Remote Sensing of Environment 218 (2018) 119–131

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Invasive tree species detection in the Eastern Arc Mountains biodiversity hotspot using one class classification

Rami Piiroinen[a,b,*], Fabian Ewald Fassnacht[c], Janne Heiskanen[a,b], Eduardo Maeda[a,d], Benjamin Mack[e], Petri Pellikka[a,b,f]

[a] *Earth Change Observation Laboratory, Department of Geosciences and Geography, University of Helsinki, P.O. Box 64, FI-00014 Helsinki, Finland*
[b] *Institute for Atmospheric and Earth System Research, Faculty of Science, University of Helsinki*
[c] *Institute of Geography and Geoecology, Karlsruhe Institute of Technology, Kaiserstraße 12, 76131 Karlsruhe, Germany*
[d] *Ecosystems and Environment Research Programme, Faculty of Biological and Environmental Sciences, University of Helsinki, P.O. Box 68, FI-00014 Helsinki, Finland*
[e] *GAF AG, Arnulfstr. 199, 80634 Munich, Germany*
[f] *Helsinki Institute of Sustainability Science, University of Helsinki*

ARTICLE INFO

Keywords:
Biased support vector machine
Invasive tree species
Imaging spectroscopy
Africa
One class classification

ABSTRACT

Eucalyptus spp. and *Acacia mearnsii* are common exotic tree species in eastern Africa that have shown (strong) invasive behavior in some regions. *Acacia mearnsii* is considered a highly invasive species that is replacing native species and *Eucalyptus* spp. are known to consume high amounts of groundwater with suspected effects on native flora. Mapping the occurrence of these species in the Taita Hills, Kenya (part of the Eastern Arc Mountains Biodiversity Hotspot) is important as there is lack of knowledge on their occurrence and ecological impact in the area. Mapping methods that require a lot of fieldwork are impractical in areas like the Taita Hills, where the terrain is rugged and the infrastructure is poor. Our aim was hence to map the occurrence of these tree species in a 100 km^2 area using airborne imaging spectroscopy and laser scanning. We used a one class biased support vector machine (BSVM) classifier as it needs labeled training data only for the positive classes (*A. mearnsii* and *Eucalyptus* spp.), which potentially reduces the amount of required fieldwork. We also introduce a new approach for parameterizing and setting the threshold level simultaneously for the BSVM classifier. The second aim was to link the occurrence of these species to selected environmental variables. The results showed that the BSVM classifier is suitable for mapping *Acacia mearnsii* and *Eucalyptus* spp., holding the potential to improve the efficiency of field data collection. The introduced parametrization/threshold selection method performed better than other commonly used approaches. The crown level F1-score was 0.76 for *Eucalyptus* spp. and 0.78 for *A. mearnsii*. We show that *Eucalyptus* spp. and *A. mearnsii* trees cover 0.8% and 1.6% of the study area, respectively. Both species are particularly located on steeper slopes and higher altitudes. Both species have significant occurrences in areas close to the biggest remaining native forest patch (Ngangao) in the study area. Nonetheless, follow-up studies are needed to evaluate their impact on the native flora and fauna, as well as their impact on the water resources. The maps created in this study in combination with such follow-up studies could serve as base data to generate guidelines that authorities can use to take action in handling the problems these species are causing.

1. Introduction

A plant species is considered non-native or exotic if it is found in an ecosystem where it did not evolve. On the other hand, invasive plant species are defined as non-native plants that produce reproductive offspring in large numbers and at considerable distances from parent plants (Richardson et al., 2000). Woody plants, in general, were not widely recognized as invasive species until fairly recently (Richardson and Rejmánek, 2011). In contrast, nowadays invasive trees and shrubs are considered in some cases among the most conspicuous and damaging life-forms, threatening local flora and fauna. In Africa, some alien tree species serving as backbone of the local plantation forestry have high economic significance, but at the same time decimate land and water resources (Chenje and Mohamed-Katerere, 2006).

* Corresponding author at: Earth Change Observation Laboratory, Department of Geosciences and Geography, University of Helsinki, P.O. Box 64, FI-00014, Helsinki, Finland.

E-mail addresses: rami.piiroinen@helsinki.fi (R. Piiroinen), fabian.fassnacht@kit.edu (F.E. Fassnacht), benjamin.mack@gaf.de (B. Mack).

https://doi.org/10.1016/j.rse.2018.09.018
Received 22 May 2018; Received in revised form 18 September 2018; Accepted 21 September 2018
Available online 28 September 2018

One of the most invasive alien tree species in Africa is *Acacia mearnsii*, featured also in the list of '100 of the World's Worst Invaders' (Lowe et al., 2000). This species, native in Australia, has been shown to compete with native species, to reduce native biodiversity, and to reduce water availability in riparian zones (Boudiaf et al., 2013; Richardson and Rejmánek, 2011). For instance, in South-Africa *A. mearnsii* was originally planted on 107,000 ha, but is estimated now to have spread to a total area of 2,500,000 ha (Nyoka, 2003). Another genus of trees known to cause environmental problems in sub-Saharan Africa and also considered invasive in some areas is *Eucalyptus* (Richardson and Rejmánek, 2011). For instance, conversion of grassland by afforestation with alien *Eucalyptus* spp. affects negatively the catchment runoff (Turpie et al., 2008). In some cases *Eucalyptus* spp. plantations have even completely dried up rivers (Rodriguez-Suarez et al., 2011). The leaf litter of *Eucalyptus* spp. (including *Eucalyptus saligna*) also contain phytotoxic compounds, that inhibit germination and initial growth of certain grassland species, and possible allelopathic effects (Silva et al., 2017).

While the adverse impacts of *Eucalyptus* spp. and *A. mearnsii* are well understood in South Africa (Nyoka, 2003; Turpie et al., 2008), fewer assessments have been conducted elsewhere in Africa. For instance, in Kenya, detailed maps of the current occurrences of these invasive species are still missing. In this study, we address this research gap by developing an efficient approach for assessing the occurrence of *A. mearnsii* and *Eucalyptus* spp. in the Taita Hills, Kenya, with limited field data.

The Taita Hills are part of the Eastern Arc Mountains biodiversity hotspot, which is known to host many endemic species (Burgess et al., 2007). However, according to a recent study, only 0.8% of the Taita Hills region are still covered with the native indigenous cloud forests which contain a large share of the endemic species and biodiversity of the area (Thijs et al., 2015). Many exotic tree species have been introduced to produce lumber (*Eucalyptus* spp., *Grevillea robusta*), tannin (*A. mearnsii*) and food (*Mangifera indica*, *Persea americana*). Aside from pure plantation forests, tree cover has increased on the croplands as treeless fields have been converted to agroforestry systems (Pellikka et al., 2018). These agroforestry systems often include exotic tree species with some of them being considered invasive. Mapping the occurrence of these invasive species would hence be highly valuable as the current spread and the impact of these species on the ecosystem in the study area is not well known. One existing study on the occurrence of tree species in the Taita Hills was based on field sampling (Thijs et al., 2015), which is an accurate but time-consuming approach that is not a practical solution for mapping the species at a broad scale and with high spatial accuracy.

An alternative approach for inventorying tree species over the entire region is provided by remote sensing (RS) techniques (Fassnacht et al., 2016). Imaging spectroscopy (IS) and airborne laser scanning (ALS) are the most common RS data sources used for the classification of tree species in the research literature (Fassnacht et al., 2016). Using these two data sources together (data fusion) has yielded the best results in many cases (Fassnacht et al., 2016). The data fusion is often performed at object level as it enables calculating smooth spectral features but also features that depict the structure and shape of a tree. However, in the tropics, the canopy structure is often complex and the automatic delineation of tree crowns is challenging (Feret and Asner, 2013; Piiroinen et al., 2017). Thus, pixel-based mapping approaches have also been presented (Baldeck et al., 2015). Most of the studies utilizing IS and ALS data for mapping tree species have been conducted in temperate forests, while fewer studies have been located in tropical or sub-tropical areas, and those mainly in Central America, South America and southern Africa (Baldeck and Asner, 2015; Baldeck et al., 2015; Cho et al., 2012; Fassnacht et al., 2016; Graves et al., 2016). Only one recent study was conducted in Kenya (Piiroinen et al., 2017).

Tree species classification and mapping studies typically use supervised classification approaches, where the classifier is trained using field measurements of all the tree species that are present in the study area. This approach is sometimes impractical. This particularly applies for tropical regions where a single study site may have dozens or hundreds of different tree species, which makes collecting representative training and validation data very laborious, particularly in areas with limited infrastructure. Furthermore, only a few species might be relevant for the application or research question. If the latter applies, the use of a one class classification (OCC) approach, where labeled data is needed only for the positive class (that is, a single tree species) might be an efficient alternative (Mũnoz-Marí et al., 2010).

In RS studies, OCCs have been used, for example, to detect focal tree species in tropical rainforests (Baldeck et al., 2015), Natura 2000 habitats and high nature value grassland habitats (Stenzel et al., 2014, 2017), for invasive species detection (Skowronek et al., 2017a, 2017b), and detecting savanna tree species in Africa (Baldeck and Asner, 2015). From the plethora of available OCC algorithms, particularly one class support vector machine (OCSVM), biased support vector machine (BSVM) and Maxent have been frequently used (Mack and Waske, 2017). OCSVM (Scholkopf et al., 1999) uses only data from the class of interest to train the classifier, while BSVM is a semi-supervised classification algorithm that utilizes also unlabeled samples (Liu et al., 2003). In a recent study conducted in Panama, very high classification accuracies were achieved with BSVM for detecting non-flowering focal tree species at the pixel level (Baldeck et al., 2015). Mack and Waske (2017) showed in their comparison of different OCC algorithms that BSVM had the highest discriminative potential followed by Maxent (with parameter tuning), Maxent (with default parameters) and OCSVM. Similarly, Stenzel et al. (2017) reported that BSVM outperformed Maxent (with default parameters) and OCSVM in the classification of high nature value grassland areas. Stenzel et al. (2017) concluded that the results could have been further improved by more sophisticated parameter tuning.

One of the benefits of using OCCs in invasive species mapping is that many governmental organizations in charge of nature conservation and management keep record of the known locations of certain invasive species and this information can be readily used to make initial maps. For instance, Wakie et al. (2014) collected 143 observations of invasive *Prosopis juliflora* in Ethiopia using targeted field sampling that was based on the pre-existing knowledge of heavily infested sites that the local communities and government employees had. Wakie et al. (2014) then used MODIS data and Maxent to model the occurrence of *P. juliflora*. The same approach does not work in supervised classification methods as information on all other species (the negative class) is often missing.

The first aim of this study was to examine the potential of the OCC approach and a BSVM classifier in combination with pixel-level data fusion of IS and ALS data for mapping common invasive tree species, namely *Eucalyptus* spp. and *A. mearnsii*, in the Taita Hills, Kenya. We selected BSVM (Liu et al., 2003) based on its good performance in previous studies (Mack and Waske, 2017; Stenzel et al., 2017). Furthermore, Mack et al. (2014) suggested that better results can be achieved for BSVM when the threshold (cut-off value) is selected manually based on diagnostic plots in case the automatic procedures fail. However, in practice, the manual tuning of the threshold might be challenging. To respond to this challenge, we introduce a new approach for selecting the model and tuning the threshold simultaneously.

