Data and resolution requirements in mapping vegetation in spatially heterogeneous landscapes

Aleksi Räsänen[ab]*, Tarmo Virtanen[a]

[a]Ecosystems and Environment Research Programme, Faculty of Biological and Environmental Sciences, and Helsinki Institute of Sustainability Science (HELSUS), P.O. Box 65, FI-00014 University of Helsinki, Finland; AR: aleksi.rasanen@helsinki.fi, TV: tarmo.virtanen@helsinki.fi

[b]Department of Geography, Norwegian University of Science and Technology, NO-7491 Trondheim, Norway

*Corresponding author, aleksi.rasanen@helsinki.fi

Abstract

It has been argued that even centimeter-level resolution is needed for mapping vegetation patterns in spatially heterogeneous landscapes such as northern peatlands. However, there are few systematic tests for determining what kind of spatial resolution and data combinations are needed and what the differences in mapping accuracy are when different datasets are omitted or included. We conducted 78 different object-based supervised random forest classifications on a patterned fen and its surroundings in Kaamanen, northern Finland, using remotely sensed optical imagery, topography, and vegetation height datasets from different platforms (unmanned aerial vehicle (UAV), aerial, satellite) with spatial resolution ranging from 5 cm to 3 m. We compared differences in classification performance when we altered (1) classification and segmentation input data and features calculated from the data, or (2) the segmentation scale. We constructed training data with the help of transect-based field sampling and UAV imagery and tested classification accuracy using 412 field-surveyed vegetation plots. The most accurate classifications (75.7% overall accuracy) were obtained when we segmented a 5 cm resolution UAV image with a small segmentation scale and calculated features from all datasets. Classification accuracy was 2.2 percentage points (*pp*) lower with the most accurate aerial image (50 cm resolution) based classification, and 7.6 *pp* and 11.9 *pp* lower with the most accurate WorldView-2 (2 m resolution) and PlanetScope (3 m resolution) satellite image based classifications respectively. Classification accuracies were low (46.7–56.0%) when we used only spectral data from one dataset. The inclusion of gray-level co-occurrence matrix textural features increased classification accuracy by 0.4–12.1 *pp* and inclusion of multiple datasets by 8.2–25.0 *pp*. Segmentation scale had a minor effect on classification accuracy (2.5–7.3 *pp* difference between the finest and coarsest segmentation scale); however, both too small and large segmentation scale might lead to suboptimal classification. The differences in land cover type areal coverage were relatively small between classifications with multiple datasets, but if classifications included features from only one dataset, the differences were larger. We conclude that multiple different optical, topographical, and vegetation height datasets should be used when mapping vegetation in spatially heterogeneous landscapes, and that sub-meter resolution data (e.g. UAV or aerial) are necessary for the most accurate maps. Although UAV data is not essentially needed for classification, it is useful for training dataset construction and especially helpful in areas lacking other sub-meter resolution data.

Keywords: Arctic; data fusion; drone; land cover classification; lidar; northern boreal; object-based image analysis; peatland; UAS; ultra-high spatial resolution; very-high spatial resolution

1. Introduction

Land cover and vegetation maps are among the most important products derived from remotely sensed data. Thematic classifications of vegetation and land cover are usually constructed for a specific purpose, such as linking them to carbon stocks and fluxes, biodiversity, or some other environmental question (Goetz et al. 2009; Gong et al. 2013; Jung et al. 2006; Pettorelli et al. 2016). In land cover mapping, key issues include what kind of datasets are used and what is their spatial resolution (Chen et al. 2017b; Chen et al. 2015; Räsänen et al. 2014). These issues are important in spatially heterogeneous landscapes such as northern peatlands and tundra (Bartsch et al. 2016; Virtanen and Ek 2014). These landscapes are fragmented and patchy in terms of their vegetation, land cover, and hydrology (Middleton et al. 2012; Palace et al. 2018; Räsänen et al. 2019b, Treat et al. 2018), and biogeochemical cycles of e.g., carbon, nitrogen, and water vary greatly between different land cover types, creating an urgent need to classify them accurately (Lehmann et al. 2016; Treat et al. 2018).

There have been contrasting claims about what kind of spatial resolution is needed for accurate mapping of land cover and vegetation patterns in spatially heterogeneous landscapes. Some have argued that Landsat-scale resolution (ca 30 m) is sufficient for mapping tundra-peatland environments if the objective is to track the relative abundance of different land cover types and carbon fluxes related to these types (Bartsch et al. 2016; Schneider et al. 2009; Treat et al. 2018). Others have claimed that very high spatial resolution satellite imagery (< 5 m) is needed for constructing realistic maps in these environments (Laidler and Treitz 2003; Räsänen et al. 2019b; Siewert et al. 2015; Virtanen and Ek 2014). Finally, some have argued that there is a need to move into centimeter-level spatial resolution, obtained with unmanned aerial vehicles (UAVs) or airborne data when mapping peatland vegetation (Palace et al. 2018).

Related to this discussion, several studies have been conducted using very high spatial resolution satellite imagery (spatial resolution < 5 m) in tracking vegetation and biogeochemical patterns in heterogenic northern landscapes such as tundra and peatlands (Laidler and Treitz 2003; Räsänen et al. 2019b; Siewert et al. 2015; Virtanen and Ek 2014), and these have been followed by a recent increase in using UAVs in similar tasks (Anderson and Gaston 2013; Arroyo-Mora et al. 2017; Lehmann et al. 2016; Lovitt et al. 2017; Palace et al. 2018). Many of these studies note that there is a trade-off between spatial resolution and areal extent when using these data: only a relatively small extent can be covered if dataset resolution is enhanced to centimeters or meters (Laidler and Treitz 2003). Therefore, coarser resolution datasets may be preferred in tasks covering a larger extent, but the trade-offs in upscaling finer resolution data to coarser resolution are generally understudied (Treat et al. 2018).

When utilizing high resolution datasets, object-based methods instead of pixel-based methods are usually preferred (Blaschke et al. 2014; Dronova 2015; Ma et al. 2017; Mahdavi et al. 2018). Firstly, when using high resolution data, the vegetation patch size is usually larger than the data pixel size; therefore, pixels can be merged into homogeneous segments before the classification or other mapping step (Blaschke et al. 2014; Castilla and Hay 2008). In particular, several land cover types have a large internal heterogeneity in very high resolution images, often due to shadow effects caused by higher vegetation, which hamper pixel-based classifications. Secondly, the generated homogeneous segments are a more realistic construction of the landscape elements and they mimic human interpretation of the landscape more intuitively than pixels (Castilla and

Hay 2008). However, the segmentation step adds uncertainty to classification and other tasks. Segmentation should delineate the areas well; therefore, there should be careful choice of the segmentation method and its parameterization (Clinton et al. 2010; Costa et al. 2018; Georganos et al. 2018; Räsänen et al. 2013). In parameterization, one of the most important choices is to select correct segmentation scale (i.e., the size of the segment). The choice of the segmentation scale is related to resolution requirements and areal extent: coarser scale segmentation allows mapping of larger areas but small-sized patches may be missed when the resolution is too coarse. Thirdly, classification accuracies are often higher with object-based than pixel-based methods (Amani et al. 2017; Dronova 2015; Sibaruddin et al. 2018). However, also other factors such as the selection of input data have an effect on the classification accuracy.

It has been shown that the inclusion of multiple images, in terms of extra spectral and phenological information, increases classification accuracy (Chen et al. 2017a; Chen et al. 2017b; Halabisky et al. 2018; Lu et al. 2017; Lucas et al. 2011). A single image is only a snapshot of one time point, and multiple images taken in different phenological or seasonal phases may allow the finding of differences between land cover or vegetation types (Chen et al. 2017b; Dudley et al. 2015; Halabisky et al. 2018; Lu et al. 2017; Lucas et al. 2011). In particular, northern landscapes are typically characterized by high seasonal variation, and phenological development differs between land cover types (Juutinen et al. 2017), and especially in peatlands, water levels vary a lot seasonally. Different sensors have different spectral resolution and details; therefore, inclusion of extra spectral data, including hyperspectral data, may reveal patterns invisible to one sensor (Chen et al. 2017a; Chen et al. 2017b; Lu et al. 2017). Moreover, instead of using only average pixel values, textural features representing spatial variation in pixel values have been shown to increase classification accuracies (Chen et al. 2018; Hall-Beyer 2017; Mishra et al. 2018). It has also been shown that when optical datasets are combined with features characterizing topographical and vegetation structure elements, classification accuracies can be boosted (Franklin and Ahmed 2017; Luo et al. 2016; Prošek and Šímová 2019; Räsänen et al. 2014; Sankey et al. 2018; Shadaydeh et al. 2017). However, some results have indicated that inclusion of lidar data does not increase classification accuracy when wetland vegetation is mapped with aerial hyperspectral data (Stratoulias et al. 2018). Although there have been multiple arguments for including different types of data in a single mapping approach, quite often UAV-based mapping includes features calculated only from the optical UAV data (Lehmann et al. 2016; Palace et al. 2018). Additionally, there are few systematic tests for determining what kind of data mixtures are needed and what the changes in mapping accuracy are when different datasets are omitted or included.

Our objectives were to test what kind of spatial resolution and dataset combination are needed for mapping land cover patterns in a patchy peatland landscape in Kaamanen, northern Finland. Earlier research in the area has concentrated on carbon dioxide (CO_2) exchange, its spatial and temporal heterogeneity, and the linkages between it and vegetation. The landscape is characterized by strong seasonal patterns, with high amount of snow in the winter and a short growing season in the summer (Aurela et al. 1998, 2001, 2002, 2004). There is also some interannual variation e.g. in the timing of snow melt (Aurela et a. 2004) and in the wetness conditions during the growing season. It has been reported that there is fine-scale variation in vegetation, land cover and topography (Räsänen et al. 2019c), and the distinct plant community types within the fen have diverging CO_2 exchange patterns (Maanavilja et al. 2011). Overall, the chosen study area is an ideal location to test how land cover maps differ when the input data and its resolution are altered.

We conducted 78 different classifications using optical imagery, topography, and vegetation height remote sensing datasets from different platforms (UAV, aerial, satellite) with spatial resolution ranging from 5 cm to 3 m. We asked what kind of changes there are in classification accuracy and in areal cover and patchiness of land cover types when (1) spatial resolution of segmented and classified data is changed, (2) segmentation scale is changed, and (3) classification input data and features calculated from the data are changed.

