

LETTER • **OPEN ACCESS**

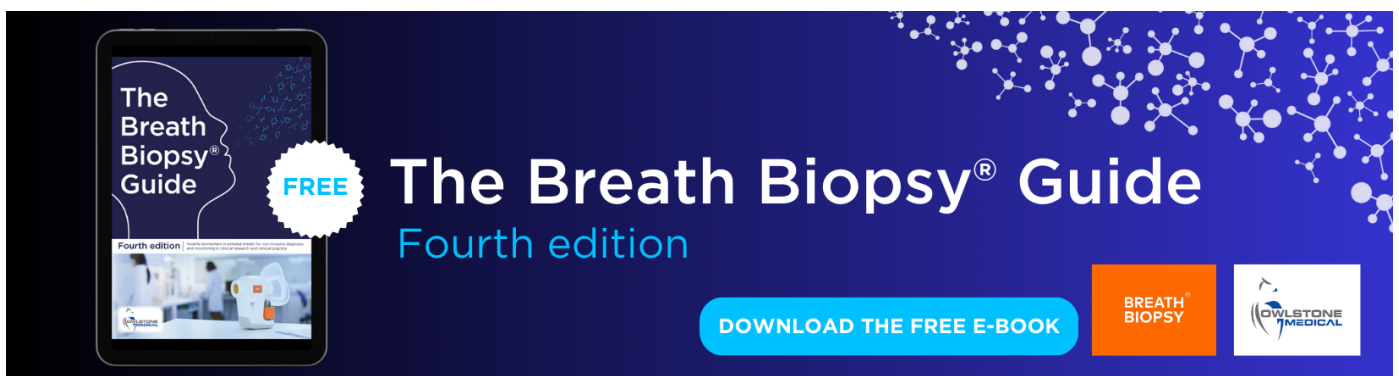
Maps of active layer thickness in northern Alaska by upscaling P-band polarimetric synthetic aperture radar retrievals

To cite this article: Jane Whitcomb *et al* 2024 *Environ. Res. Lett.* **19** 014046

View the [article online](#) for updates and enhancements.

You may also like

- [How does soil water content influence permafrost evolution on the Qinghai-Tibet Plateau under climate warming?](#)
Fang Ji, Linfeng Fan, Xingxing Kuang *et al.*
- [Inference of the impact of wildfire on permafrost and active layer thickness in a discontinuous permafrost region using the remotely sensed active layer thickness \(ReSALT\) algorithm](#)
Roger J Michaelides, Kevin Schaefer, Howard A Zebker *et al.*
- [Changes in the 1963–2013 shallow ground thermal regime in Russian permafrost regions](#)
Dmitry A Streletskiy, Artem B Sherstiukov, Oliver W Frauenfeld *et al.*



The Breath Biopsy® Guide
Fourth edition

FREE

DOWNLOAD THE FREE E-BOOK

BREATH BIOPSY

OWLSTONE MEDICAL

ENVIRONMENTAL RESEARCH
LETTERS

LETTER

Maps of active layer thickness in northern Alaska by upscaling
P-band polarimetric synthetic aperture radar retrievals

OPEN ACCESS

RECEIVED
28 June 2023REVISED
6 November 2023ACCEPTED FOR PUBLICATION
5 December 2023PUBLISHED
15 December 2023

Original Content from
this work may be used
under the terms of the
[Creative Commons
Attribution 4.0 licence](#).

Any further distribution
of this work must
maintain attribution to
the author(s) and the title
of the work, journal
citation and DOI.

Jane Whitcomb^{1,*}, Richard Chen², Daniel Clewley³, John S Kimball⁴, Neal J Pastick⁵, Yonghong Yi⁶
and Mahta Moghaddam¹¹ Department of Electrical Engineering, University of Southern California, Los Angeles, CA, United States of America² Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, United States of America³ Plymouth Marine Laboratory, Centre for Geospatial Applications, Plymouth, United Kingdom⁴ Numerical Terradynamic Simulation Group, University of Montana, Missoula, MT, United States of America⁵ U.S. Geological Survey, Earth Resources Observation and Science Center, Sioux Falls, SD, United States of America⁶ Joint Institute for Regional Earth System Science and Engineering, University of California, Los Angeles, CA, United States of America

* Author to whom any correspondence should be addressed.

E-mail: jbwhitco@usc.edu**Keywords:** ABoVE, AirMOSS, ALT, SAR, upscaling, random forests, machine learningSupplementary material for this article is available [online](#)**Abstract**

Extensive, detailed information on the spatial distribution of active layer thickness (ALT) in northern Alaska and how it evolves over time could greatly aid efforts to assess the effects of climate change on the region and also help to quantify greenhouse gas emissions generated due to permafrost thaw. For this reason, we have been developing high-resolution maps of ALT throughout northern Alaska. The maps are produced by upscaling from high-resolution swaths of estimated ALT retrieved from airborne P-band synthetic aperture radar (SAR) images collected for three different years. The upscaling was accomplished by using hundreds of thousands of randomly selected samples from the SAR-derived swaths of ALT to train a machine learning regression algorithm supported by numerous spatial data layers. In order to validate the maps, thousands of randomly selected samples of SAR-derived ALT were excluded from the training in order to serve as validation pixels; error performance calculations relative to these samples yielded root-mean-square errors (RMSEs) of 7.5–9.1 cm, with bias errors of magnitude under 0.1 cm. The maps were also compared to ALT measurements collected at a number of *in situ* test sites; error performance relative to the site measurements yielded RMSEs of approximately 11–12 cm and bias of 2.7–6.5 cm. These data are being used to investigate regional patterns and underlying physical controls affecting permafrost degradation in the tundra biome.

1. Introduction

Active layer thickness (ALT) is defined as the annual maximum thaw depth observed at the end of the summer [1]. Detailed knowledge of the spatial and temporal distribution of ALT throughout northern Alaska is useful to assess the impacts of climate change in the region, including any increases in greenhouse gas emissions that might occur as a consequence of permafrost degradation [2–5]. Many studies have been conducted in order to collect detailed data on ALT, sometimes over periods of many years [6–8]. The vastness and inaccessibility of northern Alaska, however, have so far made it impossible to map ALT

and its temporal dynamics over extensive regions by conventional means.

Current knowledge of the state of ALT in northern Alaska has relied primarily on extrapolation from direct *in situ* ALT measurements made at many scattered test sites. Although increasing numbers of these sites are being established, they are necessarily located in areas that frequently experience severe weather and their development and maintenance are highly labor-intensive. The difficulty of assembling and maintaining *in situ* measurement sites in the face of harsh and sometimes dangerous conditions of climate and topography has limited the total number of *in situ* sites and also resulted in existing sites tending to

cluster along more easily accessible arteries such as the Dalton Highway. Overall, the arrangement of *in situ* measurement sites throughout northern Alaska is thus unavoidably sparse, irregular, and potentially unrepresentative.

The largest network of *in situ* ALT measurement sites is the Circumpolar Active Layer Monitoring (CALM) network [6, 7, 9–11], which maintains dozens of long-term ALT measurement sites in northern Alaska. Additional sets of test sites have been maintained by the U.S. Geological Survey [12–14]. Numerous spot measurements of ALT have also been collected using metal probing rods and ground penetrating radar [8]. Although all of these measurements are extremely valuable in assessing the condition of the permafrost at the locations where the measurements are made, attempts to extrapolate from these point samples over even moderate distances can be difficult due to the spatial heterogeneity associated with small scale variations in topography, soil properties, and vegetative land cover [15]. Studies have extrapolated ALT by using lidar and high-resolution normalized difference vegetation index (NDVI) data to sense and account for this spatial heterogeneity throughout spans of several kilometers, but not over extensive regions [16].

In one previous study [17], piecewise linear regression tree models were applied to produce a state-wide map of estimated ALT based on numerous *in situ* measurements and a large number of spatial data layers, but only as a single snapshot and subject to the uncertain spatial extensibility of *in situ* measurements.

Another approach capable of estimating ALT uses spaceborne interferometric synthetic aperture radar (InSAR) to sense surface deformations indicative of permafrost thaw [18–20]. This approach yields swaths of estimated relative surface deformation, but independent information is needed to obtain values of absolute surface deformation, and the relationship between absolute surface deformation and ALT is not unique. The ability to produce reliable estimates of ALT using this technique can be negatively affected by radar phase decorrelation and surface erosion; variations in soil properties throughout the active layer can result in biased estimates of ALT [20].

Model-based methods have been used to develop broader-based maps of ALT throughout Alaska. In one carefully validated technique, soil process models were developed with inputs from field observations and several different types of satellite-based optical sensing inputs [21]. This model-based approach was used to successfully produce low resolution (~ 1 km) estimates of ALT for the entire state of Alaska, but it is less able to capture the effects of small regional variations in ALT.

Given the limitations of existing approaches for mapping ALT over broad regions, exploring new methods for developing regional maps of ALT based

on *in situ* and remote sensing measurements would be beneficial. We therefore seek to develop a new technique for forming large-scale maps of ALT in northern Alaska capable of delineating local (30 m) scale spatial variations in ALT at regular time intervals.

In pursuit of this goal, we attempt to test the following *hypothesis*:

Swaths of estimated ALT derived from low frequency (P-band) airborne polarimetric SAR (PolSAR) can be upscaled over extensive regions of northern Alaska to produce large-scale maps of ALT of sufficient spatial resolution, temporal frequency, and accuracy to yield useful insights into the spatial and temporal dynamics of ALT throughout these regions. This is done through the use of machine learning with the support of numerous ancillary data layers known to be related to ALT. We believe this approach to be more flexible than reliance on a fixed collection of relatively small and difficult-to-service measurement sites. SAR flights can be directed to survey regions of interest more easily than constructing *in situ* measurement sites throughout those regions, and airborne SAR swaths can sense ground and subsurface conditions over far broader areas than is possible at any *in situ* measurement site.