The second aim was to map the occurrence of these species and relate their occurrence with selected environmental variables. The goal was to achieve a better understanding of the locations that are most heavily affected by these species. The results can serve as a baseline for studying the impacts of these species on the ecosystem, biodiversity and water resources.

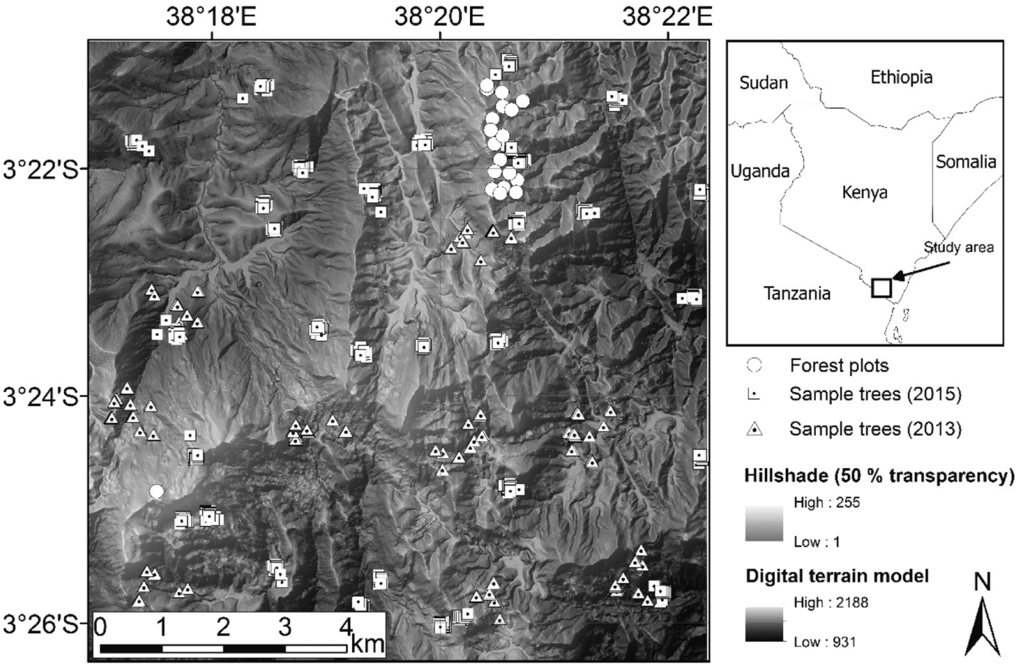

Fig. 1. Location of the study area in the Coast Province of Kenya. Locations of field measurements are shown on top of a digital terrain model with a hillshade overlay (50% transparency) produced from the laser scanning data.

2. Material and methods

2.1. Study site

The study area (10 km × 10 km) is located in the elevation range of 1100–2200 m a.s.l. in the Taita Hills (3° 25′ S, 38° 19′ E) in southeast Kenya (Fig. 1). The Taita Hills are known for its exceptionally high degree of endemism and conservation value (Aerts et al., 2011; Burgess et al., 2007). The potential natural vegetation for the Taita Hills is moist Afromontane forest or cloud forest (Aerts et al., 2011). However, most of the forested areas have been cleared for agricultural use already > 100 years ago (Clark and Pellikka, 2009; Pellikka et al., 2013). The largest patches of remaining forest and the highest levels of above-ground biomass are located on hilltops and steep slopes which have been too steep to clear for agriculture (Adhikari et al., 2017). Currently, only around 4.2 km^2 of montane forests persist in 12 forest relicts (Pellikka et al., 2009).

During the field campaign, *Eucalyptus* spp. (mostly *Eucalyptus saligna*) individuals were found especially in plantation forests (Fig. 2) which were often located on steep slopes. *Eucalyptus* spp. were originally brought to the area for producing lumber. It is suspected that *Eucalyptus* spp. individuals have been spreading from the original plantation sites, but no accurate information on their current occurrence is available prior to this study. Eucalyptuses can grow up to 40 m of height and the plantations are significant carbon stocks in the area (Pellikka et al., 2018). *A. mearnsii* trees (Fig. 2) have originally been brought to the area to produce tannins for leather production. Presently, the leather production has only small economic importance and many consider the species as a weed and cut it for use as firewood.

2.2. Field data

There were two campaigns for collecting tree level measurements and two campaigns for collecting plot level measurements from the native forests.

The first tree level campaign was organized between 17 January and 8 February 2013. The 100 km^2 study area was divided into 16 tiles (each covering 2.5 km × 2.5 km), which were each sampled by one 100 ha cluster selected randomly. Each cluster had ten circular 0.1 ha study plots (17.84 m radius). Ten clusters were selected for detailed tree sampling, and as one plot was treeless, this resulted in 99 study plots. From each study plot, every tree that had a diameter at breast height > 10 cm was measured. The central point of each study plot was measured with GNSS (Trimble GeoExplorer GeoXH 6000, Trimble Inc., Sunnyvale, CA, USA). Measuring tape and compass were used to measure the relative position of each tree from the plot center. To enable the differential correction of the data, a GNSS base station (Trimble Pro 6H receiver, Trimble Inc., Sunnyvale, CA, USA) was logging in a known position during the field measurements. The data from 2013 contained 531 trees.

The second tree level campaign was organized during 1–30 October 2015. This time, the study area was divided into 1 km × 1 km tiles and 30 tiles were randomly selected. Each tile was further divided into rectangular 1 ha study plots and one study plot was selected at random within each tile, with the exception of one tile that had two study plots. In total, there were 31 study plots. Within each study plot, nine sampling points with 33.3 m intervals were established. At each sampling point, two trees were selected using the T-square plotless sampling method (Engeman et al., 1994; Thijs et al., 2015). The same GNSS receiver and base station were used as in 2013 but each tree was measured directly with GNSS. A tree was defined as any woody plant taller than five meters. The data from 2015 contained 538 trees. From these, we excluded 98 trees that were either located under higher trees (not visible from the air) or that had GNSS positional accuracy < 4 m, which resulted in 440 crowns. In total, there were 971 tree level measurements (data from years 2013 and 2015 combined) from 64 different tree species. A total of 65 of the trees were *A. mearnsii* and 62 were *Eucalyptus* spp.

The two forest plot measurements were collected in January and February of 2013 and 2014. The number of different tree species in the circular 0.1 ha-sized field plots was recorded, but the exact position of each tree was not recorded. Schäfer et al. (2016) have described this dataset in detail (Schäfer et al., 2016). The plot level measurements were used for validating the model in the closed-canopy native forests. The forest plots did not include any *Eucalyptus* spp. or *A. mearnsii* trees. The forest plots had 58 different species. In total there were 97 different

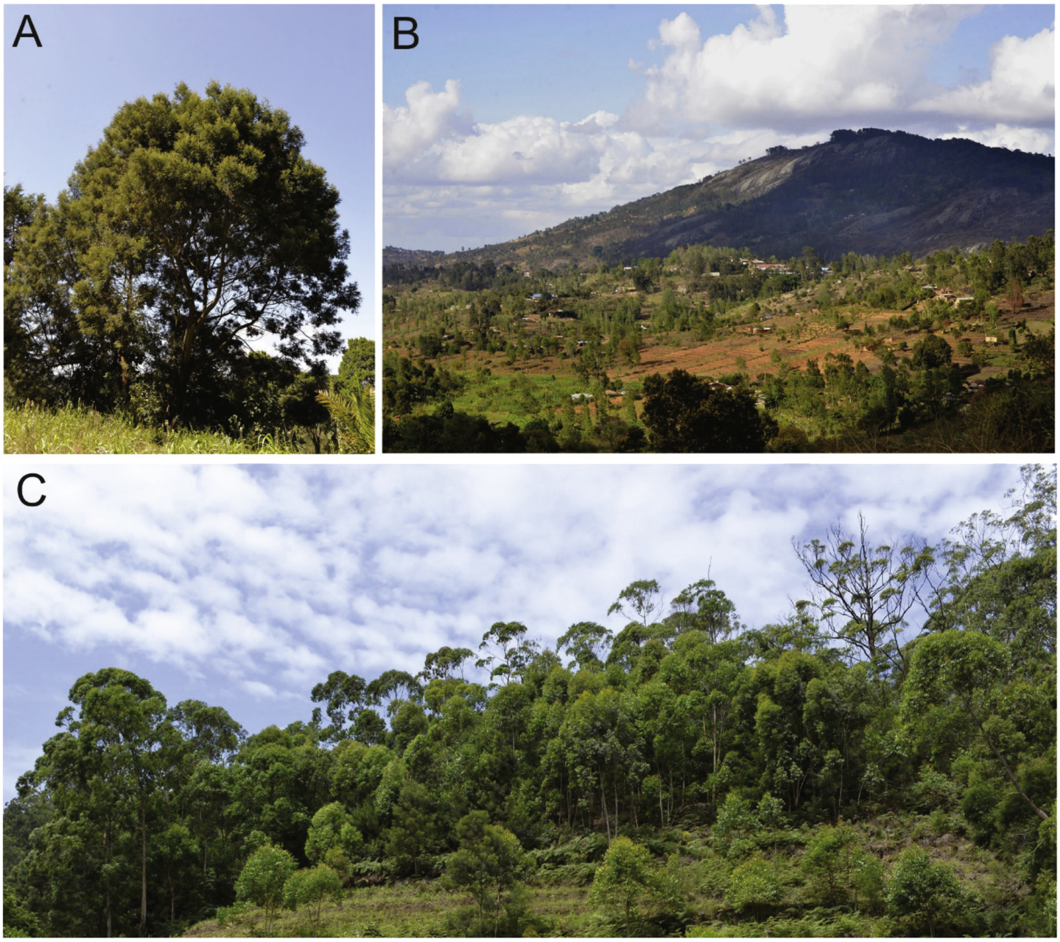

Fig. 2. A) *Acacia mearnsii* by a roadside, B) Agricultural landscape with *Eucaluptus* ssp., *Acacia mearnsii*, *Grevillea robusta* among fruit trees, C) *Eucaluptus* ssp. plantation behind terraced agricultural fields. Photos: P. Pellikka, 2018.

tree species in all datasets (tree level measurements and forest plots) combined.

2.3. Remote sensing data

The airborne remote sensing data acquisition campaign was conducted during the dry season in 2013 (3–8 February). Two sensors were used for collecting the IS and ALS data from a mean flying height of 750 m above ground. The IS data were acquired with an AisaEAGLE (Spectral Imaging Ltd., Oulu, Finland) sensor, a pushbroom scanner with an instantaneous field of view of 0.648 mrad and field of view of 36.04°. The sensor was used with four times spectral binning mode that produced output images with 129 bands and a full width at half maximum of 4.5–5.0 nm in the spectral range of 400–1000 nm. The output pixel size was one meter. ALS data was collected with an Optech ALTM 3100 sensor (Teledyne Optech, Vaughan, Ontario, Canada). The ALTM 3100 is an oscillating mirror laser scanner capable of recording up to four echoes (returns). The sensor was operated at a pulse rate of 100 kHz and a scan rate of 36 Hz. Scan angle was ± 16°.

ALS data were preprocessed by the data vendor (Topscan Gmbh, Rheine, Germany) and delivered as a georeferenced point cloud in UTM37S/WGS84 coordinate system with ellipsoidal heights. Buildings and power lines were excluded and some erroneous measurements from steep slopes were removed using TerraScan software (Terrasolid Ltd., Helsinki, Finland). The point cloud was then used to create a rasterized canopy height model (CHM).