2. Materials and methods

2.1 Study area

The study area of 0.4 km^2 is 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 area is dominated by a treeless mesotrophic patterned fen characterized by a strong pattern of strings (less than 1 m high) with dwarf shrub vegetation, and flarks with sedge and wet brown moss vegetation (Fig. 1). A small stream runs through the study area; the riparian areas are characterized by tall sedge vegetation. The study area includes also upland pine forests, shrub-dominated pine peatland in the ecotone between the upland forest and open peatland, and a small lake. In the middle of the circular study area, there is an eddy covariance tower that has been measuring ecosystem CO_2 exchange since 1997 (Aurela et al. 1998, 2001, 2002, 2004). The study area, determined by the extent of the UAV image and by the main footprint area of the eddy covariance tower, extends to a distance of 300–330 m from the tower in each direction. Similar types of peatlands and pine dominated forest vegetation can be found extensively in the region surrounding the study area.

Figure 1. The studied fen landscape is characterized by a strong string–flark pattern. <2-column fitting image>

2.2 Fieldwork data

We collected transect data of land cover distribution in 2017. Eight 250 m transects were set up in cardinal and intercardinal directions from the flux tower. Land cover along the transects was classified into ten types (Table 1). The transect data were used for training the classifiers.

Table 1. Classified land cover types. The four first land cover types are described in more detail in Maanavilja et al. (2011).

Land cover type	Description
Wet flark	Water table aboveground most of the time; field layer dominated by sedges (*Carex* 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 *Trichophorum* spp. tussocks, and other sedges (*Carex* 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 *Betula nana*, other dwarf shrubs (e.g., *Vaccinium uliginosum*, *Vaccinium oxycoccos*), and some sedges (especially *Carex* 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., *Rhododendron tomentosum*, *Vaccinium vitis-idaea*, *V. uliginosum*, *Empetrum nigrum*), as well as herbs (especially *Rubus chamaemorus*); ground layer covered by sphagnum and feather mosses; some lichen
Riparian fen	Field layer dominated by dense and tall sedge growth (*Carex* spp.), deciduous shrubs (e.g., *B. nana*, *Salix* spp.), and herbs (*Comarum palustre*); ground layer covered by sphagnum, wet mosses, and open water
Pine bog	Scots pine (*Pinus sylvestris*) with 1–30% canopy cover and ca 5 m dominant height; field layer dominated by evergreen and deciduous shrubs (e.g., *R. tomentosum*, *V. vitis-idaea*, *V. uliginosum*, *E. nigrum*), as well as herbs (especially *R. chamaemorus*); ground layer covered by sphagnum and feather mosses; some lichen
Pine forest	Forest area on mineral soil dominated by Scots pine (*P. sylvestris*), canopy cover > 10%, dominant height ca 10 m; field layer dominated by evergreen and deciduous shrubs (e.g., *Calluna vulgaris*, *V. vitis-idaea*, *Vaccinium myrtillus*); ground layer covered by feather mosses and lichen
Clear-cut	Open mineral soil forest patches or areas where trees have been cut, canopy cover < 10%; field layer dominated by evergreen and deciduous shrubs (e.g., *C. vulgaris*, *V. vitis-idaea*, *V. myrtillus*); ground layer covered by feather mosses and lichen
Water	Open water, includes lakes, ponds, and streams
Non-vegetated	Sand and other non-vegetated surfaces. Mostly consists of forest roads covered by gravel/sand and boardwalks

For validation data, we used land cover information collected in 412 vegetation plots in 2017 and 2018. In 2017, a total of 210 rectangular plots with 50 cm side length, and 18 circular plots with 40 cm diameter, were used. Rectangular plots were sampled systematically at distances of 25 to 150 m from the flux tower in cardinal, intercardinal, and secondary intercardinal directions. Circular plots were situated at distances of 7 to 100 m from the flux tower and represented the major land cover types found in the study area. In 2018, data were collected in 141 rectangular plots with 50 cm side length in the fen. We sampled the plots using stratified random sampling and used the following land cover types of a preliminary classification as strata: string top, string margin, wet flark, tussock flark, riparian fen, and pine bog. Finally, we visually interpreted the

UAV image, and set a total of 42 extra validation points for the following land cover types which were not well covered in our peatland targeted field sampling: water, pine forest, clear-cut, and non-vegetated surfaces.

Transects in 25–100 m intervals and vegetation plots were located with a Trimble R10 GPS device with ± 5 cm accuracy, and a Garmin eTrex 30 GPS device was used when transitions between the land cover types in transects were located. The location of the vegetation plots in the UAV image was double-checked with visual interpretation to verify that the vegetation description and visual interpretation in the field matched that in the UAV image.

2.3 Remote sensing datasets

We used optical UAV, aerial, and satellite imagery, as well as digital elevation and digital surface models at 5 cm to 3 m spatial resolution (Table 2) to test what kind of data and resolution are needed for mapping vegetation. A DJI phantom 4 pro UAV flight was conducted, and the UAV image was georeferenced using 14 ground control points measured with a Trimble R10 GPS device with ± 5 cm accuracy. An image mosaic, as well as a digital terrain and digital surface models were computed using Pix4D software (Pix4D SA, Lausanne, Switzerland). We calculated a vegetation height model by subtracting the digital terrain model from the digital surface model. In addition to the UAV image, we used coarser resolution aerial orthophoto and lidar data from the National Land Survey of Finland (Table 2). The spatial alignment between the orthophoto and UAV data was verified with visual interpretation. From the lidar, we used a digital terrain model calculated by the National Land Survey, as well as a vegetation height model in which we subtracted the digital terrain model from a digital surface model and in which calculation we used all lidar returns. We also used the following satellite image data sources: WorldView-2 image (WV-2, DigitalGlobe Inc., Westminster, CO, USA) and four PlanetScope images (PS, Planet Labs Inc., San Francisco, CA, USA (Planet Team 2017)). The WV-2 image was orthocorrected with the help of the aerial orthophoto and 18 ground control points. The spatial accuracy of the orthocorrected PS images was verified using visual interpretation.

Table 2. Details of the remote sensing data and layers calculated from the data. B refers to blue, G to green, GLCM to gray-level co-occurrence matrix, NDVI to normalized difference vegetation index, NDWI to normalized difference water index, NIR to near-infrared, R to red, RGI to red-green index, TPI to topographical position index, TWI to topographical wetness index, UAV to unmanned aerial vehicle, and VHM to vegetation height model. The Classifications column indicates to which dataset segmentations and further classifications the features were linked.

Dataset	Date	Producer	Spatial resolution	Number and list of layers	Classifications
UAV image	Jul 1, 2017	Finnish Meteorological Institute & authors	0.05 m	27: B, G, R, and 8 GLCM layers from all spectral bands	UAV
UAV digital elevation model	Jul 1, 2017	Finnish Meteorological Institute & authors	0.08 m	7: Elevation, slope, TPIs (1 m, 2 m, and 5 m distance), TWI, VHM	UAV
Aerial image	Jun 26, 2016	National Land Survey of Finland	0.5 m	39: B, G, R, NIR, NDVI, NDWI, RGI, and 8 GLCM layers from all spectral bands	UAV, aerial (GLCM features only in aerial image classifications)
WorldView-2	Jun 6, 2013	DigitalGlobe Inc.	2 m	75: coastal B, B, G, yellow, R, red-edge, NIR1, NIR2, NDVI, NDWI, RGI, and 8 GLCM layers from all spectral bands	UAV, aerial, WorldView-2 (GLCM features only in WorldView-2 image classifications)
Four PlanetScope images	Jun 11, 2017 Jul 25, 2017 Aug 8, 2017 Sep 7, 2017	Planet Labs Inc.	3 m	60: B, G, R, NIR, NDVI, NDWI, RGI from all images, and 8 GLCM layers from all spectral bands of the July image	UAV, aerial, WorldView-2, PlanetScope (GLCM features only in PlanetScope image classifications)
Lidar data	Jul 12, 2016	National Land Survey of Finland	0.5 points m^{-2} (point cloud), 2 m (layers)	9: Elevation, slope, TPIs (5 m, 10 m, 20 m, 50 m, 100 m distances), TWI, VHM	UAV, aerial, WorldView-2, PlanetScope

2.4 Land cover classification

We classified the land cover types with an object-based approach (Blaschke et al. 2014). First, we conducted a full lambda schedule segmentation and compared four different segmentation scale options for four different images. Second, we carried out random forest classifications (Breiman 2001) for the different segmentations and compared six different feature set options.

Visual interpretation is often the most meaningful way to parameterize segmentations in natural environments (Räsänen et al. 2013). Based on parameter combination testing and visual interpretation, we gave the relative weights 0.7, 0.5, 0.3, and 0.3 to color, texture, size, and shape, respectively. We segmented the following datasets one by one: UAV image, aerial image, WV-2 image, and PS image from July. We tested the following segmentation scales (i.e., mean size of segments): 2.5 m^2, 5 m^2, 10 m^2, and 20 m^2 with a minimum segment size of 1 m^2, 2m^2, 4m^2, and 8 m^2 respectively. As the pixel size of the WV-2 and PS images was 4 m^2 and 9 m^2, respectively, we could not conduct the classifications with the lowest segmentation scale for them. Instead, the highest resolution classifications for these was a pixel-based classification, and we carried out three classifications for WV-2 and two for PS. Segmentations were conducted in Erdas Imagine 2016 (Hexagon Geospatial, Madison, AL, USA).

For each segment, we calculated the mean value of all layers from different datasets (Table 2). In addition to the spectral bands, we calculated the following spectral indices for the aerial and satellite images: normalized difference vegetation index (Rouse et al. 1973), normalized difference water index (McFeeters 1996), and red-green index (Coops et al. 2006). For each spectral band of the segmented images, we calculated the following eight gray-level co-occurrence matrix textural images (Haralick et al. 1973): energy (texture uniformity), entropy (texture randomness), correlation (pixel's correlation with its neighborhood), inverse difference moment (texture homogeneity), inertia (intensity contrast between a pixel and its neighborhood), cluster shade, cluster prominence, and Haralick correlation. These were calculated with eight quantization levels, and a moving window technique with the neighborhood distance set to five for the UAV image, two for the aerial image, and one for the satellite images. For the digital elevation models, we calculated slope, topographical position indices with different neighborhood distances (Guisan et al. 1999), and topographical wetness index (Böhner and Selige 2006). Texture layers were calculated with Orfeo Toolbox (Grizonnet et al. 2017), and topographical layers were calculated with SAGA-GIS (Conrad et al. 2015).