To test this hypothesis, we extrapolate from long, narrow swaths of PolSAR-derived ALT that are collected and processed in an extensive airborne campaign called the Arctic-Boreal Vulnerability Experiment (ABOVE). ABOVE is a coordinated multi-disciplinary research effort studying the dynamics of permafrost in Alaska and western Canada. As described in [22], it is designed ‘to address the overarching science question: How vulnerable or resilient are ecosystems and society to environmental change in the Arctic and boreal region (ABR) of western North America?’

Extrapolation is achieved by drawing pools of randomly selected samples from the PolSAR-derived ALT swaths to train a regression expressing ALT in terms of a large set of spatial data layers. The trained regression is then applied throughout northern Alaska to produce high-resolution (30 m) maps of upscaled ALT in the ABOVE projection (i.e. a particular variant of Canada Albers Conical Equal Area projection) for study years 2014, 2015, and 2017. The resulting maps offer new insights into the vulnerability and resilience of Arctic and boreal ecosystems to environmental change in western North America, thereby helping to address a key science objective of the ABOVE [23].

In the remainder of this paper, section 2 describes the PolSAR imagery that provides the method’s primary data source. Section 3 presents our upscaling technique, while section 4 presents results obtained using the new method. Section 5 addresses outstanding limitations of the technique and other issues, followed by section 6 that summarizes the results of the study.

2. Primary data source

The regression used in this method is trained based on samples taken from swaths of estimated ALT that are formed by processing swaths of P-band (430 MHz, ~ 70 cm) PolSAR imagery. The source PolSAR imagery was collected during numerous flights over 12 tracks within our North Alaska study area, which consisted of the portion of Alaska north of 64° latitude. The flights were conducted as part of the Airborne Microwave Observatory of Subcanopy and Subsurface (AirMOSS) NASA Interdisciplinary Science program (2014 and 2015) and the ABoVE program Foundational Flights (2017), programs that have supported multiple aircraft deploying a variety of airborne sensors, including both P-band and L-band polarimetric SARs. The trajectories of the PolSAR tracks have been chosen to span landscape environmental gradients and a diversity of permafrost conditions and ALT levels, resulting in a data source far more broadly representative than scattered *in situ* test sites [1, 24].

P-band frequencies, in particular, are chosen for their ability to penetrate through low to moderate vegetation and topsoil to sense the depth to the permafrost. As described in [22], polarimetric SARs operating at P-band are sensitive to the geometrical and material properties of vegetation, surface properties, and sub-surface soil/permafrost profiles. The use of P-band provides enhanced sensitivity to the root zone (10–40 cm) portions of the sub-surface profile [25, 26], with sensing depths of as much as 65 cm possible under some circumstances. Although this is deep enough to reach the permafrost table in most locations in northern Alaska, it is too shallow to reach the permafrost in many areas south of the Brooks Range.

The SAR flight lines used in this study are shown in figure 1 as gray strips. Most of the flight paths are located within the continuous permafrost zone, where the aboveground vegetation, consisting mainly of dwarf shrub and tussock/sedge/moss tundra, has little effect on P-band radar backscatter [25]. The flight paths cover a diverse array of land cover, topography, soil properties, vegetation, and climate conditions (within the confines of what the region offers), as has been confirmed through histograms of data layer values within the flight paths. Each swath of PolSAR imagery collected is typically about 7–12 km wide and 100 or more km long and contains layers for all four polarizations of backscatter signal. Additionally, each flight path was flown twice in each of 3 years (2014, 2015, and 2017): once in mid-to-late August and once in early October. This was done in order to construct a two-element time-series, with the August element generally capturing conditions of maximum thaw (in some regions maximum thaw occurs in early-mid September), and the October

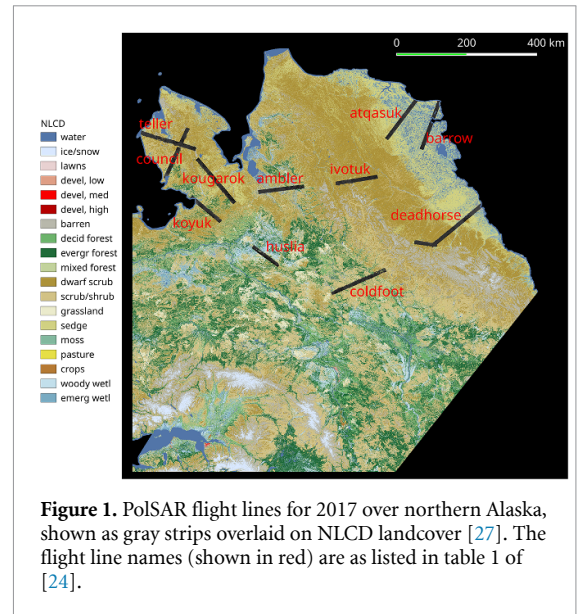


Figure 1. PolSAR flight lines for 2017 over northern Alaska, shown as gray strips overlaid on NLCD landcover [27]. The flight line names (shown in red) are as listed in table 1 of [24].

element capturing conditions of partial freezing. Additional details on the PolSAR data collection campaign are provided in the supplementary materials; the campaign is described more fully in [22].

SAR inversion processing was applied to each PolSAR swath to retrieve estimated ALT as well as the soil moisture profile and other soil properties [24]. Boxplots showing the range of PolSAR-derived ALT values for various landcover types are shown for 2014 and 2015 in figure 19 of [1]. Plots (a) and (b) in that figure show substantial variations in the range of ALT values between the different landcover types. Herbaceous wetlands and barren areas exhibit higher values of retrieved ALT (as much as 55 cm) and relatively less uncertainty, whereas sedge and scrub/shrub locations have lower values of retrieved ALT with a slightly broader range of uncertainty.

The SAR inversion algorithm is described in detail in [1], with a very brief summary provided in section 2.1 below.

2.1. SAR inversion processing for retrieving ALT and other soil properties from PolSAR

The central component of the method consists of a forward model of radar backscattering from arctic soil defined in terms of the soil's physical and electrical properties, and the transmitting radar's signal and geometric parameters. Only horizontal (HH) and vertical (VV) co-polarized radar backscatter components were modeled; cross-polarized backscatter signals were not used because the calibration accuracy of the cross-polarized radar backscatter was uncertain.

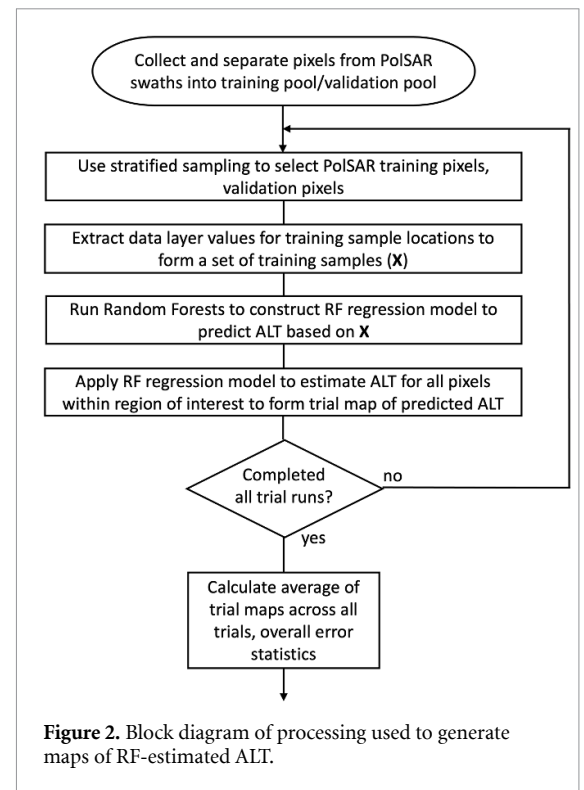
The forward model incorporates a three-layer soil dielectric model ([1], figure 3) including Layer 1: a surface active layer, Layer 2: a subsurface active layer, and Layer 3: the permafrost layer. Layer 1 is assumed to be fully thawed in August but frozen in October;

Layer 2 is assumed to be fully thawed and saturated with water throughout the time series; Layer 3 is permanently frozen. The total depth to permafrost (i.e. the total thickness of Layers 1 and 2) is assumed constant for the duration of the time series, because the upward freezing front extending from the upper permafrost layer has been found to be generally stable between late August and early October [26]. because Layer 2 is fully saturated, its water content, and so dielectric constant, depends solely on the soil porosity, which is constant. because Layer 3 is permanently frozen, its dielectric constant is a small known constant. Thus seven independent variables drive the forward model: Layer 1 dielectric constant and depth for August and October, Layer 2 dielectric constant and depth, and surface soil roughness. Based on these variables, the forward model calculates HH and VV radar backscatter for August and October.

The forward model is incorporated into an inversion technique [1] that estimates six independent variables, i.e. all but Layer 2 depth, through iterative use of simulated annealing optimization to minimize differences between the forward-model's radar backscatter predictions and the radar backscatter measurements for August and October. The remaining independent variable, Layer 2 depth, that is, ALT, is then found through an innovative process [1], the largest possible depth-assisted time-series approach, that resolves ambiguities to yield a final estimate of ALT. This inversion processing was applied to each pixel in the PolSAR swath, resulting in extensive and representative swaths of estimated ALT at 30 m resolution from which thousands of training and validation ALT samples can be extracted.

