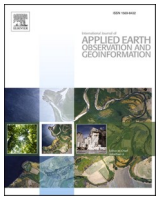
International Journal of Applied Earth Observation and Geoinformation 144 (2025) 104892

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag



Mapping wall-to-wall fractional cover of Arctic tundra plant functional types in Alaska using 20-m spatial resolution satellite imagery and harmonized plot observations

Tianqi Zhang [a,*], Morgan R. Steckler [a], Amy L. Breen [b], Forrest M. Hoffman [a], William W. Hargrove [c], Verity G. Salmon [a], Colleen M. Iversen [a], Stan D. Wullschleger [a], Jitendra Kumar [a]

[a] Oak Ridge National Laboratory, Oak Ridge, TN, United States
[b] University of Alaska Fairbanks, Fairbanks, AK, United States
[c] USDA Forest Service, Asheville, NC, United States

ARTICLE INFO

Keywords:
Plant functional types
Fractional cover
Land surface model
Arctic tundra
Machine learning
SHapley Additive exPlanations (SHAP)

ABSTRACT

Estimates of fractional cover (fCover) across given land surfaces are used to assess, and often model, vegetation composition and diversity, which are crucial for understanding the health and functioning of terrestrial ecosystems. Remote sensing provides a useful means for scaling local, plot-measured fCover estimates to regional scales. Leveraging a recently synthesized and harmonized plot database, this study generated wall-to-wall maps of fCover for six Alaskan-Arctic plant functional types (PFT), including non-vascular plants, forbs, graminoids, and deciduous and evergreen shrubs, using 20-m satellite data (Sentinel-1, Sentinel-2, ArcticDEM) using a machine learning regression approach, specifically the random forest (RF) algorithm, which is well-suited for handling nonlinear relationships and high-dimensional satellite datasets. This study additionally addressed the spatio-temporal inconsistencies e.g., sampling scale, plot size, and collection year in plot measured fCover by adopting a multivariate outlier detection approach—Cook's distance—to identify high-quality plots for model training and validation. Our approach achieves high accuracy (R^2 = 0.59–0.93, root mean squared errors = 0.02–0.10 for all PFTs) between plot-observed and satellite-derived fCover when using high-quality plot samples. The mapped fCover characterizes the spatial patterns of different PFTs across the tundra biome at a 20-m resolution, providing key information needed for improved representation of Arctic tundra vegetation in terrestrial biosphere models to better understand climate-vegetation feedback across the Arctic tundra.

1. Introduction

The Arctic region is warming approximately four times faster than lower latitude regions (Intergovernmental Panel on Climate Change (IPCC) report, Shukla et al., 2019; Rantanen et al., 2022) due to polar amplification (Graversen and Wang, 2009). Warming has led to an increase in vegetation productivity—particularly of tall shrubs that have extended their range, grown taller, and increased in biomass (Myers-Smith et al., 2020, 2011). In turn, shrubification may amplify regional warming via the entrainment of insulating snow by tall shrubs and also decreased albedo from leaves or branches that extend above the melting snowpack (Sturm et al., 2005, 2001). Changing vegetation community composition may also impact biogeochemical cycling. The increased presence of shrubs that associate with nitrogen-fixing bacteria may alter hillslope nitrogen cycling (Salmon et al., 2019), while the colonization of aquatic graminoids in inundated thaw ponds or thermokarst slumps may lead to increased methane release to the atmosphere (Joabsson and Christensen, 2001; Bao et al., 2021). Thus, understanding the fractional cover (fCover) of plant species, as aggregated into groupings of species with similar function (i.e., plant functional types, PFTs), across arctic tundra landscapes is valuable for both modeling and analysis purposes. Specifically, introducing detailed fCover of tundra PFTs into the Energy Exascale Earth System Model (E3SM; Golaz et al., 2019) Land Model (ELM) has been effective in quantifying biodiversity and biomass that were more comparable to field measurements (Sulman et al., 2021). Aside from modeling, fCover also conveys sub-pixel/plot details, which

* Corresponding author at: 1 Bethel Valley Road, Oak Ridge, TN 37830, United States.
E-mail address: zhangt1@ornl.gov (T. Zhang).

https://doi.org/10.1016/j.jag.2025.104892
Received 21 March 2025; Received in revised form 10 July 2025; Accepted 2 October 2025
Available online 7 October 2025

can be used to detect short-term vegetation structural responses to climate anomalies, such as wildfires (Fernández-Guisuraga et al., 2023), faster than categorical PFT maps in ecological applications. Additionally, it improves monitoring of vegetation degradation and desertification compared with normal satellite-derived vegetation indices (Chu, 2020).

Spatially continuous remote sensing data offer a means for mapping wall-to-wall PFT-level fCover by leveraging plot-scale fCover observations collected from boots-on-the-ground field observations or remotely-sensed from Unoccupied Aerial Vehicles (UAVs) (Macander et al. 2017, 2022; Yang et al., 2023). For instance, Yang et al. (2023) mapped fCover for twelve PFTs in western Alaska by combining UAV-estimated fCover with NASA's Airborne Visible / Infrared Imaging Spectrometer-Next Generation (AVIRIS-NG) hyperspectral imagery (Chapman et al., 2019) with mean absolute errors (MAEs) < 0.13. Despite high accuracy, UAV observations are often limited to small spatial scales (tens to hundreds of square kilometers) and study regions. One step towards enabling large-scale mapping could be building regressions between AVIRIS-NG flight paths (usually a few kilometers wide by hundreds of kilometers long; Chapman et al., 2019) where larger-scale fCover is mapped by training on plot fCover, and explanatory variables derived from downscaled climate models and topographic properties (Konduri et al., 2022). However, this requires sufficient high-quality plot samples overlapped with AVIRIS-NG flights for the initial AVIRIS-NG-wide mapping, which is not always feasible given the sparse spatial coverage of AVIRIS-NG over the Arctic Alaska region.

Synthesizing and harmonizing plot observations across areas with similar site characteristics can resolve the data scarcity issue in plot observations. With the widely available Landsat archive and plot samples spanning the North Slope of Alaska, Macander et al. (2017) successfully mapped wall-to-wall fCover for nine PFTs including shrubs subcategorized into low, dwarf, and tall classes at 30-m spatial resolution. Macander et al. (2022) then extended fCover mapping to the entirety of Alaska and part of the Yukon, Canada, based on an expanded plot network and 30-year Landsat time series, which aimed to understand long-term PFT fCover changes. Regional scale fCover maps by Macander et al. (2017, 2022) are expected to greatly benefit the parameterization of land surface models for more reliable and realistic simulations of vegetation distribution and dynamics. However, Macander et al. (2022) emphasized the development of a time series of fCover to detect change and yield insight into landscape-level responses to a changing climate. The time series approach necessitated fitting training models to yearly data rather than the full dataset, leading to a lower accuracy. However, the goal of some model simulations may be better suited to a one-time, high accuracy fCover map as opposed to a time series derived product. Moreover, across the Arctic Alaska tundra landscape, there is large variability in plot-level sampling strategies used for satellite-based fCover regression modeling, e.g., different sampling crews, plot sizes, or year of collection. These inherent inconsistencies among plot observations potentially affect the spatio-temporal representativeness of the training data for the development of remote-sensing vegetation data products, and thus the accuracy. The scan-line-corrector-off issue in the Landsat product (Scaramuzza and Barsi, 2005) may also be inherited in any Landsat-based fCover mapping, creating a spatial discontinuity (i.e., having a striped pattern).

The primary objective of our study was to build upon previous observational and remote-sensing studies to develop PFT-level fCover mapping of the Arctic Alaska region at a higher spatial resolution than previously possible. First, we harmonized plot-level data sources spanning the North Slope and the Seward Peninsula of Alaska into a consistent PFT scheme for wall-to-wall PFT fCover mapping of the entire Arctic Alaska tundra realm. Second, to ensure that the mapped fCover represented contemporary vegetation composition, we resolved inconsistencies between the plots sampled for fCover and satellite observations using a quality control approach (i.e., Cook's distance; Cook, 1977; Pinho et al., 2015). The quality control step allows us to separate two sets of plots: high-quality and low-quality samples that correlated with satellite observations well and poorly, respectively. Third, we produced spatially continuous, high-resolution (20-m) wall-to-wall maps of fCover for several target tundra PFTs across the Alaskan Arctic. These PFTs are compatible with the structure of the ELM model (Sulman et al., 2021). The mapping was achieved using a RF regression model trained on high-quality plot observations and corresponding satellite-derived predictors from Sentinel imagery and ArcticDEM data (Noh and Howat, 2017). Fourth, we identified the influential factors that contribute to inconsistencies between plot and satellite observations. Although these low-quality plots were not used for model training, they were retained to help inform improvements in future vegetation surveys, such as optimizing plot placement and enhancing measurement consistency. Finally, we analyzed the importance of satellite-derived features using SHapley Additive exPlanations (SHAP, Lundberg and Lee, 2017) to identify the variables that primarily account for the spatial variability of each PFT's fCover.

2. Materials and method

2.1. Study area

This study focused on mapping the fCover for eight common PFTs in an Alaska tundra region (i.e., the Alaskan Arctic) that encompasses both the North Slope and Seward Peninsula above the Arctic treeline as defined by Circumpolar Arctic Vegetation Map (CAVM) (Walker et al., 2005). The entire study region spans three CAVM bioclimate subzones including subzones C, D, and E (Fig. 1) with mean July temperatures reaching about 5–7C (coldest), 7–9 °C (moderate), and 9–12 °C (warmest) respectively (Walker et al., 2005; Raynolds et al., 2019) and topography ranging from coastal plains to inland mountains. The vegetation composition in Arctic Bioclimate Subzones C, D, and E varies significantly with climate and soil conditions (Raynolds et al., 2019). Specifically, Subzone C is characterized by graminoid and prostrate dwarf-shrub tundra, with sedges being dominant along with prostrate shrubs less than 5 cm tall. This zone also includes prostrate and hemi-prostrate dwarf-shrub, lichen tundra, which thrives in maritime, acidic regions. Subzone D consists primarily of non-tussock sedge, dwarf-shrub, and moss tundra. It is dominated by sedges and dwarf shrubs under 40-cm tall, with a well-developed moss layer. Frost boils and periglacial features are common. Erect dwarf-shrubs and moss tundra are also prevalent in continental areas with acidic soils. Wetland complexes dominated by sedges, mosses, and dwarf shrubs under 40-cm tall are also found here. Subzone E supports taller vegetation, including tussock-sedge, dwarf-shrub, and moss tundra, dominated by tussock cottongrass (*Eriophorum vaginatum L.*) and dwarf shrubs under 40-cm tall. Low-shrub, moss tundra dominates warmer, maritime areas with deep, moist active layers. Wetland complexes in this zone include sedges, mosses, and shrubs over 40-cm tall. This gradient from prostrate shrubs in Subzone C to low shrubs in Subzone E reflects increasing temperatures and moisture availability.

2.2. Plot data synthesis and harmonization

We used observations from the Pan-Arctic Vegetation Cover (PAVC) database (Steckler et al., 2024, 2025) that synthesized and harmonized plot observations of PFT fCover from five publicly available data sources (Fig. 1) covering the entire tundra landscape of the North Slope and the Seward Peninsula of Alaska, including the Alaska Biological Research, Incorporated – Environmental Research Services (hereafter referred to as ABR, total observations: 107, Macander et al., 2017), the Alaska Vegetation Plots Database (AKVEG, total observations: 185, Nawrocki et al., 2020), Alaska Arctic Vegetation Archive (AKAVA, total observations: 275, Davidson et al., 2016; Sloan et al., 2014 (under the Next Generation Ecosystem Experiments in the Arctic (NGEE Arctic) project); Villarreal, 2013; Villarreal et al., 2012; Walker et al., 2015, 2016), the

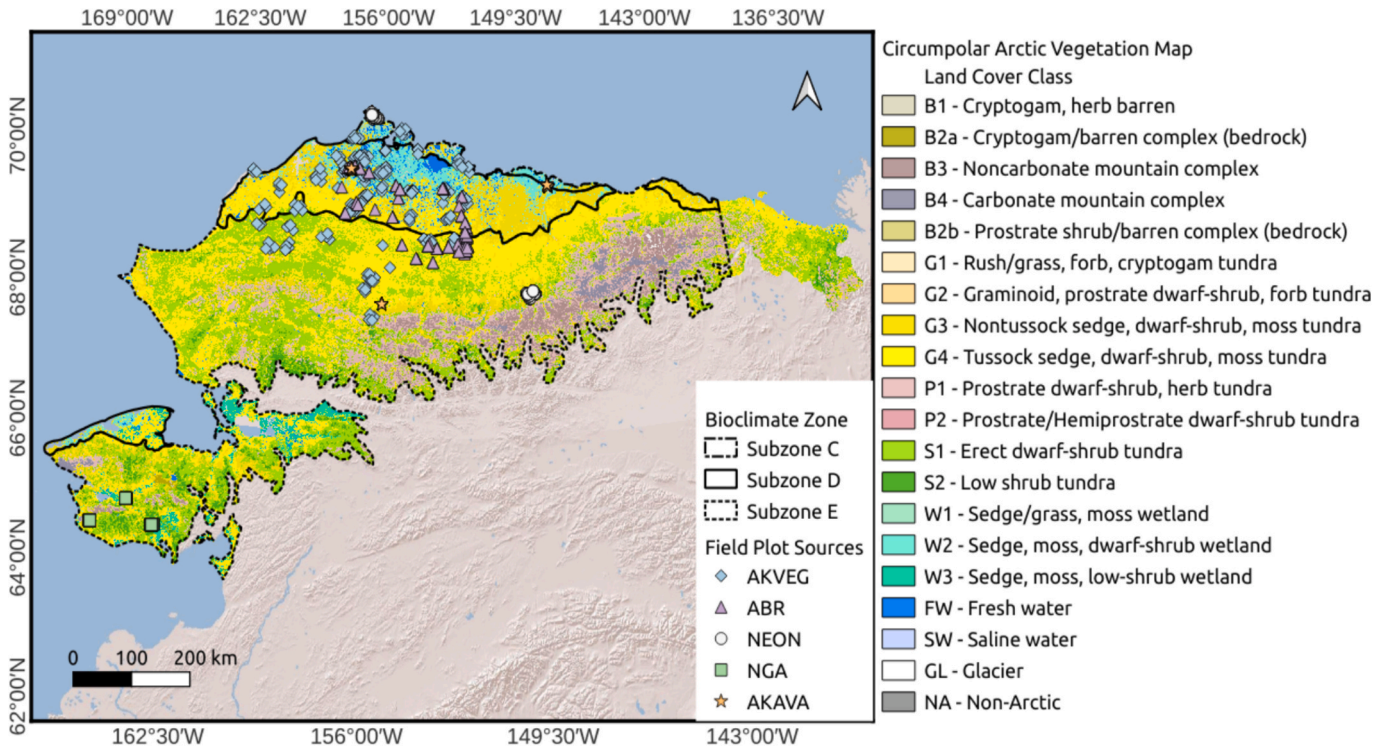


Fig. 1. Overview of field observations collected from plots; land cover and bioclimate zones developed by the Circumpolar Arctic Vegetation Map (Walker et al., 2005; Raynolds et al., 2019) for the Alaskan Arctic region.

