

Predicting plants in the wild: Mapping arctic and boreal plants with UAS-based visible and near infrared reflectance spectra

Peter R. Nelson^{a,b,*}, Kenneth Bundy^a, Kevaghna. Smith^a, Matt. Macander^c, Catherine Chan^d

^a University of Maine, United States

^b Laboratory of Ecological Spectroscopy – lecospec, LLC, United States

^c ABR Inc, United States

^d University of Nebraska, United States

ARTICLE INFO

Keywords:

Adaboost
Alaska
Bagged Regression Trees
Drone
Classification
Ground validation
Fairbanks
Hyperspectral
Image processing
Machine learning
Optimization
Open-source
Partial Least-Squares Linear Discriminant
Analysis (PLS-DA)
Ranger
R
Spectral Library
Transferability
Vegetation indices

ABSTRACT

Biophysical changes in the Arctic and boreal zones drive shifts in vegetation, such as increasing shrub cover from warming soil or loss of living mat species due to fire. Understanding current and future responses to these factors requires mapping vegetation at a fine taxonomic resolution and landscape scale. Plants vary in size and spectral signatures, which hampers mapping of meaningful functional groups at coarse spatial resolution. Fine spatial grain of remotely sensed data (<10 cm pixels) is often necessary to resolve patches of many Arctic and boreal plant groups, such as bryophytes and lichens, which are significant components of terrestrial vegetation cover. Separation of co-occurring small vegetation patches in images also requires high spectral resolution. Our goal here was to test the capabilities of UAS-based imaging spectroscopy for mapping plant functional types (PFT) using high spatial and spectral resolution data over Arctic and boreal vegetation at four sites in central Alaska. We then tested several Machine and Deep learning models of PFT cover using the reflectance spectra. The best models were very simple, balancing both bias (overfitting caused by imbalance sample sizes) and variance (fit to the independent validation data), explaining > 50 % of the independent ground cover estimation and > 84 % accuracy in estimating validation pixels. We explored the impact of spectral resolution on PFT mapping by including vegetation indices and a gradient of narrow (5 nm) to wide (50 nm) band features in our classification models across. Vegetation indices were the most important predictors for classifying PFTs, while including band features improved models, with narrow and wide bandwidths having similar importance but models with wide bandwidths performing slightly better. We conclude that Arctic and boreal PFT reflectance can be pooled across sites for mapping with relatively few labeled pixels. Underfit, simple algorithms outperformed deep learning, at least with these small sample sizes, in classifying PFTs by balancing bias and variance. Future work should aim to increase the number of labeled pixels and the detail of labels to further improve mapping taxonomic precision.

1. Introduction

1.1. Vegetation composition and the challenges of spatial and spectral scale

A vital challenge to understanding the impacts of accelerating climate remains mapping vegetation composition and function across the Arctic and boreal, which combined are the largest terrestrial biomes on earth. The identity of surfaces, especially plants, at the most precise taxonomic rank is essential to understanding ecosystem processes. While satellite remote sensing has enabled scaling ground surface vegetation measurements across the landscape, the pixel size of freely available

satellites is too large (≥ 10 m) to correctly estimate many properties of surface patches (Rocchini et al., 2016; Siewert and Olofsson 2020). Uncertainty in estimating the composition of ground surface mixtures in the Arctic and boreal is driven by the fact that most areas are fine mixtures of living vegetation, litter, gravel, rock, soil, water, snow, and ice in mixed in patches from 1 cm^2 to 100's m^2 (Räsänen and Virtanen 2019; Thomson et al., 2021). Pixels from satellite images in this region are therefore usually admixtures of spectrally heterogeneous surfaces which complicates estimates of many basic parameters, such as accurately estimating soil erosion (Kodl et al., 2024) which are then used for upscaling for ecosystem modeling Siewert and Olofsson 2020. Scale mismatches between global models and regional to local studies

* Corresponding author.

E-mail address: peter.ross.nelson@gmail.com (P.R. Nelson).

<https://doi.org/10.1016/j.jag.2024.104156>

Received 6 May 2024; Received in revised form 1 September 2024; Accepted 5 September 2024

Available online 11 September 2024

1569-8432/© 2024 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

underscore the need for measuring and understanding scale dependent properties of vegetation (Xu et al., 2009).

Arctic and boreal ecosystems have diverse reflectance spectra (Nelson et al., 2022). Vegetation diversity is one major driver of spectral diversity of vegetation, along with phenology. Green vascular vegetation spectra are relatively similar at peak phenology (i.e. midsummer) yet consist of a large number of species (Liu et al., 2023). Other surfaces, such as lichens and rocks, have highly heterogeneous reflectance signatures within the same group (Rees et al., 2004). Remotely sensed spectra within the Arctic and boreal are characterized by both high variability within and between taxonomic groups (Nelson et al., 2022). Spectral resolution of Landsat and Sentinel-2 satellite images have up to 10 bands that are usually too wide to separate similar looking surfaces, such as trees or shrubs in the same genus but different species (Pinto-Ledezma and Cavender-Bares, 2021). To achieve high taxonomic precision, recent studies needed to use structural measures of the vegetation to separate optically similar vegetation (Orndahl et al., 2022; Yang et al., 2023). Remote sensing data capable of separating tundra surfaces must therefore have sufficient spectral and spatial resolution to distinguish surfaces, which is a major challenge to collect given the challenges including a narrow phenological window, low-sun angle, and logistical access (Nelson et al., 2022). There is a long standing belief that there is a connection between spectral variation and vegetation diversity (Palmer, 2008). While this may be true and certainly of interest, this connection is tenuous depending on many factors, such as diversity metric selected and the size of the spatial window in which diversity is measured (Torresani et al., 2024). Instead of diversity metrics of plants, we focused on the straight-forward estimation of plant groups themselves so that individual patches or groups can be mapped across larger spatial scales. Our study shares the same idea as Palmer (2008) of extending the botanist's eye beyond what can be visited on foot to identify mixtures of plants in the landscape.

NASA, the National Science Foundation (NSF), and other U.S. Earth observing agencies and organizations have recently invested in collecting higher spatial and spectral resolution imaging spectroscopy datasets by fixed wing aircraft over Arctic and boreal zones. NASA's AVIRIS (Airborne Visible/Infrared Imaging Spectrometer) – Next Generation (AVIRIS-NG), NASA Goddard LiDAR Hyperspectral Thermal (G-LiHT) system and NSF National Ecological Observatory Network's Airborne Observation Platform (NEON AOP) have higher spatial (<10m) and spectral resolution than available satellite data to bridge the gaps in retrieval of surface properties. All three platforms have been deployed extensively in Alaska over the last 5 years. Our intent here is to explore higher spatial and similar spectral resolution spectroscopy in the Arctic and boreal to better understand the composition of the ground conditions. To do this, we use an airborne imaging spectrometer on UAS, which has been shown to be a key spatial scaling tool in the Arctic and boreal vegetation (Siewert and Olofsson 2020; Thomson et al., 2021; Yang et al., 2022). We also aimed to understand what kinds of models and forms of data manipulation may aid in this mapping effort.

1.2. Ecologically Informative Mappable plant functional types (PFTs)

Plant functional types (PFTs) represent our best approximation of species that distill down to a mixture of morphology, taxonomy and sizes of vegetation while still representing a meaningful taxonomic class (Ustin and Gamon, 2010). Although significant progress has been made in estimating functional traits in plants in some other biomes, only a few studies have done so in the Arctic and boreal. The best example of this used a few vegetation classes studied over small spatial extents to nicely demonstrate separation of Arctic and boreal plant groups (Thomson et al., 2021). Furthermore, estimating plant traits does not produce the same understanding as taxonomic categories used for estimating biological diversity. Nonetheless, spectral variability within and between species has prevented remote sensing collectively from being able to reliably predict species occurrence or cover within bigger pixels.

Instead, PFTs are used to represent vegetation species groups but their application has been uneven in terms of taxonomy, comprehensiveness of taxa and other co-occurring abiotic surfaces.

Definition and selection of PFTs affects the utility and accuracy of a vegetation map while varying PFT classification approaches preclude direct comparison between studies. While mapping species is the ideal, this is not yet a realistic goal in the Arctic and boreal (Nelson et al., 2022). One alternative is to use a first-principles approach proposed PFT selection driven by ecological theory balanced with empirical reflectance differences (Kattenborn et al., 2019). However, the operational limitations of detection remain at some aggregated groupings of plants. For our study, we selected PFTs classes based on these considerations, using taxonomic structure to guide our PFT classes while grouping many heterogeneous classes together, understanding that many are dominated by only a few members (e.g. most PFTs have only a few dominant species). We also selected PFTs to match general categories of recent mapping efforts (Macander et al., 2022; Orndahl et al., 2022) to maximize comparability across studies.

1.3. Translating data and imagery into maps

1.3.1. Reflectance data analysis history/challenges

Past modeling efforts to classify or estimate fractional cover of PFTs in the Arctic, boreal and elsewhere utilized statistical methods (Schaaf et al., 2011), traditional machine learning (e.g. random forest) (Macander et al., 2017; Orndahl et al., 2022; Yang et al., 2023) and deep neural networks (Hong et al., 2023; Yokoya et al., 2023). The majority of these, apart from deep learning methods, utilize site-specific classifiers. These approaches utilize data from a single image to train their statistical or ML models. As a result, inference is limited to that particular scene. This limits the ability of this approach to scale between images or across larger areas of the boreal and Arctic.

Deep learning approaches, on the other hand, offer transferable models. These approaches typically train models from satellite images and classify the landscape at coarse resolutions. Many of these models are targeted at classifying farmland or forests, which are primarily large patches of similar vegetation. This approach, in addition to these limitations of spatial resolution, are limited by the amount of data required to train them. For example, SpectralGPT, a pre-trained spectral transformer foundation model, was trained on one million images (Hong et al., 2023). Past U-net based computer vision approaches have been trained on tens or hundreds of thousands of images. There is no such comparable dataset for the Arctic and boreal yet, so training such a model would require tremendous amounts of data collection. While these models offer excellent performance across most of the globe, the lack of data representative of the Arctic and boreal zone limits their applicability to remote sensing in the tundra.

1.3.2. Vegetation indices as a means for data reduction and transferability between images

While great care is taken to collect reflectance data at the same time of year under optimal illumination and atmospheric conditions, inherent features within each image creates variation that can cause the same PFT to appear different between locations. This reflectance variation can be caused by many factors that are hard to control, such as cloud cover, bidirectional reflectance (BRDF), shadows at the local scale (nearby tree) or landscape scale (shaded valley). Correcting for these effects is ideal but often not practical, especially for very high spatial resolution imagery like used here.