As noted in [1], the retrieved ALT values are generally underestimated for the sites where the *in situ* ALT is larger than the P-band sensing depth, with a retrieval bias ranging from -0.05 m to -0.24 m as validated against the *in situ* ALT measurements collected at CALM sites. For the sites where the *in situ* ALT is smaller than 0.55 m, the retrieval errors are generally less than 0.1 m.

The mean ALT uncertainty of the PolSAR-derived swaths for the 3 years ranges from 11.4 to 12.1 cm; the root-mean-square error (RMSE) relative to CALM sites lying within the PolSAR swaths for the 3 years ranges from 11.9 to 14.2 cm (if the largest few CALM ALTs that are larger than approximately 65 cm, and beyond the sensing depth of P-band PolSAR, are dropped, the range of RMSEs becomes 11.2–13.0 cm). Understanding the errors and uncertainties associated with the PolSAR ALT retrievals is of key importance for this study as the retrievals are used to train and evaluate our upscaling algorithm. Therefore, it is expected that any errors associated with our results will be similar to those in the PolSAR-based retrievals.



3. Methodology

The upscaling technique proposed in this paper for producing large scale maps of ALT from swaths of PolSAR-derived ALT is based on a technique originally developed in [28] to upscale soil moisture from sets of individual *in situ* measurements to field scale maps of soil moisture. The method relies on repeated runs of the Random Forests (RF) ensemble decision tree algorithm [29]. Random Forests was selected as the core machine learning tool in our method because it is capable of efficiently delivering highly accurate results compared to other machine learning algorithms even for very large datasets [30], and has exhibited good results with similar problems previously [31]. As noted in [32], it is well suited to estimation and mapping applications such as this one, easily handles a large number of input variables, and does not require the specification of a large number of input parameters as do some other algorithms such as support vector machines.

3.1. Overall regression approach

In each regression run, RF is used to construct a regression model expressing ALT as a nonlinear function of 36 or more independent spatial variables or model predictors.

A block diagram of the regression algorithm is shown in figure 2. Details of the processing are provided in the supplementary materials.

3.2. Independent variable spatial data layers

Independent variables for the regression were selected to represent vegetation, topographic features, and

microclimatic features affecting ALT development [33–40]. They are taken from publicly available spatial datasets that offer coverage for our entire northern Alaska region of interest. They include National Land Cover Dataset (NLCD) land cover; elevation, slope, aspect, and potential incident radiation; soil organic carbon (SOC), bulk density, clay fraction and sand fraction; thaw degree days (TDD) and (for some runs) freeze degree days (FDD); (for some runs) mean seasonal temperatures; summer monthly NDVI; and proximity to water. Details on these data layers used in the regression are provided in the supplementary materials.

ALT can also be substantially affected by surface water balance. Surface water balance (the ratio of evapotranspiration to precipitation, ET/P) is controlled by such climate factors as the fraction of precipitation falling as snow (e.g. a higher fraction of snow reduces ET/P) and by vegetation [41]. The climate effects are captured through our use of August and October TDD (in some cases augmented by August and October FDD as well as winter and spring mean seasonal temperatures), which are calculated from daily temperature data, whereas vegetation effects are accounted for through our use of NLCD land cover and 4 months of NDVI.

3.3. Validation approach

During the course of each RF regression trial run, Random Forests calculates a ranking of the relative importance of the various spatial variable data layers indicating which data layers it relied on most heavily in generating regression results. The ranking offers insight into which characteristics most affect the estimation of ALT. Layer importances (i.e. based on Random Forests's Gini impurity index values) were averaged across all trial runs used to obtain each final map of RF-estimated ALT.

The initial accuracy assessment for the final maps of RF-estimated ALT for each year was based on validation pixels collected across all n_{runs} trial runs. It consisted of comparing the PolSAR-derived value of ALT at each validation pixel location to the RF-estimated value of ALT in the final map at that location. The results of the comparison were quantified in error performance statistics including RMSE, bias, and unbiased RMSE (ubRMSE) as follows:

$$RMSE_{\text{val}} = \sqrt{\frac{1}{N_i} \sum_i (z_{\text{RF}}(i) - z_{\text{val}}(i))^2} \quad (1)$$

$$bias_{\text{val}} = \frac{1}{N_i} \sum_i (z_{\text{RF}}(i) - z_{\text{val}}(i)) \quad (2)$$

$$ubRMSE_{\text{val}} = \sqrt{(RMSE_{\text{val}})^2 - (bias_{\text{val}})^2} \quad (3)$$

in which $z_{\text{val}}(i)$ is the value of the i th validation pixel and $z_{\text{RF}}(i)$ is the value of the pixel at that same location in the final map of RF-estimated ALT. The computation was repeated for all 3 years considered in the study.

To provide an independent accuracy assessment of the maps for each year, CALM site *in situ* ALT measurement values were compared to RF-estimated values at the site level using error statistics analogous to those just described. The processing and use of CALM site measurements in the validation of our RF-predicted maps of ALT is discussed in the supplementary materials.

The maps were also compared to a small number of individual ALT samples provided by researchers in the field [12–14].

Finally, maps of RF-estimated ALT for 2014 were compared to an Alaska-wide map of estimated ALT provided in Pastick *et al* [17] (hereafter referred to as the 'Pastick map'). The Pastick map incorporated data from multiple years up to 2013, rather than addressing a specific year. The Pastick map was developed based on a piecewise linear regression to upscale from *in situ* ALT measurements using spatial data layers including topographic, land cover, Landsat imagery, Landsat-derived indices, and climate variables.

Other measures used to assess the characteristics of the RF-estimated ALT values produced by this method include boxplots and histograms.

4. Results

4.1. Maps of RF-predicted ALT

Final maps of RF-estimated ALT are shown in figure 3.

As the maps show, the pattern in RF-predicted ALT is generally shallower on the Alaskan North Slope and deeper in more southerly locations, including patches in and near the Seward Peninsula and a swath over the Brooks Range. Temporal variations in RF-predicted ALT are also evident; throughout the North Alaska coverage area, RF-predicted ALTs increase between 2014 and 2015, then decrease between 2015 and 2017. On the North Slope, this results in an overall drop in mean estimated ALT of about 0.4 cm; in the more southerly locations, where the decrease is much smaller, it results in an overall increase in mean estimated ALT of about 1.3 cm. These variations in RF-predicted ALT are consistent with changes that can be observed in histogram plots of August TDD for the two regions, as shown in the supplementary materials.

A small pattern of seemingly deep estimated ALT near Prudhoe Bay and the Sag River delta appears to have become slightly less prominent, although in very-near coastal regions such as this there are concerns about the possibility of estimation errors resulting from the effects of seawater intrusion, a factor that was not explicitly included when developing

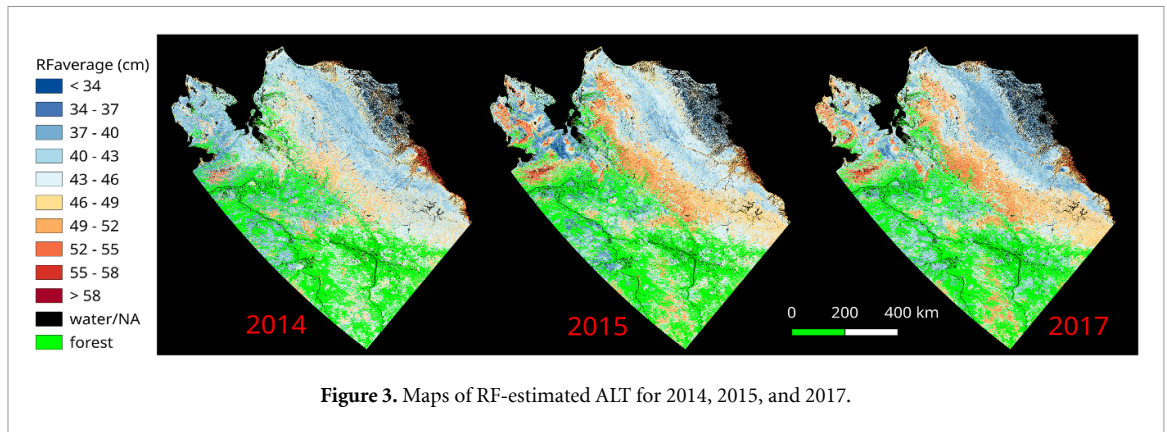


Figure 3. Maps of RF-estimated ALT for 2014, 2015, and 2017.

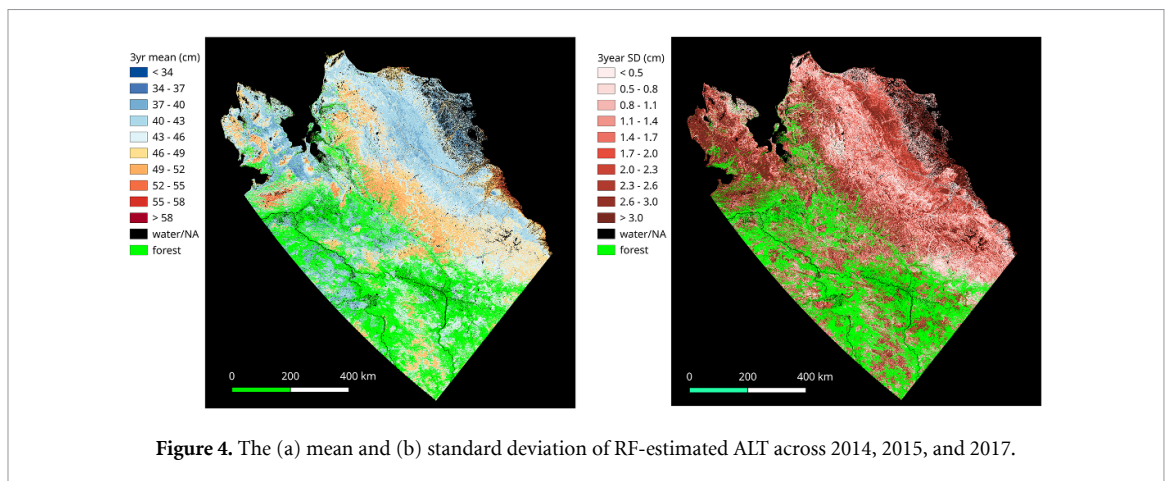


Figure 4. The (a) mean and (b) standard deviation of RF-estimated ALT across 2014, 2015, and 2017.

the retrieval algorithm that formed our swaths of PolSAR-derived ALT (based on the simulation results in figure 4(b) of [1], for example, the retrieval algorithm sets the ratio of imaginary to real parts of the complex dielectric constant to 0.15, an assumption that may be less reliable in the presence of seawater).

