Author(s): Räsänen, Aleksi; Kuitunen, Markku; Tomppo, Erkki; Lensu, Anssi

Title: Coupling high-resolution satellite imagery with ALS-based canopy height model and digital elevation model in object-based boreal forest habitat type classification

Year: 2014

Version:

Please cite the original version:

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(August 2014), 169-182. https://doi.org/10.1016/j.isprsjprs.2014.05.003

Coupling high-resolution satellite imagery with ALS-based canopy height model and digital elevation model in object-based boreal forest habitat type classification

Aleksi Räsänen[a], Markku Kuitunen[b], Erkki Tomppo[d] & Anssi Lensu[c]

[a] Corresponding author, Department of Biological and Environmental Science, P.O. Box 35, FI-40014 University of Jyväskylä, Finland, email: t.aleksi.rasanen@jyu.fi, phone number: +358-50-3719512
[b] Department of Biological and Environmental Science, P.O. Box 35, FI-40014 University of Jyväskylä, Finland, email: markku.t.kuitunen@jyu.fi
[c] Department of Biological and Environmental Science, P.O. Box 35, FI-40014 University of Jyväskylä, Finland, email: anssi.lensu@jyu.fi
[d] The Finnish Forest Research Institute, P.O. Box 18, FI-01301 Vantaa, Finland, email: erkki.tomppo@metla.fi

Abstract

We developed a classification workflow for boreal forest habitat type mapping. In object-based image analysis framework, Fractal Net Evolution Approach segmentation was combined with random forest classification. High-resolution WorldView-2 imagery was coupled with ALS based canopy height model and digital terrain model. We calculated several features (e.g. spectral, textural and topographic) per image object from the used datasets. We tested different feature set alternatives; a classification accuracy of 78.0 % was obtained when all features were used. The highest classification accuracy (79.1 %) was obtained when the amount of features was reduced from the initial 328 to the 100 most important using Boruta feature selection algorithm and when ancillary soil and land-use GIS-datasets were used. Although Boruta could rank the importance of features, it could not separate unimportant features from the important ones. Classification accuracy was bit lower (78.7 %) when the classification was performed separately on two areas: the areas above and below 1 m vertical distance from the nearest stream. The data split, however, improved the classification accuracy of mire habitat types and streamside habitats, probably because their proportion in the below 1 m data was higher than in the other datasets. It was found that several types of data are needed to get the highest classification accuracy whereas omitting some feature groups reduced the classification accuracy. A major habitat type in the study area was mesic forests in different successional stages. It was found that the inner heterogeneity of different mesic forest age groups was large and other habitat types were often inside this heterogeneity.

Keywords

habitat type mapping; multispectral imagery; ALS; object-based image analysis; random forest classifier; feature selection

1 Introduction

In boreal forests, habitat type mappings are widely used in forestry purposes but they are also valuable in conservation. In forestry, habitat type maps and other thematic maps are used e.g. for strategic analysis in forest management planning (Tomppo et al., 2008). In conservation perspective, habitat type maps can be used e.g. in mapping biodiversity patterns (e.g. Kerr & Ostrovsky, 2003; Turner et al., 2003). Habitat type mapping is often based on land use/land cover remote sensing data classification. Land cover and land use refer to biophysical surface characteristics of the Earth and land utilization respectively (e.g. Kerr & Ostrovsky, 2003; McDermid et al., 2005). Habitats, though, do not equate land cover and thus a specific approach is needed for habitat classifications (Lucas et al., 2011; McDermid et al., 2005).

Habitats are usually defined as the resources present in an area that are needed by organisms. On the other hand, habitat type is defined as a mappable land unit in which vegetation and environmental factors are fairly homogenous. However, the terms habitat and habitat type are also used interchangeably (Corsi et al., 2000). In some of the previous mapping approaches, habitat types have been mapped using only single-date satellite imagery. Yet, it has been acknowledged that mapping of detailed habitat types using only satellite imagery is challenging, since the spectral differences between different habitat types are often minor (Díaz Varela et al., 2008). To tackle this problem, multi-temporal imagery and ancillary data, such as soil map, existing land-use dataset, and digital terrain model (DTM), have been included in some of the approaches (Bock et al., 2005; Lucas et al., 2011).

For more than a decade, object-based image analysis (OBIA) has been used in constructing habitat type or other thematic maps from remotely sensed data. It has been acknowledged that OBIA gives more robust information and higher classification accuracies than pixel-based analyses (e.g. Bock et al., 2005; Díaz Varela et al. 2008; Whiteside et al. 2011; Yan et al. 2006).. OBIA combines pixels into meaningful objects which ideally mimic human perception of the analyzed image and are better representations of the landscape features. One major benefit of OBIA is that several different factors can be included into the OBIA workflow more easily and efficiently than into pixel-based analyses. These factors include several different types of data, contextual and textural information and multi-scale analysis (Benz et al., 2004; Blaschke, 2010; Bock et al., 2005). Finally, OBIA has become increasingly popular, because very high spatial resolution remote sensing data and software tools for doing OBIA have become more common (Blaschke, 2010).

So far, the main data sources in OBIA have been aerial or satellite images (Blaschke, 2010). From the spectral images, several different layers and several derived features have been used in the OBIA analyses. For instance, the usage of textural features such as the Gray-Level Co-occurrence Matrix (GLCM, Haralick, 1979; Haralick et al., 1973) is almost a standard in the OBIA analyses (e.g. Han et al., 2012; Johansen et al., 2007; Kim et al., 2009, 2011; Murray et al., 2010; Sasaki et al., 2012; Yu et al., 2006). Additionally, the promise of the wavelet features in the texture analysis has been noted also when combined with the GLCM (Arivazhagan & Ganesan, 2003; Ouma et al., 2008; Ruiz et al., 2004; Su et al., 2012; Wang et al., 2012). Wavelets have also been used in the data pattern or structure analysis (Falkowski et al., 2008; James et al., 2011; Strand et al., 2006). In this manner, Morgan et al. (2010) note that the GLCM is mainly used for a fine-scale textural analysis, whereas wavelets can extract coarse-scale patterns from the spectral images. The inclusion of different textural features in classification has produced higher classification accuracies (Han et al., 2012; Kim et al., 2009, 2011; Murray et al., 2010; Ruiz et al., 2004).

In addition to the multispectral images, several types of data have been used in the OBIA analyses. Especially, the usage of airborne laser scanning (ALS) data has become more popular (Blaschke, 2010). Yet, studies that combine ALS and spectral images in OBIA are still rather few (e.g. Arroyo et al., 2010; Breidenbach et al., 2010; Geerling et al., 2007, 2009; Ke et al., 2010; Sasaki et al., 2012; Wang et al., 2012). From ALS, the vertical and horizontal structure of vegetation or buildings, and a high-resolution DTM can be accurately quantified. Hence, ALS complements spectral images by revealing details that cannot be seen visually from above (Lefsky et al., 2002; Vierling et al., 2008). Mostly, ALS has been used for vegetation structure quantification (Antonorakis et al., 2008; Bar Massada et al., 2012; Breidenbach et al., 2010; Ke et al., 2010; Sasaki et al., 2012) but also an ALS based DTM has been in use (Bar Massada et al., 2012; Ke et al., 2010).

From the DTM, several different topographic features can be calculated, and features such as slope, aspect and curvature, are widely used in OBIA (Ke et al., 2010; Morgan & Gergel, 2010; Thompson & Gergel, 2008; Thompson et al., 2008; Yu et al., 2006). Moreover, from the DTM, different hydrological features can be calculated. One of the most used hydrological features has been topographical wetness index (TWI), originally proposed by Beven & Kirkby (1979), which has also been used in the OBIA studies (Ke et al., 2010; Morgan & Gergel, 2010; Thompson & Gergel, 2008; Thompson et al., 2008). Although many different features have been included in the OBIA studies, thorough tests of the importance of different features in classifying different habitat types are few.

1.1 Aims of the study

The main objectives in this study were: 1) Develop a working classification workflow applicable to boreal forest habitat type mapping. 2) Study, which features and layers are important in mapping different habitat types. 3) Examine the internal variation of habitat types and types' similarities with each other. Finally, we used the Finnish multisource National Forest Inventory (MS-NFI, Tomppo & Halme, 2004; Tomppo et al., 2008; Tomppo et al., 2012) as a benchmark against which we compared the results of our method.

2 Materials and methods

2.1 Used data

Our primary datasets were a multispectral 2-meter resolution WorldView-2 (WV-2) satellite image and ALS data. The WV-2 image was taken by Digital Globe Inc. in July 14th 2010 and was a subarea of one scene. The spectral range of WV-2 image was 400–1040 nm and it consisted of eight bands: coastal blue (center wavelength 425 nm), blue (480 nm), green (545 nm), yellow (605 nm), red (660 nm), red-edge (725 nm), near infra-red 1 (NIR1, 835 nm), and NIR2 (950 nm). The ALS data was provided by the National Land Survey of Finland, had at least 0.5 point per 1 m^2, and was collected in May 2010. The data was delivered as point clouds, automatically classified to ground hits, low vegetation hits, low error hits, and unclassified hits. The ALS point clouds were first triangulated and after that rasterized to construct three primary layers: a DTM, a digital surface model, and an intensity layer using LAStools (rapidlasso, Gilching, Germany). Additionally, we used 20 cm resolution aerial images (orthophotos) obtained from the city of Jyväskylä taken in 2007, a 1:10 000 resolution topographic database from the National Land Survey of Finland from the year 2010, a 1:20

000 resolution digital soil map from the Geological Survey of Finland, a 1:50 000 SLICES land use database from the NLS Finland, and a 20 m resolution MS-NFI from the Finnish Forest Research Institute from the year 2009 (Tomppo et al., 2012). Datasets and preprocessing are explained in more detail in Räsänen et al. (2013).

2.2 Study area and field work

We studied a 15 km^2 rural area southwest of the city of Jyväskylä divided into three sub-areas (Fig. 1). A part of the study area was classified into 26 different habitat or land-use types (Table 1) which were mapped with the help of field work and aerial imagery during June-July 2011. Three meadows that were mapped during the summer of 2010 were included into the analysis as well as a sand pit that was digitized using visual interpretation of WV-2 imagery. These data were included into the analysis since they were extremely close to the training dataset and because there were few meadows inside the training dataset. The field work covering in total 7 km^2 consisted of 632 patches in three contiguous sub-areas inside our study area (Fig. 1). The field work area was used for the training of the classifiers and for the classification accuracy assessments. The study area and the field work are explained in detail in Räsänen et al. (2013).

2.3 Habitat type classification system

Our habitat type classification system included natural, semi-natural and man-made habitat types. The classification system was based on the work by Rossi and Kuitunen (1996) and was modified to make it useful with remotely sensed data. Rossi and Kuitunen (1996) used habitat types as surrogates for potential species existence, and their classification was as detailed as it could be based on the species' habitat type preferences given in the species identification literature. Habitat types followed the Finnish classification systems for forests (Cajander, 1949) and mires (Eurola et al., 1995). The classification system was on one hand rather detailed, e.g. we mapped three different forest types in different successional stages. On the other hand, it included some easily mappable habitat types as well, especially water bodies and fields.

2.4 Classification approach

A detailed flow chart of the used classification approach is presented in Fig. 2. The approach was divided into two steps: segmentation and a supervised classification using random forest classifier. Only parts of the data were used in segmentation. After segmentation, a feature vector was constructed for each segment (Table 2). The supervised classification was then performed and different classification alternatives, which included different sets of features, were compared. After classification, the usage of ancillary data in post-classification adjustment was evaluated. Finally, the accuracies of classifications were assessed, the importance of different features was analyzed, and habitat type similarities were studied using feature vector analyses.