The raw IS data were radiometrically corrected and orthorectified with CaliGeoPro 2.2 (Spectral Imaging Ltd. Oulu, Finland). Atmospheric correction was applied using ATCOR-4 (ReSe Applications Schläpfer, Wil, Switzerland), (Richter and Schläpfer, 2002). After the orthorectification, it was noted that there were geometric mismatches between IS and ALS data. As the LiDAR sensor system had a higher quality inertial measurement unit and the IS data had obvious distortions, we co-registered the ALS and IS data using control points collected manually from the CHM. The processed IS scanning lines were clipped so that the side overlap was minimized to reduce the distortions on the sides of the flight lines. In total, 50–100 control points were collected for each flight line and first order polynomial transformation was applied to co-register the images. After the co-registration, RMSE at ground level was 1.06 m.

As we were only interested in trees, we generated masks to remove non-tree pixels from the IS data. First, we delineated manually the areas that were most heavily shadowed by clouds. Next, we masked all pixels with NIR (836 nm) reflectance < 20% and NDVI (Rouse et al., 1973) < 0.5 to remove non-vegetation pixels and remaining shadows. This process removed the majority of shadows resulting from the varying sun zenith angle as well as shadows resulting from the clouds. We then masked all pixels with heights < 3 m (based on CHM) to remove remaining non-tree pixels. Cloud masking was not required as the aircraft was flying at a low altitude, below potential clouds.

As there were 129 highly correlated bands in our hyperspectral dataset, we used minimum noise fraction (MNF) transformation (Green et al., 1988) to reduce noise and to pack the coherent information in a smaller set of features. The MNF algorithm was implemented in ENVI software (version 5.0, Research Systems Inc., Boulder, CO, USA) (RSI, 2004). The MNF transformation was applied to the reflectance data

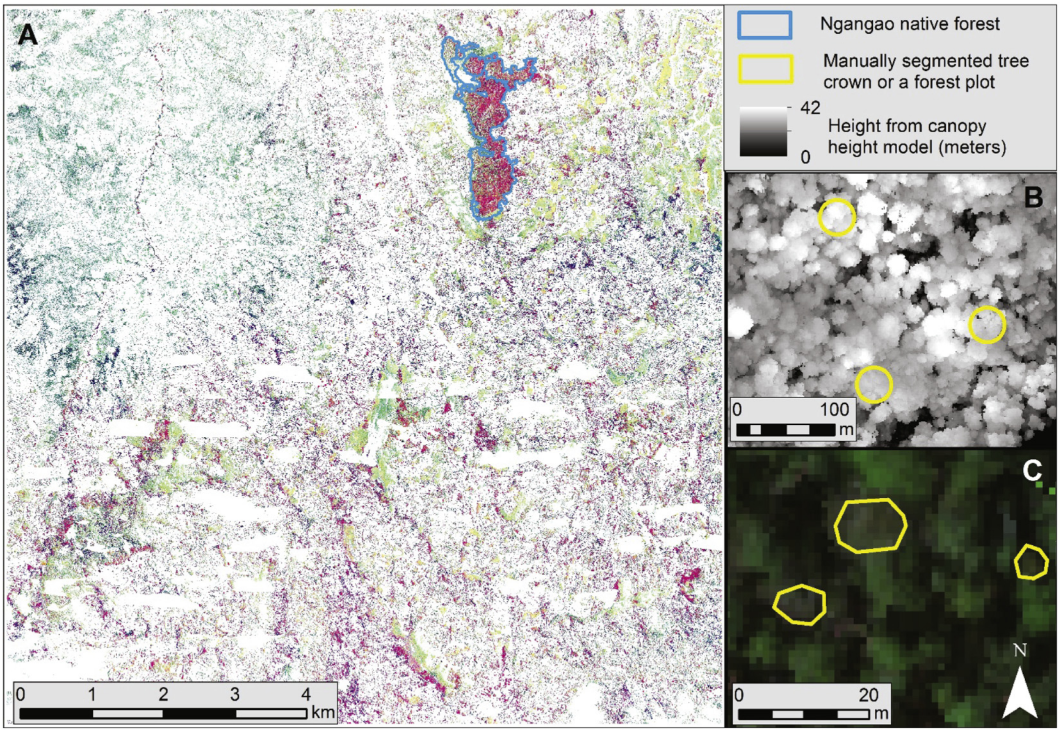
A
B
C
Ngangao native forest
Manually segmented tree crown or a forest plot
42
0
Height from canopy height model (meters)
0
100
m
0
20
m
0
1
2
3
4
km
N

Fig. 3. a) MNF components 1, 2 and 3 (white areas have been masked with the shadow/tree mask). b) An example of the forest plots on top of the canopy height model and c) an example of manually segmented tree crowns on top of true color reflectance data.

where all the non-tree pixels were first masked out (Fig. 3).

2.4. Samples and features

First, we plotted all tree level measurements (n = 971) on top of the IS data and manually delineated the tree crowns (Fig. 3c). If we could not match the field measurement to a visible tree crown, the tree was omitted. If many field measurements from the same species were close to each other and only one tree crown could be identified, we segmented only one crown covering all these field measurements. The forest plots did not need manual segmentation as they covered 0.1 ha-sized circular area around the center of the plot. We then extracted features from the MNF (Fig. 3a) and CHM (Fig. 3b) data for each manually segmented tree crown and forest plot. We included the MNF components 1–10 in the classification model and omitted the rest based on their low eigenvalues and noisiness in visual interpretation (Fassnacht et al., 2014; Piiroinen et al., 2015). From CHM we derived the height of each pixel, and focal mean and variance (Table 1). Some of the tree crowns contained only pixels that were masked away (shadow, height and NDVI) and were omitted. In the end, 707 tree crowns from 65 different tree species were available for the classification.

2.5. Classification

For the classification, we used BSVM (Liu et al., 2003) which is an adaptation of the binary SVM (Mountrakis et al., 2011; Vapnik, 1998). SVMs construct hyperplanes that maximize the margin between two classes. In supervised binary SVM, training data from two classes with available samples (i.e. positive and negative) are used to train the algorithm. The basic principle of BSVM is the same as in binary SVM as it finds an optimal separation between two classes. The main difference is that in BSVM the negative (absence) class is replaced by an unlabeled class (random samples). As the unlabeled class contains samples also from the positive class, two cost terms are used for the positive (C_+) and unlabeled classes (C_U) (Lee and Liu, 2003; Mack and Waske, 2017). The sample size of the unlabeled data should be large enough to hold a significant amount of positive samples. Then, the misclassification on the unlabeled training samples can be penalized less strongly (Liu et al., 2003; Mack et al., 2014).

Table 1
List of the features (17 in total) used in the classification process.

Data source	Feature names
Hyperspectral	MNF components 1–10
Canopy height model	Pixel height, focal mean and variance with window sizes 5 × 5 and 25 × 25

Gaussian radial basis function was applied as kernel to create a nonlinear classifier by fitting the separating hyperplane in a transformed feature space (Mack et al., 2014). The inverse kernel width σ was tuned in addition to C_+ and C_U using a grid search during a 10-fold cross-validation (378 different parameter combinations in total). The tested values were in the range of σ 0.05–0.55, C_U 0.1–1.9, C_+ 1–25 (exact values: Supplementary Fig. 1). We used the implementation of BSVM from the R package “oneClass” (Mack, 2015) during the classification process as it is openly available and open source.

The classification (Fig. 4) was conducted separately for *Eucalyptus* spp. and *A. mearnsii*. First, all labeled samples (PN-data) (Table 2) were divided into training (2/3 of samples) and testing datasets (1/3 of samples) (Fig. 4). The negative samples included in the training data were omitted from the classification. The data was divided at tree crown level to prevent samples (pixels) from the same tree crown being included in the training and test datasets. The positive samples are here on after referred to as P-data and negative samples as N-data. Additionally, we took a random sample of 20,000 unlabeled tree pixels to serve as the unlabeled data (U-data). We used only positive and unlabeled data (PU-data) during the model training and selection. The PN-test data was used only to validate the selected model to verify the results.

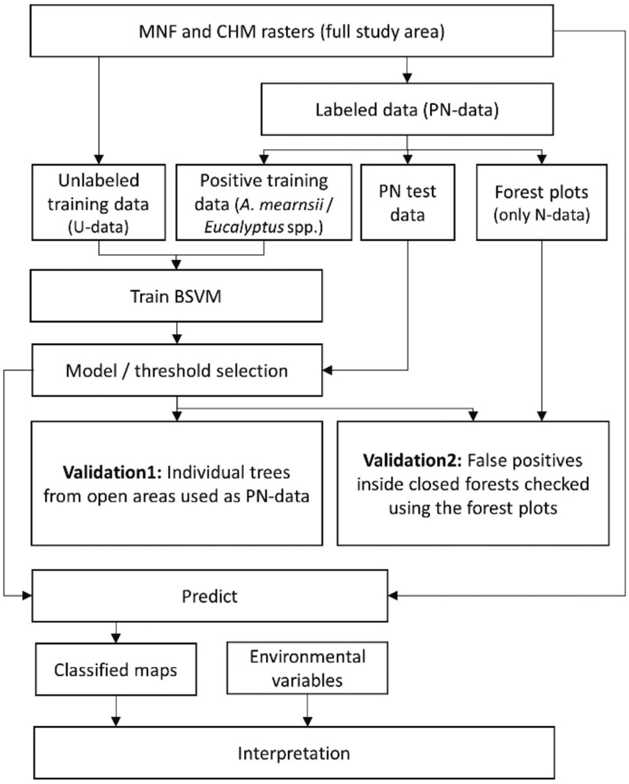
MNF and CHM rasters (full study area)
Labeled data (PN-data)
Unlabeled training data (U-data)
Positive training data (A. mearnsii / Eucalyptus spp.)
PN test data
Forest plots (only N-data)
Train BSVM
Model / threshold selection
Validation1: Individual trees from open areas used as PN-data
Validation2: False positives inside closed forests checked using the forest plots
Predict
Classified maps
Environmental variables
Interpretation

Fig. 4. The classification, validation and prediction setup. P = positive, N = negative and U = unlabeled sample.

2.6. Model and threshold selection

The true positive rate (TPR) also known as recall or producer's accuracy is a standard measure in the evaluation of classification results. It estimates the probability that a positive sample is classified correctly. In supervised classifications we could also use precision (user's accuracy) which depicts the probability that a sample classified as positive truly belongs to this class. In OCC classification, we cannot calculate precision as we do not know for certain which samples are negative. In our case, we have this information available, but it was not used during model selection so that we do not violate the OCC principles. Instead of precision, we use the probability of a positive prediction (PPP) that estimates the probability that a sample is classified as positive in relation to all samples (positive and unlabeled). For instance, if the TPR is 1 then we know that we have identified all the positive samples correctly (e.g. all eucalyptuses have been classified as eucalyptus). Now, let's assume that we have a model that yields a TPR of 1 and a PPP of 1. In this case, the classifier has simply classified all the samples to the positive class and the result is useless. Thus, we can argue that if we have two models with the same TPR but different PPP, the model with lower PPP is more accurate, because TPR is the same but the false positive rate (FPR) is necessarily lower.