The mean and (unbiased) standard deviation of RF-predicted ALT calculated across all 3 years in the study period are shown in figure 4. The map of mean RF-predicted ALT shows that the deepest predicted values occur near Prudhoe Bay and the Sag River delta, in patches on and near the Seward Peninsula, and in a swath over the Brooks Range. The average value of RF-predicted ALT across all valid pixels in the 3 year mean map is 44.4 cm. The map of standard deviations of RF-predicted ALT indicates that the greatest inter-annual variability exists on the North Slope between Barrow and Deadhorse as well as on and near the Seward Peninsula.

4.2. Importance of data layers in the generation of RF results

The relative importances of the various predictor variable data layers listed in table 2 of the supplementary materials to the results of each Random Forests run are automatically calculated based on the run's Gini impurity index values. The importances are then averaged across 10 trial runs for 2014, 2015, and 2017,

and plotted in figure 5. Only the 12 most important layers for each year are shown. As the plot indicates, the relative order and importance of the predictors are largely consistent across the three different years. The most important layers affecting estimated ALT were SOC, particularly from the second horizon (denoted as Carbon2 in figure 5 at 5–15 cm depth [42]), slope, and elevation. Further discussion of the data layer importances is provided in the supplementary materials.

4.3. Accuracy of upscaling relative to PolSAR-derived ALT pixels

Error performance statistics for the final RF-estimated maps of ALT relative to set-aside PolSAR-derived validation samples are shown in the left-hand columns of table 1. As the table shows, RMSE values relative to these samples run from approximately 7.5 to 9.1 cm. These values constitute about 20% of the overall 25–65 cm range of measurable ALT values.

Additionally, to provide a better sense as to whether our results are overfitted, error statistics relative to the PolSAR-derived training data pixels are shown in the right-hand columns of table 1. Perhaps expectedly, the error performance relative to the training pixels is similar to, but slightly better than, that relative to the validation samples. This confirms that Random Forests is not overfitting the data.

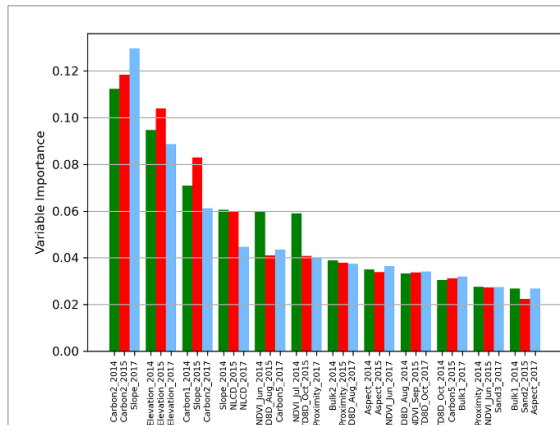


Figure 5. The relative importances of the various predictor variable data layers in generating the results of RF runs for 2014 (green), 2015 (red), and 2017 (blue). The data layers used in the study are as listed in table 2 of the supplementary materials. Integer suffixes for the various soil properties layers denote horizon number; for example, Carbon2 denotes the SOC for horizon 2 (i.e. 5–15 cm depth).

4.4. Accuracy of upscaling relative to CALM *in situ* measurements

CALM *in situ* measurement locations and plots of RF-predicted ALT versus CALM-measured ALT are shown in figure 6. The standard deviation of each site's CALM measurements, calculated across all *in situ* measurement locations included in the site mask, is shown as horizontal error bars. The error range of each site's RF-estimated value of ALT, calculated as the RMS combination of uncertainty in PolSAR-derived ALT swath values and standard deviation of RF-estimated ALT values within the site mask, is shown as vertical error bars. The latter are only included, however, for the case of CALM sites overlying a swath of PolSAR-derived ALT, because the uncertainty in PolSAR-derived ALT swath values is only available within the swaths. These bars are included to offer the reader a general sense for how large the overall errors tend to run. For comparison, the figure also includes plots of PolSAR-derived ALT relative to CALM-measured ALT for CALM sites overlying a swath.

For the case of CALM sites lying within a swath of PolSAR-derived ALT, figures 6(b) through (d) can be compared to figures 6(e) through (g) to show that the values of RF-predicted ALT for each CALM site are generally similar to the corresponding values of PolSAR-derived ALT, and within a range of uncertainty dominated by that of the PolSAR-derived ALT. The error statistics for the RF-predicted cases are slightly improved relative to those of the PolSAR-derived cases. Although all six plots exhibit significant negative biases relative to the CALM measurements, the higher RF-predicted results show somewhat of an upward trend with increasing CALM values.

Table 2 presents summary error performance statistics for the final RF-estimated maps of ALT relative to the *in situ* measurements of at least 23 CALM sites.

Values of RF-estimated ALT at the CALM sites exhibit RMS errors relative to the CALM *in situ* measurements that are comparable to the uncertainties of the ALT values in the PolSAR-derived swaths. A strong negative bias relative to the *in situ* CALM measurements is also evident in table 2. These may have been enlarged somewhat by the fact that *in situ* ALT measurements at three CALM sites (Deadhorse, Franklin Bluff, and Council Grid) are at or beyond the range of ALTs capable of being sensed through inversion of P-band PolSAR (these sites are visible as the three rightmost points in figure 6 parts (c), (d), (f), and (g)). Even for these very large values of CALM-measured ALT, however, the plots show that the corresponding values of RF-predicted ALT trend upwards as the value of CALM-measured ALT increases.

Boxplots comparing the range of CALM-measured values to the range of RF-estimated ALT values at the CALM sites within our northern Alaska coverage region are shown in figure 7. As expected, the plots show that RF-predicted maps are unable to capture the upper range of values measured at the CALM site locations, which is a problem stemming from the PolSAR-derived ALT.

Although the year-over-year changes in both CALM-measured and RF-predicted ALT are small, it can be seen that the median value of the CALM-measured ALT increases for every year included in the study, whereas the median value of RF-predicted ALT dips slightly in 2015, then rises to its largest value in 2017. In this regard, the RF-predicted results diverge from the CALM measurement results within the limited 3 year sampling period of this study. The inter-annual variations are small, however, and our CALM data, collected at only a few dozen irregularly-dispersed sites, may not be entirely representative of temporal trends throughout the North Alaska coverage region that factor into the formation of the RF-predicted maps.

Averaged over all included CALM sites for the 3 years of the study, the mean CALM-measured ALT, at about 51.1 cm, was about 4.8 cm deeper than the mean RF-predicted ALT.

Our maps of RF-estimated ALT were also compared to independent ALT measurements provided by researchers in the field [12–14]. Most of the independent ALT measurements were in the southernmost forested reaches of the coverage area and thus were either in areas where the RF-estimated ALT maps are masked out or were found to exceed 65 cm, i.e. outside of the scope of the study defined from the characteristic P-band sensing depth. Only five independent measurements for 2015 were found to be usable. When those were compared to values taken

Table 1. Error performance statistics for RF-estimated maps of ALT relative to PolSAR-derived ALT samples. Error statistics in the left-hand columns are relative to the PolSAR-derived validation samples; error statistics in the right-hand columns are relative to the PolSAR-derived training samples. All values are in centimeters.

Year	Relative to Validation			Relative to Training		
	RMSE	Bias	ubRMSE	RMSE	Bias	ubRMSE
2014	7.45	-0.01	7.45	7.27	0.0	7.27
2015	8.50	-0.04	8.50	8.35	0.0	8.35
2017	9.12	-0.03	9.12	9.01	0.0	9.01