We constructed training data for classifications with the help of the transect field data and visual interpretation of the UAV image. We constructed the training data using the 2.5 m^2 resolution UAV segmentation. We selected 3479 training segments (102 to 831 for each class).

In UAV segmentation based classifications, we used features calculated for all datasets; in aerial image segmentation based classifications, UAV features were excluded; in WV-2 segmentation based classifications, UAV and aerial image features were excluded; in PS segmentation based classifications, UAV, aerial image, and WV-2 features were excluded (Table 2). Furthermore, for each segmentation, we tested six different feature set options: (1) spectral bands and indices for the segmented image, (2) spectral bands and indices as well as textural features for the segmented image, (3) spectral bands and indices for the segmented image and topographical/vegetation height features, (4) spectral bands and indices for multiple images, (5) spectral bands and indices for multiple images and topographical/vegetation height features, and (6) spectral bands and indices for multiple images, topographical/vegetation height features, and textural features for the segmented image. We conducted altogether 78 classifications (13 segmentations and six different feature sets for each segmentation).

It has been shown that random forest is insensitive to parameterization (Du et al. 2015; Rodriguez-Galiano et al. 2012); thus, we used the default parameter values: number of trees was set to 500 and number of tested variables at each tree node was set to the square root of variables in the classification. Classifications were computed in R (R Core Team 2018) using package randomForest (Liaw and Wiener 2002).

2.5 Accuracy assessment and classification comparison

We tested classification accuracy using the 412 validation plots as reference data. For each point, we set a polygon circle either with a 25 cm (rectangular plots) or 20 cm (circular plots and extra visually interpreted plots) radius. We then cross-tabulated pixel-based classification accuracy with 5 cm accuracy (corresponds to the pixel size of UAV classifications). We compared different classifications based on overall accuracy as well as class-specific user's and producer's accuracies which have been suggested to be used as primary measures (Liu et al. 2007). Following the suggestion and equation by Foody (2008), we calculated 95% confidence intervals for the overall accuracy of each classification. In confidence interval calculations, we set the sample size to the number of 5 cm pixels within reference polygons (n = 30495 for UAV classifications and 30475 for other classifications). We also calculated the areal cover of each land cover type in each classification. To study the patchiness of the landscape, we calculated the mean patch size for each land cover type and measured patch complexity with mean shape index (i.e., patch perimeter divided by the smallest possible patch perimeter) for the classifications with the highest classification accuracy for each segmentation using V-LATE (Lang and Tiede 2003).

3. Results

The highest classification accuracy (76.7%) was achieved when we segmented the UAV image with 2.5 m^2 or 5 m^2 mean segment size and derived features from all datasets but excluded textural features (Table 3, Fig. 2). Almost as high classification accuracies were obtained (2.2 percentage points (*pp*) lower) when the segmented image was the aerial image instead of the UAV image. The classification accuracies were notably lower (7.6 *pp* with WV-2 and 11.9 *pp* with PS) when satellite imagery was segmented instead of the UAV. Confidence interval was ± 0.5 *pp* for classifications with > 60% overall classification accuracy and ± 0.6 *pp* for classifications with < 60% accuracy (Table S1); hence, the differences between different segmented image types can be considered statistically significant. Irrespective of the segmented image, visually acceptable classifications were obtained (Fig. 3). The classification accuracy decreased when the mean size of the segment increased. However, there was little difference between the two smallest segment sizes. At all segmentation scales, UAV or aerial image based classifications had the highest accuracies. Depending on the segmented data, the classification accuracy difference between the finest and coarsest segmentation scale was between 2.5 and 7.3 *pp* (Table 3, Figs 2 and 4).

Table 3. Overall classification accuracies (± confidence interval) for each segmentation with the classifications with highest classification accuracies. UAV refers to unmanned aerial vehicle.

Segment size (m^2)	UAV (%)	Aerial image (%)	WorldView-2 (%)	PlanetScope (%)
2.5	76.7±0.5	74.5±0.5	–	–
5	76.7±0.5	73.7±0.5	69.1±0.5	–
10	73.8±0.5	72.8±0.5	67.7±0.5	64.8±0.5
20	70.2±0.5	72.0±0.5	63.9±0.5	57.5±0.6

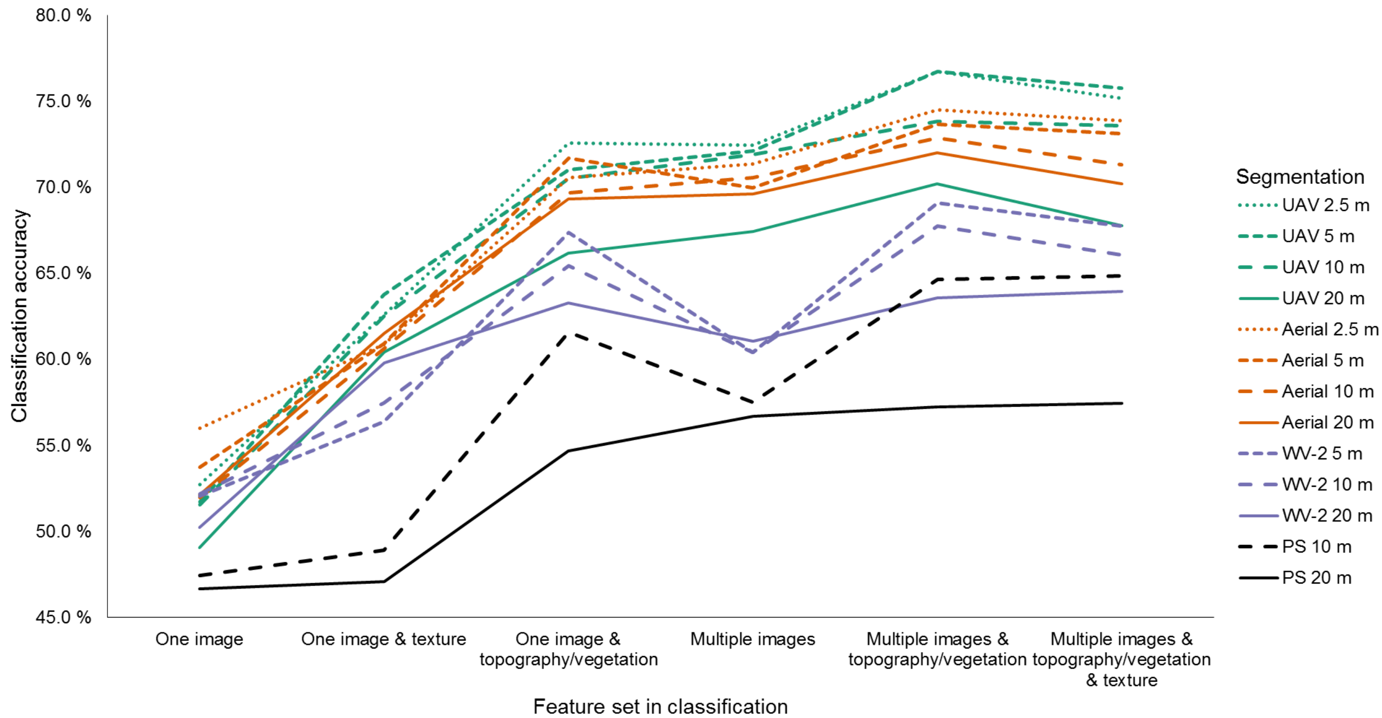


Figure 2. Classification accuracies (y-axis) of the 78 different classifications. Different feature sets used in the classification are presented on x-axis, the segmented image is visualized with different colors, and used segmentation scale is shown with line dash type. <2-column fitting image>

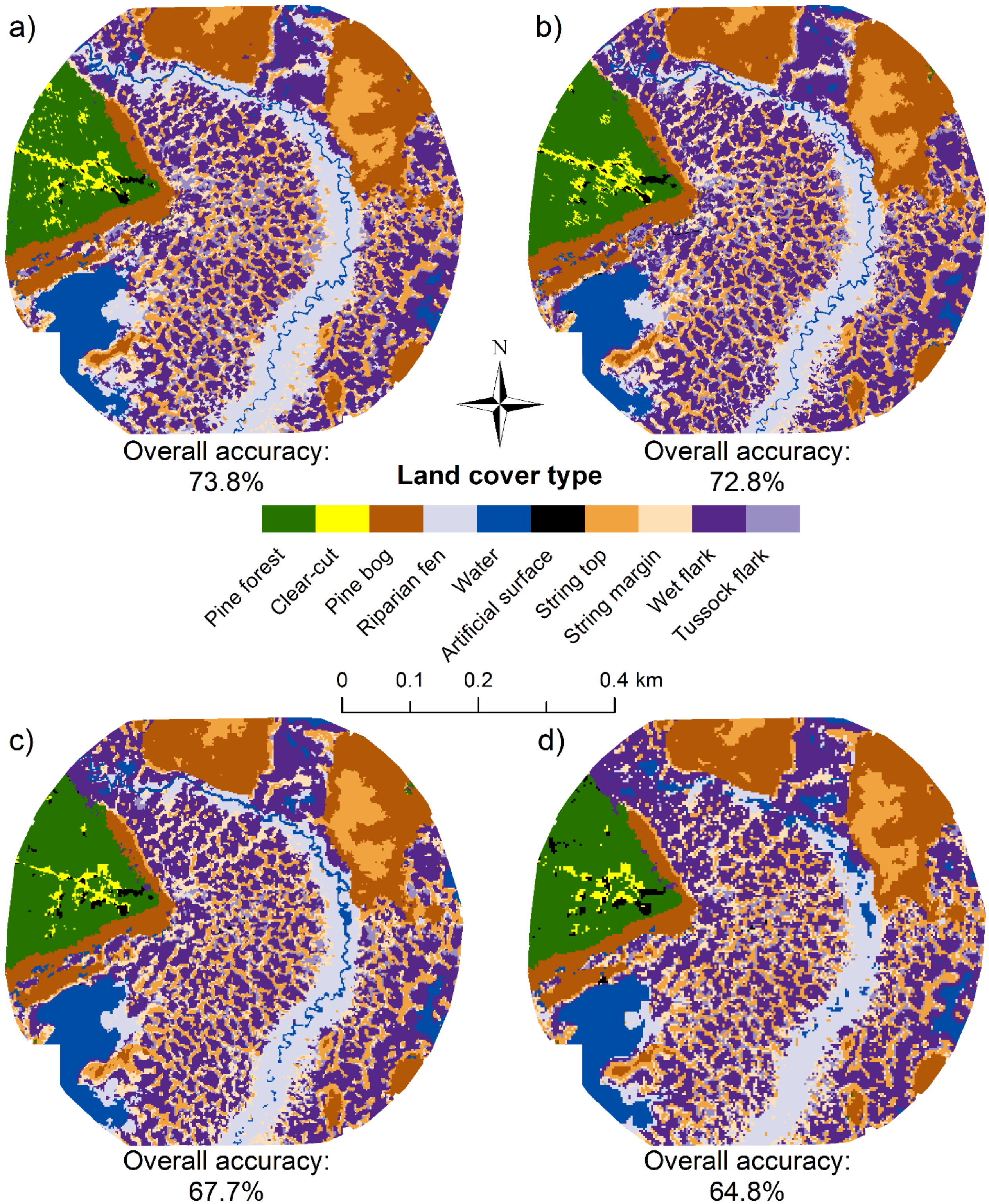


Figure 3. Classifications with 10 m² segmentation scale and with the following segmented images: a) unmanned aerial vehicle, b) aerial, c) WorldView-2, d) PlanetScope (9 m² pixels instead of segments as a basis). In all classifications, the feature set which yielded the highest classification accuracy is used. This includes features

calculated from multiple images as well as topographical and vegetation height features for all subfigures, excludes texture features for a, b, and c, and includes them for d. <2-column fitting image>

