

## Estimation of fine-scale vegetation distribution information from RPAS-generated imagery and structure to aid restoration monitoring

Rik J.G. Nuijten <sup>a,\*</sup>, Nicholas C. Coops <sup>a</sup>, Dustin Theberge <sup>b</sup>, Cindy E. Prescott <sup>a</sup>

<sup>a</sup> Faculty of Forestry, University of British Columbia, Vancouver, BC, V6T 1Z4, Canada

<sup>b</sup> TC Energy, Calgary, AB, T2P 5H1, Canada



### ARTICLE INFO

#### Keywords:

Forest regeneration  
Ecological restoration  
Plant composition maps  
Boreal forests  
Logistic regression  
Remotely piloted aircraft systems (RPAS)

### ABSTRACT

Detailed maps of vegetation composition are vital for restoration planning, implementation, and monitoring, particularly at early stages of succession. This is usually accomplished through ground surveys, which can be costly and impractical depending on extent and accessibility, or conducted at too broad a spatial scale. In this study, we propose a methodology for mapping regenerating vegetation composition at  $2 \times 2 \text{ m}^2$  spatial resolution, using very high spatial resolution ( $< 1 \text{ m}$ ) remote sensing imagery obtained from remotely piloted aerial systems (RPAS) in conjunction with digital aerial photogrammetry (DAP) techniques for reconstructing vegetation structure. We applied logistic regression on multispectral orthomosaics, clusters of vegetation structure, and local illumination estimates to develop presence-absence models for eight key plant groups at various taxonomic levels as well as six plant functional types (conifer tree seedlings, grasses, tall- and low-growing forbs, shrubs, and mosses). Our results show higher accuracies for plant functional types (mean F-score = 0.67) compared to lower taxonomic levels (0.57). Notably, shrubs (F-score = 0.79), low-growing forbs (0.70), and mosses (0.69) exhibited the highest accuracies, while grasses (0.46), the aster family (Asteraceae spp; 0.48), and spruce seedlings (Picea spp; 0.54) demonstrated lower accuracies. Vegetation structure variables were identified as the most influential in the models, with mean NIRv ranking highest among spectral variables. High average ranks of spectral variation metrics (e.g., standard deviation of NIRv) implied the influence of environmental determinants such as plant co-occurrences and micro-habitat conditions, which drive spectral variation. Discrete composition maps were produced for three restoration sites and analogous wildfire-disturbed sites. Plant compositions found at one site pair exhibited similarity (Bray-Curtis = 0.28), however, certain key plant groups covered larger extents of the restoration site than anticipated. Willows (Salix spp; 25.4% vs. 9.3%), which are typically planted for soil stabilization and obstruction, and clovers (Trifolium spp; 11.1% vs. 3.6%), which represent non-native agronomic vegetation, were prominent. The developed methodology facilitates the generation of detailed plant composition maps, aiding evaluations of vegetation patterns that are difficult to discern visually or through conventional field sampling. This approach can effectively help assess restoration goals and guide adaptive management strategies, especially when incorporating the expertise of restoration ecologists in understanding how different vegetation types affect habitat quality.

### 1. Introduction

#### 1.1. Information requirements for ecological restoration

Biodiversity and ecosystem stability are under continuous pressure due to a range of anthropogenic drivers associated with an increasing human population and the subsequent demand for goods and services (Bowler et al., 2020; Pyšek et al., 2020). Resource extraction, as well as

correlated processes such as land cover and land use change, and non-native species invasion, are among the most important anthropogenic drivers of local biodiversity change (DeMars and Boutin, 2018; Sanderson et al., 2012). Substantial land areas may be affected. For instance, in the Province of Alberta, Canada, seismic lines are among the leading causes of forest fragmentation, with densities reaching  $10 \text{ km/km}^2$ , which, together with other industrial developments including forest clear-cuts, roads and oil wells, increase movement and

\* Corresponding author.

E-mail addresses: [rik.nuijten@alumni.ubc.ca](mailto:rik.nuijten@alumni.ubc.ca) (R.J.G. Nuijten), [nicholas.coops@ubc.ca](mailto:nicholas.coops@ubc.ca) (N.C. Coops), [dustin\\_theberge@tcenergy.com](mailto:dustin_theberge@tcenergy.com) (D. Theberge), [cindy.prescott@ubc.ca](mailto:cindy.prescott@ubc.ca) (C.E. Prescott).

displacement of threatened woodland caribou (*Rangifer tarandus caribou*; Dyer et al., 2001; Lee and Boutin, 2006; Smith et al., 2000). Due to increased public awareness, land managers are increasingly concerned about protecting, enhancing, and restoring directly affected environments in order to maintain their social license to operate, regardless of legal mandates (Young et al., 2022). Ecological restoration, which aims at restoring biodiversity and ecosystem functioning towards baseline levels, is the most far-reaching process to be implemented on degraded lands (Prach et al., 2019). However, ecological restoration is complex as it requires adequate planning, should align with local ecological processes, and must involve monitoring to evaluate if objectives are met and inform adaptive management strategies.

Detailed maps of vegetation are critical to facilitate restoration planning, implementation, and monitoring (Robinson et al., 2022). The acquisition of spatial information of ecological site conditions, including plant composition, structure and function (e.g., productivity and wildlife habitat quality), are critical for monitoring before and after restorative activities (Ruiz-Jaén and Aide, 2005). In the context of resource extraction, where financial resources are relatively large, on-ground field surveys are common (Young et al., 2022). However, on-ground field surveys are time-consuming and depend on extrapolation of data collected in small field plots. Surveys may also disturb sensitive areas, disrupt organic layers, compact soil and damage vegetation (Cruzan et al., 2016). As a result, remote sensing, which has become an important tool to map biodiversity, is highly applicable (Skidmore et al., 2021). Broad-scale mapping of plant functional types (i.e., forests, grasslands, marshes, etc.) using remote sensing is possible due to correlations between reflectance and functional traits including pigmentation, canopy structure, as well as leaf water and nitrogen content (He et al., 2015). Vegetation distribution models may include remotely sensed metrics describing habitat suitability such as spatial climate variables and topographic features derived from, for example, NASA's Soil Moisture Active Passive (SMAP) mission and Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) digital elevation data (Entekhabi et al., 2009; Fujisada et al., 2011) in addition to spectral information. However, ecological restoration generally requires information at high spatial and thematic resolutions including presence-absence or abundance maps of a taxonomic group (e.g., species, genus or family), maps of plant assemblages based on ecological function (e.g., nutrient cycling, height stratum, or fuel load), as well as plant productivity (Robinson et al., 2022). Using traditional remote-sensing platforms (i.e., piloted airborne and spaceborne) such information can be difficult to acquire, particularly in areas with heterogeneous mosaics of early successional vegetation types and high species richness.

## 1.2. Fine-scale vegetation mapping using drones

Drones, or remotely piloted aerial systems (RPAS), can provide data at previously unavailable spatial resolutions, which has been identified as a key innovation in enabling the development of plant composition maps at high thematic and spatial resolutions required by plant ecologists and restoration managers (Robinson et al., 2022). Specifically, image-based cameras alongside digital aerial photogrammetry (DAP) are relevant for vegetation at early successional stages, as they can characterize detailed morphologies of the canopy and provide spectral information, thereby enabling the identification of vegetation communities (Cao et al., 2018; Nuijten et al., 2021). These cameras include red-green-blue (RGB), multispectral, and hyperspectral types. Multispectral cameras are particularly cost-effective, covering a wide range of the electromagnetic spectrum, typically 4–10 spectral bands, including the near-infrared region. They offer a balance of high spatial resolution, affordability, and ease of use where the finer spatial resolutions reduce mixing of spectral signals from different features, compared to hyperspectral cameras. The flexibility of RPAS platforms also allows data collection to be synchronized with phenological and weather conditions,

enhancing the data's relevance.

A growing body of literature utilizes DAP data products to detect plant functional types, or even taxa, in heterogeneous environments with high species richness, such as grasslands and early successional forests. For example, conifer seedlings have successfully been detected using RGB-derived orthomosaics and deep learning approaches in regenerating boreal forests in Alberta, Canada (Chadwick et al., 2020; Fromm et al., 2019). Alexandridis et al. (2017) mapped the distribution of milk thistle (*Silybum marianum*) on abandoned farmland colonized by various grass and forb species with an overall accuracy of 96%. Other applications mapping various plant functional types, generally 3–7 types, are more common (Cao et al., 2018; Cruzan et al., 2016; Cunliffe et al., 2016; van Iersel et al., 2018; Villoslada et al., 2020). According to Bradter et al. (2020), classifications of functional types yield higher accuracies than those of dominant species using 13-band multispectral imagery. This is attributed to the ability of remote sensing data to detect not only the plant species present but also the site conditions (e.g., plant co-occurrence and exposed soil areas). However, Cao et al. (2018) classified various dominant mangrove species with an overall accuracy of 88.7%, including three shrub and one swamp fern species, using RPAS-derived digital surface models and hyperspectral imagery.

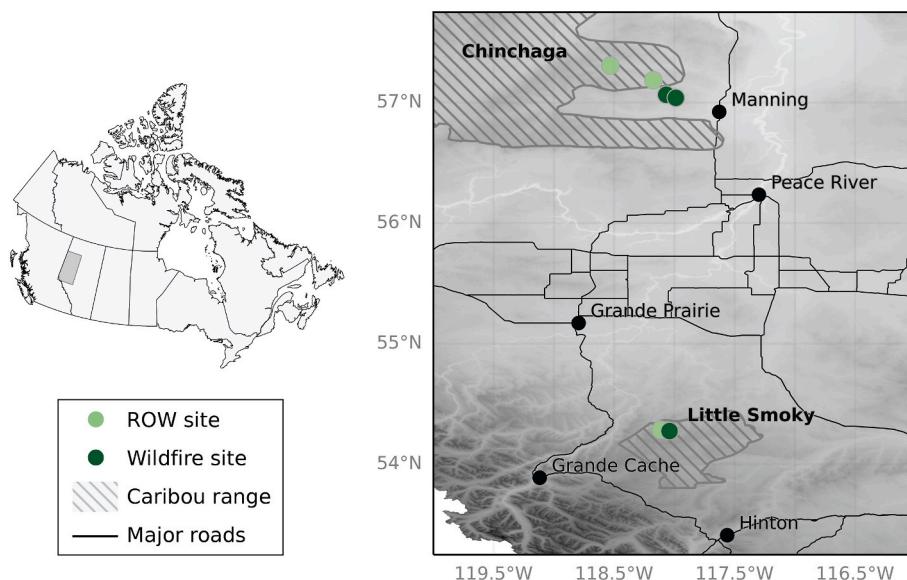
## 1.3. Research objectives

For vegetation maps to be relevant for ecological restoration practitioners, high thematic resolution is essential, particularly for baseline surveys in restoration sites and complex early seral forests, known for their species richness and heterogeneous vegetation mosaics. The challenge in these environments lies in creating detailed thematic classes from remote-sensing spectral reflectance data, due to similar spectral signatures across species and the large number of potential classes present (e.g., numerous plant species and visible conditions like exposed soil). RPAS with integrated multispectral cameras currently stand out as one of the most adaptable and practical remote-sensing technologies for this purpose. However, optimized use of data provided by DAP approaches, in conjunction with field data, is necessary to improve class separability.

This paper explores vegetation mapping using RPAS-based DAP data at two aggregation levels: (1) broader plant functional types, and (2) key plant groups at the genus or family level. Our study investigates which aggregation level yields the most accurate results and whether the number of vegetation classes derived from such RPAS photogrammetric data can be expanded beyond previous studies. A novel workflow is developed, incorporating RPAS-based multispectral orthomosaics and layers describing vegetation structure and microclimate (i.e., local ground illumination conditions), provided by DAP approaches, to potentially enhance class separability. Additionally, the study assesses the importance of spectral variation in the presence-absence models. After developing and testing the models, we evaluate the utility of resulting maps for assessing restoration sites in comparison to wildfire-disturbed sites, which served as natural analogues, to inform adaptive management strategies.

## 2. Study sites

Our focus sites in northwest Alberta included areas impacted by both anthropogenic and non-anthropogenic disturbances. Some sites were impacted by wildfire while others were disturbed by oil and gas exploration that creates pipeline right of ways (ROW), which are part of an extensive natural-gas transmission system (Fig. 1). The ROW sites are within the habitat of the woodland caribou, a species that is threatened and can be impacted by linear disturbances that may change predator-prey interactions and movement patterns (Sanderson et al., 2012; DeMars and Boutin, 2018). Covered ecoregions include the Lower Boreal Highlands and the Rocky Mountain's Upper Foothills (Alberta Parks, 2015). The Lower Boreal Highlands are characterized by



**Fig. 1.** Locations of the right-of-way (ROW) and wildfire-disturbed analogous sites in northwest Alberta, Canada, including current woodland caribou (*Rangifer tarandus caribou*) ranges.

extensive peatlands and mixedwood forests, whereas the Upper Foothills are dominated by a more diverse mix of fens and shrubby grasslands as well as conifer, deciduous and mixedwood forest stands. The climax mixedwood forests in the Lower Boreal Highlands include aspen poplar (*Populus tremuloides*), balsam poplar (*Populus balsamifera*), black spruce (*Picea mariana*) and white spruce (*Picea glauca*) (Willoughby et al., 2016, 2018). In the Upper Foothills, conifer forest stands are most common which include lodgepole pine (*Pinus contorta*), black spruce and white spruce (Willoughby et al., 2018). A more detailed description of the common vegetation communities can be found in Nuijten et al. (2021).

