

Comparing ultra-high spatial resolution remote sensing methods in mapping peatland vegetation

Running title: Remote sensing of peatland vegetation

Aleksi Räsänen^{1,2}, Sari Juutinen¹, Eeva-Stiina Tuittila³, Mika Aurela⁴, Tarmo Virtanen¹

¹Ecosystems and Environment Research Programme, Faculty of Biological and Environmental Sciences, and Helsinki Institute of Sustainability Science (HELSUS), University of Helsinki, Helsinki, Finland

²Department of Geography, Norwegian University of Science and Technology, Trondheim, Norway

³School of Forest Sciences, University of Eastern Finland, Joensuu, Finland

⁴Finnish Meteorological Institute, Helsinki, Finland

Correspondence

1. Aleksi Räsänen, Ecosystems and Environment Research Programme, Faculty of Biological and Environmental Sciences, and Helsinki Institute of Sustainability Science (HELSUS), University of Helsinki, Helsinki, Finland
Email: aleksi.rasanen@helsinki.fi

Funding information: The research was supported by the Academy of Finland (grants 296423 [AR, SJ, TV], 296888 and 314799 [MA], and 287039 [EST])

Abstract

Questions

How to map floristic variation in a patterned fen in an ecologically meaningful way? Can plant communities be delineated with species data generalized into plant functional types? What are the benefits and drawbacks of the two selected remote sensing approaches in mapping vegetation patterns, namely: (1) regression models of floristically defined fuzzy plant community clusters, and (2) classification of predefined habitat types that combine vegetation and land cover information?

Location

Treeless 0.4 km² mesotrophic string-flark fen in Kaamanen, northern Finland.

Methods

We delineated plant community clusters with fuzzy c-means clustering based on two different inventories of plant species and functional type distribution. We used multiple optical remote sensing datasets, digital elevation models and vegetation height models derived from drone, aerial and satellite platforms from ultra-high to very-high spatial resolution (0.05 to 3 m) in an object-based approach. We mapped spatial patterns for fuzzy and crisp plant community clusters using boosted regression trees, and fuzzy and crisp habitat types using supervised random forest classification.

Results

Clusters delineated with species-specific data or plant functional type data produced comparable results. However, species-specific data for graminoids and mosses improved the accuracy of clustering in the case of flarks and string margins. Mapping accuracy was higher for habitat types (overall accuracy 0.71) than for fuzzy plant community clusters (R^2 values between 0.28 and 0.67).

Conclusions

For ecologically meaningful mapping of a patterned fen vegetation, plant functional types provide enough information. However, if the aim is to capture floristic variation in vegetation as realistically as possible, species-specific data should be used. Maps of plant community clusters and habitat types complement each other. While fuzzy plant communities appear to be floristically most accurate, crisp habitat types are easiest to interpret and apply to different landscape and biogeochemical cycle analyses and modeling.

Keywords: digital elevation model, drone, floristic analysis, fuzzy, northern boreal, object-based image analysis, plant community, plant functional types, UAS, UAV, very-high spatial resolution satellite imagery

1 Introduction

Northern peatlands have fine-scale spatial variation in vegetation, land cover and topography (Lovitt, Rahman, & McDermid, 2017; Middleton et al., 2012; Palace et al., 2018). This variation is also reflected in ecosystem functioning and responses (Lehmann et al., 2016). Spatial patterns of peatlands can be tracked with remotely sensed data, but it has been argued that the spatial resolution in common mapping approaches is too coarse (Palace et al., 2018). Ultra-high spatial resolution (UHSR) remote sensing, which provides data with cm-level pixel size, can reveal such patterns in vegetation composition that are lost in coarser resolution (Díaz-Varela, Calvo Iglesias, Cillero Castro, & Díaz Varela, 2018; Gonçalves et al., 2016; Lehmann et al., 2016; Mora, Vieira, Pina, Lousada, & Christiansen, 2015). In particular, the benefits of UHSR are evident in fragmented landscapes such as peatlands (Arroyo-Mora, Kalacska, Lucanus, Soffer, & Leblanc, 2017; Lehmann et al., 2016; Lovitt et al., 2017; Palace et al., 2018).

Vegetation patterns have been mapped using remote sensing with multiple methods and levels of detail. A typical approach has been to map land cover, vegetation, or habitat types, such as different forest or peatland types (Díaz-Varela et al., 2018; Foody, 1997; Gonçalves et al., 2016; Mora et al., 2015; Tapia, Stein, & Bijker, 2005). There have also been efforts to map the presence of a single plant species (Dudley, Dennison, Roth, Roberts, & Coates, 2015; Roth, Roberts, Dennison, Peterson, & Alonzo, 2015), plant species grouped to communities (Duff, Bell, & York, 2014; Middleton et al., 2012; Rapinel, Rossignol, Hubert-Moy, Bouzillé, & Bonis, 2018) or plant functional types (PFTs) (Cole, McMorro, & Evans, 2013; Harris, Charnock, & Lucas, 2015; Kattenborn et al., 2017; Schmidtlein, Feilhauer, & Bruehlheide, 2012; Ustin & Gamon, 2010). In PFTs, plant species are grouped based on their traits (Chapin III, Bret-Harte, Hobbie, & Zhong, 1996; Duckworth, Kent, & Ramsay, 2000; Hartley, MacBean, Georgievski, & Bontemps, 2017; Ustin & Gamon, 2010), and the focus is often on plant functions and biogeochemical properties. PFTs can be delineated in multiple different ways and they are a generalization of species data (Duckworth et al., 2000). Commonly, PFTs represent similar growth forms and responses to environmental and biological conditions (Chapin III et al., 1996; Duckworth et al., 2000; Ustin & Gamon, 2010).

In remote sensing applications, species, PFTs or communities are distinguished by their optical and structural features, e.g. spectral reflectance and height. Some PFTs (such as shrubs, forbs, graminoids and mosses) are easier and faster to record in the field when collecting large validation datasets (Harris et al., 2015), and it can be assumed that their spectral signature is similar to that of key single species (Cole et al., 2013). On the other hand, PFTs can be an oversimplification of the species data (Duckworth et al., 2000; Ustin & Gamon, 2010). We argue that the applicability of PFTs could be judged based on the similarity of plant community cluster membership, i.e., are the field study plots modeled to corresponding clusters when different input data is used?

In addition to the level of units targeted in mapping, the approach used to map certainty can also differ. Most commonly, maps presenting habitat types and vegetation patterns are crisp: each pixel in a final map represents a single habitat type (Foody, 1997), or each plant species is assigned to one plant community only (Rocchini, 2014). As pure habitat types rarely exist (Foody, 1997), fuzzy maps that indicate probabilities of habitat types are considered to reflect environmental variation in nature more realistically (e.g. de Klerk, Burgess, & Visser, 2018; Foody, 1997; Tapia et al., 2005). Similar to habitat types, there have been some attempts to model fuzzy plant communities instead of crisp ones, because species assemblages vary in space and consequently plant communities seldom have clear boundaries (Duff et al., 2014; Rapinel et al., 2018; Rocchini, 2014). However, comparisons among different remote

sensing-based methods (e.g. between land cover-based classifications and plant community-based regressions) are still lacking, especially when using fuzzy methods. The main novelty of this article is to compare above mentioned remote sensing-based vegetation mapping approaches at UHSR.