BSVM gives continuous output values for each predicted sample. These values can be either positive or negative. A discrimination threshold is thus set to get a crisp (categorical) classification result that discriminates the positive and negative predictions. The selection of an appropriate threshold is a critical decision that also influences typical model selection criteria. Commonly used model selection criteria for BSVM include F_{PU} (Lee and Liu, 2003; Liu et al., 2003) and AUC_{PU} (Phillips et al., 2006; Phillips and Dudík, 2008). F_{PU} is calculated as TPR^2/PPP (Lee and Liu, 2003; Liu et al., 2003) and it aims to maximize TPR and minimize PPP. It resembles the PN-metric F-score and is thought to work similarly, as the F-score is high when recall and precision are high (Lee and Liu, 2003). One problem, when F_{PU} is used, is that it is often estimated at only one threshold level (commonly at zero). Thus, a model could have a high discriminative power at a certain threshold, but it is ranked low because the default threshold of zero is not suitable. Contrarily, AUC_{PU} is calculated independently of the threshold level. AUC_{PU} resembles area under the receiver operator characteristic curve (AUC) that is commonly used in supervised classification setting. The receiver operating characteristic (ROC) curve is calculated by plotting TPR and FPR at different threshold levels (Phillips et al., 2006), while the AUC is the surface area under the ROC curve. In AUC_{PU} the negative samples used to calculate the FPR are replaced by unlabeled random samples. Thus, AUC_{PU} can be interpreted as the probability that a randomly chosen presence sample is ranked above a random background sample (Phillips et al., 2006; Phillips and Dudík, 2008). As AUC_{PU} (like AUC) is calculated independent of the threshold level and using randomly sampled observations, it is insensitive to class imbalance (Fawcett, 2006). However, the AUC_{PU} (like AUC) also considers thresholds that are most likely unsuitable and which may result in a misleading interpretation of the results (Lobo et al., 2008).

Adjusting the threshold manually after selecting the model based on F_{PU} or AUC_{PU} can lead to better results (Mack et al., 2014), but this process requires an experienced user with knowledge of the classification task at hand. Thus, the procedure is impractical for people that are not experienced with OCC methods and the corresponding workflow. To address this challenge, we introduce a new, automated method for simultaneous parameterization and threshold selection for BSVM and compare the results with the models selected based on F_{PU} and AUC_{PU}.

The introduced method is based on the idea of finding Pareto optimal solutions introduced by Persello and Bruzzone (2009). We apply the same idea to the OCC classification problem and include simultaneous threshold selection in the workflow. We first identify the models that produce so called non-dominated solutions at a given threshold level for PPP and TPR. A given model + threshold (M + T) combination is considered non-dominated if TPR cannot be increased without increasing PPP and the identical PPP + TPR values have been removed (Persello and Bruzzone, 2009). The non-dominated M + T combinations are considered as the Pareto front. The concept is clarified in Fig. 5.

We calculated TPR and PPP for each model at 50 different threshold levels. Thus, we had 18,900 (378 models × 50 thresholds) model + threshold (M + T) combinations. Next, we assessed if the M + T combinations were dominated or non-dominated. Then, we find the M + T combination that produces TPR and PPP that are located as close as possible to the upper left corner of the introduced diagnostic plot (upper left corner is the point where TPR is 1 and PPP 0). Similar min. dist. approach has been used earlier to determine cut-off (threshold) value in ROC analysis in a supervised setting earlier by Habibzadeh et al. (2016). From here on after, this M + T combination is

Table 2
Training and validation samples.

Positive tree species	Positive training samples (pixels)	Positive tree crowns in training	Unlabeled training pixels	Positive test samples (pixels)	Positive tree crowns in testing	Negative test samples (pixels)	Negative tree crowns in testing	Forest plots
Eucalyptus spp.	512	40	20,000	194	22	2120	212	23
Acacia mearnsii	252	42	20,000	167	23	2087	213	23

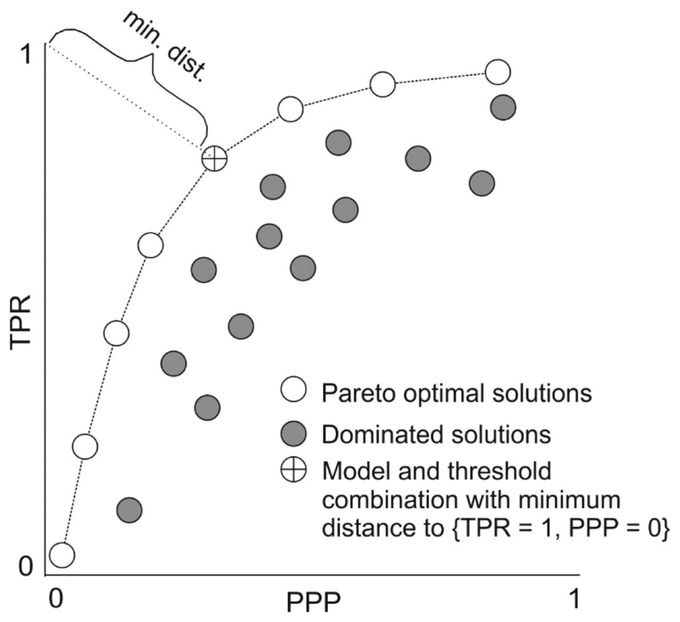

Fig. 5. Example of Pareto-optimal solutions and dominated solutions in one-class classification setting using TPR and PPP as the parameters, and the solution with minimum distance to upper left corner of the diagnostic plot.

referred to as “min. dist”. Selecting this min. dist. M + T combination that produced TPR and PPP located at the Pareto front, and as close as possible to TPR = 1 and PPP = 0, gives us a model that is maximizing TPR and PPP without favoring either, thus leading to a balanced classification result.

2.7. Validation

The selected M + T combinations were validated with PN-data. We used precision (user's accuracy), recall (producer's accuracy) and F-score to evaluate the performance of the models at the selected threshold levels. First, we used only the tree level measurements. The validation was repeated 25 times by taking a random sample (70% of test samples) during each validation round (means reported at tree crown level). If > 50% of the pixels of each crown were classified correctly we considered that crown correctly classified. Next, we used the forest plots to check for false positives (percentage of pixels inside the forest plots misclassified as *Eucalyptus* spp. or *A. mearnsii*) inside the closed forest. Including the forest plots into the same validation scheme would have skewed the distribution of the test data in favor of the negative class and the validation results would have been biased.

2.8. Analysis of the species occurrence in relation to environmental variables

The selected M + T combination was used to predict species occurrence over the full study area. The prediction results were then aggregated to a grid with a cell size of 30 × 30 m for visualization. First, we calculated the cover of *Eucalyptus* spp. and *A. mearnsii* within these grid cells (positives/all pixels). Next, we calculated the cover of *Eucalyptus* spp. and *A. mearnsii* for the same grid cells in relation to tree pixels (positives/tree pixels). The results were aggregated on a grid as the pixel level results (1 m resolution) are not easy to interpret intuitively on a study area of this size (10 × 10 km). The aggregated results also highlight the areas that are most heavily impacted by these species and are hence under the most severe environmental threat.

The occurrence patterns of the two species were then studied together with four environmental variables, which we suspected to influence the occurrence patterns. These variables included slope, aspect, elevation and proximity to main rivers. First, a digital terrain model (DTM) was generated from the ALS point cloud at one meter resolution. This DTM was then resampled to 30 m spatial resolution. Next, slope and aspect were calculated from the 30 m DTM using Horn's algorithm (Horn, 1981) included in the “raster” package (Hijmans, 2016) in R. We used 30 m spatial resolution as we were interested in how the forest level topographic conditions affected the occurrence patterns. Using higher spatial resolution DTM would have introduced micro level noise in the data that would not have helped the interpretation of the general occurrence patterns of these species. The river networks were extracted in SAGA GIS (v. 2.1.2) using the “channel network” module. Adhikari et al. (2017) have described this process in detail. The resulting river network was then edited manually and only the main rivers were kept. The distance to the main rivers were then calculated by converting the river network into a raster surface and by calculating the Euclidian distances from each raster cell to the closest river at 30 m spatial resolution.

The relationship of the species and the environmental variables were studied by taking a random sample of the pixels classified as *Eucalyptus* spp. or *A. mearnsii* and calculating the kernel density estimates in relation to the selected environmental variables. The results were compared to the distribution of all tree pixels in the study area and a random sample covering the whole study area (also non-tree targets). An unpaired Wilcoxon test was used to test if the differences in the distributions are statistically significant. This test was not conducted for aspect as it is not suitable for the distribution of aspect values.

3. Results

3.1. Pareto optimal OCC model and threshold selection

Recall that our aim was to select a model and threshold combination that has a high TPR rate and at the same time low PPP. The Fig. 6 shows how the PPP is increasing as the TPR increases. For *Eucalyptus* spp. the PPP starts to increase faster around TPR 0.9, and for *A. mearnsii* around TPR 0.8. The Pareto fronts (blue) are seen as M + T combinations that have the lowest PPP for the given TPR. The threshold zero solutions for both species move away from the Pareto front at higher TPR. The models that had the highest F_{PU} at threshold zero or the highest AUC_{PU} yielded lower TPR than the models selected based on min. dist. The min. dist. models are located at the Pareto front. The min. dist. model for *A. mearnsii* has higher TPR than any of the models at threshold zero. The min. dist. model for *Eucalyptus* spp. has the same TPR than the best threshold zero solution, but with lower PPP. The min. dist. M + T combinations for both species have TPR and PPP that are very close to the M + T combinations that produced the highest F-score on the PN test data. There were three models that produced exactly the same highest F-score with test data for *Eucalyptus* spp. and two for *A. mearnsii*. All of these were located close to the Pareto fronts and the M + T combination selected based on min. dist.

There were 73 models (out of 378) for *Eucalyptus* spp. and 63 for *A. mearnsii* that had a solution (at any of the 50 threshold levels) at the Pareto front (Fig. 6). When only the Pareto optimal M + T combinations were considered, the amount of possible M + T combinations were reduced from 18,900 to 264 and 130, *for A. mearnsii* and *Eucalyptus* spp., correspondingly.

