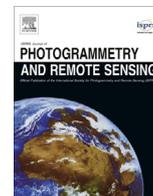




Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

Detecting *Sirex noctilio* grey-attacked and lightning-struck pine trees using airborne hyperspectral data, random forest and support vector machines classifiers



Elfatih M. Abdel-Rahman^{a,b,*}, Onesimo Mutanga^a, Elhadi Adam^a, Riyad Ismail^a

^a School of Agriculture, Earth and Environmental Sciences, Pietermaritzburg Campus, University of KwaZulu-Natal, Scottsville P/Bag X01, Pietermaritzburg 3209, South Africa

^b Department of Agronomy, Faculty of Agriculture, University of Khartoum, Khartoum North 13314, Sudan

ARTICLE INFO

Article history:

Received 22 May 2013

Received in revised form 24 November 2013

Accepted 27 November 2013

Available online 20 December 2013

Keywords:

Sirex grey stage

Lightning damage

Pinus spp.

Hyperspectral data

Random forest

Support vector machines

ABSTRACT

The visual progression of siren (*Sirex noctilio*) infestation symptoms has been categorized into three distinct infestation phases, namely the green, red and grey stages. The grey stage is the final stage which leads to almost complete defoliation resulting in dead standing trees or snags. Dead standing pine trees however, could also be due to the lightning damage. Hence, the objective of the present study was to distinguish amongst healthy, siren grey-attacked and lightning-damaged pine trees using AISA Eagle hyperspectral data, random forest (RF) and support vector machines (SVM) classifiers. Our study also presents an opportunity to look at the possibility of separating amongst the previously mentioned pine trees damage classes and other landscape classes on the study area. The results of the present study revealed the robustness of the two machine learning classifiers with an overall accuracy of 74.50% (total disagreement = 26%) for RF and 73.50% (total disagreement = 27%) for SVM using all the remaining AISA Eagle spectral bands after removing the noisy ones. When the most useful spectral bands as measured by RF were exploited, the overall accuracy was considerably improved; 78% (total disagreement = 22%) for RF and 76.50% (total disagreement = 24%) for SVM. There was no significant difference between the performances of the two classifiers as demonstrated by the results of McNemar's test (chi-squared; $\chi^2 = 0.14$, and 0.03 when all the remaining ASIA Eagle wavebands, after removing the noisy ones and the most important wavebands were used, respectively). This study concludes that AISA Eagle data classified using RF and SVM algorithms provide relatively accurate information that is important to the forest industry for making informed decision regarding pine plantations health protocols.

© 2013 International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS) Published by Elsevier B.V. All rights reserved.

1. Introduction

The invasive woodwasp, *Sirex noctilio* Fabricius (Hymenoptera, Siricidae), has caused widespread damage to South African commercial *Pinus* species. The parasite's highest impact is in the province of KwaZulu-Natal with a 6% infestation rate (Hurley et al., 2007). Mitigation measurements introduced in KwaZulu-Natal have to date yielded inaccurate *S. noctilio* management strategies (Hurley et al., 2008). Consequently, detection and monitoring of the wasp is still considered a vital protocol in the identification of remediation sites. Currently, methods implemented for the spatial identification and quantification of *S. noctilio* infestations include broad scale visual aerial reconnais-

sance (Carnegie et al., 2005) and field based enumerations (Ismail et al., 2006). Researchers have recently advocated utilizing remotely-sensed data for the detection and mapping of damage caused by *S. noctilio* (Dye et al., 2008; Ismail et al., 2007, 2008). These studies focused on correlating the visual symptoms of *S. noctilio* infestations with high spectral and spatial resolution remotely-sensed data.

In line with other studies, the visual progression of *S. noctilio* infestation symptoms has been categorized into three distinct infestation phases, namely the green, red and grey stages (Coops et al., 2006; Ismail et al., 2007). The initial green stage is evidenced by the appearance of resin droplets on the trunk (Ciesla, 2003; Tribe and Cillié, 2004) and the tree canopy still appears green and healthy and there is minimal needle loss. The red stage is characterized by turning the color of the tree canopy from green to reddish brown (Corley et al., 2007). The final stage is referred to as the grey stage. Round exit holes appear on the trunk and needle fall occurs, leading to almost complete defoliation resulting in a

* Corresponding author at: School of Agriculture, Earth and Environmental Sciences, Pietermaritzburg Campus, University of KwaZulu-Natal, Scottsville P/Bag X01, Pietermaritzburg 3209, South Africa. Tel.: +27 738799363; fax: +27 33 2605344.

E-mail address: elfatihabdelrahman@gmail.com (E.M. Abdel-Rahman).

dead standing tree or snag (Neumann and Minko, 1981). Existing *S. noctilio* damage detection and mapping studies utilizing remotely-sensed data have focused primarily on the green and red damage stages (Dye et al., 2008; Ismail et al., 2006, 2007). However, there are operational limitations that restrict the successful detection and mapping of green and red stages of attack. Coutts (1969) showed that *Pinus radiata* trees can start to transform from the healthy to the red stage within two weeks after the wasp introduces mucus and fungus into the tree. Subsequently there is needle loss with complete senescence and splitting of the bark in *P. radiata* within three months (Spradbery, 1973). This relatively small bio-window makes the acquisition and processing of the remotely-sensed data required to quantify the total mortality associated with *S. noctilio* infestations challenging. Additionally, *S. noctilio* emerges from infested trees between October and January peaking in November in KwaZulu-Natal (Hurley et al., 2008), with chlorosis visible in late December and early January. The spring and summer period for the eastern parts of South Africa is characterized by heavy rainfall, particularly between November and February (Schulze and Maharaj, 1997; Schulze et al., 1997) making the acquisition of cloud free remotely-sensed data difficult. Consequently, this study proposes mapping the grey stage of infestation in order to provide a more complete quantification of the damage caused by the wasp. It is envisaged that grey stage mapping will overcome current operational limitations because (i) the grey stage of infestation is prevalent over a longer period of time (i.e. from May till September) corresponding with the drier winter months, thus allowing for relatively easier acquisition of cloud free imagery, (ii) and more accurate correspondence between the field surveys and image data is possible, and (iii) tree damage will also be easier to associate symptomatically with the wasp due to the presence of exit holes after emergence, which is more readily visible than resin droplets and ovipositors required for the field verification of the green stage of infestation.

However, detection and mapping of pine trees of grey stage of attack is not without challenges because snags could also be a result of other pine damaging agents. For example, pine trees struck by lightning show similar visual symptoms of damage to grey stage trees. From a pest management perspective there is a need to differentiate between pine trees that are damaged by sirex and those struck by lightning.

According to Marcus et al. (2002) further investigation into the identification of dead wood, using hyperspectral and hyperspatial remotely-sensed data is required. Unlike the broadband multispectral data, the narrow band contiguous hyperspectral data provide information about vegetation biochemical makeup and can therefore be useful in detecting any biophysiological changes in a tree foliage or wood due to abiotic and/or biotic stressing factor (Goetz, 2009; Kumar et al., 2003). On the other hand, the increased spatial resolution of aerial imagery offers more finely spatial detail of potential use in tree-level studies (Meddens et al., 2011). Pasher and King (2009) utilized fine spatial multispectral aerial image data for mapping dead wood in natural forest. They successfully detected canopy-level dead wood objects with an accuracy of 94%. Notwithstanding, the study of Marcus et al. (2002) was unsuccessful to map forest woody debris because the multispectral aerial image employed could not provide the ideal spatial resolution for distinguishing the narrow woods from their surrounding objects.