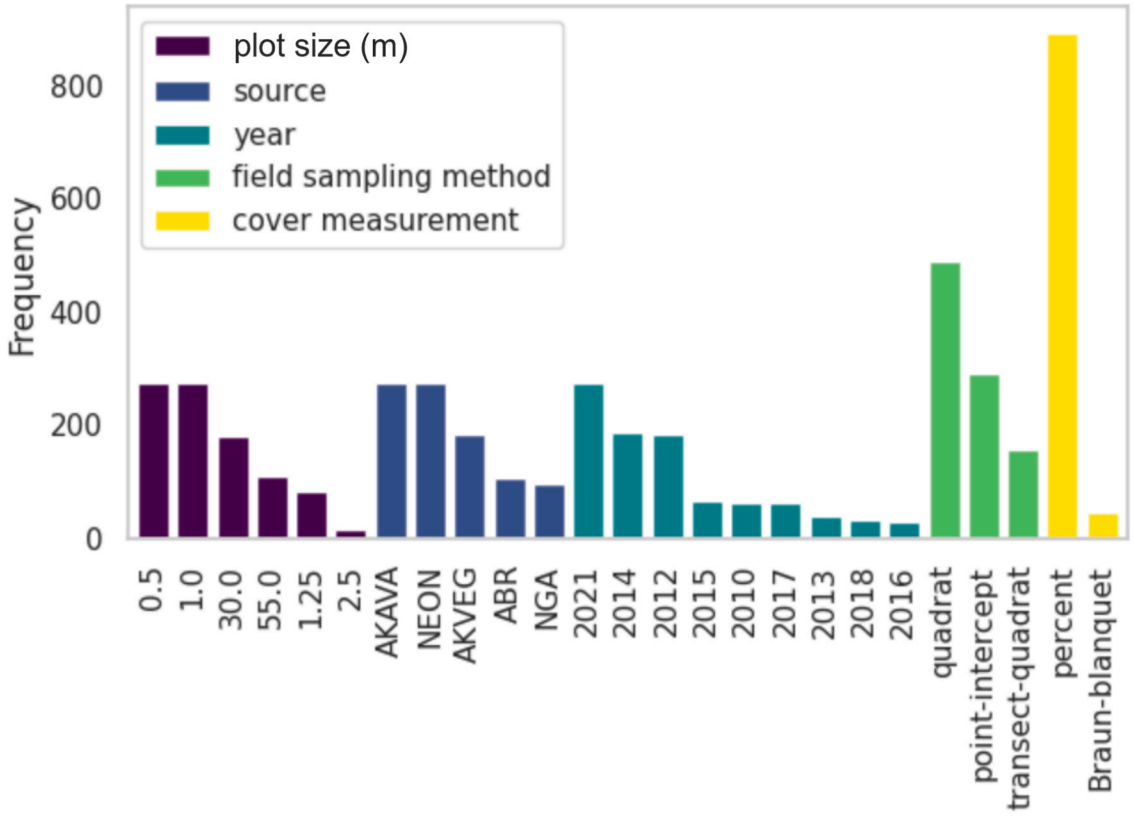


Fig. 2. Frequency of plot characteristics among all child plot sources (AKVEG, ABR, AKAVA, NEON, NGA), including plot size in meters (plot size (m)), year of collection (year), methods used for field sampling (field sampling method) and fCover measurement (cover measurement).

National Ecological Observatory Network (NEON, total observations: 275; NEON, 2023), and plots collected by the NGEE Arctic project across the Seward Peninsula (NGA, total observations: 98, Breen et al., 2020). Plots older than 2010 (mainly from AKAVA) were excluded from our modeling to ensure only contemporary observations were included. In addition, no historical fires (since 1983) are reported at the plots used in our study (Alaska wildland fire information map series, accessed April 2024).

Spatio-temporal inconsistencies are, however, identified among plot sources (Fig. 2 and Fig. S1), which could affect the representativeness of their fCover observations in conjunction with recent satellite observations. First, ABR and NEON are more consistent than the other sources regarding years of plot data collection, with all plots sampled in 2012 and 2021 respectively, whereas collection years for AKVEG, AKAVA, and NGA range from 2010 to 2018. Second, to match the spatial resolution of satellite pixels (e.g., Landsat) and reduce potential geolocation mismatches, plot-level observations in ABR and AKVEG were averaged within circular buffers of 30-m or 55-m radius (hereby referred to as parent plots) (Macander et al., 2017). In contrast, plots from the other sources are primarily original, non-aggregated transect-based observations—referred to here as child plots—with sizes (side or radius) ranging from 0.5 to 2.5 m. To address the inconsistency in plot radii, we spatially aggregated child plots that were close to one-other (within 55-m to match ABR's plot radius) and averaged the fCover to generate parent plots for AKAVA, NGA, and NEON. Third, in ~ 20 % of the AKAVA plots, the fCover was recorded using the 7-step Braun-Blanquet codes (a semi-quantitative scale used to estimate plant species' abundance and cover in vegetation surveys, ranging from rare or occasional to dominant presence), while percent cover estimates are used at the remaining plots. Fourth, AKVEG and ABR use the center-staked point-intercept method for sampling species to capture the spatial heterogeneity, while the rest plots all used quadrats or transect-quadrats. In addition, ABR and AKVEG provide both total and top fCover while the other plots primarily recorded the total cover information. Total fCover estimates the percent area cover of a species at the plot regardless of the canopies above or below, whereas top fCover only counts the areas that are not shielded by other species. Note that top fCover may underestimate the species' signal received by satellite under low-density canopy conditions where the signal may be able to penetrate the multilevel canopy. Given these differences, this study specifically focused on wall-to-wall mapping of total fCover.

Aside from various sampling and fCover measurement approaches, inconsistency in species naming also challenges plot harmonization for our regional-scale modeling. To address this issue, we used the comprehensive Alaska species checklist provided by AKVEG and leaf retention data documented in Macander et al. (2022) to associate PFT information with each species and then aggregate the species-level fCover to the PFT level following a harmonized scheme: litter, lichen, bryophyte, forb, graminoid, deciduous and evergreen shrub (Table 1). More details regarding plot synthesis and harmonization, differences among field sampling methods, and fCover measurement approaches were reported in Steckler et al. (2025). The density distribution of the harmonized plots at each PFT used in our study (Fig. 3) indicates that the fCover distribution of all PFTs, with the exception of bryophytes, is right-skewed with dominance towards low values. Moreover, forb-dominated plots seldom exist in the included plot sources, with relatively low median fCover (<0.1).

Table 1
Tundra PFTs used for fCover mapping at the study region.

PFT or others	Description
Deciduous shrub	Leaves are deciduous, lost after the summer season; deciduous shrubs were mostly low or tall in the Arctic region we examined, with canopy heights greater than or equal to 20 cm
Evergreen shrub	Leaves are evergreen, retained throughout the year; evergreen shrubs were mostly dwarf in the Arctic region we examined, with canopy heights less than 20 cm
Graminoid	Herbaceous plants including grasses and sedges; perennial plants
Forb	Non-graminoid herbaceous flowering plants; often annual plants
Lichen	A symbiotic partnership of a fungus and an alga, grouped into non-vascular; perennial plants
Bryophyte	Non-vascular perennial plants, including liverworts and mosses
Non-vascular	Combination of lichens and bryophytes
Litter	Dead plant matter or fallen leaves

2.3. Satellite-derived explanatory variables

For explanatory variables as input to fCover regression modeling, we collected the ArcticDEM (Noh and Howat, 2017) mosaic to characterize topography; pre-processed and downloaded Sentinel-1 Synthetic Aperture Radar (SAR) polarizations in the SeNtinel Application Platform (SNAP); and pre-processed and downloaded Sentinel-2 multispectral data via the Google Earth Engine platform (Gorelick et al., 2017). The ArcticDEM mosaic product, a high-resolution (2-m) and high-quality DEM product covering the entire pan-Arctic area, was compiled from the best-quality 2-m ArcticDEM strip files generated from very-high-resolution optical stereo imagery to reduce the void areas and edge-matching artifacts (Noh and Howat, 2017). Despite demonstrating effectiveness in upscaling land cover and sparse lidar canopy heights to regional and global scales (Zhang and Liu, 2023; Korhonen et al., 2017; Li et al., 2020), both Sentinel-1 (S1) and Sentinel-2 (S2) data have remained under-explored in mapping fCover at the PFT level.

S1 is a dual-polarization C-band SAR instrument that provides 2 single co-polarization bands (*VV* or *HH*) or 2 dual cross-polarization bands (*HV* or *VH)*. In our study, the available polarization bands *VV* and *VH* were collected from the S1 Ground Range Detected scenes, which have been calibrated and orthorectified using the S1 Toolbox (Veci et al., 2014). For S2, we collected the Level-2A surface reflectance product, which has been atmospherically corrected by the sen2cor model (Louis et al., 2016). Only the 10 spectral bands that have spatial resolutions less than or equal to 20-m were introduced in this study, including blue, green, red, red edge 1–4 (*redEdge1–4*), near infrared (*nir*), and shortwave infrared 1–2 (*swir1–2*). We developed median composites for both S1 and S2 images collected over the growing seasons (June to September) for the year 2019. Before image compositing, all cloudy pixels were removed using the quality assessment (QA60) band provided with the Sentinel-2 data, which flags pixels affected by clouds, cirrus, and other atmospheric effects.

To enhance the separability of PFTs using Sentinel and ArcticDEM, we additionally calculated *slope*, *hillshade*, and *aspect* in addition to *elevation* from ArcticDEM to characterize the topographic variability and 19 vegetation indices from S2 multispectral bands accounting for soil background, leaf water, and chlorophyll content that vary among PFTs. The examined vegetation indices (Table 2 and Table S1) included: (1) two normalized difference water indices (*ndwi*) using two *swir1–2*, Gao, 1996); (2) modified soil-adjusted vegetation index (*msavi*, Qi et al., 1994); (3) visible atmospherically resistant index (*vari*, Gitelson et al., 2002); (4) ratio vegetation index (*rvi*, Jordan, 1969); (5) optimized soil-adjusted vegetation index (*osavi*, Rondeaux et al., 1996); (6) triangular greenness index (*tgi*, Hunt Jr et al., 2011); (7) green leaf index (*gli*, Louhaichi et al., 2001); (8) normalized green red difference index (*ngrdi*, Tucker, 1979); (9) chlorophyll index-green (*ci_g*, Gitelson et al., 2003); (10) green normalized difference vegetation index (*gNDVI*, Gitelson et al., 1996); (11) chlorophyll vegetation index (*cvi*, Vincini et al., 2008); (12) second modified triangular vegetation index (*mtvi2*, Haboudane et al., 2004); (13) transformed chlorophyll absorption reflectance index (*tcari*, Haboudane et al., 2002); (14) triangular chlorophyll index (*tci*, Haboudane et al., 2008); (15) normalized anthocyanin reflectance index (*nari*, Bayle et al., 2019); we also included three tasseled cap indices—(16) *brightness*, (17) *greenness*, and (18) *wetness*—which are orthogonal transformations of spectral bands originally developed by Kauth and Thomas (1976) for vegetation analysis.

To ensure the spatial consistency between satellite and plot

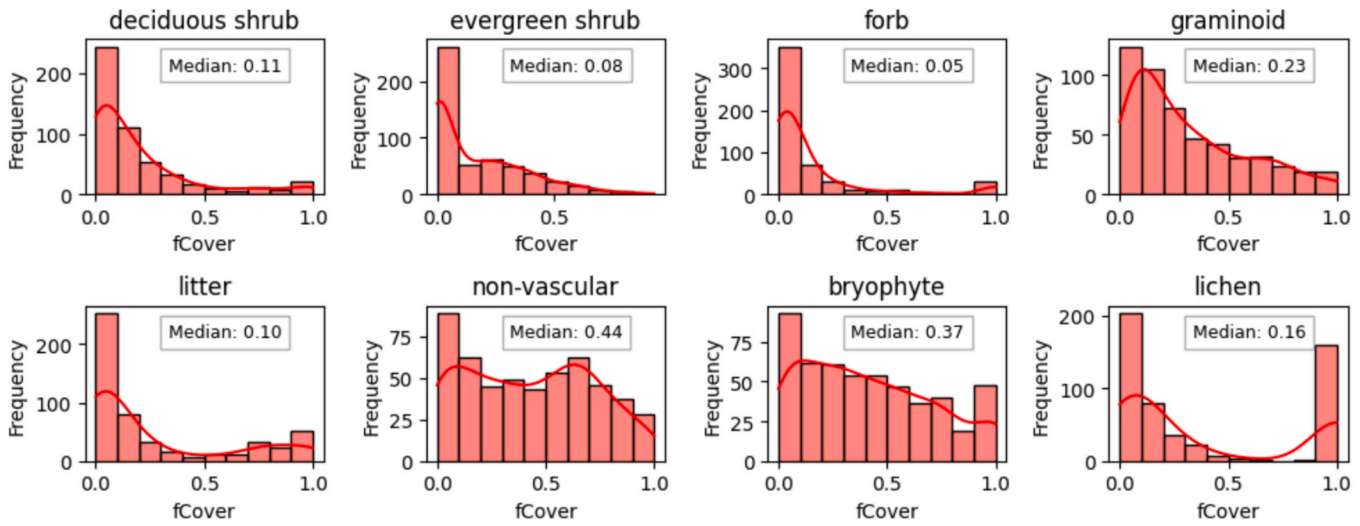


Fig. 3. Fractional cover distribution of all PFTs at the plots after spatial aggregation. Only a few plots are observed with overlapped canopies among different species, leading to > 1 fCover for some PFTs. In most cases, the PFT-level fCover does not exceed 1. Hence, all plots with > 1 fCover were rounded down to 1.