In lieu of model-based atmospheric and BRDF corrections, we used the simple but effective empirical line method per flight to standardize reflectance across sites. We then use traditional machine learning with a wide array of vegetation indices as features in addition to reflectance as predictors. Our rationale comes from the well-established approach of band ratios, in the simplest for one band divided by another, then to normalized vegetation indices (e.g. NDVI), and to more complex,

nuanced shape metrics (e.g. 1st derivative of a segment of the reflectance profile). These vegetation indices (VIs) utilize band ratios and relative differences in reflectance across a range of the bandpass available, therefore limiting the effects of differences between scenes and flights. Models using VIs are therefore likely to be more transferable across scenes, flights, and locations. Machine learning models including VIs as covariates can be trained with much smaller amounts of data than the deep learning approach while mitigating the primary challenge of the traditional machine learning approach.

1.4. Overall study goals

We developed a transferable (pooling data across sites), efficient, and scalable approach to map PFTs that maintains high spatial and spectral resolution of remotely sensed data. We used an imaging spectrometer on a UAS and detailed rapid ground cover measurements. Our workflow of mapping using remotely sensed imagery involves three main steps; 1)

taking ground measurements (cover by PFT or species) on a series of plots or sample units, 2) acquiring imagery of the locations where ground measurements occurred and acquiring a spectral signature of the plot or sample unit from the remotely sensed image and 3) modeling the ground measured surface attributes (e.g. presence of a plant) using the spectral signatures acquired from the imagery and interpolating the surface attribute to areas not directly measured. We apply these methods to map Arctic and boreal vegetation into different classes in imagery in central Alaska.

2. Methods

2.1. Field methods

2.1.1. Site, plot and transect Layout

Site locations were in the Fairbanks North Star and Denali Boroughs in central Alaska. Each site was selected to be located within (planned or

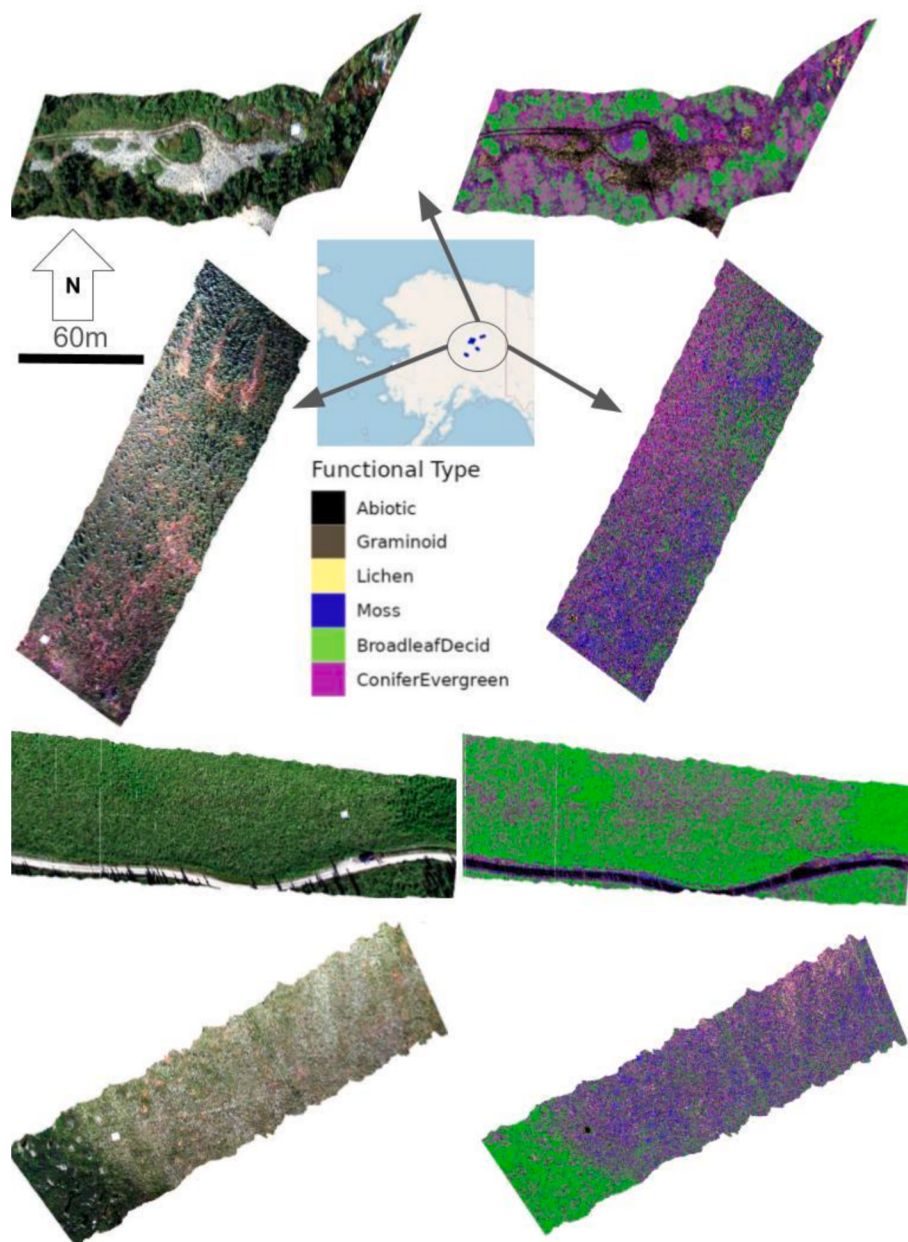


Fig. 1. Study area map of Alaska showing UAS collections with RGB images showing ground reference sampling locations, RGB images and predicted PFT maps for each site. The coordinate system is WGS84 UTM Zone 9.

recent) AVIRIS-NG (“AVIRIS”) flight lines which were accessible to ground crews, some of which were also known research areas. Site locations were sampled as close in time to AVIRIS overflights as possible for future use in ground validation. Sites included long term ecological research (LTER) areas and other sites in central Alaska, accessed in day trips from Fairbanks representing Arctic or alpine tundra to boreal forests, including Bison Gulch (adjacent Denali National Park and not under AVIRIS), Eight Mile Lake area (north of Healy, Alaska) near other teams long term research plots, Bonanza Creek Experimental Forest (southwest of Fairbanks) and Caribou-Poker Experimental Forest (northeast of Fairbanks, which we call Chatanika in this study because that is the nearby village).

At each site, we sampled a plot consisting of a ground calibration transect 100 m long delimited by a tape with 1 m white plastic pipes for ground control laid perpendicular to the tape at every 10 m mark (Fig. 1). Transects were laid out along the contour of any hillslope so that there was little elevation change along the transect but were otherwise randomly constrained by access (within a few hundred meters of road access). Starting at 0 m and at each 10 m increment along the transect, a 1 m² quadrat was used to constrain the ocular estimation of vegetation cover by species for both vascular and non-vascular species. See Figs. 1 and 6 for a map of locations and detailed view of the ground validation at one site.

2.1.2. In Situ vegetation cover Estimates

For each quadrat along the transect at each site, ground cover by taxa or abiotic cover class was estimated by a single, expert observer using the ground photos. Cover was estimated as visible from above (“top cover”) and therefore summed to 100 % for each quadrat. PFTs that were observed but less than 1 % of the image were counted as not present. Cover was estimated at the most precise taxonomic level possible, typically to species or genus. Abiotic surfaces were estimated in the images as litter, wood, rock, gravel, or bare soil but were lumped as “Abiotic” in the modeling and mapping. Cover estimates at the high precision (e.g. species or genus level) were aggregated to a single PFT level for a standard set of categories for all vegetation and abiotic surfaces in images.

2.1.3. UAS imaging spectroscopy

Over each site, airborne images were collected with a Headwall Micro A- series push-broom imaging spectrometer (Headwall Photonics, Bolton, MA, USA) covering the VNIR range (324 spectral bands from 400–1000 nm with 1004 spatial bands/pixel) hard-mounted on UAS (DJI Matrice 600). Geometric position of each image was recorded on-board with an APX-15 GPS/IMU without any post-processing due to the lack of base stations nearby and no satellite correction available at that time (e.g. PP-RTX). Expected accuracy of location and orientation with this configuration was 1–3 m. Our imagery had pixels ~ 4 cm ground surface distance (GSD) at 50 m flight above ground level (AGL).

All images were collected using the same process for radiometric calibration. Right before each flight, we collected a white reference image of a Spectralon panel covering the entire FOV under ambient light, typically by tilting the UAS to avoid shadows, which was used to assess illumination conditions and set flight speed. We also selected pixels from a 55 % reflectance tarp as white reference for radiometric calibration (Fig. 1). We used this tarp reflectance for converting radiance to reflectance in sensor manufacturer’s software that uses the empirical line approach compared to the white reference (SpectralView v.1.3, Headwall Photonics, Bolton, MA, USA). Images were orthorectified to WGS 84 UTM Zone 6 (EPSG code 32606) using Alaska 2 Arc-second Digital Elevation Models (DEMs) from the USGS.

Images were collected late July–early Aug in 2018–2019, primarily under AVIRIS footprints, and as close in time to AVIRIS flights as possible. Each day’s flights were conducted as close to solar noon as possible although a couple flights were much later (Table 1). These missions resulted in 750 GB of orthorectified reflectance across 12 sites,

Table 1

Site name, date, conditions of each flight and number of ground validation quadrats.

Site name	Date and time (UTC-8)	Atmospheric Conditions	Ground wetness	# validation quadrats
Bison Gulch (Healy)	8/12/2019, 12:07:28–12:17:41	Clear with a small area of light scattered clouds on the horizon	Dry	9
Eight Mile (Healy)	7/28/2018, 14:37:17–14:45:04	Partly cloudy with breaks of full illumination	Dry	11
Chatanika (Caribou-Poker)	7/29/2018, 11:32:59–11:47:40	Clear	Dry	11
Bonanza Creek	7/25/2018, 11:54:03–14:12:21	Scattered high cirrus	Moist	11

each with 5–20 image strips, from ABoVE domain under AVIRIS will soon be available from the Oak Ridge National Lab (ORNL) Distributed Active Archive Center (DAAC). Images used here are a subset of this full dataset.