The sheer volume and basic nature of hyperspectral data has necessitated the development of specific hyperspectral processing procedures (Lillesand and Kiefer, 2001). While numerous supervised classification methods have been developed and successfully implemented on multispectral data, these methods are not effective when dealing with hyperspectral data sets due to the high number of spectral bands. The main difficulty with hyperspectral data is the over-fitting “Hughes phenomenon” also known as the

curse of dimensionality (Hughes, 1968). With hyperspectral data, as the number of spectral bands increases, the size of the required training samples for a specific classifier increases exponentially as well (Hsu, 2007). Recently, a number of machine learning approaches have been utilized to analyze the multidimensional hyperspectral data. Random forest (RF) (Breiman, 2001) and support vector machines (SVM) (Cortes and Vapnik, 1995) are well known machine learning algorithms that perform excellent in reducing the complexity of ill-posed classification problems associated with hyperspectral data (Adam et al., 2012; Bandos et al., 2009; Plaza et al., 2009; Sesnie et al., 2010; Waske et al., 2009). It has been argued that RF and SVM are both suited to hyperspectral image classification as these algorithms can deal with large input spaces, handle noisy datasets efficiently and produce fair classification accuracies (Camps-Valls and Bruzzone, 2005). The performance of the two classifiers has been compared for classifying various land forms using hyperspectral data (e.g., Pal, 2005; Sesnie et al., 2010; Waske et al., 2009). Different results were obtained and the superiority of either RF or SVM was controversial. To the best of our knowledge no study has compared the capabilities of RF and SVM for detecting pine snags. The evaluation of the performance of the two classifiers for identifying pine trees of sirex grey infestation and those struck by lightning is needed. Hence, the objective of the present study was to examine the utility of RF and SVM for discriminating between sirex grey stage and lightning damaged pine trees using AISA Eagle hyperspectral data. Our study also presents an opportunity to look at the possibility of separating sirex and lightning damage classes and other landscape classes in the study area, namely *Eucalyptus spp.*, *Acacia spp.*, bugweed, bare soil, and shadow.

2. Methodology

2.1. Study area

The study was conducted at the Hodgsons Sappi plantation area (latitude 29.227°S and longitude 30.499°E) near Greytown which is located in the province of KwaZulu-Natal, South Africa (Fig. 1). Hodgsons is a mountainous area with elevation ranges from 1030 to 1590 m above sea level. The soil in the area is dominated by apedal and plinthic classes from the eccla group (Sappi, 1993). The rainy season is during summer (October–February) with annual rainfall ranging from 730 to 1280 mm. As described by Rutherford et al. (2006), the Hodgsons plantation is located within the midlands mistbelt grassland bioregion of South Africa. Thus, a significant amount of additional moisture is present in the area. The average annual temperature is 15.8 °C. Various *Acacia*, *Eucalyptus* and *Pinus* trees are present in the study area. However, the dominant species consists of *Pinus patula* trees that are mainly grown for pulpwood. Surrounding areas are mainly dominated by Ngongoni veld of the Natal mist-belt and Southern tall grassveld (Sappi, 1993). Alien invasive plants are a serious problem in the area, with bugweed infestations common within the planted compartments (Dobyn, 2009; So et al., 2002).

2.2. Image acquisition and pre-processing

AISA (Airborne Imaging System for different Applications) hyperspectral image was acquired for the study area on the 11th of March, 2009 using the pushbroom Eagle sensor. The image was acquired during sunny, low wind and clear sky day conditions at 11:10 am (South Africa local time) at a mean geographical position system (GPS) altitude of ca 2.728 km. The image spatial resolution was about 2 m and there were 272 spectral bands ranging

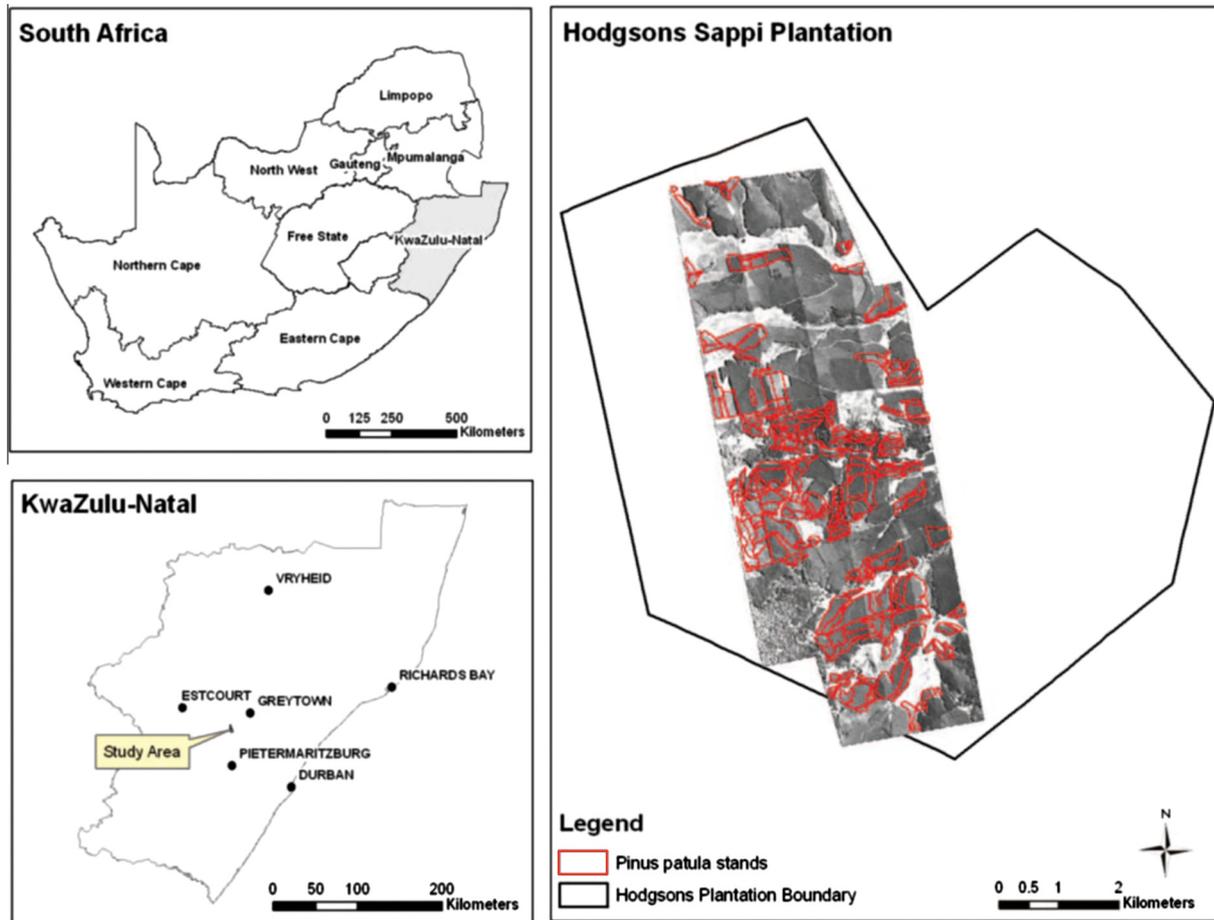


Fig. 1. Location of study area in KwaZulu-Natal (KZN), Province of South Africa.

from 393.23 to 994.09 nm (VNIR: Visible Near-Infrared) with bandwidths between 2 to 4 nm.

Atmospheric normalization reduces the effect of the atmospheric noises in the image spectral data. Consequently, the AISA Eagle image was atmospherically corrected using the quick atmospheric correction (QUAC) algorithm (Bernstein et al., 2005) built into the Environment for Visualizing Images (ENVI) software (ENVI, 2006). QUAC is VNIR-SWIR (Shortwave infrared) in-scene-based atmospheric correction module for multi- and hyperspectral imagery. It determines the required atmospheric measurements from the image (pixel spectra) without any ancillary information (Agrawal and Sarup, 2011) and converts the radiance data to surface relative reference.