Table 2
Formulas for calculating the input vegetation indices for fCover regression modeling. For implications of each index, please refer to Table S1.

Index	Abbrev.	Formula	Reference
Normalized difference water index using swir1-2	*ndwi*	$(nir - swir)/(nir + swir)$	Gao (1996)
Modified soil adjusted vegetation index	*msavi*	$0.5*\left(2*nir + 1 - \sqrt{(2*nir+1)^2 - 8*(nir - red)}\right)$	Qi et al. (1994)
Visible atmospherically resistant index	*vari*	$(green - red)/(green + red - blue)$	Gitelson et al. (2002)
Ratio vegetation index	*rvi*	nir/red	Jordan (1969)
Optimized soil adjusted vegetation index	*osavi*	$1.16*(nir - red)/(nir + red + 0.16)$	Rondeaux et al. (1996)
Triangular greenness index	*tgi*	$0.5*(120*(red - blue) - 190*(red - green))$	Hunt Jr et al. (2011)
Green leaf index	*gli*	$(2*green - red - blue)/(2*green + red + blue)$	Louhaichi et al. (2001)
Normalized green red difference index	*ngrdi*	$(green - red)/(green + red)$	Tucker (1979)
Chlorophyll index-green	*ci_g*	$nir/green - 1$	Gitelson et al. (2003)
Green normalized difference vegetation index	*gNDVI*	$(nir - green)/(nir + green)$	Gitelson et al. (1996)
Chlorophyll vegetation index	*cvi*	$(nir*red)/green^2$	Vincini et al. (2008)
Second modified triangular vegetation index	*mtvi2*	$1.5*(1.2*(nir - green) - 2.5*(red - green))/\sqrt{(2*nir+1)^2 - \left(6*nir - 5\sqrt{red}\right) - 0.5)}$	Haboudane et al. (2004)
Transformed Chlorophyll Absorption Reflectance Index	*tcari*	$3*((redEdge1 - red) - 0.2*(redEdge1 - green)*(redEdge1/red))$	Haboudane et al. (2002)
Triangular chlorophyll index	*tci*	$1.2*(redEdge1 - green) - 1.5*(red - green)*\left(\sqrt{redEdge1/red}\right)$	Haboudane et al. (2008)
Normalized Anthocyanin Reflectance Index	*nari*	$(1/green - 1/redEdge1)/(1/green + 1/redEdge1)$	Bayle et al. (2019)
Tasseled cap brightness	*brightness*	$0.3037*blue + 0.2793*green + 0.4743*red + 0.5585*nir + 0.5082*swir1 + 0.1863*swir2$	Kauth and Thomas (1976)
Tasseled cap greenness	*greenness*	$0.7243*nir + 0.0840*swir1 - 0.2848*blue - 0.2435*green - 0.5436*red - 0.1800*swir2$	
Tasseled cap wetness	*wetness*	$0.1509*blue + 0.1973*green + 0.3279*red + 0.3406*nir - 0.7112*swir1 - 0.4572*swir2$	

observations, we re-projected all data to the geographic coordinate system (EPSG: 4326). The plot sizes of the parent plots are 30-m or 55-m after the spatial aggregation, so we evaluated two search radii (i.e., 30-m or 55-m) when sampling the explanatory variables for fCover regression model training together with the plot observed fCover.

2.4. Quality control

The uncertainties (i.e., inconsistencies due to varying years of plot collection, plot size, and sampling approaches) in our plot data sources may affect the overall regression modeling, introduce uncertainties in the representativeness of the mapped fCover, and bias the model evaluation results if not properly addressed. However, without a systematic quality analysis, it is difficult to identify the plots that led to low correlations between plot and satellite observations when performing the fCover regression analysis.

To address this, we introduced an outlier detection approach known as the Cook's distance (Cook, 1977) to identify and exclude outliers and low-quality plots from fCover regression modeling. Here the low-quality plots were defined as plots where fCover observations correlated poorly

with satellite-observed vegetation conditions, which could be due to several reasons, including natural vegetation growth over time, geolocation errors and uncertainty, inconsistent sampling approaches, or a plot's lack of representativeness. The Cook's distance has an advantage over other outlier detection approaches (Kannan and Manoj, 2015) in that it quantifies a sample's quality based on its impacts on the overall regression modeling, which has been proven effective in enhancing ICESat-2's quality in predicting canopy heights comparable to airborne lidar derived counterparts (Zhang and Liu, 2023).

For each PFT, the Cook's distance (D_i) of the i-th sample is quantified as the change it brought to the least-squares regression model fitted between the plot-observed fCover and the satellite derived explanatory variables (Table 2) when excluding it from the analysis (Equation (1)–(2)):

$$D_i = \frac{\sum_{j=1}^{n}\left(\hat{y}_j - \hat{y}_{j(i)}\right)^2}{ps^2} \quad (1)$$

where $\hat{y}_j$, $\hat{y}_{j(i)}$ represent the predictions of j-th sample by the least-squares regression model fitted on observations including and excluding i-th sample respectively, n stands for the total number of samples, p denotes the number of unknown coefficients, and s^2 is the mean squared error of the regression model calculated as

$$s^2 = \frac{\sum_{j=1}^{n}\left(y_j - \hat{y}_j\right)^2}{n-p} \quad (2)$$

where y_j is the observed value of the j-th sample.

Intuitively, samples with larger distances often indicate that they introduce more variability to the regression analysis. In this study, we replaced the least-squares regression model with a Gaussian generalized linear model linked by the identity function, where the D distribution for all samples appears more separable between the "outliers" and the "non-outliers", compared with the result estimated using the ordinary least squares regression. Based on the calculated distances, we then randomly selected half of the samples with D less than the first quartile ($Q1$) of the distance distribution of all samples as the validation dataset. The remaining plots were regarded as the original samples used for fCover regression modeling. To evaluate the effectiveness of the proposed quality control, we also presented the results from a refined model that only input high-quality plots identified after a second round of quality control on the remaining training samples (excluding the validation data), selecting those with $Ds <$ the critical distance (D^*):

$$D^* = Q3 + k \times IQR \quad (3)$$

where IQR is the interquartile range ($Q3$–$Q1$), $Q1$ marks the 25th percentile, and $Q3$ (the third quartile) marks the 75th percentile of the data distribution, k was tested from 1 to 3 in 0.5 increments to optimize the model training. This quality filtering was conducted on the training data (excluding the validation set) and was applied separately for each PFT. The method targets influential statistical outliers rather than relying on spatial or temporal alignment and improves the robustness and reproducibility of the modeling process.

2.5. Machine learning based fractional cover regression

In this study, we employed an RF model to perform the regression analysis between the plot-observed fCover and the satellite-derived explanatory variables, as it leverages an ensemble of decision trees to minimize the overall model variance. RF also demonstrates its effectiveness in PFT-level fCover mapping studies (Macander et al., 2017, 2022) and outperforms other machine learning regression models on forest attribute predictions (Zhang and Liu, 2023; Zhao et al., 2019).

Given the uncertainty of the plot observations, we could not perform regular cross-validation to shuffle the training and testing datasets, as that assumed all observations were of equal quality. Instead, we performed a grid search to identify the best RF model parameter setting (number of trees used for fitting and maximum features used for splitting the trees) for each PFT using the same validation dataset (randomly chosen half from samples with $D < Q1$). The tested values for the number of trees used for fitting and the maximum features used for splitting the trees with the model regression results were [50, 100, 150, 200] and [5, 10, 15], respectively. The model with the highest accuracies was then selected to map the wall-to-wall fCover based on spatially continuous satellite imagery for each PFT using all high-quality plot samples having distances shorter than D^*. Hereon we refer to the gridded fCover estimates as "PAVC-Gridded" product (Zhang et al., 2025).

2.6. Model evaluation

The model evaluation in this study focused on three aspects: (1) fCover regression (model training and regional-scale mapping), (2) plot quality analysis, and (3) feature importance analysis of fCover regression modeling. For fCover regression modeling, we performed a scatterplot analysis between the predicted ($\hat{y}$) and observed (y) values and reported three accuracy metrics, including the root mean square error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2) using a simple linear regression. Specifically, RMSE and MAE were used for quantifying the average errors in the model predictions, whereas R^2 describes the percentage of variance in y explained by $\hat{y}$.

For regional-scale fCover mapping, we visually compared the spatial patterns of PAVC-Gridded fCover with Sentinel-2 RGB imagery. To exclude non-vegetated areas, open water and ice were masked using $ndwi > 0$ and $ndvi < 0.3$. Given the widespread use of the CAVM in Arctic vegetation studies, we summarized zonal fCover statistics for all plant functional types (PFTs) across each CAVM zone. Additionally, we compared PAVC-Gridded fCover with the product from Macander et al. (2017) using scatter plot analysis, spatial pattern assessments, and zonal statistics over the CAVM classes. While both studies employed RF regression for fCover mapping, our approach incorporated a more diverse set of plot sources (NEON, AKVEG, NGA), including coverage of the Seward Peninsula. Furthermore, our satellite predictors were primarily derived from Sentinel imagery, which offers higher spatial resolution and is unaffected by the scan-line corrector issue present in Landsat imagery used by Macander et al. (2017). To enhance model reliability, we additionally implemented quality control measures to mitigate spatial and temporal inconsistencies in plot samples when training the regression model with recent satellite observations. Furthermore, to evaluate model's transferability, we conducted a leave-one-site-out validation. Specifically, for each plot source, we excluded all plots from that site during model training and then used them exclusively for validation, rotating through all available plot sources. This procedure assesses how well the models trained on high-quality plots (after Cook's distance filtering) can generalize to new regions that were not seen during training. The results of this transferability analysis complement our standard validation, helping to demonstrate that the Cook's distance quality control improves the spatial generalizability of the fCover regression models by reducing site-specific biases.

For the plot quality analysis, we sought to disentangle the factors that contributed to uncertainty in the regression modeling between plot-observed fCover and recent satellite observations. Thus, we plotted the distribution of plot characteristics (plot data source, plot size, plot collection year, method used for sampling, and fCover measurement) of low-quality plots identified by the Cook's distance method. This understanding could help inform the design of future field campaigns for plot-level fCover sampling for regional scale mapping.

Lastly, feature importance analysis can inform both model interpretability and feature selection in model design. To quantify the relative contribution of the included explanatory variables (Table 2) on fCover regression analysis, we performed SHapley Additive

exPlanations (SHAP, Lundberg and Lee, 2017) analysis based on the best-trained RF model. This approach computes the average marginal contribution of each variable across all possible feature coalitions, providing a consistent and theoretically grounded measure of how each variable influences the model output. Compared with permutation-based importance, SHAP offers a more robust and interpretable estimate of each variable's impact without relying on random permutations, thereby improving the transparency of model-driven decisions.

3. Results

3.1. PFT level fractional cover mapping

The comparisons between the plot-observed fCover and the predictions made by the best-trained models (Tables S2-S9) are shown in Fig. 4. Overall, with high-quality plot samples, the correlations (R^2) between the predicted and observed fCover for all PFTs are substantially improved (to between 0.59 and 0.93) with relatively low bias (MAEs = 0.02–0.07, RMSEs = 0.02–0.09) in comparison to the predictions made from original non-enhanced plots (Fig. S2). PFTs predicted with highest accuracy are shrubs ($R^2 > 0.90$, MAEs and RMSEs are <=0.04), followed by non-vascular plants ($R^2 > 0.7$, RMSEs = 0.02–0.09) and graminoids ($R^2 = 0.75$, RMSE = 0.08, MAE = 0.05). While it has a positively skewed distribution, the forb fCover prediction also reaches R^2 of 0.59 (RMSE = 0.02). Litter is generally harder to predict, with the lowest R^2 (0.36) (RMSE = 0.07, MAE = 0.06) in comparison to other PFTs, likely due to its heterogeneous and temporally dynamic nature (Facelli and Pickett, 1991). The model transferability analysis results using a leave-one-site-out cross-validation framework were summarized in Tables S10-S13. Overall, the results indicate that the modeling framework demonstrates promising transferability across most sites and PFTs, with mean R^2 values of 0.77 for deciduous shrubs, 0.70 for evergreen shrubs, and 0.50 for lichens, suggesting reasonable generalizability. However, some limitations emerge. Like model training results (Fig. 4), litter shows consistently poor transferability across all leave-out sites (mean R^2 = 0.08), graminoids perform poorly when the AKAVA site was excluded ($R^2 = 0.04$), and bryophytes struggle with transferability at the NGA and NEON sites ($R^2 < 0.15$). These site-specific weaknesses likely reflect overall small sample sizes in our plot inventory, fCover underrepresentation of certain PFTs (such as predominant bryophyte in NEON compared with other sites) (Fig. S3-S7), and differences in sampling strategies between sites (Fig. S1).

The wall-to-wall PAVC-Gridded fCover maps are presented in Figs. 5–8. In general, the spatial patterns of fCover for each PFT align well with the vegetation distribution observed in the Landsat imagery, demonstrating the effectiveness of our fCover mapping approach. In addition, we find consistent spatial distribution patterns between PAVC-Gridded fCover and the CAVM dominant vegetation classes. Specifically, our maps indicate that deciduous shrubs grow in warmer areas (towards the treeline), whereas evergreen shrubs are spread out in regions of bioclimate zone E characterized by mild topography. With the fine-scale fCover map, we also identify deciduous shrubs in river valleys (Fig. 5). In comparison to the discrete CAVM vegetation classes, PAVC-Gridded fCover products reveal detailed and contrasting patterns of the spatial distribution of PFTs across the study region. For instance, PAVC-Gridded fCover maps indicate that graminoids occupy most of the northern cold and wet areas (bioclimate zone C and D), while forbs primarily exist in warmer bioclimate zone E with overall low fractional dominance across the entire study region (Fig. 6). Similarly, our maps suggest that lichens are only present in mountainous areas, yet bryophytes are pervasive across most of the bioclimate zone E except the mountains (Fig. 8).

