Usability of one-class classification in mapping and detecting changes in bare peat surfaces in the tundra

Aleksi Räsänen, corresponding author

Ecosystems and Environment Research Programme, Faculty of Biological and Environmental Sciences, and Helsinki Institute of Sustainability Science (HELSUS), PO Box 65, FI-00014 University of Helsinki, Finland, aleksi.rasanen@helsinki.fi

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

Vladimir Elsakov

Institute of Biology of Komi Scientific Centre of the Ural Branch of the Russian Academy of Sciences, Kommunisticheskaya, 28, Syktyvkar, Komi Republic, Russia, elsakov@ib.komisc.ru

Tarmo Virtanen

Ecosystems and Environment Research Programme, Faculty of Biological and Environmental Sciences, and Helsinki Institute of Sustainability Science (HELSUS), PO Box 65, FI-00014 University of Helsinki, Finland, tarmo.virtanen@helsinki.fi

word count: 10,357

Usability of one-class classification in mapping and detecting changes in bare peat surfaces in the tundra

Arctic areas have experienced greening and changes in permafrost caused by climate change during recent decades. However, there has been a lack of automated methods in mapping changes in fine-scale patterns of permafrost landscapes. We mapped areal coverage of bare peat areas and changes in them in a peat plateau located in north-western Russia between 2007 and 2015. We utilized QuickBird and WorldView-3 satellite image data in an object-based setting. We compared four different one-class classifiers (one-class support vector machine, binary support vector machine, random forest, rotation forest) both in a fully supervised binary setting and with positive and unlabelled training data. There was notable variation in classification performance. The bare peat area *F*-score varied between 0.77 and 0.96 when evaluated by cross-validated training data and between 0.22 and 0.57 when evaluated by independent test data. Overall, random forest performed the most robustly but all classifiers performed well in some classifications. During the 8 year period, there was a 21%–26% decrease in the bare peat areal coverage. We conclude that (1) tested classifiers can be used in one-class settings and (2) there is a need to develop methods for tracking changes in single land cover types.

Keywords: Arctic, change detection, land cover, one-class classification, remote sensing, very high spatial resolution

1. Introduction

Arctic areas in general and Arctic peatlands in particular have been heavily influenced by recent global warming. It has been reported that the Arctic is greening, usually meaning that the amount of green vegetation, especially bush coverage, is increasing in tundra areas (Tape, Sturm and Racine 2006; Elmendorf et al. 2012; Myers-Smith et al. 2015; Yu et al. 2017), and that climate warming and geomorphic disturbances have caused changes in permafrost areas (Beck et al. 2015; Segal, Lantz and Kokelj 2016; Borge et al. 2017; Obu et al. 2017). It has been found that bare peat surfaces, the identification of which is the target of our study, can be significant nitrous oxide, carbon

dioxide and methane sources in the tundra (Repo et al. 2009; Marushchak et al. 2011; Voigt et al. 2016). Furthermore, their areal coverage seems to be dynamic in time and their abundance in the Arctic is still poorly known (Repo et al. 2009; Marushchak et al. 2011; Voigt et al. 2016).

Remote sensing methods are widely used in mapping Arctic land cover and permafrost (Pflugmacher et al. 2011; Westermann et al. 2014; Bartsch et al. 2016) and tracking changes in land cover patterns (Singh 1989; Stow et al. 2004; Beck et al. 2015; Tewkesbury et al. 2015; Jorgenson and Grosse 2016). Due to fine-scale heterogeneity in thermokarst and tundra environments, very high spatial resolution (VHSR, spatial resolution < 5.00 m) satellite images are needed to map the extent and spatial patterns of small land cover patches such as bare peat surfaces (Laidler and Treitz 2003; Belshe, Schuur and Grosse 2013; Virtanen and Ek 2014). It has been argued that more high-resolution studies are needed to map changes in Arctic environments, and there is a need to develop standardized methods for tracking the changes (Jorgenson and Grosse 2016).

There are some but not many previous studies evaluating fine-scale changes in permafrost landscapes. Borge et al. (2017) combined visual interpretation with an analysis of climate data when they evaluated the decrease in palsa mires in northern Norway. Similarly, visual interpretation has been used in the mapping of changes in thermokarst lakes and lithalsas in Quebec (Beck et al. 2015) and in thaw slump activity in the western Canadian Arctic (Segal et al. 2016). While these studies have incorporated visual interpretation, there has been a lack of automated approaches for mapping changes in fine-scale patterns of single land cover classes.

To map patches of single land cover classes, there has been a recent rise in the use of one-class, binary and positive and unlabelled (PU) classification methods as

opposed to multi-class classification methods (Boyd, Sanchez-Hernandez and Foody 2006; Sanchez-Hernandez, Boyd and Foody 2007; Li and Guo 2010; Li, Guo and Elkan 2011; Mack, Roscher and Waske 2014; Baldeck and Asner 2015; Baldeck et al. 2015; Barbosa et al. 2016; Chen et al. 2016; Mack et al. 2016; Song et al. 2016; Mack and Waske 2017; Stenzel et al. 2017; Araya-López et al. 2018; Deng et al. 2018; Liu et al. 2018). In one-class classification, training data are needed only for the focal class (i.e. presence data). This makes it easier to use than multi-class classification in which training data are needed for all classes. In PU classifications, there are also training data for unlabelled (so-called pseudo-outlier) points, whereas binary classifiers are performed in a one-against-all setting.

Typical methods in one-class, PU and binary classification include support vector machines (SVMs) (Boyd et al. 2006; Sanchez-Hernandez et al. 2007; Baldeck and Asner 2015; Mack et al. 2016; Stenzel et al. 2017), a maximum entropy approach (Maxent) (Li and Guo 2010; Lin et al. 2014; Stenzel et al. 2014; Amici et al. 2017; Andrew and Shephard 2017; Stenzel et al. 2017; Chignell et al. 2018; Noviello et al. 2018), as well as PU and presence and background learning algorithms (Li et al. 2011; Xu et al. 2016; Ao et al. 2017; Deng et al. 2018; Liu et al. 2018). Also, regression methods such as random forest (RF) and boosted regression trees have been used in mapping the presence of single land cover types (Chignell et al. 2018), while state-of-the-art classification methods other than SVMs, Maxent and PU learning have been little used.

SVM is a non-parametric statistical learning technique in which observations are divided into classes with the help of hyperplanes (Vapnik 1982; Mountrakis, Im and Ogole 2011). In one-class SVM (OCSVM), only positive data are used in the training of the classifier; in biased SVM, PU data are used, and positive data are given a larger

misclassification cost; and in binary SVM, data from two classes are used (Baldeck and Asner 2015; Baldeck et al. 2015; Barbosa et al. 2016; Mack et al. 2016; Mack and Waske 2017; Stenzel et al. 2017). Maxent is a one-class machine learning method in which relative entropy between two probability densities (i.e. presence data and environmental covariates) is trying to be minimized (Elith et al. 2011). Maxent has been typically used in the modelling of suitable habitats or species distributions of specific species using remote sensing and other data (Phillips, Anderson and Schapire 2006; Elith et al. 2011), but it has also been used in other kinds of remote sensing tasks, such as mapping archaeological sites (Noviello et al. 2018), eagle nests (Andrew and Shephard 2017) and habitat types (Stenzel et al. 2014; Chignell et al. 2018). As the name of the method indicates, the PU learning algorithm uses PU data in model building. The PU learning algorithm is carried out in two steps: the model is first trained with PU data, and a specific constant factor is estimated from validation data (Li et al. 2011). PU learning needs to be implemented with a classification algorithm that can estimate conditional probabilities. Usually, PU has been implemented with back-propagation neural networks (Li et al. 2011; Xu et al. 2016; Ao et al. 2017; Deng et al. 2018) but also model comparisons have been conducted (Liu et al. 2018).

When comparing different classification algorithms, it has been found that biased and binary SVMs in particular are promising methods for mapping patches of rare habitat type (Boyd et al. 2006; Sanchez-Hernandez et al. 2007; Mack et al. 2016; Stenzel et al. 2017) and that they usually outperform other one-class classification methods (Boyd et al. 2006; Baldeck and Asner 2015; Baldeck et al. 2015; Barbosa et al. 2016; Mack and Waske 2017; Stenzel et al. 2017). In some comparisons against biased or binary SVMs, slightly higher classification accuracy has been achieved with iterative

biased SVM (Mack et al. 2016) and the PU learning algorithm (Li et al. 2011; Xu et al. 2016; Ao et al. 2017; Deng et al. 2018; Liu et al. 2018).

Here, we address three gaps in the remote sensing literature. First, we assess one-class classification in tracking fine-scale changes in tundra environments, which have been studied relatively little. Second, there have been few attempts to use one-class and PU classification methods to track changes, and these studies have employed Maxent (Lin et al. 2014; Amici et al. 2017) or OCSVM (Li and Xu 2010) in rather coarse resolution mapping. Here, we map fine-scale changes in bare peat surfaces in Russian tundra using VHSR satellite images at two time points (2007 and 2015). Third, although there has been a notable increase in one-class, biased and binary classifications, there has been little evaluation of how methods other than SVM, Maxent and PU learning handle the binary and PU classification situations. Hence, we test the performance of RF and rotation forest (ROF) against binary/biased SVM (BSVM) and OCSVM. We evaluate these methods in an object-based setting by comparing fully supervised binary classification (i.e. peat surfaces against all other land cover classes) with PU classification in which we used training data from bare peat areas and unlabelled random points.

2. Materials and methods

2.1.Study area and bare peat surfaces

Our study area is in the Seida peat plateau (67°04′00″N, 62°55′00″E, Figure 1) located in a discontinuous permafrost zone in the Usa River Basin in the north-eastern corner of European Russia, 70 km south-west of the town of Vorkuta. The area is characterized by low Arctic shrub vegetation and lies just north of the tree line. The mean annual temperature in the region is approximately -6 °C and the annual precipitation is just

above 505 mm (Repo et al. 2009; Marushchak et al. 2011; Hugelius et al. 2012; Voigt et al. 2016). The area is composed of a mosaic of peat plateaus, mineral moraine deposits, and thermokarst lakes. Peat plateaus with permafrost bogs and fens rise above the surrounding landscape due to high ground ice content, and the peat layer is in many locations over 4 m thick (Hugelius et al. 2012).