3.2. Validation with positive/negative test data

The models with the highest F_{PU} and AUC_{PU} had very high precision and low recall at the threshold level zero (Table 3). The M + T combination selected based on min. dist. had the highest recall and F-score for both species and the lowest precision. No pixels inside the forest plots were classified as *A. mearnsii* when F_{PU} or AUC_{PU} were used to select the model (Table 4). The M + T combination selected based on min. dist. misclassified 1.36 and 0.02% of the pixels inside the forest

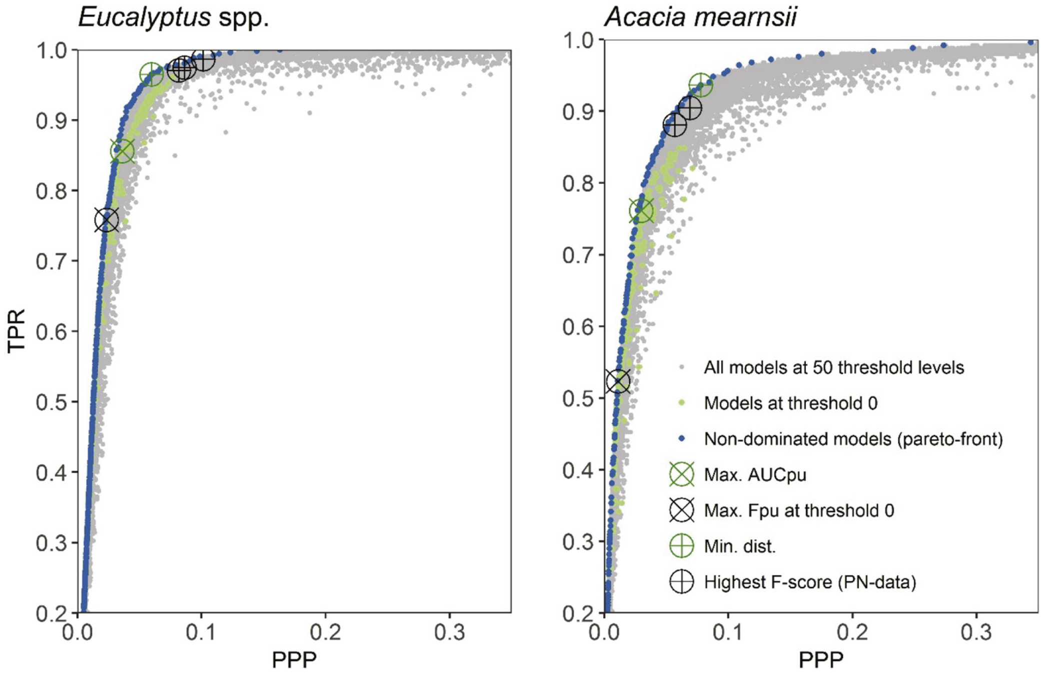

Fig. 6. True positive rate (TPR) and probability of positive prediction (PPP) for all models at 50 threshold levels. The non-dominated models (Pareto front), models at threshold 0, models with the highest F_{PU}, AUC_{PU} and minimum distance to the top left corner (min. dist.) and the models that produced the highest F-score with the PN test data are indicated separately. The PPP values extend to 1, but these are not shown in the plots.

Table 3
Validation results with the independent test data at the tree crown level. The values are means from 25 iterations with different subsamples of the test data.

Model selection	Metric	*Eucalyptus* spp.	*Acacia mearnsii*
F_{PU}	precision	**1.0**	**1.0**
AUC_{PU}	precision	0.98	0.92
Min. dist	precision	0.93	0.75
F_{PU}	recall	0.25	0.29
AUC_{PU}	recall	0.49	0.49
Min. dist	recall	**0.64**	**0.81**
F_{PU}	F-score	0.39	0.44
AUC_{PU}	F-score	0.65	0.64
Min. dist	F-score	**0.76**	**0.78**

Highest precision, recall and F-score values have been highlighted with bold.

Table 4
The percentage of pixels inside the forest plots that were falsely classified as positives (*Eucalyptus* spp. or *Acacia mearnsii*).

Metric	Native forest pixels classified as *Eucalyptus* spp. (%)	Native forest pixels classified as *Acacia mearnsii* (%)
F_{PU}	0.04	0
AUC_{PU}	0.38	0
Min. dist.	1.36	0.02

plots as *Eucalyptus* spp. and *A. mearnsii*, correspondingly.

3.3. Occurrence of Acacia mearnsii and Eucalyptus spp.

Eucalyptus spp. and *A. mearnsii* cover 0.8% and 1.6% of the study area, respectively. Both species occur especially in higher altitudes (Figs. 7, 8 and 9). *Eucalyptus* spp. occurs especially on steep South-East facing slopes, while fewer individuals are found on North-West facing slopes. *A. mearnsii* can be found throughout the higher altitude areas (Figs. 7a and 9) and is following the general trend of the occurrence of trees in the area. Un-paired Wilcoxon test results showed that the occurrence of the predicted *Eucalyptus* spp. and *A. mearnsii* pixels differed from the distribution of all tree pixels and random pixels with statistical significance in all instances (p-value < 0.05). Both species have notable occurrences close to the largest remaining native forest patch Ngangao within the study area (Fig. 7d). *A. mearnsii* was dominant (over 50% of the tree pixels classified as *A. mearnsii*) in many locations scattered throughout the higher altitude areas (Fig. 8a). *Eucalyptus* spp. were dominant in fewer areas and the dominant areas were often surrounded by areas with low occurrence rates (Fig. 8b).

4. Discussion

In this study, we applied a one class classifier to map two potentially invasive species in the Eastern Arc mountain biodiversity hotspot from combined airborne hyperspectral and LiDAR data. In the following, we will first discuss the advantages of the newly suggested model selection approach. Then, we will discuss the identified occurrence patterns of the two target species and draw some potential ecological implications. Finally, we will discuss the suitability of our work-flow in an operational context.

4.1. OCC model optimization and classification

The introduced diagnostic plot (Fig. 6) helped to evaluate the performance of the models. Considering each model at 50 threshold levels revealed the model + threshold (M + T) combinations that had the lowest possible PPP for a given TPR. When min. dist. based M + T selection approach was used, the resulting F-scores (on test set) were higher than for the models selected based on F_{PU} or AUC_{PU} (at threshold zero).

The models selected based on the highest F_{PU} and AUC_{PU} had high precision and low recall. Mack and Waske (2017) have shown earlier that BSVM does not perform well with commonly used threshold selection methods when only zero thresholds are considered. More balanced classification results using these metrics could be achieved by tuning the thresholds manually as suggested by Mack et al. (2014). However, this adds subjectivity to the model selection and requires more in-depth knowledge of the OCC workflow from the person performing the classification.

The introduced min. dist. based combined M + T selection provides an alternative approach that does not require manual tuning of the threshold to achieve balanced results. Also, considering only models that are at the Pareto front reduces the amount of meaningful M + T combinations to consider as possible solutions. Visualizing the Pareto

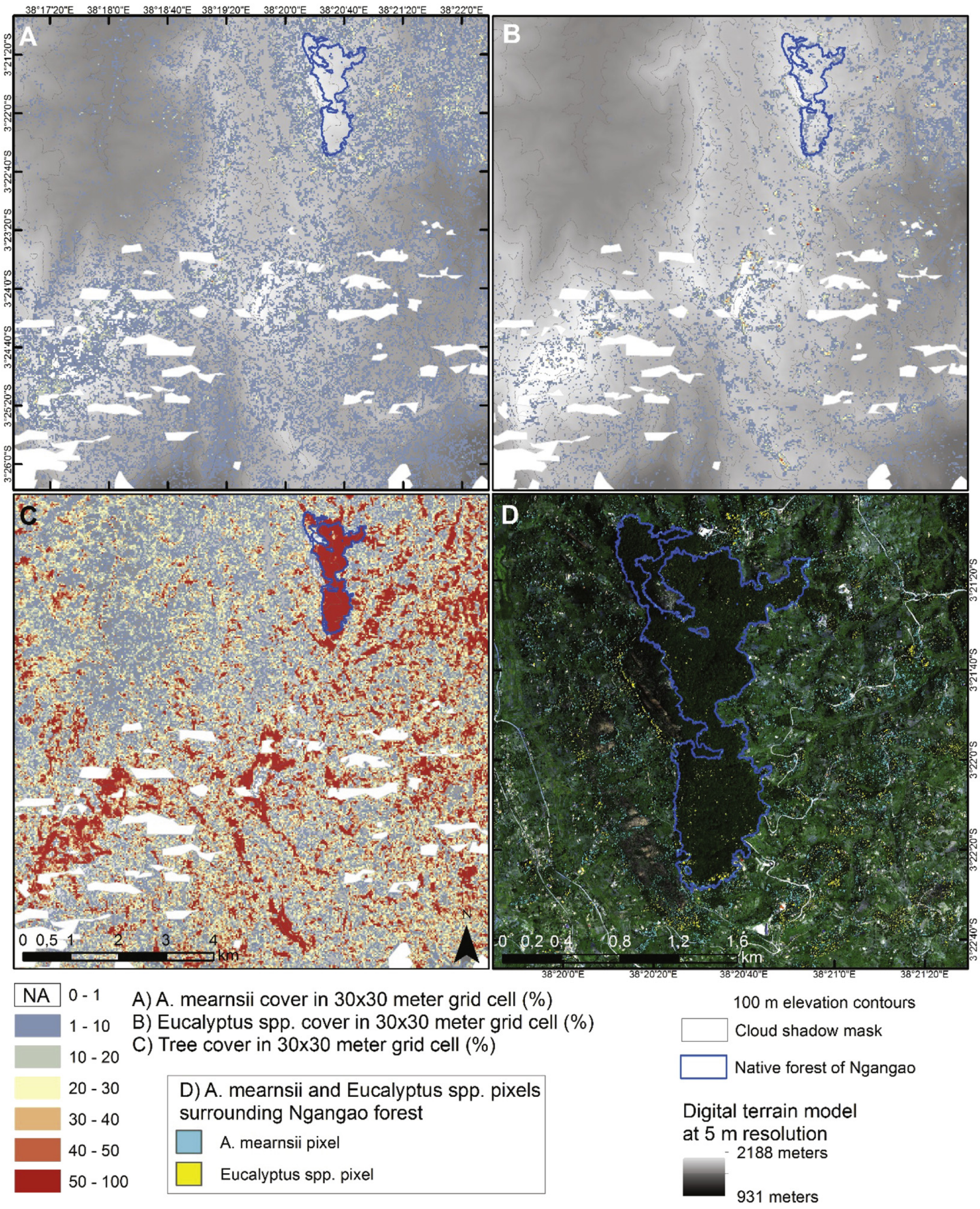

Fig. 7. a) Map of *A. mearnsii*, b) *Eucalyptus* spp. and c) tree cover. The cover is, for example, pixels classified as *A. mearnsii* divided by all pixels in 30 × 30 m grid. d) Ngangao native forest fragment and pixels classified as *A. mearnsii* and *Eucalyptus* spp.

front together with all the possible M + T combinations is a powerful way to understand the potential performance of BSVM in solving classification problems.

As we focus on detecting potentially invasive and harmful tree species to the environment, it may make more sense to minimize the number of false negatives (pixels that are actually *A. mearnsii* or *Eucalyptus* spp., but are classified as negatives). This would ensure that most individuals of invasive species are detected and potential countermeasures against further spreading can be efficiently implemented. On the other hand, there is an obvious trade-off between identifying all individuals of a potentially harmful species and unnecessary and costly fieldwork (people checking false positives). The M + T combination selected based on min. dist. produced balanced results for both species (high F-score). For *Eucalyptus* spp. the precision was very high and the recall was mediocre. For *A. mearnsii* the precision was lower than for the models selected based on F_{PU} or AUC_{PU}, but the recall increased substantially. This also means that there are more false positives for *A. mearnsii*, which in part can explain the higher occurrence rate of this species. If the negative dataset would not be available, these observations could not be analyzed without going into the field. In these circumstances, the M + T combination selected based on min. dist. is a viable option as it does not favor high precision at the cost of low recall like models selected based on F_{PU} or AUC_{PU}.

