National Laboratory (ORNL), Oak Ridge, TN (United States), https://doi.org/10.5440/1346196, 2019.

Jaedicke, Ch. and Sandvik, A. D.: High resolution snow distribution data from complex Arctic terrain: a tool for model validation, Nat. Hazards Earth Syst. Sci., 2, 147–155, 2002.

Jafarov, E. E., Coon, E. T., Harp, D. R., Wilson, C. J., Painter, S. L., Atchley, A. L., and Romanovsky, V. E.: Modeling the role of preferential snow accumulation in through talik development and hillslope groundwater flow in a transitional permafrost landscape, Environ. Res. Lett., 13, 105006, https://doi.org/10.1088/1748-9326/aadd30, 2018.

Jenness, J.: Topographic Position Index (tpi_jen. avx) extension for ArcView 3. x, v. 1.3 a. Jenness Enterprises, 2006.

Jonas, T., Marty, C., and Magnusson, J.: Estimating the snow water equivalent from snow depth measurements in the Swiss Alps, J. Hydrol., 378, 161–167, https://doi.org/10.1016/j.jhydrol.2009.09.021, 2009.

Jorgenson, T., Yoshikawa, K., Kanevskiy, M., Shur, Y., Romanovsky, V., Marchenko, S., Grosse, G., Brown, J., and Jones, B.: Permafrost Characteristics of Alaska – A new permafrost map of Alaska, Ninth International Conference on Permafrost. Abstract., 2008.

Karimi, S. S., Saintilan, N., Wen, L., and Valavi, R.: Application of Machine Learning to Model Wetland Inundation Patterns Across a Large Semiarid Floodplain, Water Resour. Res., 55, 8765–8778, https://doi.org/10.1029/2019WR024884, 2019.

King, F., Erler, A. R., Frey, S. K., and Fletcher, C. G.: Application of machine learning techniques for regional bias correction of snow water equivalent estimates in Ontario, Canada, Hydrol. Earth Syst. Sci., 24, 4887–4902, https://doi.org/10.5194/hess-24-4887-2020, 2020.

Kirnbauer, R. and Blöschl, G.: How similar are snow cover patterns from year to year?, Dtsch. Gewasserkundliche Mitteilungen, 37, 113–121, 1994.

Konduri, S. and Kumar, J.: Hyperspectral remote sensing based vegetation communities around NGEE-Arctic intensive research watersheds at Seward Peninsula, Alaska, 2019. Next Generation Ecosystem Experiments Arctic Data Collection, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tennessee, USA. Dataset accessed on Nov 4, 2021 at https://doi.org/10.5440/1828604., 2021.

König, M. and Sturm, M.: Mapping snow distribution in the Alaska Arctic using aerial photography and topographic relationships, Water Resour. Res., 34, 3471–3483, 1998.

Kouki, K., Räisänen, P., Luojus, K., Luomaranta, A., and Riihelä, A.: Evaluation of Northern Hemisphere snow water equivalent in CMIP6 models with satellite-based SnowCCI data during 1982–2014, Cryosphere Discuss., 1–32, https://doi.org/10.5194/tc-2021-195, 2021.

Léger, E., Dafflon, B., Robert, Y., Ulrich, C., Peterson, J. E., Biraud, S. C., Romanovsky, V. E., and Hubbard, S. S.: A distributed temperature profiling method for assessing spatial variability in ground temperatures in a discontinuous permafrost region of Alaska, The Cryosphere, 13, 2853–2867, https://doi.org/10.5194/tc-13-2853-2019, 2019.

Liaw, A. and Wiener, M.: Classification and regression by randomForest, R News, 2, 18–22, 2002.

Liston, G. E.: Representing subgrid snow cover heterogeneities in regional and global models, J. Clim., 17, 1381–1397, 2004.

Liston, G. E. and Elder, K.: A distributed snow-evolution modeling system (SnowModel), J. Hydrometeorol., 7, 1259–1276, 2006.

Liston, G. E. and Sturm, M.: A snow-transport model for complex terrain, J. Glaciol., 44, 498–516, 1998.

Liston, G. E., Haehnel, R. B., Sturm, M., Hiemstra, C. A., Berezovskaya, S., and Tabler, R. D.: Instruments and Methods. Simulating complex snow distributions in windy environments using SnowTran-3D, J. Glaciol., 53, 241–256, https://doi.org/10.3189/172756507782202865, 2007.

Liu, C., Huang, X., Li, X., and Liang, T.: MODIS Fractional Snow Cover Mapping Using Machine Learning Technology in a Mountainous Area, Remote Sens., 12, 962, https://doi.org/10.3390/rs12060962, 2020.

López-Moreno, J. I. and Nogués-Bravo, D.: A generalized additive model for the spatial distribution of snowpack in the Spanish Pyrenees, Hydrol. Process., 19, 3167–3176, https://doi.org/10.1002/hyp.5840, 2005.

López-Moreno, J. I., Latron, J., and Lehmann, A.: Effects of sample and grid size on the accuracy and stability of regression-based snow interpolation methods, Hydrol Process, 15, 2009.

López-Moreno, J. I., Leppänen, L., Luks, B., Holko, L., Picard, G., Sanmiguel-Vallelado, A., Alonso-González, E., Finger, D. C., Arslan, A. N., Gillemot, K., Sensoy, A., Sorman, A., Ertaş, M. C., Fassnacht, S. R., Fierz, C., and Marty, C.: Intercomparison of measurements of bulk snow density and water equivalent of snow cover with snow core samplers: Instrumental bias and variability induced by observers, Hydrol. Process., 34, 3120–3133, https://doi.org/10.1002/hyp.13785, 2020.

Louppe, G., Wehenkel, L., Sutera, A., and Geurts, P.: Understanding variable importances in forests of randomized trees, Adv. Neural Inf. Process. Syst., 26, 2013.

Manning, J. A. and Garton, E. O.: Reconstructing historical snow depth surfaces to evaluate changes in critical demographic rates and habitat components of snow-dependent and snow-restricted species, Methods Ecol. Evol., 3, 71–80, https://doi.org/10.1111/j.2041-210X.2011.00144.x, 2012.

Mauritz, M., Bracho, R., Celis, G., Hutchings, J., Natali, S. M., Pegoraro, E., Salmon, V. G., Schädel, C., Webb, E. E., and Schuur, Edward. A. G.: Nonlinear CO2 flux response to 7 years of experimentally induced permafrost thaw, Glob. CHANGE Biol., 23, 3646–3666, https://doi.org/10.1111/gcb.13661, 2017.

Meloche, J., Langlois, A., Rutter, N., McLennan, D., Royer, A., Billecocq, P. and Ponomarenko, S., High-resolution snow depth prediction using Random Forest algorithm with topographic parameters: a case study in the Greiner Watershed, Nunavut. *Hydrological Processes*, p.e14546.

McCaully, R. E., Arendt, C. A., Newman, B. D., Salmon, V. G., Heikoop, J. M., Wilson, C. J., Sevanto, S., Wales, N. A., Perkins, G. B., Marina, O. C., and Wullschleger, S. D.: High Temporal and Spatial Nitrate Variability on an Alaskan Hillslope Dominated by Alder Shrubs, Cryosphere Discuss., 1–24, https://doi.org/10.5194/tc-2021-166, 2021.

McFadden, J. P., Liston, G. E., Sturm, M., Pielke, R. A., and Chapin, F. S.: Interactions of shrubs and snow in arctic tundra: measurements and models, Sixth scientific assembly of the International Association of Hydrological Sciences, Maastricht, The Netherlands, 317–325, 2001.