3.2. Comparison with existing fCover products

Fig. 9 shows the paired scatter-histogram analysis results between our PAVC-Gridded fCover and published products by Macander et al. (2017). Overall, the distribution of the predicted fCover for both products does not always follow the plot-observed distributions, e.g., graminoids and litter, which suggests both model predictions are not biased by the plot samples. Our modeled fCover distribution is generally consistent with Macander et al. (2017) for shrubs ($R^2 > 0.5$). Our model slightly overpredicts the fCover of deciduous shrubs, while underpredicts the evergreen shrubs and non-vascular plants in the regions with higher fractional dominance (fCover = 0.2–0.5). They however, differ substantially from each other for forb, graminoid, and litter ($R^2 <$ 0.2, RMSEs = 0.06–0.14).

We further plotted the zonal statistics of the mapped fCover across the CAVM vegetation classes in Fig. 10 and Fig. S8. Not surprisingly, due to major differences in underlying methodology and spatial resolution, large differences are found in these two products except for the forbs that are consistently predicted with low dominance across all CAVM classes. In addition, we notice higher separability among shrubs and graminoids in our products compared to Macander et al. (2017). In comparison to Macander et al. (2017), PAVC-Gridded fCover tends to

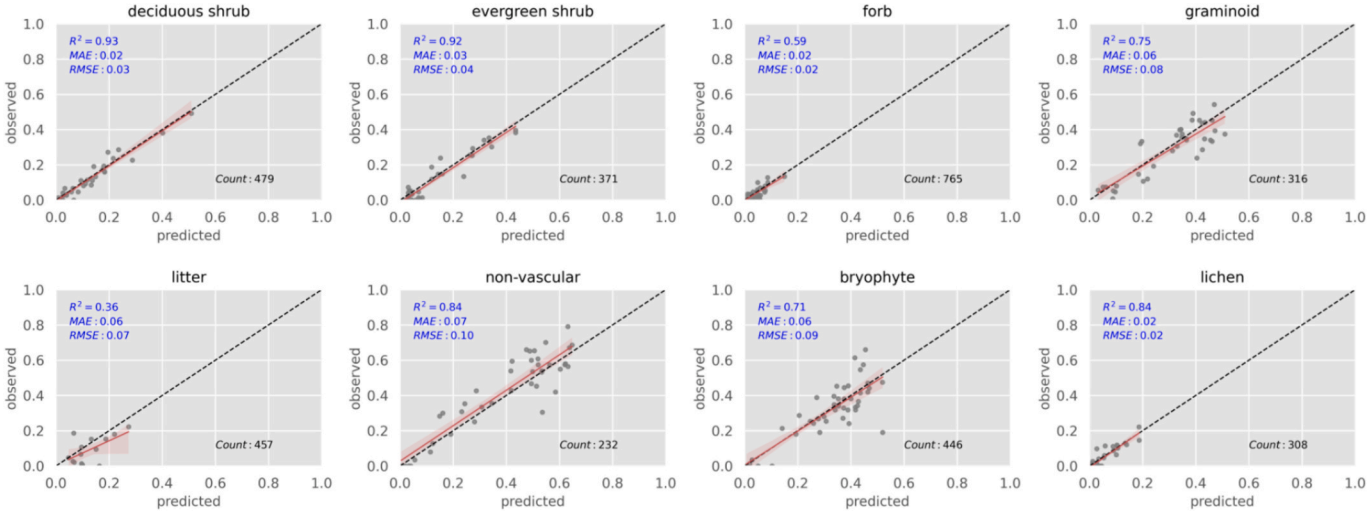


Fig. 4. Scatter plot analysis between the predicted and observed fCover using the best tuned RF model, where "count" indicates the number of the high-quality samples used for model training. The 1:1 and regression lines are indicated by black dashed and red solid lines respectively, with red shading covering the 95% confidence interval for the regression estimate.

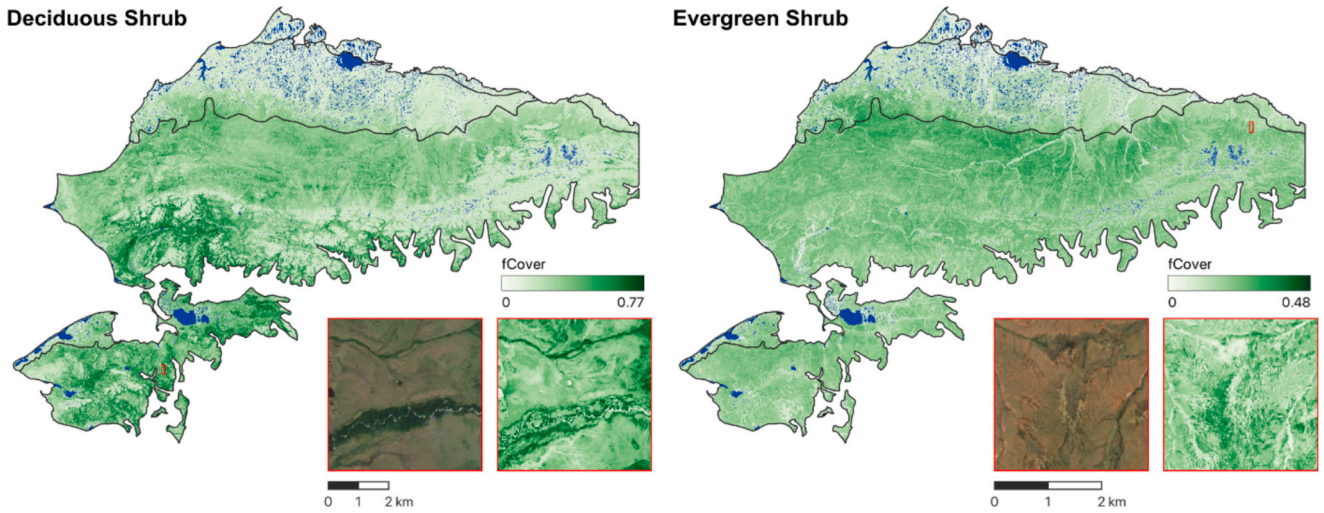


Fig. 5. Wall-to-wall fCover map for deciduous and evergreen shrub. Here, open water areas or ice are colored blue. The lower right corner shows a zoomed-in comparison between the Sentinel-2 RGB imagery and mapped fCover at the red outlined grid.

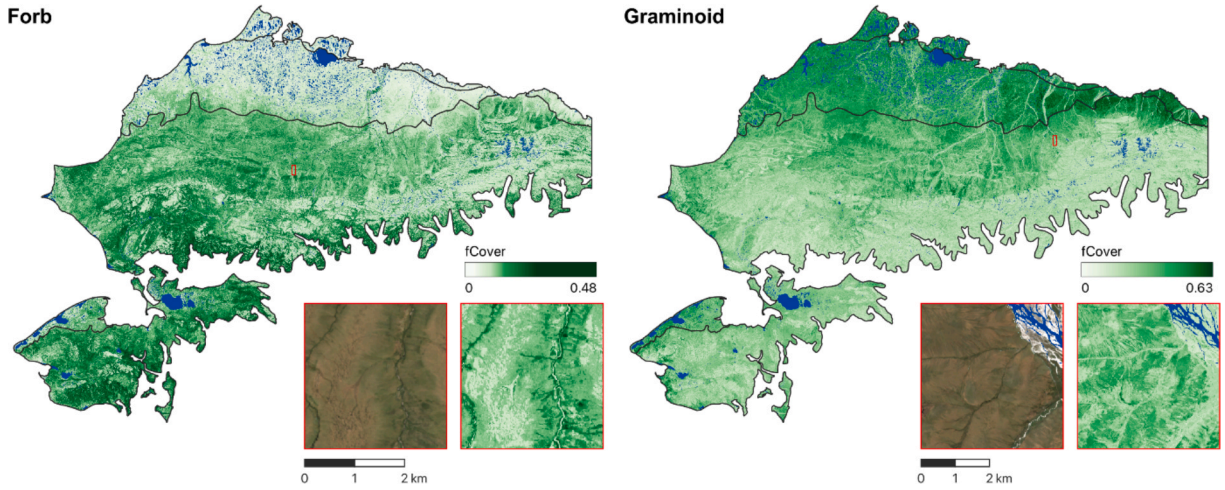


Fig. 6. Wall-to-wall fCover map for forbs and graminoids. Here, open water areas or ice are colored blue. The lower right corner shows a zoomed-in comparison between the Sentinel-2 RGB imagery and mapped fCover at the red outlined grid.

conform better with CAVM with higher fCover over the CAVM shrub classes (S1–2) and over the CAVM graminoid classes (G1–2) for shrubs and graminoids classes respectively. The zoomed-in plots (Fig. S9) at same locations also suggest more separable and spatially more continuous results from our mapped fCover than Macander et al. (2017) due to quality control approach employed and no known scan-line-corrector-off issue in Sentinel-2 as opposed to Landsat.

3.3. Quality analysis on plot outliers having low consistency with satellite data

The results associated with plot samples having low correlations to satellite observations are shown in Fig. 11. In general, we observe variability among different PFTs in terms of the contributing factors. Plot size is an important factor affecting the fCover modeling, where small plots – especially the 2.5-m ones– lead to the most variability in the estimated fCover for almost all PFTs except evergreen shrubs and bryophyte, where 1.25-m plots have a marginally larger impact. 1.25-m and 2.5-m plots among the observations are primarily from the NGA datasets located across Seward Peninsula. Located in shrub to tundra transition zone, Seward Peninsula landscape consists of highly diverse and heterogeneous spatial distribution of PFTs, and thus the plot observations from the region (primarily from NGA) are found to have high variability in terms of their spectral response in the satellite remote sensing products and hence are associated with higher Cook's distance when developing regression models. Plot observations from NGA are often a source of bias for fCover estimation of especially forbs, shrubs, and lichens. NEON plots are located in relatively homogeneous areas, with relatively consistent spectral signatures for same PFT types, and thus are robust for fCover modeling with lower fractions of low-quality plots except for bryophytes. Despite having consistent plot sizes, ABR plots show high variability in terms of their spectral response in satellite observations, likely due to a mixed pixel effect in the large plots (55-m in radius), especially for litters.

For all PFTs, the year of plot data collection is a large source of variability, due to growth and change in vegetation conditions by the time of satellite remote sensing images. The fCover collected between 2016 and 2018 suggests more uncertainties for almost all PFTs. Plots collected in 2012 additionally contribute to highest variability for litter fCover estimate. Field plot sampling methods do not contribute any significant variability to the modeling, and comparatively among the different sampling methods. The point-intercept method has more impacts on graminoids, forbs, and litters. Aside from deciduous shrubs and graminoids, the fCover quality of all other PFTs is more affected by the

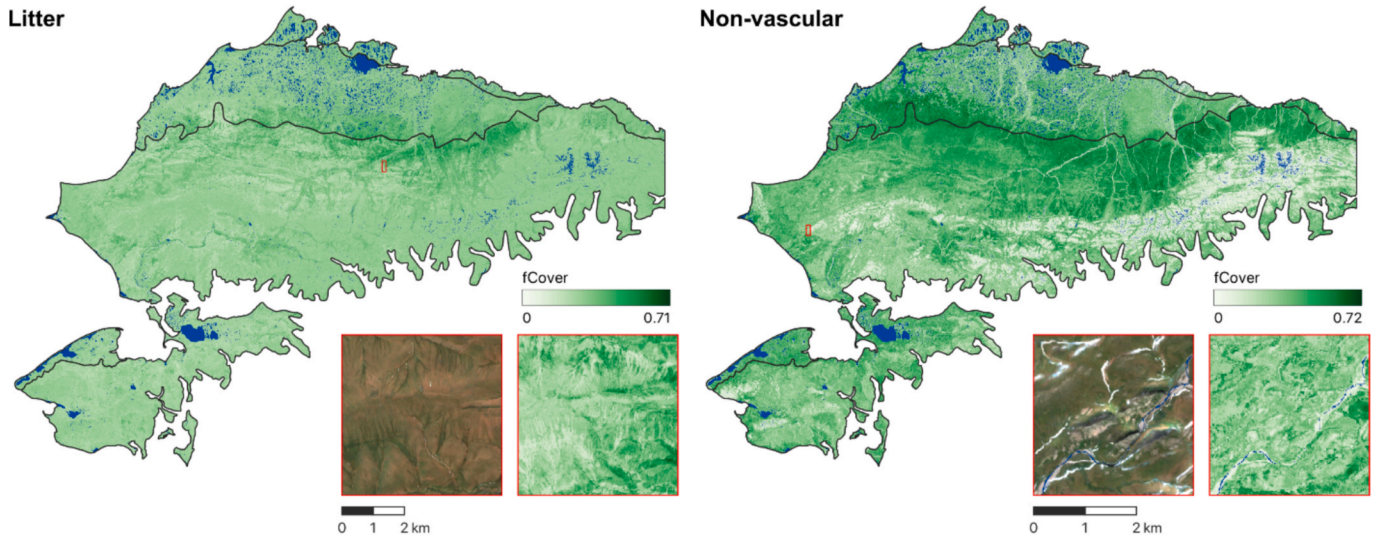


Fig. 7. Wall-to-wall fCover map for litter and non-vascular plants (lichen + bryophyte). Here, open water areas or ice are colored blue. The lower right corner shows a zoomed-in comparison between the Sentinel-2 RGB imagery and mapped fCover at the red outlined grid.