2.2. Processing methods

2.2.1. Image Acquisition and processing

Orthorectification of ground control was performed by iteratively adjusting flight altitude for each flight to match ground features (shorelines, vegetation patches) in Google Earth imagery to match expected Ground Surface Distance (GSD). Radiometric correction from digital number (DN) to radiance to reflectance followed manufacturer’s recommendation. DN were first converted to radiance based on a dark reference measurement (lens cap on) prior to flight. Radiance to reflectance conversion followed the empirical line method based using a single pixel from a 55 % reflectance gray scale tarp in an image (Fig. 1) in the SpectralView software (Headwall Photonics, Bolton, MA). Gray reflectance tarp values were close to expected values (Supplemental Fig. 1), varying largely due to the tarp not being flat due to vegetation and microtopography.

2.2.2. Image-based PFT spectral library

Ground photos of validation quadrats were georeferenced to the UAS VNIR imagery in ArcMap 10.8 (ESRI Redlands, CA) and exported as images. Patches of PFTs and the extent of each quadrat were digitized in ENVI geospatial software (ENVI 5.5.1 Harris Geospatial Solutions, Inc., 2018) using the georeferenced ground photos and coincident higher resolution RGB imagery from UAS (ca. 2 cm resolution) when available. PFTs located in images were hand digitized based on ground reference photos or obvious characteristics (e.g. gravel road, large conifer trees), primarily outside quadrats used for validation, although some patches were digitized inside quadrats. At each site and quadrat, pure patches of a PFTs were in a georeferenced ground photo. The minimum patch size was 3 pixels for any PFT to be digitized. PFTs were digitized from images across as many of the sites as possible to account for known differences in reflectance and environmental conditions among locations and images. Reflectance by PFT is summarized in Fig. 2 showing the variability within PFTs and in Fig. 3 showing how the median reflectance differed between PFTs.

2.2.3. Spectral resampling and vegetation indices

Ninety (90) narrow-band vegetation indices (VIs) were calculated using the hsdar package (Lehnert et al., 2019) in R (R Core Team, 2023). The VIs used range from Chlorophyll to carotenoids to soil indicators and were chosen as an efficient means to account for differences in the magnitude of the reflectance between the images while preserving shape

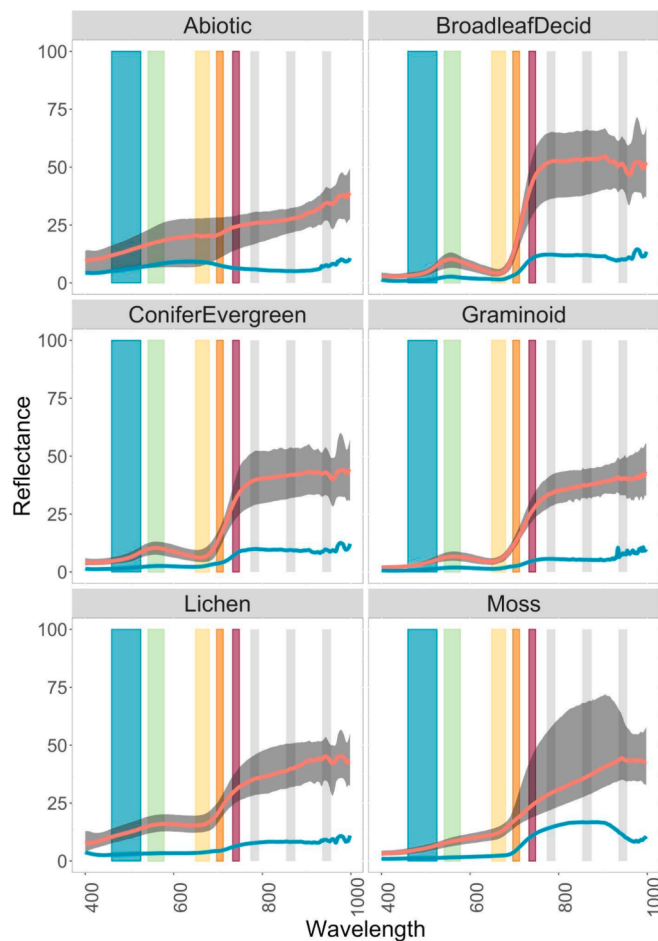


Fig. 2. Reflectance (5 nm resampling) by plant functional type (PFT). Red lines are the median reflectance, gray 90 % quantiles and blue line showing Standard Deviation based on 17,201 pixels from 193 PFT patches hand digitized in UAS VNIR imagery. For reference only, Sentinel-2 band passes are depicted behind the spectra. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

and relative magnitudes.

After VI calculation, we tested the influence of spectral resolution on the classification of PFTs by resampling the spectral library to 5, 10, 25 and 50 nm (last bandpass not used in PLS since it would have too few bands to warrant use of PLS) to reduce noise in the image spectra. Resampling was performed using cubic spline interpolation in the spectrolab package (Meireles et al., 2017). The 90 VIs (based on the raw data at 1 nm resolution) and bands resampled at different widths were the independent variables (“predictors”) in our models of PFT occurrence below.

2.2.4. Balancing of the sampling between PFTs, patches and sites

Bias from PFTs with many patches was reduced by building balanced datasets. We randomly sampled pixels one patch at a time until the specified sample size was reached, both for the training and testing splits. We set the testing sample size to 20 pixels per PFT and used a gradient of sample sizes for training. The fully balanced train/test split had 125 pixels per PFT. We also tested a range of sample sizes up to 2000 pixels, with the majority of PFTs having far fewer samples. This gradient represented the tradeoff between coverage of large PFTs with bias associated with differing sample sizes by PFT. Most sites had pixels for most PFTs, which ensured the data for each PFT has adequate representation of each site to prevent site-based and flight-based differences confounding the analysis (Table 2). After balancing the number of pixels

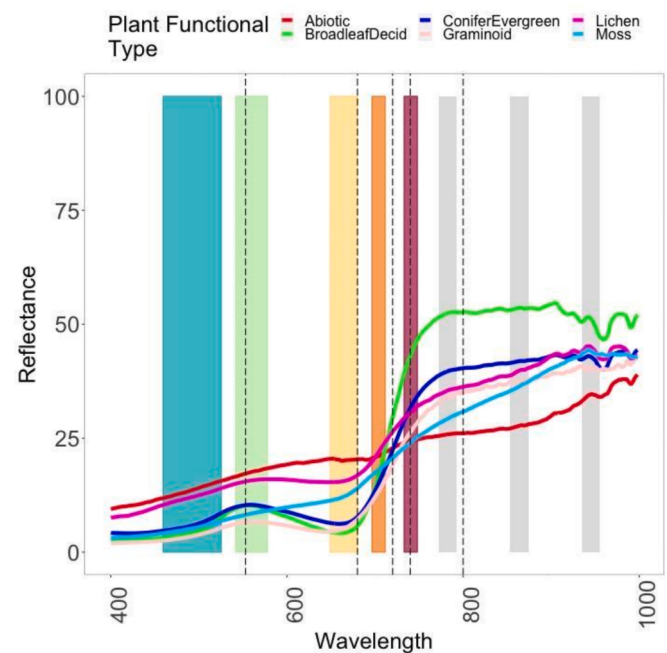


Fig. 3. Median reflectance (5 nm resampling) by plant functional type (PFT) based on 17,201 pixels from 193 PFT patches hand digitized in UAS VNIR imagery. Sentinel-2 band passes are depicted behind the spectra. Vertical dashed lines represent bands used in calculating the most important vegetation indices in the best Adaboost model.

by PFT across sites, we tested for differences between PFTs, sites and PFT by site interaction using PerMANOVA (Anderson, 2001) using a Euclidean distance and 1000 iterations.

2.3. Training and implementation of machine learning

2.3.1. Model experimental conditions and image to image transferability

Models were trained with a variety of preprocessing, feature extraction methods and hyperparameter tuning (Supplemental Table 1). We performed a grid search of preprocessing methods and features utilizing a containerized version of lecospec running R 4.3.0 on Ubuntu 22.04 on a high-performance cluster at University Maine’s Advanced Computing Group. AdaBoost model training in particular took significant amounts of time.

2.3.2. Bagged regression trees using Ranger and Adaboost and linear models PLS-LDA

We used four types of classification models built in the modeling framework caret (Kuhn 2008):

- 1) Random forest (RF) as implemented by Ranger (Wright and Ziegler, 2017). We trained RF models with 2, 4, 8, 16, 32, 64, 128, 256 and 512 trees. Exploratory analysis on this data with RF models suggested that adding max depth reduced accuracy, and alpha did not have a significant effect on model performance. While max depth and alpha typically reduce model overfitting, we noticed no associated increase in test set performance or quadrat-based validation R^2 across levels of alpha and max depth. These therefore were not included in the final hyperparameter tuning step.
- 2) Adaptive Boosting (Adaboost) (Alfaro et al., 2013). For each Adaboost model, we varied the number of weak learners (Classification and regression Trees/CART) from two to 1024 and the maximum depth of each learner from five to 30 splits.
- 3) Partial Least Square Linear Discriminant Analysis (PLS-LDA) (Liland et al., 2016) across gradients of the wavelength and sample size per

Table 2

Full sample sizes of pixels by site by PFT across 197 patches across four sites in central Alaska.

Site	Abiotic	Broadleaf Acid	ConiferEvergreen	Lichen	Graminoid	Moss
Bison Gulch	466	1085	2397	1482	0	0
Bonanza	448	972	1824	787	182	697
Chatanika	518	940	1937	0	116	321
Eight Mile	14	366	672	1039	34	904

PFT. The number of components varied from four to 40, increasing by two each time. Models were trained with bandwidths of 5, 10 and 25 nm and no vegetation indices, since PLS assumes strong inter-correlation between columns and VIs are far less intercorrelated than reflectance.

- 4) Deep Learning (DL): 1-dimensional ResNet (He et al, 2016) and Jigsaw (Moraga & Duzgun, 2022), with bands only, vegetation indices and both.

2.3.3. Deep learning methods

Some deep learning models were also developed using these data and processing systems. The training data was kept similar between the tree-based and deep learning based models, utilizing the same pixel-based approach (eg. 1-dimensional). This required selecting hyperparameters for the models using only a single pixel and providing them with no spatial information. We tested traditional linear feed-forward models, Residual Connection Convolutional Neural Networks (ResNet) and the Jigsaw geospatial model identified in Moraga & Duzgun (2022). All were trained with the Adam (Kingma, 2014) optimizer for a maximum of 2000 epochs using cross entropy loss. All were trained with the Torch deep learning framework using its R language binding.