Mendoza, P. A., Shaw, T. E., McPhee, J., Musselman, K. N., Revuelto, J., and MacDonell, S.: Spatial distribution and scaling properties of lidar-derived snow depth in the extratropical Andes, Water Resour. Res., 56, e2020WR028480, 2020.

Mott, R., Schirmer, M., and Lehning, M.: Scaling properties of wind and snow depth distribution in an Alpine catchment, J. Geophys. Res. Atmospheres, 116, 2011.

Mott, R., Vionnet, V., and Grünewald, T.: The Seasonal Snow Cover Dynamics: Review on Wind-Driven Coupling Processes, Front. Earth Sci., 6, https://doi.org/10.3389/feart.2018.00197, 2018.

Mudryk, L., Santolaria-Otín, M., Krinner, G., Ménégoz, M., Derksen, C., Brutel-Vuilmet, C., Brady, M., and Essery, R.: Historical Northern Hemisphere snow cover trends and projected changes in the CMIP6 multi-model ensemble, The Cryosphere, 14, 2495–2514, https://doi.org/10.5194/tc-14-2495-2020, 2020.

Niittynen, P., Heikkinen, R. K., and Luoto, M.: Snow cover is a neglected driver of Arctic biodiversity loss, Nat. Clim. Change, 8, 997–1001, https://doi.org/10.1038/s41558-018-0311-x, 2018.

Overland, J., Dunlea, E., Box, J. E., Corell, R., Forsius, M., Kattsov, V., Olsen, M. S., Pawlak, J., Reiersen, L.-O., and Wang, M.: The urgency of Arctic change, Polar Sci., 21, 6–13, 2019.

Painter, S. L., Coon, E. T., Atchley, A. L., Berndt, M., Garimella, R., Moulton, J. D., Svyatskiy, D., and Wilson, C. J.: Integrated surface/subsurface permafrost thermal hydrology: Model formulation and proof-of-concept simulations, Water Resour. Res., 52, 6062–6077, https://doi.org/10.1002/2015WR018427, 2016.

Parr, C., Sturm, M., and Larsen, C.: Snowdrift Landscape Patterns: An Arctic Investigation, Water Resour. Res., 56, https://doi.org/10.1029/2020WR027823, 2020.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., and Dubourg, V.: Scikit-learn: Machine learning in Python, J. Mach. Learn. Res., 12, 2825–2830, 2011.

Peel, M. C., Finlayson, B. L., and McMahon, T. A.: Updated world map of the Köppen-Geiger climate classification, Hydrol. Earth Syst. Sci. Discuss., 4, 439–473, 2007.

Pomeroy, J., Gray, D., Brown, T., Hedstrom, N., Quinton, W., Granger, R., and Carey, S.: The cold regions hydrological model: a platform for basing process representation and model structure on physical evidence, Hydrol. Process., 21, 2650–2667, 2007.

Pulliainen, J., Luojus, K., Derksen, C., Mudryk, L., Lemmetyinen, J., Salminen, M., Ikonen, J., Takala, M., Cohen, J., Smolander, T., and Norberg, J.: Patterns and trends of Northern Hemisphere snow mass from 1980 to 2018, Nature, 581, 294–298, https://doi.org/10.1038/s41586-020-2258-0, 2020.

Rees, A., English, M., Derksen, C., Toose, P., and Silis, A.: Observations of late winter Canadian tundra snow cover properties, Hydrol. Process., 28, 3962–3977, https://doi.org/10.1002/hyp.9931, 2014.

Revuelto, J., López-Moreno, J. I., Azorin-Molina, C., and Vicente-Serrano, S. M.: Topographic control of snowpack distribution in a small catchment in the central Spanish Pyrenees: intra-and inter-annual persistence, The Cryosphere, 8, 1989–2006, 2014.

Revuelto, J., Billecocq, P., Tuzet, F., Cluzet, B., Lamare, M., Larue, F., and Dumont, M.: Random forests as a tool to understand the snow depth distribution and its evolution in mountain areas, Hydrol. Process., 2020.

Revuelto, J., López-Moreno, J. I., and Alonso-González, E.: Light and shadow in mapping alpine snowpack with unmanned aerial vehicles in the absence of ground control points, Water Resour. Res., 57, e2020WR028980, 2021.

Riseth, J. Å., Tømmervik, H., Helander-Renvall, E., Labba, N., Johansson, C., Malnes, E., Bjerke, J. W., Jonsson, C., Pohjola, V., Sarri, L.-E., Schanche, A., and Callaghan, T. V.: Sámi traditional ecological knowledge as a guide to science: snow, ice and reindeer pasture facing climate change, Polar Rec., 47, 202–217, https://doi.org/10.1017/S0032247410000434, 2011.

Rogers, M. C., Sullivan, P. F., and Welker, J. M.: Evidence of nonlinearity in the response of net ecosystem CO2 exchange to increasing levels of winter snow depth in the High Arctic of Northwest Greenland, Arct. Antarct. Alp. Res., 43, 95–106, 2011.

Salmon, V. G., Soucy, P., Mauritz, M., Celis, G., Natali, S. M., Mack, M. C., and Schuur, E. A. G.: Nitrogen availability increases in a tundra ecosystem during five years of experimental permafrost thaw, Glob. Change Biol., 22, 1927–1941, https://doi.org/10.1111/gcb.13204, 2016.

Schaefer, J. A. and Messier, F.: Scale-dependent correlations of Arctic vegetation and snow cover, Arct. Alp. Res., 27, 38–43, 1995.

Scott, P. A. and Rouse, W. R.: Impacts of increased winter snow cover on upland tundra vegetation: a case example, Clim. Res., 5, 25–30, 1995.

Servén, D., Brummitt, C., and Abedi, H.: Generalized Additive Models in Python., 2018.

Shook, K. R.: Simulation of the ablation of prairie snowcovers., 1997.

Stuefer, S., Kane, D. L., and Liston, G. E.: In situ snow water equivalent observations in the US Arctic, Hydrol. Res., 44, 21–34, https://doi.org/10.2156/nh.2012.177, 2013.

Sturm, M. and Holmgren, J.: Effects of microtopography on texture, temperature and heat flow in Arctic and sub-Arctic snow, Ann. Glaciol., 19, 63–68, https://doi.org/10.3189/1994AoG19-1-63-68, 1994.

Sturm, M. and Holmgren, J.: An automatic snow depth probe for field validation campaigns, Water Resour. Res., 54, 9695–9701, 2018.

Sturm, M. and Stuefer, S.: Wind-blown flux rates derived from drifts at arctic snow fences, J. Glaciol., 59, 21–34, https://doi.org/10.3189/2013JoG12J110, 2013.

Sturm, M. and Wagner, A. M.: Using repeated patterns in snow distribution modeling: An Arctic example, Water Resour. Res., 46, W12549, https://doi.org/10.1029/2010WR009434, 2010.

Sturm, M., Racine, C., and Tape, K.: Climate change: increasing shrub abundance in the Arctic, Nature, 411, 546–547, 2001a.

Sturm, M., McFadden, J. P., Liston, G. E., Chapin III, F. S., Racine, C. H., and Holmgren, J.: Snow-shrub interactions in arctic tundra: a hypothesis with climatic implications, J. Clim., 14, 336–344, 2001b.

Sturm, M., Douglas, T., Racine, C., and Liston, G. E.: Changing snow and shrub conditions affect albedo with global implications, J. Geophys. Res. Biogeosciences, 110, 2005.

Sturm, M., Taras, B., Liston, G. E., Derksen, C., Jonas, T., and Lea, J.: Estimating snow water equivalent using snow depth data and climate classes, J. Hydrometeorol., 11, 1380–1394, https://doi.org/10.1175/2010JHM1202.1, 2010.