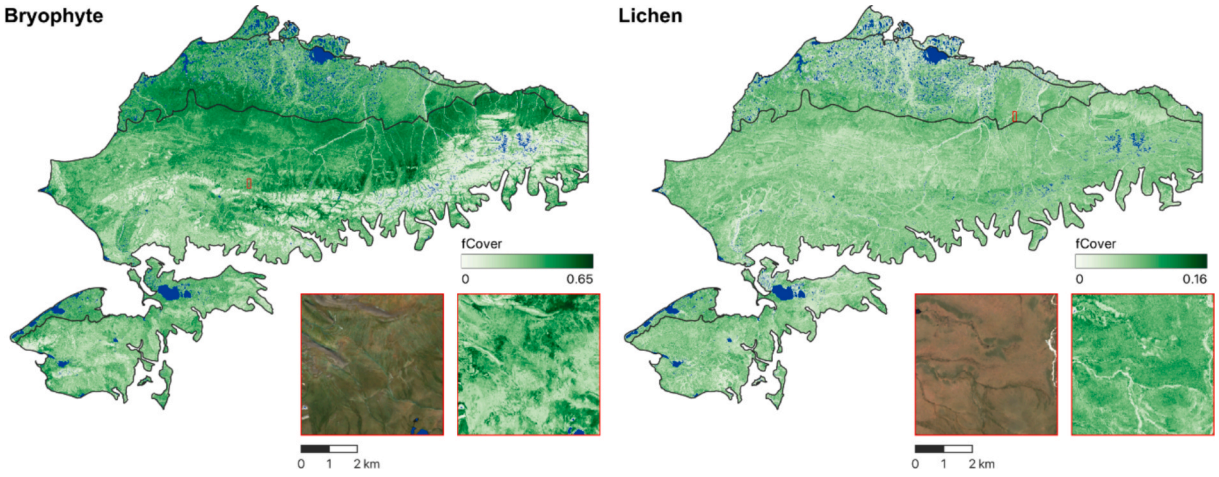


Fig. 8. Wall-to-wall fCover map for bryophyte and lichen. Here, open water areas or ice are colored blue. The lower right corner shows a zoomed-in comparison between the Sentinel-2 RGB imagery and mapped fCover at the red outlined grid.

percent-based measurement. In addition to factors affecting the quality of sampled fractional cover (fCover), herbaceous and shrub PFTs exhibit contrasting patterns. For example, compared with graminoids, forbs are more influenced by small plots (2.5-m) from NGA collected over the Seward Peninsula, well away from the North Slope where the other plot sources for forbs were from. Likewise, we find that deciduous shrubs are more sensitive to spatio-temporal inconsistencies compared with evergreen shrubs. For non-vascular plants, most low-quality plot samples also primarily arise from NGA, especially for lichens, where an additional concentration of low-quality data is associated with plots collected in 2010 (from AKAVA). In contrast, for bryophytes, low-quality plots are more evenly spread across sites and years, with moderate contributions from NEON.

3.4. Feature importance analysis on fCover regression modeling

Fig. 12 summarizes the mean absolute SHAP values for each explanatory variable, calculated from the best-performing model configuration, with dot color encoding the underlying feature value (dark purple = low, yellow = high). Similar to the quality analysis, contrasting outcomes are observed across all PFTs. For S2-derived indices, *msavi* is the top predictor for forb cover (its yellow dots clustered at the highest SHAP values suggest that higher *msavi* values are associated with higher forb predictions), yet it has minimal influence on other PFTs. Similarly, *ngrdi*'s relative importance varies between forb and graminoid models, and the red-edge bands (*rededge2* and *rededge3*) show yellow points at their highest SHAP values for forbs—indicating that higher red-edge reflectances are linked to higher forb cover estimates. In contrast, the greenness index reaches its highest SHAP values (with yellow–orange at the apex) for both forbs and litter. For deciduous shrubs, *greenness* still shows yellow dots at its SHAP maximum, implying that higher greenness is associated with higher predicted cover—whereas *wetness* leans toward dark purple at its maximum SHAP, suggesting that lower wetness (drier conditions) is associated with higher deciduous-shrub cover.

Among topographic features, *elevation* is the dominant predictor for most PFTs (except for forbs, graminoids, and lichens), but its highest SHAP values correspond to low elevations (dark purple dots), indicating that lower altitudes are linked to higher predicted fractional cover for those types. *Aspect* contributes more broadly than *slope* or *hillshade* (the latter only showing relevance for graminoids). Within the non-vascular group, bryophytes show their largest SHAP contributions from *elevation*,

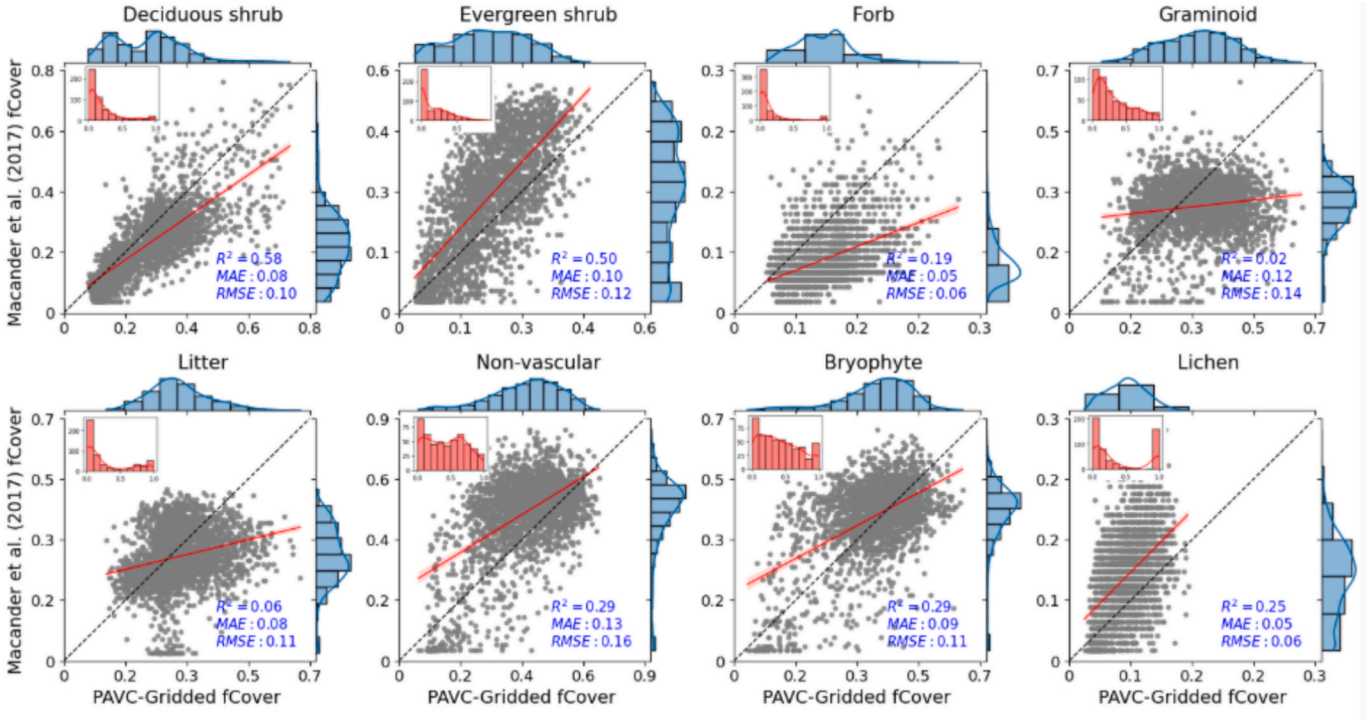


Fig. 9. Paired scatter-histogram analysis results between our PAVC-Gridded fCover and that mapped by Macander et al. (2017). The inset histograms show the plot-observed fCover distribution. Here 1:1 line and the regression line are marked in black-dashed and red solid lines respectively.

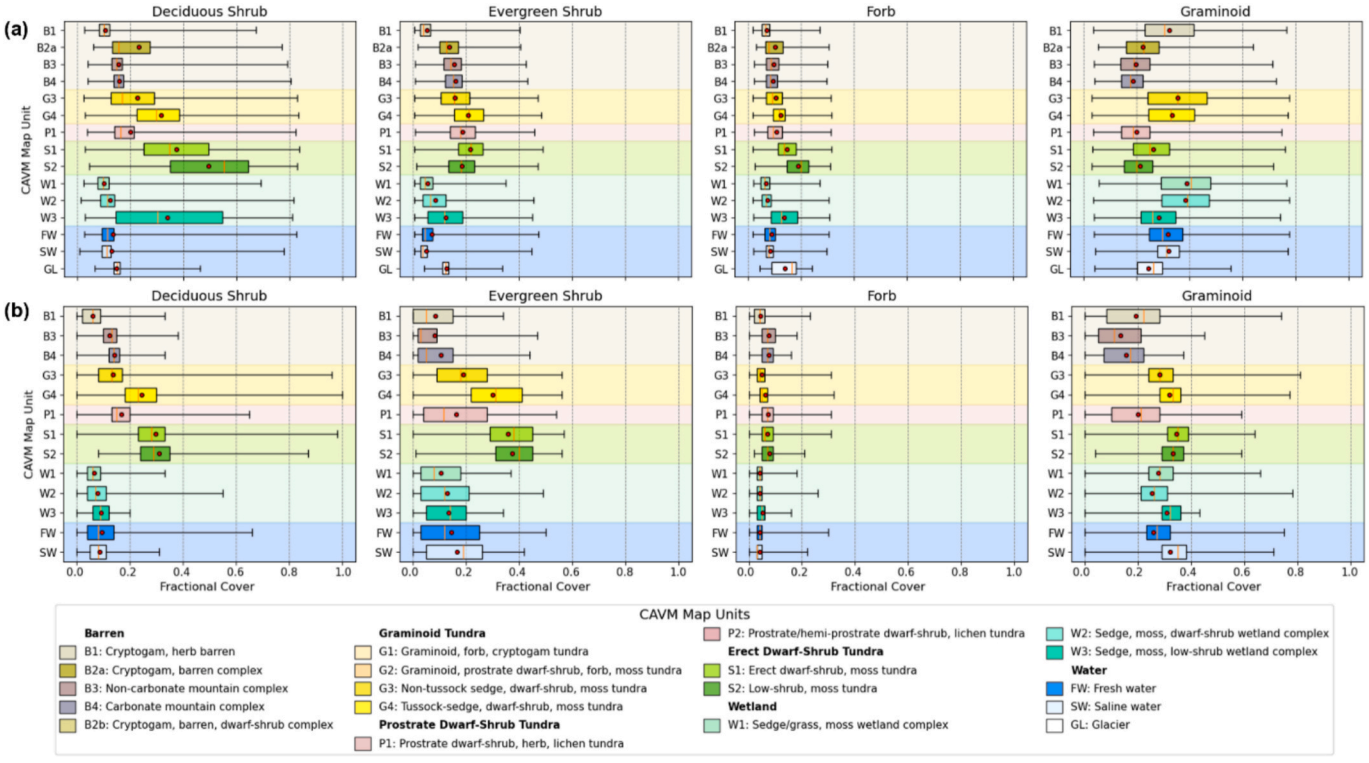


Fig. 10. Boxplots over the CAVM vegetation classes based on our PAVC-Gridded fCover (a) and that mapped by Macander et al. (2017) (b). Because our product spans the entire Alaskan Arctic tundra region, it preserves a greater number of CAVM vegetation classes (such as B2a and GL) compared to the product by Macander et al. (2017).

tgi, *osavi*, and *ci_g* (all with yellow dots at their SHAP maxima), whereas lichens are more strongly associated with *swir1*, brightness, and wetness. Finally, in the SAR domain, *VV* backscatter shows its brightest yellow dots on the left (negative SHAP) side for both evergreen shrubs and lichens, suggesting that high VV values are linked to lower predicted cover. The largest positive SHAP values (furthest right) instead correspond to lower-to-intermediate VV values (orange through purple). Conversely, *VH* shows yellow clusters at its positive SHAP extreme

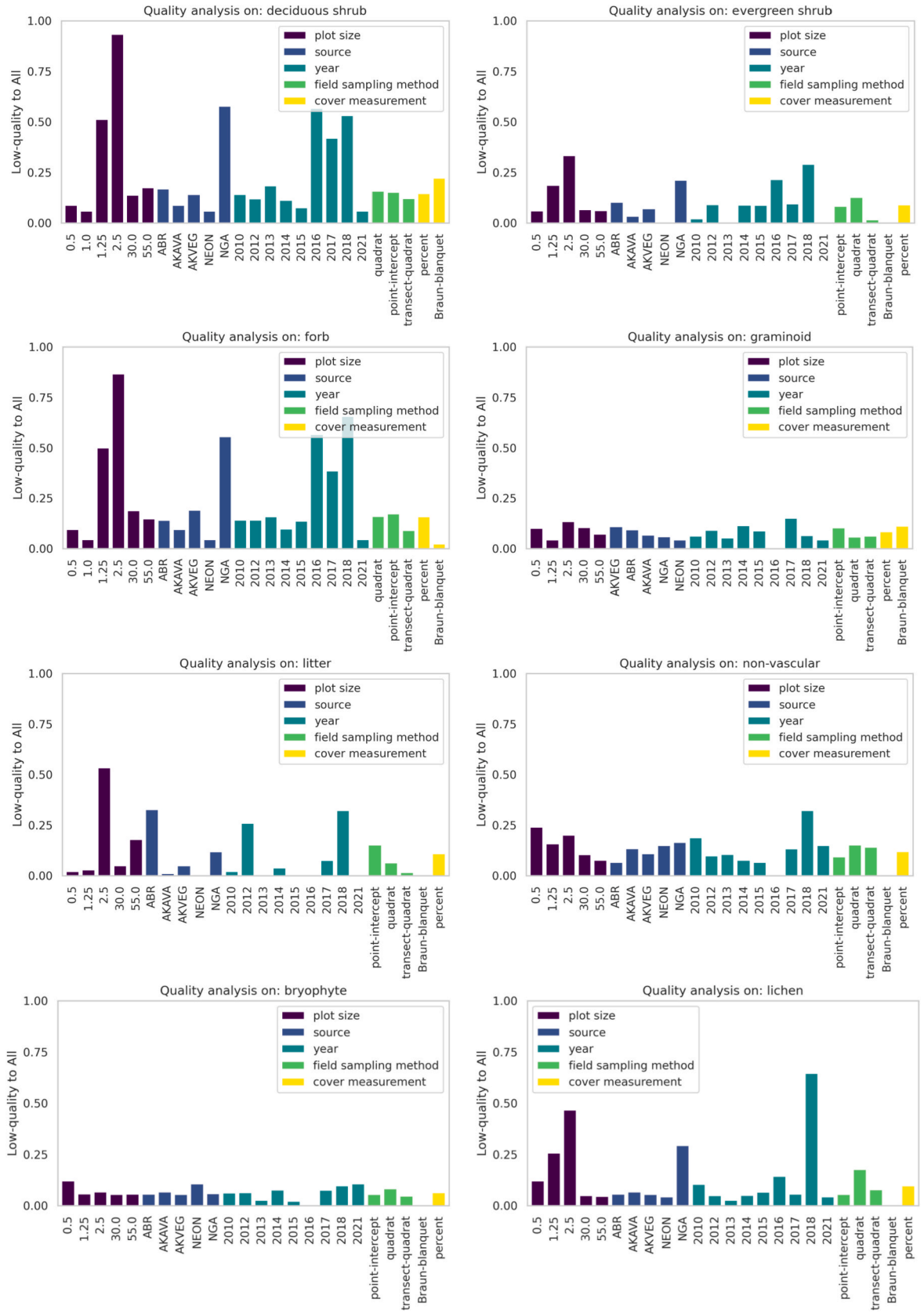


Fig. 11. Quality analysis on plots having low consistency with satellite data for all target PFTs, where plot size is measured in meters and the y-axis indicates the fraction of low-quality plots (with Cook's distance $> D^*$) within each category listed on the x-axis. In general, taller bars indicate more plots contributing variability to fCover regression from this category.