There were large differences in classification accuracy when different feature sets were used (Fig. 2). The lowest accuracies were obtained when using only spectral bands and indices for the segmented image. The inclusion of textural features increased classification accuracy (0.4 to 12.1 *pp* increase), but a higher increase was achieved when multiple remote sensing datasets were used. Inclusion of multiple images increased accuracy by 8.2–20.4 *pp*, inclusion of topographical and vegetation height data by 8.0–19.9 *pp*, and inclusion of both multiple images and topographical and vegetation height data by 10.6–25.0 *pp*. When all datasets were included in the classification, classification accuracy usually slightly decreased when textural features were included in the classification (0.4 *pp* increase to 2.5 *pp* decrease). In visual interpretation of the different classifications, it was observed that inclusion of multiple datasets was needed for visually acceptable classifications and their inclusion decreased random noise in the classifications (Fig. 5, Fig. 4a).

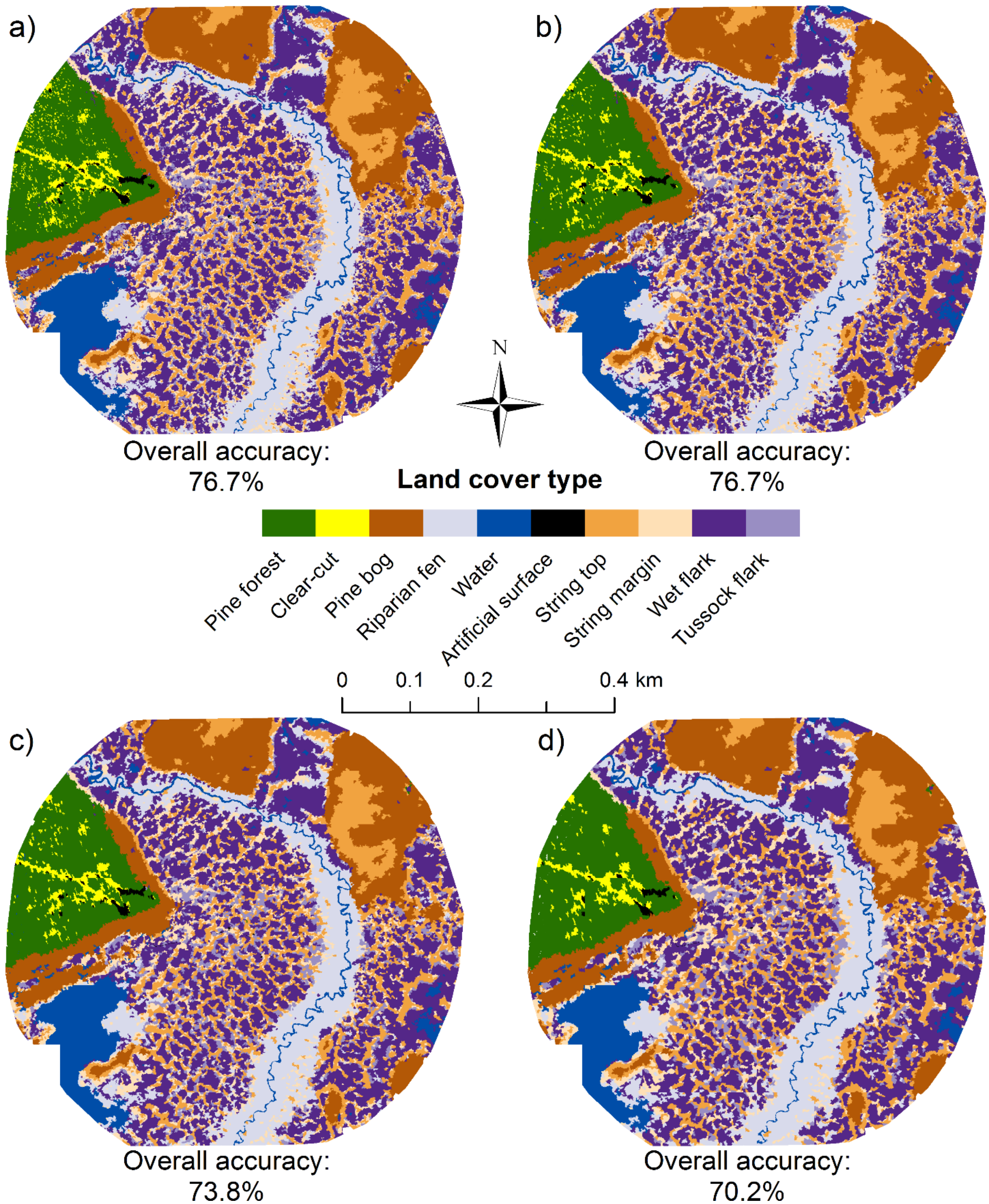
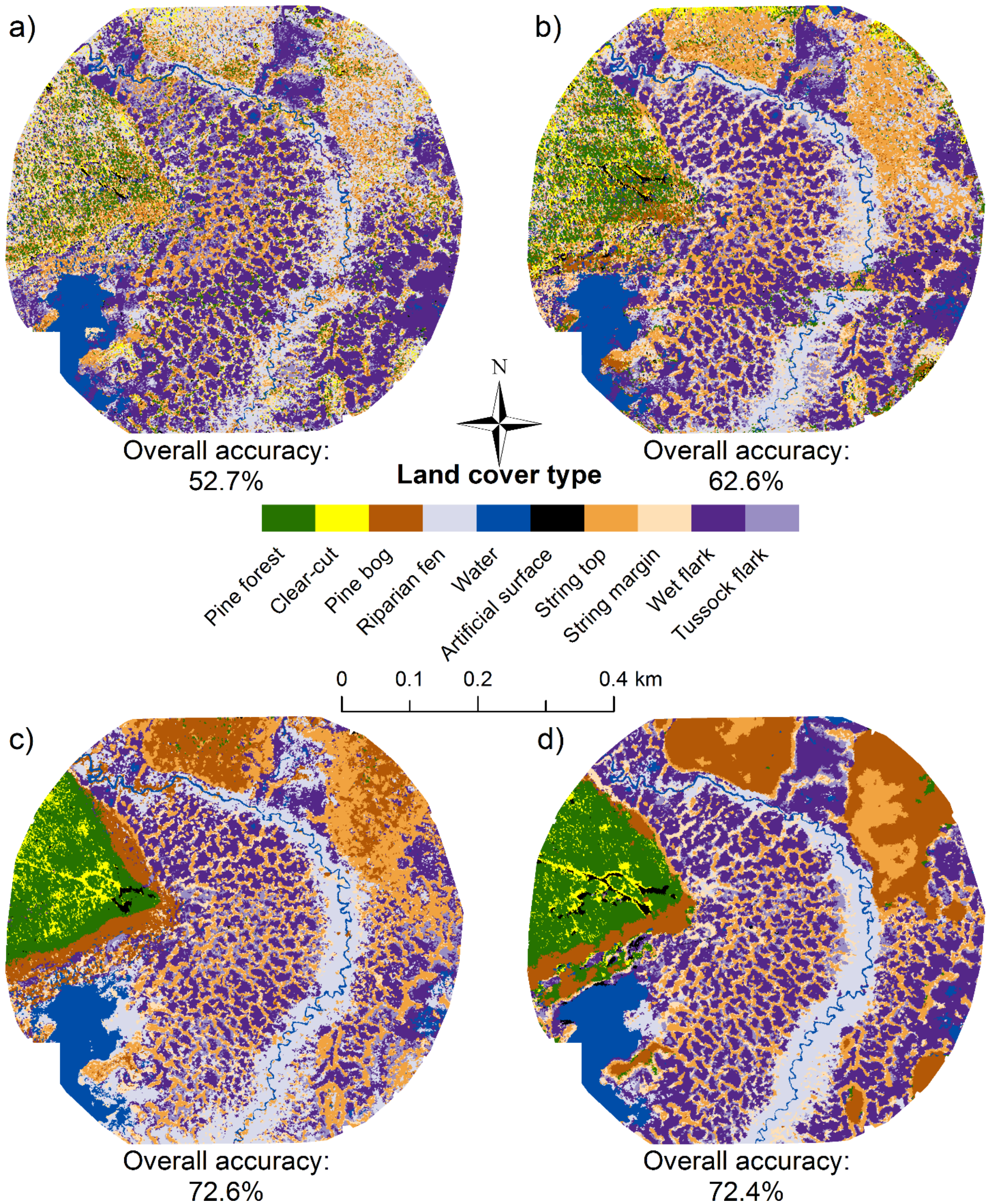


Figure 4. Unmanned aerial vehicle image classifications with the following segmentation scales: a) 2.5 m^2, b) 5 m^2, c) 10 m^2, and d) 20 m^2. The feature set is the one which yielded the highest classification accuracy (includes features

calculated from multiple images as well as topographical and vegetation height features, but excludes texture features). <2-column fitting image>



346

347 *Figure 5. Unmanned aerial vehicle (UAV) image classifications with 2.5 m*^{2} *segmentation scale and with the*
348 *following feature sets: a) UAV spectral bands only, b) UAV spectral bands and texture, c) UAV spectral bands,*
349 *topography, and vegetation height, and d) multiple images. <2-column fitting image>*

In the classification with the highest classification accuracy, wet flark had the largest areal coverage (28.5%) followed by pine bog (14.3%), string top (14.1%), and riparian fen (12.2%) (Tables 4 and S2). When compared with other classifications with the highest classification accuracy for each segmentation, the changes in areal coverage of different land cover types were generally small to moderate (between 3.0 *pp* decrease and 3.2 *pp* increase). However, when compared with classifications which included features only from one dataset (either spectral bands and indices or spectral and textural features), the differences in class-specific classification areal extent were larger (between 11.2 *pp* decrease and 10.0 *pp* increase).