Tarboton, D. G., Blöschl, G., Cooley, K., Kirnbauer, R., and Luce, C.: Spatial snow cover processes at Kühtai and Reynolds creak, in: Spatial patterns in catchment hydrology: observations and modelling, edited by: Grayson, R. and Blöschl, G., Cambridge University Press, Cambridge, 158–186, 2000.

Trujillo, E., Ramírez, J. A., and Elder, K. J.: Topographic, meteorologic, and canopy controls on the scaling characteristics of the spatial distribution of snow depth fields, Water Resour. Res., 43, https://doi.org/10.1029/2006WR005317, 2007.

Uhlemann, S., Dafflon, B., Peterson, J., Ulrich, C., Shirley, I., Michail, S., and Hubbard, S.: Geophysical Monitoring Shows that Spatial Heterogeneity in Thermohydrological Dynamics Reshapes a Transitional Permafrost System, Geophys. Res. Lett., 48, e2020GL091149, 2021.

U.S. Fish and Wildlife Service, R. 7 on behalf on the W. A. L. C. C.: Develop an existing vegetation layer for the western Alaska LCC region, U.S. Fish and Wildlife Service, Region 7 on behalf on the Western Alaska Landscape Conservation Cooperative, Anchorage, AK, 2015.

Wainwright, H. M., Liljedahl, A. K., BDafflon, B., Ulrich, C., Peterson, J. E., Gusmeroli, A., and Hubbard, S. S.: Mapping snow depth within a tundra ecosystem using multiscale observations and Bayesian methods, The Cryosphere, 11, 857–875, https://doi.org/10.5194/tc-11-857-2017, 2017.

Weiss, A.: Topographic position and landforms analysis, Poster presentation, ESRI user conference, San Diego, CA, 2001.

Westergaard-Nielsen, A., Lund, M., Pedersen, S. H., Schmidt, N. M., Klosterman, S., Abermann, J., and Hansen, B. U.: Transitions in high-Arctic vegetation growth patterns and ecosystem productivity tracked with automated cameras from 2000 to 2013, Ambio, 46, 39–52, https://doi.org/10.1007/s13280-016-0864-8, 2017.

Wilson, C., Bolton, R., Busey, R., Lathrop, E., and Dann, J.: End-of-Winter Snow Depth, Temperature, Density and SWE Measurements at Kougarok Road Site, Seward Peninsula, Alaska, 2018, https://doi.org/10.5440/1593874, 2020a.

Wilson, C., Bolton, R., Busey, R., Lathrop, E., Dann, J., Charsley-Groffman, L., and Benentt, Katrina E.: End-of-Winter Snow Depth, Temperature, Density and SWE Measurements at Teller Road Site, Seward Peninsula, Alaska, 2016-2018, https://doi.org/10.5440/1592103, 2020b.

Winstral, A. and Marks, D.: Long-term snow distribution observations in a mountain catchment: Assessing variability, time stability, and the representativeness of an index site, Water Resour. Res., 50, 293–305, 2014.

Winstral, A., Elder, K., and Davis, R. E.: Spatial Snow Modeling of Wind-Redistributed Snow Using Terrain-Based Parameters, J. Hydrometeorol., 3, 524–538, https://doi.org/10.1175/1525-7541(2002)003<0524:SSMOWR>2.0.CO;2, 2002.

Woo, M. and Young, K. L.: Modeling arctic snow distribution and melt at the 1 km grid scale, Nord. Hydrol., 35, 295–307, 2004.

Young, K. L., Brown, L., and Labine, C.: Snow cover variability at Polar Bear Pass, Nunavut, Arct. Sci., 4, 669–690, https://doi.org/10.1139/as-2017-0016, 2018.

Zhu, X., Lee, S.-Y., Wen, X., Wei, Z., Ji, Z., Zheng, Z., and Dong, W.: Historical evolution and future trend of Northern Hemisphere snow cover in CMIP5 and CMIP6 models, Environ. Res. Lett., 16, 065013, https://doi.org/10.1088/1748-9326/ac0662, 2021.

Zimmerman, D., Pavlik, C., Ruggles, A. and Armstrong, M.P., 1999. An experimental comparison of ordinary and universal kriging and inverse distance weighting. Mathematical Geology, 31(4), pp.375-390.

Zona, D., Gioli, B., Commane, R., Lindaas, J., Wofsy, S. C., Miller, C. E., Dinardo, S. J., Dengel, S., Sweeney, C., Karion, A., Chang, R. Y.-W., Henderson, J. M., Murphy, P. C., P., G. J., Moreaux, V., Liljedahl, A., Watts, J. D., Kimball, J. S., Lipson, D. A., and Oechel, W. C.: Cold season emissions dominate the Arctic tundra methane budget, PNAS, 113, 40–45, https://doi.org/10.1073/pnas.1516017113, 2016.

Figures

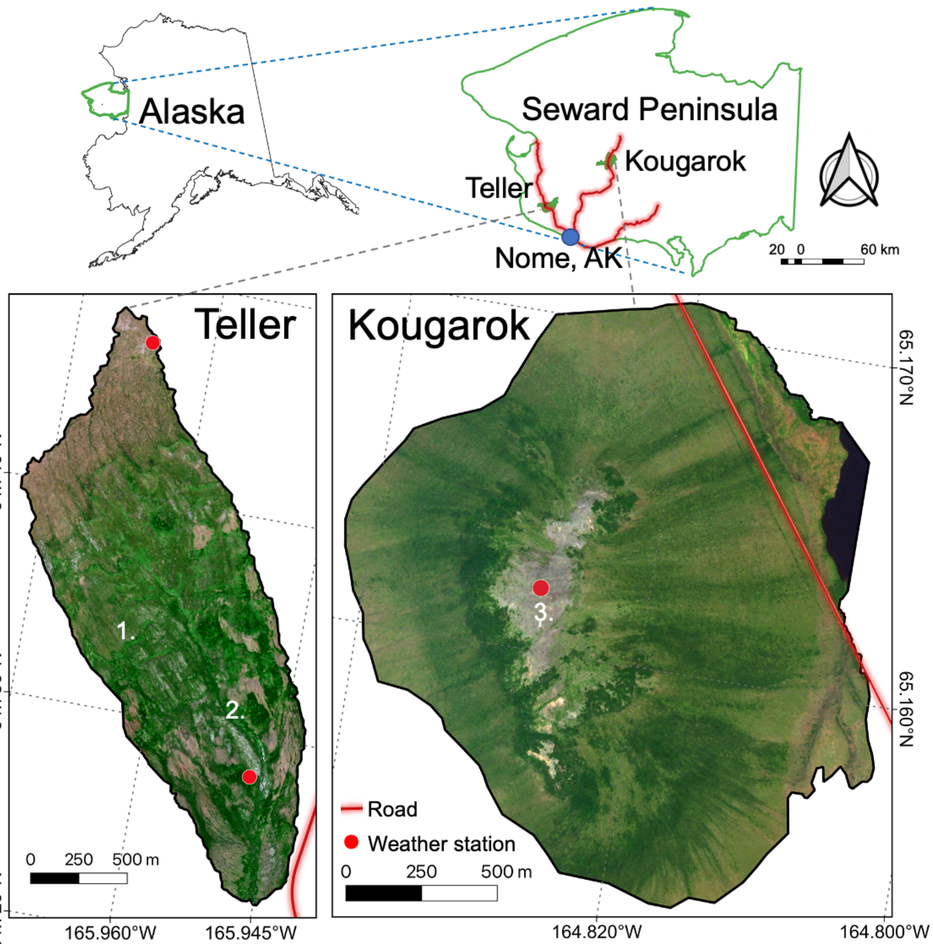


Figure 1. Location and WorldView-2 RGB imagery of the Teller watershed (lower left) and Kougarok Hillslope study site (lower right). Teller has two weather stations (red dots), located near the top and bottom of the watershed, and the Kougarok Hillslope has one weather station located near the top of the rocky dome. The sites are located on the Seward Peninsula of Alaska, with the HUC12 basins for the Teller watershed and the Kougarok Hillslope shown in green and the Seward Peninsula road system in red (upper right). RGB composite from the 8-band WorldView-2 images obtained on July 27, 2011 (Teller) and July 14, 2017 (Kougarok) at 1.5m resolution downloaded from the DigitalGlobe website (https://www.digitalglobe.com/) Imagery © 2011/2017 MAXAR. Example of features discussed in the text are denoted on the map as 1. terraces and risers, 2. the Teller watershed stream bed, and 3. the Kougarok Hillslope dome.