2.5 Segmentation data and method

We segmented WV-2 bands blue, green, red, and NIR1 (bands 2, 3, 5, and 7) together with two ALS layers: a SAGA wetness index (SWI) and a canopy height model (CHM). The SWI is a modification of a standard TWI. In the SWI, the high TWI values are predicted to larger

areas than in the TWI (Böhner & Selige 2006). The CHM was calculated by subtracting a DTM from a digital surface model. A more detailed clarification of the SWI and the CHM is given in Räsänen et al. (2013).

Data was segmented using Fractal Net Evolution Approach (FNEA) segmentation (Baatz & Schäpe, 2000, Benz et al., 2004) using TerraLib 4.2.0 C++ GIS-library (Câmara et al., 2008). The scale parameter was set to 10, whereas the weights for color and smoothness were set to 0.5 and 0.5 respectively. This segmentation layer-method-parameter-combination was visually most appealing and gave the highest classification accuracy in our segmentation evaluation study, which was performed in the same study area using the same datasets (Räsänen et al., 2013).

2.6 Used layers and features in the classification

In the classification, we used several layers derived from ALS and WV-2 data. From the layers, in total 328 features (331 if ancillary data is included) were calculated (Table 2). From remotely sensed data layers, mean and standard deviation values for each segment were calculated. To ease data interoperability and analysis, all remote sensing data was in 2 m (or 10 m) resolution which was suitable in our analysis since patches of interest were larger than 40 m^2 or ten pixels (c.f. Lechner et al., 2009). The aerial images were disregarded in the analysis because of the three year time difference. There exist differences in the landscape because of forestry actions, especially regeneration cutting.

In addition to WV-2 spectral bands, a normalized difference vegetation index (NDVI) was calculated using bands 5 (red) and 7 (NIR1). From the ALS data we calculated an intensity layer, a CHM layer and several topographic or hydrological layers.

Texture and wavelet features were calculated for each segment from the 2 m resolution NDVI, WV-2 bands and the CHM. Texture features were calculated according to a GLCM (Haralick et al., 1973) using the mean intensity of the pixel's 4 neighboring pixels (0˚, 45˚, 90˚, and 135˚) with the package EBimage 2.2.0 (Pau et al., 2010) in R 2.13.0 (R Development Core Team, 2012). The GLCM features estimate different features of pixel brightness value combinations from the neighboring pixels. Overall, 13 GLCM features given by Haralick et al. (1973) were calculated: angular second moment (asm), contrast (con), correlation (cor), variance (var), inverse difference moment (idm), sum average (sav), sum variance (sva), sum entropy (sen), entropy (ent), difference variance (dva), difference entropy (den), and two measures of correlation (f12 and f13). For the GLCM calculations, the pixel brightness values were quantized to 16 gray levels using equal intervals.

A maximum overlap discrete wavelet analysis was performed in R 2.13.0 using the package waveslim 1.7.1 (Whitcher, 2012). The wavelets calculate texture in larger neighborhoods than the GLCM and quantify local frequency, i.e. if same kind of orderliness is seen in the changes of image brightness values as in the wavelet function. The base function was Daubechies orthonormal compactly supported wavelet of length $L = 8$ (Daubechies, 1992). The wavelets were calculated in horizontal, vertical and diagonal directions in four decomposition lengths.

In the hydrological and topographic analysis, the DTM was filled to remove uncertainties, missing values and incorrect values from the data. We modeled a stream network using a D∞ flow direction method and a threshold value of 20 000 m^2 for shaping a stream with TauDEM

tools (Tarboton, 2012). From all locations in the area, a vertical distance to the nearest stream was calculated. The vertical distance to a stream was not used as a feature in the classification, but it was used in the clipping of the study area into two parts as described in Sections 2.7 and 2.11.

In addition to a percentage slope raster and the SWI, following raster layers were calculated from the filled DTM. In a distance to water (DTW) layer, a cost distance from each pixel to the nearest stream or water body is calculated with the slope raster as a cost surface (Murphy et al., 2007, 2009, 2011). A terrain ruggedness index (TRI) calculates the sum of change in altitude locally (Riley et al., 1999), a topographic position index (TPI) measures the relative altitudinal position of a pixel, i.e. if it is below or above the average local altitude (Guisan et al., 1999) and a multiresolution index for valley bottom flatness (MRVBF) calculates a probability that a pixel is in the bottom of a valley (Gallant & Dowling, 2003). The TRI was calculated using 3 different window sizes, the TPI was calculated with 4 different radiuses, and in the MRVBF, the initial threshold for slope was given two values: the default value 16, and 75 as suggested by Gallant & Dowling (2003). The other parameters were kept as default. TRI, TPI and MRVBF were calculated in SAGA-GIS and other topographic layers in ArcGIS 9.3.1 (Esri, Redlands, CA, USA).

All segments which were inside field work blocks were used in the training phase if they had at least 60 percent coverage of one habitat type. In the training dataset, each segment was given a feature vector which included the features given in Table 2.

2.7 Random forest classification

We classified the data with random forest (Breiman 2001) in R with the package randomForest (Liaw & Wiener, 2002). Random forest is an ensemble classifier, in which a majority vote over several bootstrapped classification trees is made. When each tree is built, approximately 2/3 of the data is used for training the classifier and the rest is called out of bag (OOB) data. Because of the OOB, independent test data or cross-validation is not required when random forest is used (Breiman, 2001; Breiman & Cutler, 2007).

Random forest has given good classification accuracies in remote sensing studies (Duro et al., 2012; Lawrence et al., 2006; Rodriguez-Galiano et al., 2012; Smith, 2010). However, it has been noted that when data is imbalanced, random forest can underestimate rare observations (Breidenbach et al., 2010) and that there should be abundant training data for all classes (Smith, 2010). To get the data more balanced, we tested splitting the data into two parts: areas that were inside or outside a 1 m vertical distance to the nearest stream. This was performed especially to map mire habitat types more effectively. While mires are much rarer than forests on mineral soils in our area, the percentage of mire cover is much higher close to the streams.

In random forest classification, two parameters can be modified: the number of trees built and the number of features tested at each split (mtry). Different values for the both parameters were tested.

2.8 Feature importance and selection

Random forest gives two different feature importance measures. In the first measure, permutation importance, the OOB error rate is compared to an OOB error rate when the focal

feature is not used. The second measure, gini importance, calculates the overall decrease in gini impurity when the focal feature is used as a split. The feature importance measures were used in analyzing which layers and features are useful in our purposes and what features were important in mapping each habitat type.

In feature selection, we used the R package Boruta (Kursa & Rudnicki, 2010) which tests if the features are significant or not. Boruta is a wrapper algorithm in which several runs of random forest classification are performed. Before each run, a shadow feature is created for each feature. The values of the shadow feature are derived by shuffling the values of the original feature across data items, which are segments in our case. After each random forest run, each feature's importance is tested against the shadow feature with the highest importance. All the features which have significantly lower importance than the shadow feature with the highest importance are classified unimportant and removed from the following runs. The attributes that have significantly higher importance are classified as important. These features are included in the following runs but their importance is not tested anymore. Boruta ends when all features are classified either as unimportant or important or when a specified limit of random forest runs is reached. If some variables are left tentative, a tentative rough fix test can be taken. In this test, all those features that have higher median ZScore than the median ZScore of maximal shadow attribute are set confirmed. The rest of the features are deemed rejected (Kursa & Rudnicki, 2010). In Boruta, we used 0.999 confidence level, "z-scores of mean decrease accuracy measure" i.e. permutation importance and we did not set a limit for random forest runs.

2.9 Classification accuracy assessment

When the final classification map was derived, its classification accuracy was calculated using cross tabulation matrices. We decided not to use Kappa indices, since the use of them has been sometimes criticized. The main critique is that Kappa compares observed accuracy with an expected accuracy due to randomness, although randomness is an unimportant baseline in remote sensing tasks (Foody, 2008; Pontius & Millones, 2011). Instead, in addition to the user's accuracy (the error of commission) and the producer's accuracy (the error of omission), we quantified allocation and quantity disagreement as suggested by Pontius and Millones (2011). These two measures quantified if the amount of predicted habitat types differs from the reference or if the habitat types were allocated to different locations than in the reference. The accuracy assessment was performed cross-tabulating pixel based classification data in raster format and field work data in vector format. Segment based classification accuracy was not calculated, since many segments were divided between different training data classes.

2.10 Feature vector analysis

A feature vector analysis was used to complement the classification accuracy analysis. Mean values of all the features in the training dataset as well as in the final classified datasets were calculated. Euclidean distances between class centroids were then produced. Sammon's mapping (Sammon, 1969) using all training data segments and Euclidean distance was used to illustrate which classes are close to each other. In the Sammon's mapping, multidimensional data is reduced to fewer dimensions. Instead of using linear combinations as in a principal component analysis, Sammon's mapping uses a non-linear approach. Finally, a random forest proximity measure, which calculates the mean distance between two cases in all trees, was also calculated and illustrated using a proximity image.

2.11 Classification alternatives

We tested 16 different classification alternatives to see how different alterations influence the final classification accuracy (Table 3). In eight alternatives, some of the features were omitted from classification and Boruta feature selection was used for selecting either only important or the 100 highest scoring features. Additionally, in four alternatives, data was classified separately in two zones according to a 1 m vertical distance to the nearest stream. Furthermore, ancillary data was included for the classifications of full data and split data that had the highest classification accuracy. The forest segments whose majority soil class was silt were reclassified as herb-rich forests. Segments whose majority NLS topographic database class was fields, meadows, rocky areas, waters, or sand pits were reclassified to the respective class. The xeric forest segments whose majority topographic database class was mire were reclassified to pine mire. In a similar manner, mesic and herb-rich forests together with yards were reclassified as spruce mires, while open areas (water, fields, meadows, roads) were reclassified as open mires. All the mires that had over 50 % of their area inside a 25 m buffer of small streams (width < 5 m) in the topographic database were classified as drained mires. Segments whose majority SLICES class was yard were classified as yards. Roads were updated from the SLICES database.

Finally, two habitat type classifications were made using only already available land use/land cover data. Forest and mire types as well as forest succession stage for forests were derived from the MS-NFI data while ancillary information was obtained from the NLS topographic database (mires, bare rock, lakes, meadows, fields, sand pits, partly yards and roads) and SLICES (yards, roads). In the first alternative, MS-NFI data was used as such. In the second alternative, modifying the classification according to Geneletti and Gorte (2003), such that for each segment a majority MS-NFI class was given.

3 Results

The study area was segmented into 12026 segments. When the data was split into two parts, the number of segments was 8969 and 3680 for the areas above and below the 1 m vertical distance to a stream respectively. Of the segments, 4883, 3580, and 1529 were used as the training data for the full area, the above 1 m proportion and the below 1 m proportion respectively. The numbers of segments that did not have a 60 % majority of any habitat type in the reference dataset were 606, 375, and 234 respectively. The rest of the segments were included in the classification but over 60 % of those segments were outside the area covered by the reference dataset. There were some differences in the number of segments belonging to different classes in the three different datasets (Table 4). Especially, the proportion of segments belonging to different mire classes, water bodies and streamside habitat was much higher in below 1 m data than in the two other datasets.