Table 4. Areal coverage, and user's and producer's accuracies for the classification with highest overall accuracy (unmanned aerial vehicle segmentation with 2.5 m^2 segment size and features calculated from all datasets excluding texture) as well as minimum, mean, and maximum estimates over all classifications.

		Wet flark	Tussock flark	String top	String margin	Riparian fen	Pine bog	Pine forest	Clear-cut	Water	Non-vegetated
Areal coverage (%)	Best classification	28.5	5.4	14.1	9.7	12.2	14.3	8.3	1.3	5.8	0.3
	Minimum	20.5	2.9	13.5	4.8	4.9	3.1	4.9	0.5	4.3	0.1
	Mean	29.5	6.0	16.7	9.2	11.4	11.0	8.3	1.1	6.2	0.6
	Maximum	37.3	9.9	24.0	13.9	22.1	14.6	10.1	2.5	7.8	2.1
Producer's accuracy (%)	Best classification	82.1	54.6	82.8	43.8	84.7	94.7	92.5	90.0	100.0	79.6
	Minimum	59.0	13.2	30.5	15.0	12.7	7.1	10.6	6.7	73.3	43.9
	Mean	75.6	37.4	65.3	33.2	63.1	67.2	79.6	52.7	95.5	71.7
	Maximum	84.6	58.3	82.8	49.4	88.6	98.6	100.0	100.0	100.0	96.6
User's accuracy (%)	Best classification	89.9	33.5	78.0	52.6	78.2	99.9	100.0	89.0	87.7	89.0
	Minimum	59.9	8.1	34.6	16.5	16.7	10.2	6.6	9.4	37.4	42.1
	Mean	81.0	27.9	61.8	36.3	63.7	79.2	64.7	78.4	77.2	74.7
	Maximum	89.9	44.2	79.6	53.1	81.9	100.0	100.0	100.0	100.0	100.0

In the classification with the highest classification accuracy, class-specific user's and producer's accuracies varied between 33.5% and 100% (Tables 4 and S1). Lowest accuracies were obtained for string margin and tussock flark (33.5–54.6%), whereas for other land cover types, accuracies were > 78.0%. In other classifications with the highest classification accuracy for each segmentation, most of the classes had reasonable classification accuracies (lowest accuracy 44.6% when string margin, tussock flark, and clear-cut were excluded from the comparison). However, in the other classifications, some of the class-specific accuracies were extremely low (lowest user's accuracy 6.6% and lowest producer's accuracy 6.7%) (Tables 4 and S1).

In the patchiest land cover types, mean patch sizes were two orders of magnitude smaller than in the least patchy ones (Table 5, Table S3). Land cover types with the lowest classification accuracies (tussock flark and string margin) had the smallest mean patch sizes, whereas other fen land cover types (wet flark, string top, and riparian fen) had intermediate patch sizes, and pine bog and pine forest had the largest patch sizes (Table 5). Patch sizes were the smallest in the classifications with the smallest segmentation scale, and segmented image did not have a large effect on mean patch size (Table 5). The patch complexity was dependent on the spatial resolution of the segmented data: patches were the most complex in UAV segmentation based classifications and the least complex in classifications utilizing satellite image segmentation. The complexity increased when the segmentation scale increased, and there was relatively little difference in patch complexity between land cover types (Fig. 6).

Table 5. Mean patch size in m^2 for land cover classes in the classifications with the highest classification accuracy for each segmentation. Additionally, mean patch size over all land cover classes (furthest right column) and mean patch size for different land cover classes over all classifications (bottom row) are shown. UAV refers to unmanned aerial vehicle, WV-2 refers to WorldView-2, and PS refers to PlanetScope.

Segmentation	Wet flark	Tussock flark	String top	String margin	Riparian fen	Pine bog	Pine forest	Clear-cut	Water	Non-vegetated	**Mean**
UAV 2.5 m	113	10	41	10	58	410	830	11	97	23	**39**
UAV 5 m	159	16	53	15	110	711	1855	21	91	33	**61**
UAV 10 m	233	26	69	24	168	1106	3352	41	112	34	**93**
UAV 20 m	297	43	102	42	272	2040	7726	76	408	68	**155**
Aerial 2.5 m	88	7	39	9	42	265	650	19	56	14	**33**
Aerial 5 m	142	13	52	14	75	526	853	43	109	25	**55**
Aerial 10 m	215	23	68	22	144	873	1340	76	191	36	**88**
Aerial 20 m	323	39	94	37	213	1313	2585	123	320	41	**141**
WV-2 5 m	192	11	57	15	90	709	978	64	134	25	**59**
WV-2 10 m	229	12	62	19	119	741	1426	63	174	32	**72**
WV-2 20 m	408	26	114	45	318	1650	4026	74	362	40	**159**
PS 10 m	236	20	69	24	125	1229	3830	69	247	47	**89**
PS 20 m	362	25	85	67	238	1639	10346	112	353	69	**153**
Mean	**231**	**21**	**70**	**26**	**152**	**1016**	**3061**	**61**	**204**	**37**	

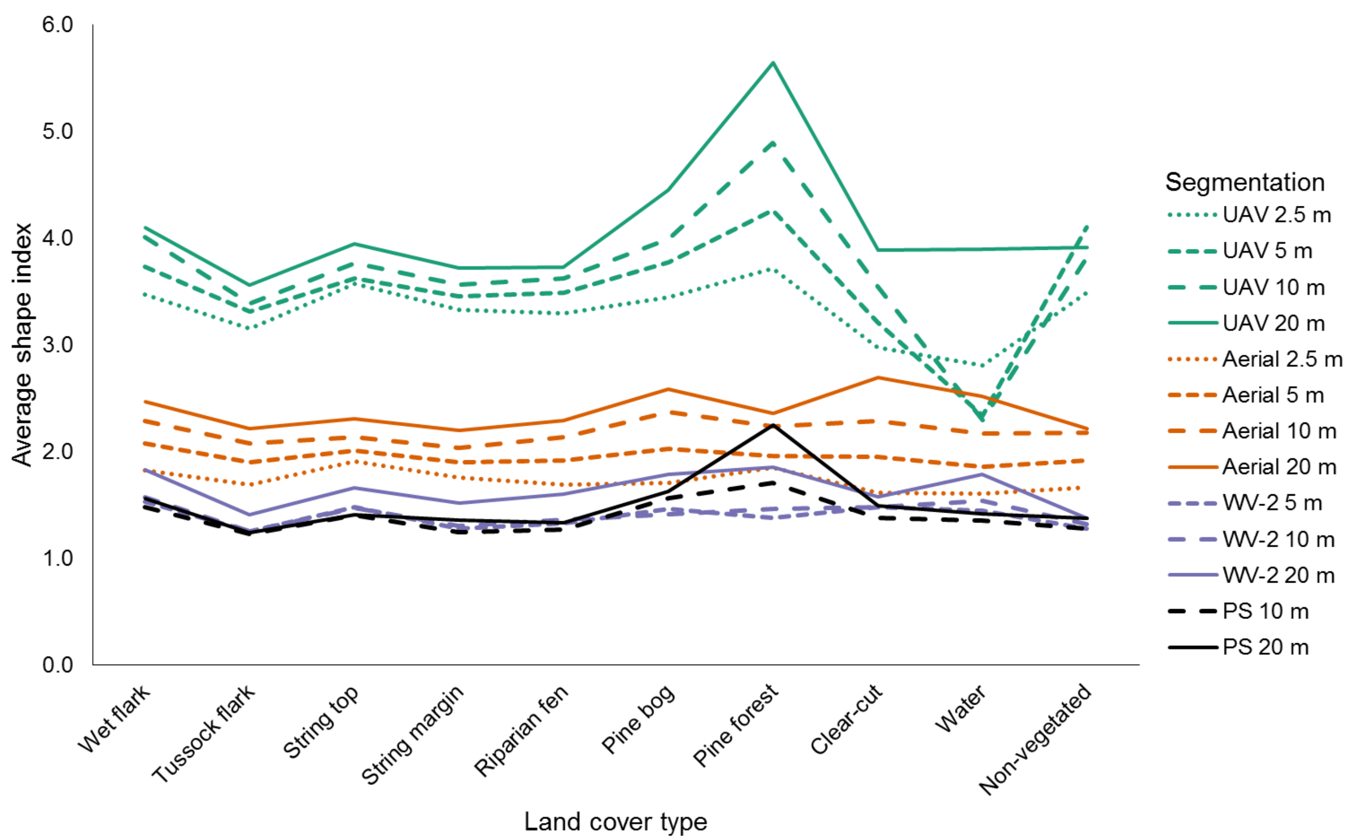


Figure 6. Patch complexities (shape index, y-axis) for the classification with the highest classification accuracy for each segmentation (lines) and land cover types (x-axis). <2-column fitting image>

4. Discussion and conclusions

Our results show that the highest classification accuracies are obtained when using features calculated from multiple datasets (Figs 2 and 5). This means that there is a need at least for multiple optical datasets or one optical dataset and data about topography and vegetation height when mapping vegetation spatially heterogeneous landscapes. However, in order to have the highest classification accuracies, both multiple optical datasets and topography/vegetation height features are needed. According to our results, textural features increase classification accuracy notably when the feature set is otherwise quite limited, such as when features are calculated from one dataset only (Palace et al. 2018). However, textural features do not increase classification accuracy when multiple optical datasets as well as topography and vegetation height features are used in classification (Fig. 2). Less useful textural and other features could also be removed from the classification using feature selection algorithms which include e.g., random forest wrappers such as Boruta (Kursa and Rudnicki 2010). Feature selection could thus remove the not useful or even harmful textural features and leave useful textural features in the final classification. However, in our case, Boruta runs indicated that all features were important in different classifications, and also random forest out-of-bag error rates did not change when we tested a different amount of the most important features. Earlier, it has been shown that classification accuracy might slightly increase when only the most important features are left in the classification and some of the less important features which are deemed important are left out (Räsänen et al. 2014).