We compare different methods to delineate and map fuzzy and crisp plant communities and habitat types at UHSR in a patterned treeless peatland. Firstly, we assess if species-specific percentage cover data can be generalized with PFTs when clustering species communities. We hypothesize that a meaningful compromise could be a combination of PFTs and species (i.e. the data includes species-specific information only for some PFTs). Secondly, we compare two different approaches to map vegetation patterns in the peatland using remote sensing methods: (1) classification of predefined habitat types that combine vegetation and land cover information, and (2) regression of floristically defined fuzzy plant community clusters. We assess how these two approaches represent vegetation patterns. We hypothesize that plant communities are floristically more accurate, but habitat types are easier to map using remote sensing data.

2 Materials and methods

2.1 Study area

The studied peatland is a treeless mesotrophic patterned fen, located in Kaamanen, northern Finland (69.14°N, 27.27°E; 155 m a.s.l.) in a northern boreal vegetation zone and subarctic climate zone. The study area (0.4 km²) is characterized by a strong pattern of strings (less than 1 m high), dominated by dwarf shrubs, and wet flarks with sedges and brown mosses. A small stream runs through the study area; the riparian areas are characterized by tall sedge vegetation. The study area is surrounded by an upland pine forest, shrub-dominated peatland types in the ecotone between the upland forest and open peatland, and a small lake. Our study focus is the footprint area of an eddy covariance tower that has been measuring ecosystem carbon dioxide exchange since 1997 (see e.g. Aurela, Laurila, & Tuovinen, 2001; Aurela, Tuovinen, & Laurila, 1998; Maanavilja et al., 2011).

2.2 Overview of methods

We compared two different remote sensing based methods for mapping vegetation in the fen. We mapped predefined habitat types with random forest classification, and we mapped plant community clusters delineated from floristic data using boosted regression trees. We used field work inventory data collected in 2006 and 2017. Plot-based data collected in 2006 was used to delineate plant community clusters and compare how clusters change when the following species data generalization options are used: (1) species-specific data, (2) species-PFT-combination (called hereafter extended PFTs), and (3) PFTs. Plot data collected in 2017 was used for remote sensing-based mapping of plant community clusters and to validate remote sensing-based habitat type mapping. Transect data collected in 2017 was used to construct training data for the habitat type classification. During both years, field work was conducted by experts capable of identifying the local plant species. The study area has been visited frequently, and no obvious changes in spatial distribution of plants and vegetation have been observed between 2006 and 2017.

2.3 Vegetation and habitat type inventories

For delineating plant community clusters, we utilized species-specific vegetation data collected in 2006 (see Maanavilja et al., 2011). The data consist of 278 circular plots, 30 cm in diameter, where the projection cover (%) of vascular plant, moss and lichen species was visually estimated. The plots were drawn by systematic transect-based sampling at distances of 5 to 220 m from the eddy covariance tower (Riutta et al., 2007) (Appendix S1). Due to the lack of accurate enough location information, this data could not be used in remote sensing analysis.

Based on the vegetation survey in 2006 and a TWINSpan analysis (Hill, 1979), Maanavilja et al. (2011) divided vegetation in the field plots into four plant community types, i.e. *Carex–Scorpidium* wet flarks, *Trichophorum* tussock flarks, *Betula–Sphagnum* string margins and *Ericales–Pleurozium* string tops, but they did not construct maps of these types. We used these four plant community types as predefined habitat types when conducting field work in 2017 and later in our habitat type classification. Therefore, in our case, these four types were not based on floristic analysis but on visual interpretation of the vegetation in the field. We also included a fifth habitat type – riparian fen – found adjacent to the small brook (Table 1).

For the remote sensing analysis, we collected extended-PFT data with accurate georeferencing in 2017. A total of 202 square plots with 50 cm side length were used. The plots were sampled systematically at distances of 25 to 150 m from the flux tower in cardinal, intercardinal and secondary intercardinal directions (Appendix S1). In these plots, forbs and shrubs were identified to the species level. In addition, we identified the following species groups or plant functional groups: *Eriophorum* spp., *Trichophorum* spp., Poaceae, other graminoids (primarily *Carex* spp.), feather mosses, wet brown mosses, *Sphagnum* spp. and lichen. The projection cover (%) of these species and PFTs was visually estimated similarly to the 2006 inventory. Each plot was also assigned to one of the five habitat types based on visual interpretation. The plots were situated between 0.6 to 4.1 m from points geolocated with a Trimble R10 GPS device with ± 5 cm accuracy, and their exact distance and angle from the GPS point were measured. The location of the plots in the drone image was double-checked to verify that the vegetation description and visual interpretation in the field matched that in the drone image.

We collected transect data of habitat type distribution in 2017 (Appendix S1). Nine 250 m transects were set up in cardinal, intercardinal, and secondary intercardinal directions from the flux tower and marked using a Trimble R10 GPS device with ± 5 cm accuracy, at 25 to 100 m intervals. Vegetation along the transects was classified into the five habitat types (Table 1) based on visual interpretation, and transitions between the habitat types were located with a Garmin eTrex 30 GPS device.

2.4 Remote sensing data and its preprocessing

We used optical drone, aerial, and satellite imagery, as well as digital elevation and digital surface models at several spatial resolutions (Table 2). We utilized different remote sensing datasets to capture information about the factors that are known to affect accuracy in remote sensing-based mapping of vegetation; namely topography, vegetation height and phenology, as well as spectral properties and their spatial variability (Dudley et al., 2015; Räsänen, Kuitunen, Tomppo, & Lensu, 2014). We used a drone image with 5 cm spatial resolution as the main data in the remote sensing analyses while coarser spatial resolution datasets were ancillary data to increase the accuracy of vegetation mapping. A DJI phantom 4 pro drone flight was conducted, and we georeferenced these data using 14 ground control points measured with a Trimble R10 GPS device with ± 5 cm accuracy. We computed an image mosaic, as well as a digital terrain and digital surface models with Pix4d software. We also calculated a vegetation height

model by subtracting the digital terrain model from the digital surface model. In addition to the drone image, we used coarser resolution aerial orthophoto and LiDAR data from the National Land Survey of Finland. To capture topographical variability, we calculated topographic position indices with different neighborhood distances (Guisan, Weiss, & Weiss, 1999) and topographic wetness index (Böhner & Selige, 2006) for the digital elevation models. To capture phenological dynamics, we included four different PlanetScope satellite images from summer 2017. The spatial accuracy of these orthorectified images was verified using visual interpretation. More spectral information was obtained from a WorldView-2 satellite image, which was orthorectified with the aerial orthophoto and 18 ground control points. For aerial and satellite images, we calculated the following spectral indices: normalized difference vegetation index (Rouse, Haas, Schell, & Deering, 1973), normalized difference water index (McFeeters, 1996) and red-green index (Coops, Johnson, Wulder, & White, 2006) to gather information about the spatial variability in the amount of green vegetation and water.

2.5 Classification of habitat types

We conducted a supervised classification, in which we constructed habitat type maps of the study area with the aid of field inventory-based transect data and remote sensing datasets. In total, we classified ten habitat types, but in this article, we focus on five of them which are found in studied fen (Table 1). When input data pixel size (in our case 5 cm) is smaller than the size of the vegetation patches in the study area, object-based image analysis approaches are recommended (Blaschke et al., 2014). Therefore, we combined full lambda schedule (FLS) segmentation with supervised random forest (RF) classification (Breiman, 2001). FLS is a region-based segmentation algorithm in which pixels are merged with their neighboring pixels to form homogeneous and contiguous segments (i.e. patches). In RF step, we classified segments into habitat types. RF is an ensemble of multiple classification trees, and final decisions are made based on a majority vote. In each tree, approximately two thirds of the data is used for training and the rest one third is called out-of-bag data and is used for evaluation.