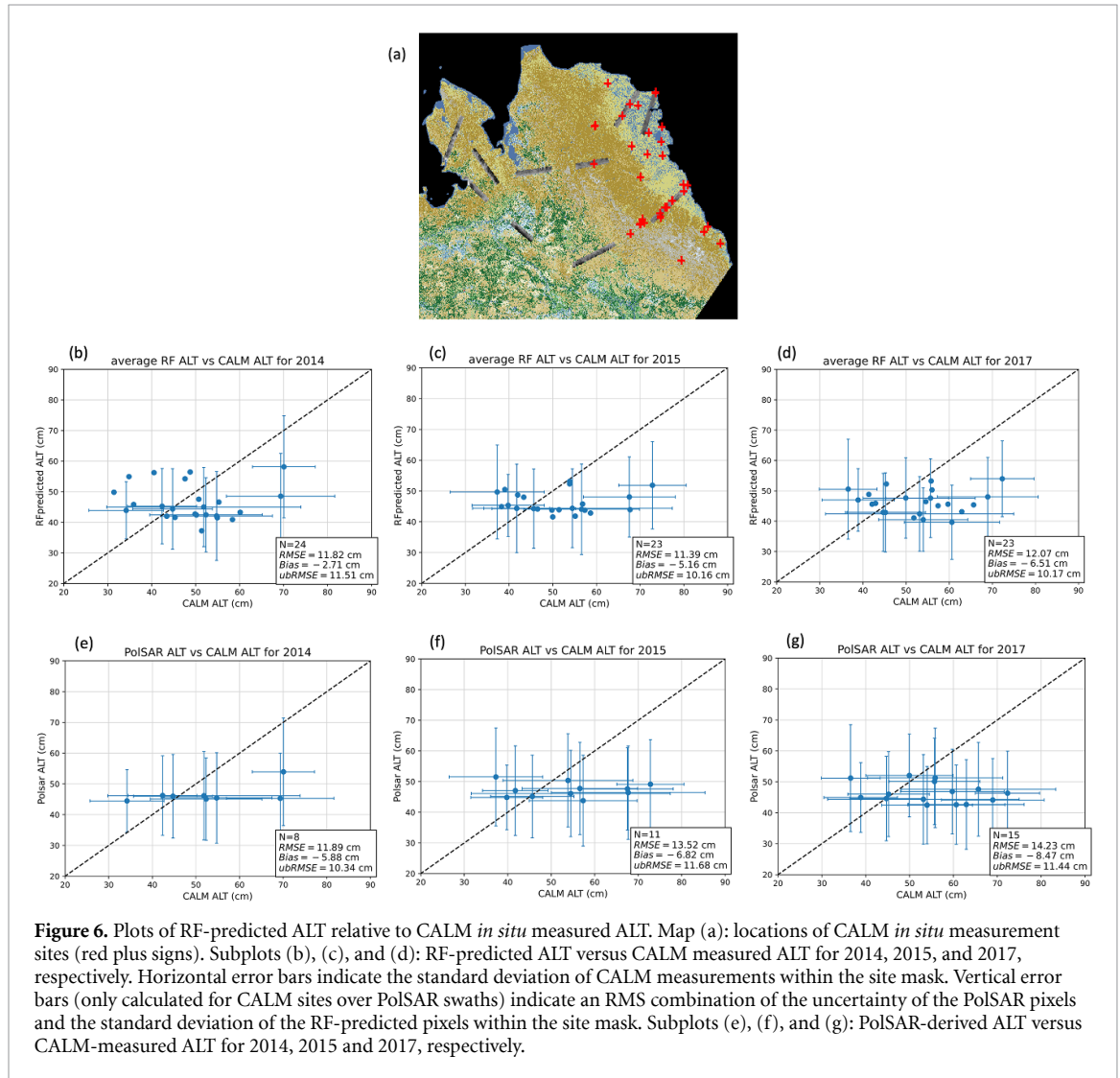


Table 2. Error performance statistics for RF-estimated maps of ALT relative to *in situ* measurements collected at 23 CALM sites. All values are in centimeters.

Year	RMSE	Bias	ubRMSE
2014	11.82	-2.71	11.51
2015	11.39	-5.16	10.16
2017	12.07	-6.51	10.17

from the 2015 RF-predicted ALT map for the same locations, the mean measurement value was found to exceed the mean RF-predicted value by 9.2 cm.

4.5. Histograms of ALT values throughout the North Alaska study area

Histograms of RF-estimated ALT throughout the study region are shown in figure 8.

The histograms for both 2015 and 2017 exhibit a bulge on the high end of the distribution indicating that these years each had stronger representation for deeper ALT values than 2014. The plots for 2017 show greater representation of shallow ALT values and deep ALT values, with a slight dip in the distribution of mid-range ALT values. Comparing this distribution to the spatial pattern of ALT seen in

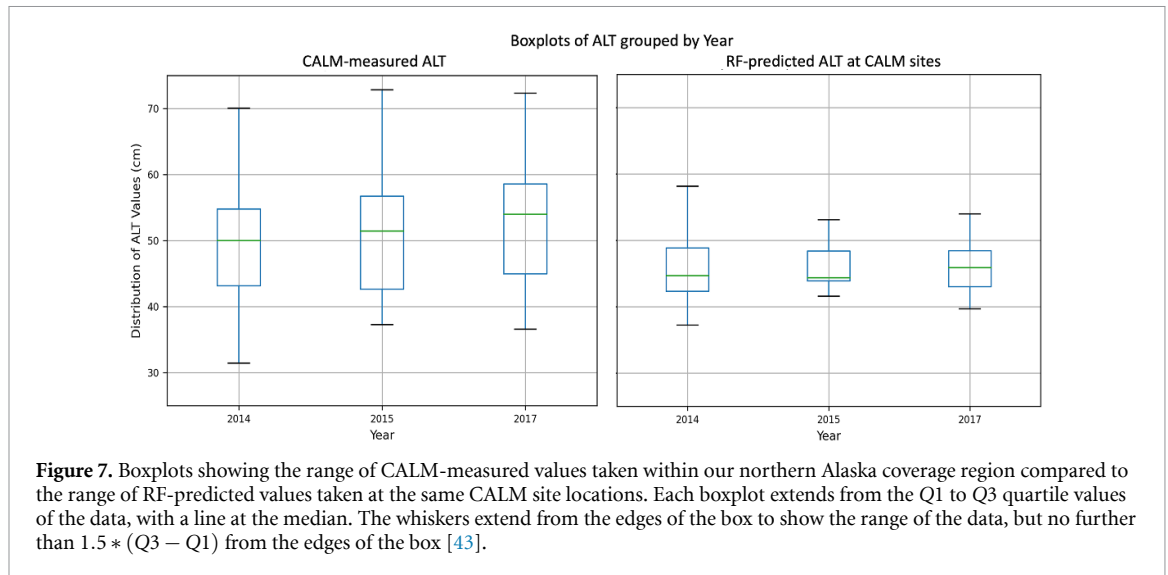


Figure 7. Boxplots showing the range of CALM-measured values taken within our northern Alaska coverage region compared to the range of RF-predicted values taken at the same CALM site locations. Each boxplot extends from the Q1 to Q3 quartile values of the data, with a line at the median. The whiskers extend from the edges of the box to show the range of the data, but no further than $1.5 * (Q3 - Q1)$ from the edges of the box [43].

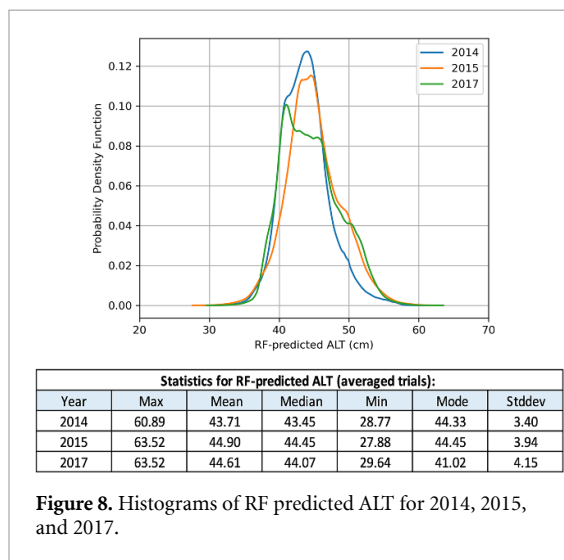


Figure 8. Histograms of RF predicted ALT for 2014, 2015, and 2017.

figure 3, it seems plausible that the peak at the lower end of the distribution is a consequence of the extensive areas of relatively shallow values of ALT seen on the North Slope in 2017, whereas the peak at the upper end of the distribution reflects the deeper values of ALT seen over the Brooks Range and in patches on and near the Seward Peninsula. Overall, the distribution of RF-predicted ALT seems to have shifted towards deeper values between 2014 and 2015, then partially retreated between 2015 and 2017. This is consistent with thermal variations apparent in histograms of TDD and mean seasonal temperature, which show TDD/temperature running low in 2014, then much higher in 2015, then backing off slightly in 2017. Histograms of TDD and mean seasonal temperature are included in the supplementary materials.