The AISA Eagle image was then geometrically registered (Universal Transverse Mercator projection, zone 36 South) to a color aerial photograph of high spatial resolution (10 cm) captured in early April 2009 from the same study area using a nearest-neighbour algorithm and a third-order polynomial transformation method. A pixel root mean square error less than a pixel ($RMSE < 1$) was obtained, indicating a good geometric rectification (Ferencz et al., 2004). AISA Eagle reflectance data located after 900 nm were removed and excluded from the analysis due to high levels of noise associated with these regions (Dye et al., 2011; So et al., 2002). Therefore, only 230 out of 272 spectral wavebands were analyzed in this study.

2.3. Ground truth data collection

A stratified sampling method was followed for collecting the reflectance spectra of sires grey stage pine trees ($n = 85$). The

hyperspatial 10 cm color aerial photograph was used to identify sires grey stage trees with the aid of expert forest specialists from Sappi Forests (a paper and pulp company) as well as spatial and timely referenced ancillary datasets. Subsequently, a field campaign was conducted to verify the occurrence of the sires sample trees (points) using a GPS with sub-meter accuracy. Additionally during the field campaign, ground control points ($n = 85$ for each class) were also collected from (i) healthy pine trees with similar ages and (ii) lightning damaged trees that were identified by the plantation foresters and (iii) additional landuse classes that included *Eucalyptus spp.*, *Acacia spp.*, bugweed, bare soil, and shadow. Reflectance spectra were extracted from the AISA Eagle data using the collected ground control points and considered as variables when the classification algorithms were employed.

2.4. Statistical analysis

The utility of random forests (RF) and support vector machines (SVM) to discriminate sires grey stage and lightning damaged pine trees was examined in this paper. The classifiers were trained on 70% ($n = 60$) of a randomly selected holdout sample and final accuracy assessments were determined using the remaining 30% of the data. The parameters of RF and SVM classifiers were optimized and then input into Interactive Data Language (IDL)-based ImageRF (Waske et al., 2012) and ImageSVM (Rabe et al., 2010) tools to delineate the classes on ASIA Eagle image. These tools are license-free platforms that can be integrated into commercially-available IDL/ENVI software and can also be run as add-ons to

the EnMAP-Box (Held et al., 2012) which is an open-source and platform-independent software interface for image processing.

2.4.1. Random forest (RF)

The RF ensemble (Breiman, 2001) grows multiple unpruned trees (*ntree*) on bootstrap samples of the original data. Each tree is grown on a bootstrap sample (2/3 of the original data known as “in-bag” data) taken with replacement from the original data. Trees are split to many nodes using random subsets of variables (*mtry*); the default *mtry* value is the square root of the total number of variables. From the *mtry* selected variables, the variable that yields the highest decrease in impurity is chosen to split the samples at each node (Breiman, 2001). A tree is grown to its maximum size without pruning until the nodes are pure. That is, the nodes hold samples of the same class or contain certain number of samples. A prediction of the response variable (e.g., sired grey trees) is made by aggregating the prediction over all trees (Biau, 2012). In a classification application, a majority vote from all the trees in the ensemble determines the final prediction (Breiman, 2001). RF is a distribution-free (non-parametric) method that does not encounter any over-fitting (Hughes phenomenon) problem and it is robust to outliers and noise (Breiman, 2001). A single tree in RF is a weak classifier, because a random subset sample is used to train a tree, while the aggregation from all trees is considered as a strong classifier. As the variables are randomly selected in each tree node, a low correlation amongst the trees is expected, and over-fitting is therefore prevented (Breiman, 2001). A useful by-product of RF is variables importance which is a measure of the strength of a variable in the final model. In other words a variable importance is a metric of how much classification accuracy would decrease if data of a particular variable are removed while all variables remain the same (Prasad et al., 2006; Verikas et al., 2011). RF variable importance is a useful output for gaining better insight of which variable or a set of variables (e.g., spectral wavebands) are the most relevant for the classification of the classes. The measure of classification accuracy is based on an internal estimate known as “out-of-bag” (OOB) error which is calculated from the prediction of the data that are not used for growing the classification trees (OOB data are one-third of the original data) (Breiman, 2001). RF algorithm is easy to implement as only two parameters (*ntree* and *mtry*) need to be optimized (Breiman, 2001; Díaz-Uriarte and De Andres, 2006; Touw et al., 2012). It is recommended that the node size should be set to its default value (one, for classification) (Liaw and Wiener, 2002; Ogutu et al., 2011). A more detailed discussion on RF can be found elsewhere (e.g., Breiman, 2001; Touw et al., 2012).

The randomForest library (Liaw and Wiener, 2002) of R statistical packages version 2.15.2 (R Development Core Team, 2012) was used to optimize RF parameters.

2.4.2. Support vector machines (SVM)

Like RF, Support vector machines (SVM) requires no assumption about the data distribution (Everingham et al., 2007) and uses very efficient principles not to over-fit the test or new data samples (Brown et al., 1999; Burges, 1998; Cortes and Vapnik, 1995). SVM finds an optimal classification hyperplane through minimizing the upper bound of the classification error (Cortes and Vapnik, 1995; Vapnik, 1995). Basically, SVM iteratively locates hyperplanes amongst the training data and thereafter optimizes them according to error associates with each hyperplane. Hyperplanes are built from axes that represent each variable (e.g., ASIA Eagle spectral bands) and exist in a multidimensional space. In a 2-class experiment, the algorithm sets two supporting hyperplanes in the boundaries and searches to maximize the margin between them. Data points lying on the supporting hyperplanes are called support vectors and in the middle of the margin is the optimal hyperplane.

SVM uses slack variables to limit the violation of the restrictions set by the two supporting hyperplanes. Slack variable is a measure of how far the outlying sample is on the wrong side of the supporting hyperplane that holds the support vectors of its class (Yu et al., 2012). In addition, SVM attempts to maximize the distance from the data of each class to the optimum linear hyperplane (Petropoulos et al., 2011). However, most classes are not linearly separable, hence SVM is tuned for finding a non-linear (e.g., polynomial, radial etc.) separating hyperplane in a high-dimensional feature space using a kernel function (Karatzoglou et al., 2006). This is typically the case for most real world hyperspectral data (Chi and Bruzzone, 2007; Chi et al., 2008). Polynomial and radial basis kernels are the most commonly used functions for classifying remotely-sensed data (Huang et al., 2002; Oommen et al., 2008). A radial basis function performs better for classifying remotely-sensed data when compared with the polynomial kernel and requires optimizing only two parameters (Huang et al., 2008; Kavzoglu and Colkesen, 2009; Oommen et al., 2008; Waske and Benediktsson, 2010). These are the cost “sigma (C)” parameter which is a value for regularizing the error of misclassifying training dataset samples, and “gamma (γ)” which is the kernel width parameter. For a multi-class problem, a one-against-one or one-against-all procedures are followed to run all possible classifiers and assign the correct class by using a voting mechanism (Karatzoglou et al., 2006; Mazzoni et al., 2007). Interested readers are referred to, for example, Cortes and Vapnik (1995), Burges (1998), Karatzoglou et al. (2006), and Mathur and Foody (2008) for comprehensive description on SVM theory, principles and mathematical formulation.

We used all the 230 AISA Eagle bands and a radial basis kernel function to find an optimal hyperplane that can distinguish amongst healthy, sired grey- and lightning-damaged pine trees. The cost and gamma parameters of the radial basis function were optimized in order to avoid over-fitting and under-fitting problems (Karatzoglou et al., 2006; Waske et al., 2009). The one-against-one scheme was used to implement a multiclass SVM model as recommended by Hsu and Lin (2002) who reported that this strategy is more symmetric than one-against-all with regard to class sizes. In addition, the use of a one-against-all scheme does not always produce a complete classification matrix (Mathur and Foody, 2008). The e1071 library (Meyer, 2001; Meyer et al., 2012) of R statistical packages version 2.15.2 (R Development Core Team, 2012) was utilized to optimize the SVM parameters.