FIGURE 1 APPROXIMATELY HERE

We tracked the areal coverage and changes in the extent of bare peat surfaces, whose coverage has been estimated to be around 1% of the peat plateau (Marushchak et al. 2013). These surfaces have been developed by late Holocene permafrost aggradation (Väliranta et al. 2018), wind erosion (Seppälä 2003) and soil mixing due to frost action, i.e. cryoturbation (Bockheim and Tarnocai 1998; Peterson and Krantz 2003; Repo et al. 2009). According to recent findings, the surfaces are remnants of larger non-vegetated areas in the past (Väliranta et al. 2018). Bare peat surfaces are of two types. Peat circles are flat surfaces with no vascular plants but are partly covered by thin moss and algal crust. Peat mounds are drier than peat circles and partly vegetated by lichens and graminoids. Peat mounds do not have a frozen core but are located within a frozen peat plateau (Marushchak et al. 2011).

2.2. Used data

In the classifications, we used two VHSR satellite images taken during different growing seasons. The first one was a QuickBird (QB) image (DigitalGlobe, Westminster, CO, USA) with 2.40 m spatial resolution and four bands (blue, green, red, near-infrared) taken on 6 July 2007. The second one was a WorldView-3 (WV-3) image (DigitalGlobe, Westminster, CO, USA) with 1.60 m spatial resolution and eight bands (coastal blue, yellow, red-edge and near-infrared 2 in addition to the QB bands) taken on 9 August 2015. We first atmospherically corrected both images using the dark-object

177 subtraction method (Chavez 1988) and then georeferenced them to match the field data
178 and each other.
179 To acquire the ground reference data, we used (1) visual interpretation of
180 satellite images and (2) a DJI (Dà-Jiāng Innovations) Phantom 2 Vision Plus drone
181 image with 0.04 m resolution taken on 2 August 2015 as well as (3) fieldwork data sets
182 from July–August 2007 and 2016. The mosaic of drone images (nearly 615 images) was
183 developed using Agisoft PhotoScan software (Agisoft, St Petersburg, Russia). In the
184 fieldwork, we collected data mostly for a purpose other than mapping bare peat areas
185 but got a good overall view of the bare peat areas and received help in how they could
186 be interpreted from satellite images.

187 ***2.3.Overview of the classification approach***

188 It has been shown that object-based methods are more applicable than pixel-based
189 methods in classifying VHSR images (Blaschke et al. 2014). Therefore, we employed
190 an object-based classification approach in which we first segmented images and then
191 classified them in two different settings (Figure 2). In the first classification setting, we
192 carried out a fully supervised binary classification and in the second setting, we
193 classified the image with the help of PU segments (PU classification setting). In both
194 settings, we tested four different classifiers and reduced the number of features using a
195 feature selection algorithm. Finally, we assessed classification accuracy in a different
196 area than the training area. Classifications were carried out separately for the QB and
197 WV-3 images. For WV-3, we carried out two classification alternatives: a low-
198 resolution (we use the acronym WV-3LR) version with data resampled at 2.40 m
199 ground resolution and four bands identical to QB classification, and a high-resolution
200 (WV-3HR) version with 1.60 m resolution and all eight bands.
201 FIGURE 2 APPROXIMATELY HERE

2.4. Image segmentation

We segmented images using full lambda schedule (FLS) segmentation in Erdas Imagine 2014 (Intergraph, Madison, AL, USA). FLS segmentation is a region-based segmentation in which pixels are merged with their neighbouring pixels if certain conditions are met. FLS uses spectral (mean pixel value in the segment), textural (standard deviation of pixel values in the segment), shape (areal complexity of the segment) and size information when decisions about merging are carried out. This information can be given relative weights. Furthermore, the average size of the segment (i.e. pixel/segment ratio) and the minimum and maximum segment sizes can also be adjusted.

We parameterized segmentations with the help of visual interpretation as it is often the most meaningful method for segmentation evaluation in natural environments (Räsänen et al. 2013). When setting the parameters, we had three objectives: (1) the segments should be as large as possible, but (2) even the smaller bare peat areas should be delineated, and (3) the average size of the segment should be the same for all classifications. We ended up giving weights of 0.7, 0.5, 0.3 and 0.3 to spectral, textural, shape and size information respectively. Hence, we gave the largest relative weight to spectral information as the bare peat surfaces usually have a distinct colour in the images. We set the pixel/segment ratio to 50 for 2.40 m resolution classifications and to 112 for the 1.60 m resolution classification, and we gave minimum and maximum segment sizes the default values of 10 and 2000 respectively.

2.5. Calculated features

For each segment and image band, we calculated 34 features using package EBImage (Pau et al. 2010) in R 3.2.2 (R Core Team 2015). To analyse texture, we calculated the

following 13 grey-level co-occurrence matrix (GLCM) features: angular second moment, contrast, correlation, sum of squares: variance, inverse difference moment, sum average, sum variance, sum entropy, entropy, difference variance, difference entropy, and two information measures of correlation (Haralick, Dinstein and Shanmugam 1973) which are among the most widely used textural features (Benz et al. 2004; Pacifici, Chini and Emery 2009; Blaschke et al. 2014). When calculating the GLCM features, we quantized data to 32 levels. We calculated all GLCM features at two scales; in other words, we calculated the adjacency of pixels to their closest and second closest neighbouring pixels. In addition, we calculated mean, standard deviation, mean absolute deviation, and 1%, 5%, 50%, 95% and 99% quantiles. We calculated features from each image band as well as from three band indices: normalized difference vegetation index (NDVI, Rouse et al. 1973), red-green index (RGI, Coops et al. 2006) and normalized difference water index (McFeeters 1996). Hence, we ended up with 374 features for the WV-3HR classification and 238 features for the QB and WV-3LR classifications.

2.6.Classification settings

We used two classification settings: binary and PU. For the binary classification, we manually classified an area of the images for bare peat segments and other segments. We carried out this classification in the area where most of the fieldwork in 2016 was conducted. We used our knowledge of the terrain collected during the fieldwork and visual interpretation of the VHSR images when constructing the training data. In binary classifications, we ended up with 122 (WV-3HR), 92 (WV-3LR) and 98 (QB) segments for the bare peat class, and 1112 (WV-3HR), 1027 (WV-3LR) and 1067 (QB) for the other class (Figure 3). For the PU classification, we used the same bare peat areas as for the binary classification. For the unlabelled points, we randomly drew 1000 segments

from the overall study area.

FIGURE 3 APPROXIMATELY HERE

2.7. Used classifiers and their parameterization

We used four different classifiers: OCSVM, BSVM, ROF and RF in both classification settings.

SVMs have been used widely in binary, PU and one-class settings, and BSVM has performed the best in various comparisons (Baldeck and Asner 2015; Baldeck et al. 2015; Barbosa et al. 2016; Mack and Waske 2017; Stenzel et al. 2017). In SVMs, observations are divided into a discrete number of classes with the help of hyperplanes, and the closest training samples to the hyperplanes are called support vectors (Vapnik 1982; Mountrakis et al. 2011). In its simplest form, SVM is a linear binary classifier. However, there are also techniques such as kernels that modify hyperplanes so that separability between classes does not need to be linear and also further adjustments that enable classification into multiple classes (Mountrakis et al. 2011). SVM generally needs only small training data sets and handles multidimensional data sets making it appealing for remote sensing classification tasks (Mountrakis et al. 2011). We used a radial basis function kernel (Vert, Tsuda and Schölkopf 2004) when conducting SVM classifications in one-class, biased and binary settings. SVMs are sensitive to changes in parameter values and typically a grid search is used in parameterization (Mack et al. 2014; Baldeck and Asner 2015; Baldeck et al. 2015; Barbosa et al. 2016; Mack et al. 2016; Mack and Waske 2017; Stenzel et al. 2017).

OCSVM uses only data from the positive class and searches for the optimal configuration that separates the positive class (Schölkopf et al. 2000). In OCSVM with radial basis function kernel, two parameters need to be set: inverse kernel width (σ) (Vert et al. 2004) and the upper bound for the training error (ν) (Schölkopf et al. 2000).

After initial evaluation we tested the following parameter values: $\sigma \in \{1 \times 10^{-7}, 1 \times 10^{-6}, \ldots, 1 \times 10^{-2}\}$ and the upper bound for the training error $v \in (0.01, 0.05, 0.10, \ldots, 0.25\}$.

Binary and biased SVM use data from two classes in training. In binary SVM, it is known that training samples originate from two different classes. The classifier tries to maximize the distance between the hyperplane and support vector and minimize misclassifications. Misclassifications are penalized using a cost parameter (*C*) (Vapnik 1982). In biased SVM, unlabelled samples may include samples from the positive class (Liu et al. 2003). Therefore, two classes are given a different weight through *C*; misclassifications of the focal class are penalized more than misclassifications in the unlabelled samples using an extra cost parameter *C+* (Liu et al. 2003). We tested the following parameter values with BSVM: $\sigma \in \{1 \times 10^{-7}, 1 \times 10^{-6}, \ldots, 1 \times 10^{-2}\}$, $C \in \{1 \times 10^{3}, 1 \times 10^{4}, \ldots, 1 \times 10^{7}\}$ and $C+ \in \{1 \times 10^{0}, 1 \times 10^{1}, \ldots, 1 \times 10^{5}\}$. We thus allowed equal misclassification costs for both classes. For both OCSVM and BSVM, we tuned parameters using tenfold cross-validation.

RF has generally performed well in various comparisons and it is capable of handling multidimensional data (Duro, Franklin and Dubé 2012; Rodriguez-Galiano et al. 2012; Adam et al. 2014; Belgiu and Dragut 2016). It is an ensemble classifier which combines multiple bootstrapped classification trees, and final classification is based on a majority vote over all trees. When a tree is built, two thirds of the data are used for training and the rest is used for evaluation and is called out-of-bag data. At each tree node, the best split is chosen among a randomized subset of input features and the number of tested features can be adjusted (Breiman 2001). In order to deal with imbalanced data, one can increase the misclassification cost of a specific class by giving a higher value to the class weight (*W*) parameter (Chen, Liaw and Breiman 2004). We tuned *W* using out-of-bag assessment. After initial testing to find a suitable range for the

positive class W, we tested the following values: $W \in \{1 \times 10^5, 2 \times 10^5, \ldots, 19 \times 10^5\}$. Otherwise, RF has been found to be insensitive to parameterization (Rodriguez-Galiano et al. 2012; Du et al. 2015), which we also verified in our initial evaluations. Thus, we used the default values for the other parameters. We set the number of trees grown to 500 and the number of parameters tested in each node to the square root of the total number of variables.