4.2. The occurrence and threat by Eucalyptus spp. and A. mearnsii

Eucalyptus spp. and *A. mearnsii* covered 0.8% and 1.6% of the study area, respectively. This result should be interpreted with care as the classification is based on very high one meter spatial resolution data. For example, some pixels within a tree crown were misclassified even though the whole crown was classified correctly (> 50% of pixels classified correctly). On the other hand, there were random pixels that

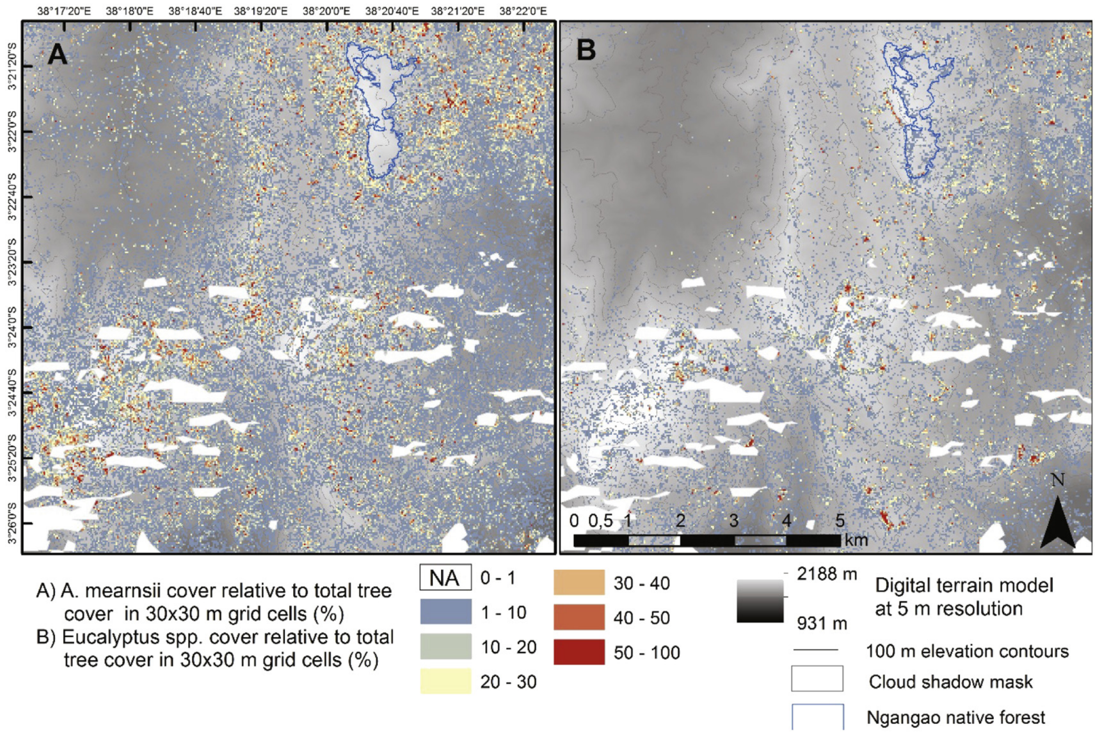

Fig. 8. a) *A. mearnsii* and b) *Eucalyptus* spp. tree cover relative to total tree cover in 30 × 30 m grid cells (%).

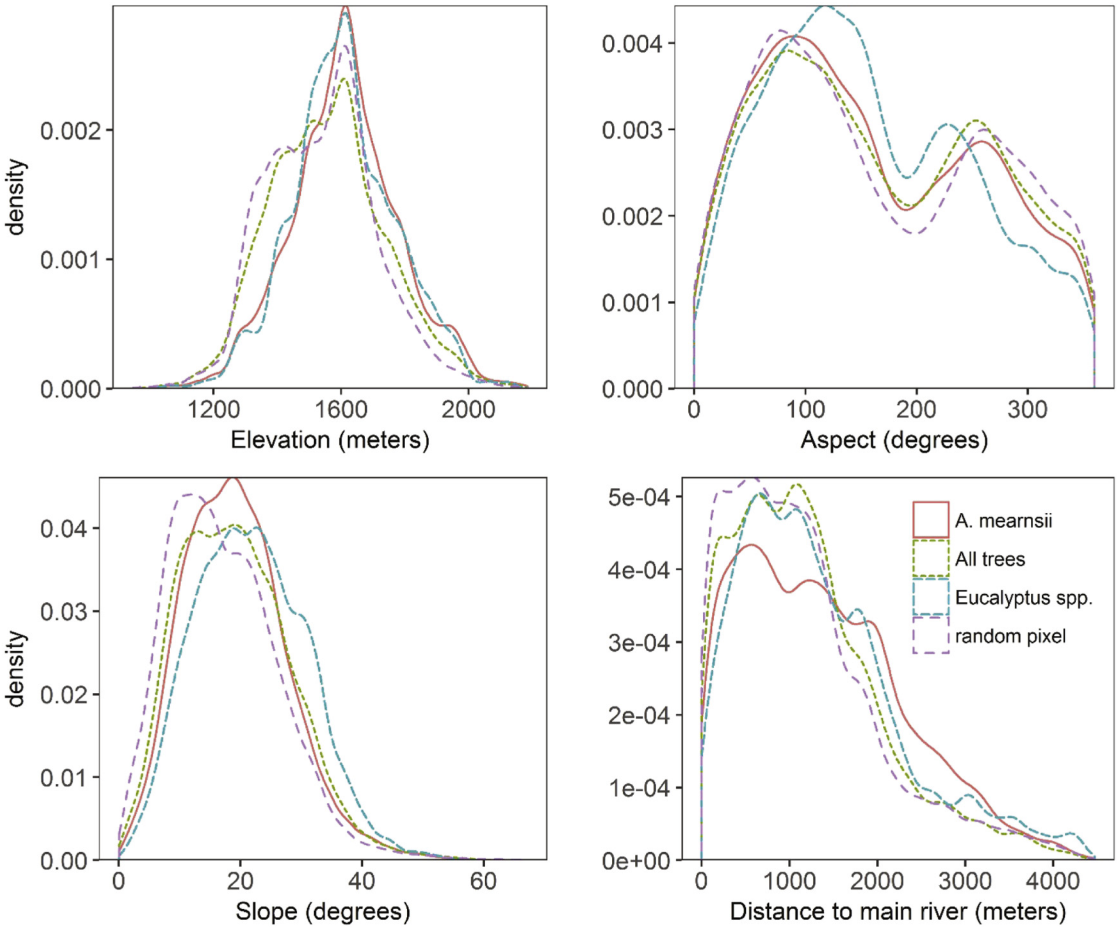

Fig. 9. The occurrence (Kernel densities) of *Eucalyptus* spp., *A. mearnsii*, all trees and a random pixel set for selected environmental factors. Random pixels were randomly sampled from the full study area covering all land cover types.

were falsely classified as positives. To allow for a meaningful interpretation of the occurrence patterns, we created percentage cover maps from the binary occurrence datasets. These cover maps ease the interpretation of the results, as they clearly show the areas with the highest concentrations of the harmful species. Dense *Eucalyptus* spp. stands were often found in proximity to known native forests. Some *Eucalyptus* trees were found also inside the native forests. Although some of those are false positives like in Ngangao forest.

The frequent proximity of *Eucalyptus* trees to native forest stands can be considered problematic as the native forests of Eastern Arc Mountains play a significant role in the provision of ecosystem services (Fisher et al., 2011). For example, they provide habitat for endemic animals and plants. The trees themselves also capture atmospheric moisture through fog deposit (Räsänen et al., submitted), store water in the foliage, epiphytes and trunk, and create infiltration favoring soil type (Cardwell, 2017). They also have the highest density of aboveground carbon stocks in the landscape (Pellikka et al., 2018). The plantation forest with exotic tree species, on contrary, may lower the biodiversity (Bremer and Farley, 2010) and the water table (Rodriguez-Suarez et al., 2011). For instance, *Eucalyptus* spp. are blamed often by local people to be the main reason for unavailability of water resources in the Taita Hills (Hohenthal and Minoia, 2018). Kenya Forest Service has also acknowledged that the *Eucalyptus* spp. plantations are harmful and should be gradually replaced with native, or other less harmful species (Hohenthal and Minoia, 2018).

A. mearnsii trees were found all over the higher altitude areas, which could be explained by its high invasiveness (Lowe et al., 2000) which enable the tree to quickly establish in remote areas where the management influence is low. In the Taita Hills, *A. mearnsii* is known to spread easily on rocky and sandy areas, like on roadsides. However, our mapping results would benefit from further validations as the *A. mearnsii* classification had comparably low precision, and hence a comparably high false positive rate. Similarly, as in the case of *Eucalyptus* spp., our results suggest that *A. mearnsii* can be frequently found close to remaining native forest patches. As *A. mearnsii* infestations have been linked to decreasing biodiversity (Samways et al., 1996) and to negative impacts on soil function and indigenous vegetation growth (Boudiaf et al., 2013), our results suggest a potential ecological threat.

Being a biodiversity hotspot (Burgess et al., 2007), the native forests need to be protected from invasive species and the frequent proximity of both species to the few remaining forest patches may suggest a need for management interventions. Another threat related to both *Eucalyptus* ssp. and *A. mearnsii* is the increased fire risk. Both exotic species catch fire easily (Supplementary Fig. 2). *A. mearnsii* produces large amounts of long-lived seeds that could be triggered after fire (Strydom et al., 2017) helping them to spread to native forests. For *Eucalyptus* spp. regenerative fire is an important factor that reduces competition with other plant species (da Silva et al., 2016). Fire disturbance can also accelerate the naturalization of *Eucalyptus* spp. around the plantation forests of the Taita Hills, as observed in Portugal with *E. globulus* (Águas et al., 2014) and suspected in Brazil (da Silva et al., 2016).

4.3. Operational invasive tree species mapping

We presented a framework to efficiently classify invasive tree species with an OCC algorithm and limited field data. This approach holds potential for operational use, as only limited fieldwork is required. In accordance with our results, Baldeck and Asner (2015) used a OCC approach in mapping savanna tree species in South Africa with good classification accuracies (F-scores 0.4–0.72) comparable to our results. Moreover, Baldeck et al. (2015) obtained very good classification results (recall 0.94–0.97 and precision 0.94–1.0) when applying BSVM for mapping tropical tree species in Panama. However, there were notable differences in the classification setup and the way the validation was conducted compared to our setup. The achieved classification results depend highly on the species that is classified, the tree species diversity in the study area, and the amount of field data available (Alonzo et al., 2013; Feret and Asner, 2013; Piiroinen et al., 2017). Overall, we aimed to build an OCC classification workflow that can be implemented without extensive experience in tree species classification and mapping in the tropics. For instance, segmenting individual tree crowns (ITC) could be done first, but it is known to be challenging in the tropics (Feret and Asner, 2013; Piiroinen et al., 2017). Automatic delineation of ITCs also adds complexity to the classification workflow, while very good results have been achieved also with pixel-based classification approaches (Baldeck et al., 2015) which are easier to implement, and were hence applied here.

The newly introduced diagnostic plot for the OCC (Fig. 6) helps in evaluating the potential performance of BSVM and the min. dist. based M + T selection approach provides a straightforward way to select the model and the threshold without the need to tune the threshold manually to achieve sensible results. However, further research should be conducted to test this approach with other datasets where PN-data is available to draw more robust conclusions on the generalizability of our findings. Nevertheless, our results suggest that this OCC classification approach has potential in mapping tree species in high species diversity systems when there is interest in only one or a few key species. The mapping results could be used in the Taita Hills for managing protection measures for the benefit of native forests, while the same method could be used elsewhere in East Africa and globally for invasive species mapping.