First, we segmented the drone image with the FLS method. After visual interpretation of segmentations with different parameter combinations, we set the average size of the segment to 5 m² and gave the relative weights 0.7, 0.5, 0.3 and 0.3 to color, texture, size and shape, respectively; hence, we formed such segments that had relatively homogeneous spectral and textural characteristics. Second, to aid classification of the segments, we calculated 236 features from 65 remote sensing dataset layers for each segment (Table 2). The features consisted of mean and standard deviation value of the different dataset pixels within each segment. Additionally, to track variability of spectral properties within segments, we calculated 13 grey-level co-occurrence matrix textural features (Haralick, Dinstein, & Shanmugam, 1973) with 32 quantization levels and for first and second order neighbors for the drone spectral bands. Third, we constructed training data for classification with the help of the 2017 transect field data and visual interpretation of the drone image by selecting 134 to 389 training segments for each class (Appendix S1). Fourth, as the performance of RF classification can be boosted when reducing the number of features (Räsänen et al., 2014), we selected features using the Boruta algorithm (Kursa & Rudnicki, 2010). In Boruta, several (in this case 1000) RF runs are conducted and the importance of features is compared against each other and shadow features with random values. We determined the optimal number of features by comparing out-of-bag error rates of classifications with different amount of the most important features. Fifth, we classified the dataset using RF classification with the top 100 most important features by using 500 trees and tested ten features at each tree node. We computed (i) *a crisp random forest output*, in which each segment was classified to the habitat type class with the largest proportion of the individual tree votes, (ii) *fuzzy maps of each habitat type* by calculating the proportion

of tree votes, (iii) a *confusion map*, which was a ratio between the proportion of votes of the habitat type with largest proportion of votes and habitat type with second largest proportion of votes (Tapia et al., 2005). Sixth, we calculated the accuracy of the classification by using the 2017 vegetation plots as the reference data. Analyses were carried out in R (R Core Team, 2018) using packages randomForest (Liaw & Wiener, 2002), Boruta (Kursa & Rudnicki, 2010) and EBImage (Pau, Fuchs, Sklyar, Boutros, & Huber, 2010) as well as in ArcGIS 10.3.1 and ERDAS Imagine 2014.

2.6 Fuzzy clustering of plant communities

We conducted a regression analysis, in which we constructed plant community cluster maps of the study area with the aid of field plot data and remote sensing datasets. First, we clustered the species-specific data collected in 2006 and compared how the clusters differ from each other when using the following modifications of the data: (1) using the raw data, i.e. species-specific data, (2) extended PFTs (i.e. similar classification as in 2017 data), (3) delineation to the following PFTs: evergreen dwarf shrubs, deciduous dwarf shrubs, *Betula nana*, *Salix* spp., deciduous trees, forbs, graminoids, wet brown mosses, feather mosses, *Sphagnum* spp. and lichens. Second, we clustered the 2017 data and predicted the spatial patterns of the clusters using boosted regression trees (BRTs) (Elith, Leathwick, & Hastie, 2008; Ridgeway, 1999). BRTs are an ensemble of regression trees which are gradually boosted by minimizing the loss function. The methodological steps are described in greater detail below.

Plant communities were clustered following Duff et al. (2014). First, to measure dissimilarities between the field inventory plots, we calculated a distance matrix of the species or PFT data using Bray-Curtis distances (Bray & Curtis, 1957). Second, to reduce noise in the distance matrix and to simplify the data, we carried out a principal coordinates analysis (PCoA) of the distance matrix (Gower, 1966) and chose the optimal number of axes by visually estimating the point where the scree plot of coordinate influence levels off. Third, we chose the optimal number of clusters with cascading K-means clustering of the most important PCoA axes and Calinski criterion (Calinski & Harabasz, 1974). Fourth, we clustered the data using fuzzy c-means clustering with fuzziness parameter set to 2 (Bezdek, 1981). In the clustering, each plot was given a membership value for each cluster as well as a crisp cluster result. Fifth, we evaluated the indicator species/PFTs for each crisp cluster using Dufrene-Legendre indicator value analysis (Dufrêne & Legendre, 1997). Sixth, using permutational multivariate analysis of variance (Anderson, 2001), we tested if the differences between clusters in the species/PFT distance matrix are statistically significant and how large amount of variation of the distance matrix the cluster membership values and pairwise comparisons of crisp clusters explain.

Seventh, we predicted spatial patterns of the cluster membership values using BRTs. The BRTs were applied to predict the cluster membership values to each FLS segment. Plot-based vegetation data collected in 2017 were used as training and model evaluation data. Features listed in Table 2 were used for each plot and segment. For each BRT model, we included the features deemed important by Boruta analysis with 100 RF runs (Kursa & Rudnicki, 2010). We assumed a Gaussian distribution, and evaluated model performance and tuned parameters using ten-fold cross-validation. We chose the optimal parameter combinations based on RMSE values and a grid search. After initial parameter value screening, we tested the following parameter values: number of regression trees in the model $\in \{1000, 2000, \dots, 10000\}$, number of splits in each tree (i.e. interaction depth or tree complexity) $\in \{1, 2, \dots, 5\}$, shrinkage of added tree contribution in the model (i.e. learning rate) $\in \{0.001, 0.002, \dots, 0.005\}$, and minimum number of observations in the trees' terminal nodes $\in \{1, 2, \dots, 5\}$. For further description of these parameters, see Elith et al. (2008) and Ridgeway (2017). We produced maps of (i) fuzzy

membership values of each cluster, (ii) crisp cluster map, in which segments were mapped to the cluster with the highest membership value, and (iii) a confusion map, which was a ratio between membership value of the cluster with highest value and the cluster with second highest membership value (Tapia et al., 2005). These analyses were carried out in R (R Core Team, 2018) with packages *vegan* (Oksanen et al., 2018), *pairwiseAdonis* (Martinez Arbizu, 2019), *e1071* (Meyer, Dimitriadou, Hornik, Weingessel, & Leisch, 2017), *labdsv* (Roberts, 2016), *gbm* (Ridgeway, 2017), *caret* (Kuhn et al., 2018) and *Boruta* (Kursa & Rudnicki, 2010).

3 Results

3.1 Effect of level of detail in plant community clustering

Optimal number of clusters was five for each of the datasets (Appendix S2). The level of complexity was reflected in the number of PCoA axes needed to explain variation in vegetation composition; four for species-specific and extended-PFT data, and three for PFT data (Appendix S3).

Cluster 1 (i.e. wet brown moss cluster) was characterized by wet brown mosses, especially *Scorpidium scorpioides*. In the species-specific clustering, some *Carex* species were also representative for this community. The most important indicator species or PFTs for cluster 2 (deciduous tall shrub) were *Sphagnum* mosses and *Betula nana*. Clusters 3 (graminoid) and 4 (graminoid/*Utricularia*) were characterized by graminoid species, and there was some variation between the datasets in terms of which indicators were significant in these clusters. While the 2017 data differentiated *Carex* spp. and *Utricularia* spp. clusters, in the 2006 data there were different clusters for *Carex* spp. and *Trichophorum* spp. Finally, cluster 5 (evergreen dwarf shrub) was relatively similar when using different datasets, and was dominated by string-top species such as *Rubus chamaemorus*, evergreen shrubs and feather mosses.