The highest classification accuracies were obtained with UAV image based classifications. However, we argue that UAV image is not necessarily needed for classifying fine-resolution vegetation patterns in patchy landscapes, because almost as high classification accuracies were obtained when using a 0.5 m pixel size aerial image as a basis for the classification (Table 3, Fig. 2 and 3). Actually, when using only spectral features calculated only from dataset, aerial image-based classifications had slightly higher classification accuracies than UAV-based classifications (Fig. 2). In turn, in UAV-based classifications, the inclusion of texture boosted classification accuracy more than in aerial image-based classifications. Classification accuracies were notably smaller when both UAV and aerial image were excluded from the classification (Table 3, Fig. 2), although visually acceptable maps were produced also with a combination of very high resolution satellite imagery and aerial lidar (Fig. 3).

Our results do not necessarily suggest that UAV mappings are not useful. Firstly, in our case, the UAV image was especially useful for training dataset construction, and the use of a coarser resolution aerial image in training dataset construction would have been more demanding. Of course, the training dataset could be constructed using field observations and field-measured GPS information only, but also in this case the UAV image was useful in double checking the relative positional accuracy of the field observations. Secondly, in many areas across the globe, aerial imagery and lidar data are not available and data collection of such data is expensive. In these areas, UAV offers a cheaper and easier solution to collect data from areas with limited areal extent (Anderson and Gaston 2013; Palace et al. 2018). Considering the first two points, our results indicate that the highest spatial resolution UAV images over small areas could be used for training or validation dataset construction (Räsänen et al. 2019a), and lower spatial resolution UAV data over a larger area could be collected for classification and other mapping purposes. Thirdly, related to the two first points, UAV data can be used for upscaling purposes, and utilized as a training data for satellite imagery based mappings (Riihimäki et al. 2019). Fourthly, we used data collected only from one UAV flight. Results could have been different if we had used multiple UAV images, as it has been shown that inclusion of images taken at different phenological stages boost classification accuracy (Chen et al. 2017b; Dudley et al. 2015; Halabisky et al. 2018; Lu et al. 2017; Lucas et al. 2011). Fifthly, our UAV flight had only an RGB camera onboard. Classification accuracies could have been higher if we had used visible and near-infrared (VNIR) or hyperspectral cameras (Cao et al. 2018; Sankey et al. 2018) or UAV lidar (Sankey et al. 2018). These instruments would have allowed more detailed mapping of spectral and structural properties of different land cover types. Already in our case, classification accuracies were considerably higher when we combined spectral UAV data with vegetation height and topography data collected using airborne lidar and UAV. However, the inclusion of hyperspectral or lidar data would have increased the cost and time required for data collection and processing (Palace et al. 2018). Sixthly, based on visual inspection, patch boundaries delineated from the UAV image followed the actual patch boundaries in the field more accurately than patch boundaries delineated using other data. This was also supported by the fact that patches were the most complex when classifications were based on UAV segmentations (Fig. 6). Although the classification accuracy was only slightly lower with more general patch boundaries in our case, it could be more useful to delineate patches as realistically as possible in some other tasks (Lang et al. 2014).

According to our results, segmentation scale has an effect on classification accuracy, but this effect is mostly minor (Table 3, Figs 2 and 4). Our results suggest that there might be a lower limit for optimal segmentation scale, probably in our case 2.5 m^2. Below this limit, finer scale segmentations do not increase classification accuracy any further but might instead lead to noise in the classification and lower classification accuracies (Dronova et al. 2012; Räsänen et al. 2013; Yue et al. 2012). On the other hand, when segmentation scale is slightly increased from the upper limit of the optimal scale (in our case 5 m^2), the decreases in classification accuracy are generally small. When the segmentation scale grows too large (in our case 20 m^2 and above), decreases in classification accuracy can be larger. However, we tested only four different segmentation scales and did not test how the changes in the other segmentation parameters affect classification accuracy. Earlier, it has been shown that changing segmentation scale has a large effect on classification accuracy (Dronova et al. 2012), but also the segmentation method and other parameters have an effect (Dronova et al. 2012; Räsänen et al. 2013). Furthermore, also multi-resolution segmentations could be conducted in which different segmentation scale is used for delineating patches of different land cover types (Blaschke et al. 2014; Dronova 2015), but classification based on a single-scale segmentation is easier to implement.

It is evident that optimal segmentation scale for classification depends on what the real patchiness of vegetation and land cover types in the study area is. Northern peatlands are extremely mosaicked in their structure, and this is the case also with our study area. A mean segment size as small as 2.5 m^2 was found to produce the most accurate classification results, although the difference in classification accuracy was very small when compared to 5 m^2 segment size. The patchiness of the peatland landscape is also illustrated by the fact that some of the fen land cover types, especially tussock flark and string margin, had very low mean patch size while the mean patch size for forest and pine bog was many times larger (Table 5). This indicates that smaller segmentation scale and higher resolution data are needed for mapping fen than for mapping forest vegetation. This is an important finding from a carbon dynamics research point of view, as fens are very critical especially in methane circulation (Marushchak et al. 2016). However, before making a strong generalization about the landscape patchiness, the optimal segmentation scale in several different landscapes should be tested. In any case, nowadays, there are tools and images to study this question at a fine scale, while this was not possible some years ago when very high resolution data were not widely available.

We calculated confidence intervals for each classification, although we could have also tested if differences in classification performance are statistically significant. However, the tests of significance, such as the widely used McNemar test (Foody 2004) are mostly based on pairwise comparisons, and such comparisons would have been challenging in our case with approximately 3000 comparison pairs. Overall, both confidence intervals and statistical tests are extremely sensitive to sample size (Foody 2009), and confidence intervals we reported should be treated with caution. We set the sample size to the number of 5 cm pixels within our reference polygons (ca. 30000). If we had set sample size to the number of reference polygons (412), confidence limits would have been approximately nine times wider. In that case, each classification would have been allowed to have only one value within each reference polygon. However, the land cover type boundaries of different classifications are often located within reference polygons, and classifications can thus be partly correct per each reference polygon (Fig. S1). In these cases, choosing the suitable reference unit (polygon vs. pixel vs. aerial unit such as m^2) is somewhat arbitrary. Although the chosen reference unit has small to moderate effect for commonly used accuracy metrics such as user's, producer's, and overall accuracy, its effect can be disproportionally large for statistical tests. This highlights the difficulty of evaluating classifier performance for classifications with differing pixel sizes and boundaries, and also for object-based classifications. Numerous polygon or object-based accuracy assessment methods have been suggested, but those methods have unresolved conceptual challenges (Ye et al. 2018).

When classifying vegetation or other patterns using a fine-resolution approach, there are strict requirements for high locational and geometrical precision (Müllerová et al. 2017). If the pixel size is some centimeters, also locational accuracy should be some centimeters and high-precision GPS devices should be used. The need for high positional accuracy is evident especially if one is merging multiple different remote sensing datasets and/or field-measured data in the mapping. In practice, each dataset should be in the same correct position. Although UAV images can be orthocorrected with ground control points and small markers in the field, similar methods are more difficult to implement for satellite images, as their pixel size is usually meters instead of centimeters. Therefore, it might be that satellite images are not exactly in the same position as the

UAV data, which might affect mapping accuracy. Also in our case, we could not verify the exact positional accuracy of the satellite imagery due to coarser pixel size and few easily mappable (man-made) features in the study area. However, classifications using satellite imagery were still feasible, which suggests that positional accuracy was sufficient.

When land cover classification is linked to biogeochemical cycles such as carbon flux data measured with chambers or eddy covariance towers, it is important that the relative proportion of different land cover types is predicted accurately and that the patches of different land cover types are approximately in the correct position (Davidson et al. 2017; Treat et al. 2018). However, small errors in patch location or form are not that worrisome. Considering the requirement that relative proportions of land cover types are predicted accurately, our results suggest that it is important to include multiple datasets in the classification. However, according to our results, if only one dataset (i.e. UAV, aerial imagery, WV-2 or PS) is used in classification, the relative proportion of different land cover types may not be accurately predicted. Hence, our results suggest that finer resolution data (such as UAV or aerial imagery) may be left out from the classification if the goal is to map relative proportions of different classes and there is no need to maximize classifier performance. Coarser resolution datasets and segmentations provide sufficient mapping accuracy for relative proportions of land cover types, especially if mapping is conducted in areas with rather large areal extent. In the high northern latitudes, widely available very-high resolution satellite datasets such as PS and ArcticDEM (Porter et al. 2018) could thus be used for different fine-scale mapping approaches. Nevertheless, we concentrated only on one study area and did not test what the implications of the different classification options is for applications such as carbon flux modeling. Therefore, more research should be conducted to test what kind of datasets and what spatial resolution should be used in different tasks.

It has been reported that there have been changes in the high-latitude vegetation patterns during the past decades (Guay et al. 2014; Jorgenson and Grosse 2016; Macias-Fauria et al. 2012). Also in the future, vegetation and land cover patterns in the north will probably change rapidly due to warming climate. Previously, it has been argued that there should be standardized approaches for fine-scale change detection (Jorgenson and Grosse 2016). Our results imply that sub-meter resolution data is required for tracking changes in vegetation patches and their spatial location, but very high resolution satellite data (< 5 m) may be sufficient for detecting changes in areal cover of different land cover or vegetation types. Overall, repeated standardized UAV mappings could offer a low-cost method for tracking fine-scale changes. Furthermore, it has been discussed that UAVs provide a powerful approach to track fine-scale phenology (Berra et al. 2019; Klosterman et al. 2018).

Finally, instead of using crisp maps of land cover or habitat types, fuzzy or continuous maps could be used in mapping vegetation patterns (Foody 1997; Rapinel et al. 2018; Rocchini 2014, Räsänen et al. 2019c). In these maps, boundaries between different land cover types are not exact, and/or specific areas might be a mixture of multiple mapped properties such as vegetation communities. These methods might also help in mapping land cover types with low classification accuracy such as tussock flark and string margin in our case (Table 4). Although the continuous and fuzzy maps are often more realistic, they might be less intuitive to use and less straightforward to interpret. They could also be produced from coarser pixel sized data, which would allow land cover products with a larger extent but lower accuracy. Therefore, it seems that the most feasible way is to produce multiple maps showing spatial patterns of different environmental properties and use the different maps flexibly for different purposes.