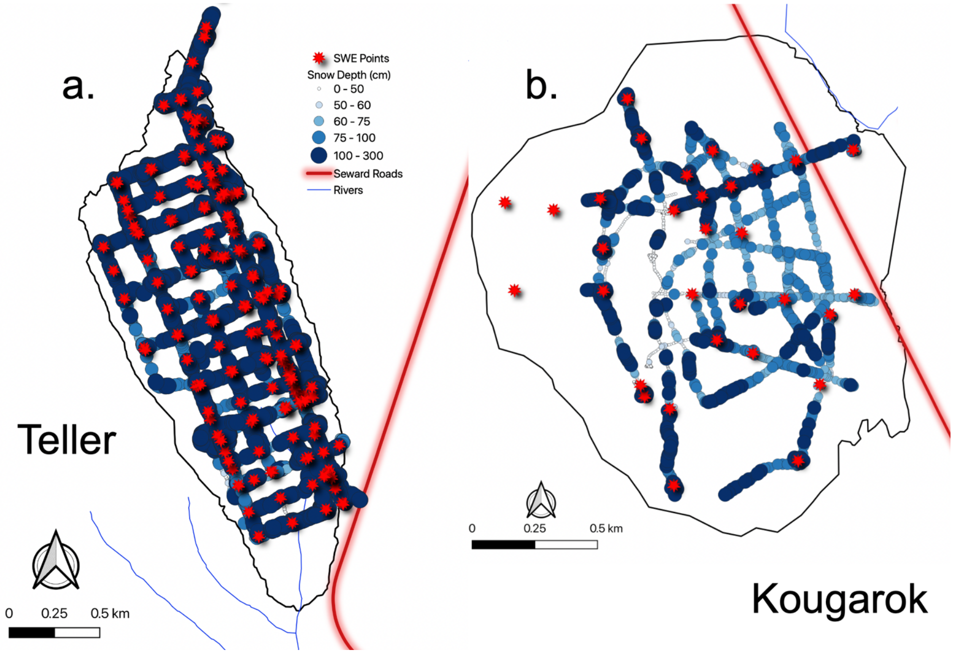
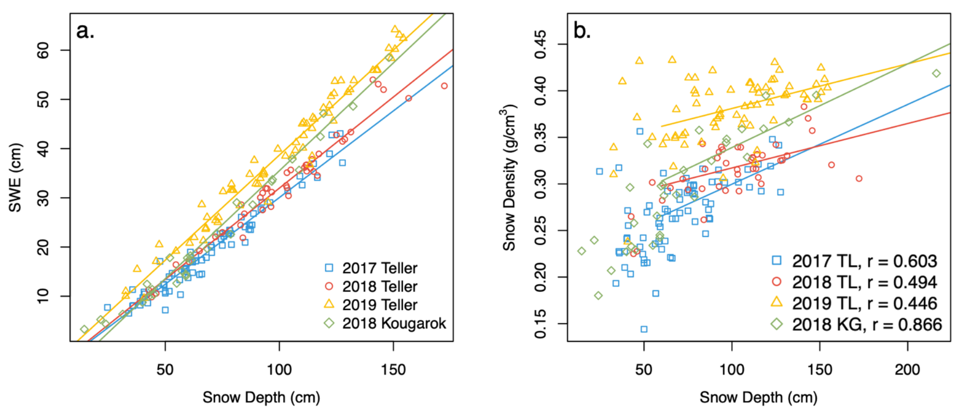


Figure 2. Measured snow depth (cm) and snow density (SWE Points) at a.) the Teller watershed (2017-2019) and b.) the Kougarok Hillslope (2018).

985



990 **Figure 3. Scatter plots of (a.) snow depth vs SWE and (b.) snow depth vs snow density. The linear regressions in panel b. are for snow depth > 60 cm.**

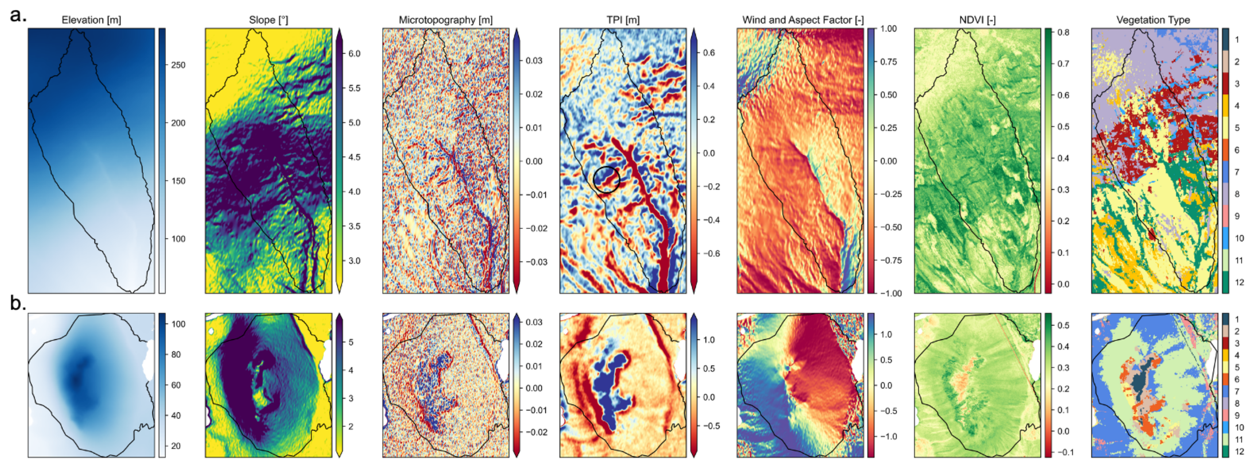
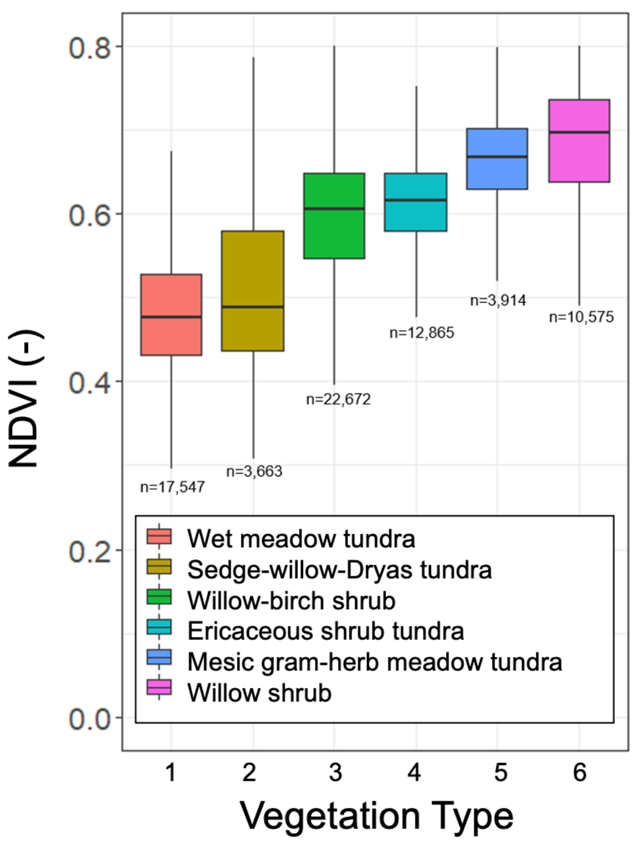