ROF has proven to be an efficient method for classifying remote sensing data (Kavzoglu and Colkesen 2013) and it has outperformed SVM and RF in some comparisons (Du et al. 2015). ROF is also an ensemble classifier of multiple classification trees. In each tree, features are randomly split into subsets, and principal component analysis is conducted for each subset so that the full feature set can be reconstructed with subset axis rotation. Confidence for each class is calculated using an average combination over all trees, and samples are assigned to the class with the highest confidence (Rodriguez and Kuncheva 2006). Similarly to RF, ROF is insensitive to parameterization (Du et al. 2015), and we verified this when testing the classifier. We thus used default parameter values; in more detail, we chose the number of variable subsets so that there were three features in each subset, and we set the number of trees to ten.

For SVMs and RF, we chose the best parameter values based on the *F*-score, which has been found to be applicable in binary and one-class classification settings (Baldeck and Asner 2015; Baldeck et al. 2015). We conducted classifications in R 3.2.2 (R Core Team 2015) using packages oneClass (Mack et al. 2014), randomForest (Liaw and Wiener 2002) and rotationForest (Ballings and Van der Poel 2015).

2.8.Feature selection

In previous studies, it has been found that feature selection improves the classification accuracy in SVM and RF classifications (Low et al. 2013; Rasanen et al. 2014; Li, XJ et al. 2016). For selecting features, we used the RF wrapper algorithm Boruta (Kursa and Rudnicki 2010) in R 3.2.2 (R Core Team 2015). Boruta and other RF-based feature selection are compatible with SVMs (Low et al. 2013; Li, XJ et al. 2016) and have performed well in different comparisons (Samsudin et al. 2015; Li, Tran and Siwabessy 2016; Poona et al. 2016). Boruta feature selection proceeds as follows. Before each RF run, shadows for each feature are created. Shadow feature values are obtained by disarranging the values of the original feature across samples. After the run, the features which have significantly lower importance than the shadow feature with the highest importance are judged to be unimportant, and features having significantly higher importance are judged important. Unimportant features are removed from the following runs and Boruta ends when all features are deemed either important or unimportant or when the maximum amount of RF runs have been conducted (Kursa and Rudnicki 2010), which in our case was after 1000 runs. For those variables which were still tentative after 1000 runs, we carried out a tentative rough fix in which the median importance of tentative features was compared with the median importance of the most important shadow feature (Kursa and Rudnicki 2010).

2.9.Accuracy assessment and change detection

We tested the accuracy of the classifications in an area covered by the drone image. In constructing the reference against which classifications were evaluated, we visually interpreted the drone image and delineated all bare peat areas with > 20 m^2 extent. In addition, we visually estimated whether the bare peat areas visible in the drone image

348 could be seen in the satellite images. We evaluated the performance of the methods by
349 both pixel-based and object-based metrics. First, we calculated the user's accuracy (i.e.
350 precision, the likelihood that the mapped area is a bare peat surface) and the producer's
351 accuracy (i.e. recall, the proportion of correctly classified bare peat surfaces). We also
352 calculated the *F*-score (i.e. the harmonic mean of the user's and producer's accuracy)
353 based on (1) the test data and (2) the out-of-bag (RF) or tenfold cross-validated (other
354 classifiers) estimate on the training data. Furthermore, we calculated the proportion of
355 positively classified segments having at least 10% and 50% bare peat coverage based on
356 the reference data. Then, we calculated the proportion of bare peat patches in the
357 reference of which 10% or 50% were classified as bare peat surfaces in the remote
358 sensing classifications. Finally, we calculated the number and areal coverage of bare
359 peat surfaces in the validation area based on each classifier and the reference, and in the
360 whole 10.9 km^2 study area of which approximately 5.5 km^2 is peat plateau (see
361 Hugelius et al. 2012 for the peat plateau area).
362 To track changes in bare peat areal coverage, we compared QB and WV-3LR
363 classifications, since they had similar spectral and spatial resolutions. We compared the
364 areal coverage and number of bare peat area patches in the whole study area using the
365 following metrics. First, we calculated the consensus of 3, 5 and 7 classifiers; in other
366 words, we mapped those segments that were classified as bare peat based on the
367 prediction of 3, 5 or 7 classifiers. Second, we calculated the average values of areal
368 coverage and number of patches over all classifications and reported also the lowest and
369 highest estimate of single classifiers.

3. Results

3.1.Feature and parameter selection

The Boruta feature selection reduced the number of features to 88, 80, 66, 54, 78 and 81 for the binary and PU QB, binary and PU WV-3LR, and binary and PU WV-3HR classifications respectively. Features deemed important consisted usually of GLCM sum average, GLCM sum variance, and mean and quantile features. Furthermore, features calculated from the NDVI, RGI and the near-infrared band in the QB and WV-3LR classifications, and features calculated from the red-edge band, NDVI and RGI followed by red, green, and near-infrared 1 bands in the WV-3HR classifications were over-represented in the final set of features. Nevertheless, some features from all bands were included and all types of features were included in the final classifications.

In OCSVM classifications, the optimal value for σ varied between 1×10^{-4} and 1×10^{-3} and for v between 0.01 and 0.15. In BSVM classifications, the optimal value for σ varied between 1×10^{-7} and 1×10^{-3}, for C between 1×10^{3} and 1×10^{5}, and for $C+$ between 1×10^{0} and 1×10^{5}, but in most of the classifications, $C+$ was 1×10^{0}. In all BSVM classifications, F-scores close to the best F-score were obtained with several parameter combinations including higher values for $C+$. In RF classifications, the optimal W varied between 1×10^{5} and 10×10^{5}.

3.2.Classification and change detection

When looking at the cross-validation or out-of-bag performance of the training data, F-scores for the different classifications varied between 0.77 and 0.96, with the highest F-scores for ROF and lowest for OCSVM (Table 1). However, the F-scores were notably smaller (0.22–0.57) when independent test data were used (Table 1). In the six comparisons based on test data, RF had the highest F-score in five cases. The

394 classification accuracies were the highest for WV-3HR classifications and the lowest for
395 WV-3LR classifications, while the situation was the opposite in the evaluation based on
396 cross-validated training data.
397 TABLE 1 APPROXIMATELY HERE
398 There was considerable variation in the classifier performance based on
399 independent test data. All classifiers performed well in some classifications but not as
400 well in some other classifications (Tables 2 – 4). User accuracies varied between 22%
401 and 64% and producer accuracies between 22% and 60%. OCSVM had high producer
402 accuracies and low user accuracies especially in QB classifications and suffered thus
403 from over-classification in some cases. BSVM was more prone to under-classification
404 but it performed well especially in the PU WV-3 classifications. RF performed the most
405 robustly of the classifiers but had some under-prediction problems in the WV-3LR
406 classifications. ROF performed quite poorly in the WV-3LR classifications and under-
407 predicted in some classifications but had also high classification accuracies in other
408 classifications. In visual interpretation, the performance of the classifiers varied, and all
409 classifiers produced visually acceptable classifications (Figure 4).
410 TABLE 2 APPROXIMATELY HERE
411 TABLE 3 APPROXIMATELY HERE
412 TABLE 4 APPROXIMATELY HERE
413 FIGURE 4 APPROXIMATELY HERE
414 According to the consensus of five classifiers and average estimates, the number
415 of bare peat areas as well as their areal coverage decreased by 21%–26% between 2007
416 and 2015 (Table 5). Nevertheless, there was some under-prediction in many of the WV-
417 3LR classifications (Table 3). WV-3HR gave a similar area estimate and higher patch
418 number estimate than WV-3LR of on average 5.3 ha (4.5 ha – 7.8 ha) in 554 patches

(319–951), which could be judged as a better estimate due to higher classification accuracy. Overall, the areal coverage of bare peat was around 0.5% of the study area and 1.0% of the peat plateau.

TABLE 5 APPROXIMATELY HERE

4. Discussion

According to our results, there was considerable variation in classifier performance. Although the *F*-scores were high when evaluated using cross-validated training data, they were low when using independent test data. Actually, the cross-validated *F*-scores were similar to the *F*-scores or accuracies received in some other studies using one-class or binary classifications in relatively simple classification tasks in urban or agricultural areas (Mack et al. 2014; Ao et al. 2017; Deng et al. 2018). On the other hand, the classifier performance based on independent test data was comparable to earlier classifier performance in more challenging classification tasks in natural environments (Baldeck and Asner 2015; Stenzel et al. 2017), although higher classification accuracies have also been obtained in natural environments (Baldeck et al. 2015; Mack et al. 2016). The reason for the high disparity between the classification accuracies can be related to the fact that training data and test data were derived from different areas. Although the distance between the areas is only 1 km – 2 km, the composition of land cover classes is to some extent different. Another possibility would have been to use part of the data from both areas for training and part of the data for testing. However, in that case, we could not have used the drone image and manually delineated bare peat areas in the test data, and we could only have used object-based metrics when evaluating classifier performance. We judge the classifier performance based on the test data as a more reliable estimate, because independent test data are usually recommended (Mack and Waske 2017; Liu et al. 2018). Later in this section, we mostly

444 discuss the classifier performance against the independent test data.

445 When evaluated using the independent test data, RF performed most robustly in
446 classifying bare peat areas, and its performance was often slightly better than OCSVM
447 or BSVM. Quite surprisingly, the performance of OCSVM was approximately on par
448 with BSVM when evaluated using independent test data, which is in contrast to
449 previous research (Li et al. 2011; Baldeck and Asner 2015; Baldeck et al. 2015; Barbosa
450 et al. 2016; Ao et al. 2017; Mack and Waske 2017; Stenzel et al. 2017; Deng et al.
451 2018). However, the performance of OCSVM was lower compared with other
452 classifiers when evaluated by the cross-validated training data. It is also notable that
453 OCSVM produced different classification results in binary and PU settings, although it
454 uses only positive data for training. These dissimilarities were probably caused by
455 different final parameter sets.