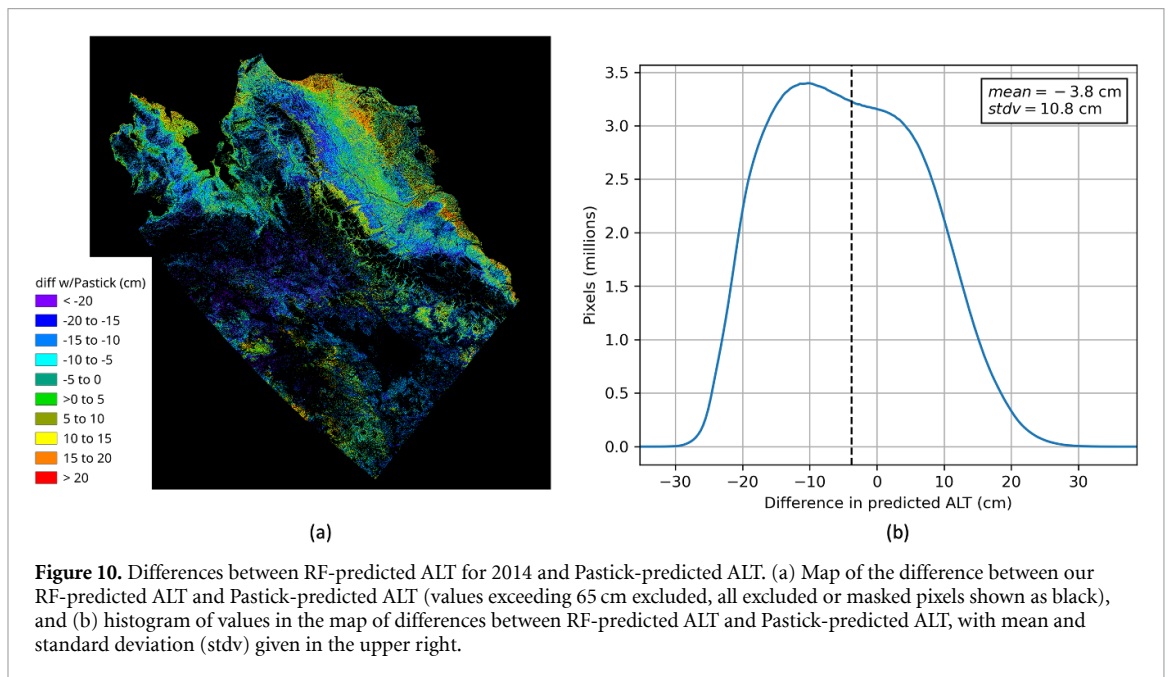
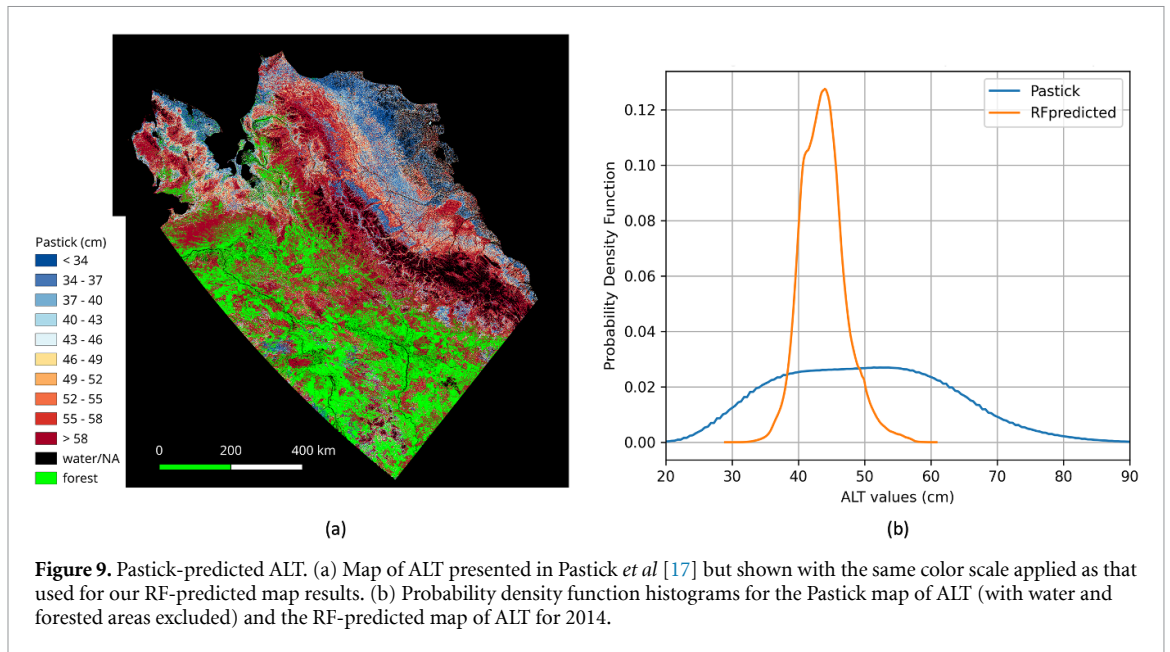
The effect of the Brooks Range on the overall statistics of our RF-predicted ALT is discussed in the supplementary materials.

4.6. Comparison with Pastick map of estimated ALT

Our maps of RF-estimated ALT were also compared to a pre-existing map of ALT in Alaska, namely, the map of ALT presented as figure 3(c) in [17] (referred to as the ‘Pastick map.’). This map was based on inventory data collected up to 2013. The portion of the map overlapping our coverage area is shown in figure 9(a), but with the color mapping set to match that used to present our RF-predicted results. Numerous pixels in the Pastick map are assigned a value of 101, which denotes a value of ALT greater than 1 m. A probability density function histogram for the Pastick map is shown in figure 9(b). The range of the plot does not extend to 101 so as to not distort the vertical scale of the distribution; water and forested areas are excluded such that the histogram only represents pixels containing valid data in the RF-predicted maps. A histogram for the RF-predicted ALT from 2014 (i.e. the closest year to when the Pastick map data were collected) is also included in the plot for comparison.

The geographic patterns of ALT in the Pastick map are similar to those in the 2014 RF-predicted map, with both maps similarly affected by, particularly, topographic spatial variations, in most places. However, the Pastick map predicts values of ALT to be deeper than those in the RF-predicted maps in and south of the Brooks Range. There are also many places where the Pastick map shows shallower ALT values than those of the RF-predicted map. The histogram of the Pastick map also displays a broader distribution than that of the RF-predicted ALT map, which is to be expected because the predicted ALT range of the RF product is set by the P-band radar sensing depth [1].

A plot of Pastick-estimated ALT versus CALM-measured ALT is presented in the supplementary materials.



A difference map was generated between our RF-predicted map for 2014 (closest year to data in the Pastick map) and the Pastick map, as is shown in figure 10(a). Pixels with values of ALT exceeding 65 cm were eliminated from the Pastick map prior to forming a difference raster because those values are beyond the range of ALTs that can be measured using PolSAR (and thus beyond the range of values in our RF-predicted maps). The resulting difference map and a histogram of the ALT difference values from the map are shown in figures 10(a) and (b), respectively.

The difference map shows differences in the range of ± 25 cm between the Pastick and RF-predicted maps, as well as numerous gaps south of the Brooks Range resulting from the decision to exclude pixels

for which the Pastick map value exceeds 65 cm. This is expected because the process used to develop the RF-predicted map is substantially different from that used to develop the Pastick map. The RF maps were trained based on swaths of PolSAR-derived ALT rather than *in situ* ALT measurements, and our decision tree regression approach differed from that implemented in [17]. Additionally, the RF map was calculated based on a different set of independent variable data layers from those used to produce the Pastick map. This may partly reflect the fact that the Pastick map and our maps sought to realize different goals, with [17] developing a single state-wide map of ALT based on data collected over multiple years up to the time of that study, whereas we sought to develop a method for spatially extrapolating

airborne SAR-based ALT retrievals over northern Alaska independently for each year included in our study. The Pastick map relied on late-season Landsat images, four Landsat-derived vegetation index layers, six infrared Landsat-derived index layers, and seven decadal average seasonal and annual weather variables based on climate data from the Scenarios Network for Alaska and Arctic Planning (SNAP). TDD and FDD thermal layers, monthly NDVI layers, soil data layers, and proximity to water were not used in [17]. Also, the land cover types masked out of the RF-predicted maps (i.e. open water and forested areas) differ from those masked out of the Pastick map (open water, barren land, perennial ice/snow, cultivated, and developed land covers, all per the NLCD).

Statistics accompanying the histogram of the difference map indicate a standard deviation of almost 11 cm around a mean of about -3.8 cm.

5. Discussion

5.1. Choice of data layers

Some considerations involved in our selection of data layers to include in the regression are discussed in the supplementary materials.

5.2. Observations from maps of RF-estimated ALT

One experiment carried out was evaluating the RF regression model trained based on PolSAR data collected in 2015 to generate RF-predicted results for all three study years. This was done in order to determine how much year-to-year differences in PolSAR flight trajectories might be contributing to the apparent temporal variations in RF-estimated ALT. The resulting sequence of maps did not exhibit a clear temporal trend.

In some locations, the maps can highlight the effect on ALT of local spatial variations in certain data layers. In one example of this on the North Slope, low elevation areas of mixed sedge and emergent wetland near Barrow have shallower RF-estimated ALTs, while similar areas near Deadhorse have deeper RF-predicted ALTs, as can be seen in figure 11.

Although the difference in ALTs may be partly explained by the fact that Barrow is about 1 degree of latitude farther north than Deadhorse, the difference in mean annual temperature between the two sites is small ($\ll 1$ °C). Nonetheless, the wide river-/drainage corridors around Deadhorse exhibit substantially larger values of ALT than those in the relatively flat Barrow area. The boxplots in figure 11(c) show the dependence of ALT on slope, with ALT becoming shallower with increasing slope for lower values of slope and deeper with increasing slope for higher values of slope. A contributing factor to the behavior for higher values of slope may be the fact that the Deadhorse corridors are surrounded by slightly steeper areas, which may channel subsurface

water flow towards them thereby resulting in greater groundwater flow and less freezing; this effect may be minimal for lower values of slope.

ALT could also be influenced by differences in the characteristics of surficial deposits (e.g. young marine [peat, pebbly silty sand] versus old fluvial deposits [peat, silt, sand, gravel]), differences that could be reflected in the organic carbon and sand data layers. Boxplots for the fraction of organic carbon in the second (i.e. 5–15 cm depth) horizon, figure 11(d), show Barrow with higher overall values of organic carbon than Deadhorse. Also, the plots for both sites exhibit a distinct negative dependence of ALT on organic carbon, although the trend is less pronounced for Barrow. This relation between organic carbon and ALT perhaps reflects the slower decomposition of organic matter, and so increased fraction of soil carbon, in the coldest regions with the smallest values of ALT. Boxplots for the fraction of sand in the second horizon, figure 11(e), show a positive dependence of ALT on sand fraction in the case of Deadhorse, but no clear trend in the case of Barrow. The positive dependence for the Deadhorse location may result from a likely increase in hydraulic conductivity, with corresponding increase in subsurface water flow, that accompany increasing sand fraction; as mentioned above, such increases in water flow may slow/reduce freezing. The relatively static variation of ALT with sand fraction for the Barrow location may indicate that the far larger values of organic carbon fraction present at Barrow are sufficient to dominate the effect of soil composition on permafrost formation.