Companies involved in resource extraction in the region aim to reduce their impact on the landscape. This involves the recovery of the land's ability to support land types that existed prior to disturbance (e.g., pipeline construction), as well as the restoration of vegetation composition and reinforcement of wildlife-habitat characteristics where appropriate. Restorative activities are based on vegetation data collected from on-ground and helicopter surveys at target and proximate analogue sites. Common activities include erosion control (e.g., planting willow species around streams), access management (e.g., planting conifers and positioning logs), improving natural regeneration potential (e.g., minimizing soil disturbance) and removing of some invasive species according to provincial regulatory requirements (Government of Alberta, 2021; TC Energy, 2020). Other site characteristics that may be corrected based on survey data include the composition of plant functional types and height strata, seedling densities and line-of-sight (Fili-cetti et al., 2019; TC Energy, 2020). In contrast, forested areas outside of resource extraction activities typically comprise public land, where restorative activities following natural disturbances, primarily fire, are generally lacking.

We selected three sites on ROWs as well as three wildfire-disturbed sites to serve as natural analogues, which were paired by region

(Chinchaga and Little Smoky). Details are summarized in Table 1 including ecoregion, elevation, aspect, years since disturbance at time of data collection, and total area covered by RPAS imagery. Post-wildfire analogues were within 8–70 km of the ROW sites and had an age difference of 1–7 years from the ROW. To ensure that the matched post-wildfire analogues had similar site conditions and had not been impacted by timber harvesting prior to the fires, we used Alberta's vegetation inventory (Alberta Agriculture and Forestry, 2020), digital elevation model (Alberta Environment and Parks, 2017) and timber harvesting polygons (Government of Alberta, 2021). Note that during pair-wise comparisons of compositions, the wildfire sites in the Chinchaga range were treated as a single site due to their comparable age and site conditions.

### 3. Methods

Fig. 2 demonstrates our overall workflow, which included (1) data collection and pre-processing, (2) evaluation of logistic regression models predicting presence-absence of various plant functional types and key plant groups at lower taxonomic levels, and (3) model deployment across study sites and implementation of a novel mapping routine based on probability scores.

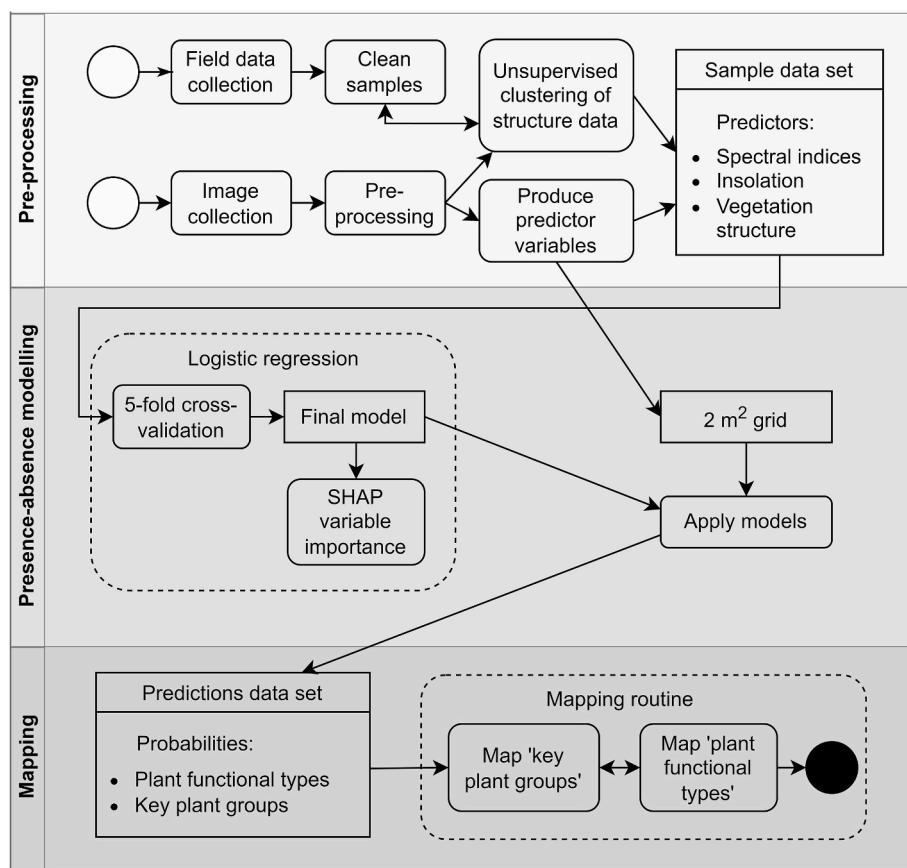
#### 3.1. On-ground surveys

Plant species, height and cover percentage were sampled within 2 x 2-m<sup>2</sup> quadrats along 20-m transects by an experienced field crew (Nuijten et al., 2021). In some cases, plants were sampled at higher taxonomic levels or grouped after sampling due to uncertainty and time constraints. Cover of exposed soil and dead wood were also recorded. Transect locations were determined in the field based on an optimal

**Table 1**

Summary of the study sites. LBH = Lower Boreal Highlands ecoregion; and UF = Upper Foothills ecoregion.

Study site	Years since disturbance	Ecoregion	Elevation (m)	Slope and Aspect	Site area (ha)	No. of field transects
ROW Chinchaga A	10 years	LBH	720	Level	2.6	2
ROW Chinchaga B	5 years	LBH	785	Level	8.2	3
ROW Little Smoky	14 years	UF	1180	7.5, NW	1.6	2
Wildfire Chinchaga A	3 years	LBH	765	Level	17	7
Wildfire Chinchaga B	2 years	LBH	760	Level	28	5
Wildfire Little Smoky	15 years	UF	1210	1.5, NW	24	4



**Fig. 2.** Methodological flow diagram of data collection and pre-processing, the development of various binary presence-absence models, as well as mapping of plant composition.

representation of plant diversity and site conditions visually observed at each study site. To do so, sketch maps of plant functional types were drawn in the field based on visual inspection from a micro drone (DJI Mini). Transects were distributed across the identified plant functional types, while capturing as much variation along the transects as possible (*i.e.*, high species richness, varying compositions). Start and end locations of each transect were recorded using a Trimble Geo7X ground station, which were post-processed using differential GNSS. These locations were then marked with paint prior to RPAS image collections at the wildfire sites. On the ROW sites, transects' positioning was verified by using relative distances from recorded easily identifiable features such as trees, shrubs, and logs. The chosen quadrat size of 2 × 2 m<sup>2</sup>, which determines the spatial resolution of our RPAS-based assessments, aimed to minimize the effects of transects' positioning error compared to finer resolutions.

The sampling data set contained 98 taxa, sampled at the lowest possible taxonomic level, of which 27 taxa were found in at least 20 quadrats. We refined the list of modelled species based on their abundance in the site set, their common usage in restorative activities such as willow (*Salix* spp) and spruce species (*Picea* spp), and their ability to characterize site conditions and native plant communities as described by Willoughby et al. (2016 and 2018). For example, native communities on wet soils often include black spruce (*Picea mariana*), bog cranberry (*Vaccinium vitis-idaea*) and Labrador tea (*Rhododendron groenlandicum*), as well as peat and Sphagnum mosses (*bryophytes*). Native flowering forbs, of which aster species (*Asteraceae* spp) are a large component, are often found on slightly drier and richer sites (Willoughby et al., 2016, 2018). One of the selected groups, clover (*Trifolium* spp), characterizes agronomic plant communities. In the past, restorative activities merely focussed on rapid revegetation and used agronomic seed mixes. Such communities are now undesirable but remain a challenge for newly

reclaimed sites due to ingress from adjacent disturbance and competitive advantages (TC Energy, 2020, 2014).

In total 8 key plant groups were selected for vegetation mapping, including five at genus level, one at family level, as well as bryophytes and grasses, as shown in Table 2. Based on findings by Bradter et al. (2020), classifications based on plant functional types were expected to be more accurate than classifications of dominating species. Therefore, taxa were also grouped into six plant functional types including

**Table 2**

A list of assessed plant functional types and key plant groups at various taxonomic levels, which includes the number of present samples after data cleaning. The last column lists the structure classes, corresponding to descriptions found in Table 3, indicative of the plant functional types and key groups.

	No. of species included	No. of present samples	Key plant groups	Plant func. types	Indicative struc. classes
Low-growing forbs	3 spp	96	–	Included	3, 4, 5, 6
Shrubs	≥ 10 spp	171	–	Included	1, 2, 3, 5, 6
Tall-growing forbs	≥ 2 spp	91	–	Included	2, 3, 5, 6
Trees	3 spp	83	–	Included	1, 2, 3, 5, 6
Bryophytes	≥ 4 spp	104	Included	Included	3, 4, 5, 6
Grasses	≥ 7 spp	85	Included	Included	2, 3, 5, 6
Bare soil	None	19	Included	Included	4
<i>Asteraceae</i> spp	≥ 2 spp	71	Included	–	2, 3, 5, 6
<i>Picea</i> spp	2 spp	79	Included	–	1, 2, 3, 5, 6
<i>Rhododendron</i> spp	1 spp	73	Included	–	2, 3, 5, 6
<i>Salix</i> spp	≥ 1 spp	51	Included	–	1, 3, 5, 6
<i>Trifolium</i> spp	2 spp	38	Included	–	3, 4, 6
<i>Vaccinium</i> spp	3 spp	70	Included	–	2, 3, 5, 6

**Table 3**

The structure classes and corresponding to descriptions from Nuijten et al. (2021).

Class	Structural descriptors	Observed vegetation types
1	Tall structures and most complex canopy	Woody (e.g., willows)
2	Mid-tall and moderate complexity	Tall grasses, forbs and dwarf shrubs (e.g., <i>Vaccinium</i> spp)
3	Short and moderate complexity	Short grasses and forbs (e.g., clovers)
4	No structures and no complexity	Short grasses, moss, and bare soil
5	No structures and minimal complexity	Short grasses, moss, and conifer seedlings
6	Very short and lack of complexity	Short grasses, forbs, and conifer seedlings
7	Structures >2 m height were excluded	N.A.

**Table 4**

Parameters of RGB and multispectral image collection including date of flight, height above ground level (HAGL) and ground sample distance (GSD).

	Month and year	RGB		Multispectral	
		HAGL	GSD orthom.	HAGL	GSD orthom.
ROW Chinchaga A	Mid-Aug. and Early-Sept. 2019	90 m	2.5 cm	60 m	4.4 cm
ROW Chinchaga B	Late-Aug. 2019	125 m	3.4 cm	65 m	4.7 cm
ROW Little Smoky	Mid-Aug. 2019	110 m	2.4 cm	70 m	5 cm
Wildfire Chinchaga A	Late-Aug. 2021	50 m	1.4 cm	45 m	3 cm
Wildfire Chinchaga B	Late-Aug. 2021	38 m	1 cm	40 m	2.2 cm
Wildfire Little Smoky	Mid-Aug. 2021	54 m	1.5 cm	42 m	2.9 cm

bryophytes, grasses, low-growing forbs (largely representing agronomic species), shrubs (deciduous), tall-growing forbs, and tree seedlings (conifer). In addition, bare soil was included in the vegetation maps and treated as an additional key plant grouping and plant functional type (see Table 4).