Figure 4. Input features included in the snow models (elevation, slope, microtopography, TPI, wind and aspect factor, NDVI and vegetation type) for (a.) Teller and (b.) Kougarok. The white areas shown on the Kougarok maps are lakes. An open circle on the TPI figure denotes one of the terrace and riser. The vegetation types are (1) Dryas-lichen dwarf shrub tundra, (2) Birch-Ericaceous-lichen shrub tundra, (3) Ericaceous dwarf shrub tundra, (4) Sedge-willow-Dryas tundra, (5) Willow-birch shrub, (6) Alder-willow shrub, (7) Tussock-lichen tundra, (8) Wet meadow tundra, (9) Wet sedge bog-meadow, (10) Mesic graminoid-herb meadow tundra, (11) Mixed shrub-sedge tussock tundra, and (12) Willow shrub.



995 **Figure 5. Teller watershed NDVI by major vegetation type. Mesic gram-herb is short for mesic graminoid-herbaceous tundra. Table A1 details each vegetation type.**

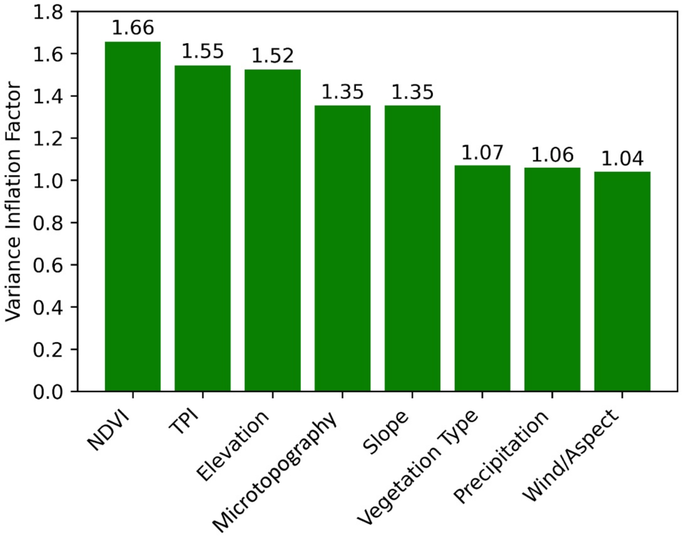


Figure 6. Variance inflation factors for the model inputs.

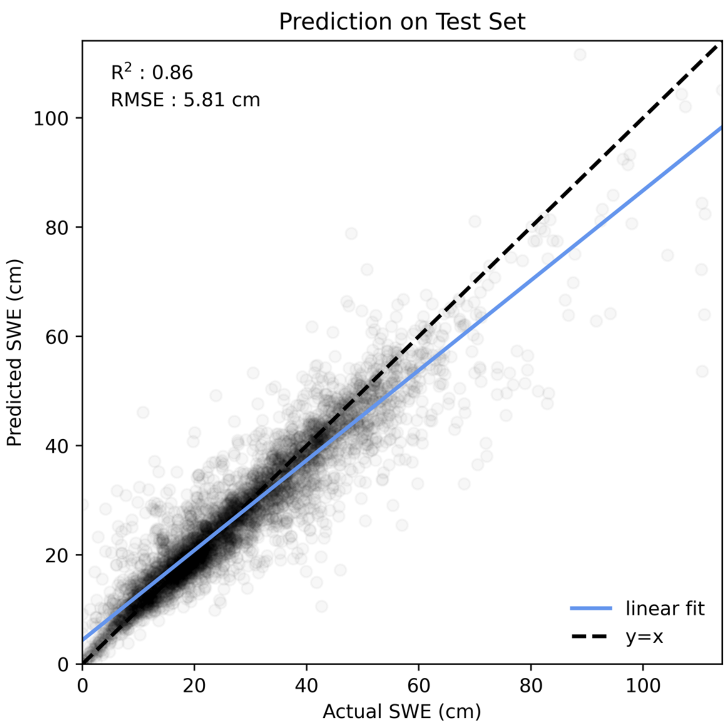


Figure 7. Random forest results of predicted SWE vs actual SWE for the test set when using all year and site datasets, with a test set of 20%. R^2 equal to 0.86 and the RMSE is equal to 5.81 cm. The solid blue line is a linear fit to the scatter points, and the dashed black line is a y=x line.

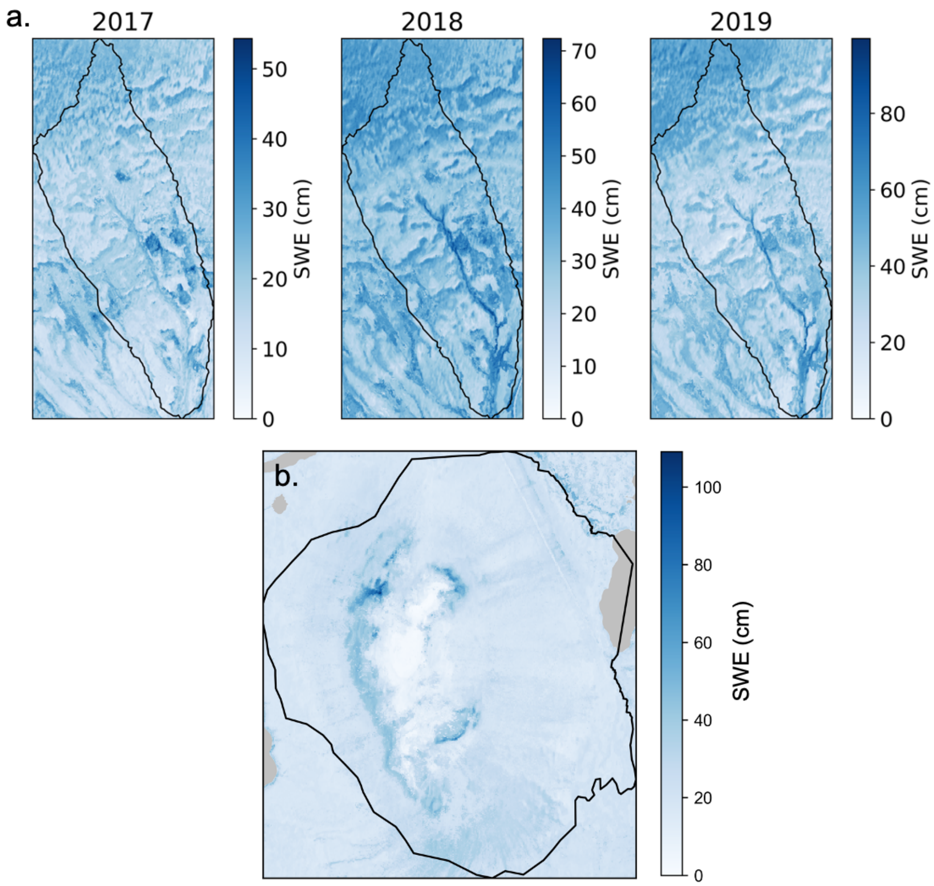


Figure 8. Spatially predicted SWE for the final model for (a.) Teller 2017, 2018, and 2019 and (b.) Kougarok 2018. The gray areas on the Kougarok map are small lakes. Note that the scales change for each year and study location across the panels.