2.3.4. Accuracy assessment

To assess model predictive accuracy for pure PFT patches, we randomly sampled 20 pixels for each PFT from the 193 high confidence pure patches digitized in the imagery. We compared the classification accuracy of this independent validation data across models to inform model selection and better understand model performance (hereafter “accuracy”). To assess overall model fit for mixed PFT patches, we aggregated cover estimates by PFT for each quadrant at each site. We then computed the correlation between the ML predicted proportion of pixels by PFT to the ground validation proportion by PFT to estimate the proportion of explained variance (R^2). We also compared the Relative Percent Difference (RPD) between the model estimates and the ground truth values, as used in Orndahl et al., (2022). Overall model performance was assessed using the three above metrics (accuracy, R^2 , and RPD), in which the best models had the highest R^2 , highest accuracy, and the lowest RPD. After selecting a few well-performing models, we created large area ground cover maps from a single, whole image per site (Fig. 1). These predictions were then compared to the original images and high-resolution aerial photographs to assess the map quality.

2.3.5. Variable importance

The best models were examined for variable importance. For RF, we retrained models with more iterations and correcting for node impurity, after which ranked predictors on the corrected gini impurity. For Ada-boost, the most important predictors were averaged CART variable importance scores across all the trees in the model. The 12 most important variables for both models are listed with their respective importance metrics in Supplemental Table 2.

Another important consideration in the variable importance is the level of intercorrelation between variables. Variable inter-correlation was assessed using pairwise Pearson correlation coefficients. We compared models with the full predictor set to models with a reduced predictor matrix, removing predictors at $r = 0.99$. We reduced the predictor matrix sequentially starting with the variables correlated with the most important RF predictors. We chose this approach because it

retained the most important predictors.

3. Results

3.1. Image-based PFT spectra

Pure patches ($n = 193$) of positively identified patches of different vegetation at various levels of taxonomic specificity were distinct enough to be digitized in the USA images based on the ground photos. The sample size in the image spectral library was 17,201 pixels. Graminoid had the fewest pure pixels ($n = 332$ pixels) while conifer trees had the most ($n = 6830$ pixels), which we grouped as a single class called ConiferEvergreen. No patches were large or distinguishable enough to be digitized. PFTs were statistically different based on a balanced dataset of 125 pixels per PFT at 1 nm wide bands accounting for differences in sites ($F=6.94$, $df = 5$, $p < 0.001$) based on PerMANOVA. The interaction between site and PFT was weak and non-significant ($F=0.96$, $df=12$, $p = 0.43$).

3.2. Exploratory analysis

In order to better understand the data and PFTs therein, principal component analysis (PCA) was used to examine the data (Fig. 4). Abiotic and biotic classes appear separable based on the PCA results, with the two classes forming distinct groups. Lichens appear between abiotic and green plant clusters. Broadleaf, deciduous plants are largely superimposed on Axes 1 and 2 in the PCA, showing strong spectral similarity of these groups relative to the other classes.

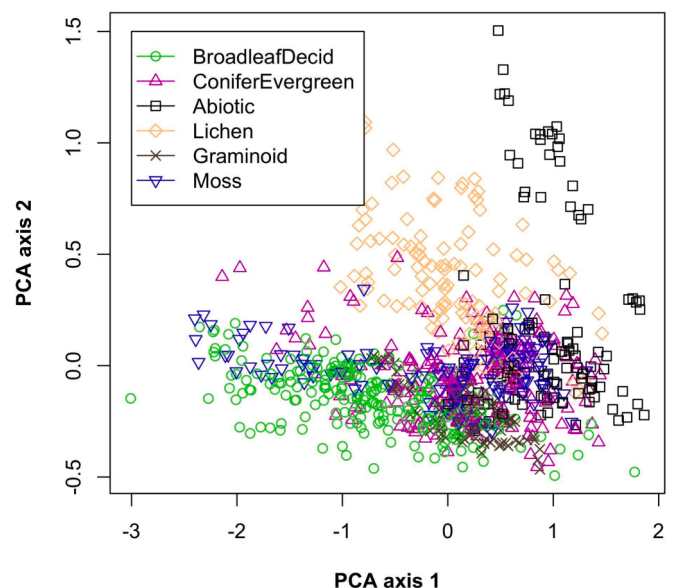


Fig. 4. Principal Components Analysis (PCA) biplot using image spectra showing separation of PFTs using a balanced sample across sites and PFTs (125 pixels per PFT) and 5 nm wide bands. Axes 1 explains 78.7 % of the variance and Axis 2 9.5 %.

3.3. Validation cover estimation by PFT

We used 42 quadrats (1x1 m) across 4 sites (Bonanza Cr, Bison Gulch, Chatanika, Eight Mile) in central Alaska as our ground validation set (Fig. 5, Table 1). Cover estimation by PFT in the ground photos presented difficulties due to the spatial distribution of patches in each image. PFTs like graminoids or litter could often occur scattered individuals. Some PFTs were difficult to separate from each other, such as gravel (abiotic) that is covered with a fine layer of lichen.

3.4. Image-based ML predictions

3.4.1. Overall model results summary

Model accuracy estimating PFTs in the quadrats (Fig. 5), PFT patches and features in landscape broadly aligned with the known surfaces present at each location (Fig. 6). At the quadrat level, individual pure patches of PFTs were often visible both in the ground validation photos and the predicted PFT images (Fig. 5). However, it was also clear from the ground photos that many of the pixels in the quadrats were mixtures in the UAS imagery. At the landscape level, model predictions showed recognizable features, such as a calibration tarp in each image (Fig. 1) correctly classified as “Abiotic” in most of the sites.

Model confusion/error was high for Graminoids across sites, especially Bonanza, likely due to a small sample size. Broadleaf deciduous trees and shrubs were underestimated at Eight Mile while lichen had the reverse case, with one high estimated cover and low observed cover in one quadrant at Eight Mile (Fig. 1). Moss in one Bonanza quadrat had very high observed cover and low estimated cover, likely from the intermixed overstory above. Conifer evergreens were accurately estimated across sites as was Abiotic (Fig. 6).

The best models from Adaboost and RF were close in their fit (ca. $R^2 = 0.5$) and accuracy (ca. 0.85), while PLS-LDA was lower for both metrics ($R^2 < 0.4$, accuracy < 0.7) (Table 3). Overall, the most important feature selection option in modeling was model complexity (e.g. number of trees or components). Resampling bands to coarser spectral resolution improved RF and Adaboost models although several narrow band models were nearly as good in terms of accuracy and fit (Figs. 7 and 8).

3.4.2. Adaboost model results

Adaboost models with 4 bagged CARTs with a max depth of 20 had the highest R^2 with relatively very high accuracy. This model used 500 pixels max sample size per PFT and 50 nm wide bands. Deeper trees produced better accuracy but lower quadrant R^2 . After 12 trees, adding

more increased model accuracy at the cost of R^2 . Most of the calibration tarp, road surfaces and gravel were regularly correctly estimated to be Abiotic class by Adaboost models compared to RF. Accuracy by PFT for this model is shown in a confusion matrix (Table 4). The validation cover estimates from quadrats was more challenging for the model than the validation pixel set. The fit of the model predicted cover to the observed cover varied across sites and PFTs (Table 5) from near 0 to almost 1 for lichens at a couple sites. The intercept showed strong bias for most PFTs at each site. However, overall pooling the data across sites into a single model resulted in quadrats clustered around the 1:1 observed vs predicted for most PFTs with an intercept of near zero (Fig. 6).

3.4.3. Ranger model results

The best ranger model used 50 nm wide bands, 4 trees and removal of variables at the 0.99 level of Pearson intercorrelation and moderate imbalance ($n = 500$ max pixels per PFT). It explained 52 % of the variation in the ground validation data ($R^2 = 0.52$) across the four sites and was 80 % accurate on the test set of pixels. Several other simple RF models with few trees (< 10) with different combinations of wavelengths (10 and 25 nm) and sample size ($n = 125$ or 2000 max pixels per PFT) were close in overall fit and accuracy ($R^2 > 0.5$ and accuracy > 0.8). The calibration tarp was often partially or completely estimated to be lichen by RF models. Increasing the number of trees increased model accuracy up to a point and then plateaued (Fig. 7).

3.4.4. Partial least Squares Discriminant analysis (PLS-LDA) model results

PLS-LDA image-based models produced a best model ($R^2 = 0.39$) with 65.6 % accuracy using 28 components derived from 10 nm band-width data from 2000 samples max per PFT. The model accuracy peaked at 24 components, with performance on the independent quadrat-based validation peaked at 28 components. Observed vs predicted scatterplots by PFT and site (like Fig. 6 but for PLS) show overfitting for the PFTs with larger sample sizes, with very low fit for lichens and mosses across all sites (results not shown).

3.4.5. Deep learning (DL) model results

Deep learning models were also investigated, but these models were limited in their effectiveness, with the highest observed accuracy of 46 % for the Jigsaw model (Moraga & Duzgun, 2022). This model was far less accurate than the comparable tree-based methods. We believe that this is due to a lack of training data, and the use of these models at the pixel level rather than patch level. This lack of spatial information may significantly weaken the performance of these model types. Future work

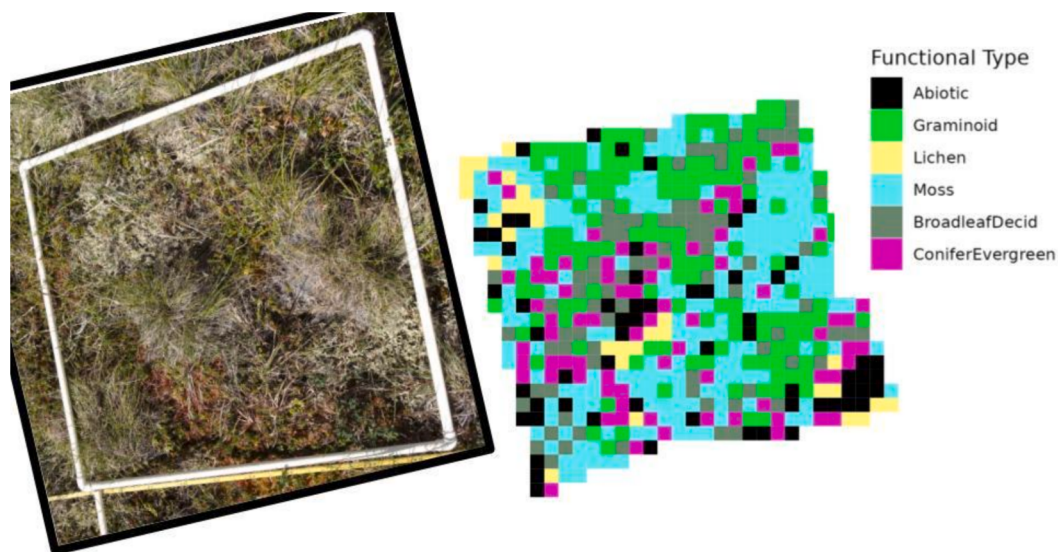


Fig. 5. Ground photo of a quadrat and model estimated cover by plant functional type (PFT) using the UAV-based VNIR image.