All plant functional types and key plant groups were found on both

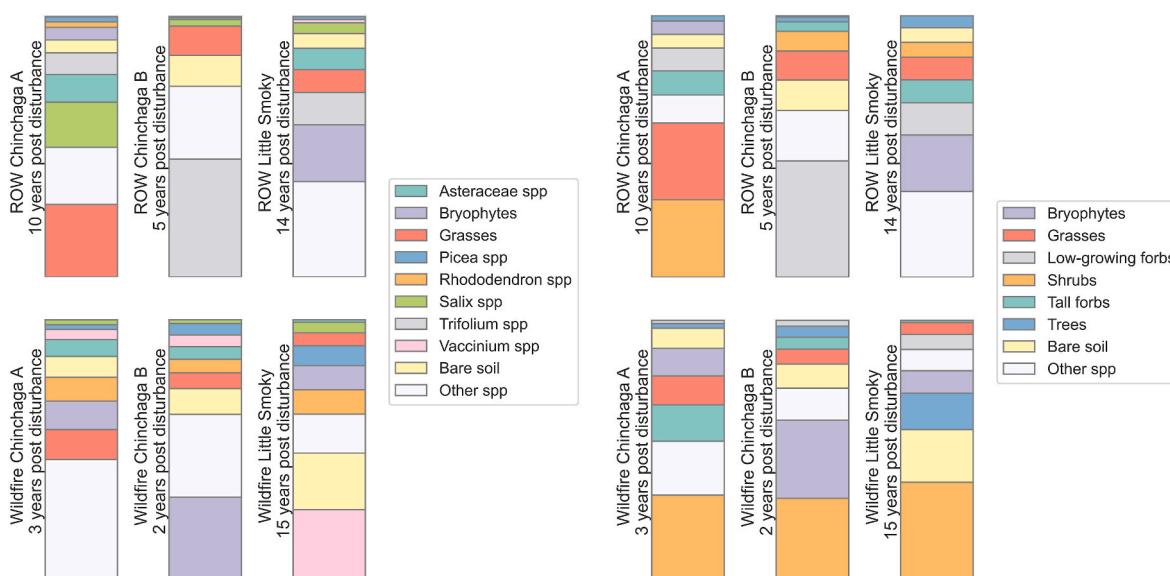
ROW and wildfire sites, as shown in Fig. 3. However, low-growing forbs and grasses were more abundant on right-of-ways, whereas bryophytes and shrubs were generally more abundant on the wildfire sites. High co-occurrence (>50 % of samples co-occurring with another grouping) was found between bryophytes and *Picea* spp, *Rhododendron* spp, as well as *Vaccinium* spp (Fig. 4). In addition, bryophytes were often found together with bare soil.

### 3.2. Image collection and pre-processing

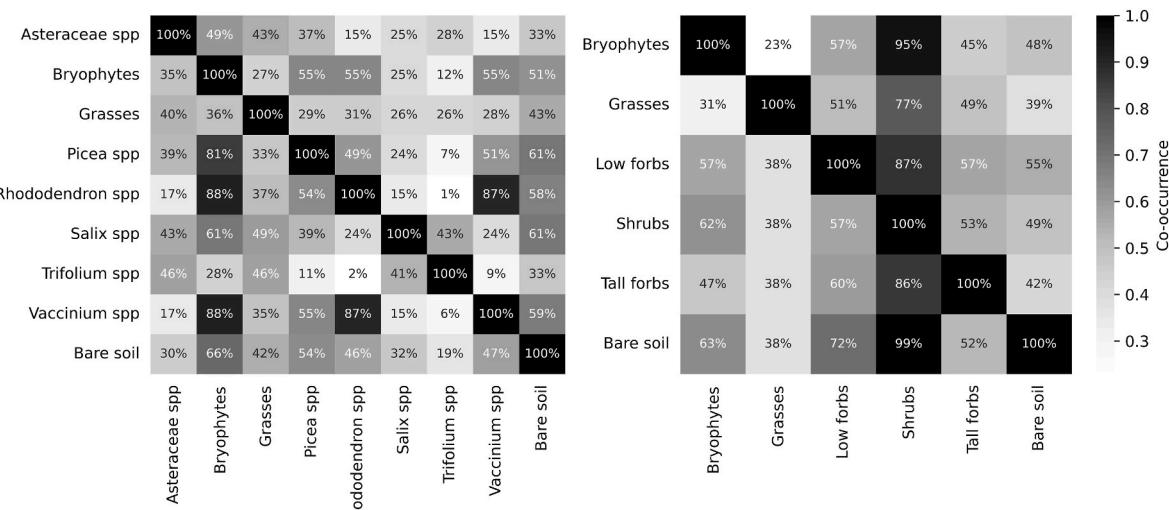
In late summer (August–September) of 2019 and 2021, images were captured using a quadcopter drone (DJI Phantom 4 Pro/Phantom 4 RTK/DJI Matrice M210 RTK V2). We used either a drone with real-time kinematic (RTK) correction or, alternatively, ground-control-points (GCPs) that were post-processed with a Precise-Point-Positioning service, as described by Nuijten et al. (2021). A 20-megapixel RGB imagery was used to acquire imagery at 1–3.3 cm resolution while a Micasense RedEdge-MX 1.2-megapixel multispectral monochromatic sensor collected imagery at 2.5–5 cm resolution, capturing reflectance values at five distinct spectral bands: blue (443–507 nm), green (533–587 nm), red (652–684 nm), red edge (705–729 nm) and near infrared (785–899 nm). For radiometric correction of the multispectral images, irradiance was measured during flights using a downwelling light sensor. Additionally, a calibrated reflectance panel was used to take calibration images before and after each flight. For the multispectral sensor, both image sidelap and frontlap were configured to yield an 80 % overlap which is necessary to account for reduced light energy reaching the camera at wide angles, and required for the construction of detailed point clouds of short-stature vegetation as demonstrated by Cruzan et al. (2016) and Tu et al. (2020). Most images were collected during overcast weather, however, unavoidable sunny conditions at the ROW Little Smoky site resulted in shading.

To obtain a point cloud and two orthomosaics per site, images were processed in Agisoft Metashape (v 1.6.3) as described in Nuijten et al. (2021). Where available, image locations were corrected using correction factors from an RTK base station. For imagery at the ROW study sites where RTK-enabled drones were not available, we incorporated GCPs during the structure-from-motion photogrammetric processing (*i.e.*, DAP), resulting in alignment off-sets of up to 5 cm, similar to the post-wildfire analogues.

To preprocess the point data we undertook ground filtering, height normalization, point cloud thinning, and noise filtering. The ground



**Fig. 3.** Cover percentages by site of selected key plant groups, at various taxonomic levels (left), and plant functional types (right) based on the field sample data set.



**Fig. 4.** Relative co-occurrence matrices for selected key plant groups, at various taxonomic levels (left), and plant functional types (right) based on the field sample data set.

filtering approach included a pre-processing step which involved using the excessive greenness index (ExGI), calculated point-wise from RGB color values, to identify non-vegetation points, specifically those with ExGI values lower than the 35th percentile. Ground detection primarily utilized triangulated irregular network densification followed by an iterative surface lowering approach (*i.e.*, ground smoothing; [Anders et al., 2019](#); [Isenburg, 2012](#); [Nuijten et al., 2021](#)). After ground filtering and height normalization, we performed point cloud thinning, retaining the highest point per  $4 \times 4 \text{ cm}^2$ . Noise filtering, conducted by examining  $2 \times 2 \text{ m}^2$  cells, removed erroneous points and those from overhanging branches, further refining our data sets ([Isenburg, 2012](#)).

We used several sources of data in our analysis, which included multispectral imagery and insolation layers derived from a digital surface model. In addition, we used a fine-scale vegetation structure map of the region, which identified six structural types of regenerating vegetation and was created by [Nuijten et al. \(2021\)](#). The map was created using an unsupervised clustering approach that incorporated various point-cloud-based height metrics, such as maximum height. These metrics were originally derived at a  $1 \cdot \text{m}^2$  spatial resolution and then aggregated to  $2 \times 2 \text{ m}^2$  by counting the number of classes (up to four classes per pixel) and creating binary presence-absence variables for each class. To account for local illumination affecting growing conditions, local shading maps were generated using digital surface models and the location of the sun ([Morgan-Wall, 2023](#)). Insolation surfaces, aggregated at  $2 \times 2 \text{ m}^2$ , were then derived for every first day of the week

during the core growing season (June, July, and August).

### 3.3. Predictor variables

A total of 15 predictor variables was produced at 2-m spatial resolution including 10 spectral variables, from cameras listed in [Table 5](#), as well as 5 variables based on 3D reconstructions of vegetation structure describing structural composition and growing conditions. The produced insolation surfaces were generated for 9 a.m., noon and 3 p.m. across the core growing season. Spectral variables include the normalized difference vegetation index (NDVI) which is a commonly used index related to vegetation cover and photosynthetic activity ([Cruzan et al., 2016](#)). In addition to NDVI, we selected three other multispectral-derived indices including near-infrared reflectance of vegetation (NIRv), normalized difference red edge index (NDRE) and plant senescence reflectance index (PSRI). NIRv, which is the product of NDVI and total NIR reflectance, better reflects the proportion of reflectance attributable to the vegetation in the pixel ([Badgley et al., 2017](#)). NDRE incorporates the red-edge region, which may improve classification accuracies due to its sensitivity to carotenoids, chlorophylls and other pigments ([Villoslada et al., 2020](#)). Finally, PSRI is sensitive to species-specific plant phenology based on the ratio between carotenoids and leaf chlorophylls, which changes during fruiting and senescence. In addition to multispectral reflectance data, the RGB-derived normalized green ratio index (GRI) was included, which is typically used to detect

**Table 5**

The predictor variables used in the species distribution models (SDMs). For variables indicated with <sup>1</sup> both mean and standard deviation were calculated. Variables indicated with <sup>2</sup> are based on the outputs of unsupervised clustering analysis.

Predictor variables	Type	Data source	Equation	Reference
Green ratio index (GRI) <sup>1</sup>	Biophysical	RGB ortho.	$\frac{G}{B+G+R}$	<a href="#">Ide and Oguma (2010)</a>
Near-infrared reflectance of vegetation (NIRv) <sup>1</sup>	Biophysical	Multispectral ortho.	$\frac{NIR-R}{NIR+R} * NIR$	<a href="#">Badgley et al. (2017)</a>
Normalized Difference Red Edge Index (NDRE) <sup>1</sup>	Biophysical	Multispectral ortho.	$\frac{NIR-RE}{NIR+RE}$	<a href="#">Gitelson and Merzlyak (1994)</a>
Normalized Difference Vegetation Index (NDVI) <sup>1</sup>	Biophysical	Multispectral ortho.	$\frac{NIR-R}{NIR+R}$	<a href="#">Tucker (1979)</a>
Plant Senescence Reflectance Index (PSRI) <sup>1</sup>	Biophysical	Multispectral ortho.	$\frac{R-G}{NIR}$	<a href="#">Merzlyak et al. (1999)</a>
Insolation 9 a.m.	Microclimate	Digital surface model (DSM)	N.A.	<a href="#">Morgan-Wall (2023)</a>
Insolation 12 p.m.	Microclimate	Digital surface model (DSM)	N.A.	<a href="#">Morgan-Wall (2023)</a>
Insolation 3 p.m.	Microclimate	Digital surface model (DSM)	N.A.	<a href="#">Morgan-Wall (2023)</a>
Structure class count <sup>2</sup>	Structure classes	Point cloud	N.A.	<a href="#">Nuijten et al. (2021)</a>
Structure class occurrence <sup>2</sup>	Structure classes	Point cloud	N.A.	<a href="#">Nuijten et al. (2021)</a>

phenological changes and canopy photosynthesis rate (Ide and Oguma, 2010). Mean and standard deviation within the  $2 \times 2 \text{ m}^2$  cells were derived for all spectral vegetation indices. The Pearson correlation coefficients ( $r$ ) for all variables were below 0.75, with an average  $r$  of 0.21 across all variables and 0.36 specifically for the spectral indices. The NDVI-NDRE and NDVI-PSRI index combinations exhibited highest correlations ( $r \approx 0.75$ ).

### 3.4. Logistic regression modelling and mapping routine

Logistic regression was used to generate plant-occurrence models, followed by a mapping routine based on probability scores. Logistic regression models have high computational efficiency, reduce the likelihood of overfitting using small data sets, and produce probability distributions which can be compared across different models (James et al., 2021). To account for class imbalance, models used class weights inversely proportional to class frequencies. Our implementation used the limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) method and maximum 1000 iterations (Pedregosa et al., 2011). Models were assessed using 5-fold cross validation, each randomly withholding 20 percent of the samples for independent model evaluation. Performance metrics reported in this study include precision, recall, specificity, and the F-score:

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{Specificity} = \frac{TN}{TN + FP}$$

$$F \text{ score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

where FP is false positives, FN is false negatives, TP is true positives and TN is true negatives. The F-score, which ranges from 0 to 1, describes the models' balanced ability to accurately capture actual positive cases (i.e., recall) as well as the ratio of correct positives produced by the model out of all predictions (i.e., precision). Final presence-absence models, from which predictor variable importance and wall-to-wall probability predictions were derived, were trained using all available samples.

The vegetation survey data were filtered on the seven structure clusters to maintain only observations of key plant groups and plant functional types having corresponding structural classes, as illustrated in Fig. 5 and defined in Table 2. Likewise, when applying the models over the entire area, presence-absence predictions were not undertaken on  $2 \times 2 \text{ m}^2$  cells with inconsistent structure classes. SHapley Additive exPlanation (SHAP), a model-agnostic tool for explaining the magnitude

and direction of the variable's global effect on model outputs, was used to calculate variable importance to understand the overall contribution of classes of vegetation structure, which were represented by seven binary predictor variables, i.e., presence-absence, as well as other predictor variables listed in Table 5 (Cha et al., 2021) (see Fig. 6).