2.4.3. Optimizing the random forest and support vector machines hyperparameters

The aim of the optimization process was to determine the best parameters for each classifier in order to obtain the best classification accuracies. The user defined *ntree* and *mtry* for RF and C and γ for SVM were optimized using a grid search and a ten-fold cross validation method (Waske et al., 2009). That dataset was divided into ten subsets of equal size, RF and SVM models were then trained on nine subset samples, and tested on the omitted one and the process was repeated ten times until all subset samples have served as test samples. The pair of parameters for each classifier that minimizes the classification error was then considered as the best values for final classification. Based on the recommendation of Statnikov et al. (2008) *ntree* values up to 5000 were considered using intervals of 500 while a multiplicative factor of the default *mtry* was used (for example {1/3, 1/2, 1, 3, 2} * default *mtry*). The default value of *mtry* is based on the square root of the 230 AISA Eagle wavebands. With regards to C and γ , there was no stepwise procedure described in the literature concerning the selection criteria (Carrão et al., 2008; Petropoulos et al., 2011). However, we tested exponentially growing sequence of C and γ values (for example 10^{-3} , ..., 10^3) as suggested by Hsu

et al., 2010. On the other hand, Rakotomamonjy (2003) recommended that high C value reduces training error.

2.4.4. Variable selection

We utilized the variable importance as calculated by RF to rank the 230 AISA Eagle wavebands according to their ability to discriminate amongst healthy, sirex grey- and lightning-damaged pine trees as well as the other classes considered in this study. Subsequently, a forward selection method (Kohavi and John, 1997) was performed to identify the least number of the spectral bands that produced the highest classification accuracy. Multiple RFs were iteratively fitted using the ranked wavebands in a sequential manner. Initially, a new RF model was built using the highest ranked band and for the next iteration the two highest bands were considered. This process was repeated until all the spectral variables used in our study ($n = 230$) were considered. Finally, the subset of spectral bands that produced the lowest 10-fold cross validated error was then selected as the optimum subset of spectral bands for classification.

2.4.5. Accuracy assessment

The accuracy of each classifier was assessed using the 30% ($n = 25$) holdout sample. The overall accuracy (OA), user's accuracy (UA), and producer's accuracy (PA) were used as criteria for evaluating the generalization ability (accuracy) of the RF and SVM classifiers (Tso and Mather, 2009). OA is a ratio (%) between the number of correctly classified samples and the number of test samples, while UA represents the likelihood that a sample belongs to specific class and the classifier accurately assigns it such class. PA expresses the probability of a certain class being correctly recognized. Two useful parameters were calculated from the cross-tabulation matrices to evaluate the practicability of each classifier. These are quantity disagreement (QD) and allocation disagreement (AD) statistical metrics that were recently developed by Pontius and Millones (2011). The quantity disagreement is the absolute dissimilarity between the number of reference (test) observations and the predicted ones, while the allocation disagreement describes the number of predicted classes that have less than optimal spatial location in comparison to the reference samples. In addition, the McNemar's test was employed to test if there was any significant difference between the results of RF and SVM classifiers. McNemar's is a nonparametric test based on standardized normal test statistic calculated from error matrices of the two classifiers as follows (Foody, 2004; Leeuw et al., 2006):

$$Z = \frac{f_{12} - f_{21}}{\sqrt{f_{12} + f_{21}}} \quad (1)$$

where f_{12} denotes the number of samples that are misclassified by RF but correctly classified by SVM, and f_{21} denotes the number of samples that are correctly classified by RF but misclassified by SVM. The Z value could be referred to the tables of chi-squared distribution with one degree of freedom (Agregti, 1996). McNemar's test can therefore be expressed using a chi-squared formula computed as follows:

$$\chi^2 = \frac{(f_{12} - f_{21})^2}{f_{12} + f_{21}} \quad (2)$$

If the statistic χ^2 estimated from Eq. (2) is greater than a chi-squared table value of 3.84 at 5% level of significance, it implies that RF and SVM perform significantly different.

Following the calculation of overall and individual accuracies for all eight classes, we subsequently examined the paired accuracies of the healthy, sirex grey stage and lightning damaged pine trees as our main objective was to look at the possibility of discriminating amongst these classes.

3. Results

3.1. Random forest and support vector machines optimization

When all ASIA Eagle wavebands were used, the optimization results for RF and SVM using a 10-fold cross validation (CV) method illustrate that a combination of n_{tree} and m_{try} of 500 and 45 yielded the minimum CV error (32.91%) for the RF classifier. While the minimum CV error (24.95%) for the SVM classifier was produced by γ and C values of 0.01 and 100, respectively.

3.2. Spectral bands selection

The importance of AISA Eagle spectral bands in separating amongst the eight classes and as determined by the RF classifier is shown in Fig. 2. The most important 50 spectral wavebands are located in the red edge (670–780 nm), blue (400–500 nm) and green edge (500–600 nm) of the electromagnetic spectrum, and very few (only 4) are located at the near infrared (700–800 nm). The forward selection method selected an optimal 51 spectral bands using the ranking output of RF for discriminating amongst classes. These 51 spectral wavebands produced a minimal CV error of 31.91% using the training dataset and 25.50% using the holdout dataset (Fig. 3). These spectral wavebands were then used as the optimal input variables for the RF and SVM classification models.

3.3. Accuracy assessment

Using the 51 bands identified by the forward selection methods both classifiers were optimized. The results indicate that $\gamma = 0.1$ and $C = 10$ produced the best results for SVM while $n_{tree} = 500$ and $m_{try} = 15$ produced the best results for RF.

Figs. 4 and 5 show thematic maps obtained using RF and SVM classifiers. The main visual difference between the maps is the relatively more homogeneous maps produced when the most important 51 AISA Eagle bands were used (Figs. 4a and 5b). On the other hand, the lightning-struck pine trees are confused with bare soil. The maps show some sirex-attacked and lightning-struck classes outside the pine compartments.

The results of the overall accuracy assessment based on the test data set demonstrated that both machine learning classifiers performed relatively similar and yielded overall accuracy of 74.50% for RF and 73.50% for SVM using all usable AISA Eagle spectral bands (Tables 1 and 2). When the subset of 51 spectral bands was used, the overall accuracy was improved; 78% for RF and 76.50% for SVM (Tables 1 and 2). Moreover, all classification

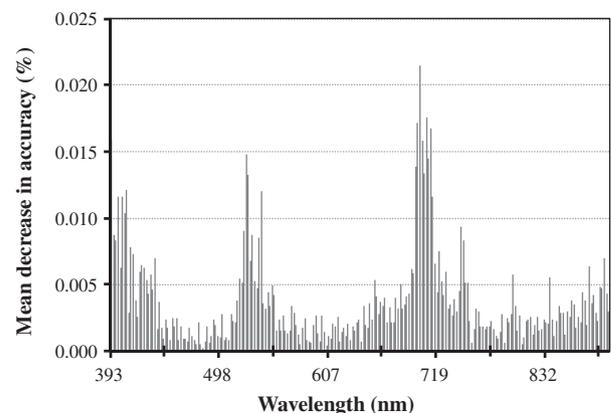


Fig. 2. The importance of ASIA Eagle wavebands in distinguishing among the studied eight classes as measure by random forest (RF).