456 In previous research, there has been mixed evidence as to whether there are
457 differences in classifier performance when comparing binary and PU settings. In one
458 comparison, higher classification accuracy has been achieved with binary SVM
459 compared with biased SVM (Mack et al. 2016); in a second comparison, the situation
460 was the opposite (Baldeck et al. 2015); and in a third comparison, the difference was
461 small between the two classifiers (Baldeck and Asner 2015). According to our results,
462 there was no clear trend in classifier performance between binary and PU settings.
463 Although there were divergences in single classifiers, equally high classification
464 accuracies could be achieved in both settings. Overall, we judge that all tested
465 classifiers can be used both in binary and PU settings, and especially RF showed
466 promising performance. Therefore, the use of the PU classification setting for all tested
467 classifiers can be recommended for future studies as this setting requires training data
468 only for the positive class.

Considering change detection, our results indicate that there has been a drastic decrease in bare peat areas during the last decade. Based on all comparisons, the areal coverage and number of bare peat surfaces is decreasing in the area. The detected decrease can be related many factors which include an actual decrease in the number and size of patches, a decrease in the number of patches that could be clearly identified (i.e. segmented and classified), uncertainties in classifications such as under-prediction in WV-3LR classifications, and goodness of the training data. Actually, already in training data of the WV-3 classifications, the areal coverage of bare peat was 16% lower than in the training data of the QB classification. Overall, the decrease is supported by our field observations. We have observed that vegetation density and area covered by vegetation seem to be at least to some extent increased around bare peat surfaces, and there seem to be no new bare peat areas. The decrease is also supported by the recent evidence that current bare peat areas are remnants of larger zones of barren land (Väliranta et al. 2018) and by the observations that Arctic areas are greening (Tape et al. 2006; Elmendorf et al. 2012; Myers-Smith et al. 2015; Yu et al. 2017). On the other hand, there has been little change in thermokarst lake coverage near the study site between the mid-1970s and mid-2000s (Sannel and Kuhry 2011), which might mean that there has been little change in the landscape dynamics also in our study area. It might also be possible that bare areas would at some time also increase, at least temporarily, if a warming climate creates quick disturbances to vegetation due to changes in permafrost conditions in the area.

In earlier research, change detection in permafrost environments has been concentrated on coarse-scale patterns, while fine-scale changes have been mostly tracked with visual interpretation (Jorgenson and Grosse 2016). Nevertheless, as visual interpretation is labour demanding, we urge for the development of an efficient and

automatic change tracking method for VHSR images. Furthermore, images could be fused with repeat lidar analysis, which has been successfully used in tracking changes (Jorgenson and Grosse 2016). Another possible avenue could be testing the resolution requirements in tracking changes; in other words, what kind of changes that can be detected using different resolution data sets could be assessed. In addition to methodological developments, more validation studies with thorough field assessments are needed to track fine-scale changes in permafrost environments.

There is a one-month difference in the timing of the QB and WV-3 images; therefore, phenology might affect the classification of bare peat areas. Although there are no phenological changes in non-vegetated surfaces such as bare peat, phenology affects the interpretation of other land cover classes and vegetation types (Anderson et al. 2016; Wang et al. 2016; Juutinen et al. 2017). This might then have a minor influence on the classification of bare peat.

Our results show that the higher spectral and spatial resolution of WV-3HR classifications gave higher classification accuracies than QB or WV-3LR classifications. Also in earlier studies, it was found that WV-3 performs better than QB in various environments (Novack et al. 2011; McCarthy and Halls 2014). Overall, WV-3HR classifications were the most reliable in our case and showed that high spectral and spatial resolution is needed in the mapping of bare peat areas.

5. Conclusion

To sum up, both binary and PU classification methods were shown to be appropriate in mapping bare peat surfaces, but there were significant differences in classifier predictions of bare peat areas. RF was the most robust classifier in the comparison, but differences between classifiers were generally small. More research is needed in evaluating the performance of RF and ROF in various one-class tasks and also in

evaluating whether single classes are better mapped in one-class than multi-class settings. The performance of one-class methods in tracking changes was mixed and more research should be carried out in the development of automated change detection methods for specific single habitat types. Finally, applying one-class methods in mapping the extent, distribution and dynamics of single habitat and land cover types at a fine scale over a large spatial extent is a promising research avenue as low-cost VHSR satellite images are becoming more widely available.

Acknowledgements: This research was supported by the Academy of Finland (projects 291736 and 296423).

References

Adam, Elhadi, Onisimo Mutanga, John Odindi, and Elfatih M Abdel-Rahman. 2014. "Land-use/cover classification in a heterogeneous coastal landscape using RapidEye imagery: evaluating the performance of random forest and support vector machines classifiers." *International journal of remote sensing* 35 (10):3440-58.

Amici, V., M. Marcantonio, N. La Porta, and D. Rocchini. 2017. "A multi-temporal approach in MaxEnt modelling: A new frontier for land use/land cover change detection." *Ecological Informatics* 40:40-9. doi: 10.1016/j.ecoinf.2017.04.005.

Anderson, H. B., L. Nilsen, H. Tommervik, S. R. Karlsen, S. Nagai, and E. J. Cooper. 2016. "Using ordinary digital cameras in place of near-infrared sensors to derive vegetation indices for phenology studies of high arctic vegetation." *Remote Sensing* 8 (10). doi: 10.3390/rs8100847.

Andrew, M. E., and J. M. Shephard. 2017. "Semi-automated detection of eagle nests: an application of very high-resolution image data and advanced image analyses to wildlife surveys." *Remote Sensing in Ecology and Conservation* 3 (2):66-80. doi: 10.1002/rse2.38.

Ao, Z., Y. Su, W. Li, Q. Guo, and J. Zhang. 2017. "One-class classification of airborne LiDAR data in urban areas using a presence and background learning algorithm." *Remote Sensing* 9 (10). doi: 10.3390/rs9101001.

Araya-López, R. A., J. Lopatin, F. E. Fassnacht, and H. J. Hernández. 2018. "Monitoring Andean high altitude wetlands in central Chile with seasonal optical data: A comparison between Worldview-2 and Sentinel-2 imagery." *ISPRS Journal of Photogrammetry and Remote Sensing*. doi: 10.1016/j.isprsjprs.2018.04.001.

Baldeck, C. A., and G. P. Asner. 2015. "Single-Species Detection With Airborne Imaging Spectroscopy Data: A Comparison of Support Vector Techniques." *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 8 (6):2501-12. doi: 10.1109/jstars.2014.2346475.

Baldeck, C. A., G. P. Asner, R. E. Martin, C. B. Anderson, D. E. Knapp, J. R. Kellner, and S. J. Wright. 2015. "Operational Tree Species Mapping in a Diverse

559 Tropical Forest with Airborne Imaging Spectroscopy." *Plos One* 10 (7). doi:
560 10.1371/journal.pone.0118403.
561 Ballings, M., and D. Van der Poel. 2015. "rotationForest: Fit and Deploy Rotation
562 Forest Models." http://CRAN.R-project.org/package=rotationForest.
563 Barbosa, J. M., G. P. Asner, R. E. Martin, C. A. Baldeck, F. Hughes, and T. Johnson.
564 2016. "Determining Subcanopy Psidium cattleianum Invasion in Hawaiian
565 Forests Using Imaging Spectroscopy." *Remote Sensing* 8 (1). doi:
566 10.3390/rs8010033.
567 Bartsch, A., A. Hofler, C. Kroisleitner, and A. M. Trofaier. 2016. "Land Cover Mapping
568 in Northern High Latitude Permafrost Regions with Satellite Data:
569 Achievements and Remaining Challenges." *Remote Sensing* 8 (12). doi:
570 10.3390/rs8120979.
571 Beck, I., R. Ludwig, M. Bernier, E. Lévesque, and J. Boike. 2015. "Assessing
572 Permafrost Degradation and Land Cover Changes (1986-2009) using Remote
573 Sensing Data over Umiujaq, Sub-Arctic Québec." *Permafrost and Periglacial*
574 *Processes* 26 (2):129-41. doi: 10.1002/ppp.1839.
575 Belgiu, M., and L. Dragut. 2016. "Random forest in remote sensing: A review of
576 applications and future directions." *ISPRS Journal of Photogrammetry and*
577 *Remote Sensing* 114:24-31. doi: 10.1016/j.isprsjprs.2016.01.011.
578 Belshe, E. F., E. A. G. Schuur, and G. Grosse. 2013. "Quantification of upland
579 thermokarst features with high resolution remote sensing." *Environmental*
580 *Research Letters* 8 (3). doi: 10.1088/1748-9326/8/3/035016.
581 Benz, U. C., P. Hofmann, G. Willhauck, I. Lingenfelder, and M. Heynen. 2004. "Multi-
582 resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready
583 information." *ISPRS Journal of Photogrammetry and Remote Sensing* 58 (3-
584 4):239-58. doi: 10.1016/j.isprsjprs.2003.10.002.
585 Blaschke, Thomas, Geoffrey J Hay, Maggi Kelly, Stefan Lang, Peter Hofmann,
586 Elisabeth Addink, Raul Queiroz Feitosa, Freek van der Meer, Harald van der
587 Werff, and Frieke van Coillie. 2014. "Geographic object-based image analysis–
588 towards a new paradigm." *ISPRS Journal of Photogrammetry and Remote*
589 *Sensing* 87:180-91.
590 Bockheim, J. G., and C. Tarnocai. 1998. "Recognition of cryoturbation for classifying
591 permafrost-affected soils." *Geoderma* 81 (3-4):281-93. doi: 10.1016/s0016-
592 7061(97)00115-8.
593 Borge, A. F., S. Westermann, I. Solheim, and B. Etzelmuller. 2017. "Strong degradation
594 of palsas and peat plateaus in northern Norway during the last 60 years."
595 *Cryosphere* 11 (1):1-16. doi: 10.5194/tc-11-1-2017.
596 Boyd, D. S., C. Sanchez-Hernandez, and G. M. Foody. 2006. "Mapping a specific class
597 for priority habitats monitoring from satellite sensor data." *International journal*
598 *of remote sensing* 27 (13):2631-44. doi: 10.1080/01431160600554348.
599 Breiman, L. 2001. "Random forests." *Machine Learning* 45 (1):5-32. doi:
600 10.1023/a:1010933404324.
601 Chavez, P. S. 1988. "An improved dark-object subtraction technique for atmospheric
602 scattering correction of multispectral data." *Remote Sensing of Environment* 24
603 (3):459-79. doi: 10.1016/0034-4257(88)90019-3.
604 Chen, C., A. Liaw, and L. Breiman. 2004. "Using Random Forest to Learn Imbalanced
605 Data." In *Statistics Technical Reports*. Berkeley, CA, USA: University of
606 California Berkeley.