The extremes of the wetness gradient, namely wet brown moss and evergreen dwarf shrub clusters, were similar in all of the clustering options, while there was more variation in the intermediate clusters (Table 3 and Appendix S4). Clusters based on the species-specific data commonly differed from those based on the two datasets with fewer details (i.e. PFT and extended PFT). These two more coarse datasets, however, yielded relatively similar clusters. We decided to cluster the 2017 data using extended PFTs because those were slightly more similar to the species clusters than the PFT clusters. We found that indicator species/PFTs in the 2017 clusters were quite similar to the clusters in the 2006 extended-PFT data (Appendix S4). According to the permutational multivariate analysis of variance, clusters differed from each other (P-value < 0.001 in each pairwise comparison for each dataset, Appendix S5), and cluster membership values explained considerable amount of the variation in species/PFT distance matrix (43% with species-specific 2006 data, 63% with extended-PFT 2006 data, 66% with PFT 2006 data, and 66% with 2017 data).

3.2 Maps of habitat types and plant communities

Habitat type mapping had an overall accuracy of 72% (Table 4). Wet flark, string top and riparian fen had the highest classification accuracies, while string margins were often confused with string tops, and tussock flarks with wet flarks. The same pattern also appeared when using the fuzzy approach in habitat type mapping; wet flark, string top and riparian fen had a clear pattern, and had the clear majority of votes in many areas (Appendix S6). String margin and tussock flark, on the other hand, usually had a low proportion of votes in segments even if they had the largest proportion of votes.

The BRT model fits were best for the evergreen dwarf shrub cluster (R^2 0.67, RMSE 0.20), followed by graminoid/*Utricularia* (R^2 0.55, RMSE 0.23), deciduous tall shrub (R^2 0.48, RMSE 0.18), graminoid (R^2 0.34, RMSE 0.17), and wet brown moss clusters (R^2 0.27, RMSE 0.23).

Cross-comparison of the habitat type and plant community maps showed that the dominant plant community in string tops was the evergreen dwarf shrub cluster (Fig. 1). Shrub clusters dominated in string margins, whereas wet brown moss and graminoid clusters had the highest membership values in wet and tussock flarks. Finally, in the riparian fen, the graminoid cluster was dominant. These findings were confirmed by visual interpretation of fuzzy plant community maps (Appendix S7).

The crisp outputs of habitat type and plant community maps revealed the string-flark pattern of the fen (Fig. 2). Nevertheless, the confusion in the plant community map was generally much higher than in the habitat type map, which was largely explained by the difference between classification and regression methods. In the habitat type map, confusion was the largest in the margins between strings and flarks, whereas in the plant community map, confusion was relatively high in all areas and lowest in areas dominated by evergreen dwarf shrub cluster. In the habitat type map (Fig. 2a), the landscape was dominated by large wet flark patches, followed by string tops (Table 5). When looking at the crisp clusters (Fig. 2c), graminoid, graminoid/*Utricularia*, and evergreen dwarf shrub clusters covered the largest extent and had largest patch sizes (Table 5).

4 Discussion and conclusions

Based on our results, the indicator species or PFTs were more or less similar when using different input data options (Appendix S4). However, there were still some differences between the clusters and modeled membership values, especially when species-specific input data was compared with generalized input data (Table 3). Contrary to our hypothesis, extended PFTs (i.e. combination of species and PFT information) provided little benefit in comparison to PFTs. Therefore, it is important that in addition to forbs and shrubs, graminoids, mosses and lichens should also be identified at the species level when carrying out floristic analyses in patterned peatlands. This result is in line with earlier findings concerning peatland vegetation mapping that show how different moss and graminoid species locate differently on ordination axes (Harris et al., 2015; Middleton et al., 2012). However, in other types of environments with fragmented vegetation patterns but higher forb and shrub species richness, such as tundra, an extended-PFT approach can be more valuable (Mikola et al., 2018; Virtanen & Ek, 2014). Overall, the value of the PFT generalization depends on the studied system and the choice of traits which are used to form PFTs.

The 2006 field work data was sampled with circular plots, whereas in the 2017 inventory, square plots were used. In previous studies, it has been found that plot shape has an effect on how many species are observed within the plot (Bacaro et al., 2015; Güler et al., 2016). Thus, the differently shaped plots might have had an effect on the observed species and PFTs, their %-coverages and modeled clusters, but as indicator species and PFTs were closely similar for both datasets (see Appendix S4), we do not see that the differing plot shapes had any major impact on our results.

The driest and most compositionally unique cluster dominated by evergreen dwarf shrubs was the least sensitive to the input data (Table 3). This cluster also had the highest R^2 in remote sensing-based regressions, and corresponded with string top habitat type in terms of spatial patterns, average patch size

and total extent. String tops also had a considerably high classification accuracy (Figs. 1&3, Table 4), and could thus be modeled well both using plant community clusters and habitat types. In this case, PFTs were a relevant generalization.

For other clusters and plant functional types, more information is lost, and species-specific input data is more preferable. In our case, these include wetter habitat types (i.e. flarks, string margins and riparian areas). First, there was variation in how much information was lost when PFTs rather than species were used in plant community delineation. The wettest cluster dominated by wet brown mosses was consistent regardless of the input data used, but deciduous tall shrub cluster and graminoid clusters had a larger variation (Table 3). In regression models using remote sensing datasets, R^2 values for these clusters were low to moderate. Instead, in habitat type classification, wet flarks and riparian fens could be classified with reasonably high accuracy, but accuracies were lower for tussock flarks and string margins (Table 4). These disparities in the success of mapping different habitat types and plant community clusters are supported by the fact that peatlands have both continuous and abrupt transitions between vegetation patterns (Middleton et al., 2012).

In the patterned fen, the two extremes along the moisture gradient had the most specialized plant species. However, the intermediate habitat types displayed a clear overlap of species within communities, resulting in continuous transitions and less clear borders. This highlights the continuous and fuzzy nature of fen vegetation. For instance, when cross-comparing habitat type classification and plant community cluster regression, it becomes evident that string margins and riparian areas have partly similar vegetation yet distinct topographical and land cover patterns. As another example, water is commonly a dominant feature in flarks. The water cover is better mapped with habitat type classification, which accounts for land cover. At the same time, habitat type classification misses fine-scale vegetation transitions in flarks. This is exemplified by average patch sizes and total extent of classes (Table 5). On the one hand, wet flarks had large patches and a large total extent according to habitat type classification. On the other hand, flark vegetation clusters defined according to plant community composition had generally smaller patches and smaller total extent. Therefore, the cluster regressions suggest that wet flarks – which actually have largely uniform land cover – may not be floristically that uniform. This highlights how crisp land cover classifications often miss fine-scale variation in vegetation patterns (Harris et al., 2015).

Habitat type classification and plant community regressions each have strengths and weaknesses. On the one hand, as hypothesized, classification accuracy for mapping habitat types was higher. This is probably caused by the fact that habitat type classification was developed to detect differences between types with remote sensing methods. On the other hand, habitat type maps do not reveal the fuzziness of fen vegetation patterns, which is better modeled with fuzzy plant community clusters (Fig. 2, Appendices S5&S6). Hence, the approaches complement rather than compete against each other. These approaches could be further accompanied by and compared to other remote sensing-based vegetation mapping approaches, such as direct mapping of multiple different types of PFTs (Cole et al., 2013; Schmidtlein et al., 2012), mapping of floristic gradients (i.e. ordination axes) (Harris et al., 2015), and mapping of dominant species (Roth et al., 2015). In each of these approaches, methods that capture continuity in species distribution, such as regressions and fuzzy methods, should be prioritized (Duff et al., 2014; Rapinel et al., 2018; Rocchini, 2014).