3.1 Random forest parameterization

In random forest classification, default parameter values were used but several options were tested. The number of trees was set to 500 and mtry to the square root of all features. The OOB error rate decreased up to around 100 trees and then it stabilized. Similar results have been obtained in other studies (Lawrence et al., 2006; Rodriguez-Galiano et al., 2012). A smaller parameter value for mtry increased the OOB error rate a bit and a larger value

decreased the error rate a bit. Nevertheless, the highest decrease in the OOB error rate compared to the error rate with the default mtry parameter value was not more than 1.3 percentage points. Different mtry parameter values were not tested more thoroughly, since our focus was more on testing how omission or inclusion of different types of features affects classification accuracy. According to Rodriguez-Galiano et al. (2012), random forest is not sensitive to the value of mtry when enough trees are built. They even suggested giving a small value for mtry since it decreases the correlation between trees, although then the strength of each tree decreases (see also Breiman 2001).

3.2 Classification accuracies

In the first classification alternative, using all data and all features, a classification accuracy of 78 % was obtained (Fig. 3, Tables 5 and 6). The accuracy increased to 79 % with the 100 highest scoring features by Boruta and ancillary data. This classification accuracy was further improved to 86 % when the misclassifications between different forest successional stages as well as drained and non-drained mires were not considered to be errors. Nonetheless, the misclassifications between different forest and mire habitat types were regarded as errors. The lowest classification accuracy of 73 % was obtained when only WV-2 data was used. Omitting the extra topographical features or the GLCM and wavelet features reduced the classification accuracy but omitting only either the GLCM or wavelet features did not. Splitting the area into two parts did not improve the overall classification accuracy but ancillary data increased it by 0.5 percentage points. Furthermore, all our classifications had higher classification accuracy compared to the accuracy of 52 % (55 % with segmentation) obtained using MS-NFI, the topographical database and SLICES. The classification accuracy of the MS-NFI based classification, however, improved to 78 % (80 % with segmentation) when the misclassifications between different forest successional stages as well as drained and non-drained mires were not considered errors.

The differences in the classification accuracies of single classes had higher variations. For instance, clipping the area into two parts increased producer's accuracy in mapping mires and streamside habitats. Ancillary data raised the producer's accuracy especially in mapping yards, roads, spruce mires, and bare rock, and in lesser extent in mapping herb-rich forests, meadows, and fields (Fig. 3d). On the other hand, the user's accuracy was lower for spruce and pine mires when the area was split and for meadows, bare rock, herb-rich forests, drained mires and roads when ancillary data was used.

Some of the classes (springs, xeric mature forests, herb-rich natural forests) could not be classified at all probably due to the low amount of training data. Other low classification accuracies among classes were produced for herb-rich forests, spruce mires, streamside habitats and meadows. Common or easily separated habitat types, such as mesic forests, water, fields, and sand, got high classification accuracies in all classifications. The MS-NFI based benchmark classification had as high or nearly as high classification accuracy as other approaches for many of the classes. Forest habitat types were, however, more poorly classified both in regards of the user's and the producer's accuracy.

In the classification with the highest accuracy, the allocation disagreement was slightly larger than the quantity disagreement (Table 7). Most of the total disagreement in the classification accuracy was caused by the classification of mesic forest classes. Although these classes were fairly well classified compared to some other classes, they were among the most common classes inside the study area. The largest total quantity disagreement was caused by mesic2

which was over-predicted i.e. the error of commission. Also other mesic forests were over-predicted. Some of the classes, such as streamside habitats and spruce mires, were significantly under-predicted in regards to the total area. This under-prediction was lower though, when the area was split.

When omission and commission were evaluated relative to habitat type abundance (Table 7), xeric and herb-rich forests, non-drained pine and spruce mires, bare rock, meadows, streamside habitats and some mesic forests were more under-predicted. Conversely, some mesic forests, open and drained mires, fields, roads, and yards were more over-predicted. The proportions of omission and commission were slightly different for the other classifications.

In the MS-NFI classification, allocation disagreement (38 %, with segmentation 34 %) was higher than quantity disagreement (10 %, with segmentation 12 %). In more detail, most of the forest was, also in MS-NFI, predicted as mesic forests and mesic3 was over-predicted (quantity disagreement 4 % of total study area). Different mesic forest classes also had highest allocation disagreements which were 26 % for mesic2, 17 % for mesic3, and 13 % for mesic4.

3.3 Feature importance

In the Boruta feature selection runs using all data and above 1 m vertical distance data, all 328 features were confirmed important after 44 and 73 random forest runs respectively. For the below 1 m areas, 130 random forest runs were performed, and 275 features were confirmed important, 21 unimportant and 32 were left tentative. These features had highly fluctuating ZScores probably due to tricky data. For tentative features, a tentative rough fix test was performed using 100 last rf runs. After test, 300 and 28 features were confirmed important and unimportant respectively. The rejected features consisted of GLCM and wavelet features of WV-2 bands and NDVI together with standard deviation values of 10 m resolution WV-2 bands.

In the feature importance analysis, it was found that many different types of features were regarded as important using the different measures (Table 8). There were rather large differences between different habitat types as well as between the permutation importance, the Gini importance, and the Boruta mean Z-score. The GLCM, wavelet as well as topographical features were among the highest ranking features. The mean features got almost without exception higher importance values than the standard deviation features. In a similar manner, the 2 m resolution features were generally more important than the 10 m resolution features. Of the WV-2 bands, the band number 1 (coastal blue) got the highest scores of all three overall measures and its mean value was among the most important features when all measures were taken into account. Additionally, CHM, DTW and MRVBF75 got high ranks from the different measures. The wavelet and the GLCM features got both high and low importance values. When only the WV-2 bands were taken into account, sum average (sav) and sum variance (sva) were the most important GLCM features. However, among the features derived from the NDVI and the CHM, also other GLCM features such as sum entropy (sen), variance (var), entropy (ent), inverse difference moment (idm), contrast (con) got high importance values. Furthermore, the results were quite contradictory. Although many wavelet features got high permutation importance values, omitting them from analysis increased classification accuracy by 0.1 percentage points. One reason could be that wavelet and GLCM features gave overlapping information. In other words, it is possible that wavelet features did not produce any extra information compared to

GLCM features, and vice versa. In addition, the randomness in random forest classification might also result in small changes in classification accuracies.

When correlations between the different features were examined (Fig. 4), it was found that the WV-2 visible color bands (1-5) are strongly correlated. Also, infrared bands (6-8) are correlated but not as heavily as color bands. Furthermore, the GLCM and wavelet features of the correlating bands were correlated. All the wavelet features were giving more or less same information which was quite close to information given by the GLCM features. The GLCM features were not as heavily correlated to each other as wavelet features, and some features (idm, asm) correlated negatively, and some features (sav, sva) did not correlate with the other GLCM features. ALS and WV-2 features were not highly correlated overall. The topographic features with different parameter settings were again highly correlated to each other but different topographic features did not correlate strongly with each other. Finally, the mean and standard deviation features correlated usually negatively or not at all.

3.4 Feature vector analysis

In the Sammon's map (Fig. 5), different mesic forests covered almost the entire plot. Notable exceptions were water areas and fields which both had rather clearly separable areas. Other classes were more or less inside the area occupied by the mesic forests. Inside the mesic forests, the variation seemed to be quite large and some successional stages, especially 1 and 4, were separable from each other.

In the proximity image (Fig. 6, Table 4), some classes (mesic4, water, sand) stood out exceptionally well and some classes (mesic0, mesic1, mesic2, mesic3, field, road) can be distinguished. However, it can also be seen that segments in the classes of mires, herb-rich and xeric forests have rather high proximities to segments of mesic forests. Although the proximities inside different mesic forest classes are larger than the proximities between classes, also some of the latter proximities are rather large.

4 Discussion

4.1 Mapping habitat types in boreal forests

In the Finnish forest classification system (Cajander, 1949); habitat types are classified based on ground vegetation, nutrient status, soil permeability and soil grain size, not on tree species. This problem has been accounted for in making more remote sensing specific habitat type classification systems (e.g. Tuominen et al., 2001) but these systems are not widely used.

Tree species are not, though, the only problem in habitat type mapping. In our study area, most of the forest is mesic spruce dominated forest. In our initial evaluations changing the classification basis from nutrient status to tree species did not improve classification accuracy. The feature vector analysis suggested that other habitat types were confused with different mesic forests. As well, the inner heterogeneity of mesic forests is rather large, as it was illustrated by the feature vector analysis. One possible explanation is that both mesic forests and spruce dominated forests are heterogeneous in the terms of species composition, density of stand, proximity to bedrock and soil types, for instance. Therefore, although other habitat types might be distinguishable in the field; their characteristics are often inside the variation of dominating classes in remotely sensed data.

Better training data, in which only clear examples of different classes are selected, could improve the classification accuracy. This kind of training data might not, on the other hand, represent the reality out in the forest. Furthermore, the nature itself is not easy to interpret. For instance, Cherrill and McClean (1995, 1999) achieved very weak accuracy rates when they compared habitat type maps drawn by different human mappers. Therefore, it is not always important to know the habitat type as such but to have a broader idea of the site. In this manner, Lucas et al. (2011) offered three different levels of information for their habitat maps. Additionally, the inner variability of the classes as well as patches can be accounted using feature vector analysis or fuzzy analysis (e.g. Benz et al. 2004).

There are numerous reasons why the MS-NFI performed more poorly than our classifications in our purposes. First, our approach was local while the MS-NFI is nation-wide. We used training and test data, explained in Section 2.2, developed for this study locally while the MS-NFI was only reclassified for our purposes. In the MS-NFI, one field sample plot represents forest area whose size is more than 3 km^2 (Tomppo et al., 2012). Hence, there are mathematically less than five plots in our study area. The comparison is; thus, a bit unfair and heavily favors our classification approach. Second, there might be a slight temporal mismatch since the MS-NFI is a couple of years older than our field work and remotely sensed datasets. This should have effects, however, only on open regeneration forest areas and young forests. Nevertheless, the classification accuracy of the MS-NFI based classification was remarkably better when the misclassifications between forest successional stages as well as drained and non-drained mires were not regarded as errors. The difference in the classification accuracies between the classification with the highest accuracy and the MS-NFI-based classification with segmentation reduced from 24 to 6 percentage points. Third, the spatial resolution of the MS-NFI was coarser which might have effects on its classification accuracy. Also the fact that allocation disagreement was notably higher than quantity disagreement in the MS-NFI based classification might be partly because of its coarser spatial resolution.

4.2 Mapping different habitat types

As we discuss in our segmentation goodness evaluation (Räsänen et al., 2013); boreal forest habitat type classification using only OBIA does not give fully satisfactory results. Especially, the prediction of location of mires, riparian habitats and streams is difficult using only objects. It is complicated to include hydrological or topographical layers to segmentation so that riparian areas, for instance, could be better segmented.

Delineating mires inside mineral soil forest matrix is difficult even when other methods than OBIA are used. The 1 meter vertical distance to stream did not delineate mires but it gave a zone inside which many mire patches exist. Inside this zone, data were more balanced, i.e. mires were not as rare, which helped in the random forest prediction (Breidenbach et al., 2010; Smith, 2010). Data splitting also had its drawbacks in our case, however, since the overall classification accuracy did not improve. There exist also other methods for data balancing (Chen et al., 2004) which could be tested in remote sensing tasks.