In an operational setting, OCC results without N-data might serve also as an initial map product for directing the fieldwork. The initial classification could be generated with only a few observations of the species of interest. The results could be used to locate areas with a high occurrence probability of the species and hence increase the efficiency in subsequent field campaigns to collect presence data. This process could be iterated and adjusted to locate all species of interest.

A strong limitation of the approach, as presented in this study, is that it relies on relatively expensive airborne RS data. Alternative options have recently been presented by, for example, Kganyago et al. (2018) who mapped invasive tree species in KwaZulu-Natal, South-Africa using Landsat and SPOT data and a supervised classification approach. This indicates that certain invasive tree species could be identified with satellite-based data. In another recent study, Ng et al. (2017) presented promising results for mapping the invasive *Prosopis* spp. and the indigenous *Vachellia* spp. trees in Baringo, Kenya using Sentinel-2 data. In future studies, the introduced OCC approach should be tested with these satellite-based data sources. Ideally, airborne IS-based classification would be compared with the satellite-data based classification in the same study to draw further conclusions on the performances of these two data sources in invasive tree species mapping in tropics. Our occurrence maps could even serve as training data for the Sentinel-2 based analysis. Another possible adaptation to the suggested OCC approach in operational tree species mapping could be to combine satellite data with data acquired from UAVs. The high spatial resolution UAV imagery could be an efficient way to collect presence data on species of interest through manual image interpretation. These data could then be used for training OCC algorithms applied to Sentinel-2 data. Required fieldwork would be minimal and relatively cheap as UAV data acquired with a normal RGB camera is likely to suffice to reliably identify *Eucalyptus* spp. and *A. mearnsii* through manual image interpretation.

5. Conclusions

This study showed how a BSVM classifier can be used to detect and map common invasive tree species in the Taita Hills, Kenya. In areas where tree species diversity is very high, the terrain is rugged and infrastructure is poor, the OCC approach is useful as labeled training data is needed only for the positive class. The newly suggested diagnostic plot and the minimum distance to the upper left corner (of the diagnostic plot) based model + threshold selection approach makes it easier for an unexperienced user to apply OCC in RS case studies. The amount of manual work is reduced, compared to approaches that require manual tuning of the threshold level to achieve good results or a very large grid search of potential parameters that would increase the computational costs.

A. mearnsii was found throughout the higher altitudes of the study area, which suggests possibly invasive behavior in the Taita Hills, Kenya. However, we did not have information about whether the trees have been planted on purpose or not. *Eucalyptus* spp. were found especially in the higher altitudes and steeper slopes, but it was not as widely spread as *A. mearnsii*. Further assessments of its invasiveness and impact on the local ecosystem is needed. Generally, the two species are very common in the study area. They were found in large quantities close to the biggest remaining native forest patch (Ngangao) in the study area. This highlights the need to monitor the occurrence of these two species as they might spread even more and endanger the last remaining native forest patches in the Taita Hills. Finally, this study provides valuable information for officials to take action in controlling these potentially harmful and invasive tree species, especially in sites that hold great ecological value.

Acknowledgments

The authors would like to acknowledge the funding from the Ministry for Foreign Affairs of Finland for CHIESA and BIODEV projects, from Academy of Finland for TAITAWATER and TAITASMART projects, and Finnish Culture Foundation for Petri Pellikka as a PI. The authors would also like to acknowledge everyone who participated in collecting the field data and the remote sensing data and in data preprocessing, including Jessica Broas, Jesse Hietanen, Elisa Schäfer, Kirsi Kivistö, Arto Viinikka and Hari Adhikari. The Taita Research Station of the University of Helsinki provided logistics, accommodation and field assistants, from who we would like to mention Darius Kimuzi, Francis Namisiko and Mwadime Mjomba. The research was carried out with a kind permission from National Council for Science and Technology of Republic of Kenya (NACOSTI: NCST/RCD/17/012/33). Eduardo Maeda acknowledges funding from the Academy of Finland (decision numbers: 318252 and 319905).

Conflicts of interest

The authors declare no conflict of interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.rse.2018.09.018.

References

Adhikari, H., Heiskanen, J., Siljander, M., Maeda, E., Heikinheimo, V., Pellikka, P.K.E., 2017. Determinants of aboveground biomass across an afromontane landscape mosaic in Kenya. Remote Sens. 9, 1–19. https://doi.org/10.3390/rs9080827.

Aerts, R., Thijs, K.W., Lehouck, V., Beentje, H., Bytebier, B., Matthysen, E., Gulinck, H., Lens, L., Muys, B., 2011. Woody plant communities of isolated Afromontane cloud forests in Taita Hills, Kenya. Plant Ecol. 212, 639–649. https://doi.org/10.1007/s11258-010-9853-3.

Águas, A., Ferreira, A., Maia, P., Fernandes, P.M., Roxo, L., Keizer, J., Silva, J.S., Rego, F.C., Moreira, F., 2014. Natural establishment of Eucalyptus globulus Labill. In burnt stands in Portugal. For. Ecol. Manag. 323, 47–56. https://doi.org/10.1016/j.foreco.2014.03.012.

Alonzo, M., Roth, K., Roberts, D., 2013. Identifying Santa barbara's urban tree species from AVIRIS imagery using canonical discriminant analysis. Remote Sens. Lett. 4, 513–521. https://doi.org/10.1080/2150704X.2013.764027.

Baldeck, C., Asner, G., 2015. Single-species detection with airborne imaging spectroscopy data: acomparison of support vector techniques. Sel. Top. Appl. Earth Obs. Remote Sensing, IEEE J. 8, 2501–2512. https://doi.org/10.1109/JSTARS.2014.2346475.

Baldeck, C.A., Asner, G.P., Martin, R.E., Anderson, C.B., Knapp, E., Kellner, J.R., Wright, S.J., 2015. Operational tree species mapping in a diverse tropical forest with airborne imaging spectroscopy. PLoS One 10, 1–21. https://doi.org/10.1371/journal.pone.0118403.

Boudiaf, I., Baudoin, E., Sanguin, H., Beddiar, A., Thioulouse, J., Galiana, A., Prin, Y., Le Roux, C., Lebrun, M., Duponnois, R., 2013. The exotic legume tree species, Acacia mearnsii, alters microbial soil functionalities and the early development of a native tree species, Quercus suber, in North Africa. Soil Biol. Biochem. 65, 172–179. https://doi.org/10.1016/j.soilbio.2013.05.003.

Bremer, L.L., Farley, K.A., 2010. Does plantation forestry restore biodiversity or create green deserts? A synthesis of the effects of land-use transitions on plant species richness. Biodivers. Conserv. 19, 3893–3915. https://doi.org/10.1007/s10531-010-9936-4.

Burgess, N.D., Butynski, T.M., Cordeiro, N.J., Doggart, N.H., Fjeldså, J., Howell, K.M., Kilahama, F.B., Loader, S.P., Lovett, J.C., Mbilinyi, B., Menegon, M., Moyer, D.C., Nashanda, E., Perkin, a, Rovero, F., Stanley, W.T., Stuart, S.N., 2007. The biological importance of the Eastern Arc Mountains of Tanzania and Kenya. Biol. Conserv. 134, 209–231. https://doi.org/10.1016/j.biocon.2006.08.015.

Cardwell, A., 2017. Master's thesis: the effect of land use on infiltration in Taita Hills, Kenya. In: University of Helsinki, . http://urn.fi/URN:NBN:fi-fe2017112251801.

Chenje, M., Mohamed-Katerere, J., 2006. Invasive alien species. In: Africa Environment Outlook 2. United Nations Environment Programme, Nairobi, pp. 1–542.

Cho, M.A., Mathieu, R., Asner, G.P., Naidoo, L., van Aardt, J., Ramoelo, A., Debba, P., Wessels, K., Main, R., Smit, I.P.J., Erasmus, B., 2012. Mapping tree species composition in South African savannas using an integrated airborne spectral and LiDAR system. Remote Sens. Environ. 125, 214–226. https://doi.org/10.1016/j.rse.2012.07.010.

Clark, B., Pellikka, P., 2009. Landscape analysis using multi-scale segmentation and objectoriented classification. In: Roeder, A., Hill, J. (Eds.), Recent Advances in Remote Sensing and Geoinformation Processing for Land Degradation Assessment. Taylor & Francis Group, London, pp. 323–341.

da Silva, P.H.M., Bouillet, J.P., de Paula, R.C., 2016. Assessing the invasive potential of commercial eucalyptus species in Brazil: germination and early establishment. For. Ecol. Manag. 374, 129–135. https://doi.org/10.1016/j.foreco.2016.05.007.

Engeman, R.M., Sugihara, R.T., Pank, L.F., Dusenberry, W.E., 1994. A comparison of plotless density estimators using Monte Carlo simulation. Ecology 75, 1769–1779. https://doi.org/10.2307/1939636.

Fassnacht, F.E., Neumann, C., Forster, M., Buddenbaum, H., Ghosh, A., Clasen, A., Joshi, P.K., Koch, B., 2014. Comparison of feature reduction algorithms for classifying tree species with hyperspectral data on three central European test sites. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 7, 2547–2561. https://doi.org/10.1109/JSTARS.2014.2329390.

Fassnacht, F.E., Latifi, H., Sterenczak, K., Modzelewska, A., Lefsky, M., Waser, L.T., Straub, C., Ghosh, A., 2016. Review of studies on tree species classification from remotely sensed data. Remote Sens. Environ. 186, 64–87. https://doi.org/10.1016/j.rse.2016.08.013.

Fawcett, T., 2006. An introduction to ROC analysis. Pattern Recogn. Lett. 27, 861–874. https://doi.org/10.1016/j.patrec.2005.10.010.

Feret, J.-B., Asner, P.G., 2013. Tree species discrimination in tropical forests using airborne imaging spectroscopy. IEEE Trans. Geosci. Remote Sens. 51, 73–84. https://doi.org/10.1109/TGRS.2012.2199323.

Fisher, B., Turner, R.K., Burgess, N.D., Swetnam, R.D., Green, J., Green, R.E., Kajembe, G., Kulindwa, K., Lewis, S.L., Marchant, R., Marshall, A.R., Madoffe, S., Munishi, P.K.T., Morse-Jones, S., Mwakalila, S., Paavola, J., Naidoo, R., Ricketts, T., Rouget, M., Willcock, S., White, S., Balmford, A., 2011. Measuring, modeling and mapping ecosystem services in the Eastern Arc Mountains of Tanzania. Prog. Phys. Geogr. 35, 595–611. https://doi.org/10.1177/0309133311422968.

Graves, S.J., Asner, G.P., Martin, R.E., Anderson, C.B., Colgan, M.S., Kalantari, L., Bohlman, S.A., 2016. Tree species abundance predictions in a tropical agricultural landscape with a supervised classification model and imbalanced data. Remote Sens. 8, 1–21. https://doi.org/10.3390/rs8020161.

Green, A., Berman, M., Switzer, P., Craig, M.D., 1988. A transformation for ordering multispectral data in terms of image quality with implications for noise removal. IEEE Trans. Geosci. Remote Sens. 26, 65–74.

Habibzadeh, F., Habibzadeh, P., Yadollahie, M., 2016. On determining the most appropriate test cut-off value: the case of tests with continuous results. Biochem. Med. 26, 297–307. https://doi.org/10.11613/BM.2016.034.

Hijmans, R.J., 2016. raster: Geographic Data Analysis and Modeling.