The benefits of UHSR drone images were obvious in mapping vegetation patterns in a patterned fen. Some of the vegetation patterns that could be mapped with object-based UHSR methods could not be seen in visual interpretation of very high spatial resolution satellite imagery. Our work thus further

highlights the advantages of UHSR remote sensing approaches for mapping not only vegetation patterns (Arroyo-Mora et al., 2017; Lehmann et al., 2016; Palace et al., 2018) but also spatial dynamics in biogeochemical processes such as carbon cycling in peatlands with patchy vegetation and topography (Lehmann et al., 2016). However, we well understand that UHSR image based classifications cannot be produced for large geographical areas due to limitations of image acquisition and data processing. In future studies, there is a need to consider what kind of resolution and data is needed in mapping vegetation patterns in environments with fine-scale variability. In these studies, comparisons could be made as to (1) what types of input data should be included, (2) what kind of spatial and spectral resolution these data should have, and (3) what kind of segmentation scale should be used. When remote sensing based datasets are used in other studies, such as biogeochemical process modeling and upscaling, it is relevant to know what details are lost when coarser spatial resolution datasets are used.

Finally, the way how PFTs are categorized affects how well they can be detected using remote sensing methods and how well they represent species' variable traits and growth forms (Duckworth et al., 2000; Ustin & Gamon, 2010). Therefore, PFTs may complement but not replace species-specific data (Duckworth et al., 2000), because mapping at the species level provides the possibility to later generalize data to PFTs using different types of groupings, if knowledge on species traits increases or different traits appear important. Species-level identification allows also to test different types of PFT categorization when analyzing the data. However, identifying plants at the species level takes much more time in the field, which consequently reduces the number of sampled plots, as fieldwork time is always limited. Moreover, species-level identification often requires that trained botanist conduct or guide the field work. Therefore, careful consideration must be taken when deciding whether to sample a smaller number of plots identified at the species level or more plots identified at PFT or another similar level of generalization. In addition, some species are almost impossible to determine reliably in the field, especially if field work is not conducted during an optimal phenological stage or if plants are of very small size.

Acknowledgements

We thank Terhi Riutta, Holtti Hakonen and Jani Antila for carrying out the fieldwork in 2006 and 2017, and Kari Mäenpää and Pauli Heikkinen for conducting the drone flight and image pre-processing.

Data accessibility

Data is available upon request from the authors.

References

- Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecology*, 26(1), 32-46. doi:10.1046/j.1442-9993.2001.01070.x
- Arroyo-Mora, J. P., Kalacska, M., Lucanus, O., Soffer, R., & Leblanc, G. (2017). *Spectro-spatial relationship between UAV derived high resolution DEM and SWIR hyperspectral data: Application to an ombrotrophic peatland*. Paper presented at the Proceedings of SPIE - The International Society for Optical Engineering.
- Aurela, M., Laurila, T., & Tuovinen, J. P. (2001). Seasonal CO₂ balances of a subarctic mire. *Journal of Geophysical Research Atmospheres*, 106(2), 1623-1637.
- Aurela, M., Tuovinen, J. P., & Laurila, T. (1998). Carbon dioxide exchange in a subarctic peatland ecosystem in northern Europe measured by the eddy covariance technique. *Journal of Geophysical Research Atmospheres*, 103(D10), 11289-11301. doi:10.1029/98JD00481

- Bacaro, G., Rocchini, D., Diekmann, M., Gasparini, P., Gioria, M., Maccherini, S., . . . Chiarucci, A. (2015). Shape matters in sampling plant diversity: Evidence from the field. *Ecological Complexity*, 24, 37-45. doi:<https://doi.org/10.1016/j.ecocom.2015.09.003>
- Bezdek, J. C. (1981). *Pattern recognition with fuzzy objective function algorithms*. New York, NY, USA: Plenum.
- Blaschke, T., Hay, G. J., Kelly, M., Lang, S., Hofmann, P., Addink, E., . . . van Coillie, F. (2014). Geographic object-based image analysis—towards a new paradigm. *ISPRS Journal of Photogrammetry and Remote Sensing*, 87, 180-191.
- Bray, J. R., & Curtis, J. T. (1957). An ordination of the upland forest communities of Southern Wisconsin. *Ecological Monographs*, 27(326-349).
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32. doi:10.1023/a:1010933404324
- Böhner, J., & Selige, T. (2006). Spatial prediction of soil attributes using terrain analysis and climate regionalisation. In J. Böhner, K. R. McCloy, & J. Strobl (Eds.), *SAGA – Analysis and modelling applications. Göttinger Geographische Abhandlungen 115*. (pp. 13-28).
- Calinski, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics - Theory and Methods*, 3(1), 1-27.
- Chapin III, F. S., Bret-Harte, M. S., Hobbie, S. E., & Zhong, H. (1996). Plant functional types as predictors of transient responses of arctic vegetation to global change. *Journal of Vegetation Science*, 7(3), 347-358.
- Cole, B., McMorow, J., & Evans, M. (2013). Empirical modelling of vegetation abundance from airborne hyperspectral data for upland peatland restoration monitoring. *Remote Sensing*, 6(1), 716-739. doi:10.3390/rs6010716
- Coops, N. C., Johnson, M., Wulder, M. A., & White, J. C. (2006). Assessment of QuickBird high spatial resolution imagery to detect red attack damage due to mountain pine beetle infestation. *Remote Sensing of Environment*, 103(1), 67-80. doi:10.1016/j.rse.2006.03.012
- de Klerk, H. M., Burgess, N. D., & Visser, V. (2018). Probabilistic description of vegetation ecotones using remote sensing. *Ecological Informatics*, 46, 125-132. doi:10.1016/j.ecoinf.2018.06.001
- Díaz-Varela, R. A., Calvo Iglesias, S., Cillero Castro, C., & Díaz Varela, E. R. (2018). Sub-metric analysis of vegetation structure in bog-heathland mosaics using very high resolution rps imagery. *Ecological Indicators*, 89, 861-873. doi:10.1016/j.ecolind.2017.11.068
- Duckworth, J. C., Kent, M., & Ramsay, P. M. (2000). Plant functional types: An alternative to taxonomic plant community description in biogeography? *Progress in Physical Geography*, 24(4), 515-542. doi:10.1191/030913300701542778
- Dudley, K. L., Dennison, P. E., Roth, K. L., Roberts, D. A., & Coates, A. R. (2015). A multi-temporal spectral library approach for mapping vegetation species across spatial and temporal phenological gradients. *Remote Sensing of Environment*, 167, 121-134. doi:10.1016/j.rse.2015.05.004
- Duff, T. J., Bell, T. L., & York, A. (2014). Recognising fuzzy vegetation pattern: the spatial prediction of floristically defined fuzzy communities using species distribution modelling methods. *Journal of Vegetation Science*, 25(2), 323-337. doi:10.1111/jvs.12092
- Dufrêne, M., & Legendre, P. (1997). Species assemblages and indicator species: The need for a flexible asymmetrical approach. *Ecological Monographs*, 67(3), 345-366.
- Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, 77(4), 802-813. doi:10.1111/j.1365-2656.2008.01390.x
- Foody, G. M. (1997). Fully fuzzy supervised classification of land cover from remotely sensed imagery with an artificial neural network. *Neural Computing & Applications*, 5(4), 238-247. doi:10.1007/bf01424229