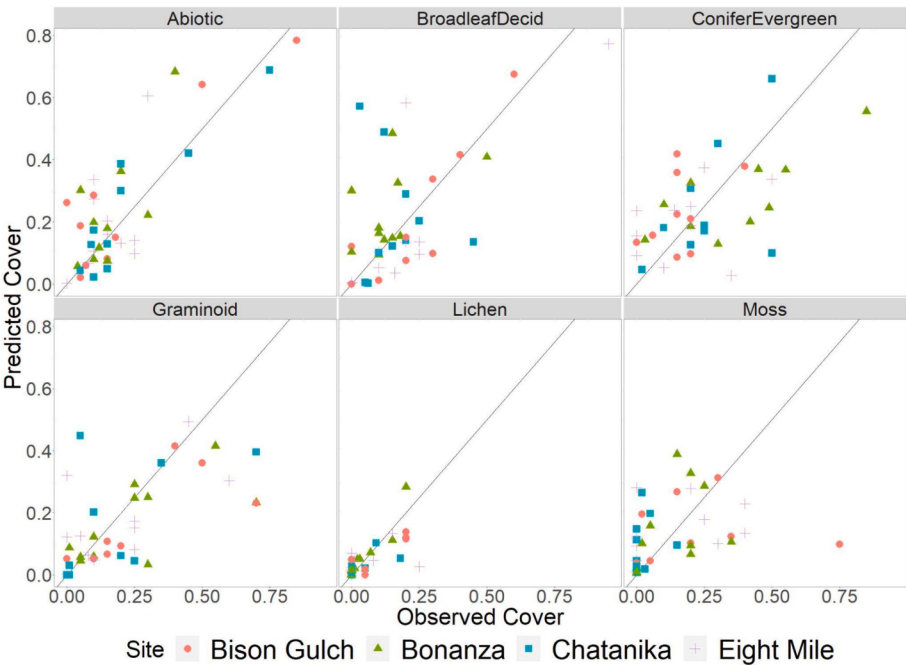


Fig. 6. Observed (x-axis) vs Predicted (y-axis) fractional cover by PFT from best Adaboost model (see Table 3) with dots representing individual quadrats from the four sites. See Table 6 for linear fit of observed vs predicted, slope and intercept by site by.

Table 3
Model results with the best in accuracy and fit with run conditions (best model shown with *).

Model	Accuracy	R ²	Training Conditions
Random Forest	0.8	0.522	n = 4 trees, unbalanced, max depth = None, max variable correlation = 0.99, 500 samples per PFT, bandwidth = 50 nm
AdaBoost (M1)*	0.843	0.518	n = 4, max depth = 20, balanced, max correlation = 0.99, bandwidth = 50 nm, 500 samples
PLS-DA	0.656	0.396	n = 28 components, bandwidth = 10 nm, 2000 samples

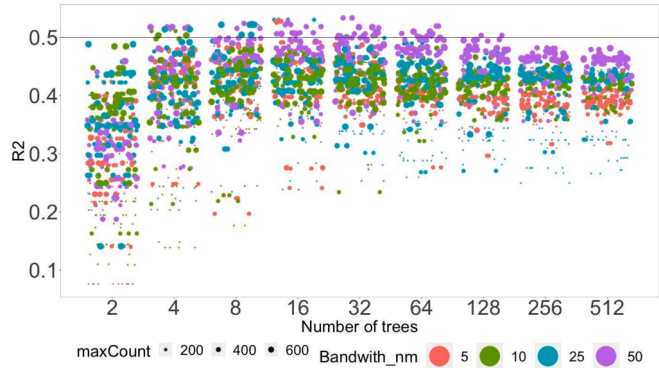


Fig. 7. Model fit (R^2) vs complexity (number of trees) in Adaboost Models jittered along X axis. Horizontal black line shows cut off for acceptable models ($R^2 \geq 0.5$). Circle color varies by bandwidth of reflectance used. Circle size varies by model bias measured as the maximum number of pixels per PFT in the training dataset.

in this area using image patches is needed and outside the scope of this work.

Based on the investigation of deep learning methods herein, the tree-based models outperform the more complex deep learning networks. We

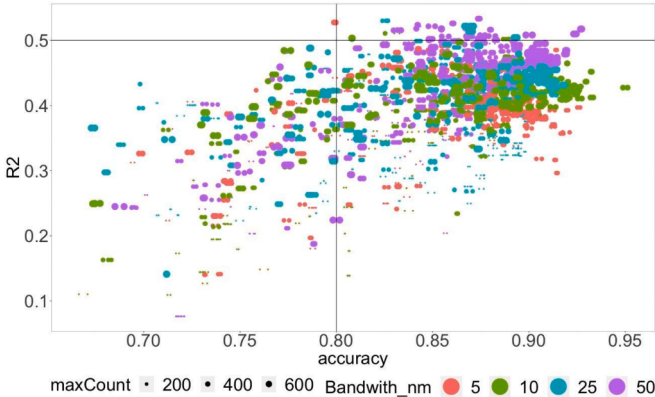


Fig. 8. Model fit (R^2) vs Accuracy for Adaboost Models. Horizontal and vertical lines show cut offs for acceptable models (upper right) at $R^2 = 0.5$ and accuracy = 0.8. Each circle represents a single model run that varies in color by bandwidth of reflectance used. Circle color varies by bandwidth of reflectance used. Circle size varies by model bias measured as the maximum number of pixels per PFT in the training dataset.

believe additional spatial information (eg. 2-dimensional) would better suit the convolutional neural networks, and patch-level training data would potentially enable the performance of the deep learning approaches to surpass tree-based models used here. For comparison to ML, DL model accuracy statistics are in Table 6.

3.5. Variable importance of vegetation indices over bands

The most important VIs in the RF models were almost entirely different from those in the best AdaBoost model (Supplemental Table 2). Adaboost’s five most important variables were Vogelmann (Vogelmann et al., 1993), Pigment-Specific Normalized Difference (PSND, Blackburn 1998), Normalized Difference Vegetation Index 3 (NDVI3, Gandia et al., 2004), 652.5–702.5 nm (50 nm wide band) and transformed chlorophyll absorption ratio index 2 (TCARI2) divided by optimized soil-adjusted

Table 4Confusion matrix from best Adaboost model ($\kappa = 0.807$) tested on an independent validation pixel set.

	Abiotic	BroadleafDecid	Conifer Evergreen	Graminoid	Lichen	Moss
Abiotic	16	0	2	1	2	1
BroadleafDecid	0	35	2	2	0	1
ConiferEvergreen	2	5	33	2	0	0
Graminoid	0	0	1	15	0	0
Lichen	1	0	2	0	18	0
Moss	1	0	0	0	0	18

Table 5Statistics of the relationship between observed and predicted PFT cover by site including fit (R^2 = covariance squared), slope and intercept (bias).

Site	PFT	R^2	Slope	Intercept
Bison Gulch	Abiotic	0.879	0.917	43.883
Bison Gulch	BroadleafDecid	0.784	1.000	-23.119
Bison Gulch	ConiferEvergreen	0.279	0.486	68.026
Bison Gulch	Graminoid	0.572	0.456	20.100
Bison Gulch	Lichen	0.949	0.646	-1.818
Bison Gulch	Moss	0.319	0.485	33.442
Chatanika	Abiotic	0.929	0.896	14.213
Chatanika	BroadleafDecid	0.000	0.002	94.732
Chatanika	ConiferEvergreen	0.076	0.270	88.470
Chatanika	Graminoid	0.215	0.459	44.397
Chatanika	Lichen	0.512	0.409	6.115
Chatanika	Moss	0.030	0.583	50.097
Eight Mile	Abiotic	0.501	1.453	-11.836
Eight Mile	BroadleafDecid	0.850	0.781	7.535
Eight Mile	ConiferEvergreen	0.148	0.381	75.082
Eight Mile	Graminoid	0.336	0.380	46.129
Eight Mile	Lichen	0.243	0.222	13.272
Eight Mile	Moss	0.044	0.137	68.510
Bonanza	Abiotic	0.648	1.248	11.262
Bonanza	BroadleafDecid	0.210	0.474	78.338
Bonanza	ConiferEvergreen	0.517	0.381	63.969
Bonanza	Graminoid	0.305	0.367	40.388
Bonanza	Lichen	0.899	1.166	3.485
Bonanza	Moss	0.270	0.658	36.197

Table 6

Deep Learning model accuracy results by architecture.

Model	Accuracy	Kappa
Jigsaw	0.4375	0.3579
Linear, Bands Only	0.0625	0.0054
Linear, Indices Only	0.1125	0
Linear, Bands + Indices	0.0875	0.0487
ResNet25	0.2625	0.1525

vegetation index 2 (OSAVI2), (TCARI2/OSAVI2, Wu et al., 2008). The five most important variables for RF were Disease–Water Stress Index 4 (DWSI4, Apan et al., 2004), Datt5 (Datt 1999), simple ratio pigment index (SRP)I, normalized pigment chlorophyll ratio index (NPCL, Peñuelas et al., 1994) and Greenness Index (GI, Smith et al., 1995). None of the five most important variables were the same between RF and Adaboost. However, both Photochemical Reflective Index (PRI, Gamon et al., 1992) and Simple Ratio 7 (SR7, Lichtenthaler et al., 1996) were amongst the 10 most important variables for both models. The only wavelength/bandpass combination of reflectance that was among the most important predictors was 650–700 nm (50 nm wide bands).