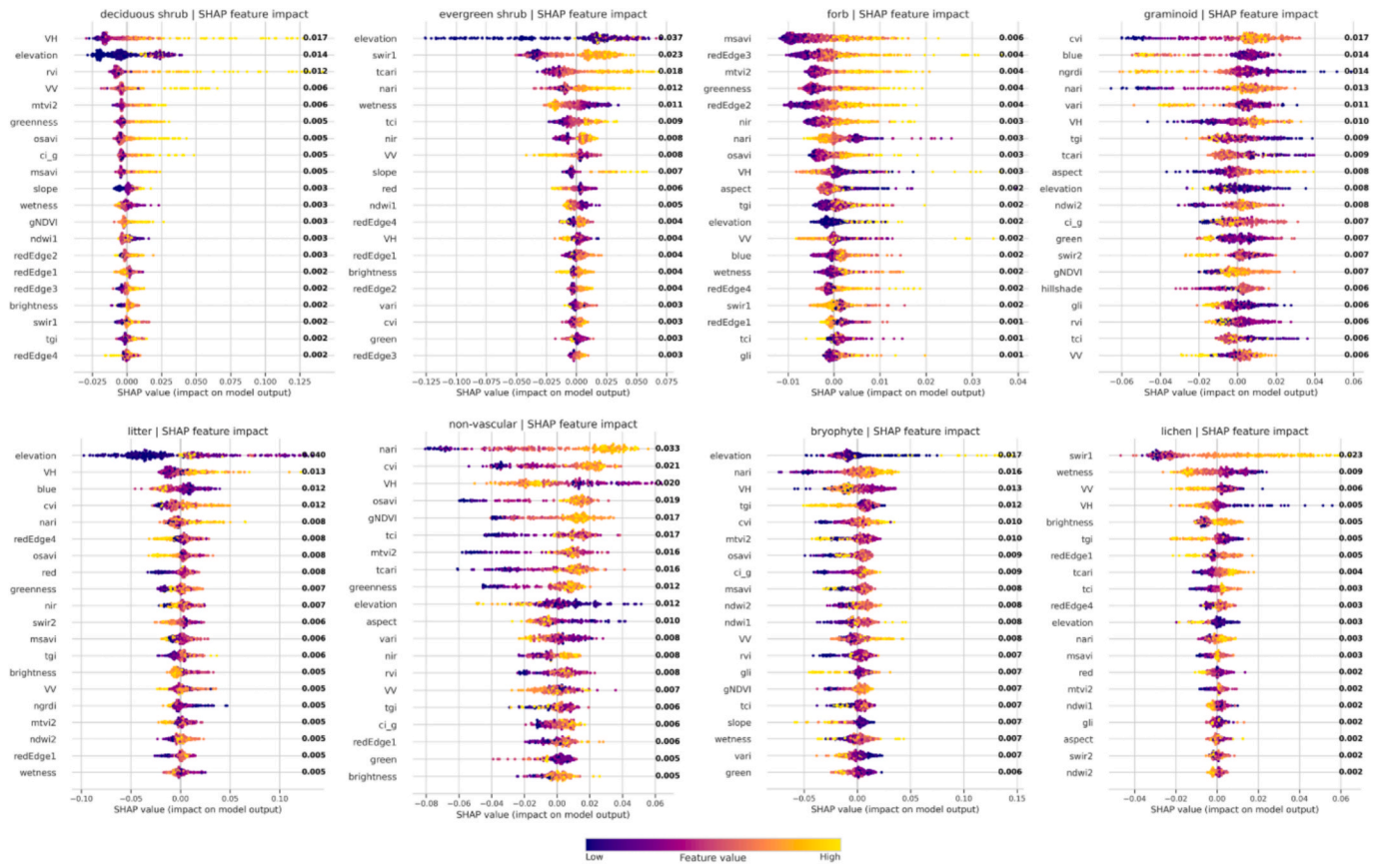


Fig. 12. SHAP analysis on all target PFTs, where bold numbers on the right indicate the mean absolute SHAP value. Only the top 20 most influential features are shown.

across most other PFTs (except for bryophytes), indicating that higher VH values are associated with higher fractional-cover predictions.

4. Discussion

4.1. Comparison between our mapped products and the published fCover products

High-quality PFT-level fCover is critical for accurate representation of tundra vegetation diversity and physiology (e.g., carbon capture, biomass allocation, snow trapping) in terrestrial ecosystem modeling (Sulman et al., 2021). By leveraging high-resolution Sentinel data and five plot sources spanning the entire North Slope and Seward Peninsula of Alaska, our machine learning-based fCover regression framework predicts reliable fCover for the targeted PFTs, including litters, non-vascular plant species (bryophytes and lichen), herbaceous species (forbs and graminoid) and woody shrubs (deciduous and evergreen). Compared with Macander et al. (2017), we not only expanded the study extent to the entire Alaskan tundra region to capture broader ecological variability, but also examined the spatial heterogeneity over different CAVM bioclimatic zones. The mapped fCover successfully reveals sub-pixel details of the spatial distribution of the PFTs of interest, which is crucial for understanding short-term disturbances that are often missed in pixel-based maps.

Moreover, we note that the fCover of nearly all of the PFTs is underestimated by Macander et al. (2017, 2022), especially for forbs, likely due to the lack of treatment for temporal inconsistencies among plot-sampled fCover (ABR and AKVEG) and satellite data. To address these inconsistencies, we employed a statistical quality control approach using Cook's distance to filter out low-quality plots that contributed a large amount of variability to regression modeling with recent satellite observations. With the high-quality plots, we substantially reduced the bias and enhanced the correlation between the satellite-predicted fCover and plot-observed counterparts for all PFTs.

In addition, we incorporated SHapley Additive exPlanations (SHAP) to interpret model predictions and quantify the relative contribution and directionality of each satellite-derived variable. These improvements, including robust feature importance analysis and data-driven plot screening, collectively advance the generation of PFT-level fCover maps and provide a better representation of the Alaskan Arctic tundra landscape. This, in turn, enhances our understanding of how local environmental factors (e.g., topography) shape PFT distributions and responses to climate change. Ultimately, the resulting maps can support initialization and benchmarking of terrestrial land surface models, improving the realism of future climate projections.

4.2. Importance of consistent plot sampling for fCover regression modeling

Our quality analysis suggests that influential factors responsible for the low consistency (e.g., phenological changes, vegetation growth) between plot and satellite observations are not consistent for different PFTs. In general, we observe greater temporal variability in forbs and deciduous shrubs compared to graminoids and evergreen shrubs, likely due to their more pronounced seasonal phenological changes such as rapid shifts in greenness, biomass, or moisture content during the growing season. The limited dynamics in evergreen shrubs and graminoids are also supported by the fact that their growth is largely allocated belowground to roots compared with leaves and other aboveground components where the latter can persist for multiple seasons (Iversen et al., 2015; Sulman et al., 2021). Likewise, litter is more affected by large temporal differences compared with living PFTs as it is primarily derived from deciduous PFTs. Although lichens and bryophytes are both

non-vascular plants, the difference in their fractional cover is not as substantial as that observed between other groups such as shrubs and herbs, apart from lichen samples collected from NGA. This is likely because both lichens and bryophytes are low-lying ground cover types whose fractional cover tends to remain relatively stable from year to year, except in cases of abrupt disturbances. Nevertheless, we notice the fCover consistency of lichens is more impacted by spatial heterogeneity (NGA) than that of bryophytes. Among all plot sources, NEON is more reliable in almost all PFT fCover estimation due to more recent and consistent sampling (2021). NGA, in contrast, contributes more bias than the other plot sources due to high spatial heterogeneity on the Seward Peninsula, and different plant communities from the North Slope, and hence could be under-represented in our plot inventory. Despite more consistent plot sizes that match the resolution of the satellites, ABR also contributes larger variation in deciduous PFTs and litters, likely because all their plots were collected in 2012. All these findings highlight the importance of consistent sampling of PFT fCover, especially those subject to seasonality (deciduous shrubs, forbs, litters) in future field work. Our quality control approach demonstrates its effectiveness in identifying high-quality plot samples that correlate well with satellite observations for building reliable fCover regression models when plot sampling ambiguity (e.g., plot size, approach used for sampling and fCover measurement, collection year) is present.

While our approach to filtering plots improved overall model performance, it is important to acknowledge potential tradeoffs. Cook's distance, originally designed for linear regression diagnostics, was used here as a conservative screening method to flag highly influential samples likely tied to measurement errors, geolocation issues, or outlier reflectance values. However, in the context of machine learning, such influential samples may not always be erroneous. Some may represent valid but complex or transitional vegetation conditions that are important for model generalization. Excluding such "hard" examples may reduce the model's ability to capture transitional or edge-case vegetation conditions, especially in heterogeneous landscapes. Future work may benefit from incorporating model-based uncertainty quantification (e.g., ensemble disagreement or residual spread) or adaptive weighting schemes that retain these samples while downweighting their influence, rather than excluding them entirely.

In addition, while the modeling framework integrates satellite imagery and plot data collected across multiple years, the current product is designed as a static representation tied specifically to the 2019 satellite observation window. This decision was made due to the limited availability and uneven temporal distribution of high-quality plot samples, which precludes robust year-to-year generalization. Temporal inconsistencies between plot measurements and satellite acquisition dates were instead addressed through a targeted quality control strategy (see Section 2.4), which improves the coherence between in-situ and remote sensing data by filtering samples with large statistical influence or mismatched conditions. As a result, the fractional cover maps generated in this study should be interpreted as snapshot estimates representative of Arctic vegetation conditions around 2019. Extending the model temporally for monitoring interannual dynamics would require more densely distributed and time-aligned plot measurements, as well as careful validation across years to ensure reliability under shifting environmental and phenological conditions. Nevertheless, the core framework (e.g., feature engineering, PFT-specific modeling, quality control) remains adaptable and could be expanded for future temporal applications with improved field data support.

4.3. Relative importance of input explanatory variables for modeling fCover

The feature importance analysis conveys key information regarding the controlling factors that explain the most about the fCover variations for each PFT. Due to differences in plant physiognomy and physiology, the relative importance of satellite-derived features differs across PFTs. For example, *VH* polarization is generally more influential than *VV*, particularly for annual and deciduous plants (e.g., forbs, deciduous shrubs), compared with perennial and evergreen plants (e.g., evergreen shrubs). This is likely because *VH* is more sensitive to changes in vegetation structure than *VV* (Bousbih et al., 2017; Vreugdenhil et al., 2020). Seasonal changes in deciduous plants may reflect more on *VH* variability. Among all topography related features, elevation consistently emerges as a key factor associated with fCover, likely due to its close link with temperature gradients. For tasseled cap indices, annual and deciduous plants (forbs and deciduous shrubs) are more sensitive to greenness than perennial and evergreen PFTs (graminoids, evergreen shrubs), as the changes in the former group mainly occur at the canopy layer. Although lichens and bryophytes are both non-vascular, lichens are essentially fungal symbionts attached to the ground surface, making them more sensitive to soil brightness and wetness. Regarding vegetation indices, though *ngrdi* has been found to outperform solar-induced chlorophyll fluorescence in end of season extraction of evergreen needleleaf forest (Yin et al., 2022), it is found less influential to other PFTs aside from graminoids for fCover regression modeling. A much higher SHAP sensitivity to *msavi* observed in forbs compared with deciduous and evergreen shrubs suggests that forbs are more exposed to soil background during leaf development stages. In contrast, deciduous shrubs, with their broader canopy, tend to show higher importance for *ci_g* and *gNDVI*, which relate more directly to leaf chlorophyll content and photosynthetic activity.

4.4. Future work

This study demonstrates the effectiveness of employing the Cook's distance in mitigating the spatio-temporal inconsistencies among plot sources that span the entire Alaskan-arctic tundra landscape with different field sampling approaches. The high-quality plots in combination with the spatially continuous satellite observations successfully estimate fCover that is highly correlated with plot-observed counterparts with low bias. We expect PAVC-Gridded fCover product containing spatially detailed fractional cover distribution for the included PFTs (litter, forb, graminoid, lichen, bryophyte, deciduous and evergreen shrubs) can greatly benefit the parameterization in terrestrial ecosystem models for more accurate simulation of vegetation and carbon cycle above and below ground and their projections under changing climate (Sulman et al., 2021). Nevertheless, a few limitations are acknowledged here which should be considered as a priority in future work.

First of all, due to limited plot samples, we chose to harmonize all plot sources to augment the training data for reliable predictions rather than performing regression modeling based on plots collected over every year to capture the PFT changes like Macander et al. (2022). To account for temporal inconsistency, we limited the training data to plots that were collected after 2010 with no fire history. Nevertheless, these older plots (earlier than 2015) still account for ~ 1/3 of the total training data with 5–9-year temporal gaps. Fortunately, our quality control effectively filtered out low-quality plots that are poorly correlated with satellite observations. Based on our quality analysis, we found the spatio-temporal inconsistencies affect the plot' quality more in annual and deciduous (forb, litter, and deciduous shrub) than in perennial and evergreen plants (graminoid, non-vascular plants, evergreen shrub), which highlights the necessity for expanding deciduous especially forb sampling in future field efforts. In addition, plots sampled from Seward Peninsula were appended to the north slope plot sources for sample augmentation. Yet in general, we find large spatial heterogeneity in these samples, which are mostly filtered out from our regional regression analysis. To address this spatial heterogeneity, we recommend performing a spatial representation analysis (Hoffman et al., 2013) to ensure the regional representativeness of the sampled plots and the trained regression models. To augment the field measured fCover, AVIRIS-NG (Chapman et al., 2019) can be introduced to leverage its spatial-and-spectral advantages over plot and multispectral satellite

imagery respectively as an intermediate product on fCover regression.