Our results in mire or wetland mapping are in line with previous results. It has been pointed out that including topographical features and other ancillary information such as soil data gives better results than using only satellite imagery (Corcoran et al., 2011; Maxa & Bolstad, 2009; Ozesmi & Bauer, 2002, Tomppo & Halme, 2004; Tomppo et al., 2012; Wright & Gallant, 2007). It has also been discussed that it is hard to find a balance in wetland under-

and over-prediction. While wetlands are better mapped when their proportion in training data is larger, they are also more over-predicted (Wright & Gallant, 2007). We also got weak support that DTW is better than TWI in wetland prediction as suggested by Murphy et al. (2009). Yet, all topographical features seemed to be more or less important in wetland mapping. Therefore, they may complement each other as Murphy et al. (2011) pointed out.

It became evident in our study that some fine-scale habitat types such as springs are demanding to map. Although they may have very distinct characteristics that can be seen in the field, they do not necessarily stand out easily in remotely sensed data or even in soil or bedrock data. Since the patches of these habitat types are small and quite rare in larger habitat mosaic, they are easily missed into noise in the automated approaches. Moreover, they are hardly present in training data. Also some not as rare habitat types, such as streamside habitats were difficult to map. Some of them were, however, correctly mapped and streams themselves can be mapped using a hydrological analysis. A more thorough visual or manual analysis performed by a skilled interpreter could bring better results. An expert analysis implemented also in an OBIA workflow has given good results in the mapping of rare habitat types (Thompson & Gergel, 2008).

4.3 Feature importance and selection

As other authors (Arroyo et al., 2010; Geerling et al., 2007, 2009; Ke et al., 2010; Sasaki et al., 2011) have already pointed out, coupling ALS and spectral images gives better classification accuracies than using only one of them. In our analysis, using only ALS data gave slightly better classification accuracy than using only WV-2 data. In contradiction to previous remote sensing studies (Ouma et al., 2008; Ruiz et al. 2004), although the GLCM and wavelet features separately enhanced the classification accuracy in our study, they did not further improve it together. Additionally, omitting some of the topographical features (DTW, MRVBF, TPI, TRI) decreased the classification accuracy. Topographical features were needed especially in mapping mires.

The highest classification accuracy was achieved when the 100 highest ranked features were selected from the full feature set. Boruta was not, however, able to distinguish unimportant features from important features with the exception of the below 1 m vertical distance area. This may be due to the quite large variability inside habitat types and due to the quite large set of segments in training data compared to the number of features. The below 1 m area had clearly the smallest set of segments and thus also the highest feature per segment ratio (0.21). In a previous study when Boruta was used in remote sensing, the feature per segment ratio was far higher (1.65); and a large proportion of features (62 %) were confirmed unimportant or tentative (Duro et al., 2012). Hence, it may be questioned that does feature selection performed with Boruta help in situations when the feature per segment ratio is small. Moreover, the features were seemingly highly correlated and clearly not all of them are useful in classification. On the other hand, Boruta confirmed that almost all features gave meaningful information in our case.

Gini and permutation importance highlighted different features. It has even been argued that the random forest feature importance calculations are biased, Gini being the most biased. The bias favors correlating and continuous features and features with many categories. Based on this, alternative methods for measuring feature importance and selecting features have been suggested (Hapfelmeier & Ulm, 2013; Strobl et al., 2007, 2008). The feature importance bias does not have an effect on random forest predictive capacity. Instead, the predictive capacity

might be even better when many (and possibly correlating) features are included in random forest (Hapfelmeier & Ulm, 2013).

5 Conclusions

In our study, we developed a classification workflow to map different habitat types in a Finnish boreal forest landscape. The classification accuracy of our approach was considerably higher than the classification accuracy of the MS-NFI based classification. Our study area was dominated by different mesic forests. Their heterogeneity was large and other habitat types were partly inside this heterogeneity. This was confirmed using both classification accuracy assessments and feature vector analyses. The feature vector analysis showed its usefulness in our study, illustrated the closeness of different habitat types and complemented information given by the classification accuracy assessments. It became evident in our analysis that different types of data and features are needed in a boreal forest habitat type classification. Finally, a working boreal forest habitat type classification should include (1) both spectral images and other data, (2) meaningful habitat type classes that can be classified based on the used data, (3) a specific approach for mapping mires and rare habitat types, (4) good training data, (5) a classifier that works well with the used data, and (6) segmentation or some other technique for habitat patch boundary delineation.

Acknowledgements

This research was funded by Maj and Tor Nessling Foundation. We are grateful to Antti Rusanen who did the enormous field work during summer of 2011. As well, we included some of the field work by Paula Kinanen from summer of 2010. We also thank the city of Jyväskylä for the aerial orthophotos. MK received from Jenny and Antti Wihuri foundation a sabbatical scholarship for the year 2013 that was partly funded also by the EU IMPERIA-project (LIFE11 ENV/FI/905) and the University of Jyväskylä.

6 References

Ahti, T., Hämet-Ahti, L., Jalas, J., 1968. Vegetation zones and their sections in northwestern Europe. Annales Botanici Fennici 5, 169-211.

Antonorakis, A.S., Richards, K.S., Brasington, J., 2008. Object-based land cover classification using airborne LiDAR. Remote Sensing of Environment 112, 2988-2998.

Arivazhagan, S., Ganesan, L., 2003. Texture classification using wavelet transform. Pattern Recognition Letters 24, 1513-1521.

Arroyo, L.A., Johansen, K., Armston, J., Phinn, S., 2010. Integration of LiDAR and Quickbird imagery for mapping riparian biophysical parameters and land cover types in Australian tropical savannas. Forest Ecology and Management 259, 598-606.

Baatz, M., Schäpe, A., 2000. Multiresolution segmentation – an optimization approach for high quality multi-scale image segmentation. In: Strobl, J., Blaschke, T., Griesebner, G. (Eds.), Angewandte Geographische Informationsverarbeitung XII. Wichmann, Heidelberg, pp. 12-23.

Bar Massada, A., Kent, R., Blank, L., Perevolotsky, A., Hadar, L., Carmel, Y., 2012. Automated segmentation of vegetation units in a Mediterranean landscape. International Journal of Remote Sensing 33(2), 346-364.

Benz, U.C., Hofmann, P., Willhauck, G., Lingenfelder, I., Heynen, M., 2004. Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. ISPRS Journal of Photogrammetry and Remote Sensing 58, 239-258.

Beven, K.J., Kirkby, M.J., 1979. A physically based variable contributing area model of catchment hydrology. Hydrological Sciences Bulletin 24, 43-69.

Blaschke, T., 2010. Object based image analysis for remote sensing. ISPRS Journal of Photogrammetry and Remote Sensing 65, 2-16.

Bock, M., Xofis, P., Mitchley, J., Rossner, G., Wissen, M., 2005. Object-oriented methods for habitat mapping at multiple scales – Case studies from Northern Germany and Wye Downs, UK. Journal of Nature Conservation 13, 75-89.

Böhner, J., Selige, T., 2006. Spatial prediction of soil attributes using terrain analysis and climate regionalisation. In: Böhner, J., McCloy, K.R., Strobl, J. (Eds.), SAGA – Analysis and modelling applications. Göttinger Geographische Abhandlungen, Vol.115, pp. 13-28.

Breidenbach, J., Næsset, E., Lien, V., Gobakken, T., Solberg, S., 2010. Prediction of species specific forest inventory attributes using a nonparametric semi-individual tree crown approach based on fused airborne laser scanning and multispectral data. Remote Sensing of Environment 114, 911-924.

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

Breiman, L., Cutler, A., 2007. Random forests – classification description, http://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm. (Accessed 10 September, 2013)

Cajander, A.K., 1949. Forest types and their significance. Acta Forestalia Fennica 56, 1-71.

Câmara, G., Vinhas, L., Ferreira, K.R., De Queiroz, G.R., De Souza, R.C.M., Monteiro, A.M.V., De Carvalho, M.T., Casanova, M.A., De Freitas, U.M., 2008. TerraLib: An open source GIS library for large-scale environmental and socio-economic applications. In: Hall, G.B., Leahy, M.G. (Eds.), Open source approaches in spatial data handling. Springer, Berlin, pp. 247-270.

Chen, C., Liaw, A., Breiman. L., 2004. Using random forest to learn imbalanced data. Technical Report 666, Statistics Department, University of California at Berkeley, http://statistics.berkeley.edu/sites/default/files/tech-reports/666.pdf. (Accessed 11 September, 2013)

Cherrill, A., McClean, C., 1995. An investigation of uncertainty in field habitat mapping and the implications for detecting land cover change. Landscape Ecology 10, 5-21.

Cherrill, A., McClean, C., 1999. The reliability of 'Phase 1' habitat mapping in the UK: the extent and types of observer bias. Landscape and Urban Planning 45, 131-143.

Corcoran, J., Knight, J., Brisco, B., Kaya, S., Cull, A., Murnaghan, K., 2011. The integration of optical, topographic, and radar data for wetland mapping in northern Minnesota. Canadian Journal of Remote Sensing 37, 564-582.

Corsi, F., de Leeuw, J., Skidmore, A.K., 2000. Modeling species distribution with GIS. In Boitani, L., Fuller, T.K. (Eds.), Research techniques in animal ecology: Controversies and consequences. Columbia University Press, New York, pp. 389-434.

Daubechies, I., 1992. Ten lectures on wavelets (CBMS-NSF Regional Conference Series in Applied Mathematics; 61). Society for Industrial and Applied Mathematics, Philadelphia.

Díaz Varela, R.A., Rego, P.R., Iglesias, S.C., Sobrino, C.M., 2008. Automatic habitat classification methods based on satellite images: A practical assessment in the NW Iberia coastal mountains. Environmental Monitoring and Assessment 144, 229-250.

Duro, D.C., Franklin, S.E., Dubé, M.G., 2012. Multi-scale object-based image analysis and feature selection of multi-sensor earth observation imagery using random forests. International Journal of Remote Sensing 33 (14), 4502-4526.

Eurola, S., Huttunen, A., Kukko-Oja, K., 1995. Suokasvillisuusopas, second ed. Oulanka Reports 14. Oulanka Biological Station, University of Oulu, Oulu.

Falkowski,M.J., Smith, A.M.S., Gessler, P.E., Hudak, A.T., Vierling, L.A., Evans, J. S., 2008. The influence of confifer forest canopy cover on the accuracy of two individual tree measurement algorithms using lidar data. Canadian Journal of Remote Sensing 34 (S2), S338-S350.

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

Gallant, J.C., Dowling, T.I., 2003. A multiresolution index of valley bottom flatness for mapping depositional areas. Water Resources Research 39, 1347-1360

Geerling, G.W., Labrador-Garcia, M., Clevers, J.G.P.W., Ragas, A.M.J., Smits, A.J.M., 2007. Classification of floodplain vegetation by data fusion of spectral (CASI) and LiDAR data. International Journal of Remote Sensing 28, 4263-4284.

Geerling, G.W., Vreeken-Buijs, M.J., Jesse, P., Ragas, A.M.J., Smits, A.J.M., 2009. Mapping river floodplain ecotopes by segmentation of spectral (CASI) and structural (LiDAR) remote sensing data. River research and applications 25, 795-813.

Geneletti, D., Gorte, B.G.H., 2003. A method for object-oriented land cover classification combining Landsat TM data and aerial photographs. International Journal of Remote Sensing 24, 1273-1286.

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

Han, N., Wang, K., Yu, L., Zhang, X., 2012. Integration of texture and landscape features into object-based classification for delineating Torreya using IKONOS imagery. International Journal of Remote Sensing 33, 2003-2033.

Hapfelmeier, A., Ulm, K., 2013. A new variable selection approach using Random Forests. Computational Statistics & Data Analysis 60, 50-59.

Haralick, R.M., Shanmugam, K., Dinstein, I., 1973. Textural features for image classification. IEEE Transactions of Systems, Man and Cybernetics 3, 610-621.