Chen, X. H., D. M. Yin, J. Chen, and X. Cao. 2016. "Effect of training strategy for positive and unlabelled learning classification: test on Landsat imagery." *Remote Sensing Letters* 7 (11):1063-72. doi: 10.1080/2150704x.2016.1217437.

Chignell, S. M., M. W. Luizza, S. Skach, N. E. Young, and P. H. Evangelista. 2018. "An integrative modeling approach to mapping wetlands and riparian areas in a heterogeneous Rocky Mountain watershed." *Remote Sensing in Ecology and Conservation* 4 (2):150-65. doi: 10.1002/rse2.63.

Coops, N. C., M. Johnson, M. A. Wulder, and J. C. White. 2006. "Assessment of QuickBird high spatial resolution imagery to detect red attack damage due to mountain pine beetle infestation." *Remote Sensing of Environment* 103 (1):67-80. doi: 10.1016/j.rse.2006.03.012.

Deng, X., W. Li, X. Liu, Q. Guo, and S. Newsam. 2018. "One-class remote sensing classification: One-class vs. Binary classifiers." *International journal of remote sensing* 39 (6):1890-910. doi: 10.1080/01431161.2017.1416697.

Du, P. J., A. Samat, B. Waske, S. C. Liu, and Z. H. Li. 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. doi: 10.1016/j.isprsjprs.2015.03.002.

Duro, Dennis C, Steven E Franklin, and Monique G Dubé. 2012. "A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery." *Remote Sensing of Environment* 118:259-72.

Elith, J., S. J. Phillips, T. Hastie, M. Dudík, Y. E. Chee, and C. J. Yates. 2011. "A statistical explanation of MaxEnt for ecologists." *Diversity and Distributions* 17 (1):43-57. doi: 10.1111/j.1472-4642.2010.00725.x.

Elmendorf, S. C., G. H. R. Henry, R. D. Hollister, R. G. Bjork, N. Boulanger-Lapointe, E. J. Cooper, J. H. C. Cornelissen, et al. 2012. "Plot-scale evidence of tundra vegetation change and links to recent summer warming." *Nature Climate Change* 2 (6):453-7. doi: 10.1038/nclimate1465.

Haralick, R. M., I. Dinstein, and K. Shanmugam. 1973. "Textural Features for Image Classification." *IEEE Transactions on Systems, Man and Cybernetics* SMC-3 (6):610-21. doi: 10.1109/TSMC.1973.4309314.

Hugelius, G., J. Routh, P. Kuhry, and P. Crill. 2012. "Mapping the degree of decomposition and thaw remobilization potential of soil organic matter in discontinuous permafrost terrain." *Journal of Geophysical Research: Biogeosciences* 117 (2). doi: 10.1029/2011JG001873.

Jorgenson, M. T., and G. Grosse. 2016. "Remote Sensing of Landscape Change in Permafrost Regions." *Permafrost and Periglacial Processes* 27 (4):324-38. doi: 10.1002/ppp.1914.

Juutinen, S., T. Virtanen, V. Kondratyev, T. Laurila, M. Linkosalmi, J. Mikola, J. Nyman, A. Räsänen, J-P. Tuovinen, and M. Aurela. 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. doi: https://doi.org/10.1088/1748-9326/aa7f85.

Kavzoglu, T., and I. Colkesen. 2013. "An assessment of the effectiveness of a rotation forest ensemble for land-use and land-cover mapping." *International journal of remote sensing* 34 (12):4224-41. doi: 10.1080/01431161.2013.774099.

Kursa, M. B., and W. R. Rudnicki. 2010. "Feature Selection with the Boruta Package." *Journal of Statistical Software* 36 (11):1-13.

Laidler, G. J., and P. Treitz. 2003. "Biophysical remote sensing of arctic environments." *Progress in Physical Geography* 27 (1):44-68. doi: 10.1191/0309133303pp358ra.

Li, J., M. Tran, and J. Siwabessy. 2016. "Selecting Optimal Random Forest Predictive Models: A Case Study on Predicting the Spatial Distribution of Seabed Hardness." *Plos One* 11 (2). doi: 10.1371/journal.pone.0149089.

Li, P., and H. Xu. 2010. "Land-cover change detection using one-class support vector machine." *Photogrammetric Engineering and Remote Sensing* 76 (3):255-63.

Li, W., Q. Guo, and C. Elkan. 2011. "A positive and unlabeled learning algorithm for one-class classification of remote-sensing data." *Ieee Transactions on Geoscience and Remote Sensing* 49 (2):717-25. doi: 10.1109/TGRS.2010.2058578.

Li, W. K., and Q. H. Guo. 2010. "A maximum entropy approach to one-class classification of remote sensing imagery." *International journal of remote sensing* 31 (8):2227-35. doi: 10.1080/01431161003702245.

Li, X. J., W. T. Chen, X. W. Cheng, and L. Z. Wang. 2016. "A Comparison of Machine Learning Algorithms for Mapping of Complex Surface-Mined and Agricultural Landscapes Using ZiYuan-3 Stereo Satellite Imagery." *Remote Sensing* 8 (6). doi: 10.3390/rs8060514.

Liaw, A., and M. Wiener. 2002. "Classification and Regression by randomForest." *R News* 2 (3):18-22.

Lin, J., X. Liu, K. Li, and X. Li. 2014. "A maximum entropy method to extract urban land by combining MODIS reflectance, MODIS NDVI, and DMSP-OLS data." *International journal of remote sensing* 35 (18):6708-27. doi: 10.1080/01431161.2014.960623.

Liu, B., Y. Dai, X. Li, W. S. Lee, and P. S. Yu. 2003. Building Text Classifiers Using Positive and Unlabeled Examples. Paper presented at the Proceedings of the Third IEEE International Conference on Data Mining (ICDM '03).

Liu, R., W. Li, X. Liu, X. Lu, T. Li, and Q. Guo. 2018. "An Ensemble of Classifiers Based on Positive and Unlabeled Data in One-Class Remote Sensing Classification." *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 11 (2):572-84. doi: 10.1109/JSTARS.2017.2789213.

Low, F., U. Michel, S. Dech, and C. Conrad. 2013. "Impact of feature selection on the accuracy and spatial uncertainty of per-field crop classification using Support Vector Machines." *ISPRS Journal of Photogrammetry and Remote Sensing* 85:102-19. doi: 10.1016/j.isprsjprs.2013.08.007.

Mack, B., R. Roscher, S. Stenzel, H. Feilhauer, S. Schmidtlein, and B. Waske. 2016. "Mapping raised bogs with an iterative one-class classification approach." *ISPRS Journal of Photogrammetry and Remote Sensing* 120:53-64. doi: 10.1016/j.isprsjprs.2016.07.008.

Mack, B., R. Roscher, and B. Waske. 2014. "Can I Trust My One-Class Classification?" *Remote Sensing* 6 (9):8779-802. doi: 10.3390/rs6098779.

Mack, B., and B. Waske. 2017. "In-depth comparisons of MaxEnt, biased SVM and one-class SVM for one-class classification of remote sensing data." *Remote Sensing Letters* 8 (3):290-9. doi: 10.1080/2150704x.2016.1265689.

Marushchak, M. E., I. Kiepe, C. Biasi, V. Elsakov, T. Friborg, T. Johansson, H. Soegaard, T. Virtanen, and P. J. Martikainen. 2013. "Carbon dioxide balance of subarctic tundra from plot to regional scales." *Biogeosciences* 10 (1):437-52. doi: 10.5194/bg-10-437-2013.

Marushchak, M. E., A. Pitkämäki, H. Koponen, C. Biasi, M. Seppälä, and P. J. Martikainen. 2011. "Hot spots for nitrous oxide emissions found in different types of permafrost peatlands." *Global Change Biology* 17 (8):2601-14. doi: 10.1111/j.1365-2486.2011.02442.x.

McCarthy, M. J., and J. N. Halls. 2014. "Habitat mapping and change assessment of coastal environments: An examination of worldview-2, quickbird, and ikonos satellite imagery and airborne lidar for mapping barrier island habitats." *ISPRS International Journal of Geo-Information* 3 (1):297-325. doi: 10.3390/ijgi3010297.

McFeeters, S. K. 1996. "The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features." *International journal of remote sensing* 17 (7):1425-32. doi: 10.1080/01431169608948714.

Mountrakis, G., J. Im, and C. Ogole. 2011. "Support vector machines in remote sensing: A review." *ISPRS Journal of Photogrammetry and Remote Sensing* 66 (3):247-59. doi: 10.1016/j.isprsjprs.2010.11.001.

Myers-Smith, I. H., S. C. Elmendorf, P. S. A. Beck, M. Wilmking, M. Hallinger, D. Blok, K. D. Tape, et al. 2015. "Climate sensitivity of shrub growth across the tundra biome." *Nature Climate Change* 5 (9):887-+. doi: 10.1038/nclimate2697.

Novack, T., T. Esch, H. Kux, and U. Stilla. 2011. "Machine learning comparison between WorldView-2 and QuickBird-2-simulated imagery regarding object-based urban land cover classification." *Remote Sensing* 3 (10):2263-82. doi: 10.3390/rs3102263.

Noviello, M., B. Cafarelli, C. Calculli, A. Sarris, and P. Mairota. 2018. "Investigating the distribution of archaeological sites: Multiparametric vs probability models and potentials for remote sensing data." *Applied Geography* 95:34-44. doi: 10.1016/j.apgeog.2018.04.005.

Obu, J., H. Lantuit, I. Myers-Smith, B. Heim, J. Wolter, and M. Fritz. 2017. "Effect of terrain characteristics on soil organic carbon and total nitrogen stocks in soils of Herschel Island, western Canadian Arctic." *Permafrost and Periglacial Processes* 28 (1):92-107. doi: 10.1002/ppp.1881.