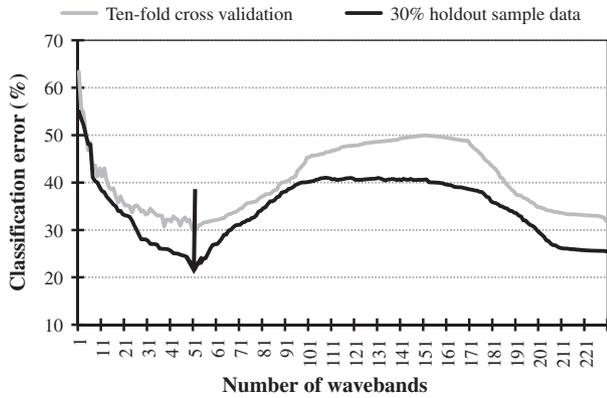


Fig. 3. Result of forward selection function using the importance of ASIA Eagle wavebands in separating among the studied classes as measured by random forest (RF). The classification error was calculated from the ten-fold cross validation and independent data sets. The arrow shows the optimal number of wavebands that produced the least classification error.

methods obtained quantity disagreements of 3%, except when SVM was employed for discriminating amongst the studied classes using the 230 AISA Eagle bands (Tables 1 and 2). Relatively high allocation disagreements (19%–23%) were recorded.

When the results of the PA and UA were examined, it was shown that values of the PA were high for shadow, bare soil,

Eucalyptus spp., lightning-struck trees and bugweed classes irrespective of the classifier and AISA Eagle wavebands (Table 3). On the other hand, all classes obtained fairly high UA, except siren grey-attack, *Eucalyptus spp.*, and *Acacia spp.* classes.

The results of the pairwise comparisons of healthy, siren grey- and lightning-damaged pine trees are shown in Figs. 6 and 7. The overall accuracies of all class pairs were above 90%, except when siren grey-attacked and lightning-struck trees were compared (86.05%) using the SVM classification algorithm and the 51 most important AISA Eagle spectral bands. Low disagreements (0–10%) were obtained by most pairwise comparisons (Figs. 6 and 7).

The difference between the performance of the RF and SVM classifiers was not significant according to the McNemar’s test (chi-squared; $\chi^2 = 0.14$, all the AISA Eagle bands were utilized, while chi-squared; $\chi^2 = 0.03$ when the optimal subset of bands ($n = 51$) was used). Nonetheless, RF produced the highest OA (79%) and less quantity (3%) and allocation (19%) disagreements when the optimal subset of bands ($n = 51$) was utilized (Table 1).

4. Discussion

The main focus of the present study was to look at the possibility of distinguishing siren grey-attacked and lightning-struck pine trees using airborne hyperspectral data and two machine learning classifiers, namely RF and SVM. Other classes on the image of the study area were also classified in order to map the attacked and

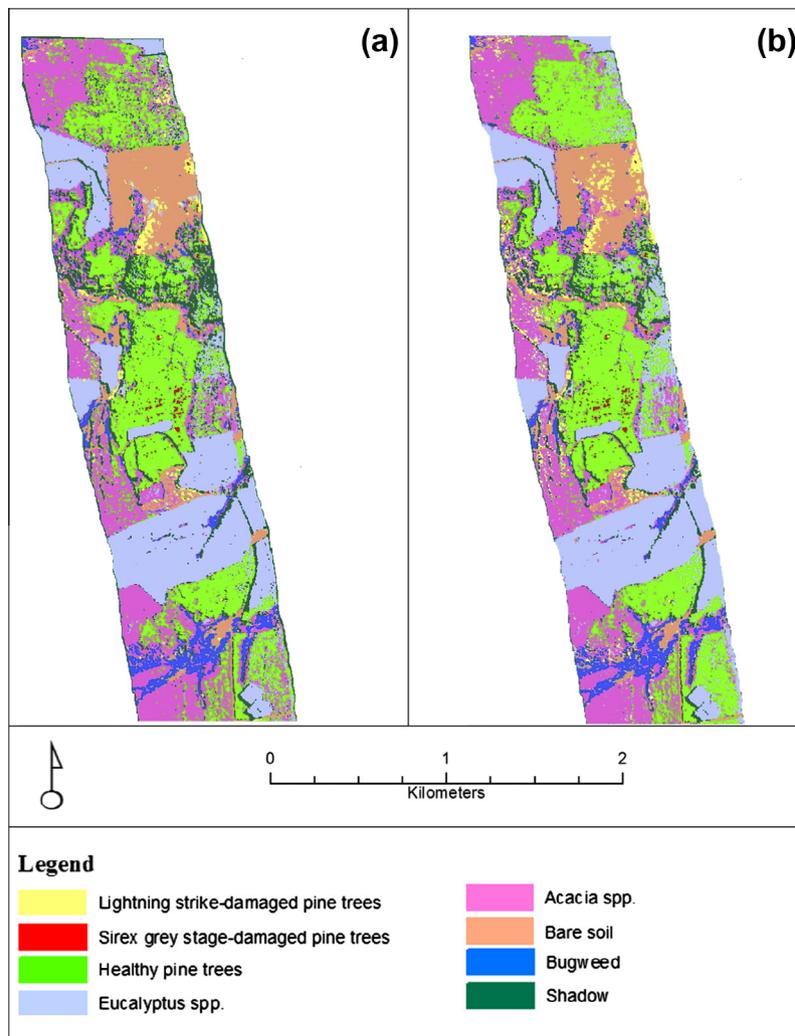


Fig. 4. Classification maps obtained using random forest classification algorithm, all (a) and the 51 most important (b) AISA Eagle spectral bands.

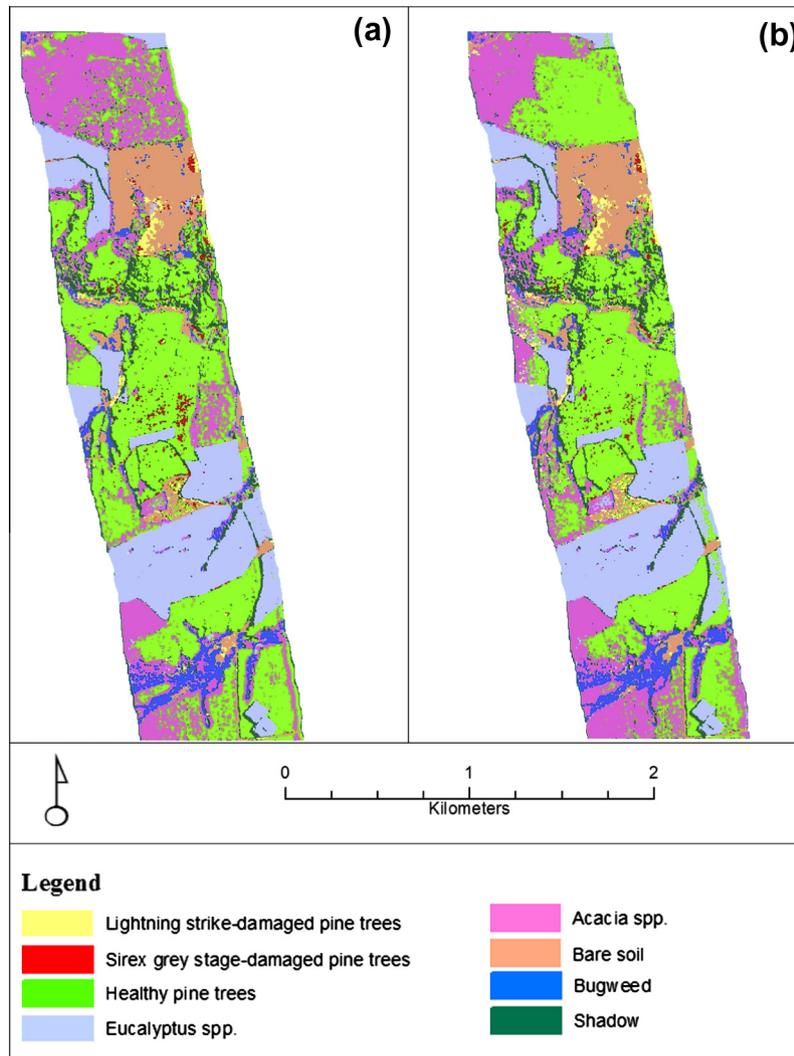


Fig. 5. Classification maps obtained using support vector machine classification algorithm, all (a) and the 51 most important (b) AISA Eagle spectral bands.