- Gonçalves, J., Henriques, R., Alves, P., Sousa-Silva, R., Monteiro, A. T., Lomba, A., . . . Honrado, J. (2016). Evaluating an unmanned aerial vehicle-based approach for assessing habitat extent and condition in fine-scale early successional mountain mosaics. *Applied Vegetation Science*, 19(1), 132-146. doi:10.1111/avsc.12204
- Gower, J. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*, 53(325-328).
- Guisan, A., Weiss, S. B., & Weiss, A. D. (1999). GLM versus CCA spatial modeling of plant species distribution. *Plant Ecology*, 143(1), 107-122. doi:10.1023/A:1009841519580
- Güler, B., Jentsch, A., Apostolova, I., Bartha, S., Bloor, J. M. G., Campetella, G., . . . Dengler, J. (2016). How plot shape and spatial arrangement affect plant species richness counts: implications for sampling design and rarefaction analyses. *Journal of Vegetation Science*, 27(4), 692-703. doi:doi:10.1111/jvs.12411
- Haralick, R. M., Dinstein, I., & Shanmugam, K. (1973). Textural Features for Image Classification. *IEEE Transactions on Systems, Man and Cybernetics*, SMC-3(6), 610-621. doi:10.1109/TSMC.1973.4309314
- Harris, A., Charnock, R., & Lucas, R. M. (2015). Hyperspectral remote sensing of peatland floristic gradients. *Remote Sensing of Environment*, 162, 99-111. doi:10.1016/j.rse.2015.01.029
- Hartley, A. J., MacBean, N., Georgievski, G., & Bontemps, S. (2017). Uncertainty in plant functional type distributions and its impact on land surface models. *Remote Sensing of Environment*, 203, 71-89. doi:10.1016/j.rse.2017.07.037
- Hill, M. O. (1979). *TWINSPAN—A FORTRAN program for arranging multivariate data in an ordered two-way table by classification of the individuals and attributes*. Retrieved from Ithaca, NY, USA:
- Kattenborn, T., Fassnacht, F. E., Pierce, S., Lopatin, J., Grime, J. P., & Schmidtlein, S. (2017). Linking plant strategies and plant traits derived by radiative transfer modelling. *Journal of Vegetation Science*, 28(4), 717-727. doi:10.1111/jvs.12525
- Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., . . . Hunt, T. (2018). caret: classification and regression training. R package version 6.0-79. Retrieved from <https://CRAN.R-project.org/package=caret>
- Kursa, M. B., & Rudnicki, W. R. (2010). Feature Selection with the Boruta Package. *Journal of Statistical Software*, 36(11), 1-13.
- Lehmann, J. R. K., Münchberger, W., Knoth, C., Blodau, C., Nieberding, F., Prinz, T., . . . Kleinebecker, T. (2016). High-resolution classification of south patagonian peat bog microforms reveals potential gaps in up-scaled CH₄ fluxes by use of Unmanned Aerial System (UAS) and CIR imagery. *Remote Sensing*, 8(3). doi:10.3390/rs8030173
- Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. *R News*, 2(3), 18-22.
- Lovitt, J., Rahman, M. M., & McDermid, G. J. (2017). Assessing the value of UAV photogrammetry for characterizing terrain in complex peatlands. *Remote Sensing*, 9(7). doi:10.3390/rs9070715
- Maanavilja, L., Riutta, T., Aurela, M., Pulkkinen, M., Laurila, T., & Tuittila, E. S. (2011). Spatial variation in CO₂ exchange at a northern aapa mire. *Biogeochemistry*, 104(1-3), 325-345. doi:10.1007/s10533-010-9505-7
- Martinez Arbizu, P. (2019). *pairwiseAdonis: Pairwise multilevel comparison using adonis*. R package version 0.3. Retrieved from
- McFeeters, S. K. (1996). The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International journal of remote sensing*, 17(7), 1425-1432. doi:10.1080/01431169608948714

- Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., & Leisch, F. (2017). e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien. R package version 1.6-8. Retrieved from <https://CRAN.R-project.org/package=e1071>
- Middleton, M., Närhi, P., Arkimaa, H., Hyvönen, E., Kuosmanen, V., Treitz, P., & Sutinen, R. (2012). Ordination and hyperspectral remote sensing approach to classify peatland biotopes along soil moisture and fertility gradients. *Remote Sensing of Environment*, 124, 596-609. doi:10.1016/j.rse.2012.06.010
- Mikola, J., Virtanen, T., Linkosalmi, M., Vähä, E., Nyman, J., Postanogova, O., . . . Aurela, M. (2018). Spatial variation and linkages of soil and vegetation in the Siberian Arctic tundra - Coupling field observations with remote sensing data. *Biogeosciences*, 15(9), 2781-2801. doi:10.5194/bg-15-2781-2018
- Mora, C., Vieira, G., Pina, P., Lousada, M., & Christiansen, H. H. (2015). Land Cover Classification Using High-Resolution Aerial Photography in Adventdalen, Svalbard. *Geografiska Annaler, Series A: Physical Geography*, 97(3), 473-488. doi:10.1111/geoa.12088
- Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., . . . Wagner, H. (2018). vegan: Community Ecology Package. R package version 2.5-1. Retrieved from <https://CRAN.R-project.org/package=vegan>
- Palace, M., Herrick, C., DelGreco, J., Finnell, D., Garnello, A. J., McCalley, C., . . . Varner, R. K. (2018). Determining subarctic peatland vegetation using an unmanned aerial system (UAS). *Remote Sensing*, 10(9). doi:10.3390/rs10091498
- Pau, G., Fuchs, F., Sklyar, O., Boutros, M., & Huber, W. (2010). EBIImage-an R package for image processing with applications to cellular phenotypes. *Bioinformatics*, 26(7), 979-981. doi:10.1093/bioinformatics/btq046
- R Core Team. (2018). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Rapinel, S., Rossignol, N., Hubert-Moy, L., Bouzillé, J. B., & Bonis, A. (2018). Mapping grassland plant communities using a fuzzy approach to address floristic and spectral uncertainty. *Applied Vegetation Science*, 21(4), 678-693. doi:10.1111/avsc.12396
- Ridgeway, G. (1999). The state of boosting. *Computing Science and Statistics*, 31, 172-181.
- Ridgeway, G. (2017). gbm: Generalized Boosted Regression Models. R package version 2.1.3. Retrieved from <https://CRAN.R-project.org/package=gbm>
- Riutta, T., Laine, J., Aurela, M., Rinne, J., Vesala, T., Laurila, T., . . . Tuittila, E. S. (2007). Spatial variation in plant community functions regulates carbon gas dynamics in a boreal fen ecosystem. *Tellus, Series B: Chemical and Physical Meteorology*, 59(5), 838-852. doi:10.1111/j.1600-0889.2007.00302.x
- Roberts, D. W. (2016). labdsv: Ordination and Multivariate Analysis for Ecology. R package version 1.8-0. Retrieved from <https://CRAN.R-project.org/package=labdsv>
- Rocchini, D. (2014). Fuzzy species distribution models: A way to represent plant communities spatially. *Journal of Vegetation Science*, 25(2), 317-318. doi:10.1111/jvs.12152
- Roth, K. L., Roberts, D. A., Dennison, P. E., Peterson, S. H., & Alonzo, M. (2015). The impact of spatial resolution on the classification of plant species and functional types within imaging spectrometer data. *Remote Sensing of Environment*, 171, 45-57. doi:10.1016/j.rse.2015.10.004
- Rouse, J. W. J., Haas, R. H., Schell, J. A., & Deering, D. W. (1973). *Monitoring vegetation systems in the Great Plains with ERTS*. Paper presented at the Third Earth Resources Technology Satellite-1 Symposium, Washington, DC.
- Räsänen, A., Kuitunen, M., Tomppo, E., & Lensu, A. (2014). Coupling high-resolution satellite imagery with ALS-based canopy height model and digital elevation model in object-based