Pacifici, F., M. Chini, and W. J. Emery. 2009. "A neural network approach using multi-scale textural metrics from very high-resolution panchromatic imagery for urban land-use classification." *Remote Sensing of Environment* 113 (6):1276-92. doi: 10.1016/j.rse.2009.02.014.

Pau, G., F. Fuchs, O. Sklyar, M. Boutros, and W. Huber. 2010. "EBImage-an R package for image processing with applications to cellular phenotypes." *Bioinformatics* 26 (7):979-81. doi: 10.1093/bioinformatics/btq046.

Peterson, R. A., and W. B. Krantz. 2003. "A mechanism for differential frost heave and its implications for patterned-ground formation." *Journal of Glaciology* 49 (164):69-80. doi: 10.3189/172756503781830854.

Pflugmacher, D., O. N. Krankina, W. B. Cohen, M. A. Friedl, D. Sulla-Menashe, R. E. Kennedy, P. Nelson, et al. 2011. "Comparison and assessment of coarse resolution land cover maps for Northern Eurasia." *Remote Sensing of Environment* 115 (12):3539-53. doi: 10.1016/j.rse.2011.08.016.

Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006. "Maximum entropy modeling of species geographic distributions." *Ecological Modelling* 190 (3-4):231-59. doi: 10.1016/j.ecolmodel.2005.03.026.

Poona, N. K., A. van Niekerk, R. L. Nadel, and R. Ismail. 2016. "Random Forest (RF) Wrappers for Waveband Selection and Classification of Hyperspectral Data." *Applied Spectroscopy* 70 (2):322-33. doi: 10.1177/0003702815620545.

R Core Team. 2015. "R: A language and environment for statistical computing." R Foundation for Statistical Computing. https://www.R-project.org/.

Rasanen, A., M. Kuitunen, E. Tomppo, and A. Lensu. 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* 94:169-82. doi: 10.1016/j.isprsjprs.2014.05.003.

Repo, M. E., S. Susiluoto, S. E. Lind, S. Jokinen, V. Elsakov, C. Biasi, T. Virtanen, and P. J. Martikainen. 2009. "Large N2O emissions from cryoturbated peat soil in tundra." *Nature Geoscience* 2 (3):189-92. doi: 10.1038/ngeo434.

Rodriguez-Galiano, Victor Francisco, Bardan Ghimire, John Rogan, Mario Chica-Olmo, and Juan Pedro Rigol-Sanchez. 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.

Rodriguez, J. J., and L. I. Kuncheva. 2006. "Rotation forest: A new classifier ensemble method." *Ieee Transactions on Pattern Analysis and Machine Intelligence* 28 (10):1619-30. doi: 10.1109/tpami.2006.211.

Rouse, J. W. Jr., R. H. Haas, J. A. Schell, and D. W. Deering. 1973. "Monitoring vegetation systems in the Great Plains with ERTS." In *Third Earth Resources Technology Satellite-1 Symposium*, 309-17. Washington, DC: NASA.

Räsänen, Aleksi, Antti Rusanen, Markku Kuitunen, and Anssi Lensu. 2013. "What makes segmentation good? A case study in boreal forest habitat mapping." *International journal of remote sensing* 34 (23):8603-27.

Samsudin, S. H., H. Z. M. Shafri, A. Hamedianfar, and S. Mansor. 2015. "Spectral feature selection and classification of roofing materials using field spectroscopy data." *Journal of Applied Remote Sensing* 9. doi: 10.1117/1.jrs.9.095079.

Sanchez-Hernandez, C., D. S. Boyd, and G. M. Foody. 2007. "Mapping specific habitats from remotely sensed imagery: Support vector machine and support vector data description based classification of coastal saltmarsh habitats." *Ecological Informatics* 2 (2):83-8. doi: 10.1016/j.ecoinf.2007.04.003.

Sannel, A. B. K., and P. Kuhry. 2011. "Warming-induced destabilization of peat plateau/thermokarst lake complexes." *Journal of Geophysical Research: Biogeosciences* 116 (3). doi: 10.1029/2010JG001635.

Schölkopf, B., R. Williamson, A. Smola, J. Shawe-Taylor, and J. Piatt. 2000. Support vector method for novelty detection. Paper presented at the Advances in Neural Information Processing Systems.

Segal, R. A., T. C. Lantz, and S. V. Kokelj. 2016. "Acceleration of thaw slump activity in glaciated landscapes of the Western Canadian Arctic." *Environmental Research Letters* 11 (3). doi: 10.1088/1748-9326/11/3/034025.

Seppälä, M. 2003. "Surface abrasion of palsas by wind action in Finnish Lapland." *Geomorphology* 52 (1-2):141-8. doi: 10.1016/S0169-555X(02)00254-4.

Singh, Ashbindu. 1989. "Review article digital change detection techniques using remotely-sensed data." *International journal of remote sensing* 10 (6):989-1003.

Song, B. Q., P. J. Li, J. Li, and A. Plaza. 2016. "One-Class Classification of Remote Sensing Images Using Kernel Sparse Representation." *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 9 (4):1613-23. doi: 10.1109/jstars.2015.2508285.

Stenzel, S., F. E. Fassnacht, B. Mack, and S. Schmidtlein. 2017. "Identification of high nature value grassland with remote sensing and minimal field data." *Ecological Indicators* 74:28-38. doi: 10.1016/j.ecolind.2016.11.005.

Stenzel, S., H. Feilhauer, B. Mack, A. Metz, and S. Schmidtlein. 2014. "Remote sensing of scattered Natura 2000 habitats using a one-class classifier." *International Journal of Applied Earth Observation and Geoinformation* 33:211-7. doi: 10.1016/j.jag.2014.05.012.

Stow, D. A., A. Hope, D. McGuire, D. Verbyla, J. Gamon, F. Huemmrich, S. Houston, et al. 2004. "Remote sensing of vegetation and land-cover change in Arctic Tundra Ecosystems." *Remote Sensing of Environment* 89 (3):281-308. doi: 10.1016/j.rse.2003.10.018.

Tape, K., M. Sturm, and C. Racine. 2006. "The evidence for shrub expansion in Northern Alaska and the Pan-Arctic." *Global Change Biology* 12 (4):686-702. doi: 10.1111/j.1365-2486.2006.01128.x.

Tewkesbury, Andrew P, Alexis J Comber, Nicholas J Tate, Alistair Lamb, and Peter F Fisher. 2015. "A critical synthesis of remotely sensed optical image change detection techniques." *Remote Sensing of Environment* 160:1-14.

Wang, P., L. Mommer, J. van Ruijven, F. Berendse, T. C. Maximov, and Mmpd Heijmans. 2016. "Seasonal changes and vertical distribution of root standing biomass of graminoids and shrubs at a Siberian tundra site." *Plant and Soil* 407 (1-2):55-65. doi: 10.1007/s11104-016-2858-5.

Vapnik, V. 1982. *Estimation of Dependences Based on Empirical Data*. Vol. 40. New York: Springer-Verlag.

Vert, J.-P., K. Tsuda, and B. Schölkopf. 2004. "A primer on kernel methods." In *Kernel Methods in Computational Biology*, edited by B. Schölkopf, K. Tsuda and J.-P. Vert, 35-70. Cambridge, MA, USA: MIT Press.

Westermann, S., C. R. Duguay, G. Grosse, and A. Kääb. 2014. "Remote sensing of permafrost and frozen ground." In *Remote Sensing of the Cryosphere*, 307-44.

Virtanen, T., and M. Ek. 2014. "The fragmented nature of tundra landscape." *International Journal of Applied Earth Observation and Geoinformation* 27:4-12. doi: 10.1016/j.jag.2013.05.010.

Voigt, C., R. E. Lamprecht, M. E. Marushchak, S. E. Lind, A. Novakovskiy, M. Aurela, P. J. Martikainen, and C. Biasi. 2016. "Warming of subarctic tundra increases emissions of all three important greenhouse gases - carbon dioxide, methane, and nitrous oxide." *Global Change Biology*. doi: 10.1111/gcb.13563.

Väliranta, M., M. Marushchak, J.-P. Tuovinen, A. Lohila, C. Biasi, T. Ronkainen, H. Zhang, et al. 2018. "Increased climate forcing of high-latitude peatlands due to permafrost initiation." *Submitted manuscript*.

Xu, X., X. Liu, X. Li, Q. Xin, Y. Chen, Q. Shi, and B. Ai. 2016. "Global snow cover estimation with Microwave Brightness Temperature measurements and one-class in situ observations." *Remote Sensing of Environment* 182:227-51. doi: 10.1016/j.rse.2016.05.010.

Yu, Q., H. Epstein, R. Engstrom, and D. Walker. 2017. "Circumpolar arctic tundra biomass and productivity dynamics in response to projected climate change and herbivory." *Global Change Biology*. doi: 10.1111/gcb.13632.

Table 1. *F*-scores of the different classifiers using the parameter combinations yielding highest *F*-score values. Values are calculated based on two different evaluations: (1) out-of-bag (RF) or tenfold cross-validation (other classifiers) estimates using the training data only, or (2) using independent test data. The highest *F*-scores for each image and classification setting are marked in bold.

Evaluation	Classification	Binary				Positive and unlabelled			
		OCSVM	BSVM	RF	ROF	OCSVM	BSVM	RF	ROF
Training data cross-validation	QB	0.85	0.89	0.88	**0.94**	0.90	0.93	0.94	**0.96**
	WV-3LR	0.87	0.92	0.92	**0.93**	0.95	**0.96**	0.96	0.96
	WV-3HR	0.77	0.86	0.86	**0.91**	0.88	0.90	0.90	**0.95**
Independent test data	QB	0.48	0.39	**0.48**	0.40	0.44	0.45	**0.46**	0.42
	WV-3LR	0.32	0.31	**0.38**	0.22	0.42	**0.45**	0.42	0.34
	WV-3HR	0.42	0.48	**0.57**	0.52	0.43	**0.56**	0.48	0.53

Table 2. Classification accuracy results for the QB classifications. The best performances for the binary and PU settings are marked in bold.