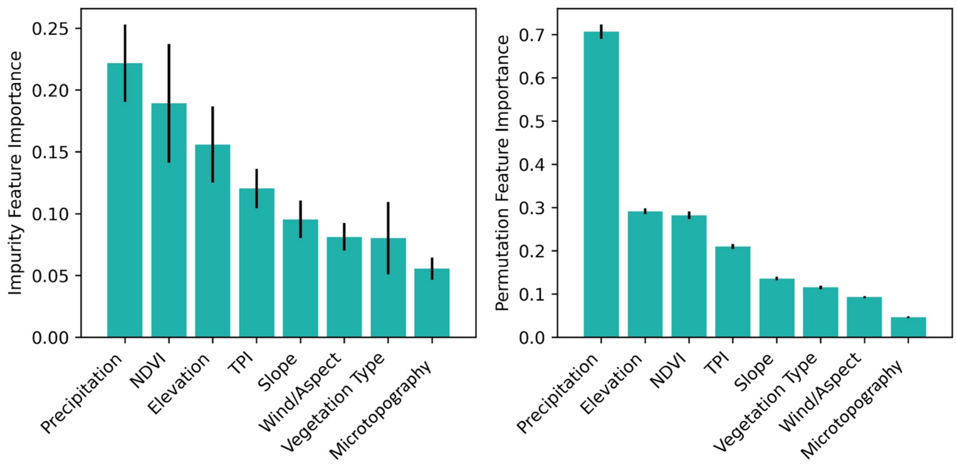


Figure 9. Random forest feature importance results (left, impurity and right, permutation) for the final model, with ± 1 standard deviation shown by the black error bars.

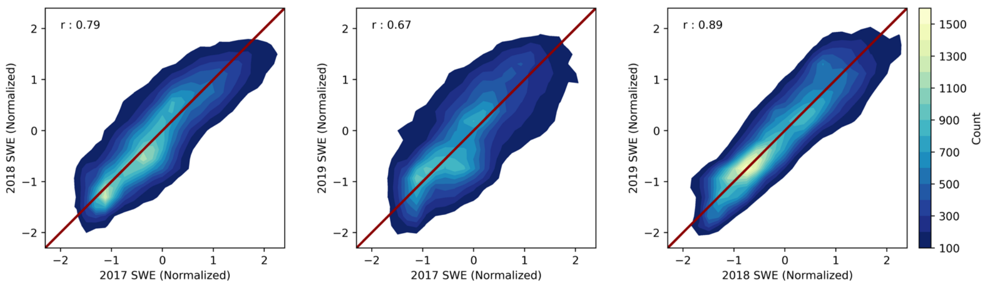


Figure 10. Heatmaps of predicted normalized SWE for Teller between 2017, 2018 and 2019 with Pearson correlation coefficients (r) showing significant correlations of SWE among years. Random forest was trained with data from all years and sites. The color scale represents the density of data points, with dark blue representing areas with the least points and light-yellow representing areas with the most points. Areas with fewer than 100 points are not plotted. The solid red line is a line of best fit using singular value decomposition.

Tables

Table 1. Topography and vegetation data sources. Data sources for the input features are listed on the left, and the description of the features are listed on the right.

Data Sources	Features	Descriptions
DEM	**Elevation**	Elevation (meters)
	Slope	Slope angle (degrees)
	Microtopography	Difference between elevation of single 5 m cell and the average elevation of cells in the surrounding box of 15 m width (m)
	Topographic Position Index (TPI)	Difference between elevation of single 5 m cell and the average elevation of cells in the surrounding box of 155 m width (m)
DEM and wind data	**Wind/Aspect**	Calculated using wind direction and aspect formula (unitless)
WorldView-2 imagery	**NDVI**	Normalized difference vegetation index (unitless)
Konduri and Kumar, 2021	**Vegetation Type**	Vegetation types: (1) Dryas-lichen dwarf shrub tundra, (2) Birch-Ericaceous-lichen shrub tundra, (3) Ericaceous dwarf shrub tundra, (4) Sedge-willow-Dryas tundra, (5) Willow-birch shrub, (6) Alder-willow shrub, (7) Tussock-lichen tundra, (8) Wet meadow tundra, (9) Wet sedge bog-meadow, (10) Mesic graminoid-herb meadow tundra, (11) Mixed shrub-sedge tussock tundra, and (12) Willow shrub

Table 2. Teller and Kougarok snow depth and density observations and SWE observations and estimates. The coefficient of variation (CV) of the observed variables are in brackets following the average value.

Year	Study Site	Observations	SWE	Density	Snow Depth	SWE
		Snow depth/SWE (SWE groupings)	Observed, cm (CV)	Observed, g/cm^3 (CV)	Observed, cm (CV)	Estimated, cm
2017	Teller	8469 / 234 (77)	21.03 (0.43)	0.27 (0.11)	73.77 (0.42)	19.37
2018	Teller	5076 / 150 (49)	32.58 (0.31)	0.32 (0.07)	108.95 (0.31)	34.03
2018	Kougarok	4655 / 96 (31)	21.15 (0.86)	0.29 (0.16)	75.27 (0.57)	23.15
2019	Teller	5376 / 199 (69)	37.94 (0.36)	0.38 (0.07)	106.38 (0.38)	39.75
Total		23481 / 653 (203)				

Formatted Table
Deleted: 19.33
Deleted: 6
Deleted: 5
Deleted: 2
Deleted: 57
Deleted: 7
Deleted: 1
Deleted: 10
Deleted: 6
Deleted: 4
Deleted: 20
Deleted: 88
Deleted: 30
Deleted: 21
Deleted: 3
Deleted: 8
Deleted: 21
Deleted: 6
Deleted: 9
Deleted: 74

Table 3. Comparison of model performance for linear regression, GAM, and random forest models that are trained on individual years and individual sites.

Study Site	Year	Model	Train R^2	Test R^2	Train RMSE	Test RMSE
Teller	**2017**	Linear Regression	0.34	0.29	7.17	7.66
		GAM	0.49	0.45	6.27	6.56
		Random Forest	0.97	0.78	1.47	4.25
	2018	Linear Regression	0.35	0.30	8.85	9.20
		GAM	0.53	0.48	7.56	7.92
		Random Forest	0.97	0.77	1.94	5.28
	2019	Linear Regression	0.30	0.36	13.58	12.69
		GAM	0.51	0.52	11.40	10.91
		Random Forest	0.96	0.72	3.15	8.31
Kougarok	**2018**	Linear Regression	0.45	0.44	12.02	13.60
		GAM	0.70	0.72	8.96	9.68
		Random Forest	0.98	0.92	2.23	5.16

Table 4. Random forest results for the training and testing data used to estimate SWE.

Study Sites	Year	Train R^2	Test R^2	Train RMSE (cm)	Test RMSE (cm)
Teller & Kougarok	All	0.98	0.86	2.18	5.81
Teller	All	0.98	0.86	2.17	5.78
	2017	0.97	0.78	1.47	4.25
	2018	0.97	0.77	1.94	5.28
	2019	0.96	0.72	3.15	8.31
Kougarok	2018	0.98	0.92	2.23	5.20

Appendix

Wind Equations

The two wind equations in the current study were derived similarly to the wind equation found in Dvornikov et al., 2015. The purpose of the wind equations is to establish an index using positive and negative values based on the topography-wind relationship. First, we used meteorological station data to calculate an average prevailing wind direction for the winter months. Second, we used the corresponding wind equation to assign positive values to the leeward side of topographical features (where wind loading or drifting of snow is likely) and negative values to the windward side of topographical features (where wind scour of snow is likely).