Final maps were created based on the probability scores provided by the logistic regression models at two thematic resolutions—plant functional types and key plant groups as listed in Table 2. Based on Bradter et al. (2020), it was assumed models of plant functional types would achieve higher accuracies compared to models based on plant assemblages at low taxonomic ranks (e.g., species or genus). This assumption was incorporated into the mapping routine, which verifies, for each pixel, agreement between the probabilities of 'key plant groups' and the associated plant functional type. In an iterative process, key plant groups sorted from highest to lowest probability are accepted based on the following criteria: (1) the associated plant functional type (e.g., shrubs) has the highest probability score amongst other functional types; and (2) the difference between probabilities of the plant functional type and the associated 'key plant grouping' (*Vaccinium* spp) is smaller than 0.35. If none of the 'key plant groups' was selected during the first iteration, the routine was repeated using the plant functional type with the second highest probability score. The threshold value of 0.35 was selected based on empirical evaluation of mapping products in a trial-and-error process, involving high resolution orthomosaics and expert knowledge.

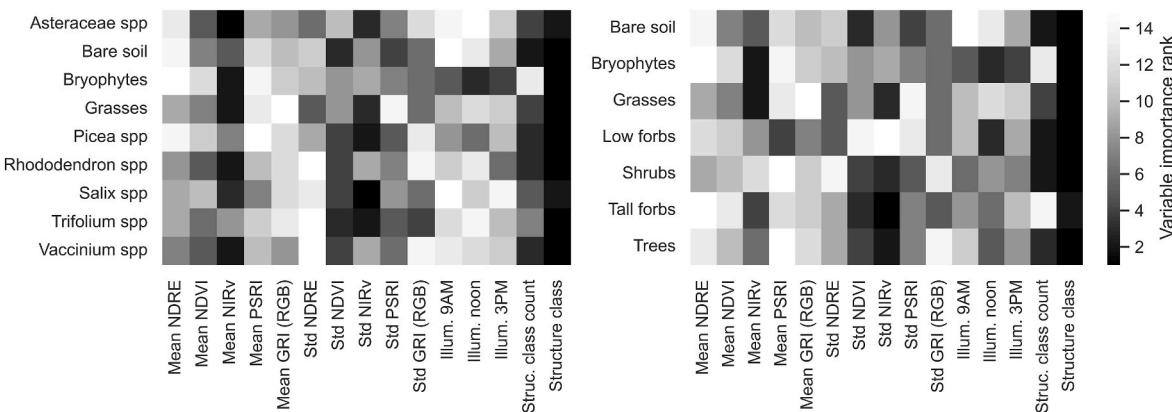
Lastly, to demonstrate how cover of key plant groups and plant functional types can be used to evaluate the state of regeneration, we produced measures of Bray-Curtis dissimilarity to quantify compositional dissimilarity between paired sites (Virtanen et al., 2020).

## 4. Results

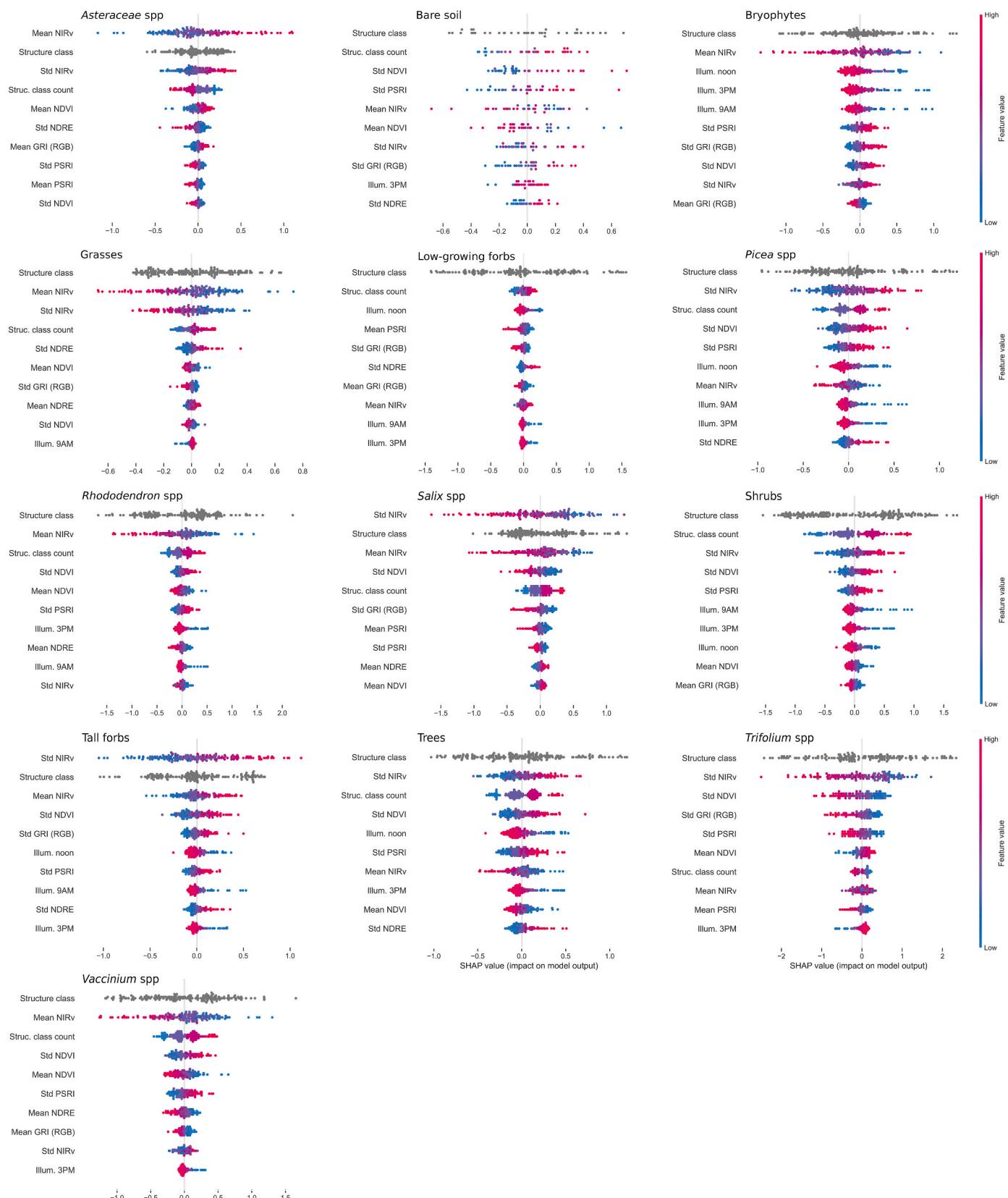
### 4.1. Performance presence-absence models

In total, we developed 13 logistic regression models for plant functional types, key plant groups, and bare soil. Six of the models have an F-score larger than 0.6, including shrubs (F-score = 0.79), low-growing forbs (0.70), bryophytes (0.69), *Trifolium* spp (0.67), bare soil (0.64), and *Rhododendron* spp (0.62). Of these models, shrubs, bryophytes and low-growing forbs have highest precision scores (0.69–0.93), whereas *Trifolium* spp, *Vaccinium* spp, shrubs, and *Rhododendron* spp have highest recall (0.68–0.79). The poorest classification models were for grasses (F-score = 0.46), *Asteraceae* spp (0.48), *Picea* spp (0.54) and conifer tree seedlings (0.60), with lowest precision for *Asteraceae* spp (0.44) and lowest recall for grasses (0.46).

The plant functional type models for tall- and low-growing forbs, shrubs, and trees have higher accuracies than the associated key plant groups at lower taxonomic levels, with an average F-score of 0.67 compared to 0.57 (Table 6). The number of false positives produced by



**Fig. 5.** Variables ranked by importance for presence-absence models of the key plant groups (left), and plant functional types (right).



**Fig. 6.** SHAP variable importance plots for all logistic regression-based presence-absence models.

**Table 6**

Mean and standard deviations of performance metrics derived from cross validation, including precision, recall (i.e., true positive rate), specificity (i.e., true negative rate) and f-score.

Classification system	Presence-absence model	Precision	Recall	Specificity	F-score
Plant func. types	Low-growing forbs	0.69 (0.09)	0.71 (0.13)	0.61 (0.12)	0.70 (0.10)
	Shrubs	0.93 (0.06)	0.70 (0.11)	0.69 (0.29)	0.79 (0.08)
	Tall-growing forbs	0.60 (0.10)	0.55 (0.15)	0.64 (0.14)	0.57 (0.11)
	Trees	0.56 (0.07)	0.65 (0.17)	0.64 (0.11)	0.60 (0.10)
	Bryophytes	0.73 (0.07)	0.66 (0.07)	0.60 (0.16)	0.69 (0.03)
	Grasses	0.48 (0.05)	0.46 (0.17)	0.59 (0.13)	0.46 (0.13)
	Bare soil	0.64 (0.11)	0.68 (0.32)	0.58 (0.33)	0.64 (0.21)
	Asteraceae spp	0.44 (0.02)	0.54 (0.10)	0.57 (0.08)	0.48 (0.05)
	Picea spp	0.50 (0.13)	0.60 (0.23)	0.63 (0.10)	0.54 (0.16)
	Rhododendron spp	0.58 (0.19)	0.68 (0.08)	0.64 (0.19)	0.62 (0.14)
Key plant groups (i.e. lower taxa)	Salix spp	0.49 (0.11)	0.65 (0.12)	0.72 (0.11)	0.55 (0.07)
	Trifolium spp	0.58 (0.15)	0.79 (0.16)	0.79 (0.10)	0.67 (0.15)
	Vaccinium spp	0.53 (0.08)	0.71 (0.11)	0.60 (0.08)	0.60 (0.09)

the models of key plant groups is relatively high, having an average precision of 0.52 compared to 0.70 for models of plant functional types. However, specificity, which describes the proportion of the data set's negatives correctly identified, is relatively high for most key plant groups, particularly for *Trifolium* spp, *Salix* spp and *Rhododendron* spp.

#### 4.2. Variable importance

Fig. 5 shows the predictor variables ordered by importance based on SHAP values, which demonstrates the inclusion of classes of vegetation structure was important for all models. In addition, structure-class count was of high importance for most models, except bryophytes and tall-growing forbs. Higher class counts were driving present predictions for bare soil, tree seedlings, low-growing forbs, shrubs, *Rhododendron* spp, *Salix* spp, *Vaccinium* spp and *Picea* spp. Low average illumination values across the core growing season, at 9 a.m., noon and 3 p.m., were driving presence of bryophytes as well as tree seedlings and *Picea* spp.

With respect to variable importance, the standard deviation of the spectral metrics at 2-m resolution were generally more important than mean values as their combined average ranks were 7.6 and 9.4, respectively, as derived from Fig. 5. Generally, mean NIRv and standard deviations of NDVI and NIRv were the most important spectral index-derived variables. Importance of mean NIRv was ranked 5th or higher for most models, except for low-growing forbs, shrubs, tree seedlings, *Trifolium* spp and *Picea* spp. Small mean PSRI values, which is sensitive to the ratio between carotenoids and leaf chlorophylls and changes during fruiting and senescence, were partly driving present-predictions of low-growing forbs. A large mean NDVI value was associated with the presence of *Trifolium* spp. Generally, NDRE, which incorporates the red-edge region, and RGB-derived GRI were important for presence-absence models of herbaceous plants, which includes *Asteraceae* spp, grasses, *Trifolium* spp, as well as low- and tall-growing. Key plant groups and plant functional types that were characterized by the inclusion of exposed soil and minimal vegetation structure, such as bryophytes, tree seedlings and bare soil, saw relatively high importance of the standard deviation of PSRI. SHAP dependency plots in Appendix I-III, illustrating

interaction effects across a selection of predictor variables, show that interaction effects are common between vegetation indices (e.g., NDVI-PSRI) as well as between illumination variables.

#### 4.3. Output mapping routine

The developed models were applied across the study sites at a 2-m resolution, as shown in Figs. 7 and 8. Based on these site summaries, compositional dissimilarity was assessed across ROW and post-wildfire analogue sites using Bray–Curtis dissimilarity (BC). Similarity of plant functional type cover was highest (BC = 0.28) between ROW Chinchaga A and its wildfire analogue, followed by Chinchaga B (0.49). Similarity was lowest between the ROW and the wildfire site in the Little Smoky range (0.69). Overall, the fractional differences across sites for the various classes were smallest for grasses, with total cover being 1.3–4.8 times higher compared to the wildfire sites. Fractional differences were largest for tall-growing forbs (11.9–71.6 times less cover) and low-growing forbs (3.1–9.7 times more cover). The direction of fractional difference varied most for bryophytes and conifer tree seedlings. For example, bryophytes had lower cover on the ROW Chinchaga B site compared to the post-wildfire analogue (1.5 times) but higher cover on the ROW Chinchaga A and Little Smoky and sites (1.2–3.5 times).