Performance metric	Binary				PU			
	OCSVM	BSVM	RF	ROF	OCSVM	BSVM	RF	ROF
User's accuracy (%)	36.9	**46.7**	43.7	39.2	34.8	42.0	46.8	**58.9**
Segment that is 50% correct (%)	41.7	**63.2**	54.7	57.4	39.2	54.5	63.0	**84.6**
Segment that is 10% correct (%)	80.6	**92.1**	88.7	68.1	74.7	81.8	91.3	**100.0**
Producer's accuracy (%)	**60.0**	42.4	53.7	43.4	**59.8**	21.5	50.0	34.4
Reference area that is 50% mapped (%)	**56.2**	37.1	49.4	31.5	**56.2**	18.0	42.7	25.8
Reference area that is 10% mapped (%)	**65.2**	40.4	52.8	34.8	**65.2**	19.1	46.1	29.2
Number of segment classified as bare peat	**72**	38	53	47	**79**	22	46	26
Number of bare peat area in the reference	89	89	89	89	89	89	89	89
Mapped area (ha)	0.866	**0.472**	0.643	0.580	0.914	0.267	**0.558**	0.308
Bare peat area in the reference (ha)	0.521	0.521	0.521	0.521	0.521	0.521	0.521	0.521
Number of segment total	636	455	418	437	741	291	436	290
Total (ha)	9.102	6.686	5.939	6.231	10.030	4.364	6.070	4.311

Table 3. Classification accuracy results for the WV-3LR classifications. The best performances for the binary and PU settings are marked in bold.

Performance metric	Binary				PU			
	OCSVM	BSVM	RF	ROF	OCSVM	BSVM	RF	ROF
User's accuracy (%)	31.7	25.7	**54.2**	21.8	43.0	45.1	**45.8**	45.0
Segment that is 50% correct (%)	34.9	46.8	**58.6**	33.3	44.7	48.1	**48.9**	53.1
Segment that is 10% correct (%)	72.1	72.3	**82.8**	52.4	**83.0**	78.8	77.8	78.1
Producer's accuracy (%)	31.3	**37.7**	28.9	22.6	40.7	**45.1**	38.6	27.7
Reference area that is 50% mapped (%)	30.3	**33.7**	22.5	20.2	37.1	**39.3**	33.7	23.6
Reference area that is 10% mapped (%)	36.0	**38.2**	25.8	22.5	41.6	**44.9**	38.2	27.0
Number of segment classified as bare peat	43	**47**	29	42	47	**52**	45	32
Number of bare peat area in the reference	89	89	89	89	89	89	89	89
Mapped area (ha)	**0.520**	0.772	0.279	0.544	0.503	**0.526**	0.443	0.323
Bare peat area in the reference (ha)	0.521	0.521	0.521	0.521	0.521	0.521	0.521	0.521
Number of segment total	339	390	288	424	365	395	368	314
Total (ha)	4.674	7.942	4.069	5.905	4.746	5.469	5.369	4.306

863 Table 4. Classification accuracy results for the WV-3HR classifications. The best
864 performances for the binary and PU settings are marked in bold.

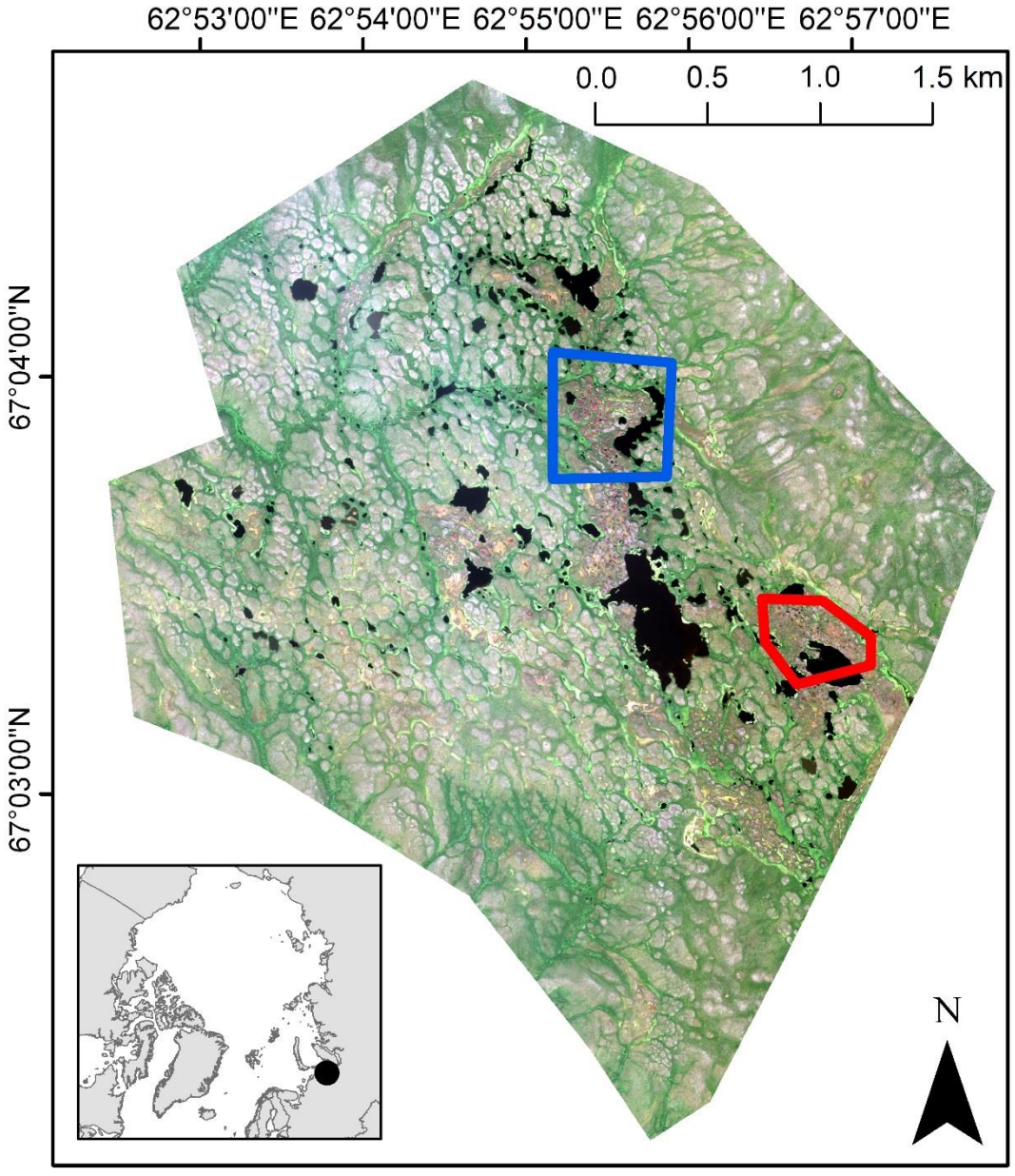
Performance metric	Binary				PU			
	OCSVM	BSVM	RF	ROF	OCSVM	BSVM	RF	ROF
User's accuracy (%)	42.9	**63.6**	56.8	58.2	33.7	**55.4**	47.2	49.1
Segment that is 50% correct (%)	50.7	**76.5**	66.3	69.6	40.2	**65.8**	54.1	61.4
Segment that is 10% correct (%)	80.0	88.2	**89.5**	84.1	68.2	**92.4**	83.5	86.4
Producer's accuracy (%)	40.6	38.4	**57.4**	47.3	**59.9**	55.7	49.2	56.8
Reference area that is 50% mapped (%)	43.8	32.6	**56.2**	41.6	**65.2**	53.9	53.9	58.4
Reference area that is 10% mapped (%)	55.1	46.1	**68.5**	52.8	**79.8**	62.9	67.4	66.3
Number of segment classified as bare peat	75	51	**86**	69	132	80	85	**88**
Number of bare peat area in the reference	89	89	89	89	89	89	89	89
Mapped area (ha)	0.498	0.318	**0.531**	0.427	0.930	**0.528**	0.550	0.606
Bare peat area in the reference (ha)	0.521	0.521	0.521	0.521	0.521	0.521	0.521	0.521
Number of segment total	513	319	454	359	951	604	685	549
Total (ha)	4.661	5.802	4.487	4.634	7.780	5.045	5.478	4.644

865

866

867 Table 5. Change detection metrics based on the QB and WV-3LR classifications.

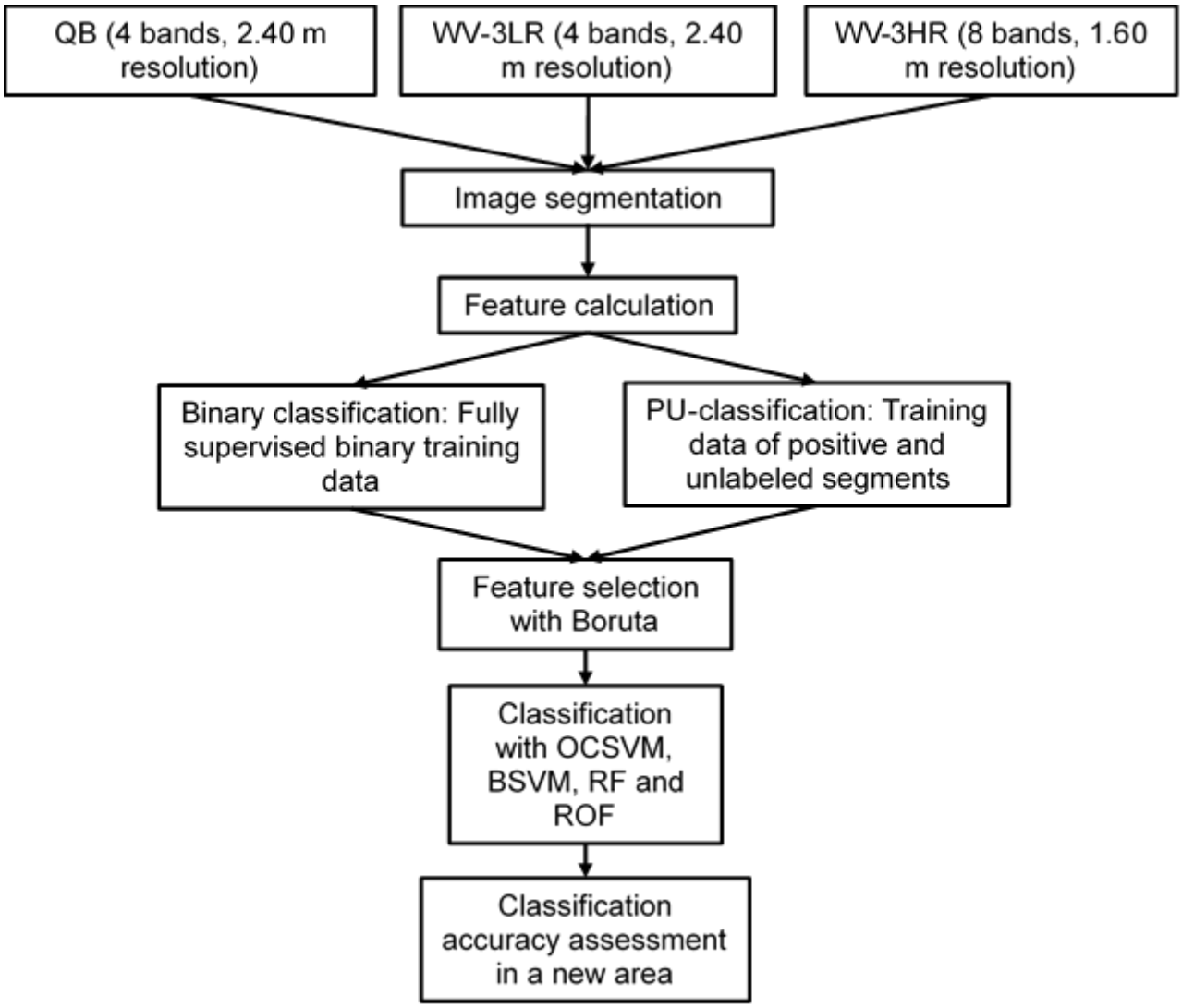
	QB		WV-3LR		Decrease (%)	
Metric	Area (ha)	Number of patches	Area (ha)	Number of patches	Area	Number of patches
Consensus of 7 classifiers	4.3	286	3.1	236	27.9	17.5
Consensus of 5 classifiers	5.7	401	4.4	318	22.8	20.7
Consensus of 3 classifiers	8.4	616	5.2	375	38.1	39.1
Lowest estimate	4.3	290	4.1	288	4.7	0.7
Average	7.0	487	5.3	360	24.3	26.1
Highest estimate	10.0	741	7.9	424	21.0	42.8