Table 1

Classification confusion matrix of random forest (RF) classifier using 230 and the 51 most important AISA Eagle wavebands for the 30% test data sets.

Class	Using 230 AISA eagles wavebands									Using the 51 most important AISA eagles wavebands								
	Ground truth									Ground truth								
	HP	SGS	LS	Eu	Aca	BW	BS	Sh	Total	HP	SGS	LS	Eu	Aca	BW	BS	Sh	Total
HP	18	00	02	01	03	01	00	00	25	19	02	02	01	01	00	00	00	25
SGS	02	17	01	01	03	01	00	00	25	01	19	01	00	03	01	00	00	25
LS	01	02	18	00	02	02	00	00	25	01	02	20	00	01	01	00	00	25
Eu	01	01	01	18	01	01	00	02	25	01	01	01	19	01	00	00	02	25
Aca	02	02	02	01	17	01	00	00	25	02	02	01	01	17	01	00	01	25
BW	01	01	01	02	00	18	01	01	25	00	00	02	01	00	21	01	00	25
BS	00	00	00	01	01	01	21	01	25	01	00	00	01	00	01	21	01	25
Sh	00	00	00	00	01	01	01	22	25	00	00	00	00	01	01	01	22	25
Total	25	23	25	24	28	26	23	26	200	25	26	27	23	24	26	23	26	200
OA (%)	74.50									79.00								
QD (%)	03.00									03.00								
AD (%)	23.00									19.00								

HP = Healthy pine trees, SGS = Sirex grey stage-damaged pine trees, LS = Lightning strike-damaged pine trees, Eu = Eucalyptus spp., Aca = Acacia spp., BW = Bugweed, BS = Bare soil, and Sh = Shadow. OA = Overall accuracy, QD = Quantity disagreement, AD = Allocation disagreement.

lightning-struck trees within different land classes and to produce a completely classified image. The relatively small overall and individual classification errors obtained on this study demonstrated the capability of the fine spatial and spectral resolutions of the

AISA Eagle sensor to detect pine trees mortality. The confusion of lightning-struck pine trees with bare soil on the maps (Figs. 4 and 5) could be due to some woody debris laying on the soil of the compartments that have been recently harvested or left as

Table 2

Classification confusion matrix of support vector machines (SVM) classifier using 230 and the 51 most important AISA Eagle wavebands for the 30% test data sets.

Class	Using 230 AISA Eagles wavebands									Using the 51 most important AISA Eagles wavebands								
	Ground truth									Ground truth								
	HP	SGS	LS	Eu	Aca	BW	BS	Sh	Total	HP	SGS	LS	Eu	Aca	BW	BS	Sh	Total
HP	20	00	01	01	03	00	00	00	25	19	02	00	01	03	00	00	00	25
SGS	01	18	00	02	03	01	00	00	25	02	18	03	00	01	01	00	00	25
LS	02	03	16	01	01	02	00	00	25	02	03	19	00	00	01	00	00	25
Eu	02	02	01	16	02	00	00	02	25	02	00	01	18	03	01	00	00	25
Aca	00	01	01	05	17	01	00	00	25	01	01	01	05	17	00	00	00	25
BW	01	01	01	01	00	20	00	01	25	00	01	03	01	00	20	00	00	25
BS	00	01	00	00	00	01	22	01	25	00	00	00	00	00	02	23	00	25
Sh	00	02	00	00	01	01	03	18	25	00	01	00	00	01	01	03	19	25
Total	26	28	20	26	27	26	25	22	200	26	26	27	25	25	26	26	19	200
OA (%)	73.50									76.50								
QD (%)	04.00									03.00								
AD (%)	23.00									21.00								

HP = Healthy pine trees, SGS = Sirex grey stage-damaged pine trees, LS = Lightning strike-damaged pine trees, Eu = *Eucalyptus* spp., Aca = *Acacia* spp., BW = Bugweed, BS = Bare soil, and Sh = Shadow. OA = Overall accuracy, QD = Quantity disagreement, AD = Allocation disagreement.

Table 3

Producer's accuracy (%), user's accuracy (%) and overall accuracy (%) of the studied eight classes using 230 and the 51 most important AISA Eagle wavebands, random forest (RF) and support vector machine (SVM) classifiers for the 30% test data sets.

Class	Using 230 AISA Eagles wavebands				Using the 51 most important AISA Eagles wavebands			
	RF		SVM		RF		SVM	
	Producer's accuracy	User's accuracy	Producer's accuracy	User's accuracy	Producer's accuracy	User's accuracy	Producer's accuracy	User's accuracy
HP	72.00	72.00	76.92	80.00	76.00	76.00	073.00	76.00
SGS	73.91	68.00	64.29	72.00	73.08	76.00	069.23	72.00
LS	72.00	72.00	80.00	64.00	74.07	80.00	070.37	76.00
Eu	75.00	72.00	61.54	64.00	82.61	76.00	072.00	72.00
Aca	60.71	68.00	62.96	68.00	70.83	68.00	068.00	68.00
BW	69.23	72.00	76.92	80.00	80.77	84.00	076.92	80.00
BS	91.30	84.00	88.00	88.00	91.30	84.00	088.46	92.00
Sh	84.62	88.00	81.82	72.00	84.62	88.00	100.00	76.00

HP = Healthy pine trees, SGS = Sirex grey stage-damaged pine trees, LS = Lightning strike-damaged pine trees, Eu = *Eucalyptus* spp., Aca = *Acacia* spp., BW = Bugweed, BS = Bare soil, and Sh = Shadow.

fallow. Since we provide thematic maps for the entire classes on the study area, the accuracy of mapping sirenx-attacked and lightning-struck pine trees should be assessed within pine compartments.

Generally speaking, the shadow and bare soil classes showed the highest producer's and user's accuracies. This was expected since all other classes represent vegetation (i.e. pine, *Eucalyptus* spp., *Acacia* spp., and bugweed), hence their spectral characteristics should accurately be discriminated from the non-vegetation classes (i.e. shadow and bare soil). Within the vegetation classes, the highest accuracies were attained by bugweed, healthy and lightning-struck pine trees. That might be attributed to the distinguishable canopy structure (geometry) and biochemical compositions (mainly chlorophyll) of bugweed. These attributes are directly related to the canopy spectral signature of a vegetation class on the image (Kumar et al., 2003; Lillesand and Kiefer, 2001). Regarding pine trees, the order of the three classes in terms of frequently high individual accuracies was healthy trees, lightning-struck and sirenx grey-damaged snags. This implies that the spectral responses of pine snags were changed by the damage of lightning strike and sirenx grey stage of attack. Furthermore, the results revealed that the pair classes of pine trees (i.e., HP, SGS, and LS) can be separated from each other with very high overall and individual accuracies (Figs. 6 and 7). That could be due to different lightning strike and sirenx grey stage mode of actions. In other words, the biophysiological components and structural parameters of sirenx grey-attacked and lightning-struck pine snags could considerably differ. This is an interesting topic for further investigation to look at for example

pine pulp structure, wood density, sapwood, cellulose, and lignin variations that might be due to sirenx infestations and lightning strikes. Studies have shown that tree wood structure and some other biophysiological constituents can be estimated using hyperspectral data (Chave et al., 2009; Popescu et al., 2007; So et al., 2002).

The relatively high allocation disagreements shown in the tables of the confusion matrices were expected since pixels covered by multiple classes could possibly be mismatched in terms of spatial pattern between test ground truth instances and predicated ones. However, we believe that our classification results are of good practical application as the amount of differences between the reference and predicted test samples ranged between 3% and 4% for all classes and between 0% and 8% for pair classes. Additionally, we were interested on the number of pine trees attacked by sirenx and/or struck by lightning.