Haralick, R.M., 1979. Statistical and structural approaches to texture. Proceedings of the IEEE 67, 786-804.

James, P.M.A., Sturtevant, B.R., Townsend, P., Wolter, P., Fortin, M., 2011. Two-dimensional wavelet analysis of spruce budworm host basal area in the Border Lakes landscape. Ecological Applications 21, 2197-2209.

Johansen, K., Coops, N.C., Gergel, S.E., Stange, Y., 2007. Application of high spatial resolution satellite imagery for riparian and forest ecosystem classification. Remote Sensing of Environment 110, 29-44.

Ke, Y., Quackenbush, L.J., Im, J., 2010. Synergistic use of QuickBird multispectral imagery and LIDAR data for object-based forest species classification. Remote Sensing of Environment 114, 1141-1154.

Kerr, J.T., Ostrovsky, M., 2003. From space to species: ecological applications for remote sensing. TRENDS in Ecology and Evolution 18 (6), 299-305.

Kim, M., Madden, M., Warner, T.A., 2009. Forest type mapping using object-specific texture measures from multispectral Ikonos imagery: Segmentation quality and image classification issues. Photogrammetric Engineering & Remote Sensing 75, 819-829.

Kim, M., Warner, T.A., Madden, M., & Atkinson, D.S., 2011. Multi-scale GEOBIA with very high spatial resolution digital aerial imagery: scale, texture and image objects. International Journal of Remote Sensing 32, 2825-2850.

Kursa, M.B., Rudnicki, W.R., 2010. Feature selection with the Boruta Package. Journal of Statistical Software 36, 1-13.

Lawrence, R.L., Wood, S.D., Sheley, R.L., 2006. Mapping invasive plants using hyperspectral imagery and Breiman Cutler classifications (RandomForest). Remote Sensing of Environment 100, 356-362.

Lechner, A.M., Stein, A., Jones, S.D., Ferwerda, J. G., 2009. Remote sensing of small and linear features: Quantifying the effects of patch size and length, grid position and detectability on land cover mapping. Remote Sensing of Environment 113, 2194-2204.

Lefsky, M.A., Cohen, W.B., Parker, G.G., Harding, D.J., 2002. Lidar remote sensing for ecosystem studies. BioScience 52, 19-30.

Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. R News 2, 18-22.

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.

Maxa, M., Bolstad, P., 2009. Mapping northern wetlands with high resolution satellite images and LiDAR. Wetlands 29, 248-260.

McDermid, G.J., Franklin, S.E., LeDrew, E.F., 2005. Remote sensing for large-area habitat mapping. Progress in Physical Geography 29, 449-474.

Morgan, J.L., Gergel, S.E., 2010. Quantifying historic landscape heterogeneity from aerial photographs using object-based analysis. Landscape Ecology 25, 985-998.

Morgan, J.L., Gergel, S.E., Coops, N.C., 2010. Aerial photography: A rapidly evolving tool for ecological management. BioScience 60, 47-59.

Murphy, P.N.C., Ogilvie, J., Connor, K., Arp, P., 2007. Mapping wetlands: A comparison of different approaches for New Brunswick, Canada. Wetlands 27, 846-854.

Murphy, P.N.C., Ogilvie, J., Arp, P., 2009. Topographic modeling of soil moisture conditions: a comparison and verification of two models. European Journal of Soil Science 60, 94-109.

Murphy, P.N.C., Ogilvie, J., Meng, F., White, B., Bhatti, J.S., Arp, P.A., 2011. Modelling and mapping topographic variations in forest soils at high resolution: A case study. Ecological Modelling 222, 2314-2332.

Murray, H., Lucieer, A., Williams, R., 2010. Texture-based classification of sub-Antarctic vegetation communities on Heard Island. International Journal of Applied Earth Observation and Geoinformation 12, 138-149.

Ouma, Y.O., Tetuko, J., Takeishi, R., 2008. Analysis of co-occurrence and discrete wavelet transform textures for differentiation of forest and non-forest vegetation in very-high-resolution optical-sensor imagery. International Journal of Remote Sensing 29, 3417-3456.

Ozesmi, S.L., Bauer, M.E., 2002. Satellite remote sensing of wetlands. Wetlands Ecology and Management 10, 381-402.

Pau, G., Fuchs, F., Sklyar, O., Boutros, M., Huber, W., 2010. EBImage – an R package for image processing with applications to cellular phenotypes. Bioinformatics 26, 979-981.

Pontius, R.G., Millones, M., 2011. Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. International Journal of Remote Sensing 32, 4407-4429.

R Development Core Team, 2012. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0, http://www.R-project.org/. (Accessed 11 September 2013)

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 (23), 8603-8627.

Riley, S.J., DeGloria, S.D., Elliot, R., 1999. A terrain ruggedness index that quantifies topographic heterogeneity. Intermountain Journal of Sciences 5, 23-27.

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.

Rossi, E., Kuitunen, M., 1996. Ranking of habitats for the assessment of ecological impact in land use planning. Biological Conservation 77, 227-234.

Ruiz, L.A., Fernandez-Sarría, A., Recio, J.A., 2004. Texture feature extraction for classification of remote sensing data using wavelet decomposition: A comparative study. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences 35 (Part B4), 1109-1115.

Sammon, J.W., 1969. A nonlinear mapping for data structure analysis. IEEE Transactions on Computers C-18, 401-409.

Sasaki, T., Imanishi, J., Ioki, K., Morimoto, Y., Kitada, K., 2012. Object-based classification of land cover and tree species by integrating airborne LiDAR and high spatial resolution imagery data. Landscape and Ecological Engineering 8, 157-171.

Smith, A., 2010. Image segmentation scale parameter optimization and land cover classification using the Random Forest algorithm. Journal of Spatial Science 55 (1), 69-79.

Strand, E.K., Smith, A.M.S., Bunting, S.C., Vierling, L.A, Hann, D.B., Gessler, P.E., 2006. Wavelet estimation of plant spatial patterns in multitemporal aerial photography. International Journal of Remote Sensing 27, 2049-2054.

Strobl, C., Boulesteix, A., Zeileis, A., Hothorn, T., 2007. Bias in random forest variable importance measures: Illustrations, sources and a solution. BMC Bioinformatics 8, 25, doi:10.1186/1471-2105-8-25.

Strobl, C., Boulesteix, A., Kneib, T., Augustin, T., Zeileis, A., 2008. Conditional variable importance for random forests. BMC Bioinformatics 9, 307, doi:10.1186/1471-2105-9-307.

Su, W., Zhang, C., Yang, J., Wu, H., Deng, L., Ou, W., Yue, A., Chen, M., 2012. Analysis of wavelet packet and statistical textures for object-oriented classification of forest-agriculture ecotones using SPOT 5 imagery. International Journal of Remote Sensing 33, 3557-3579.

Tarboton, D.G., 2012. Terrain analysis using digital elevation models (TauDEM) 5.0, http://hydrology.usu.edu/taudem/taudem5/index.html. (Accessed 11 September, 2013)

Thompson, S.D., Gergel, S.E., 2008. Conservation implications of mapping rare ecosystems using high spatial resolution imagery: Recommendations for heterogeneous and fragmented landscapes. Landscape Ecology 23, 1023-1037.

Thompson, S.D., Gergel, S. E., Coops, N.C., 2008. Classification of late seral coastal temperate rainforests with high spatial resolution QuickBird imagery. Canadian Journal of Remote Sensing 34, 460-470.

Tomppo, E., Halme, M., 2004. Using coarse scale forest variables as ancillary information and weighting of variables in k-NN estimation: a genetic algorithm approach. Remote Sensing of Environment 92, 1-20.

Tomppo, E., Olsson, H., Ståhl, G., Nilsson, M., Hagner, O., Katila, M., 2008. Combining national forest inventory field plots and remote sensing data for forest databases. Remote Sensing of Environment 112, 1982-1999.

Tomppo, E., Katila, M., Mäkisara, K., Peräsaari, J., 2012. The Multi-source National Forest Inventory of Finland - methods and results 2007. Working Papers of the Finnish Forest Research Institute 227, http://www.metla.fi/julkaisut/workingpapers/2012/mwp227-en.htm. (Accessed 11 September, 2013)

Tuominen, S., Eeronheimo, H., Toivonen, H., 2001. Yleispiirteinen biotooppiluokitus. Metsähallituksen luonnonsuojelujulkaisuja, Sarja B No 57, http://julkaisut.metsa.fi/assets/pdf/lp/Bsarja/b57.pdf. (Accessed 11 September, 2013)

Turner, W., Spector, S., Gardiner, N., Fladeland, M., Sterling, E., Steiniger, M. 2003. Remote sensing for biodiversity science and conservation. TRENDS in Ecology and Evolution 18, 306-314.

Vierling, K.T., Vierling, L.A., Gould, W.A., Martinuzzi, S., Clawges, R.M., 2008. Lidar: shedding new light on habitat characterization and modeling. Frontiers in Ecology and the Environment 6, 90-98.

Wang, Z., Boesch, R., Ginzler, C., 2012. Forest delineation of aerial images with Gabor wavelets. International Journal of Remote Sensing 33, 2196-2213.

Whitcher, B., 2012. waveslim: Basic wavelet routines for one-, two- and three-dimensional signal processing. R package version 1.7.1, http://CRAN.R-project.org/package=waveslim. (Accessed 11 September, 2013)

Whiteside, T.G., Boggs, G.S., Maier, S.W., 2011. Comparing object-based and pixel-based classifications for mapping savannas. International Journal of Applied Earth Observation and Geoinformation 13, 884-893.

Wright, C., Gallant, A., 2007. Improved wetland remote sensing in Yellowstone National Park using classification trees to combine TM imagery and ancillary environmental data. Remote Sensing of Environment 107, 582-605.

Yan, G., Mas, J.-F., Maathuis, B.H.P., Xiangmin, Z., van Dijk, P.M., 2006. Comparison of pixel-based and object-oriented image classification approaches – a case study in a coal fire area, Wuda, Inner Mongolia, China. International Journal of Remote Sensing 27, 4039-4055.

Yu, Q., Gong, P., Clinton, N., Biging, G., Kelly, M., Schirokauer, D., 2006. Object-based Detailed Vegetation Classification with Airborne High Spatial Resolution Remote Sensing Imagery. Photogrammetric Engineering & Remote Sensing 72, 799-811.

Table 1. Different habitat types mapped during field work and used in classification. Numbers and letters in parentheses refer to abbreviations used in text and other tables.

Habitat type	Number of age groups/management possibilities
xeric (pine dominated) forests	4: open regeneration area (0), sapling stand (1), young (2), mature (3)
mesic (spruce dominated) forests	5: open regeneration area (0), sapling stand (1), young (2), mature (3), natural (4)
herb-rich (mixed/deciduous) forests	4: sapling stand (1), young (2), mature (3), natural (4)
bare rock	1
pine mires	1
spruce mires	2: not drained, drained (d)
open mires	1
water (lakes and rivers)	1
small streams	1
springs	1
grasslands	1
fields	1
roads	1
yards	1
sand pits	1

Table 2. Different layers derived from datasets and features calculated from the layers to be used in classification. Layers that were used in segmentation are indicated in column titled Segmentation.