868

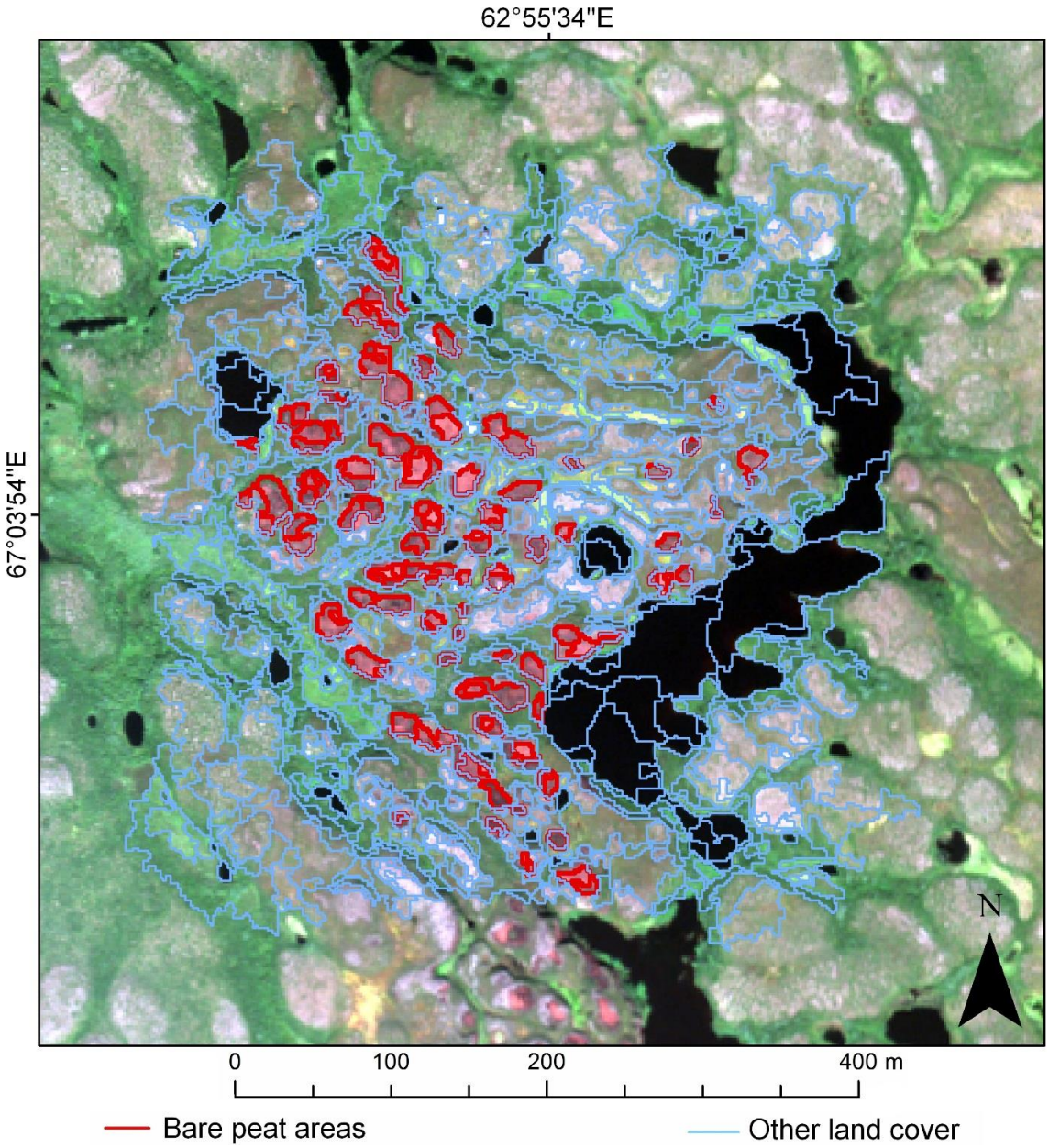
869 Figure 1. The study area and its location. The classifier training area is marked in blue

870 and the test area in red. Image: WorldView-3 ©DigitalGlobe.



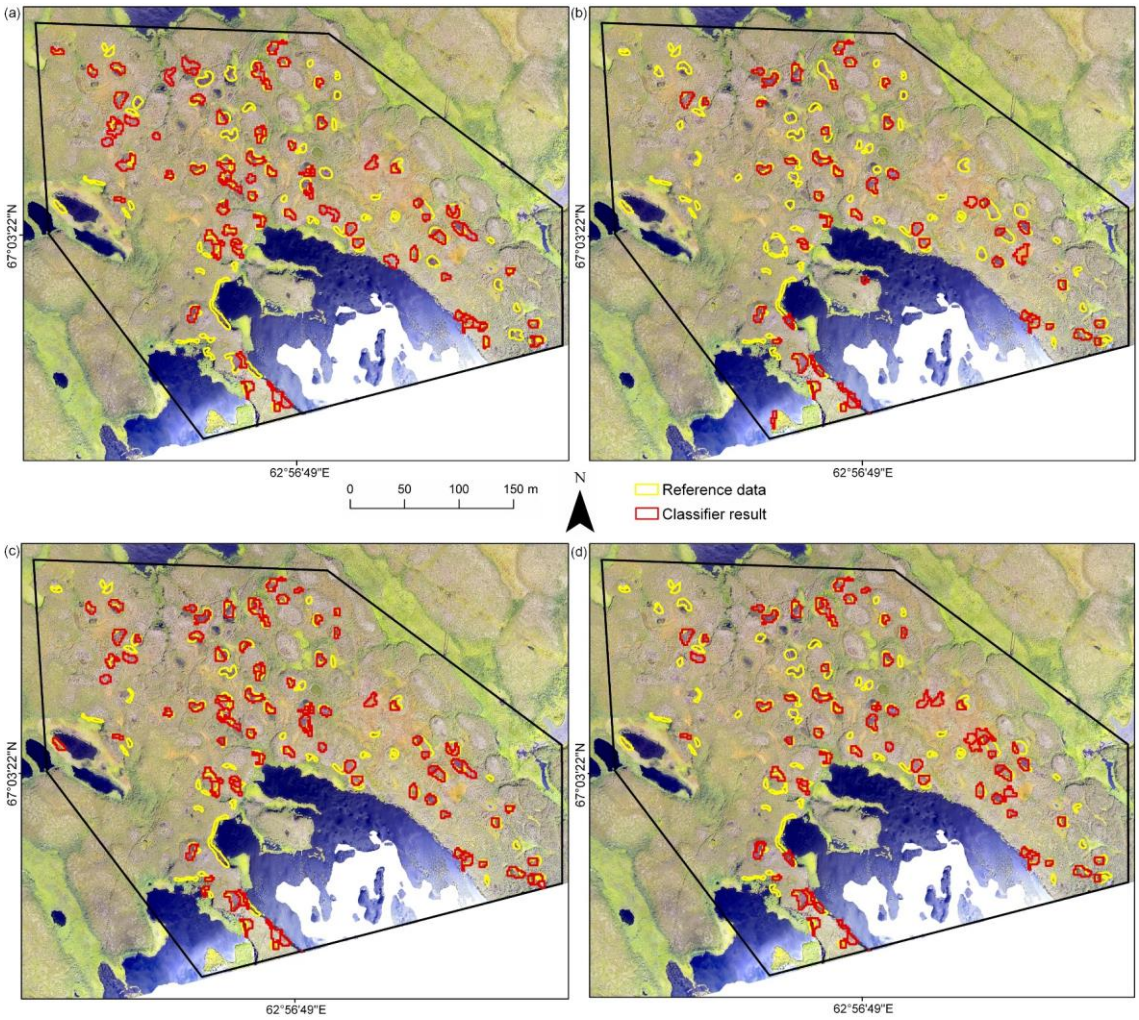
871

872 Figure 2. The workflow of the methods used.



873

874 Figure 3. The training area for the WV-3HR classifications. Bare peat areas were used
875 both in binary and positive and unlabelled settings and other land cover types in the
876 binary setting. Image ©DigitalGlobe.



877

878 Figure 4. Visualizations of the OCSVM (a), BSVM (b), RF (c), and ROF (d)
879 classifications for the WV-3HR classifications using the binary setting. Background
880 image: drone image from the validation area.

881