This example demonstrates how our method can be a useful tool for investigating how local variations in various spatial variables affect the value of ALT. Within a given locality, data layers whose spatial patterns are similar to those in the map of RF-predicted ALT can be identified as possible contributors to spatial variations in ALT for that locality. Boxplots of predicted ALT versus each of these independent variables can reveal how changes in the value of that variable influence the value of ALT, potentially increasing insight into the physical processes driving local changes in ALT.

5.3. Limitations of our approach

Several outstanding performance issues are subjects of our ongoing work.

One of the most important factors is the limited penetration depth of P-band PolSAR in arctic soils, which restricts achievable PolSAR-derived ALT estimates to no more than about 65 cm. The PolSAR-derived swaths of ALT used as training data in the RF regression are thus unable to capture ALT values exceeding this depth. Because RF cannot generate regression estimates beyond the range of the available training data, our method is unable to represent values of ALT exceeding 65 cm. Although this is deep enough for most locations in northern Alaska

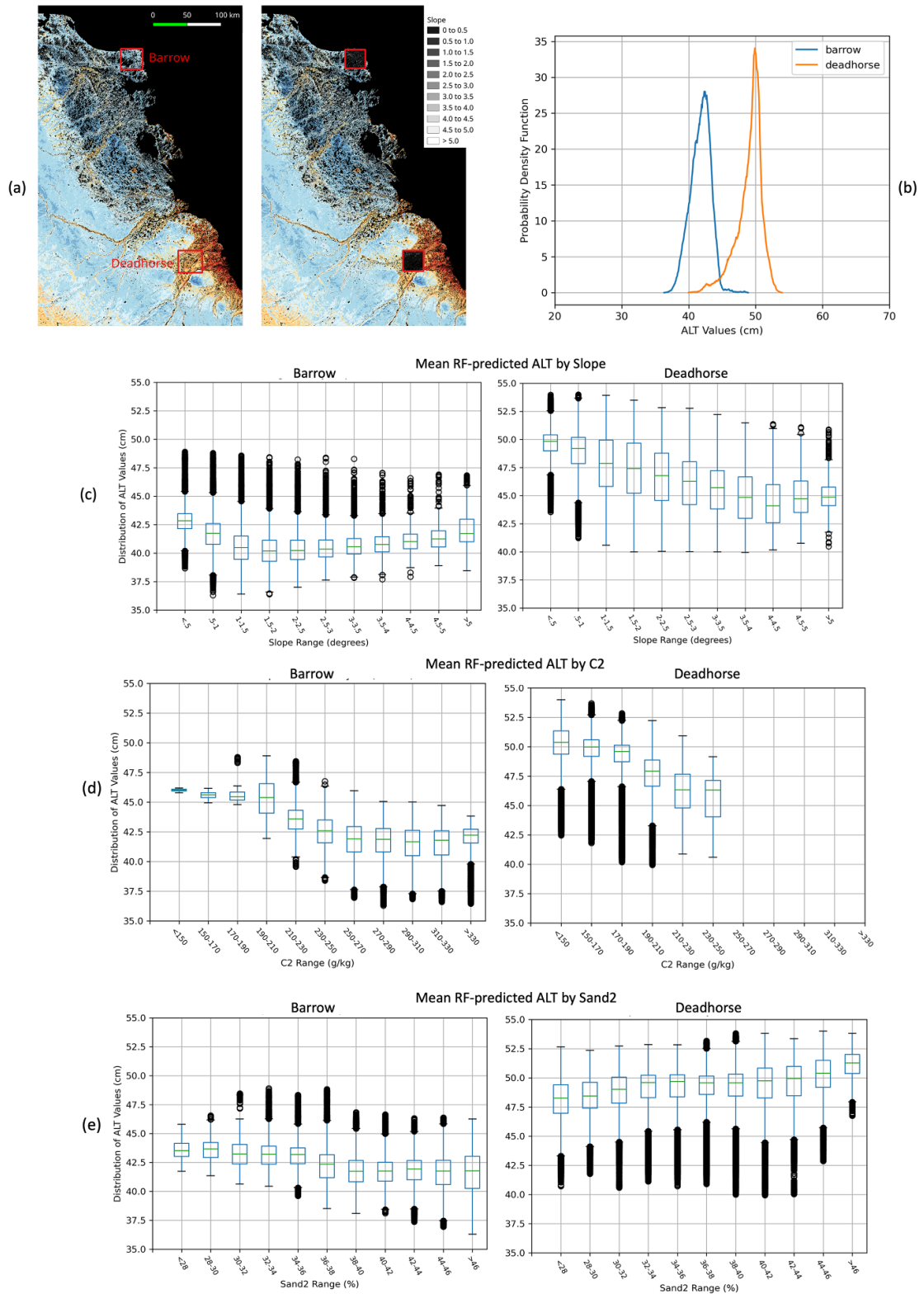


Figure 11. Detail of RF-estimated ALT for small areas in the vicinity of Deadhorse and Barrow. (a) RF-estimated ALT on the North Slope overlaid with slope images covering Deadhorse and Barrow, (b) histogram of RF ALT for the two areas, (c) Boxplots of mean RF-predicted ALT grouped by slope value, (d) boxplots of mean RF-predicted ALT grouped by C2 value, where C2 is the organic carbon value in the second horizon (i.e. 5–15 cm depth), (e) boxplots of mean RF-predicted ALT grouped by Sand2 value, where Sand2 is the sand value in the second horizon. For parts (c) through (e), each boxplot extends from the Q1 to Q3 quartile values of the data, with a line at the median. The whiskers extend from the edges of the box to show the range of the data, but no further than $1.5 * (Q3 - Q1)$ from the edges of the box. Outliers are plotted as black circles [43].

(as substantiated, for example, in figure 2 of [17]), it is too shallow for many areas south of the Brooks Range. Some solutions to address the issue include

future development of lower-frequency radars, radar systems with lower system noise floor to allow more sensitivity to faint signals scattered from subsurface

layers, and/or addition of other observation modalities such as InSAR. Several future tasks aimed at mitigating the problem are proposed in section 6.

Also, because RF forms its decision trees by minimizing the mean squared error across all training samples, its regression models tend to favor the most common values of ALT; this reduces the likelihood of predicting very low or very high values of ALT. As previously mentioned, stratified sampling of the training data as described in section 3.1 partially compensates for this effect.

The RF predictions can only be as accurate as the PolSAR-derived ALT data used for model training, which bounds the error performance achievable using our method. This is evident, for example, in the error statistics reported in section 4.4. These constraints also limit our ability to detect inter-annual variability and longer-term climate trends given that we might expect this level of ALT variability to be 2–3 cm and well within our noise floor. Therefore, our approach is expected to be more robust in detecting regional ALT spatial patterns, which cover a larger distributional range.

6. Conclusions

We have developed a method of upscaling ALT from swaths of high-resolution P-band PolSAR-derived ALT. The method has been shown capable of producing maps of estimated ALT with values of up to about 65 cm over an extensive region in northern Alaska. Error analysis of the maps evaluated relative to the values of a large number of set-aside validation pixels yields RMSEs of no more than 9.1 cm (i.e. about 23% of the full (~25–65 cm) range of ALT values retrievable using P-band PolSAR) and bias errors of magnitude less than 0.1 cm. Error analysis relative to *in situ* ALT measurements available from the CALM network leads to RMSEs of less than 12.1 cm and bias errors of magnitude no more than about 6.5 cm (i.e. about 29% and 16%, respectively, of the full (~31–73 cm) range of ALT values measured at all CALM sites included in the study).

Thus, the upscaling does a reasonably good job of predicting PolSAR-derived ALT values but is unable to capture the full range of CALM *in situ* measurements. Our maps exhibit minimal bias and high levels of accuracy comparable to the PolSAR-derived ALT retrievals [1] used for model training, and relative to more comprehensive ALT retrievals derived from combined L-band InSAR/P-band PolSAR retrievals over Alaska [44].

In summary, our maps provide estimated ALT at 30 m resolution for a succession of years throughout the northern half of Alaska in the Canada Albers Conical Equal Area projection. The maps offer a new source of insights into the influence of climate, land cover, and terrain factors affecting active layer development and permafrost stability in the Arctic.

6.1. Future work under consideration

Some options for future work that could help to further the goals of this project are considered in the supplementary materials.

Data availability statement

The data cannot be made publicly available upon publication because the cost of preparing, depositing and hosting the data would be prohibitive within the terms of this research project. The data that support the findings of this study are available upon reasonable request from the authors. Data will be available from 12 January 2024.

Acknowledgment

This work was supported by NASA as part of the Arctic and Boreal Vulnerability Experiment (ABOVE), a NASA Terrestrial Ecology project, under Grant No. 80NSSC19M0114.