Data	Layer	Resolution	Segmentation	Calculated features
WV-2	Bands 1–8	2 m	Bands: 2, 3, 5, 7	Mean, standard deviation, GLCM, wavelets
	Bands 1–8	10 m		Mean, standard deviation
	NDVI	2 m		Mean, standard deviation, GLCM, wavelets
	NDVI	10 m		Mean, standard deviation
ALS	CHM	2 m	X	Mean, standard deviation, range, GLCM, wavelets
	CHM	10 m		Mean, standard deviation, range
	Intensity	2 m		Mean, standard deviation
	Intensity	10 m		Mean, standard deviation
	Slope	2 m		Mean, standard deviation
	Slope	10 m		Mean, standard deviation
	SWI	2 m	X	Mean, standard deviation
	SWI	5 m		Mean, standard deviation
	SWI	10 m		Mean, standard deviation
	DTW	2 m		Mean, standard deviation
	DTW	10 m		Mean, standard deviation
	TRI[1]	2 m		Mean, standard deviation
	TPI[2]	2 m		Mean, standard deviation
	MRVBF[3]	2 m		Mean, standard deviation
Digital soil map	Soil type	NA		Majority
Digital map	Land use type	NA		Majority
SLICES	Land use type	NA		Majority

[1] Window sizes: 3×3, 7×7, 11×11
[2] Radiuses: 10. 25, 50, and 100 m
[3] The initial thresholds for slope: 16 and 75

Table 3. Different classification alternatives tested.

Alternative	Boruta	Omitted features	Data split	ancillary	MS-NFI
1	no	no	no	no	no
2	no	ALS	no	no	no
3	no	WV-2	no	no	no
4	no	DTW, TPI, TRI, MRVBF	no	no	no
5	no	GLCM	no	no	no
6	no	wavelet	no	no	no
7	no	GLCM & wavelet	no	no	no
8	important	no	no	no	no
9	100 highest	no	no	no	no
10[1]	no	no	no	yes	no
11	no	no	yes	no	no
12	important	no	yes	no	no
13	100 highest	no	yes	no	no
14[1]	no	no	yes	yes	no
15	NA	NA	NA	yes	yes
16	NA	NA	NA	yes	yes + segmentation

[1]Ancillary data was added to the classifications (1–9 and 10–13) with the highest classification accuracies.

Table 4. Different classes used in classification and numbers of segments in each class in training data using whole area, above and below 1 m vertical distance to the nearest stream. Columns titled Lower and Upper refer to Fig. 6.

		Number of segments				
No	Class	Full data	Above 1m	Below 1m	Lower	Upper
1	xeric0	21	15	5	0.000	0.004
2	xeric1	15	15	0	0.005	0.007
3	xeric2	119	112	19	0.008	0.032
4	xeric3	4	3	1	0.032	0.033
5	mesic0	50	49	9	0.033	0.043
6	mesic1	492	471	29	0.043	0.144
7	mesic2	1120	981	151	0.144	0.374
8	mesic3	512	436	104	0.374	0.479
9	mesic4	628	596	37	0.479	0.608
10	herb-rich1	37	29	15	0.608	0.616
11	herb-rich2	62	42	18	0.616	0.628
12	herb-rich3	51	45	8	0.629	0.639
13	herb-rich4	5	4	2	0.639	0.640
14	rock	58	63	0	0.640	0.652
15	pine_m	65	2	62	0.652	0.665
16	spruce_m	94	19	95	0.665	0.685
17	spruce_m_d	4	2	3	0.685	0.685
18	open_m	16	0	16	0.686	0.689
19	water	464	0	461	0.689	0.784
20	stream	67	9	91	0.784	0.797
	spring		1	0		
21	meadow	41	18	25	0.798	0.806
22	field	382	184	240	0.806	0.884
23	road	236	200	56	0.884	0.931
24	yard	206	181	35	0.931	0.973
25	sand	134	103	47	0.973	1.000

Table 5. Producer's accuracy (in percent) for different habitat classes in different classification alternatives. In the table, there are some classes that do not exist in training data. First, some segments could not be assigned to any class due to missing data. This was because features based on 10 m resolution data could not be calculated for all segments. Class NA is assigned to these patches. Second, mire drainage mapping was based on ancillary data; therefore, also drained pine and open mires were mapped in some classifications.

Alternative	1	2	3	4	5	6	7	9	10	11	12	13	14	15	16
NA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
xeric0	78	76	61	78	72	78	71	77	77	66	69	67	67	0	0
xeric1	61	53	67	65	55	60	55	59	59	64	64	64	64	0	0
xeric2	77	72	65	76	76	76	76	79	78	73	73	78	77	16	14
xeric3	0	0	0	0	0	0	0	0	0	0	0	0	0	10	6
mesic0	58	48	0	58	57	55	53	58	56	54	54	56	55	0	0
mesic1	83	79	78	83	82	82	80	83	79	82	82	81	79	54	58
mesic2	90	88	90	90	91	91	91	90	88	89	90	89	86	51	59
mesic3	73	55	65	71	68	71	62	73	70	71	71	71	68	23	17
mesic4	89	84	88	89	91	90	87	89	89	88	88	89	88	52	59
herb-rich1	5	2	4	0	0	14	22	10	9	2	2	7	7	0	0
herb-rich2	12	7	5	7	14	17	13	13	30	13	15	17	21	2	1
herb-rich3	5	2	0	10	2	9	0	9	9	3	3	1	1	0	0
herb-rich4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
bare rock	46	48	31	41	48	48	49	47	57	45	45	41	55	50	51
pine mire	48	42	44	46	44	50	45	51	52	65	67	66	66	54	56
pine mire d	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
spruce mire	7	0	15	2	6	12	12	11	19	18	17	23	29	14	13
spruce mire d	0	0	0	0	0	0	0	0	27	0	0	0	40	39	39
open mire	47	10	12	22	71	64	64	75	73	39	61	73	69	61	61
open mire d	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
water	97	97	97	97	98	98	98	98	98	97	97	97	98	99	99
streamside	4	0	1	0	4	6	6	6	6	27	28	26	25	0	0
spring	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
meadow	20	14	11	17	18	28	23	23	35	31	31	35	32	34	34
field	94	92	88	94	94	93	90	95	96	93	93	94	96	97	96
road	51	48	41	50	52	48	48	51	94	50	51	49	94	91	91
yard	54	53	31	53	57	46	41	54	69	57	57	55	69	66	64
sand	97	94	91	96	97	97	95	98	98	97	97	97	99	75	75
Total	78.0	73.1	73.8	77.1	78.0	78.1	76.0	78.6	79.1	77.9	78.0	78.2	78.7	51.8	54.6

Table 6. User's accuracy for different habitat classes in different classification alternatives. In the table, there are some classes that do not exist in training data. First, some patches could not be assigned to any class due to missing data. This was because features based on 10 m resolution data could not be calculated for all segments. Class NA is assigned to these patches. Second, mire drainage mapping was based on ancillary data; therefore, also drained pine and open mires were mapped in some classifications.

Alternative	1	2	3	4	5	6	7	9	10	11	12	13	14	15	16
NA	0	0	0	0	0	0	0	0	0	0	0	0	0	100	100
xeric0	91	88	92	91	92	90	92	91	91	93	93	92	92	100	100
xeric1	90	96	93	88	94	89	96	92	92	92	92	89	89	0	0
xeric2	84	82	85	82	85	90	86	83	84	86	86	85	85	22	30
xeric3	100	100	100	100	100	100	100	100	100	100	100	100	100	1	0
mesic0	91	86	100	89	89	88	90	89	91	88	86	87	91	0	0
mesic1	81	75	67	80	81	80	79	81	84	81	81	80	83	61	68
mesic2	70	64	68	69	70	71	69	71	75	72	72	72	76	49	50
mesic3	64	59	60	62	64	65	60	65	67	66	66	66	69	17	14
mesic4	88	80	85	87	87	88	84	88	90	88	88	88	89	55	57
herb-rich1	100	100	74	100	100	98	84	71	29	100	67	100	36	0	0
herb-rich2	66	69	59	50	84	66	60	66	72	74	74	68	54	6	4
herb-rich3	62	93	100	79	100	58	100	74	76	85	85	66	5	100	100
herb-rich4	100	100	100	100	100	100	100	100	100	100	100	100	0	100	100
bare rock	91	83	87	87	87	86	83	89	76	89	89	91	78	68	46
pine mire	75	66	69	82	69	72	62	72	71	58	57	63	64	57	56
pine mire d	100	100	100	100	100	100	100	100	0	100	100	100	100	0	0
spruce mire	55	0	42	46	58	57	54	50	43	42	44	39	38	22	35
spruce mire d	100	100	100	100	100	100	100	100	19	100	100	100	25	42	42
open mire	55	71	38	58	60	62	57	61	57	62	66	66	61	34	35
open mire d	100	100	100	100	100	100	100	100	0	100	100	100	0	0	0
water	99	98	95	99	99	99	98	99	99	99	99	99	99	97	98
streamside	50	100	28	100	42	50	37	52	53	34	36	44	45	100	100
spring	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
meadow	67	73	57	72	74	75	70	75	46	88	80	81	45	45	45
field	87	84	84	85	88	87	87	87	89	88	89	90	89	86	94
road	70	68	62	69	69	66	63	67	55	67	66	66	55	46	48
yard	55	50	50	54	58	49	50	60	59	56	56	56	58	61	60
sand	97	94	95	98	97	98	95	100	100	99	99	99	100	100	100
Total	78.0	73.1	73.8	77.1	78.0	78.1	76.0	78.6	79.1	77.9	78.0	78.2	78.7	51.8	54.6

Table 7. Some extra classification accuracy calculations for the classification with the highest accuracy (all data, 100 features, ancillary data adjusted). In the table, there are some classes that do not exist in training data. First, some patches could not be assigned to any class due to missing data. This was because features based on 10 m resolution data could not be calculated for all segments. Class NA is assigned to these patches. Second, mire drainage mapping was based on ancillary data; therefore, also drained pine and open mires were mapped in some classifications.

	% OF ALL AREA								OF THE CLASS		
	Proportion in reference	Proportion in classified data	Total Difference	Quantity Difference	Allocation Difference	Agreement	Commission	Omission	Figure Of Merit	Commission	Omission
NA	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0	100	0
xeric0	0.5	0.4	0.1	0.1	0.1	0.4	0.0	0.1	72	9	23
xeric1	0.5	0.3	0.2	0.2	0.0	0.3	0.0	0.2	56	8	41
xeric2	4.0	3.8	1.5	0.3	1.2	3.2	0.6	0.9	68	16	22
xeric3	0.1	0.0	0.1	0.1	0.0	0.0	0.0	0.1	0	0	100
mesic0	0.8	0.5	0.4	0.3	0.1	0.4	0.0	0.3	53	9	44
mesic1	8.8	8.3	3.2	0.5	2.7	7.0	1.3	1.8	69	16	21
mesic2	26.0	30.4	10.9	4.4	6.5	22.8	7.6	3.2	68	25	12
mesic3	11.1	11.6	7.1	0.5	6.6	7.8	3.8	3.3	52	33	30
mesic4	15.2	15.0	3.3	0.2	3.1	13.5	1.6	1.7	80	10	11
herb-rich1	0.7	0.2	0.8	0.5	0.3	0.1	0.2	0.6	8	71	91
herb-rich2	1.2	0.5	1.0	0.7	0.3	0.4	0.1	0.8	27	28	70
herb-rich3	0.9	0.1	0.8	0.8	0.1	0.1	0.0	0.8	9	24	91
herb-rich4	0.1	0.0	0.1	0.1	0.0	0.0	0.0	0.1	0	0	100
bare rock	1.7	1.2	1.0	0.4	0.6	1.0	0.3	0.7	48	24	43
pine mire	1.1	0.8	0.8	0.3	0.5	0.6	0.2	0.6	43	29	48
pine mire d	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0	100	0
spruce mire	2.6	1.2	2.8	1.5	1.3	0.5	0.7	2.1	15	57	81
spruce mire d	0.1	0.2	0.2	0.1	0.2	0.0	0.1	0.1	13	81	73
open mire	0.3	0.3	0.2	0.1	0.1	0.2	0.1	0.1	47	43	27
open mire d	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0	100	0
water	9.9	9.8	0.3	0.1	0.2	9.7	0.1	0.2	97	1	2
streamside	1.7	0.2	1.7	1.5	0.2	0.1	0.1	1.6	5	47	94
spring	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0	0	100
meadow	0.6	0.4	0.6	0.1	0.5	0.2	0.2	0.4	25	54	65
field	6.4	7.0	1.0	0.5	0.5	6.2	0.8	0.2	86	11	4
road	2.4	4.1	2.0	1.7	0.3	2.2	1.8	0.2	53	45	6
yard	1.9	2.2	1.5	0.3	1.2	1.3	0.9	0.6	47	41	31
sand	1.4	1.4	0.0	0.0	0.0	1.4	0.0	0.0	98	0	2
TOTAL	100	100	20.9	7.6	13.3	63.0	18.4	18.4		18	18

Table 8. Important features based on random forest run for all data and Boruta. Three most important features of all habitat classes that could be mapped are shown. Habitat class numbers are given in Table 3.