Secondly, shrubs subcategorized into different vegetation heights should be represented in detail in ecosystem modeling since they can function distinctly in climate-vegetation dynamics (Kropp et al., 2020; Loranty et al., 2018; Wullschleger et al., 2014; Sturm et al., 2001). For instance, low vegetation can act as insulation for snow cover, promoting higher albedo and potentially mitigating warming. In contrast, tall vegetation can deliver negative feedback to carbon flux by growing above snow cover, reducing the surface albedo, and trapping snow, which contributes to soil warming during winter and provides a cooling effect for the soils beneath during hotter summers. Although integrating shrub height stratification is highly relevant, its reliable implementation would require a substantially larger sample of in situ plots with accurately measured shrub heights, to build and validate stratified fCover models. At present, our available plot network does not include consistent canopy height data to support this finer categorization, making robust stratification beyond the scope of the current study. However, we recognize its importance and propose that a dedicated shrub height product could be developed in the future by integrating sparse canopy height observations from spaceborne lidar, such as ICESat-2, with spatially continuous satellite-derived explanatory variables (Travers-Smith et al., 2024). Alternatively, shrub heights can be mapped based on empirical relationships established between satellite-derived features and field-measured shrub heights (Bartsch et al., 2020). Future work combining such shrub height maps with fractional cover would enable a more nuanced stratification of shrub PFTs for climate–ecosystem modeling applications.

5. Conclusion

Accurate PFT-level fCover information is crucial for improving the vegetation composition and function estimate by the land surface model to a comparable level of the field measured counterpart. In this study, we successfully generated a 20-m wall-to-wall fCover for several typical PFTs at the Arctic-Alaskan tundra landscape by integrating the plot-observed fCover (i.e., the PAVC database) with Sentinel and ArcticDEM predictors using random forest regression models. The novelty of our fCover mapping work, compared with previous studies, lies in that we systematically filtered out plots having low consistency with recent satellite observations by adopting a quality control approach namely the Cook's distance. Our results suggest that PAVC-Gridded fCover estimated by the high-quality plot observations achieves much improved accuracy in comparison to the raw, unfiltered plots. Moreover, PAVC-Gridded fCover is also consistent with the distribution of CAVM vegetation classes and spatial pattern of vegetation in the satellite imagery more closely than that developed by Macander et al. (2017). We additionally find that annual and deciduous plants such as deciduous shrubs, forbs, and litters are more susceptible to spatio-temporal heterogeneity (e.g., small plots, older plots, plots that were collected from Seward Peninsula), which need more awareness in field sampling efforts. Lastly, when choosing the influential factors input for the fCover regression, we find deciduous plants more affected by variables that are related to canopies. In contrast, non-vascular lichens and bryophytes are more influenced by soil moisture content. For future work, we will incorporate additional data sources such as the AVIRIS-NG to augment the fCover training samples and introduce shrub heights information to further divide the shrub PFTs into dwarf, low, and tall subcategories that are associated with divergent eco-functions. Our high-resolution PAVC-Gridded fCover dataset provides valuable input for terrestrial ecosystem models, enabling more accurate simulations of Arctic vegetation dynamics, carbon fluxes, and climate-vegetation feedback under future climate scenarios.

Code and data availability statement

The harmonized plot inventory, Pan-Arctic Vegetation Cover (PAVC) database, is publicly accessible at Doi: 10.15485/2483557. The PAVC-Gridded fCover product at the PFT level developed by this study (across the Arctic tundra in Alaska at 20-m resolution) is available through ESS-DIVE at https://doi.org/10.15485/2513385. Satellite-derived explanatory variables used for model training and prediction, including Sentinel-1, Sentinel-2, and ArcticDEM products, can be accessed via Google Earth Engine. Scripts for mapping the PFT-level fCover and plot harmonization are available at https://github.com/climatemodeling/alaska_pft_fcover_modeling and https://github.com/climatemodeling/pavc respectively.

CRediT authorship contribution statement

Tianqi Zhang: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Formal analysis, Conceptualization. **Morgan R. Steckler:** Writing – review & editing, Visualization, Data curation. **Amy L. Breen:** Data curation, Writing – review & editing. **Forrest M. Hoffman:** Writing – review & editing, Supervision, Funding acquisition. **William W. Hargrove:** Writing – review & editing, Supervision. **Verity G. Salmon:** Writing – review & editing, Resources, Data curation. **Colleen M. Iversen:** Writing – review & editing, Resources, Funding acquisition. **Stan D. Wullschleger:** Writing – review & editing, Resources, Project administration. **Jitendra Kumar:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was partially supported by the NGEE Arctic project which is sponsored by the Biological and Environmental Research program in the Department of Energy's Office of Science. We thank the Mary's Igloo, Council, and Sitnasuak Native Corporations for their guidance and for permitting us to quantify vegetation cover on their lands.

Partial support was provided by the United States Army Corps of Engineers (USACE) Engineering Research and Development Center (ERDC) Geospatial Research Laboratory (GRL) and was accomplished under Cooperative Agreement Federal Award Identification Number (FAIN) W9132V-22-2-0001. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of USACE ERDC GRL or the U.S. Government.

This manuscript has been authored in part by UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the US Department of Energy (DOE). The publisher acknowledges the US government license to provide public access under the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jag.2025.104892.

References

Bao, T., Jia, G., Xu, X., 2021. Wetland Heterogeneity Determines methane Emissions: a Pan-Arctic Synthesis. Environ. Sci. Technol. 55, 10152–10163. https://doi.org/10.1021/acs.est.1c01616.

Bartsch, A., Widhalm, B., Leibman, M., Ermokhina, K., Kumpula, T., Skarin, A., Wilcox, E.J., Jones, B.M., Frost, G.V., Höfler, A., Pointner, G., 2020. Feasibility of

tundra vegetation height retrieval from Sentinel-1 and Sentinel-2 data. Remote Sens. Environ. 237, 111515. https://doi.org/10.1016/j.rse.2019.111515.
Bayle, A., Carlson, B.Z., Thierion, V., Isenmann, M., Choler, P., 2019. Improved Mapping of Mountain Shrublands using the Sentinel-2 Red-Edge Band. Remote Sens. (Basel) 11, 2807. https://doi.org/10.3390/rs11232807.
Bousbih, S., Zribi, M., Lili-Chabaane, Z., Baghdadi, N., El Hajj, M., Gao, Q., Mougenot, B., 2017. Potential of Sentinel-1 Radar Data for the Assessment of Soil and cereal Cover Parameters. Sensors 17, 2617. https://doi.org/10.3390/s17112617.
Breen, A., Iversen, C., Salmon, V., VanderStel, H., Busey, B., Wullschleger, S., 2020. NGEE Arctic plant traits: plant community composition, Kougarok road mile marker 64, Seward Peninsula, Alaska, 2016. Next Generation Ecosystems Experiment-Arctic, Oak Ridge National Laboratory 2020. https://doi.org/10.5440/1465967.
Chapman, J.W., Thompson, D.R., Helmlinger, M.C., Bue, B.D., Green, R.O., Eastwood, M. L., Geier, S., Olson-Duvall, W., Lundeen, S.R., 2019. Spectral and radiometric calibration of the next generation airborne visible infrared spectrometer (AVIRIS-NG). Remote Sens. (Basel) 11, 2129.
Chu, D., 2020. Fractional Vegetation Cover. In: Chu, D. (Ed.), Remote Sensing of Land Use and Land Cover in Mountain Region: A Comprehensive Study at the Central Tibetan Plateau. Springer, Singapore, pp. 195–207. https://doi.org/10.1007/978-981-13-7580-4_10.
Cook, R.D., 1977. Detection of influential observation in linear regression. Technometrics 19, 15–18.
Davidson, S.J., Santos, M.J., Sloan, V.L., Watts, J.D., Phoenix, G.K., Oechel, W.C., Zona, D., 2016. Mapping Arctic Tundra Vegetation Communities using Field Spectroscopy and Multispectral Satellite Data in North Alaska, USA. Remote Sens. (Basel) 8, 978. https://doi.org/10.3390/rs8120978.
Facelli, J.M., Pickett, S.T., 1991. Plant litter: its dynamics and effects on plant community structure. Bot. Rev. 57, 1–32.
Fernández-Guisuraga, J.M., Calvo, L., Quintano, C., Fernández-Manso, A., Fernandes, P. M., 2023. Fractional vegetation cover ratio estimated from radiative transfer modeling outperforms spectral indices to assess fire severity in several Mediterranean plant communities. Remote Sens. Environ. 290, 113542.
Gao, B.-C., 1996. NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. Remote Sens. Environ. 58, 257–266.
Gitelson, A.A., Gritz, Y., Merzlyak, M.N., 2003. Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. J. Plant Physiol. 160, 271–282.
Gitelson, A.A., Kaufman, Y.J., Merzlyak, M.N., 1996. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. Remote Sens. Environ. 58, 289–298. https://doi.org/10.1016/S0034-4257(96)00072-7.
Gitelson, A.A., Kaufman, Y.J., Stark, R., Rundquist, D., 2002. Novel algorithms for remote estimation of vegetation fraction. Remote Sens. Environ. 80, 76–87.
Golaz, J., Caldwell, P.M., Van Roekel, L.P., Petersen, M.R., Tang, Q., Wolfe, J.D., Abeshu, G., Anantharaj, V., Asay-Davis, X.S., Bader, D.C., Baldwin, S.A., Bisht, G., Bogenschutz, P.A., Branstetter, M., Brunke, M.A., Brus, S.R., Burrows, S.M., Cameron-Smith, P.J., Donahue, A.S., Deakin, M., Easter, R.C., Evans, K.J., Feng, Y., Flanner, M., Foucar, J.G., Fyke, J.G., Griffin, B.M., Hannay, C., Harrop, B.E., Hoffman, M.J., Hunke, E.C., Jacob, R.L., Jacobsen, D.W., Jeffery, N., Jones, P.W., Keen, N.D., Klein, S.A., Larson, V.E., Leung, L.R., Li, H., Lin, W., Lipscomb, W.H., Ma, P., Mahajan, S., Maltrud, M.E., Mametjanov, A., McClean, J.L., McCoy, R.B., Neale, R.B., Price, S.F., Qian, Y., Rasch, P.J., Reeves Eyre, J.E.J., Riley, W.J., Ringler, T.D., Roberts, A.F., Roesler, E.L., Salinger, A.G., Shaheen, Z., Shi, X., Singh, B., Tang, J., Taylor, M.A., Thornton, P.E., Turner, A.K., Veneziani, M., Wan, H., Wang, H., Wang, S., Williams, D.N., Wolfram, P.J., Worley, P.H., Xie, S., Yang, Y., Yoon, J., Zelinka, M.D., Zender, C.S., Zeng, X., Zhang, C., Zhang, K., Zhang, Y., Zheng, X., Zhou, T., Zhu, Q., 2019. The DOE E3SM coupled Model Version 1: Overview and Evaluation at Standard Resolution. J. Adv. Model. Earth Syst. 11, 2089–2129. https://doi.org/10.1029/2018MS001603.
Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sens. Environ. 202, 18–27.
Graversen, R.G., Wang, M., 2009. Polar amplification in a coupled climate model with locked albedo. Clim. Dyn. 33, 629–643.
Haboudane, D., Miller, J.R., Pattey, E., Zarco-Tejada, P.J., Strachan, I.B., 2004. Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. Remote Sens. Environ. 90, 337–352.
Haboudane, D., Miller, J.R., Tremblay, N., Zarco-Tejada, P.J., Dextraze, L., 2002. Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. Remote Sens. Environ. 81, 416–426.
Haboudane, D., Tremblay, N., Miller, J.R., Vigneault, P., 2008. Remote estimation of crop chlorophyll content using spectral indices derived from hyperspectral data. IEEE Trans. Geosci. Remote Sens. 46, 423–437.
Hoffman, F.M., Kumar, J., Mills, R.T., Hargrove, W.W., 2013. Representativeness-based sampling network design for the State of Alaska. Landsc. Ecol. 28, 1567–1586. https://doi.org/10.1007/s10980-013-9902-0.
Hunt Jr, E.R., Daughtry, C.S.T., Eitel, J.U., Long, D.S., 2011. Remote sensing leaf chlorophyll content using a visible band index. Agron. J. 103, 1090–1099.
Iversen, C.M., Sloan, V.L., Sullivan, P.F., Euskirchen, E.S., McGuire, A.D., Norby, R.J., Walker, A.P., Warren, J.M., Wullschleger, S.D., 2015. The unseen iceberg: plant roots in arctic tundra. New Phytol. 205, 34–58. https://doi.org/10.1111/nph.13003.
Joabsson, A., Christensen, T.R., 2001. Methane emissions from wetlands and their relationship with vascular plants: an Arctic example. Glob. Chang. Biol. 7, 919–932. https://doi.org/10.1046/j.1354-1013.2001.00044.x.
Jordan, C.F., 1969. Derivation of leaf-area index from quality of light on the forest floor. Ecology 50, 663–666.
Kannan, K.S., Manoj, K., 2015. Outlier detection in multivariate data. Appl. Math. Sci. 47, 2317–2324.
Kauth, R.J., Thomas, G.S., 1976. The tasselled cap–a graphic description of the spectral-temporal development of agricultural crops as seen by Landsat. LARS Symposia, in, p. 159.
Korhonen, L., Packalen, P., Rautiainen, M., 2017. Comparison of Sentinel-2 and Landsat 8 in the estimation of boreal forest canopy cover and leaf area index. Remote Sens. Environ. 195, 259–274.
Kropp, H., Loranty, M.M., Natali, S.M., Kholodov, A.L., Rocha, A.V., Myers-Smith, I., Abbot, B.W., Abermann, J., Blanc-Betes, E., Blok, D., 2020. Shallow soils are warmer under trees and tall shrubs across Arctic and Boreal ecosystems. Environ. Res. Lett. 16, 015001.
Konduri, Venkata, et al. "Hyperspectral remote sensing-based plant community map for region around NGEE-Arctic intensive research watersheds at Seward Peninsula, Alaska, 2017-2019." , Aug. 2022. Doi: 10.5440/1828604.
Li, W., Niu, Z., Shang, R., Qin, Y., Wang, L., Chen, H., 2020. High-resolution mapping of forest canopy height using machine learning by coupling ICESat-2 LiDAR with Sentinel-1, Sentinel-2 and Landsat-8 data. Int. J. Appl. Earth Obs. Geoinf. 92, 102163. https://doi.org/10.1016/j.jag.2020.102163.
Loranty, M.M., Abbott, B.W., Blok, D., Douglas, T.A., Epstein, H.E., Forbes, B.C., Jones, B. M., Kholodov, A.L., Kropp, H., Malhotra, A., Mamet, S.D., Myers-Smith, I.H., Natali, S.M., O'Donnell, J.A., Phoenix, G.K., Rocha, A.V., Sonnentag, O., Tape, K.D., Walker, D.A., 2018. Reviews and syntheses: changing ecosystem influences on soil thermal regimes in northern high-latitude permafrost regions. Biogeosciences 15, 5287–5313. https://doi.org/10.5194/bg-15-5287-2018.
Louhaichi, M., Borman, M.M., Johnson, D.E., 2001. Spatially located Platform and Aerial Photography for Documentation of Grazing Impacts on Wheat. Geocarto Int. 16, 65–70. https://doi.org/10.1080/10106040108542184.
Louis, J., Debaecker, V., Pflug, B., Main-Knorn, M., Bieniarz, J., Mueller-Wilm, U., Cadau, E., Gascon, F., 2016. Sentinel-2 Sen2Cor: L2A processor for users, in: Proceedings Living Planet Symposium 2016. Spacebooks Online, pp. 1–8.
Lundberg, S.M., Lee, S.I., 2017. A unified approach to interpreting model predictions. In: Advances in Neural Information Processing Systems, p. 30.
Macander, M.J., Frost, G.V., Nelson, P.R., Swingley, C.S., 2017. Regional Quantitative Cover Mapping of Tundra Plant Functional Types in Arctic Alaska. Remote Sens. (Basel) 9, 1024. https://doi.org/10.3390/rs9101024.
Macander, M.J., Nelson, P.R., Nawrocki, T.W., Frost, G.V., Orndahl, K.M., Palm, E.C., Wells, A.F., Goetz, S.J., 2022. Time-series maps reveal widespread change in plant functional type cover across Arctic and boreal Alaska and Yukon. Environ. Res. Lett. 17, 054042. https://doi.org/10.1088/1748-9326/ac6965.
Myers-Smith, I.H., Forbes, B.C., Wilmking, M., Hallinger, M., Lantz, T., Blok, D., Tape, K. D., Macias-Fauria, M., Sass-Klaassen, U., Lévesque, E., Boudreau, S., Ropars, P., Hermanutz, L., Trant, A., Collier, L.S., Weijers, S., Rozema, J., Rayback, S.A., Schmidt, N.M., Schaepman-Strub, G., Wipf, S., Rixen, C., Ménard, C.B., Venn, S., Goetz, S., Andreu-Hayles, L., Elmendorf, S., Ravolainen, V., Welker, J., Grogan, P., Epstein, H.E., Hik, D.S., 2011. Shrub expansion in tundra ecosystems: dynamics, impacts and research priorities. Environ. Res. Lett. 6, 045509. https://doi.org/10.1088/1748-9326/6/4/045509.
Myers-Smith, I.H., Kerby, J.T., Phoenix, G.K., Bjerke, J.W., Epstein, H.E., Assmann, J.J., John, C., Andreu-Hayles, L., Angers-Blondin, S., Beck, P.S.A., Berner, L.T., Bhatt, U. S., Bjorkman, A.D., Blok, D., Bryn, A., Christiansen, C.T., Cornelissen, J.H.C., Cunliffe, A.M., Elmendorf, S.C., Forbes, B.C., Goetz, S.J., Hollister, R.D., de Jong, R., Loranty, M.M., Macias-Fauria, M., Maseyk, K., Normand, S., Olofsson, J., Parker, T. C., Parmentier, F.-J.-W., Post, E., Schaepman-Strub, G., Stordal, F., Sullivan, P.F., Thomas, H.J.D., Tømmervik, H., Treharne, R., Tweedie, C.E., Walker, D.A., Wilmking, M., Wipf, S., 2020. Complexity revealed in the greening of the Arctic. Nat. Clim. Chang. 10, 106–117. https://doi.org/10.1038/s41558-019-0688-1.
Nawrocki, T.W., Carlson, M.L., Osnas, J.L.D., Trammell, E.J., Witmer, F.D.W., 2020. Regional mapping of species-level continuous foliar cover: beyond categorical vegetation mapping. Ecol. Appl. 30, e02081. https://doi.org/10.1002/eap.2081.
NEON (National Ecological Observatory Network). Plant presence and percent cover (DP1.10058.001) RELEASE-2023. NEON Data Portal Doi: 10.48443/9579-a253 (2023).
Noh, M.-J., Howat, I.M., 2017. The Surface Extraction from TIN based Search-space Minimization (SETSM) algorithm. ISPRS J. Photogramm. Remote Sens. 129, 55–76. https://doi.org/10.1016/j.isprsjprs.2017.04.019.
Pinho, L.G.B., Nobre, J.S., Singer, J.M., 2015. Cook's distance for generalized linear mixed models. Comput. Stat. Data Anal. 82, 126–136.
Qi, J., Chehbouni, A., Huete, A.R., Kerr, Y.H., Sorooshian, S., 1994. A modified soil adjusted vegetation index. Remote Sens. Environ. 48, 119–126. https://doi.org/10.1016/0034-4257(94)90134-1.
Rantanen, M., Karpechko, A.Y., Lipponen, A., Nordling, K., Hyvärinen, O., Ruosteenoja, K., Vihma, T., Laaksonen, A., 2022. The Arctic has warmed nearly four times faster than the globe since 1979. Commun. Earth Environ. 3, 1–10. https://doi.org/10.1038/s43247-022-00498-3.
Raynolds, M.K., Walker, D.A., Balser, A., Bay, C., Campbell, M., Cherosov, M.M., Daniëls, F.J.A., Eidesen, P.B., Ermokhina, K.A., Frost, G.V., Jedrzejek, B., Jorgenson, M.T., Kennedy, B.E., Kholod, S.S., Lavrinenko, I.A., Lavrinenko, O.V., Magnússon, B., Matveyeva, N.V., Metúsalemsson, S., Nilsen, L., Olthof, I., Pospelov, I.N., Pospelova, E.B., Pouliot, D., Razzhivin, V., Schaepman-Strub, G., Šibík, J., Telyatnikov, M.Y., Troeva, E., 2019. A raster version of the Circumpolar Arctic Vegetation Map (CAVM). Remote Sens. Environ. 232, 111297. https://doi.org/10.1016/j.rse.2019.111297.
Rondeaux, G., Steven, M., Baret, F., 1996. Optimization of soil-adjusted vegetation indices. Remote Sens. Environ. 55 (2), 95–107.