Disclaimer

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

References

- [1] Chen R H, Tabatabaenejad A and Moghaddam M 2019 Retrieval of permafrost active layer properties using time-series P-band radar observations *IEEE Trans. Geosci. Remote Sens.* **57** 6037–54
- [2] Zhao D and Wu S 2019 Projected changes in permafrost active layer thickness over the Qinghai–Tibet plateau under climate change *Water Resour. Res.* **55** 7860–75
- [3] Peng X, Zhang T, Frauenfeld O W, Wang K, Luo D, Cao B, Su H, Jin H and Wu Q 2018 Spatiotemporal changes in active layer thickness under contemporary and projected climate in the Northern Hemisphere *J. Clim.* **31** 251–66
- [4] Pang Q, Zhao L, Li S and Ding Y 2012 Active layer thickness variations on the Qinghai–Tibet plateau under the scenarios of climate change *Environ. Earth Sci.* **66** 849–57
- [5] Anisimov O A, Shiklomanov N I and Nelson F E 1997 Global warming and active-layer thickness: results from transient general circulation models *Glob. Planet. Change* **15** 61–77
- [6] Brown J, Hinkel K M and Nelson F E 2000 The circumpolar active layer monitoring (CALM) program: research designs and initial results *Polar Geogr.* **24** 166–258
- [7] Nelson F E and Hinkel K M 2003 Methods for measuring active-layer thickness *A Handbook on Periglacial Field Methods* ed O Humlum and N Matsuoka (University of the North in Svalbard)
- [8] Chen A, Parsekian A, Schaefer K, Jafarov E, Panda S K, Liu L, Zhang T and Zebker H A 2015 Pre-ABOVE: ground-penetrating radar measurements of ALT on the Alaska North Slope (<https://doi.org/10.3334/ORNLDAC/1265>)
- [9] Shiklomanov N I, Nelson F E, Streletskiy D A, Hinkel K M and Brown J 2008 The circumpolar active layer monitoring (CALM) program: data collection, management and dissemination strategies *Proc. 9th Int. Conf. on Permafrost* vol 1 pp 1647–52
- [10] Shiklomanov N, Streletskiy D and Nelson F 2012 Northern Hemisphere component of the global circumpolar active

- layer monitoring (CALM) program *Proc. 10th Int. Conf. on Permafrost* vol 1 pp 377–82
- [11] Luo D, Wu Q, Jin H, Marchenko S S, Lü L and Gao S 2016 Recent changes in the active layer thickness across the Northern Hemisphere *Environ. Earth Sci.* **75** 555
- [12] Minsley B J, Pastick N J, Wylie B K and Kass M A 2015 Fire impacts on permafrost in Alaska: geophysical and other field data collected in 2014: U.S. Geological Survey data release (<https://doi.org/10.5066/F7959FM0>)
- [13] Kass M A, Minsley B J, Pastick N J, Wylie B K, Brown D R N and Chase B F W 2018 Fire impacts on permafrost in Alaska: geophysical and other field data collected in 2015: U.S. Geological Survey data release (<https://doi.org/10.5066/F7F18WT>)
- [14] Minsley B J, Pastick N J, Wylie B K, Kass M A, Ebel B A, Rey D M, Chase B F W and James S R 2018 Alaska permafrost characterization: geophysical and related field data collected from 2016–2017: U.S. Geological Survey data release (<https://doi.org/10.5066/P99PTGP4>)
- [15] Engstrom R, Hope A, Kwon H, Stow D and Zamolodchikov D 2005 Spatial distribution of near surface soil moisture and its relationship to microtopography in the Alaskan Arctic coastal plain *Hydrol. Res.* **36** 219–34
- [16] Gangogadamage C et al 2014 Extrapolating active layer thickness measurements across Arctic polygonal terrain using LiDAR and NDVI data sets *Water Resour. Res.* **50** 6339–57
- [17] Pastick N J, Torre Jorgenson M, Wylie B K, Nield S J, Johnson K D and Finley A O 2015 Distribution of near-surface permafrost in Alaska: estimates of present and future conditions *Remote Sens. Environ.* **168** 301–15
- [18] Nelson F, Shiklomanov N, Hinkel K and Brown J 2008 Decadal results from the Circumpolar Active Layer Monitoring (CALM) program *9th Int. Conf. on Permafrost* vol 9 pp 1273–80
- [19] Liu L, Schaefer K, Zhang T and Wahr J 2012 Estimating 1992–2000 average active layer thickness on the Alaskan North Slope from remotely sensed surface subsidence *J. Geophys. Res.* **117** F01005
- [20] Li Z, Zhao R, Hu J, Wen L, Feng G, Zhang Z and Wang Q 2015 InSAR analysis of surface deformation over permafrost to estimate active layer thickness based on one-dimensional heat transfer model of soils *Sci. Rep.* **5** 15542
- [21] Yi Y, Kimball J S, Chen R H, Moghaddam M, Reichle R H, Mishra U, Zona D and Oechel W C 2018 Characterizing permafrost active layer dynamics and sensitivity to landscape spatial heterogeneity in Alaska *Cryosphere* **12** 145–61
- [22] Miller C E et al 2019 An overview of ABoVE airborne campaign data acquisitions and science opportunities *Environ. Res. Lett.* **14** 080201
- [23] ABOVE Science Definition Team 2018 A concise experiment plan for the Arctic-Boreal vulnerability experiment (<https://doi.org/10.3334/ORNLDAAC/1617>)
- [24] Chen R H, Tabatabaenejad A and Moghaddam M 2019 ABoVE: active layer and soil moisture properties from AirMOSS P-band SAR in Alaska (<https://doi.org/10.3334/ORNLDAAC/1657>)
- [25] Burgin M, Clewley D, Lucas R M and Moghaddam M 2011 A generalized radar backscattering model based on wave theory for multilayer multispecies vegetation *IEEE Trans. Geosci. Remote Sens.* **49** 4832–45
- [26] Chen R H, Tabatabaenejad A and Moghaddam M 2018 P-band radar retrieval of permafrost active layer properties: time-series approach and validation with *in-situ* observations *IGARSS 2018—2018 IEEE Int. Geoscience and Remote Sensing Symp.* pp 6777–9
- [27] Multiresolution Land Characteristics Consortium (MRLC) National Land Cover Database (NLCD), Alaska 2016 edition (available at: www.mrlc.gov/data/nlcd-2016-land-cover-alaska)
- [28] Clewley D, Whitcomb J, Akbar R, Silva A, Berg A, Adams J, Caldwell T, Entekhabi D and Moghaddam M 2017 A method for upscaling *in situ* soil moisture measurements to satellite footprint scale using random forests *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **10** 2663–73
- [29] Breiman L 2001 Random forests *Mach. Learn.* **45** 5–32
- [30] Fernandez-Delgado M, Cernadas E, Barro S and Amorim D 2014 Do we need hundreds of classifiers to solve real world classification problems? *J. Mach. Learn. Res.* **15** 3133–81
- [31] Forkuor G, Hounkpatin O K L, Welp G and Thiel M 2017 High resolution mapping of soil properties using remote sensing variables in south-western Burkina Faso: a comparison of machine learning and multiple linear regression *PLoS One* **12** 3133–81
- [32] Zhang R, Li Y, Goh A T C, Zhang W and Chen Z 2021 Analysis of ground surface settlement in anisotropic clays using extreme gradient boosting and random forest regression models *J. Rock Mech. Geotech. Eng.* **13** 1478–84
- [33] Fisher J P, Estop-Aragónés C, Thierry A, Charman D J, Wolfe S A, Hartley I P, Murton J B, Williams M and Phoenix G K 2016 The influence of vegetation and soil characteristics on active-layer thickness of permafrost soils in boreal forest *Glob. Change Biol.* **22** 3127–40
- [34] Hinkel K M and Nelson F E 2003 Spatial and temporal patterns of active layer thickness at circumpolar active layer monitoring (CALM) sites in Northern Alaska, 1995–2000 *J. Geophys. Res.: Atmos.* **108** 8168
- [35] Li A, Tan X, Wu W, Liu H and Zhu J 2017 Predicting active-layer soil thickness using topographic variables at a small watershed scale *PLoS One* **12** 1–17
- [36] Mazhitova G, Malkova (Ananjeva) G, Chestnykh O and Zamolodchikov D 2004 Active-layer spatial and temporal variability at European Russian Circumpolar-Active-Layer-Monitoring (CALM) sites *Permafr. Periglac. Process.* **15** 123–39
- [37] Oht M 2003 Impact of meteorological factors on active layer development in Central Spitsbergen *Permafrost* ed Phillips, Springman and Arenson (Swets and Zeitlinger) pp 845–50
- [38] Sazonova T S and Romanovsky V E 2003 A model for regional-scale estimation of temporal and spatial variability of active layer thickness and mean annual ground temperatures *Permafr. Periglac. Process.* **14** 125–39
- [39] Strand S M, Christiansen H H, Johansson M, Åkerman J and Humlum O 2021 Active layer thickening and controls on interannual variability in the Nordic Arctic compared to the circum-Arctic *Permafr. Periglac. Process.* **32** 47–58
- [40] Zhang T et al 2005 Spatial and temporal variability in active layer thickness over the Russian Arctic drainage basin *J. Geophys. Res.: Atmos.* **110** D16101
- [41] Padrón R S, Gudmundsson L, Greve P and Seneviratne S I 2017 Large-scale controls of the surface water balance over land: insights from a systematic review and meta-analysis *Water Resour. Res.* **53** 9659–78
- [42] Poggio L, de Sousa L M, Batjes N H, Heuvelink G B M, Kempen B, Ribeiro E and Rossiter D 2021 SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty *SOIL* **7** 217–40
- [43] The pandas development team 2023 pandas-dev/pandas: pandas (*Zenodo*) <https://doi.org/10.5281/zenodo.3509134>
- [44] Chen R H, Michaelides R J, Zhao Y, Huang L, Wig E, Sullivan T D, Parsekian A D, Zebker H A, Moghaddam M and Schaefer K M 2022 Permafrost Dynamics Observatory (PDO)—part II: joint retrieval of permafrost active layer thickness and soil moisture from L-band InSAR and P-band PolSAR *Earth Space Sci.* **10** e2022EA002453