		CLASSES																					OVERALL		
		1	2	3	5	6	7	8	9	10	11	12	14	15	16	18	19	20	21	22	23	24	25	Z	mZ	Gini
WV-2	b1			1			1			1								2						1	3	10
	b1_10m						2			2														8	7	18
	b1_std																				2			102	105	56
	b7																		1					56	47	29
	b8_10m																		2					16	24	31
WV-2 / GLCM	b1_sav																						2	35	43	36
	b1_sva																						1	59	51	51
	b2_sav																						3	20	29	43
	ndvi_con																				3			64	88	101
	ndvi_dva																				1			123	90	82
WV-2 / wavelet	b3_hor_l1																			2				21	28	46
	b5_ver_l1										2													83	207	158
	b6_hor_l2																					1		24	133	116
	ndvi_hor_l3																					3		14	62	166
ALS	CHM								2			3												23	15	3
	CHM_range			2					1															2	2	1
	CHM_std							2			3													28	18	12
	Intensity_std																					2		5	6	87
CHM / GLCM	chm_asm	2			2																			105	57	6
	chm_ent	3	2		1	3																		37	19	4
	chm_idm	1	3		3							1							3					101	70	7
	chm_sen			3				1	3		1	2												9	9	2
	chm_var							3																26	13	13
CHM/wavelet	chm_dia_l1																			1				158	126	16
Topography	slope																3							33	38	15
	SWI																1							61	63	11
	SWI_10m												2											17	14	23
	DTW						3						1		2			1						3	1	28
	DTW_10m																	3						6	5	35
	MRVBF16													2	1									12	31	50
	MRVBF75					1				3			3	1	3						3			4	4	30
	TRI11													3		3								43	23	24
	TRI3		1			2										1	2							15	20	9
	TRI7															2								46	22	22

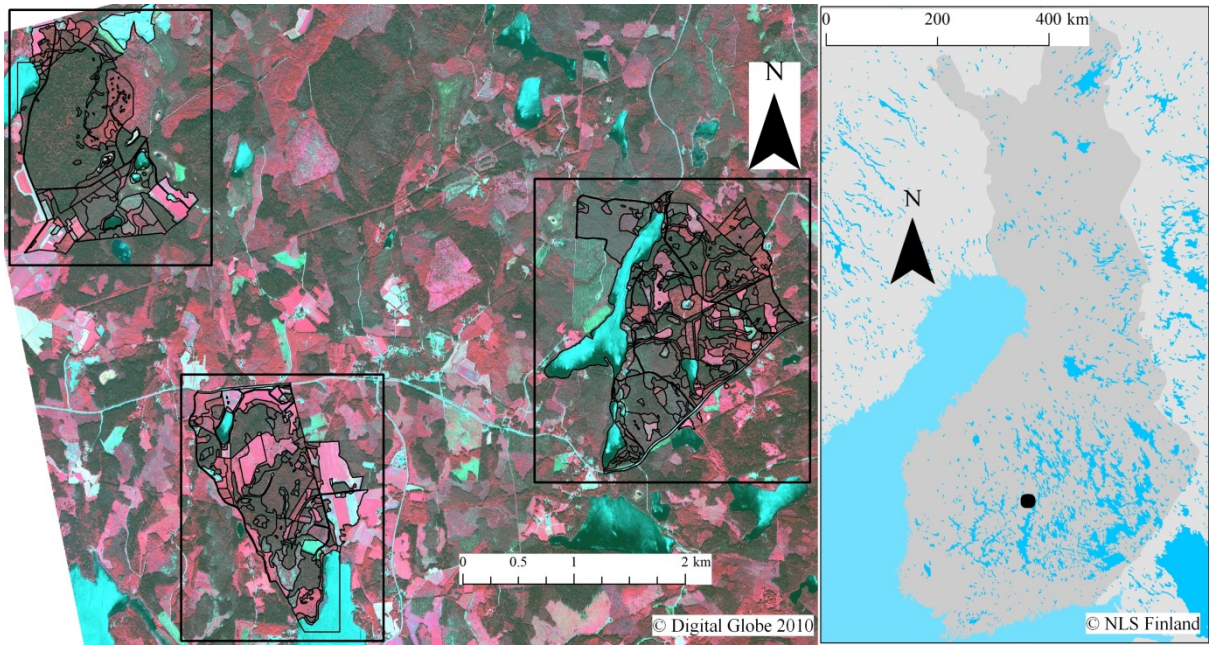

Figure 1. Our study area located in Central Finland. On the left side is a false color WorldView-2 image (red channel: band 7/NIR1, green: band5/red, blue: band3/green) from our study area. Study area is divided into three sub-areas which are marked with black rectangles. Field work reference polygons inside sub-areas are drawn in black.

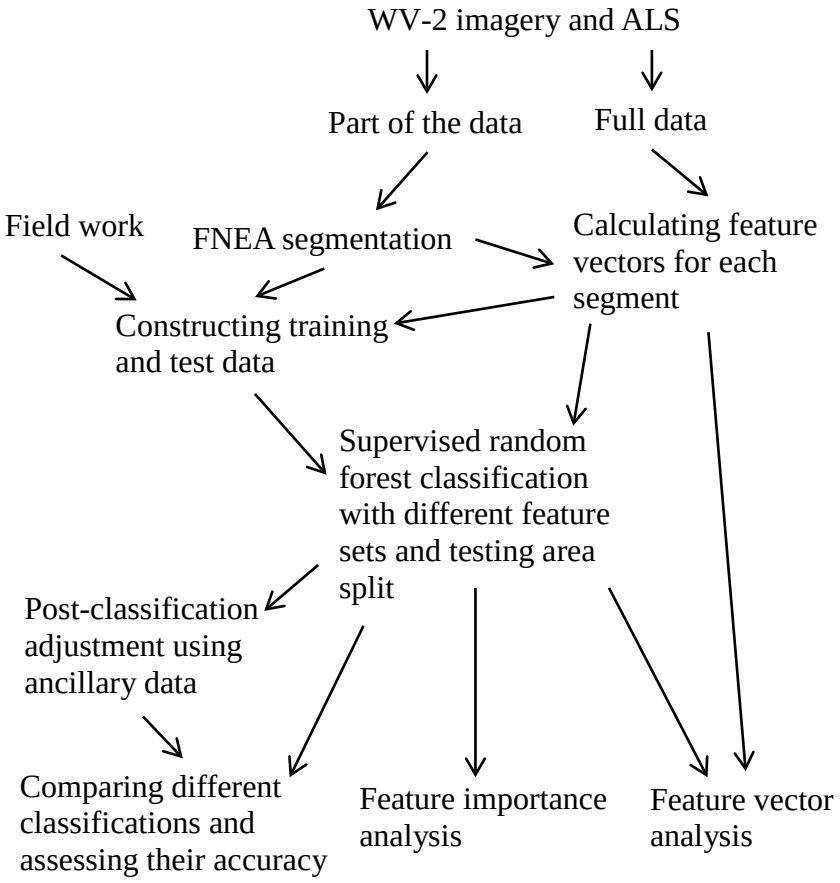

Figure 2. A flow chart of the used classification approach, in which Fractal Net Evolution Approach (FNEA) segmentation was coupled with random forest classification.

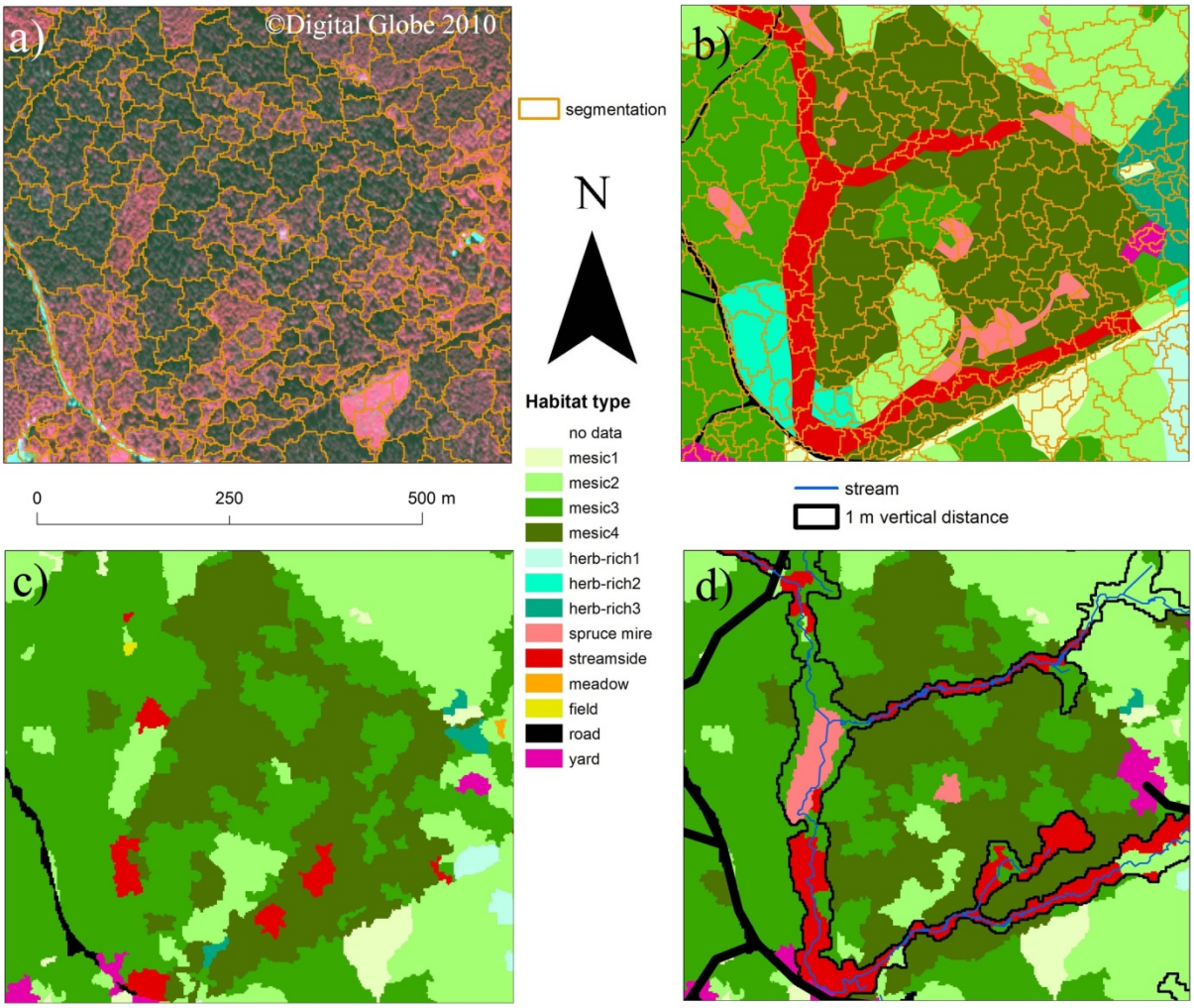

Figure 3. Classification example inside easternmost sub-area. a) WV-2 imagery (Red: band 7, G: band 5, B: band 3) shown with segmentation boundaries. b) Field work reference polygons shown with segmentation boundaries. c) Classification output with 100 features selected with Boruta. d) Classification output with data split and 100 features selected by Boruta and ancillary data. Also mapped stream network and 1 m vertical distance to the nearest stream are shown.