624 boreal forest habitat type classification. *ISPRS Journal of Photogrammetry and Remote*
625 *Sensing*, 94, 169-182. doi:10.1016/j.isprsjprs.2014.05.003
626 Schmidtlein, S., Feilhauer, H., & Bruelheide, H. (2012). Mapping plant strategy types using remote
627 sensing. *Journal of Vegetation Science*, 23(3), 395-405. doi:10.1111/j.1654-1103.2011.01370.x
628 Tapia, R., Stein, A., & Bijker, W. (2005). Optimization of sampling schemes for vegetation mapping
629 using fuzzy classification. *Remote Sensing of Environment*, 99(4), 425-433.
630 doi:10.1016/j.rse.2005.09.013
631 Ustin, S. L., & Gamon, J. A. (2010). Remote sensing of plant functional types. *New Phytologist*,
632 186(4), 795-816. doi:10.1111/j.1469-8137.2010.03284.x
633 Virtanen, T., & Ek, M. (2014). The fragmented nature of tundra landscape. *International Journal of*
634 *Applied Earth Observation and Geoinformation*, 27, 4-12. doi:10.1016/j.jag.2013.05.010

635
636

637	List of appendices:
638	Appendix S1: Drone image over the study area and location of field inventory data
639	Appendix S2: Calinski criterion plots of cascading K-means clustering
640	Appendix S3: Scree plots of principal coordinates analyses
641	Appendix S4: Indicator species and plant functional types for different clusters
642	Appendix S5: Pairwise comparisons of cluster difference using permutational multivariate analysis of
643	variance
644	Appendix S6: Maps of fuzzy habitat types
645	Appendix S7: Maps of fuzzy plant community clusters

646 *Table 1. Classified habitat types.*

Habitat type	Description
Wet flark	Water table aboveground most of the time; field layer dominated by sedges (<i>Carex</i> spp.); ground layer covered by open water, bare peat and wet brown mosses
Tussock flark	Water table aboveground most of the time; field layer covered by <i>Trichophorum</i> spp. tussocks and other sedges (<i>Carex</i> spp.); ground layer covered by open water, bare peat and wet brown mosses; more vegetation than in wet flarks
String margin	Field layer covered by <i>Betula nana</i> , other dwarf shrubs (e.g. <i>Vaccinium uliginosum</i> , <i>V. oxycoccos</i>) and some sedges (esp. <i>Carex</i> spp.); ground layer covered by sphagnum, dry and wet mosses, as well as open water
String top	Field layer covered by evergreen and deciduous shrubs (e.g. <i>Rhododendron tomentosum</i> , <i>Vaccinium vitis-idaea</i> , <i>V. uliginosum</i> , <i>Empetrum nigrum</i>), as well as forbs (esp. <i>Rubus chamaemorus</i>); ground layer covered by sphagnum and feather mosses; some lichen
Riparian fen	Field layer dominated by dense and tall sedge growth (<i>Carex</i> spp.), deciduous shrubs (e.g. <i>Betula nana</i> , <i>Salix</i> spp.) and forbs (<i>Comarum palustre</i>); ground layer covered by sphagnum, wet mosses and open water

647
648 *Table 2. Details of the remote sensing data and features calculated from the data. Features were calculated from*
649 *all layers. B refers to blue, G to green, R to red, NIR to near infra-red, TPI to topographical position index, TWI*
650 *to topographic wetness index, VHM to vegetation height model, NDVI to normalized difference vegetation index,*
651 *NDWI normalized difference water index, RGI to red-green index, GLCM to gray level co-occurrence*
652 *matrix, μ to mean, SD to standard deviation, and MAD to mean absolute deviation.*

Dataset	Date	Producer	Spatial resolution	Number and list of layers	Number and list of features
Drone image	Jul 1 2017	Finnish Meteorological Institute & authors	0.05 m	3: B, G, R	102: GLCM texture, μ , SD, MAD, percentiles (1st, 5th, 50th, 95th, 99th)
Drone digital elevation model	Jul 1 2017	Finnish Meteorological Institute & authors	0.08 m	7: Elevation, slope, TPIs (1m, 2m and 5m distance), SWI, VHM	14: μ , SD
Aerial image	Jun 26 2016	National Land Survey of Finland	0.5 m	7: B, G, R, NIR, NDVI, NDWI, RGI	14: μ , SD
WorldView-2	Jun 6 2013	Digital Globe Inc.	2 m	11: coastal B, B, G, yellow, R, red-edge, NIR1, NIR2, NDVI, NDWI, RGI	22: μ , SD
Four PlanetScope images	Jun 11 2017 Jul 25 2017 Aug 8 2017 Sep 7 2017	Planet Labs Inc.	3 m	28: B, G, R, NIR, NDVI, NDWI, RGI from all images	56: μ , SD
LiDAR data	Jul 12 2016	National Land Survey of Finland	0.5 points m ⁻² (point cloud), 2 m (layers)	9: Elevation, slope, TPIs (5m, 10m, 20m, 50m, 100m distances), SWI, VHM	18: μ , SD

653
654
655

656 *Table 3. Correlations (Pearson's r) between cluster membership values of the different clustering options across*
657 *the study plots in the 2006 data. Correlations indicate how well cluster membership values correspond to each*
658 *other when different input data is used. In the crisp cluster row, instead of correlations, numbers represent the*
659 *proportion of the plots which were modeled similar crisp cluster in both clustering options.*

Comparison 1 Comparison 2	Species Extended PFT	Species PFT	Extended PFT PFT
Wet brown moss	0.80	0.79	0.98
Deciduous tall shrub	0.61	0.59	0.96
Graminoid	0.67	0.66	0.96
Graminoid/ <i>Utricularia</i>	0.44	0.46	0.92
Evergreen dwarf shrub	0.95	0.95	1.00
Crisp cluster	0.65	0.57	0.81

660
661 *Table 4. Confusion matrix for the habitat type classification. Cell values in the inner matrix are square meters,*
662 *and accuracies are calculated on a scale of 0-1. User's accuracy is the ratio between correctly classified area of*
663 *a habitat type and the area covered by the respective class in the final map, and producer's accuracy is the ratio*
664 *between correctly classified area of a habitat type and the area covered by the respective class in the reference*
665 *data.*

		Reference data					User's accuracy
		Wet flark	Tussock flark	String margin	String top	Riparian fen	
Classified data	Wet flark	16.40	0.74	0.62	0.28	0.20	0.90
	Tussock flark	1.77	0.57	0.76	0.23	0.00	0.17
	String margin	0.92	0.25	3.47	1.49	0.20	0.55
	String top	0.50	0.00	2.02	6.39	0.00	0.72
	Riparian fen	0.79	0.20	0.20	0.00	1.57	0.57
	Producer's accuracy	0.80	0.32	0.49	0.76	0.80	0.72

666
667
668 *Table 5. Overall extent and average patch size of different habitat types and plant community clusters.*

	Type	Extent (ha)	Average patch size (m ²)
Habitat types	Wet flark	10.4	144
	Tussock flark	1.7	14
	String margin	3.3	15
	String top	5.2	39
	Riparian fen	4.1	71
Clusters	Wet brown moss	3.5	21
	Deciduous tall shrub	2.2	10
	Graminoid	6.7	55
	Graminoid/ <i>Utricularia</i>	6.2	55
	Evergreen dwarf shrub	6.0	46