4. Discussion

4.1. Data Exploration and Interpretation

4.1.1. Spectral patterns

Plant Functional Types (PFTs) VNIR reflectance signatures (Fig. 2) differed from one another based on a PerMANOVA, the separation of

clusters in the PCA (Fig. 4) and when comparing their median reflectance (Fig. 3). These spectral differences between PFTs were consistent across sites, since there was no interaction between site and PFT in the PerMANOVA. We interpreted this to indicate the reflectance measurements across sites could be pooled by PFT into a single, transferrable model rather than separate models by site. Both the reflectance profiles (Figs. 2 and 3) and the PCA (Fig. 4) clearly show abiotic surfaces were the most distinctive from the other classes. Abiotic surfaces are generally darker across all wavelengths than the other PFTs, making it very distinctive (Figs. 2 and 3). Most of the pixels for the training data came from gravel roads or logs, as those were the larger patches positively identified in the whole-image PFT predictions (Fig. 1). Smaller Abiotic patches were missing from our training data, such as mineral or organic materials like leaf litter. Our sample of Abiotic reflectance varied evenly across all wavelengths, which also differed from all the other PFTs, indicating we selected homogenous Abiotic surfaces. The best model also correctly estimated the tarp to be primarily Abiotic pixels in most of the four sites (Fig. 1), despite there being no tarp pixels in the model.

Lichens (and mosses) were the next most distinctive PFTs based on the PCA and reflectance profiles and had the overall lowest misclassification rate (Table 4), with generally darker reflectance than vascular plants in the near infrared. Lichens and mosses vary their photosynthetic activity based on environmental conditions since they do not actively control their hydration state. This large temporal and spatial variation in water content variation leads to large differences in reflectance (Salko et al., 2023; Turner et al., 2023). Therefore, it is not surprising their chlorophyll absorption features, which are partially controlled by a plant's hydration state, are different from vascular plants. Lichens (and mosses) also lacked the distinctive absorption in the yellow before the steep red edge in all the vascular plant PFTs. Lichens, and to a lesser degree mosses, have different physical and chemical components than vascular plants, with many unique chemicals which are involved in photoprotection and therefore optically active.

While our sample of lichen reflectance is within the range of lichen reflectance published (Rees et al., 2004), our sample lacked many darker species (e.g. *Cetraria* spp.) because large patches didn't occur in our imagery. Compared to lichens, mosses were much more variable in the NIR region, likely due to their similarity in physical and photochemical composition to vascular plants (Nelson et al., 2022). Mosses in our sample were largely *Sphagnum* sp., which are broadly similar to vascular plants in that their cells are primarily made of cellulose whereas lichens are primarily made of chitin (Honegger and Bartnicki-Garcia 1991). However, both mosses and lichens absorb and lose water rapidly depending on environmental conditions, which in turn drives strong changes in reflectance based on hydration state (Turner et al., 2023).

All vascular plants (BroadleafDecid, ConiferEvergreen, Graminoids) were somewhat similar in their reflectance profiles, with low variability in the visible and more variability in the near infrared. This manifested as those PFT classes largely superimposed points in the PCA. ConiferEvergreen and BroadleafDecid were more different from each other than Graminoids, which appeared intermediate between the non-vascular plants (lichens and mosses) and vascular plants. ConiferEvergreen was primarily made of *Picea* sp. (with some *Larix laricina*) and a few patches of shrubs, such as *Ledum* spp.. All these species form finely dissected or thin canopies with likely large BRDF and have overall

darker pigmentation. In contrast, BroadleafDecid included wide-leaved species, such as *Populus* sp., *Salix* sp. and *Alnus* spp.. All these fall into the stereotypical “green plant” reflectance spectrum, with the obvious green peak, steep red edge and relatively large infrared plateau (Fig. 2). Graminoids are likely different from other vascular vegetation, in part, because grass leaves are very narrow and the pixels that were selected contained whatever was on the ground surface (often litter, lichen, moss). Another reason Graminoids were different may be the standing dead leaves of graminoids (Abiotic) observed in many patches, which would dampen the reflectance of green grass leaves.

4.1.2. Spatial patterns

Despite many quadrat pixels being mixtures, many PFT patches were correctly predicted when the patch was greater than a couple mostly pure pixels and not over topped by a different PFT and spectrally distinctive to adjacent patches (contrast). For example, lichen and moss patches were often easily detected when separated from adjacent vascular plants. The opposite also occurred where the spatial pattern predicted in the PFT maps did not seem to resemble the ground pattern, despite the overall cover per PFT for that quadrant being decently accurate.

PFT patterns in the whole images showed large agreement with what can be seen in just RGB images. Features that were clearly correctly estimated included ConiferEvergreen canopies, Abiotic areas (roads and the calibration tarp) and Graminoid, Lichen and Moss patches. Since we did not separate conifer trees from evergreen shrubs, it was sometimes difficult to make out conifer tree canopies where there was a large understory component of evergreen shrubs. The same was true of broadleaf trees and deciduous shrubs.

4.1.3. Transferability

While PFTs were strongly different from each other, they were similar between sites based on no statistical support for the interaction between site and PFT in the PerMANOVA. We therefore felt comfortable using our PFT classes as distinguishable in our imagery. Reflectance of gray scale calibration tarps between sites was also similar (Supplementary Fig. 1), which also gave us confidence that images were consistently radiometrically corrected between sites. Most PFTs were present at all sites but two sites lacked one or two PFTs (Table 2), which meant that site-specific models would lack those PFTs. For these reasons, we did not attempt to build separate models for each site because we wanted a transferable model across sites and potentially usable across other sites where we don’t have ground validation data. Pooling data across sites resulted in similar error between sites in estimating the ground validation quadrat PFT cover (Fig. 6).

4.2. Models and model performance

4.2.1. Model selection

Throughout model development and hyperparameter tuning, the models were evaluated on an independent test pixel set (accuracy) and the predictions on quadrat images were compared to the human estimates via Pearson correlation and RPD. In addition to the metrics, the output images (PFT cover) were compared to the ground photographs. However, no single metric appears to adequately represent which model performed best. For example, models with higher accuracy and R^2 sometimes failed to capture features that would be obvious to a human observer (e.g. trees and boulders) while models with slightly lower performance captured spatial patterns better.

Highly accurate models were often biased, significantly overpredicting certain PFTs when applied at larger scales. Models with higher R^2 , while matching the distribution of plants on the landscape, were not necessarily correct about which plants were growing at a particular location. RPD alone, without high accuracy and R^2 , produced maps that failed to capture the spatial patterns, with similar issues to the high R^2 -low accuracy maps. For this reason, the models that produced

the best estimates across the landscape had high accuracy, high R^2 , and low RPD.

After hyperparameter tuning was complete, high-performing models were identified and used to create large area maps. These maps were evaluated to compare how well visually identifiable features (trees, calibration tarp, roads, large patches of moss, etc.) matched the model predictions. Of the best models ($R^2 > 0.5$ and accuracy > 0.8), we saw artifacts in some full image output that dissuaded us from using them (e.g. estimating gravel road or the tarp as lichen). No single model was free of such artifacts, but we selected the best model based on the calibration tarp in the four sites being mostly estimated to be Abiotic, gravel on two sites as Abiotic and conifer tree canopies at two sites mostly called ConiferEvergreen. The best models successfully captured spatial patterns, including roads, the calibration tarp, roads, creek beds, and trees. Furthermore, the models successfully predicted more than 84 % of the test set of pixels correctly.

Based on both the model statistics and qualitative assessment of full image predictions, we identified an AdaBoost Model (see Table 3) as the best performing model based on this; there were models with higher R^2 and accuracy (AdaBoost model with $R^2 = 0.53$, accuracy = 0.87) or lower RPD (AdaBoost, Accuracy = 0.70, rpd = 0.91) which appear to produce output maps of lower quality. However, the highest R^2 values do not coincide with the highest accuracy; rather for very high accuracy models the R^2 is lower than that of models with lower accuracy. One possible explanation for this discrepancy is that models overfit to the pure pixels in training data and perform worse on pixels containing mixtures in the large images. This overfitting theory is corroborated by the higher complexity of the higher- R^2 models compared to the selected model.

The stability of the training process was also investigated. Training many models at different seeds suggests that the model training is relatively stable even for small (4 weak learner) models. The results in Table 7 suggest that while there is non-trivial variance in small model accuracy and r-squared, the higher r-squared of the small models outweighs the slightly increased variance and accuracy tradeoff. It is hypothesized that this tradeoff is due to model under/overfitting, with significant bias introduced to the large models. Ultimately, the landscape maps produced by the large models appear significantly worse than the small models.

4.2.2. Overfitting and the Bias-Variance tradeoff

The independent test of ground cover in the quadrats proved a more difficult challenge (R^2) than accuracy on a reserve test set of pixels, leading us to conclude overfitting was a key problem to avoid. We optimized accuracy and R^2 through our grid search of parameters, which showed the less complex (not overfit) models were better overall. Models were then selected based on the combination of accuracy and R^2 , showing simple (lower number of components/trees) were better at estimating the ground validation while still maintaining high (but not the highest) accuracy. The relationship between accuracy, R^2 and model complexity is shown in Fig. 8. Comparing predictions across quadrat images and large area predictions, we noticed an increase in model bias with the increase in accuracy; for example, ConiferEvergreen was overpredicted across sites in many models.

Wider bands largely improved the RF models, which left the VIs as the primary data in the model. However, there were many models with

Table 7
Accuracy, R^2 and Relative Percent Difference (RPD) model performance statistics for 4 (small) and 1000 (large) tree random forest models across 15 seed iterations.

	Accuracy		R^2		RPD	
	Mean	Variance	Mean	Variance	Mean	Variance
Small model	0.8156	0.0007	0.3831	0.0035	1.0891	0.0028
Large model	0.882	0.0005	0.3296	0.0008	1.3478	0.001

smaller bands that were nearly as good based on the ground validation and test set of pixels. We were surprised that both coarse and narrow band datasets could perform similarly, at least in the company of VIs. We interpreted this to mean balancing different forms of error, such as improved prediction of one PFT or overall PFT prediction site at the cost of another PFT or site.

We expected more data (larger sample sizes) to bias models towards abundant PFTs and reduce fit and accuracy. Our results show low sample sizes (125–500 samples per PFT) balancing across PFTs were able to do as well as models with more data. The bias associated with larger sample sizes per PFT again presented a challenge and led to measures to reduce this bias by balancing class frequencies within the training data, and the use of class weights in models that support it. Importantly, the two related model types, AdaBoost and RF, performed best with similar hyperparameters. Both reached peak performance with 4 weak learners, 500 samples per class at most, 50 nm bandwidth, and the removal of the most intercorrelated features. These results suggest good consistency in training results.