Acknowledgments: We thank Jani Antila, Holtti Hakonen, and Olivia Kuuri-Riutta for field assistance, and Kari Mäenpää for conducting the UAV flight. This work was supported by the Academy of Finland [grant number 296423].

References

Amani, M., Salehi, B., Mahdavi, S., Granger, J.E., Brisco, B., & Hanson, A. (2017). Wetland Classification Using Multi-Source and Multi-Temporal Optical Remote Sensing Data in Newfoundland and Labrador, Canada. *Canadian Journal of Remote Sensing, 43*, 360-373

Anderson, K., & Gaston, K.J. (2013). Lightweight unmanned aerial vehicles will revolutionize spatial ecology. *Frontiers in Ecology and the Environment, 11*, 138-146

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. In, *Proceedings of SPIE - The International Society for Optical Engineering*

Aurela, M., Laurila, T., & Tuovinen, J.P. (2001). Seasonal CO_2 balances of a subarctic mire. *Journal of Geophysical Research Atmospheres, 106*, 1623-1637

Aurela, M., Laurila, T., & Tuovinen, J.-P. (2002). Annual CO2 balance of a subarctic fen in northern Europe: Importance of the wintertime efflux. Journal of Geophysical Research: Atmospheres, 107, ACH 17-11-ACH 17-12

Aurela, M., Laurila, T., & Tuovinen, J.P. (2004). The timing of snow melt controls the annual CO2 balance in a subarctic fen. Geophysical Research Letters, 31, L16119 16111-16114

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*, 11289-11301

Bartsch, A., Hofler, A., Kroisleitner, C., & Trofaier, A.M. (2016). Land Cover Mapping in Northern High Latitude Permafrost Regions with Satellite Data: Achievements and Remaining Challenges. *Remote Sensing, 8*

Berra, E.F., Gaulton, R., & Barr, S. (2019). Assessing spring phenology of a temperate woodland: A multiscale comparison of ground, unmanned aerial vehicle and Landsat satellite observations. Remote Sensing of Environment, 223, 229-242

Blaschke, T., Hay, G.J., Kelly, M., Lang, S., Hofmann, P., Addink, E., Feitosa, R.Q., van der Meer, F., van der Werff, H., & van Coillie, F. (2014). Geographic object-based image analysis–towards a new paradigm. *ISPRS Journal of Photogrammetry and Remote Sensing, 87*, 180-191

Breiman, L. (2001). Random forests. *Machine Learning, 45*, 5-32

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)

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. *Remote Sensing, 10*

Castilla, G., & Hay, G.J. (2008). Image objects and geographic objects. In, *Lecture Notes in Geoinformation and Cartography* (pp. 91-110)

Chen, B., Chen, L., Lu, M., & Xu, B. (2017a). Wetland mapping by fusing fine spatial and hyperspectral resolution images. *Ecological Modelling, 353*, 95-106

Chen, B., Huang, B., & Xu, B. (2017b). Multi-source remotely sensed data fusion for improving land cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing, 124*, 27-39

Chen, J., Chen, J., Liao, A., Cao, X., Chen, L., Chen, X., He, C., Han, G., Peng, S., Lu, M., Zhang, W., Tong, X., & Mills, J. (2015). Global land cover mapping at 30m resolution: A POK-based operational approach. ISPRS Journal of Photogrammetry and Remote Sensing, 103, 7-27

Chen, W., Li, X., He, H., & Wang, L. (2018). Assessing different feature sets' effects on land cover classification in complex surface-mined landscapes by ZiYuan-3 satellite imagery. *Remote Sensing, 10*

Clinton, N., Holt, A., Scarborough, J., Yan, L.I., & Gong, P. (2010). Accuracy assessment measures for object-based image segmentation goodness. *Photogrammetric Engineering and Remote Sensing, 76*, 289-299

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., & Böhner, J. (2015). System for Automated Geoscientific Analyses (SAGA) v. 2.1.4 *Geoscientific Model Development, 8*, 1991-2007

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*, 67-80

Costa, H., Foody, G.M., & Boyd, D.S. (2018). Supervised methods of image segmentation accuracy assessment in land cover mapping. *Remote Sensing of Environment, 205*, 338-351

Davidson, S.J., Santos, M.J., Sloan, V.L., Reuss-Schmidt, K., Phoenix, G.K., Oechel, W.C., & Zona, D. (2017). Upscaling CH_4 fluxes using high-resolution imagery in Arctic Tundra ecosystems. *Remote Sensing, 9*

Dronova, I. (2015). Object-based image analysis in wetland research: A review. *Remote Sensing, 7*, 6380-6413

Dronova, I., Gong, P., Clinton, N.E., Wang, L., Fu, W., Qi, S., & Liu, Y. (2012). Landscape analysis of wetland plant functional types: The effects of image segmentation scale, vegetation classes and classification methods. *Remote Sensing of Environment, 127*, 357-369

Du, P.J., Samat, A., Waske, B., Liu, S.C., & Li, Z.H. (2015). Random Forest and Rotation Forest for fully polarized SAR image classification using polarimetric and spatial features. *ISPRS Journal of Photogrammetry and Remote Sensing, 105*, 38-53

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

Foody, G.M. (1997). Fully fuzzy supervised classification of land cover from remotely sensed imagery with an artificial neural network. *Neural Computing & Applications, 5*, 238-247

Foody, G.M. (2004). Thematic Map Comparison: Evaluating the Statistical Significance of Differences in Classification Accuracy. *Photogrammetric Engineering & Remote Sensing, 70,* 627-633

Foody, G. M. (2008). Harshness in image classification accuracy assessment. *International Journal of Remote Sensing 29,* 3137-3158

Foody, G. M. (2009). Classification accuracy comparison: Hypothesis tests and the use of confidence intervals in evaluations of difference, equivalence and non-inferiority. Remote Sensing of Environment 113, 1658-1663

Franklin, S.E., & Ahmed, O.S. (2017). Object-based wetland characterization using radarsat-2 quad-polarimetric SAR data, landsat-8 OLI imagery, and airborne lidar-derived geomorphometric variables. *Photogrammetric Engineering and Remote Sensing*, *83*, 27-36

Georganos, S., Grippa, T., Lennert, M., Vanhuysse, S., Johnson, B.A., & Wolff, E. (2018). Scale matters: Spatially Partitioned Unsupervised Segmentation Parameter Optimization for large and heterogeneous satellite images. *Remote Sensing, 10*

Goetz, S.J., Baccini, A., Laporte, N.T., Johns, T., Walker, W., Kellndorfer, J., Houghton, R.A., Sun, M.J.C.B., & Management (2009). Mapping and monitoring carbon stocks with satellite observations: a comparison of methods. *Carbon Balance and Management 4*, 2

Gong, P., Wang, J., Yu, L., Zhao, Y., Zhao, Y., Liang, L., Niu, Z., Huang, X., Fu, H., Liu, S., Li, C., Li, X., Fu, W., Liu, C., Xu, Y., Wang, X., Cheng, Q., Hu, L., Yao, W., Zhang, H., Zhu, P., Zhao, Z., Zhang, H., Zheng, Y., Ji, L., Zhang, Y., Chen, H., Yan, A., Guo, J., Yu, L., Wang, L., Liu, X., Shi, T., Zhu, M., Chen, Y., Yang, G., Tang, P., Xu, B., Giri, C., Clinton, N., Zhu, Z., Chen, J., & Chen, J. (2013). Finer resolution observation and monitoring of global land cover: first mapping results with Landsat TM and ETM+ data. *International Journal of Remote Sensing*, 34, 2607-2654

Grizonnet, M., Michel, J., Poughon, V., Inglada, J., Savinaud, M., & Cresson, R. (2017). Orfeo ToolBox: open source processing of remote sensing images. *Open Geospatial Data, Software and Standards, 2*, 15

Guay, K.C., Beck, P.S.A., Berner, L.T., Goetz, S.J., Baccini, A., & Buermann, W. (2014). Vegetation productivity patterns at high northern latitudes: a multi-sensor satellite data assessment. *Global Change Biology*, *20,* 3147-3158

Guisan, A., Weiss, S.B., & Weiss, A.D. (1999). GLM versus CCA spatial modeling of plant species distribution. *Plant Ecology, 143*, 107-122

Halabisky, M., Babcock, C., & Moskal, L.M. (2018). Harnessing the temporal dimension to improve object-based image analysis classification of wetlands. *Remote Sensing*, *10*

Hall-Beyer, M. (2017). Practical guidelines for choosing GLCM textures to use in landscape classification tasks over a range of moderate spatial scales. *International Journal of Remote Sensing, 38*, 1312-1338

Haralick, R.M., Dinstein, I., & Shanmugam, K. (1973). Textural Features for Image Classification. *IEEE Transactions on Systems, Man and Cybernetics, SMC-3*, 610-621

Jorgenson, M.T., & Grosse, G. (2016). Remote Sensing of Landscape Change in Permafrost Regions. *Permafrost and Periglacial Processes, 27*, 324-338

Jung, M., Henkel, K., Herold, M., & Churkina, G. (2006). Exploiting synergies of global land cover products for carbon cycle modeling. *Remote Sensing of Environment, 101*, 534-553

Juutinen, S., Virtanen, T., Kondratyev, V., Laurila, T., Linkosalmi, M., Mikola, J., Nyman, J., Räsänen, A., Tuovinen, J.-P., & Aurela, M. (2017). Spatial variation and seasonal dynamics of leaf-area index in the arctic tundra-implications for linking ground observations and satellite images. *Environmental Research Letters, 12*, 095002

Klosterman, S., Melaas, E., Wang, J.A., Martinez, A., Frederick, S., O'Keefe, J., Orwig, D.A., Wang, Z., Sun, Q., Schaaf, C., Friedl, M., & Richardson, A.D. (2018). Fine-scale perspectives on landscape phenology from unmanned aerial vehicle (UAV) photography. *Agricultural and Forest Meteorology, 248*, 397-407

Kursa, M.B., & Rudnicki, W.R. (2010). Feature Selection with the Boruta Package. *Journal of Statistical Software, 36*, 1-13

Laidler, G.J., & Treitz, P. (2003). Biophysical remote sensing of arctic environments. *Progress in Physical Geography, 27*, 44-68

Lang, S., Kienberger, S., Tiede, D., Hagenlocher, M., & Pernkopf, L. (2014). Geons - domain-specific regionalization of space. *Cartography and Geographic Information Science, 41*, 214-226

Lang, S., & Tiede, D. (2003). vLATE Extension für ArcGIS - vektorbasiertes Tool zur quantitativen Landschaftsstrukturanalyse. In, *ESRI Anwenderkonferenz 2003*. Innsbruck

Lehmann, J.R.K., Münchberger, W., Knoth, C., Blodau, C., Nieberding, F., Prinz, T., Pancotto, V.A., & Kleinebecker, T. (2016). High-resolution classification of south patagonian peat bog microforms reveals potential gaps in up-scaled CH_4 fluxes by use of Unmanned Aerial System (UAS) and CIR imagery. *Remote Sensing, 8*

Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. *R News, 2*, 18-22

Liu, C., Frazier, P., & Kumar, L. (2007). Comparative assessment of the measures of thematic classification accuracy. *Remote Sensing of Environment*, *107*, 606-616.