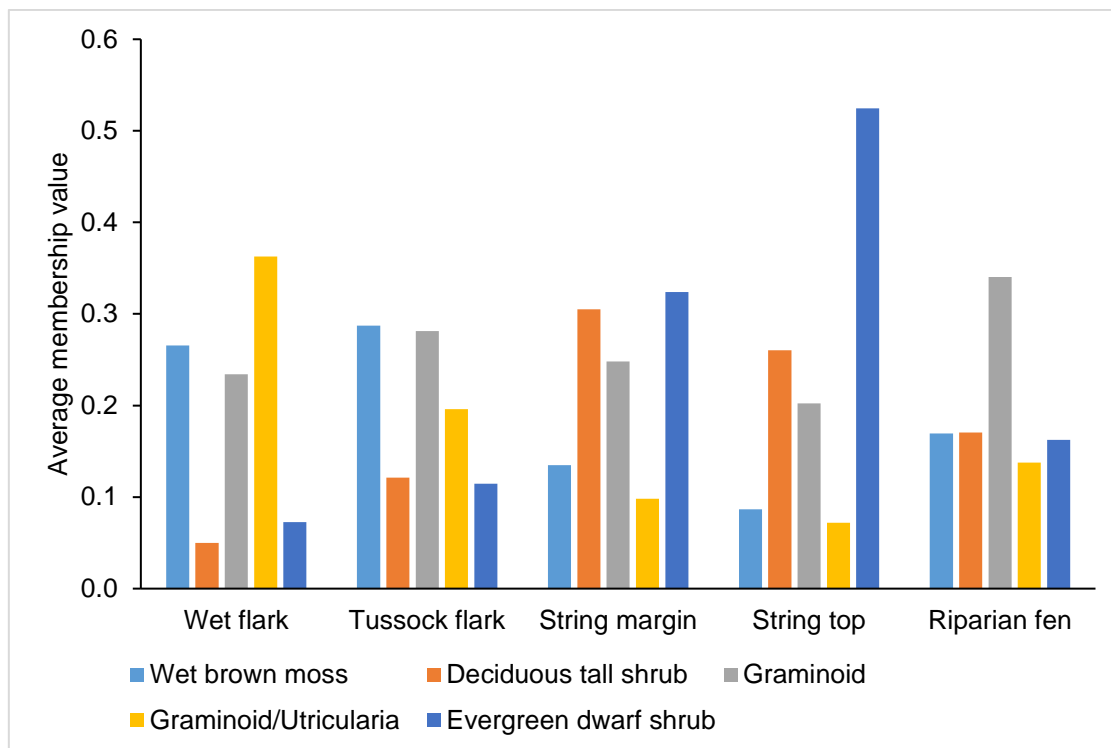
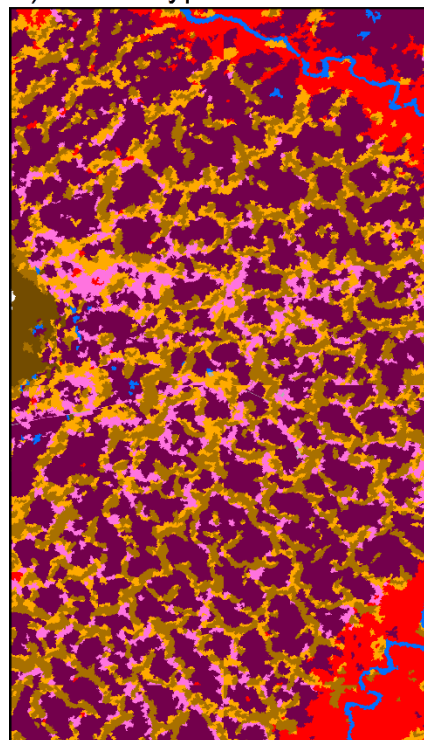


Figure 1. The average membership value (y-axis) of each cluster (different colored bars) in habitat type segments.

a) habitat types



Habitat type

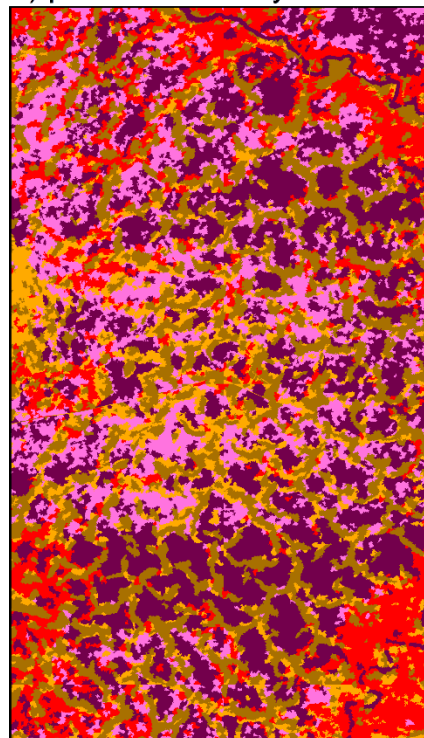
- Wet flark
- Tussock flark
- String margin
- String top
- Riparian fen
- Other habitat types:
- Pine bog
- Sand/artificial
- Water

Confusion

- <0.1
- 0.1-0.2
- 0.2-0.3
- 0.3-0.4
- 0.4-0.5
- 0.5-0.6
- 0.6-0.7
- 0.7-0.8
- 0.8-0.9
- >0.9



c) plant community clusters

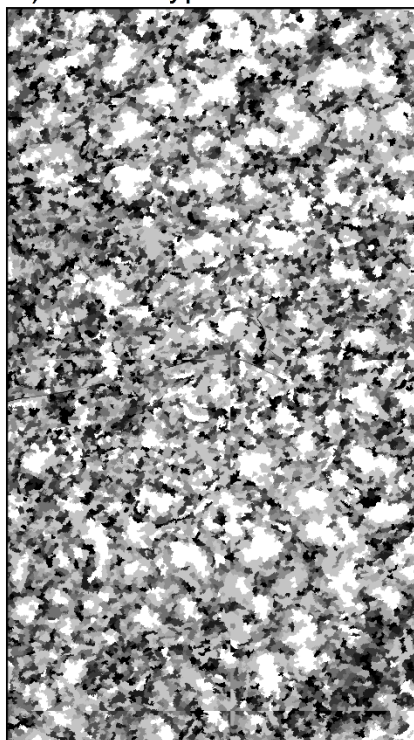


Cluster

- Wet brown moss
- Deciduous tall shrub
- Graminoid
- Graminoid/Utricularia
- Evergreen dwarf shrub

0 100 200 m

b) habitat type confusion



d) cluster confusion

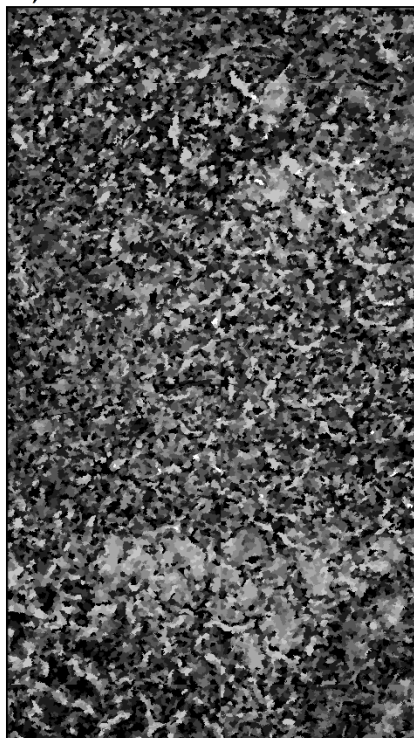


Figure 2. Crisp habitat type and plant community cluster maps and confusion maps.