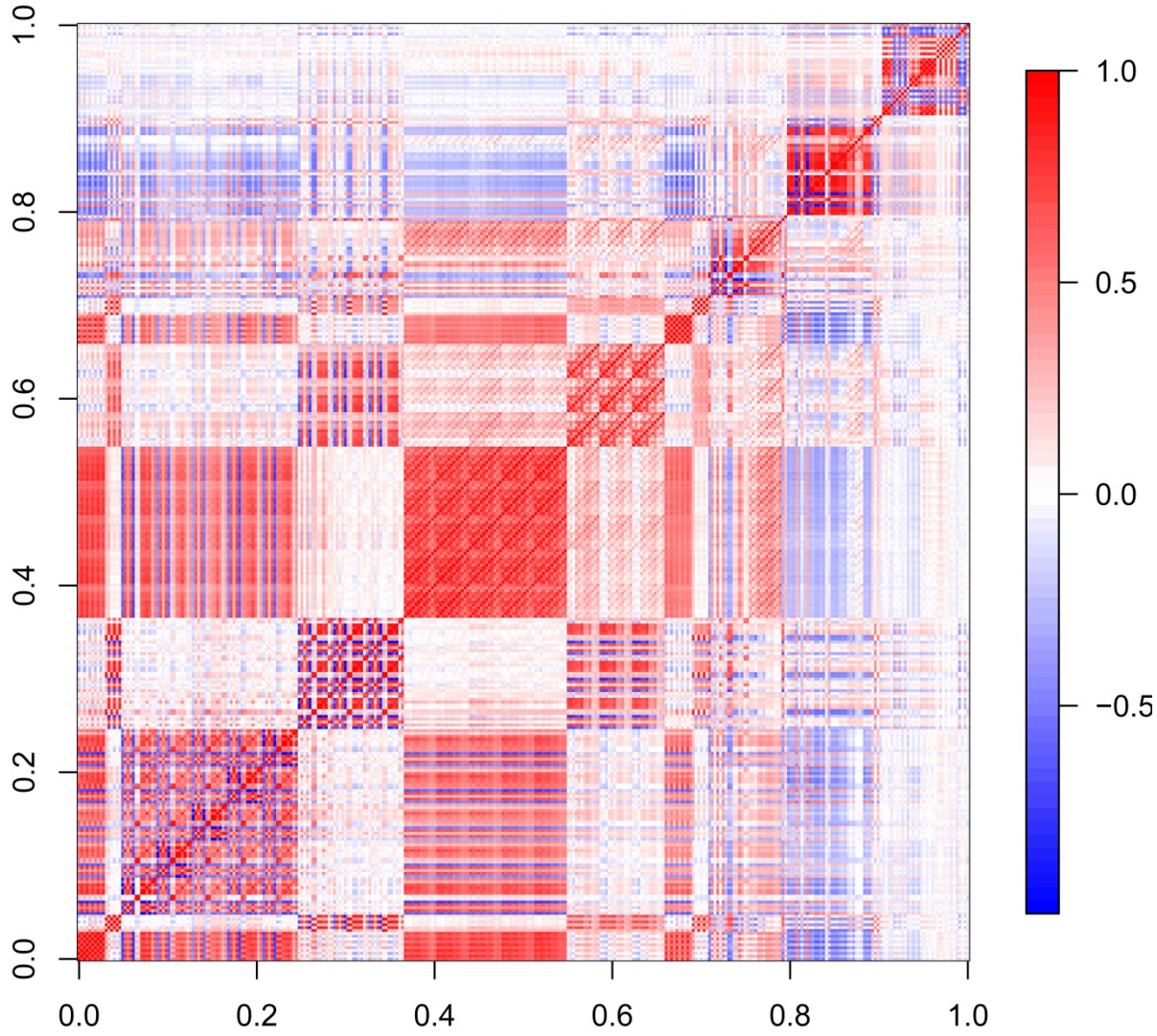

Figure 4. Correlation image of the different features used in classification. Features are in the same order as in Table 2. WV-2 layer mean and std values are between values 0 and 0.05, GLCM features between 0.05 and 0.366, and wavelet features between 0.368 and 0.66. WV-2 10 m features are between 0.66 and 0.71, NDVI features are between 0.71 and 0.796, CHM features between 0.798 and 0.89, and topographic features between 0.9 and 1.

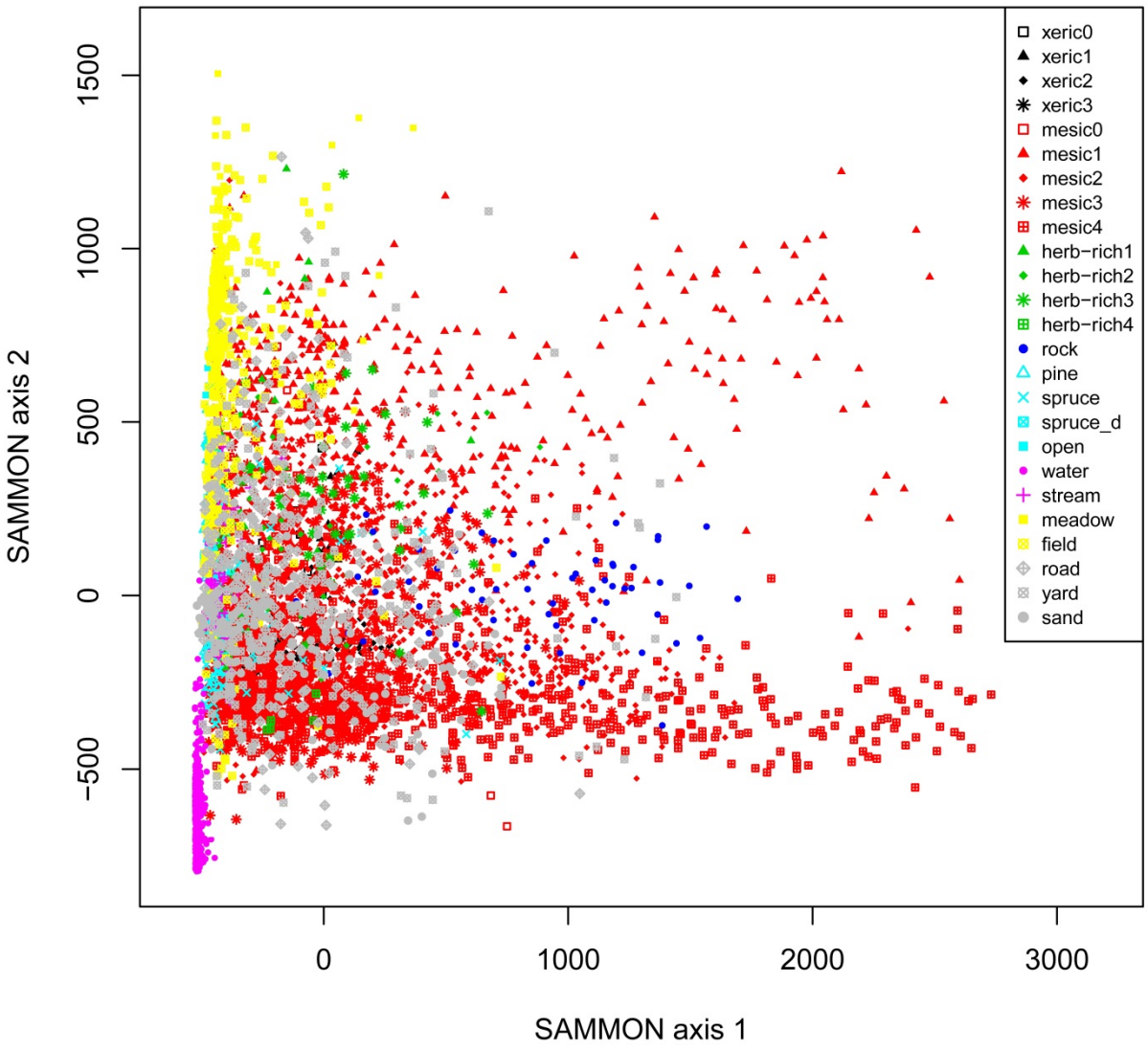

Figure 5. Sammon's map of all segments in training data using 100 features in whole study area.

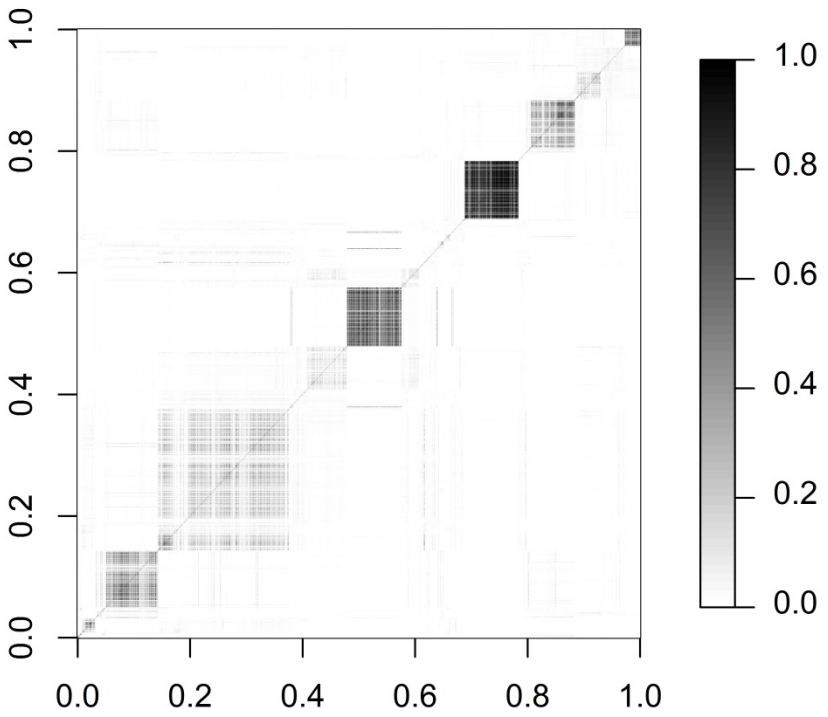

Figure 6. Proximity image of random forest classification using training data of the whole area and 100 features selected by Boruta algorithm. Values show how close to each other different segments are. Value 1 means high proximity and 0 low proximity. Different segments are in x and y axes as clarified in Table 4.