Since RF and SVM classification algorithms were run using equivalent training and test data points in the present study, we weighted their performance against each other in separating our classes of interest using hyperspectral data. RF and SVM are unique and versatile algorithms for classifying multidimensional and noisy hyperspectral data, because the algorithms are robust to outliers and do not over-fit (Biau, 2012; Breiman, 2001; Mountrakis et al., 2011; Prasad et al., 2006; Vapnik, 1995). Over-fitting is one of the obstacles that mostly hampers the classification of hyperspectral data due the high variable-to-sample ratio (Hughes effect) problem. We analyzed 230 AISA Eagle spectral wavebands for

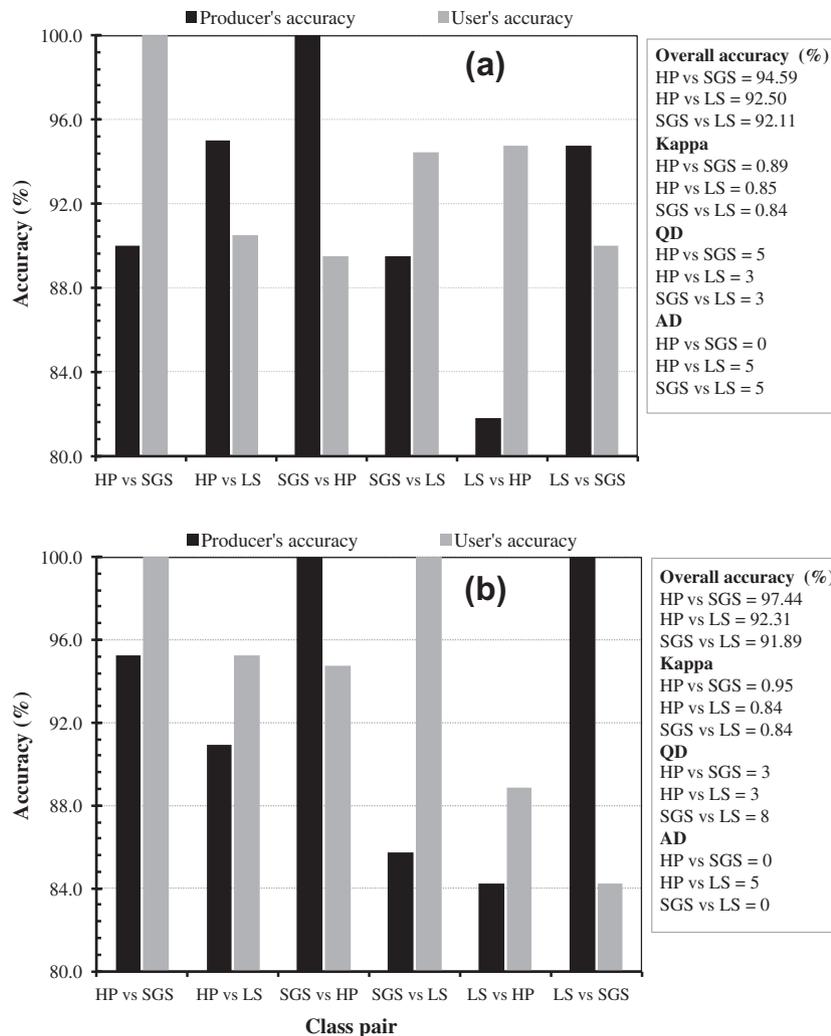


Fig. 6. Producer's and user's accuracies for healthy (HP), sirex grey-damaged (SGS) and lightning-struck (LS) pine trees pairs comparison achieved by random forest (a) and support vector machine (b) when the 230 ASIA Eagle wavebands were employed.

classifying eight classes of 85 samples each. Our sample size was about 37% of the total number of AISA Eagle wavebands. The good performance of RF and SVM in the present study proves the strength of the two algorithms against over-fitting.

The optimal number of trees ($n_{tree} = 500$) reported in the present study for RF conforms to the default n_{tree} value recommended by Liaw and Wiener (2002). However, the optimal number of wavebands that were randomly chosen to split each node in a tree ($m_{try} = 45$) exceeded the default value (about 15). Researchers have noted that fewer trees in a RF model are required to reduce the autocorrelation (multi-collinearity) amongst the RF ranked variables (Goldstein et al., 2010). On the other hand, smaller m_{try} values may result in a biased RF model (Goldstein et al., 2010). Notably, Breiman (2001) and Statnikov et al. (2008) argued that RF parameters (n_{tree} and m_{try}) should be optimized for obtaining a RF classification model of improved accuracy.

In this study, the optimal C value (100) of SVM is in consistency with the values calculated by Petropoulos et al. (2012) who utilized SVM method for land use/cover classification using hyperspectral data. A relatively high C value penalizes the training error and forces the algorithm to reduce the misclassification of samples during the training process (Petropoulos et al., 2012). Notwithstanding, Belousov et al. (2002) reported that SVM is flexible to a wide

range of C values. On the other hand, the best γ value (0.01) reported in the present study is slightly different from the recommendation (inverse of the number of AISA Eagle spectral bands; 0.004) of Petropoulos et al. (2012) and Sesnie et al. (2010). Huang et al. (2002) argued that the classification error decreased when γ decreased from 1 to 0.1. However, the authors noticed that the radial kernel is relatively less affected by γ parameter.

RF is also a known method for features selection (Breiman, 2001; Prasad et al., 2006). Therefore, it is a very valuable procedure for minimizing hyperspectral datasets that have redundant bands. RF assigned weights to the AISA Eagle wavebands according to their importance in discriminating amongst the studied eight classes. The ranked AISA Eagle wavebands provide an insight in which spectral bands or a set of spectral bands are more useful than the others. RF variable selection function reduced the dimensionality of AISA Eagle data by about 78% as only 51 wavebands were selected as the most important features for classifying the eight classes on the image.

The results of the present paper explained the similarity of the performance of SVM and RF classifiers. This is in conformity with the findings of other authors (e.g., Waske et al., 2009) who reported relatively similar performance of the two algorithms for classifying hyperspectral data. However, we suggest the employment of RF