4.2.3. Images artifacts among the best models guided selection

After selecting a few candidate models, predictions were made for the complete datacubes (entire images) and visually assessed in comparison to the original images in RGB. Of the best models ($R^2 > 0.5$ and accuracy > 0.8), we saw artifacts in some full image output that dissuaded us from using them. Example model artifacts included incorrectly estimating the calibration tarp and gravel road surface as lichen or conifer tree shadows called a different PFT than the illuminated conifer canopy. No single model was free of such artifacts but we selected the best model based on the calibration tarp in the four sites being mostly estimated to as Abiotic, gravel on two sites as Abiotic and conifer tree canopies at two sites mostly called ConiferEvergreen.

4.2.4. Variable importance

Vegetation indices were by far the most important predictors of PFTs. So, one may ask “Are vegetation indices useful in the Arctic?”, which is the title of a paper written over 25 years ago (Rees et al., 1998). While that paper was about NDVI, we can say “yes”, vegetation indices were broadly effective means in our study to reduce the dimensionality while preserving predictive power in our classification models of PFTs. Certain areas of the electromagnetic spectrum were included in many of the indices that were most important in the RF and AdaBoost models. Both the models use similar parts of the electromagnetic spectrum, utilizing bands 800, 680/682 and 550/553. They also use a band in the visible range, either 430 or 470. Overall, while combined differently to create different band ratios, they utilize similar information.

While our goal was not to understand why PFTs were different spectrally, we can look at the importance of VIs and reflectance bands selected in the models to start. AdaBoost had one very important band, Vogelmann (740/720) (Vogelmann et al., 1993), followed by PSND, (800–680/800 + 680) (Blackburn 1998) and NDVI3, (682–553/682 + 553) (Gandia et al., 2004), which were half as important. While interpreting the magnitude of importance should be done with caution, the relative importance of Vogelmann is indicative that the red edge holds a lot of information for separating PFTs (Fig. 3). In contrast, variable importance in RF showed a gradual decline from most to least important. DWSI4, a very important index in the RF models, included bands outside the bandpass of the instrument. Instead, this infrared band was approximated from the available data by the hsdar R package using the nearest bands. Datt5 was nearly as important in the RF models, which is calculated as 672/550. Taken together, data reduction by VIs seems a simple, effective means to separate PFTs despite that not being their purpose.

4.3. PFTs spectra are transferable across sites

We aimed to make a transferable model for classifying PFTs across

sites. Our initial observations of reflectance of PFTs showed differences that were consistent across sites (Fig. 2), indicating pooling spectra across sites may be appropriate. PFTs were strongly different from each other but when site was included in our test, there was no interaction between site and PFT. We therefore felt comfortable using our PFT classes as distinguishable in our imagery.

Overall separability of the PFTs varied widely, as some PFTs were very similar and overlapping in reflectance signatures, at least when a linear model of data reduction is used (Fig. 3). Fortunately, PFTs were more separable when non-linear methods, such as RF and Adaboost, were used to recover relationships between the large number of predictors with relatively simple models.

4.4. Sources of error

4.4.1. Ubiquitous mixed pixels

Pixels at 4 cm GSD from the UAS image clearly are mixtures of many different PFTs when compared to the ground photos used for estimating validation PFT cover. This effect is potentially compounded by the fact that the models were trained only with data from pure pixels — that is, data that was digitized by a human ecologist with high confidence that the pixel contains a single, known PFT. This leads to a so-called “training/serving skew” in the data when creating predictions across the landscape. In other words, the distribution of features in the data from quadrats or large site images may be significantly different than the training data. Based on the high-resolution ground photography, many pixels were clearly mixtures even at a 4 cm resolution.

4.4.2. Geometric and radiometric error between image and ground

Inaccuracies in global positioning systems and coordinate reference systems introduce additional error, such as the difference between the DEM used in orthorectification and GPS/IMU on the UAS. Many sources of radiometric noise also still exist in our data, such as atmospheric effects between and within flights/sites, bidirectional reflectance (BRDF) both at the landscape (hillslope) and plant canopy-level. These corrections are commonly applied to imaging spectrometer data, especially when using physical based models to retrieve plant traits. The instrument used for this study also has a variable signal/noise ratio, where there is more noise especially at the longer wavelengths.

4.4.3. Observer error in validation data

Validation cover estimation has observer error due to ocular cover estimation. More precise methods like validation per pixel but these are labor intensive and require highly precise co-registration of UAS and quadrat photos. The use of multiple observers with a set of iteratively refined estimates across observers may increase accuracy, mitigate human error, and allow for variance estimates. Others have used AI to build validation datasets from ground photos (Lovitt et al., 2022), which seems very promising for both increasing accuracy, decreasing observer bias and enabling rapid generation of labeled data. We opted for a single, expert observer who had estimated ground cover of Arctic and boreal plants in central Alaska in hundreds of quadrats the same style and size (1 m²) used in this study. Nonetheless, the precision of any expert observer has bias and error, which we did not measure.

4.4.4. Systematic sampling error

Sampling methodologies in this sample were not purely random. It was necessary to select training pixels that corresponded well with the high-resolution aerial photographs and were recognizable by sight to create labeled training data. This systematic, rather than random, selection of the included pixels for training may lead to bias in the data, resulting in a sampling distribution not representative of the population of pixels by PFT in an image. However, pure patches of any PFT were rare except for the largest size individuals, such as trees or large shrubs. In many cases, pixels were mixtures even at 4 cm resolution thus the pool of patches we could sample for training data may be much smaller

than at first glance. Transects for the quadrat images were also selected systematically, based on road access and then fixed at 10 m spacing once a random azimuth was selected for the transect. This potentially creates further bias in the data as well, since these 100 m long transects only sampled a very small fraction and spatial extent of each image. Several validation quadrat transects randomly per image would be much more representative of the large image footprint.

5. Conclusions

The overall goal of the analysis and modeling is to provide high quality maps of PFTs in different sites in Arctic and boreal biomes, which we demonstrated was possible within an acceptable level of error. To do this, we generated spectra of different PFTs to test whether they were statistically different based on their VNIR reflectance spectra. Median reflectance was indeed distinctly different between PFTs while within a PFT, reflectance was consistent in four sites across central Alaska, USA. Our classification models overfit when even moderately complex, reducing performance for the independent validation (R^2) but improves the accuracy on the test set of pure pixels. For these reasons, we selected a simple model that could separate test pixels and have high validation accuracy with independent ground cover estimation.

We demonstrated transferable approaches to PFT mapping across different images, sites, dates, and mixtures of vegetation by pooling pixels with known labels collected across VNIR images. While the model is transferable across our four sites, we also observed strong algorithmic differences in classification fit and accuracy. The best models, while similar statistically, varied in their prediction of some large features (e.g. Abiotic tarp estimated to be lichen). All models showed the importance of vegetation indices in reducing the dimensionality of the reflectance spectra.

Future studies should explore sampling vegetation using points on pure patches located opportunistically rather than quadrats arrayed in a rigid sampling. While this would change the statistical sample of the validation, we expect a better transfer of image-based models with this approach for generating training spectra from images by target (e.g. PFT). Even with the opportunistic sampling, the patch size of many PFTs is often not large enough for easy geolocation.

Our PFT maps from these UAS images provide ground validation for AVIRIS-NG imagery that was flown near the same for most of our flights. We aim to expand our PFT mapping to the entire archive of AVIRIS-NG imagery in the Arctic and boreal biomes of Alaska and northwestern Canada. To do this, we hope to partner with other teams of scientists who have ground validation of PFT cover under AVIRIS to extend our knowledge of PFT reflectance spectra and associate spatial patterns we can observe and interpret with imaging spectroscopy.

5.1. Data and code availability

The R code and the spectral library of training pixels by PFT available to the public for image classification of any hyperspectral imagery based on a spectral library at <https://github.com/nelsopet/lecospec>. Full bandpass image data is available to the public at ORNL DAAC (<https://doi.org/10.3334/ORNLDAAAC/1980>).

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.

Data availability

The data and code are available through Github and the Oakridge DAAC.

Acknowledgments

This work was made possible by multiple funding sources, including NASA Arctic-Boreal Vulnerability Experiment (ABOVE) Phases 1-3 (NNX17AE44G, 80NSSC19M0112, 80NSSC22K1247), Maine Economic Improvement Fund-Small Campus Initiative (awards in 2016 and 2018), NSF EPSCoR Track 2 INSPIRES (#OIA-1920908), Maine Space Grant Consortium EPSCoR Research Infrastructure Development and University of Maine Advanced Computing Group. David Paradis assisted with some flights. Dawson White, Lance Stasinski, and Duke (George) Brady assisted with ground data collection and Katie Orndahl provided useful comments on earlier drafts.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jag.2024.104156>.