To derive our wind equations, we divided the wind rose into eight cardinal directions by +/- 22.5 degree increments, N, NE, E, SE, S, SW, W, NW. The eight possible wind equations that correspond to the eight cardinal directions are listed below, along with the range of values associated with each equation. We determined that the coefficients from Dvornikov et al., 2015, simply added a scaling factor to the wind equation values, and were not necessary for our results.

We chose two wind equations (NE and E) for our study area based on the prevailing winter wind direction for each year (Figure A2).

N: $Wf(x) = -\cos(x)$; range = -1 to 1
NE: $Wf(x) = -\cos(x) - \sin(x)$; range = -1.414 to 1.414
E: $Wf(x) = -\sin(x)$; range = -1 to 1
SE: $Wf(x) = \cos(x) - \sin(x)$; range = -1.414 to 1.414
S: $Wf(x) = \cos(x)$; range = -1 to 1
SW: $Wf(x) = \cos(x) + \sin(x)$; range = -1.414 to 1.414
W: $Wf(x) = \sin(x)$; range = -1 to 1
NW: $Wf(x) = -\cos(x) + \sin(x)$; range = -1.414 to 1.414

Figures

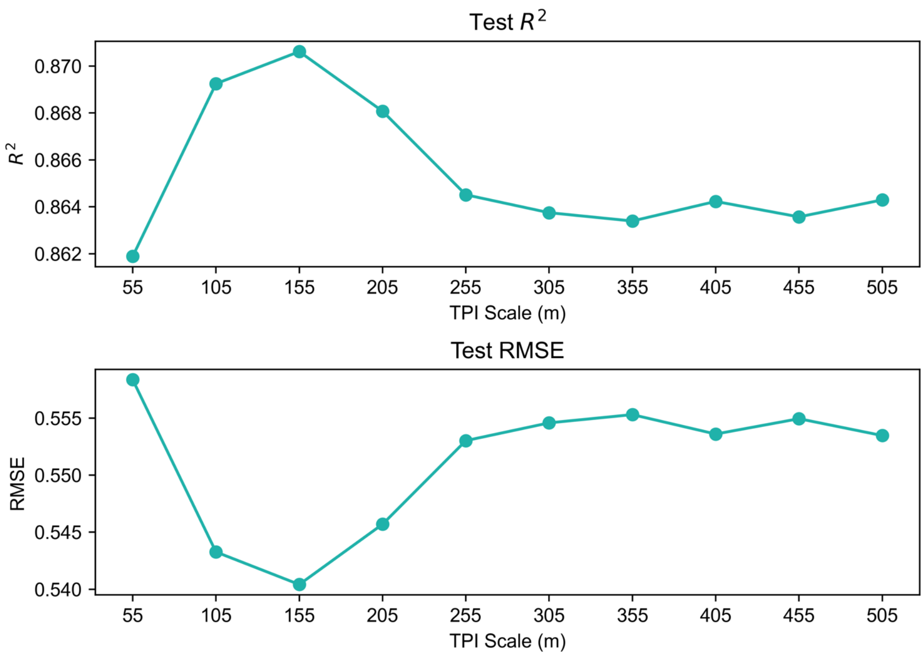


Figure A1. Random forest model performance for model runs using a range of TPI scales. The model performance is optimized when the TPI scale is 155 m.

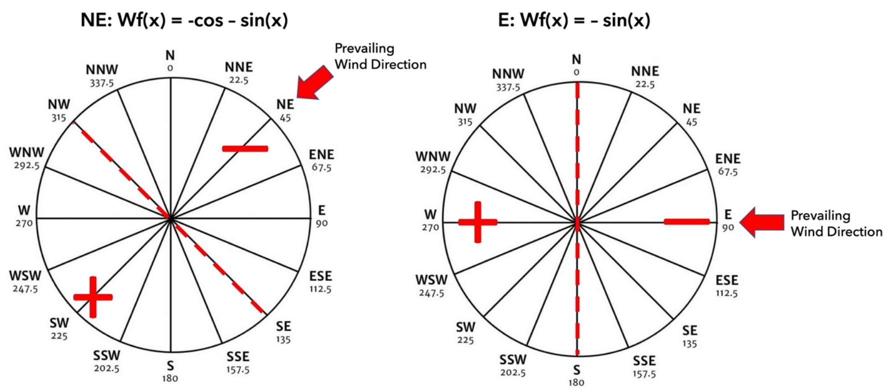


Figure A2. The prevailing wind directions (NE, E) calculated from meteorological station data and the corresponding wind equations applied in this study.

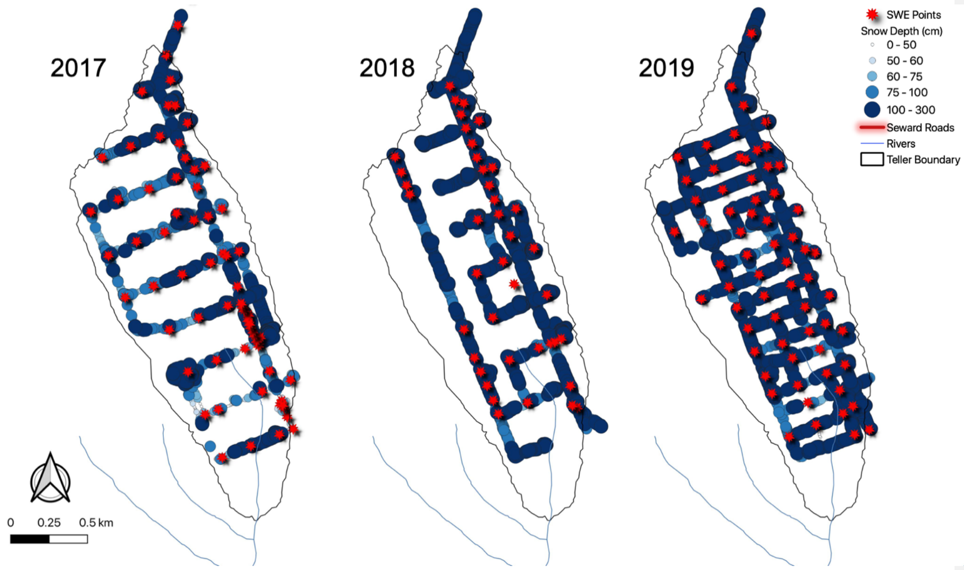


Figure A3. Teller snow depth and SWE measurements for each year of the survey.

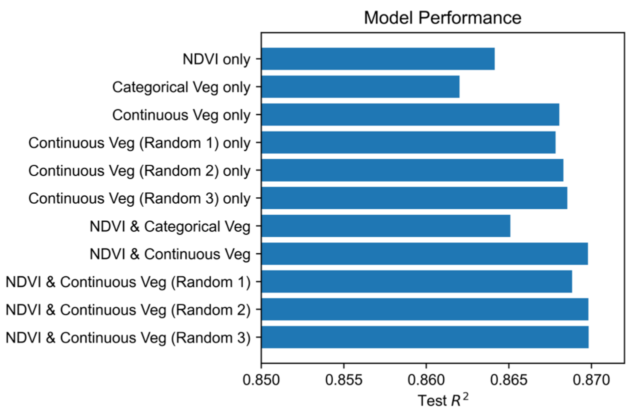


Figure A4. Random forest model performance for model runs using various combinations of vegetation features. The categorical vegetation uses the vegetation types shown in Figure 4. The continuous vegetation is a ranking ordered by which vegetation type has higher SWE. The continuous vegetation ranking is also randomized three different ways to determine if the ranking order is important.