Salmon, V.G., Breen, A.L., Kumar, J., Lara, M.J., Thornton, P.E., Wullschleger, S.D., Iversen, C.M., 2019. Alder distribution and expansion across a tundra hillslope: implications for local N cycling. Front. Plant Sci. 10, 1099.

Scaramuzza, P., Barsi, J., 2005. Landsat 7 scan line corrector-off gap-filled product development. Proceeding of Pecora. 23–27.

Shukla, P.R., Skea, J., Calvo Buendia, E., Masson-Delmotte, V., Pörtner, H.O., Roberts, D. C., Zhai, P., Slade, R., Connors, S., Van Diemen, R., 2019. IPCC, 2019: Climate Change and Land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems.

Sloan, V.L., Brooks, J.D., Wood, S.J., Liebig, J.A., Siegrist, J., Iversen, C.M., Norby, R.J., 2014. Plant community composition and vegetation height, Barrow, Alaska, ver. 1. Next Generation Ecosystem Experiments Arctic Data Collection, Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA. Data set accessed at http://dx. doi. org/10.5440/1129476.

Steckler, M.R., Kumar, J., Breen, A.L., Zhang, T., Hoffman, F.M., Hargrove, W.W., Walker, D.A., Wells, A.F., Droghini, A., Nawrocki, T.W., Wullschleger, S.D., Macander, M.J., Frost, G.V., Salmon, V.G., Iversen, C.M., 2024. Initialization of Pan-Arctic Vegetation Cover (PAVC) database containing synthesized Alaska plot data collected after 2010. Next-Generation Ecosystem Experiments (NGEE) Arctic, ESS-DIVE repository. Dataset. https://doi:10.15485/2483557.

Steckler, M.R., Kumar, J., Breen, A.L., Zhang, T., Hoffman, F.M., Hargrove, W.W., Walker, D.A., Wells, A.F., Droghini, A., Nawrocki, T.W., Wullschleger, S.D., Macander, M.J., Frost, G.V., Salmon, V.G., Barnett, D.T., Iversen, C.M., 2025. PAVC: the foundation for a Pan-Arctic Vegetation Cover database. Sci. Data. https://doi.org/10.1038/s41597025053269.

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

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

Sulman, B.N., Salmon, V.G., Iversen, C.M., Breen, A.L., Yuan, F., Thornton, P.E., 2021. Integrating Arctic Plant Functional Types in a Land Surface Model using Above- and Belowground Field Observations. J. Adv. Model. Earth Syst. 13. https://doi.org/10.1029/2020MS002396.

Travers-Smith, H., Coops, N.C., Mulverhill, C., Wulder, M.A., Ignace, D., Lantz, T.C., 2024. Mapping vegetation height and identifying the northern forest limit across Canada using ICESat-2, Landsat time series and topographic data. Remote Sens. Environ. 305, 114097. https://doi.org/10.1016/j.rse.2024.114097.

Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sens. Environ. 8, 127–150.

Veci, L., Prats-Iraola, P., Scheiber, R., Collard, F., Fomferra, N., Engdahl, M., 2014. The sentinel-1 toolbox, in. In: Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS). IEEE, pp. 1–3.

Villarreal, S., 2013. International polar year (ipy) back to the future (btf): changes in arctic ecosystem structure over decadal times scales. The University of Texas at El Paso. PhD Thesis.

Villarreal, S., Hollister, R.D., Johnson, D.R., Lara, M.J., Webber, P.J., Tweedie, C.E., 2012. Tundra vegetation change near Barrow, Alaska (1972–2010). Environ. Res. Lett. 7, 015508.

Vincini, M., Frazzi, E., D'Alessio, P., 2008. A broad-band leaf chlorophyll vegetation index at the canopy scale. Precis. Agric. 9, 303–319. https://doi.org/10.1007/s11119-008-9075-z.

Vreugdenhil, M., Navacchi, C., Bauer-Marschallinger, B., Hahn, S., Steele-Dunne, S., Greimeister-Pfeil, I., Dorigo, W., Wagner, W., 2020. Sentinel-1 Cross Ratio and Vegetation Optical Depth: a Comparison over Europe. Remote Sens. (Basel) 12, 3404. https://doi.org/10.3390/rs12203404.

Walker, D.A., Buchhorn, M., Kanevskiy, M., Matyshak, G.V., Raynolds, M.K., Shur, Y.L., Wirth, L.M., 2015. Infrastructure-Thermokarst-Soil-Vegetation interactions at Lake Colleen Site a, Prudhoe Bay, Alaska. Alaska Geobotany Center Data Report AGC 15.

Walker, D.A., Daniëls, F.J.A., Alsos, I., Bhatt, U.S., Breen, A.L., Buchhorn, M., Bültmann, H., Druckenmiller, L.A., Edwards, M.E., Ehrich, D., Epstein, H.E., Gould, W.A., Ims, R.A., Meltofte, H., Raynolds, M.K., Sibik, J., Talbot, S.S., Webber, P.J., 2016. Circumpolar Arctic vegetation: a hierarchic review and roadmap toward an internationally consistent approach to survey, archive and classify tundra plot data. Environ. Res. Lett. 11, 055005. https://doi.org/10.1088/1748-9326/11/5/055005.

Walker, D.A., Raynolds, M.K., Daniëls, F.J.A., Einarsson, E., Elvebakk, A., Gould, W.A., Katenin, A.E., Kholod, S.S., Markon, C.J., Melnikov, E.S., Moskalenko, N.G., Talbot, S.S., Yurtsev, B.A.(†), Team, T. other members of the C., 2005. The Circumpolar Arctic vegetation map. J. Veget. Sci. 16, 267–282. https://doi.org/10.1111/j.1654-1103.2005.tb02365.x.

Wullschleger, S.D., Epstein, H.E., Box, E.O., Euskirchen, E.S., Goswami, S., Iversen, C.M., Kattge, J., Norby, R.J., Van Bodegom, P.M., Xu, X., 2014. Plant functional types in Earth system models: past experiences and future directions for application of dynamic vegetation models in high-latitude ecosystems. Ann. Bot. 114, 1–16.

Yang, D., Morrison, B.D., Hanston, W., McMahon, A., Baskaran, L., Hayes, D.J., Miller, C. E., Serbin, S.P., 2023. Integrating very-high-resolution UAS data and airborne imaging spectroscopy to map the fractional composition of Arctic plant functional types in Western Alaska. Remote Sens. Environ. 286, 113430. https://doi.org/10.1016/j.rse.2022.113430.

Yin, G., Verger, A., Descals, A., Filella, I., Peñuelas, J., 2022. A broadband green-red vegetation index for monitoring gross primary production phenology. J. Remote Sens. https://doi.org/10.34133/2022/9764982.

Zhang, T., Liu, D., 2023. Improving ICESat-2-based boreal forest height estimation by a multivariate sample quality control approach. Methods Ecol. Evol. 14, 1623–1638. https://doi.org/10.1111/2041-210X.14112.

Zhang, T., Steckler M.R., Breen, A.L., Hoffman, F.M., Hargrove, W.W., Salmon, V.G., Iversen, C.M., Wullschleger, S.D., Kumar, J., 2025. Pan-Arctic Vegetation Cover (PAVC) Gridded: High resolution fractional coverage maps of plant functional types at 20-meter spatial resolution. Next-Generation Ecosystem Experiments (NGEE) Arctic, ESS-DIVE repository. Dataset. Doi: 10.15485/2513385.

Zhao, Q., Yu, S., Zhao, F., Tian, L., Zhao, Z., 2019. Comparison of machine learning algorithms for forest parameter estimations and application for forest quality assessments. For. Ecol. Manage. 434, 224–234.