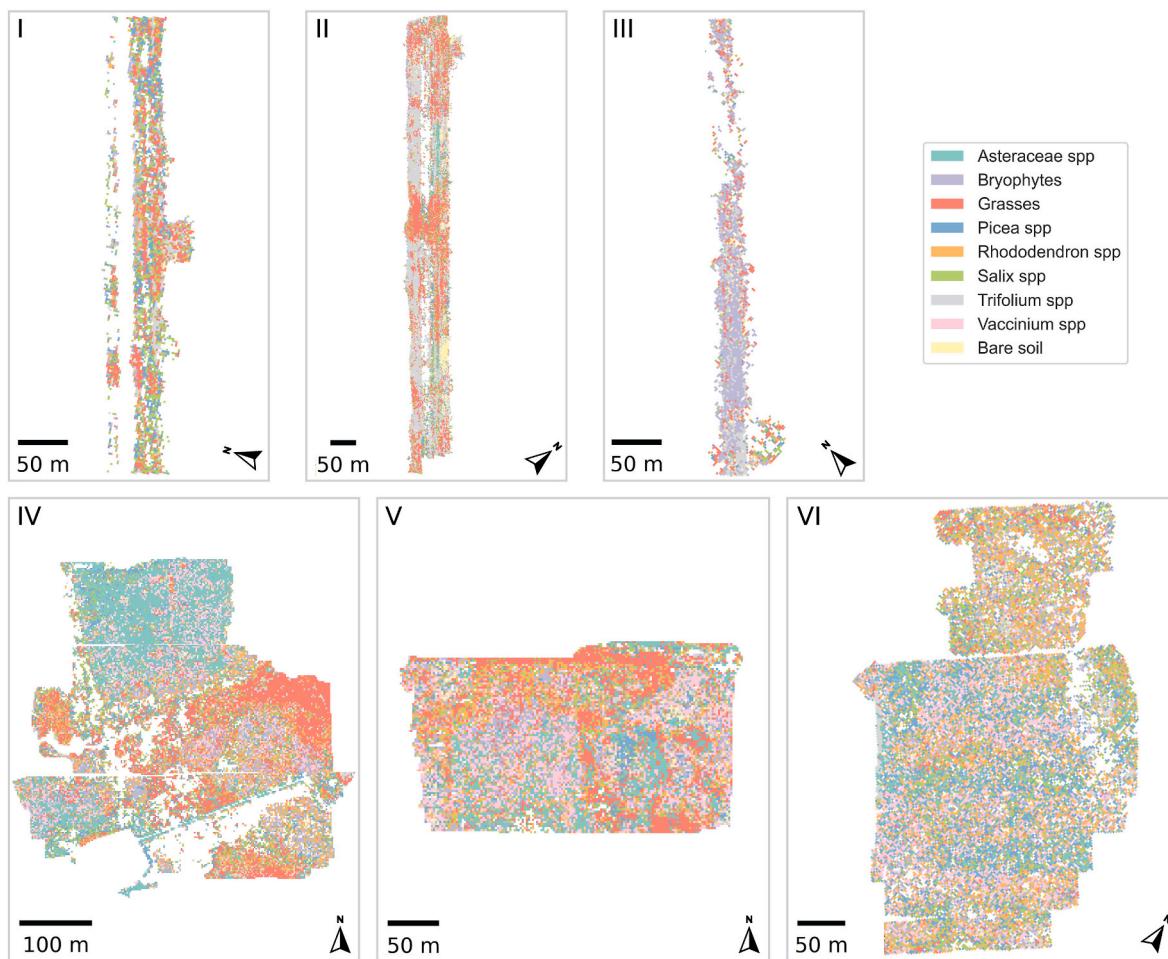
For key plant groups, the BC dissimilarity scores were 0.44 for ROW Chinchaga A, 0.49 for ROW Chinchaga B, and 0.69 for ROW Little Smoky. Some key plant groups, including *Salix* spp (1.3–2.7 times more cover) and *Trifolium* spp (3–9.7 times), cover relatively large extents of most ROW sites. For example, at the Chinchaga B site the total cover of *Trifolium* spp was 34.8% while cover at the wildfire site was 3.6%. However, *Salix* spp cover (6.9 versus 12.8%) was lower on the ROW Little Smoky site. Classes covering relatively small extents of the ROW sites include the forb class *Asteraceae* spp (11.9–71.7 times less cover) as well most woody classes, including *Picea* spp (1.2–2.5 times), *Rhododendron* spp (2.7–9 times) and *Vaccinium* spp (12.1–57.7 times). Note that *Picea* spp had greater cover on the ROW Smoky site compared to the analogue (6.3 versus 15.9%). This is largely in line with coverage estimates found in the field transects, illustrated in Fig. 3.

## 5. Discussion

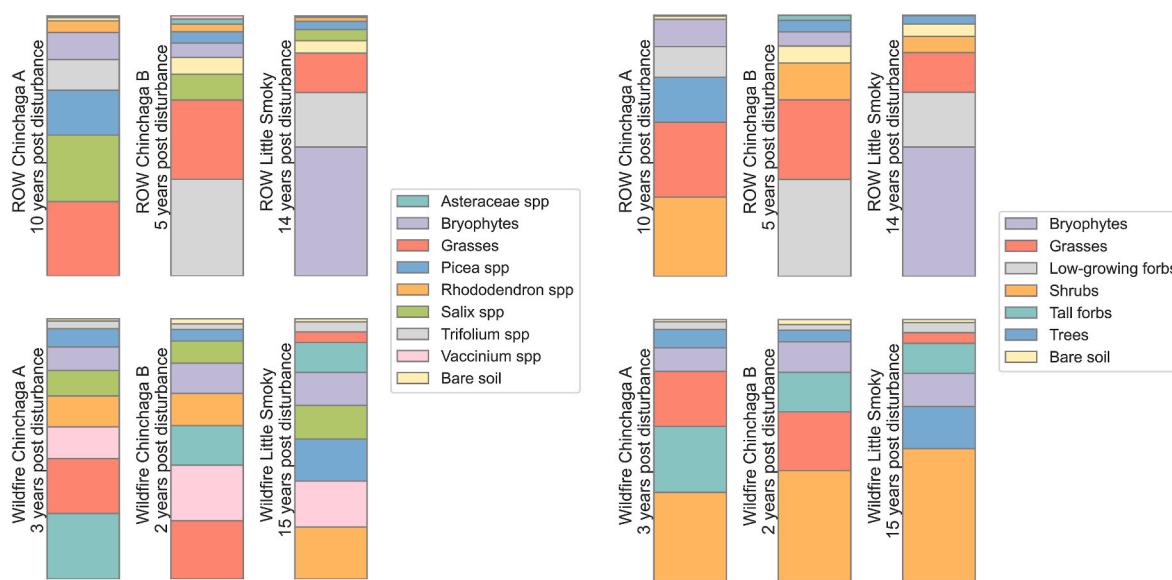
Robinson et al. (2022) recently highlighted the potential for RPAS to map vegetation composition and structure for restoration ecology and management. Several studies have demonstrated the use of high-resolution drone imagery for such purposes, typically mapping a few (2–7) ecologically relevant plant functional types or individual plants such as conifer seedlings or an invading species (Alexandridis et al., 2017; Cao et al., 2018; Chadwick et al., 2020; Cruzan et al., 2016; Cunliffe et al., 2016; van Iersel et al., 2018; Villoslada et al., 2020). Our work has integrated remote sensing products produced by high spatial resolution RPAS imagery of spectral reflectance, vegetation structure and solar illumination conditions to map plant composition. We produced maps at higher thematic resolution compared to previous studies, in order to assess a wider range of ecosystem attributes. This aids assessments of composition using an appropriate reference ecosystem and identification of unwanted vegetation encroaching a habitat, providing a basis for restoration ecologists to manage sites and implement mitigation measures (Gann et al., 2019).

#### 5.1. Presence-absence models

Presence-absence models were produced for various plant functional types and key plant groups at lower taxonomic levels (where feasible), as well as bare soil. The models of plant functional types were more accurate than key plant groups, with average F-scores of 0.67 and 0.58, respectively. This was attributed to their higher precision, reflecting a greater proportion of present predictions being correct. Spatial maps were created based on probability scores and a mapping routine which



**Fig. 7.** Logistic regression-based composition maps of key plant groups at right-of-way (ROW) Chinchaga A (i), ROW Chinchaga B (ii), ROW Little Smoky (iii), wildfire Chinchaga A (iv), wildfire Chinchaga B (v) and wildfire Little Smoky (vi).



**Fig. 8.** Cover percentages for key plant groups (left) and plant functional types (right) at each site, retrieved from the final mapping product.

verifies agreement between the associated models, for example between the model for *Rhododendron* spp and shrubs. This allows us to take advantage of the strengths of models at different thematic scales, where

the plant functional types generally produce relatively fewer false positives and the key plant groups (lower taxa) produce fewer false negatives in some models.

We found that standard deviations of NDVI and NIRv were the most important variables overall, and in some instances, those of PSRI as well. The standard deviation of PSRI, which is sensitive to the ratio of bulk carotenoids to chlorophyll, was important for discriminating native flowering species (*i.e.*, Asteraceae spp), woody vegetation (including *Vaccinium* spp, *Rhododendron* spp and *Picea* spp), and bryophytes. These higher standard deviations may be caused by blooming events observed in *Asteraceae* spp, berry production in *Vaccinium* spp, and senescing in *Rhododendron* spp. However, spectral variation captures not just variations within plants themselves but also within their habitats, such as the presence of exposed soil and co-occurring species (Fassnacht et al., 2022). For instance, exposed soil was commonly observed adjacent to *Picea* spp and bryophytes, while field samples indicated that *Rhododendron* spp often co-occurred with bryophytes.

### 5.2. Probability-based mapping routine

Our results were produced at two thematic scales — plant functional types and key plant groups — revealing patterns of emerging desired and undesired vegetation and enabling pair-wise site comparisons. For example, cells with a high probability of *Trifolium* spp occurrence, which is undesired as it represents pasture vegetation and may limit tree establishment, are more abundant on the pipeline ROWs and visually patchier. In contrast, traditional field studies do not reveal the extents and patterning of taxa and functional vegetation types unless a very dense sample network has been set up and subjective site-level cover estimates are recorded (Levin, 1992; Wheatley and Johnson, 2009).

Bray–Curtis dissimilarity scores suggest that the ROW Chinchaga A site is most similar to its post-wildfire analogue, having a score of 0.28 and 0.44 for plant functional types and key plant groups, respectively. The fractional difference in cover is relatively small, *i.e.*, smaller than a factor of 2, for 25% of the key plant groups which include bryophytes and grasses. The cover of the remaining key plant groups was relatively dissimilar between the ROW and wildfire sites, having a fractional difference larger than a factor of 2.6. Of these groups, *Asteraceae* spp (0.3 versus 21.5%) and *Vaccinium* spp (0.3 versus 14.5%) had very little cover. The low coverage of *Asteraceae* spp, an important component of native wildflower communities, is concerning as it deviates from post-wildfire site composition and has an important function for bumble bee species (Pengelly and Cartar, 2010). Other large differences were found for *Salix* spp (24.4 versus 9.3%) and *Rhododendron* spp (4.5 versus 11.9%). A larger component of *Rhododendron* spp may be desired based on post-wildfire composition, however, this may not be concerning as *Salix* spp and *Picea* spp replace some of its function. Both shrub taxa and conifers can limit line-of-sight which potentially improves caribou habitat (Filicetti et al., 2019). In addition, large *Salix* spp and *Picea* spp cover may be a result of plantings performed to stabilize soil and increase vegetation structure (TC Energy, 2020, 2014).

The cover of bryophytes, grasses, *Picea* spp and *Salix* spp was relatively similar between the ROW Chinchaga B site and its post-wildfire analogue, having a fractional difference smaller than a factor of 2. The remaining groups had a fractional difference larger than a factor of 3.8. Despite comparable *Salix* spp cover, which is desirable, the establishment of shrubs (16.1 at ROW versus 35.7% at wildfire) was limited compared to the post-wildfire analogue. The largest differences in cover were found for *Trifolium* spp (34.8 versus 3.6%), *Asteraceae* spp (1.8 versus 21.5%), *Rhododendron* spp (3.1 versus 11.9%) and bare soil (6.1 versus 1.2%). High coverage of *Trifolium* spp is particularly undesirable as it characterizes agronomic plant communities that can inhibit tree establishment (TC Energy, 2020, 2014).