Hohenthal, J., Minoia, P., 2018. Political ecology of asymmetric ecological knowledges: diverging views on the eucalyptus-water nexus in the Taita Hills, Kenya. J. Polit. Ecol. 25 (1), 19.

Horn, B.K.P., 1981. Hill shading and the reflectance map. Proc. IEEE 69, 14–47. https://doi.org/10.1109/PROC.1981.11918.

Kganyago, M., Odindi, J., Adjorlolo, C., Mhangara, P., 2018. Evaluating the capability of Landsat 8 OLI and SPOT 6 for discriminating invasive alien species in the African Savanna landscape. Int. J. Appl. Earth Obs. Geoinf. 67, 10–19. https://doi.org/10.1016/j.jag.2017.12.008.

Lee, W.S., Liu, B., 2003. Learning with positive and unlabeled examples using weighted logistic regression. Proc. Twent. Int. Conf. Mach. Learn. 3, 448–455. https://doi.org/10.1016/j.tcs.2005.09.007.

Liu, B., Dai, Y., Li, X., Lee, W.S., Yu, P., 2003. Building text classifiers using positive and unlabeled examples. In: Third IEEE International Conference on Data Mining. Melbourne, pp. 179–186. https://doi.org/10.1002/cpe.3879.

Lobo, J.M., Jiménez-valverde, A., Real, R., 2008. AUC: a misleading measure of the performance of predictive distribution models. Glob. Ecol. Biogeogr. 17, 145–151. https://doi.org/10.1111/j.1466-8238.2007.00358.x.

Lowe, S., Browne, M., Boudjelas, S., De Poorter, M., 2000. 100 of the World's Worst

Invasive Alien Species a Selection From the Global Invasive Species Database. Invasive Species Spec. Gr. (ISSG). A Spec. Gr. Species Surviv. Comm. or World Conserv. Union (IUCN). 12https://doi.org/10.1614/WT-04-126.1.

Mack, B., 2015. oneClass: One-Class Classification in the Absence of test Data, Version 0.1-1: Software.

Mack, B., Waske, B., 2017. In-depth comparisons of MaxEnt, biased SVM and one-class SVM for one-class classification of remote sensing data. Remote Sens. Lett. 8, 290–299. https://doi.org/10.1080/2150704X.2016.1265689.

Mack, B., Roscher, R., Waske, B., 2014. Can I trust my one-class classification? Remote Sens. 6, 8779–8802. https://doi.org/10.3390/rs6098779.

Mountrakis, G., Im, J., Ogole, C., 2011. Support vector machines in remote sensing: a review. ISPRS J. Photogramm. Remote Sens. 66, 247–259. https://doi.org/10.1016/j.isprsjprs.2010.11.001.

Mũnoz-Marí, J., Bovolo, F., Gómez-Chova, L., Bruzzone, L., Camp-Valls, G., 2010. Semisupervised one-class support vector machines for classification of remote sensing data. IEEE Trans. Geosci. Remote Sens. 48, 3188–3197. https://doi.org/10.1109/TGRS.2010.2045764.

Ng, W.T., Rima, P., Einzmann, K., Immitzer, M., Atzberger, C., Eckert, S., 2017. Assessing the potential of sentinel-2 and pléiades data for the detection of prosopis and vachellia spp. in Kenya. Remote Sens. 9. https://doi.org/10.3390/rs9010074.

Nyoka, B., 2003. Biosecurity in forestry: a case study on the status of invasive forest tree species in Southern Africa. In: Forest Biosecurity Working Paper FBS/1E. Rome.

Pellikka, P.K.E., Lötjönen, M., Siljander, M., Lens, L., 2009. Airborne remote sensing of spatiotemporal change (1955-2004) in indigenous and exotic forest cover in the Taita Hills, Kenya. Int. J. Appl. Earth Obs. Geoinf. 11, 221–232. https://doi.org/10.1016/j.jag.2009.02.002.

Pellikka, P.K.E., Clark, B.J.F., Gosa, A.G., Himberg, N., Hurskainen, P., Maeda, E., Mwang'ombe, J., Omoro, L.M. a, Siljander, M., 2013. Agricultural expansion and its consequences in the Taita Hills, Kenya. Dev. Earth Surf. Process. 16, 165–179. https://doi.org/10.1016/B978-0-444-59559-1.00013-X.

Pellikka, P.K.E., Heikinheimo, V., Hietanen, J., Schäfer, E., Siljander, M., Heiskanen, J., 2018. Impact of land cover change on aboveground carbon stocks in Afromontane landscape in Kenya. Appl. Geogr. 94, 178–189. https://doi.org/10.1016/j.apgeog.2018.03.017.

Persello, C., Bruzzone, L., 2009. A novel approach to the selection of spatially invariant features for classification of hyperspectral images. In: 2009 IEEE Int. Geosci. Remote Sens. Symp. 2. pp. 3180–3190. https://doi.org/10.1109/IGARSS.2009.5418001.

Phillips, S.J., Dudík, M., 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography (Cop.) 31, 161–175. https://doi.org/10.1111/j.0906-7590.2008.5203.x.

Phillips, S.J., Anderson, R.P., Schapire, R.E., 2006. Maximum entropy modeling of species geographic distributions. Ecol. Model. 190, 231–259. https://doi.org/10.1016/j.ecolmodel.2005.03.026.

Piiroinen, R., Heiskanen, J., Mõttus, M., Pellikka, P., 2015. Classification of crops across heterogeneous agricultural landscape in Kenya using AisaEAGLE imaging spectroscopy data. Int. J. Appl. Earth Obs. Geoinf. 39, 1–8. https://doi.org/10.1016/j.jag.2015.02.005.

Piiroinen, R., Heiskanen, J., Maeda, E., Viinikka, A., Pellikka, P., 2017. Classification of tree species in a diverse African Agroforestry landscape using imaging spectroscopy and laser scanning. Remote Sens. 9, 1–20. https://doi.org/10.3390/rs9090875.

Räsänen, M., Katurji, M., Pellikka, P., Rinne, J., Katul, G.G., 2018. Intermittency and SOC Scaling for Rainfall and Fog Deposition at a Tropical Cloud Forest. (submitted).

Richardson, D.M., Rejmánek, M., 2011. Trees and shrubs as invasive alien species - a global review. Divers. Distrib. 17, 788–809. https://doi.org/10.1111/j.1472-4642.2011.00782.x.

Richardson, D.M., Pyšek, P., Rejmánek, M., Barbour, M.G., Dane Panetta, F., West, C.J., 2000. Naturalization and invasion of alien plants: concepts and definitions. Divers. Distrib. 6, 93–107. https://doi.org/10.1046/j.1472-4642.2000.00083.x.

Richter, R., Schläpfer, D., 2002. Geo-atmospheric processing of airborne imaging spectrometry data. Part 2: atmospheric/topographic correction. Int. J. Remote Sens. 23, 2631–2649. https://doi.org/10.1080/01431160110115834.

Rodriguez-Suarez, J.A., Soto, B., Perez, R., Diaz-Fierros, F., 2011. Influence of Eucalyptus globulus plantation growth on water table levels and low flows in a small catchment. J. Hydrol. 396, 321–326. https://doi.org/10.1016/j.jhydrol.2010.11.027.

Rouse, J.W., Haas, R.H., Schell, J.A., Deering, D.W., 1973. Monitoring vegetation Systems in the Great Okains with ERTS. Third Earth Resour. Technol. Satell. Symp. 1, 325–333 (https://doi.org/10/citeulike-article-id:12009708).

RSI, 2004. ENVI User's Guide. Research Systems, Inc., Boulder.

Samways, M.J., Caldwell, P.M., Osborn, R., 1996. Ground-living invertebrate assemblages in native, planted and invasive vegetation in South Africa. Agric. Ecosyst. Environ. 59, 19–32. https://doi.org/10.1016/0167-8809(96)01047-X.

Schäfer, E., Heiskanen, J., Heikinheimo, V., Pellikka, P., 2016. Mapping tree species diversity of a tropical montane forest by unsupervised clustering of airborne imaging spectroscopy data. Ecol. Indic. 64, 49–58. https://doi.org/10.1016/j.ecolind.2015.12.026.

Scholkopf, B., Williamson, R., Smola, A., Shawe-Taylor, J., Platt, J., 1999. Support vector method for novelty detection. In: NIPS'99 Proc. 12th Int. Conf. Neural Inf. Process. Syst, pp. 582–588(10.1.1.71.4642).

Silva, E.R., Lazarotto, D.C., Schwambach, J., Overbeck, G.E., Soares, G.L.G., 2017. Phytotoxic effects of extract and essential oil of Eucalyptus saligna (Myrtaceae) leaf litter on grassland species. Aust. J. Bot. 65, 172–182. https://doi.org/10.1071/BT16254.

Skowronek, S., Asner, G.P., Feilhauer, H., 2017a. Performance of one-class classifiers for invasive species mapping using airborne imaging spectroscopy. Eco. Inform. 37, 66–76. https://doi.org/10.1016/j.ecoinf.2016.11.005.

Skowronek, S., Ewald, M., Isermann, M., Van De Kerchove, R., Lenoir, J., Aerts, R., Warrie, J., Hattab, T., Honnay, O., Schmidtlein, S., Rocchini, D., Somers, B., Feilhauer, H., 2017b. Mapping an invasive bryophyte species using hyperspectral remote sensing data. Biol. Invasions 19, 239–254. https://doi.org/10.1007/s10530-016-1276-1.

Stenzel, S., Feilhauer, H., Mack, B., Metz, A., Schmidtlein, S., 2014. Remote sensing of scattered natura 2000 habitats using a one-class classifier. Int. J. Appl. Earth Obs. Geoinf. 33, 211–217. https://doi.org/10.1016/j.jag.2014.05.012.

Stenzel, S., Fassnacht, F.E., Mack, B., Schmidtlein, S., 2017. Identification of high nature value grassland with remote sensing and minimal field data. Ecol. Indic. 74, 28–38. https://doi.org/10.1016/j.ecolind.2016.11.005.

Strydom, M., Veldtman, R., Ngwenya, M.Z., Esler, K.J., 2017. Invasive Australian acacia seed banks: size and relationship with stem diameter in the presence of gall-forming biological control agents. PLoS One 12, 1–16. https://doi.org/10.1371/journal.pone.0181763.

Thijs, K.W., Aerts, R., van de Moortele, P., Aben, J., Musila, W., Pellikka, P., Gulinck, H., Muys, B., 2015. Trees in a human-modified tropical landscape: species and trait composition and potential ecosystem services. Landsc. Urban Plan. 144, 49–58. https://doi.org/10.1016/j.landurbplan.2015.07.015.

Turpie, J.K., Marais, C., Blignaut, J.N., 2008. The working for water programme: evolution of a payments for ecosystem services mechanism that addresses both poverty and ecosystem service delivery in South Africa. Ecol. Econ. 65, 788–798. https://doi.org/10.1016/j.ecolecon.2007.12.024.

Vapnik, V., 1998. Statistical Learning Theory. John Wiley & Sons, New York.

Wakie, T.T., Evangelista, P.H., Jarnevich, C.S., Laituri, M., 2014. Mapping current and potential distribution of non-native prosopis juliflorain the Afar region of Ethiopia. PLoS One 9, 3–11. https://doi.org/10.1371/journal.pone.0112854.