References

- Alfaro, E., Gamez, M., Garcia, N., 2013. adabag: An R package for classification with boosting and bagging. *J. Stat. Softw.* 54, 1–35.
- Anderson, M.J., 2001. A new method for non-parametric multivariate analysis of variance. *Austral Ecol.* 26 (1), 32–46.
- Apan, A., Held, A., Phinn, S., Markley, J., 2004. Detecting sugarcane ‘orange rust’ disease using EO-1 Hyperion hyperspectral imagery. *Int. J. Remote Sens.* 25 (2), 489–498. <https://doi.org/10.1080/01431160310001618031>.
- Blackburn, G.A., 1998. Quantifying chlorophylls and carotenoids at leaf and canopy scales: An evaluation of some hyperspectral approaches. *Remote Sens. Environ.* 66 (3), 273–285.
- Datt, B., 1999. Remote sensing of water content in Eucalyptus leaves. *Aust. J. Bot.* 47 (6), 909–923.
- Gamon, J.A., Penuelas, J., Field, C.B., 1992. A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency. *Remote Sens. Environ.* 41 (1), 35–44.
- Gandia, S., Fernández, G., García, J.C., Moreno, J., 2004. Retrieval of vegetation biophysical variables from CHRIS/PROBA data in the SPARC campaign. *Esa Sp* 578, 40–48.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770–778).
- Honegger, R., Bartnicki-Garcia, S., 1991. Cell wall structure and composition of cultured mycobionts from the lichens *Cladonia macrophylla*, *Cladonia caespiticia*, and *Physcia stellaris* (Lecanorales, Ascomycetes). *Mycol. Res.* 95 (8), 905–914. [https://doi.org/10.1016/S0953-7562\(09\)80085-3](https://doi.org/10.1016/S0953-7562(09)80085-3).
- Hong, D., Zhang, B., Li, X., Li, Y., Li, C., Yao, J., Yokoya, N., Li, H., Ghamisi, P., Jia, X., Plaza, A., Paolo, G., Benediktsson, J. A., & Chanussot, J. (2023). *SpectralGPT: Spectral Foundation Model* (arXiv:2311.07113). arXiv. <https://doi.org/10.48550/arXiv.2311.07113>.
- Kattenborn, T., Fassnacht, F.E., Schmidtlein, S., 2019. Differentiating plant functional types using reflectance: Which traits make the difference? *Remote Sens. Ecol. Conserv.* 5 (1), 5–19. <https://doi.org/10.1002/rse2.86>.
- Kingma, Diederik P. “Adam: A method for stochastic optimization.” *arXiv preprint arXiv:1412.6980* (2014).
- Kodl, G., Streeter, R., Cutler, N., Bolch, T., 2024. Arctic tundra shrubification can obscure increasing levels of soil erosion in NDVI assessments of land cover derived from satellite imagery. *Remote Sens. Environ.* 301, 113935 <https://doi.org/10.1016/j.rse.2023.113935>.
- Kuhn, M., 2008. Building predictive models in R using the caret package. *J. Stat. Softw.* 28, 1–26.
- Lehnert, L.W., Meyer, H., Obermeier, W.A., Silva, B., Regeling, B., Bendix, J., 2019. Hyperspectral Data Analysis in R: The hsdar Package. *J. Stat. Softw.* 89 (12). <https://doi.org/10.18637/jss.v089.i12>.
- Lichtenthaler, H.K., Lang, M., Sowinska, M., Heisel, F., Miehe, J.A., 1996. Detection of vegetation stress via a new high resolution fluorescence imaging system. *J. Plant Physiol.* 148 (5), 599–612.
- Liland, K. H., Mevik, B. H., Wehrens, R., & Hiemstra, P. (2016). Partial least squares and principal component regression. *CRAN*, 66p. Available at [Access Date: 26.01. 2023]: <https://cran.r-project.org/Web/Packages/Pls/Pls.Pdf>.
- Liu, Y., Zhang, R., Lin, C.-F., Zhang, Z., Zhang, R., Shang, K., Zhao, M., Huang, J., Wang, X., Li, Y., Zeng, Y., Zhao, Y.-P., Zhang, J., Xing, D., 2023. Remote sensing of subtropical tree diversity: The underappreciated roles of the practical definition of forest canopy and phenological variation. *Forest Ecosystems* 10, 100122. <https://doi.org/10.1016/j.fecs.2023.100122>.
- Lovitt, J., Richardson, G., Rajaratnam, K., Chen, W., Leblanc, S.G., He, L., Nielsen, S.E., Hillman, A., Schmelzer, I., Arsenault, A., 2022. A New U-Net Based Convolutional Neural Network for Estimating Caribou Lichen Ground Cover from Field-Level RGB Images. *Can. J. Remote. Sens.* 1–24. <https://doi.org/10.1080/07038992.2022.2144179>.

- Macander, M.J., Frost, G.V., Nelson, P.R., Swingle, C.S., 2017. Regional Quantitative Cover Mapping of Tundra Plant Functional Types in Arctic Alaska. *Remote Sens. (basel)* 9(10), Article 10. <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 (5), 054042 <https://doi.org/10.1088/1748-9326/ac6965>.
- Meireles, J.E., Schweiger, A.K., Cavender-Bares, J.M., 2017. spectrolab: Class and Methods for Hyperspectral Data. R Package Version 2. <https://experts.umn.edu/en/publications/spectrolab-class-and-methods-for-hyperspectral-data-r-package-ver>.
- Moraga, J., & Duzgun, H. S. (2022). JigsawHSI: A network for hyperspectral image classification. *arXiv preprint arXiv:2206.02327*.
- Nelson, P.R., Maguire, A.J., Pierrat, Z., Orcutt, E.L., Yang, D., Serbin, S., Frost, G.V., Macander, M.J., Magney, T.S., Thompson, D.R., 2022. Remote sensing of tundra ecosystems using high spectral resolution reflectance: Opportunities and challenges. *J. Geophys. Res. Biogeophys.* 127 (2) e2021JG006697.
- Orndahl, K.M., Ehlers, L.P.W., Herriges, J.D., Pernick, R.E., Hebblewhite, M., Goetz, S.J., 2022. Mapping tundra ecosystem plant functional type cover, height, and aboveground biomass in Alaska and northwest Canada using unmanned aerial vehicles. *Arct. Sci.* 8 (4), 1165–1180. <https://doi.org/10.1139/as-2021-0044>.
- Palmer, M.W., Earls, P.G., Hoagland, B.W., White, P.S., Wohlgemuth, T., 2002. Quantitative Tools for Perfecting Species Lists. *Environmetrics* 13 (2), 121–137. <https://doi.org/10.1002/env.516>.
- Peñuelas, J., Gamon, J.A., Fredeen, A.L., Merino, J., Field, C.B., 1994. Reflectance indices associated with physiological changes in nitrogen-and water-limited sunflower leaves. *Remote Sens. Environ.* 48 (2), 135–146.
- Pinto-Ledezma, J.N., Cavender-Bares, J., 2021. Predicting species distributions and community composition using satellite remote sensing predictors. *Sci. Rep.* 11 (1), 16448.
- [Computer software]. R Foundation for Statistical Computing. <https://www.R-project.org>.
- Räsänen, A., Virtanen, T., 2019. Data and resolution requirements in mapping vegetation in spatially heterogeneous landscapes. *Remote Sens. Environ.* 230, 111207 <https://doi.org/10.1016/j.rse.2019.05.026>.
- Rees, W.G., Golubeva, E.I., Williams, M., 1998. Are vegetation indices useful in the Arctic? *Polar Rec.* 34 (191), 333–336.
- Rees, W.G., Tutubalina, O.V., Golubeva, E.I., 2004. Reflectance spectra of subarctic lichens between 400 and 2400 nm. *Remote Sens. Environ.* 90 (3), 281–292. <https://doi.org/10.1016/j.rse.2003.12.009>.
- Rocchini, D., Boyd, D.S., Féret, J.-B., Foody, G.M., He, K.S., Lausch, A., Nagendra, H., Wegmann, M., Pettorelli, N., 2016. Satellite remote sensing to monitor species diversity: Potential and pitfalls. *Remote Sens. Ecol. Conserv.* 2 (1), 25–36. <https://doi.org/10.1002/rse2.9>.
- Salko, S.-S., Juola, J., Burdun, I., Vasander, H., Rautiainen, M., 2023. Intra- and interspecific variation in spectral properties of dominant Sphagnum moss species in boreal peatlands. *Ecol. Evol.* 13 (6), e10197.
- Schaaf, A.N., Dennison, P.E., Fryer, G.K., Roth, K.L., Roberts, D.A., 2011. Mapping Plant Functional Types at Multiple Spatial Resolutions Using Imaging Spectrometer Data. *GisScience & Remote Sensing* 48 (3), 324–344. <https://doi.org/10.2747/1548-1603.48.3.324>.
- Smith, R.C.G., Adams, J., Stephens, D.J., Hick, P.T., 1995. Forecasting wheat yield in a Mediterranean-type environment from the NOAA satellite. *Aust. J. Agr. Res.* 46 (1), 113–125.
- Thomson, E.R., Spiegel, M.P., Althuisen, I.H.J., Bass, P., Chen, S., Chmurzynski, A., Halbritter, A.H., Henn, J.J., Jónsdóttir, I.S., Klanderud, K., Li, Y., Maitner, B.S., Michaletz, S.T., Niittynen, P., Roos, R.E., Telford, R.J., Enquist, B.J., Vandvik, V., Macias-Fauria, M., Malhi, Y., 2021. Multiscale mapping of plant functional groups and plant traits in the High Arctic using field spectroscopy, UAV imagery and Sentinel-2A data. *Environmental Research Letters* 16 (5), 055006. <https://doi.org/10.1088/1748-9326/abf464>.
- Torresani, M., Rossi, C., Perrone, M., Hauser, L.T., Féret, J.-B., Moudry, V., Simova, P., Foody, G.M., Kacic, P., Feilhauer, H., Malavasi, M., Tognetti, R., Rochinni, D., 2024. Reviewing the Spectral Variation Hypothesis: Twenty Years in the Tumultuous Sea of Biodiversity Estimation by Remote Sensing. *Eco. Inform.* 82, 102702 <https://doi.org/10.1016/j.ecoinf.2024.102702>.
- Turner, D., Cimoli, E., Lucieer, A., Haynes, R. S., Randall, K., Waterman, M. J., Lucieer, V., & Robinson, S. A. (2023). Mapping water content in drying Antarctic moss communities using UAS -borne SWIR imaging spectroscopy. *Remote Sensing in Ecology and Conservation*, rse2.371. <https://doi.org/10.1002/rse2.371>.
- Ustin, S.L., Gamon, J.A., 2010. Remote sensing of plant functional types. *New Phytol.* 186 (4), 795–816. <https://doi.org/10.1111/j.1469-8137.2010.03284.x>.
- Vogelmann, J.E., Rock, B.N., Moss, D.M., 1993. Red edge spectral measurements from sugar maple leaves. *International Journal of Remote Sensing* 14 (8), 1563–1575. <https://doi.org/10.1080/01431169308953986>.
- Wright, M.N., Ziegler, A., 2017. ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *J. Stat. Softw.* 77 (1). <https://doi.org/10.18637/jss.v077.i01>.
- Wu, C., Niu, Z., Tang, Q., Huang, W., 2008. Estimating chlorophyll content from hyperspectral vegetation indices: Modeling and validation. *Agric. for. Meteorol.* 148 (8–9), 1230–1241.
- Xu, Y., Dickson, B.G., Hampton, H.M., Sisk, T.D., Palumbo, J.A., Prather, J.W., 2009. Effects of Mismatches of Scale and Location between Predictor and Response Variables on Forest Structure Mapping. *Photogramm. Eng. Remote Sens.* 75 (3), 313–322. <https://doi.org/10.14358/PERS.75.3.313>.
- ..
- Yang, D., Morrison, B.D., Davidson, K.J., Lamour, J., Li, Q., Nelson, P.R., Hantson, W., Hayes, D.J., Swetnam, T.L., McMahon, A., Anderson, J., Ely, K.S., Rogers, A., Serbin, S.P., 2022. Remote sensing from unoccupied aerial systems: Opportunities to enhance Arctic plant ecology in a changing climate. *J. Ecol.* 110 (12), 2812–2835. <https://doi.org/10.1111/1365-2745.13976>.
- Yokoya, N., Xia, J., & Broni-Bediako, C. (2023). *Submeter-level Land Cover Mapping of Japan* (arXiv:2311.11252). arXiv. <https://doi.org/10.48550/arXiv.2311.11252>.