Only cover of *Salix* spp was relatively similar between the ROW Little Smoky site and its post-wildfire analogue, where cover of 50% of the key plant groups had a fractional difference larger than a factor 4. Like the paired Chinchaga B site, *Picea* spp (6.3 versus 15.9%) and *Vaccinium* spp (0.3 versus 17.3%) were almost completely absent on the ROW, however, the associated differences were largest for the paired Little Smoky

site. In addition, shrubs (9.4 versus 50%) and conifer tree seedlings (6.3 versus 15.9%) covered relatively small extents of the ROW site. In contrast, grasses (19.8 versus 4.1%), bryophytes (43.5 versus 12.5%) and *Trifolium* spp (17 versus 4.5%) covered a large extent of the ROW site. High bryophyte cover is desirable as bryophytes fulfill important ecosystem functions in temperate and boreal forests. Visual assessment of the mapping product suggests that line-of-sight has not been reduced due to a lack of regenerating woody vegetation (Bonan and Shugart, 1989).

### 5.3. Model improvement: bias and predictor variables

Spectral signals and vegetation structure change throughout the growing season due to changing photosynthetic activity, phenological events and leaf structures (Dandois and Ellis, 2013; Nuijten et al., 2019). Our RPAS imagery, collected in August which is late summer, captured species with distinct spectral signals, notably those in blooming phases or showing early signs of senescence. Further research is necessary to examine how the developed models will perform when applied to imagery collected at different times of the growing season. Model accuracies may improve by incorporating multi-temporal reflectance data at a frequency timed with phenological variations. For example, conifer seedlings including *Picea* spp, for which we found relatively low accuracies (*i.e.*, F-scores) of 0.60 and 0.54, respectively, are more distinct during dormant season (Chadwick et al., 2020).

Improving model performance could involve increasing the spatial resolution of predictor variables (*e.g.*, mean NIRv) to reduce the mixing of species' spectral signals, while still capturing adequate spectral variation. Achieving this requires improving the spatial alignment between field quadrats (*i.e.*, training and validation data) and orthomosaics, which can be as large as 10 cm. This task is particularly challenging in remote areas, where distant differential GNSS stations and limited cellular service are common. However, the use of mobile differential stations for RTK corrections during both RPAS and ground-based GNSS data collection could facilitate this by improving positioning accuracy significantly (Assiri et al., 2023).

The developed presence-absence models were based on sample data sets covering relatively small geographic extents. In such cases, models may become biased due to underrepresentation of important environmental determinants such as plant co-occurrences and microhabitat conditions (Besag, 1974). Therefore, it will be necessary to acquire additional field samples when applying our methodology to proximate sites. Note that such bias may be less likely to occur if the predictor variables do not measure vegetation directly, but instead represent characteristics of habitat suitability including surface temperatures, soil parameters, and micro topography (Bradley et al., 2012). Such predictors may not be easily derived from RPAS imagery or contain sufficient spatial detail. For example, terrain characterized by DAP may not capture small topographic features, particularly under thick vegetation (Lovitt et al., 2017). To derive these additional ecosystem attributes, other RPAS-based sensing technologies are required. For instance, LiDAR can capture small changes in micro topography, whereas thermal imaging may be used to identify surface water and elevated soil moisture content (Khanal et al., 2017; Poirier et al., 2022; Watts et al., 2023).

### 5.4. Vegetation structure as predictor

This research utilized a range of RPAS-acquired spectral and structural predictors. Canopy height models alone, which represent vegetation height, enable mapping of up to three plant functional types, for example shrubs, trees and swale/hummock habitat (Cruzan et al., 2016; Cunliffe et al., 2016). Alternatively, a variety of point cloud metrics, such as height, canopy cover, and height variability, can be used to classify vegetation. Nuijten et al. (2021) demonstrated that mapping structures representative of woody vegetation and tall grass-forb communities is possible, however, distinguishing lower-stature vegetation

types like bryophyte-dominated communities, flowering plants, and short pasture-type vegetation using only structural data remains challenging. This paper builds on their work by utilizing the six developed classes of vegetation structure, which were produced through unsupervised clustering as described in section 5.2. Derived predictor variables included structure class occurrence (*i.e.*, binary presence-absence variables) and class count. The overall importance score of structure class occurrence (*i.e.*, sum of SHAP values of binary variables) was highest or second highest for all logistic regression models. Some models including bryophytes, *Picea* spp, as well as the plant functional types shrubs and conifer trees, however, were largely driven by low estimates of received solar illumination throughout the growing season, which is another predictor variable derived from modelled vegetation structure.

### 5.5. Operability

The methodologies and results produced in this research can be readily used to assess large restoration and wildfire sites (40–120 ha) due to recent advances in camera and RPAS technology (Chadwick et al., 2022). Improved flight duration of RPAS and enhanced spatial image resolution and capture speeds have reduced data acquisition time significantly (Coops et al., 2019). Moreover, the use of very high spatial resolution multi-camera image acquisitions, where data sets from multiple cameras are fused, for mapping key plant groups at low taxonomic levels, requires near-centimeter alignment between data products. This is achievable with advancements in Global Navigation Satellite System (GNSS) technology, including RTK GNSS, and improvements in computational power and software (Forlani et al., 2018). However, the effectiveness of these methodologies depends on image resolution (<5 cm), to capture spectral variation, and optimal illumination conditions such as cloud overcast and solar noon (Cruzan et al., 2016; Tu et al., 2020). Additionally, non-sunlit pixels may need to be removed (Shen and Cao, 2017).

## 6. Conclusion

This study utilized high-resolution RGB and multispectral RPAS imagery to derive fine-scale vegetation distribution information to aid restoration monitoring. Presence-absence models were produced for plant functional types and key plant groups at lower taxonomic levels using predictor variables derived from RPAS-based maps of vegetation structure, spectral indices derived from multispectral reflectance data, and insolation surfaces. Performance metrics showed that the presence-absence of plant functional types can be more accurately predicted ( $F$ -score = 0.67) than the occurrences of associated key plant groups ( $F$ -score = 0.57). Generally, important predictor variables included the types of vegetation structure, spectral variation (*i.e.*, standard deviation of various spectral indices), and the mean value of NIRv. Using a novel

methodology, probabilistic outputs of the presence-absence models were used to produce discrete maps. The maps may allow the patchiness of the landscape to be assessed, revealing patterns of emerging desired and undesired vegetation and enabling comparisons with contemporary reference sites (*e.g.*, post-wildfire analogues). Plant composition at the 10-year-old ROW site in the Lower Boreal Highlands ecoregion (Bray-Curtis = 0.28) was most similar to its paired post-wildfire analogy, followed by the nearby 5-year-old ROW site ( $BC = 0.49$ ) and, lastly, the 14-year-old ROW site in the Upper Foothills ecoregion ( $BC = 0.69$ ). *Salix* spp was the dominant woody component on the ROW sites, while *Asteraceae* spp (a key native wildflower species) and *Picea* spp (a key conifer species) were generally lacking. *Trifolium* spp, which is undesirable as it characterizes non-native agronomic plant communities, covered relatively large areas at two of the three ROW sites (17 and 34.8%). Our approach can be readily applied to similar restoration and post-wildfire sites and be used to inform restoration programs and assess restoration success.

### CRediT authorship contribution statement

**Rik J.G. Nuijten:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. **Nicholas C. Coops:** Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing. **Dustin Theberge:** Writing – review & editing. **Cindy E. Prescott:** Writing – review & editing.

### Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Nicholas C. Coops reports financial support was provided by Natural Sciences and Engineering Research Council of Canada. Nicholas C. Coops reports financial support was provided by TC Energy Corporation.

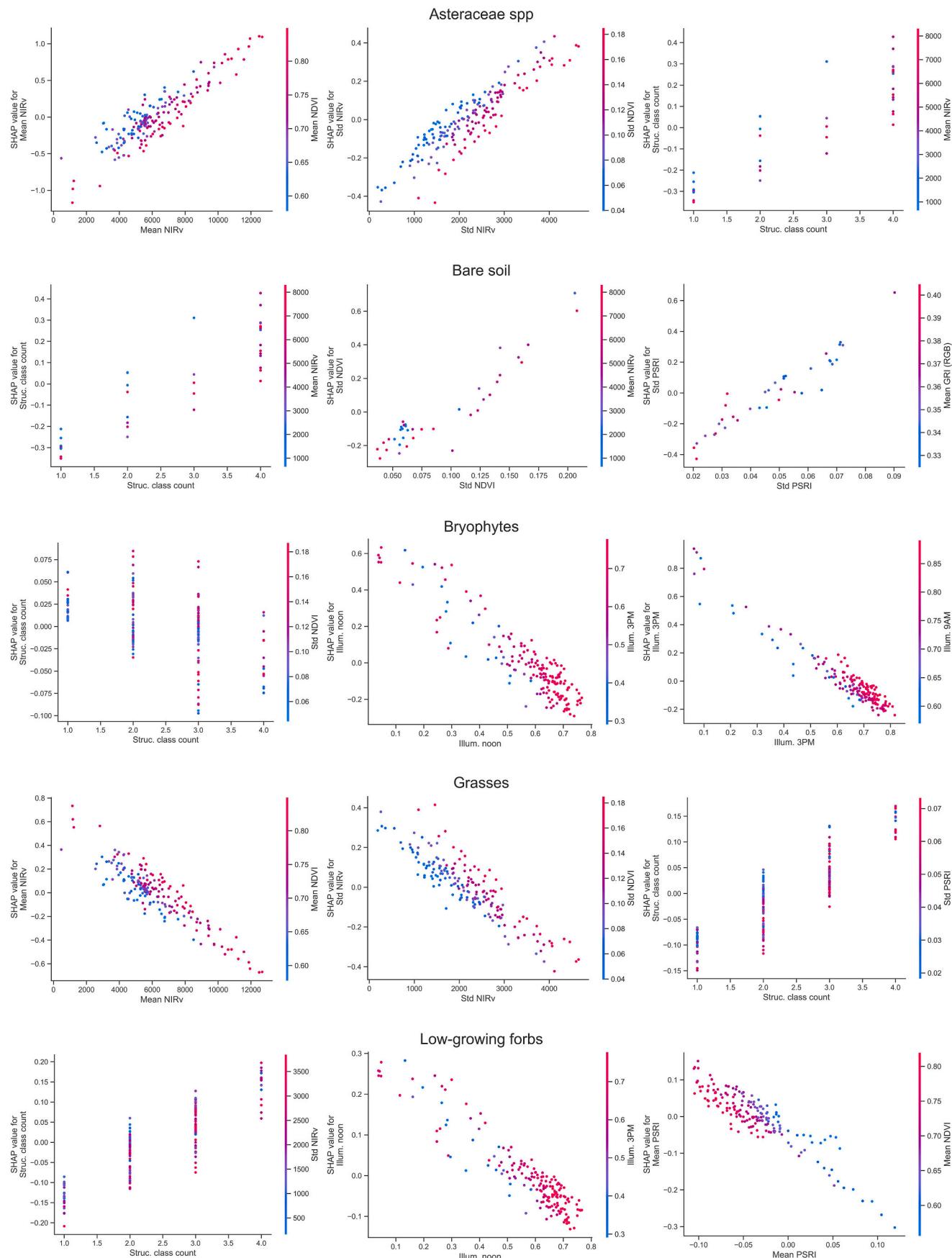
### Data availability

The authors do not have permission to share data.

### Acknowledgements

We thank all staff from TC Energy that initiated and supported our research, as well as members of the Integrated Remote Sensing Studio at UBC Forestry who contributed to data collection and the improvement of this manuscript. This research was supported by TC Energy and Natural Sciences and Engineering Research Council of Canada (NSERC) grant CRDPJ 537625–18.

## Appendices.



**Fig. I.** SHAP dependence plots for the three most important variables by model. The x-axis is the value of the feature, whereas the y-axis represents the impact on the model's output. The colors correspond to a second feature that may have an interaction with the main feature.