Deleted: 3

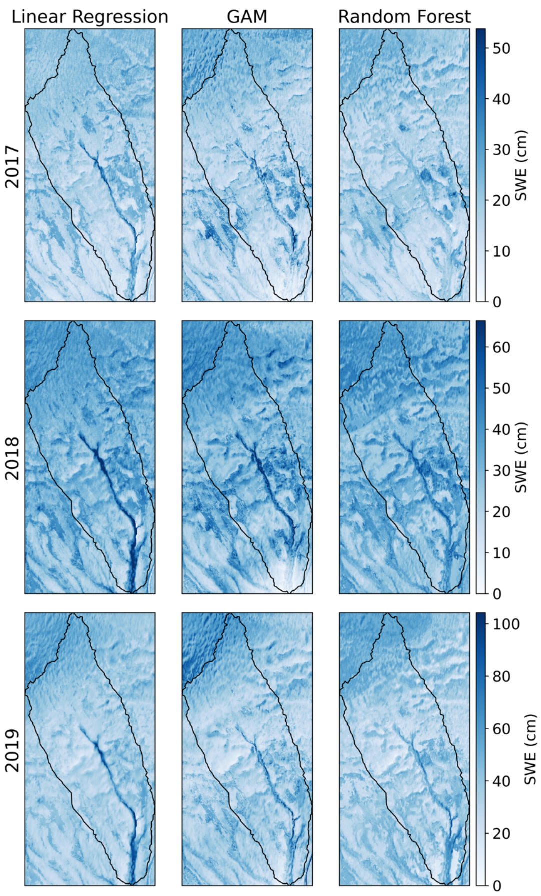


Figure A5. All three statistical models trained on individual years at Teller. The scale of SWE changes from year to year depending on the annual snowfall.

165

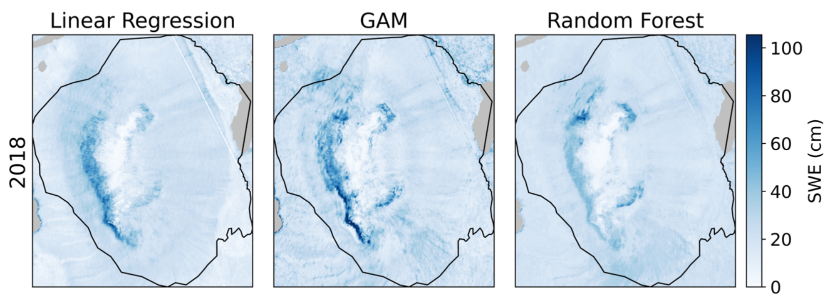
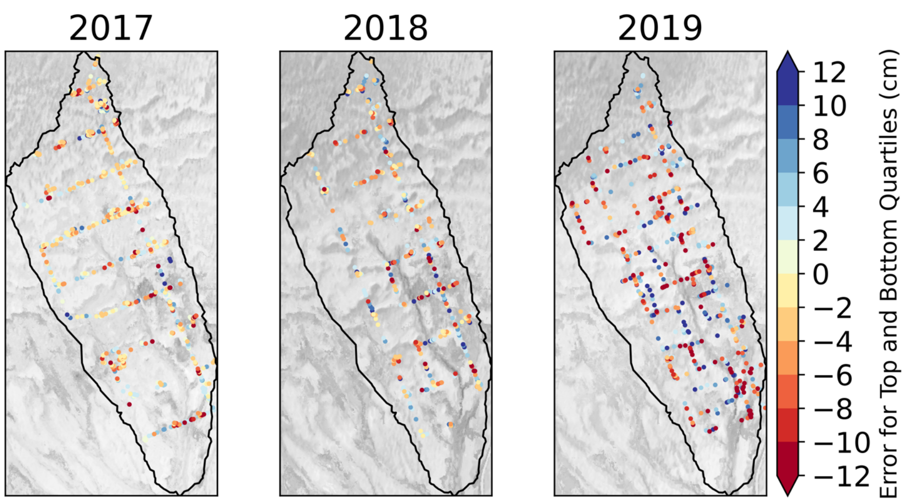


Figure A6. All three models trained on single year (2018) at Kougarok.



175 Figure A7. Spatial error (predicted minus observed) for the upper-most and lower-most quartiles of error in the random forest SWE prediction for Teller.

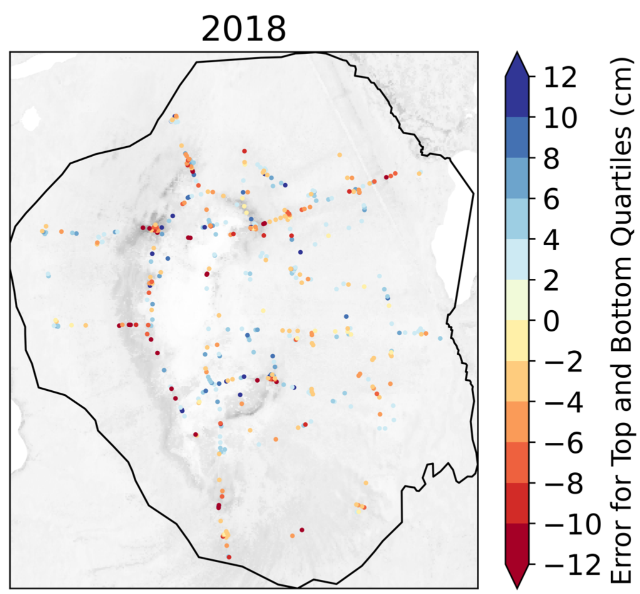


Figure A8. Spatial error (predicted minus observed) for the upper-most and lower-most quartiles of error in the random forest SWE prediction for the final model at Kougarok. The error is not distributed systematically through space.

Tables

Table A1. Vegetation classes ranked by observed SWE from lowest SWE (1) to highest SWE (12) at Teller. All binned values are significantly different from each other.

Vegetation Type	Rank	Binned Value (Teller)
Dryas-lichen dwarf shrub tundra	1	2
Birch-Ericaceous-lichen shrub tundra	2	NA
Ericaceous dwarf shrub tundra	3	4
Sedge-willow-Dryas tundra	4	2
Willow-birch shrub	5	3
Alder-willow shrub	6	NA
Tussock-lichen tundra	7	NA
Wet meadow tundra	8	1
Wet sedge bog-meadow	9	1
Mesic graminoid-herb meadow tundra	10	5
Mixed shrub-sedge tussock tundra	11	NA
Willow shrub	12	6

NA in binned values indicate vegetation types comprised <1% of total area at Teller

Table A2. Average winter (October-March) wind speed and prevailing direction.

		Wind Speed	**Wind Direction (Prevailing)**	**Wind Direction (Prevailing) > 5 m/s**
		(m/s)	(°)	(°)
Nome (measured at 5 m height)	2016-2017	10.14	60.40 (ENE)	65.16 (ENE)
	2017-2018	12.42	66.11 (ENE)	67.71 (ENE)
	2018-2019	12.06	76.57 (ENE)	78.27 (ENE)
Teller (station at top of watershed)	2016-2017	4.99	83.63 (E)	98.17 (E)
	2017-2018	5.47	94.66 (E)	115.38 (ESE)
	2018-2019	4.86	76.10 (E)	92.07 (E)
Kougarok	2018-2019	6.74	60.23 (ENE)	45.00 (NE)

Table A3. Hyperparameters for Random Forest Model

	Final Model	Teller	Kougarok
Tree density	840	1200	1165
Max depth	90	100	67
Max features	4	4	3
Min samples split	2	2	2
Min samples leaf	1	1	1