Lovitt, J., Rahman, M.M., & McDermid, G.J. (2017). Assessing the value of UAV photogrammetry for characterizing terrain in complex peatlands. *Remote Sensing, 9*

Lu, M., Chen, B., Liao, X., Yue, T., Yue, H., Ren, S., Li, X., Nie, Z., & Xu, B. (2017). Forest types classification based on multi-source data fusion. *Remote Sensing*, *9*

Lucas, R., Medcalf, K., Brown, A., Bunting, P., Breyer, J., Clewley, D., Keyworth, S., & Blackmore, P. (2011). Updating the Phase 1 habitat map of Wales, UK, using satellite sensor data. *ISPRS Journal of Photogrammetry and Remote Sensing*, *66*, 81-102

Luo, S., Wang, C., Xi, X., Zeng, H., Li, D., Xia, S., & Wang, P. (2016). Fusion of airborne discrete-return LiDAR and hyperspectral data for land cover classification. *Remote Sensing*, *8*

Ma, L., Li, M., Ma, X., Cheng, L., Du, P., & Liu, Y. (2017). A review of supervised object-based land-cover image classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, *130*, 277-293

Macias-Fauria, M., Forbes, B.C., Zetterberg, P., & Kumpula, T. (2012). Eurasian Arctic greening reveals teleconnections and the potential for structurally novel ecosystems. *Nature Climate Change*, *2*, 613

Maanavilja, L., Riutta, T., Aurela, M., Pulkkinen, M., Laurila, T., & Tuittila, E.S. (2011). Spatial variation in CO_2 exchange at a northern aapa mire. *Biogeochemistry, 104*, 325-345

Mahdavi, S., Salehi, B., Granger, J., Amani, M., Brisco, B., & Huang, W. (2018). Remote sensing for wetland classification: a comprehensive review. *GIScience and Remote Sensing, 55*, 623-658

Marushchak, M.E., Friborg, T., Biasi, C., Herbst, M., Johansson, T., Kiepe, I., Liimatainen, M., Lind, S.E., Martikainen, P.J., Virtanen, T., Soegaard, H., & Shurpali, N.J. (2016). Methane dynamics in the subarctic tundra: combining stable isotope analyses, plot- and ecosystem-scale flux measurements. *Biogeosciences, 13*, 597-608

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*, 1425-1432

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

Mishra, V.N., Prasad, R., Rai, P.K., Vishwakarma, A.K., & Arora, A. (2018). Performance evaluation of textural features in improving land use/land cover classification accuracy of heterogeneous landscape using multi-sensor remote sensing data. *Earth Science Informatics*

Müllerová, J., Bartaloš, T., Brůna, J., Dvořák, P., & Vítková, M. (2017). Unmanned aircraft in nature conservation: an example from plant invasions. *International journal of remote sensing, 38*, 2177-2198

Palace, M., Herrick, C., DelGreco, J., Finnell, D., Garnello, A.J., McCalley, C., McArthur, K., Sullivan, F., & Varner, R.K. (2018). Determining subarctic peatland vegetation using an unmanned aerial system (UAS). *Remote Sensing, 10*

Pettorelli, N., Wegmann, M., Skidmore, A., Mücher, S., Dawson, T.P., Fernandez, M., Lucas, R., Schaepman, M.E., Wang, T., O'Connor, B., Jongman, R.H.G., Kempeneers, P., Sonnenschein, R., Leidner, A.K., Böhm, M., He, K.S., Nagendra, H., Dubois, G., Fatoyinbo, T., Hansen, M.C., Paganini, M., de Klerk, H.M., Asner, G.P., Kerr, J.T., Estes, A.B., Schmeller, D.S., Heiden, U., Rocchini, D., Pereira, H.M., Turak, E., Fernandez, N., Lausch, A., Cho, M.A., Alcaraz-Segura, D., McGeoch, M.A., Turner, W., Mueller, A., St-Louis, V., Penner, J., Vihervaara, P., Belward, A., Reyers, B., & Geller, G.N. (2016). Framing the concept of satellite remote sensing essential biodiversity variables: challenges and future directions. *Remote Sensing in Ecology and Conservation*, *2*, 122-131

Planet Team (2017). Planet Application Program Interface: In Space for Life on Earth. In. San Francisco, CA

Porter, C., Morin, P., Howat, I., Noh, M.-J., Bates, B., Peterman, K., Keesey, S., Schlenk, M., Gardiner, J., Tomko, K., Willis, M., Kelleher, C., Cloutier, M., Husby, E., Foga, S., Nakamura, H., Platson, M., Wethington, M., Jr., Williamson, C., Bauer, G., Enos, J., Arnold, G., Kramer, W., Becker, P., Doshi, A., D’Souza, C., Cummens, P., Laurier, F., & Bojesen, M. (2018). ArcticDEM. *Harvard Dataverse, V1*, https://doi.org/10.7910/DVN/OHHUKH

Prošek, J., & Šímová, P. (2019). UAV for mapping shrubland vegetation: Does fusion of spectral and vertical information derived from a single sensor increase the classification accuracy? *International Journal of Applied Earth Observation and Geoinformation*, *75*, 151-162

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*, 678-693

Riihimäki, H., Luoto, M., & Heiskanen, J. (2019). Estimating fractional cover of tundra vegetation at multiple scales using unmanned aerial systems and optical satellite data. Remote Sensing of Environment, 224, 119-132

Rocchini, D. (2014). Fuzzy species distribution models: A way to represent plant communities spatially. *Journal of Vegetation Science*, *25*, 317-318

Rodriguez-Galiano, V.F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J.P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, *67*, 93-104

Rouse, J.W.J., Haas, R.H., Schell, J.A., & Deering, D.W. (1973). Monitoring vegetation systems in the Great Plains with ERTS. In, *Third Earth Resources Technology Satellite-1 Symposium* (pp. 309-317). Washington, DC: NASA

Räsänen, A., Elsakov, V., & Virtanen, T. (2019a). Usability of one-class classification in mapping and detecting changes in bare peat surfaces in the tundra. *International journal of remote sensing, 40*, 4083-4103

Räsänen, A., Juutinen, S., Aurela, M., & Virtanen, T. (2019b). Predicting aboveground biomass in Arctic landscapes using very high spatial resolution satellite imagery and field sampling. *International journal of remote sensing, 40*, 1175-1199

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 boreal forest habitat type classification. *ISPRS Journal of Photogrammetry and Remote Sensing, 94,* 169-182

Räsänen, A., Rusanen, A., Kuitunen, M., & Lensu, A. (2013). What makes segmentation good? A case study in boreal forest habitat mapping. *International journal of remote sensing, 34*, 8603-8627

Räsänen, A., Juutinen, S., Tuittila, E.-S., Aurela, M., & Virtanen, T. (2019c). Comparing ultra-high spatial resolution remote sensing methods in mapping peatland vegetation. *Journal of Vegetation Science*, accepted manuscript

Sankey, T.T., McVay, J., Swetnam, T.L., McClaran, M.P., Heilman, P., & Nichols, M. (2018). UAV hyperspectral and lidar data and their fusion for arid and semi-arid land vegetation monitoring. *Remote Sensing in Ecology and Conservation*, *4*, 20-33

Schneider, J., Grosse, G., & Wagner, D. (2009). Land cover classification of tundra environments in the Arctic Lena Delta based on Landsat 7 ETM+ data and its application for upscaling of methane emissions. *Remote Sensing of Environment, 113*, 380-391

Shadaydeh, M., Zlinszky, A., Manno-Kovacs, A., & Sziranyi, T. (2017). Wetland mapping by fusion of airborne laser scanning and multi-temporal multispectral satellite imagery. *International journal of remote sensing, 38*, 7422-7440

Sibaruddin, H.I., Shafri, H.Z.M., Pradhan, B., & Haron, N.A. (2018). Comparison of pixel-based and object-based image classification techniques in extracting information from UAV imagery data. In, *IOP Conference Series: Earth and Environmental Science*

Siewert, M.B., Hanisch, J., Weiss, N., Kuhry, P., Maximov, T.C., & Hugelius, G. (2015). Comparing carbon storage of Siberian tundra and taiga permafrost ecosystems at very high spatial resolution. *Journal of Geophysical Research-Biogeosciences, 120*, 1973-1994

Stratoulias, D., Balzter, H., Zlinszky, A., & Tóth, V.R. (2018). A comparison of airborne hyperspectral-based classifications of emergent wetland vegetation at lake Balaton, Hungary. *International journal of remote sensing, 39*, 5689-5715

Treat, C.C., Marushchak, M.E., Voigt, C., Zhang, Y., Tan, Z., Zhuang, Q., Virtanen, T.A., Räsänen, A., Biasi, C., Hugelius, G., Kaverin, D., Miller, P.A., Stendel, M., Romanovsky, V., Rivkin, F., Martikainen, P.J., & Shurpali, N.J. (2018). Tundra landscape heterogeneity, not interannual variability, controls the decadal regional carbon balance in the Western Russian Arctic. *Global Change Biology, 24*, 5188-5204

Virtanen, T., & Ek, M. (2014). The fragmented nature of tundra landscape. *International Journal of Applied Earth Observation and Geoinformation*, *27*, 4-12

Ye, S., Pontius Jr., R. G., & Rakshit, R. (2018). A review of accuracy assessment for object-based image analysis: From per-pixel to per-polygon approaches. *ISPRS Journal of Photogrammetry and Remote Sensing*, *141*, 137–147

Yue, A., Yang, J., Zhang, C., Su, W., Yun, W., Zhu, D., Liu, S., & Wang, Z. (2012). The optimal segmentation scale identification using multispectral worldview-2 images. *Sensor Letters*, *10*, 285-291