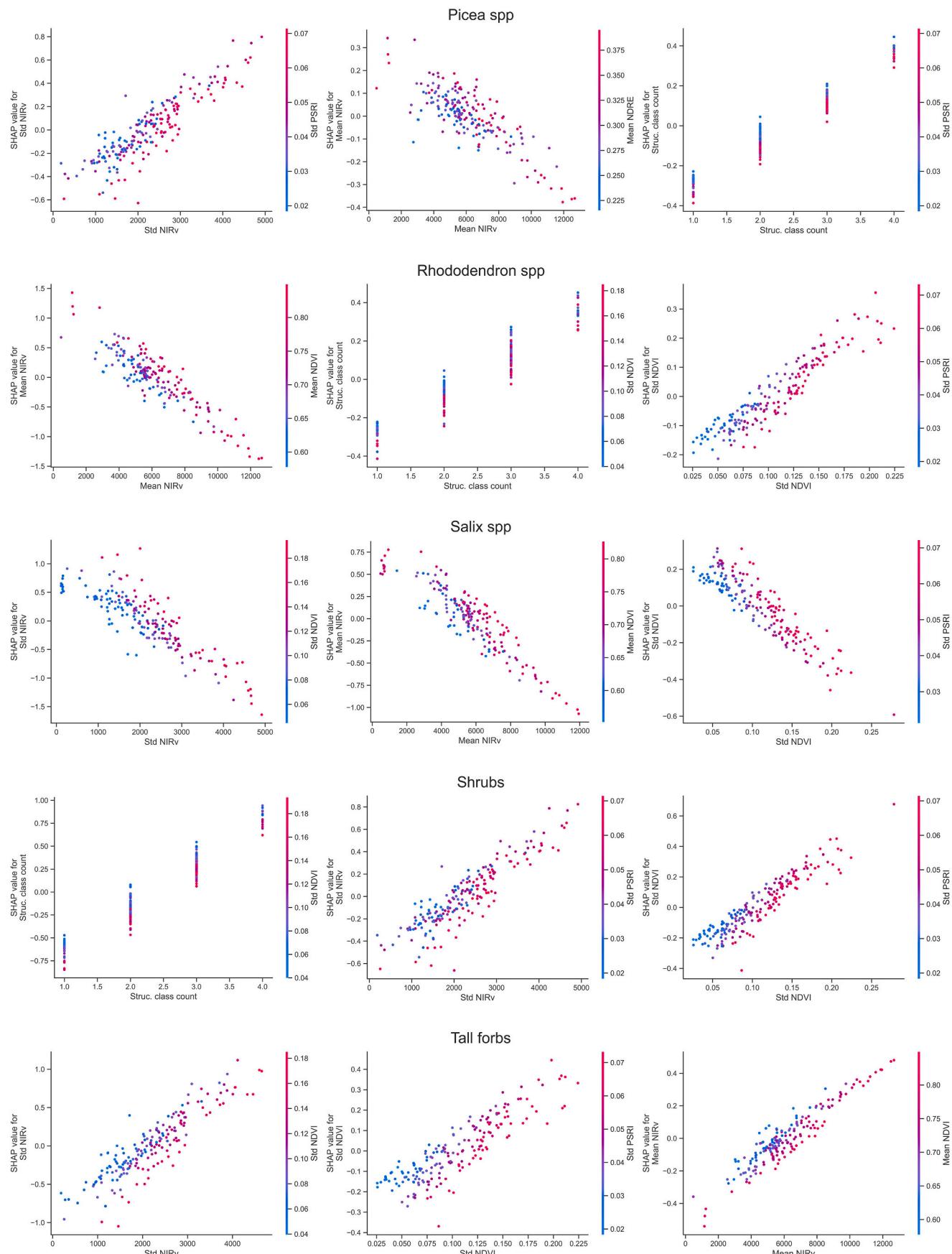


Fig. II. Additional SHAP dependence plots (1).

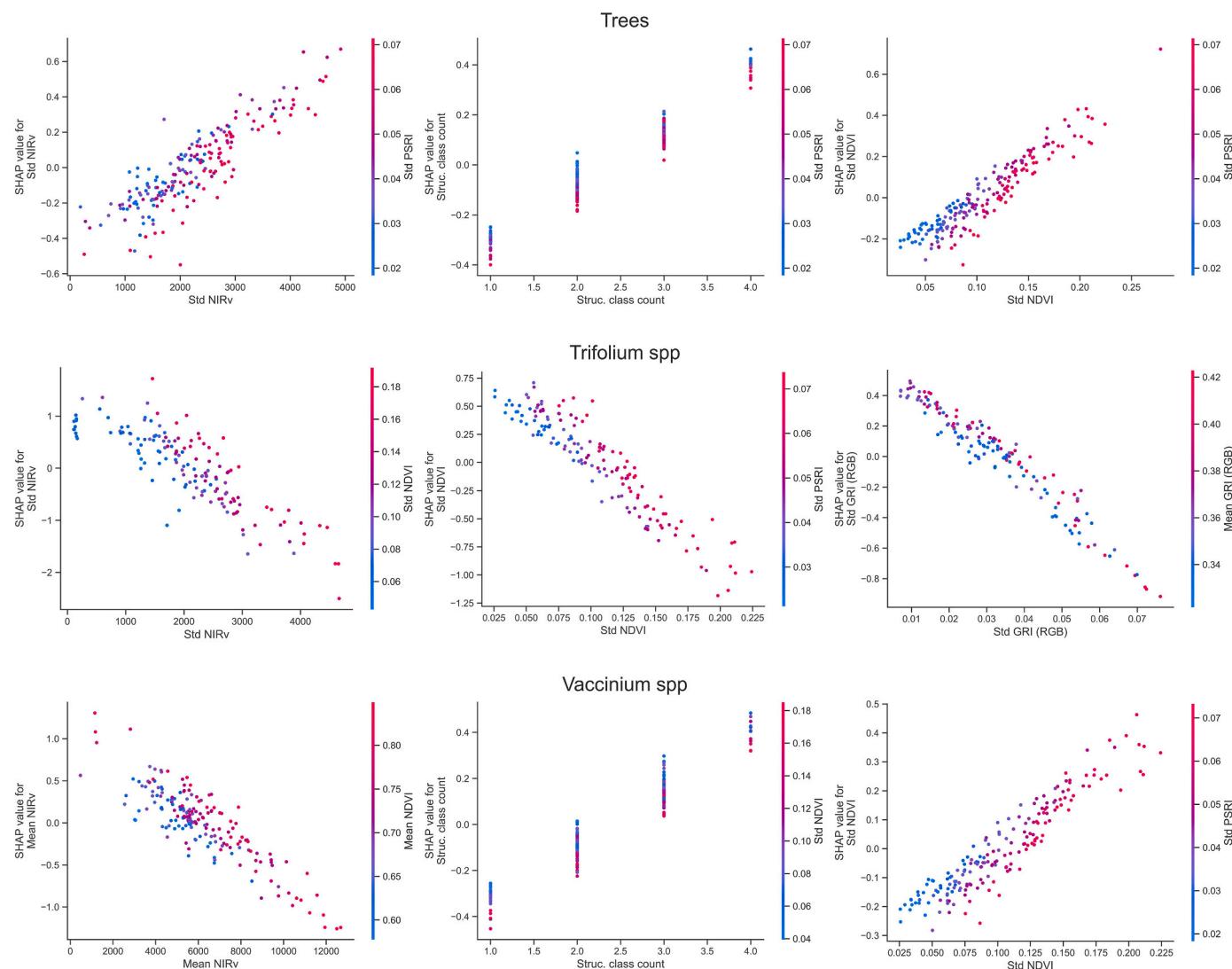


Fig. III. Additional SHAP dependence plots (2).

## References

- Alberta Agriculture and Forestry, 2020. Derived Ecosite Phase. Edmonton, AB, Canada.
- Alberta Environment and Parks, 2017. Alberta provincial digital elevation model [WWW Document]. URL: <https://open.alberta.ca/opendata/gda-c16469a2-5541-455c-bba0-63a24c0ff08a#summary>. (Accessed 18 September 2023).
- Alexandridis, T.K., Tamouridou, A.A., Pantazi, X.E., Lagopodi, A.L., Kashefi, J., Ovakoglu, G., Polychronos, V., Moshou, D., 2017. Novelty detection classifiers in weed mapping: *Silybum marianum* detection on UAV multispectral images. Sensors 17, 2–12. <https://doi.org/10.3390/s17092007>.
- Anders, N., Valente, J., Masselink, R., Keesstra, S., 2019. Comparing filtering techniques for removing vegetation from UAV-based photogrammetric point clouds. Drones 3, 61. <https://doi.org/10.3390/drones3030061>.
- Assiri, M., Sartori, A., Persichetti, A., Miele, C., Faelga, R.A., Blount, T., Silvestri, S., 2023. Leaf area index and aboveground biomass estimation of an alpine peatland with a UAV multi-sensor approach. GIScience Remote Sens. 60 <https://doi.org/10.1080/15481603.2023.2270791>.
- Badgley, G., Field, C.B., Berry, J.A., 2017. Canopy near-infrared reflectance and terrestrial photosynthesis. Sci. Adv. 3, 1–5. <https://doi.org/10.1126/sciadv.1602244>.
- Besag, J., 1974. Spatial interaction and the statistical analysis of lattice systems. J. R. Stat. Soc. Ser. B 36, 192–225. <https://doi.org/10.1111/j.2517-6161.1974.tb00999.x>.
- Bonan, G.B., Shugart, H.H., 1989. Environmental factors and ecological processes in boreal forests. Annu. Rev. Ecol. Systemat. 20, 1–28. <https://doi.org/10.1146/annurev.es.20.110189.000245>.
- Bowler, D.E., Bjorkman, A.D., Myers-smith, M.D.I.H., Navarro, L.M., Niamir, A., Supp, S.R., Waldock, C., Winter, M., Vellend, M., Blowes, S.A., Böhning-gaese, K., Bruehlheide, H., Elahi, R., Antão, L.H., Hines, J., Isbell, F., Jones, H.P., Magurran, A.E., Sarmento, J., Bates, A.E., 2020. Mapping human pressures on biodiversity across the planet uncovers anthropogenic threat complexes. People Nat 2, 380–394. <https://doi.org/10.1002/pan3.10071>.
- Bradley, B.A., Olsson, A.D., Wang, O., Dickson, B.G., Pelech, L., Sesnie, S.E., Zachmann, L.J., 2012. Species detection vs. habitat suitability: are we biasing habitat suitability models with remotely sensed data? Ecol. Model. 244, 57–64. <https://doi.org/10.1016/j.ecolmodel.2012.06.019>.
- Bradter, U., O'Connell, J., Kunin, W.E., Boffey, C.W.H., Ellis, R.J., Benton, T.G., 2020. Classifying grass-dominated habitats from remotely sensed data: the influence of spectral resolution, acquisition time and the vegetation classification system on accuracy and thematic resolution. Sci. Total Environ. 711, 134584–134598. <https://doi.org/10.1016/j.scitotenv.2019.134598>.
- Cao, J., Leng, W., Liu, K., Liu, L., He, Z., Zhu, Y., 2018. Object-based mangrove species classification using unmanned aerial vehicle hyperspectral images and digital surface models. Rem. Sens. 10, 89–109. <https://doi.org/10.3390/rs10010089>.
- Cha, Y.K., Shin, J., Go, B.G., Lee, D.S., Kim, Y.W., Kim, T.H., Park, Y.S., 2021. An interpretable machine learning method for supporting ecosystem management: application to species distribution models of freshwater macroinvertebrates. J. Environ. Manag. 291, 112719–112732. <https://doi.org/10.1016/j.jenvman.2021.112719>.
- Chadwick, A.J., Goodbody, T.R.H., Coops, N.C., Hervieux, A., Bater, C.W., Martens, L.A., White, B., Roeser, D., 2020. Automatic delineation and height measurement of regenerating conifer crowns under leaf-off conditions using uav imagery. Rem. Sens. 12, 1–26. <https://doi.org/10.3390/rs12244104>.

- Chadwick, A.J., Goodbody, T.R.H., Bater, C.W., Martens, L.A., Chadwick, A.J., Goodbody, T.R.H., Bater, C.W., Martens, L.A., Nijjten, R.J.G., Smith-Tripp, S., Grubinger, S., Irwin, L., Arkin, J., Hervieux, A., Coops, N.C., 2022. Best Practice Guide to Acquisition of 3D Imagery from RPAS. Vancouver, British Columbia.
- Coops, N.C., Goodbody, T.R.H., Cao, L., 2019. Four steps to extend drone use in research. *Nature* 572, 433–435.
- Cruzan, M.B., Weinstein, B.G., Grasty, M.R., Kohrn, B.F., Hendrickson, E.C., Arredondo, T.M., Thompson, P.G., 2016. Small unmanned aerial vehicles (Micro-UAVs, drones) in plant ecology. *Appl. Plant Sci.* 4, 1600041–1600052. <https://doi.org/10.3732/apps.1600041>.
- Cunliffe, A.M., Brazier, R.E., Anderson, K., 2016. Ultra-fine grain landscape-scale quantification of dryland vegetation structure with drone-acquired structure-from-motion photogrammetry. *Remote Sens. Environ.* 183, 129–143. <https://doi.org/10.1016/j.rse.2016.05.019>.
- Dandois, J.P., Ellis, E.C., 2013. High spatial resolution three-dimensional mapping of vegetation spectral dynamics using computer vision. *Remote Sens. Environ.* 136, 259–276. <https://doi.org/10.1016/j.rse.2013.04.005>.
- DeMars, C.A., Boutin, S., 2018. Nowhere to hide: effects of linear features on predator-prey dynamics in a large mammal system. *J. Anim. Ecol.* 87, 274–284. <https://doi.org/10.1111/1365-2656.12760>.
- Dyer, S.J., Neill, J.P.O., Wasel, S.M., Boutin, S., 2001. Avoidance of industrial development by woodland caribou. *J. Wildl. Manag.* 65, 531–542.
- Entekhabi, B.D., Njoku, E.G., Neill, P.E.O., Kellogg, K.H., Crow, W.T., Edelstein, W.N., Entin, J.K., Goodman, S.D., Jackson, T.J., Johnson, J., Kimball, J., Piepmeier, J.R., Koster, R.D., Martin, N., McDonald, K.C., Moghaddam, M., Moran, S., Reichle, R., Shi, J.C., Spencer, M.W., Thurman, S.W., Tsang, L., Zyl, J. Van, 2009. The soil moisture active passive (SMAP) mission. *Proc. IEEE* 98, 704–716. <https://doi.org/10.1109/JPROC.2010.2043918>.
- Fassnacht, F.E., Müllerová, J., Conti, L., Malavasi, M., Schmidlein, S., 2022. About the link between biodiversity and spectral variation. *Appl. Veg. Sci.* 25, 1–13. <https://doi.org/10.1111/avsc.12643>.
- Filicetti, A.T., Cody, M., Nielsen, S.E., 2019. Caribou conservation: restoring trees on seismic lines in Alberta, Canada. *Forests* 10, 185–204. <https://doi.org/10.3390/f10020185>.
- Forlani, G., Dall'Asta, E., Diotri, F., di Cellia, U.M., Roncella, R., Santise, M., 2018. Quality assessment of DSMs produced from UAV flights georeferenced with on-board RTK positioning. *Rem. Sens.* 10 <https://doi.org/10.3390/rs10020311>.
- Fromm, M., Schubert, M., Castilla, G., Linke, J., McDermid, G., 2019. Automated detection of conifer seedlings in drone imagery using convolutional neural networks. *Rem. Sens.* 11 <https://doi.org/10.3390/rs11212585>.
- Fujisada, H., Urai, M., Iwasaki, A., 2011. Advanced methodology for ASTER DEM generation. *IEEE Trans. Geosci. Rem. Sens.* 49, 5080–5091. <https://doi.org/10.1109/TGRS.2011.2158223>.
- Gann, G.D., McDonald, T., Walder, B., Aronson, J., Nelson, C.R., Jonson, J., Hallett, J.G., Eisenberg, C., Guariguata, M.R., Liu, J., Hua, F., Echeverría, C., Gonzales, E., Shaw, N., Decler, K., Dixon, K.W., 2019. International principles and standards for the practice of ecological restoration. *Restor. Ecol.* 27, S1–S46. <https://doi.org/10.1111/rec.13035>. Second edition.
- Gitelson, A.A., Merzlyak, M.N., 1994. Quantitative estimation of chlorophyll-a using reflectance spectra: experiments with autumn chestnut and maple. *J. Photochem. Photobiol. A* 22, 247–252. [https://doi.org/10.1016/1011-1344\(93\)06963-4](https://doi.org/10.1016/1011-1344(93)06963-4).
- Government of Alberta, 2021. Historical wildfire perimeter spatial data [WWW Document]. URL <https://www.alberta.ca/wildfire-maps-and-data.aspx>. (Accessed 18 September 2023).
- He, K.S., Bradley, B.A., Cord, A.F., Rocchini, D., Tuanmu, M.N., Schmidlein, S., Turner, W., Wegmann, M., Pettorelli, N., 2015. Will remote sensing shape the next generation of species distribution models? *Remote Sens. Ecol. Conserv.* 1, 4–18. <https://doi.org/10.1002/rse.2.7>.
- Ide, R., Oguma, H., 2010. Use of digital cameras for phenological observations. *Ecol. Inf.* 5, 339–347. <https://doi.org/10.1016/j.ecoinf.2010.07.002>.
- Isernburg, M., 2012. LAStools-efficient tools for LiDAR processing [WWW Document]. URL <https://rapidlasso.com/lastools/>. (Accessed 1 October 2020).
- James, G., Witten, D., Hastie, T., Tibshirani, R., 2021. An Introduction to Statistical Learning. Springer Texts. Springer, New York, NY, USA, pp. 129–195. <https://doi.org/10.1007/978-1-0716-1418-1>.
- Khanal, S., Fulton, J., Shearer, S., 2017. An overview of current and potential applications of thermal remote sensing in precision agriculture. *Comput. Electron. Agric.* 139, 22–32. <https://doi.org/10.1016/j.compag.2017.05.001>.
- Lee, P., Boutin, S., 2006. Persistence and developmental transition of wide seismic lines in the western Boreal Plains of Canada. *J. Environ. Manag.* 78, 240–250. <https://doi.org/10.1016/j.jenvman.2005.03.016>.
- Levin, S.A., 1992. The problem of pattern and scale in ecology. *Ecology* 73, 1943–1967. <https://doi.org/10.2307/1941447>.
- Lovitt, J., Rahman, M.M., McDermid, G.J., 2017. Assessing the value of UAV photogrammetry for characterizing terrain in complex peatlands. *Rem. Sens.* 9, 715–728. <https://doi.org/10.3390/rs9070715>.
- Merzlyak, M.N., Gitelson, A.A., Chivkunova, O.B., Rakitin, V.Y., 1999. Non-destructive optical detection of pigment changes during leaf senescence and fruit ripening. *Physiol. Plantarum* 106, 135–141. <https://doi.org/10.1034/j.1399-3054.1999.106119.x>.
- Morgan-Wall, T., 2023. Rayshader: create maps and visualize data in 2D and 3D [WWW Document]. URL <https://www.rayshader.com>.
- Nijjten, R.J.G., Coops, N.C., Goodbody, T.R.H., Pelletier, G., 2019. Examining the multi-seasonal consistency of individual tree segmentation on deciduous stands using digital aerial photogrammetry (DAP) and unmanned aerial systems (UAS). *Rem. Sens.* 11, 739–757. <https://doi.org/10.3390/rs11070739>.
- Nijjten, R.J.G., Coops, N.C., Watson, C., Theberge, D., 2021. Monitoring the structure of regenerating vegetation using drone-based digital aerial photogrammetry. *Rem. Sens.* 13, 1942–1965. <https://doi.org/10.3390/rs13101942>.
- Parks, Alberta, 2015. Natural Regions and Subregions of Alberta: A Framework for Alberta's Parks. Alberta Tourism, Parks and Recreation. Edmonton, AB, Canada.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830.
- Pengelly, C.J., Cartar, R.V., 2010. Effects of variable retention logging in the boreal forest on the bumble bee-influenced pollination community, evaluated 8–9 years post-logging. *For. Ecol. Manage.* 260, 994–1002. <https://doi.org/10.1016/j.foreco.2010.06.020>.
- Poirier, N., Baleux, F., Calastren, C., 2022. The mapping of forested archaeological sites using UAV LiDAR. A feedback from a south-west France experiment in settlement & landscape archaeology. *HAL* 4, 1–24. <https://doi.org/10.21494/ISTE.OP.2020.0556>.
- Prach, K., Durigan, G., Fennessy, S., Overbeck, G.E., Torezan, J.M., Murphy, S.D., 2019. A primer on choosing goals and indicators to evaluate ecological restoration success. *Restor. Ecol.* 27, 917–923. <https://doi.org/10.1111/rec.13011>.
- Pýšek, P., Hulme, P.E., Simberloff, D., Bacher, S., Blackburn, T.M., Carlton, J.T., Dawson, W., Essl, F., Foxcroft, L.C., Genovesi, P., Jeschke, J.M., Kühn, I., Liebhöfer, A.M., Mandrak, N.E., Meyerson, L.A., Pauchard, A., Pergl, J., Roy, H.E., Seebens, H., van Kleunen, M., Vilà, M., Wingfield, M.J., Richardson, D.M., 2020. Scientists' warning on invasive alien species. *Biol. Rev.* 95, 1511–1534. <https://doi.org/10.1111/brv.12627>.
- Robinson, J.M., Harrison, P.A., Mavoa, S., Breed, M.F., 2022. Existing and emerging uses of drones in restoration ecology. *Methods Ecol. Evol.* 13, 1899–1911. <https://doi.org/10.1111/2041-210X.13912>.
- Ruiz-Jaén, M.C., Aide, T.M., 2005. Vegetation structure, species diversity, and ecosystem processes as measures of restoration success. *For. Ecol. Manage.* 218, 159–173. <https://doi.org/10.1016/j.foreco.2005.07.008>.
- Sanderson, L.A., McLaughlin, J.A., Antunes, P.M., 2012. The last great forest: a review of the status of invasive species in the North American boreal forest. *Forestry* 85, 329–340. <https://doi.org/10.1093/forestry/cpr022>.
- Shen, X., Cao, L., 2017. Tree-species classification in subtropical forests using airborne hyperspectral and LiDAR data. *Rem. Sens.* 9, 1180–1204. <https://doi.org/10.3390/rs9111180>.
- Skidmore, A.K., Coops, N.C., Neinavaz, E., Ali, A., Schaepman, M.E., Paganini, M., Kissling, W.D., Viheriära, P., Darvishzadeh, R., Feilhauer, H., Fernandez, M., Fernández, N., Gorelick, N., Geijzendorffer, I., Heiden, U., Heurich, M., Hobern, D., Holzwarth, S., Muller-Karger, F.E., Van De Kerchove, R., Lausch, A., Leitão, P.J., Lock, M.C., Mücher, C.A., O'Connor, B., Rocchini, D., Turner, W., Vis, J.K., Wang, T., Wegmann, M., Wingate, V., 2021. Priority list of biodiversity metrics to observe from space. *Nat. Ecol. Evol.* 5, 896–906. <https://doi.org/10.1038/s41559-021-01451-x>.
- Smith, K.G., Ficht, E.J., Hobson, D., Sorensen, T.C., Hervieux, D., 2000. Winter distribution of woodland caribou in relation to clear-cut logging in west-central Alberta. *Can. J. Zool.* 78, 1433–1440. <https://doi.org/10.1139/cjz-78-8-1433>.
- TC Energy, 2014. Restoration Plan for the Chinchaga Lateral Loop No. 3 (Chinchaga Section). Calgary, AB, Canada.
- TC Energy, 2020. Year Three Caribou Monitoring Report Northwest Mainline Expansion Project (NWML), Leismer-Kettle River Crossover Project (LKXO), and Chinchaga Lateral Loop No. 3 (Chinchaga) Project. Calgary, AB, Alberta.
- Tu, Y.H., Phinn, S., Johansen, K., Robson, A., Wu, D., 2020. Optimising drone flight planning for measuring horticultural tree crop structure. *ISPRS J. Photogrammetry Remote Sens.* 160, 83–96. <https://doi.org/10.1016/j.isprsjprs.2019.12.006>.
- Tucker, C., 1979. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens. Environ.* 8, 127–150.
- van Iersel, W., Straatsma, M., Middelkoop, H., Addink, E., 2018. Multitemporal classification of river floodplain vegetation using time series of UAV images. *Rem. Sens.* 10, 1144–1162. <https://doi.org/10.3390/rs10071144>.
- Villoslada, M., Bergamo, T.F., Ward, R.D., Burnside, N.G., Joyce, C.B., Bunce, R.G.H., Sepp, K., 2020. Fine scale plant community assessment in coastal meadows using UAV based multispectral data. *Ecol. Indicat.* 111, 105979–105992. <https://doi.org/10.1016/j.ecolind.2019.105979>.
- Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., Walt, S.J., van der Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C., Polat, İ., Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A.H., Pedregosa, F., van Mulbregt, P., 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* 17, 261–272. <https://doi.org/10.1038/s41592-019-0686-2>.

- Watts, C.L., Hatch, C.E., Wicks, R., 2023. Mapping groundwater discharge seeps by thermal UAS imaging on a wetland restoration site. *Front. Environ. Sci.* 10, 1–16. <https://doi.org/10.3389/fenvs.2022.946565>.
- Wheatley, M., Johnson, C., 2009. Factors limiting our understanding of ecological scale. *Ecol. Complex.* 6, 150–159. <https://doi.org/10.1016/j.ecocom.2008.10.011>.
- Willoughby, M.G., Downing, D.J., Meijer, M., 2016. *Guide to Ecological Sites of the Lower Boreal Highlands*. Edmonton, AB, Canada.
- Willoughby, M.G., Archibald, J.H., Klappstein, G.D., Corns, I.G.W., Beckingham, J.D., France, T.L., Downing, D.J., 2018. *Guide to Ecological Sites of the Upper Foothills Subregion*. Edmonton, AB, Canada.
- Young, R.E., Gann, G.D., Walder, B., Liu, J., Cui, W., Newton, V., Nelson, C.R., Tashe, N., Jasper, D., Silveira, F.A.O., Carrick, P.J., Hägglund, T., Carlsén, S., Dixon, K., 2022. International principles and standards for the ecological restoration and recovery of mine sites. *Restor. Ecol.* 30, 1–47. <https://doi.org/10.1111/rec.13771>.