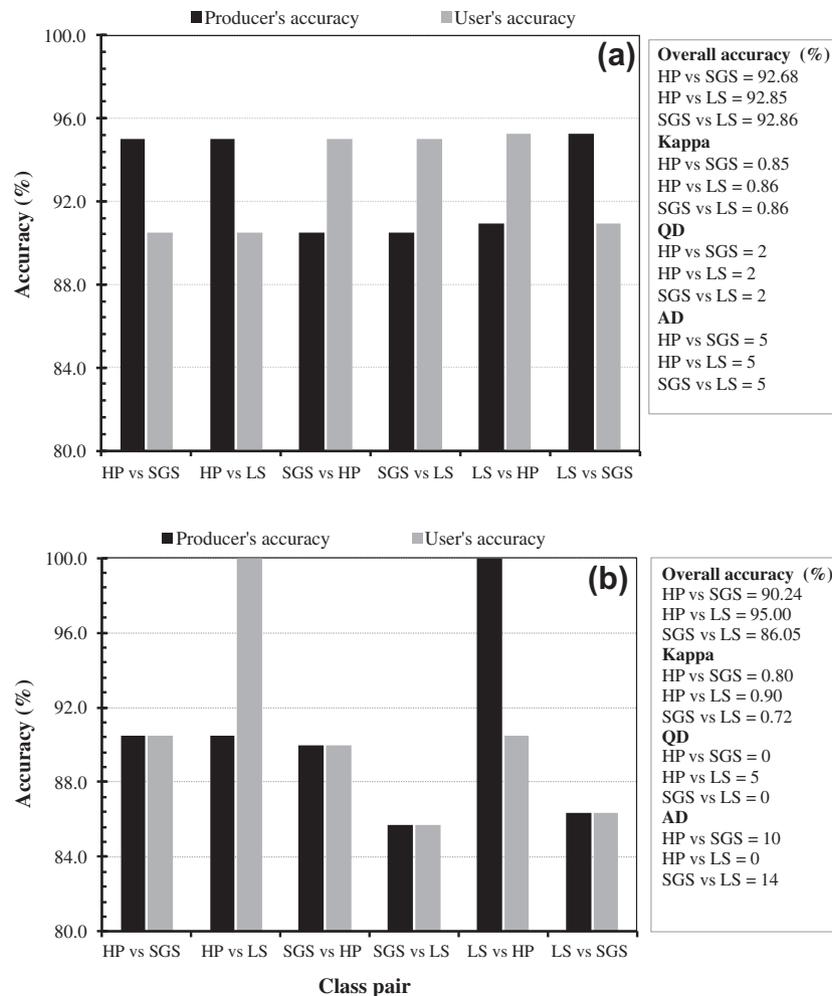


Fig. 7. Producer's and user's accuracies for healthy (HP), sirex grey-damaged (SGS) and lightning-struck (LS) pine trees pairs comparison achieved by random forest (a) and support vector machine (b) when the 51 most important ASIA Eagle wavebands were employed.

when hyperspectral data are classified due to the following advantages:

- RF requires an optimization of two parameters only (*ntree* and *mtry*), whereas SVM requires a selection of a suitable kernel function first and then setting of the specific parameters.
- The variable importance by-product of RF which is produced during the classification process makes the algorithm a robust feature selection and redundancy reduction method.
- In nonlinear separable cases, RF method generates many decision tree classifiers that construct simple decision boundaries. In contrast SVM transfers the data into higher feature space that enable a generation of a quite complicated decision boundary that appears nonlinear in the input feature space (Waske et al., 2009).
- The internal OOB error rate of RF could be used for classification accuracy assessment when there are limited samples for independent accuracy assessments.
- RF calculates proximity among samples and hence can enable the detection of outliers.

Identification of outliers provides means for unsupervised classification (Pal, 2005; Touw et al., 2012). Nevertheless, authors (see Mountrakis et al., 2011 for extensive list of references) have shown

the capability, strengths, and the reliable performance of SVM for classifying various types of remotely-sensed data.

5. Conclusions

From the results of the present study we can conclude that:

- (1) Sirex grey- and lightning-damaged pine trees could accurately be detected using AISA Eagle hyperspectral data, RF and SVM classifiers.
- (2) Other landscape classes in the pine forest plantation could also be successfully distinguished.
- (3) RF and SVM classifiers performed comparatively similar for separating amongst the eight classes, nevertheless RF achieved relatively higher overall accuracy. RF is considered to be a useful approach for reducing the dimensionality of hyperspectral data.

Overall, our study presents a successful application of hyperspectral data, RF and SVM classifiers in detecting pine trees snags. This could help in making informed decision regarding the management strategies for pine snags and therefore contributes to the concept of "arboriculture". However, caution should be taken when interpreting the results since the study was only a snapshot of specific environmental conditions.

Acknowledgements

We are grateful to Sappi Forests for providing remotely-sensed data and supporting field data collection campaign. We thank Mr. Brice Gijbertsen (Chief Cartographer at the School of Agricultural, Earth and Environmental Sciences) for help with fieldwork. Gratitude is extended to the R development core team for their very powerful open source packages for statistical analysis.

References

- Adam, E.M., Mutanga, O., Rugege, D., Ismail, R., 2012. Discriminating the papyrus vegetation (*Cyperus papyrus* L.) and its co-existent species using random forest and hyperspectral data resampled to HYMAP. *Int. J. Remote Sens.* 33, 552–569.
- Agrawal, G., Sarup, J., 2011. Comparison of QUAC and FLAASH atmospheric correction modules on EO-1 hyperion data of Sanchi. *Int. J. Adv. Eng. Sci. Technol.* 4, 178–186.
- Agresti, A., 1996. *An Introduction to Categorical Data Analysis*. John Wiley, New York.
- Bandos, T.V., Bruzzone, L., Camps-Valls, G., 2009. Classification of hyperspectral images with regularized linear discriminant analysis. *IEEE Trans. Geosci. Remote Sens.* 47, 862–873.
- Belousov, A.I., Verzakov, S.A., vonFrese, J., 2002. A flexible classification approach with optimal generalisation performance: support vector machines. *Chernometrics Intell. Lab. Syst.* 64, 15–25.
- Bernstein, L.S., Adler-Golden, S.M., Sundberg, R.L., Levine, R.Y., Perkins, T.C., Berk, A., 2005. Validation of the QUick atmospheric correction (QUAC) algorithm for VNIR SWIR multi- and hyperspectral imagery. In: Shen, S.S., Lewis, P.E. (Eds.), *Proc SPIE 5806 Algorithms and Technologies for Multispectral, Hyperspectral and Ultraspectral Imagery XI*, Orlando, Florida, 13 July, pp. 668–678.
- Biau, G., 2012. Analysis of a random forests model. *J. Mach. Learn. Res.* 13, 1063–1095.
- Breiman, L., 2001. Random forests. *Machine Learning* 45, 5–32.
- Brown, M., Gunn, S.R., Lewis, H.G., 1999. Support vector machines for optimal classification and spectral unmixing. *Ecol. Model.* 120, 167–179.
- Burges, C.J.C., 1998. A tutorial on support vector machines for pattern recognition. *Data Min. Knowl. Disc.* 2, 121–167.
- Camps-Valls, G., Bruzzone, L., 2005. Kernel-based methods for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* 43, 1351–1362.
- Carnegie, A.J., Eldridge, R.H., Waterson, D.G., 2005. History and management of siren wood wasp in pine plantations in New South Wales, Australia. *NZ J. Forest. Sci.* 35, 3–24.
- Carrão, H., Gonçalves, P., Caetano, M., 2008. Contribution of multispectral and multitemporal information from MODIS images to land cover classification. *Remote Sens. Environ.* 112, 986–997.
- Chave, J., Coomes, D., Jansen, S., Lewis, S.L., Swenson, N.G., Zanne, A.E., 2009. Towards a worldwide wood economics spectrum. *Ecol. Lett.* 12, 351–366.
- Chi, M., Bruzzone, L., 2007. Semisupervised classification of hyperspectral images by SVMs optimized in the primal. *IEEE Trans. Geosci. Remote Sens.* Lett. 45, 1870–1880.
- Chi, M., Feng, R., Bruzzone, L., 2008. Classification of hyperspectral remote sensing data with primal SVM for small-sized training dataset problem. *Adv. Space Res.* 41, 1793–1799.
- Ciesla, W.M., 2003. European woodwasp: a potential threat to North Americas conifer forests. *J. Forest.* 101, 18–23.
- Coops, N.C., Goodwin, N., Stone, C., 2006. Predicting *Sphaeropsis sapinea* damage in *Pinus radiata* canopies using spectral indices and spectral mixture analysis. *Photogramm. Eng. Remote Sens.* 72, 405–416.
- Corley, J.C., Villacide, J.M., Bruzzone, O.A., 2007. Spatial dynamics of a *Sirex noctilio* woodwasp population within a pine plantation in Patagonia, Argentina. *Entomol. Exp. Appl.* 125, 231–236.
- Cortes, C., Vapnik, V., 1995. Support-vector networks. *Mach. Learn.* 20, 273–297.
- Coutts, M.P., 1969. The mechanism of pathogenicity of *Sirex noctilio* on *Pinus radiata* I. Effects of the symbiotic fungus *Amylostereum* sp. (Thelephoraceae). *Aust. J. Biol. Sci.* 22, 915–924.
- Díaz-Urriarte, R., De Andres, S.A., 2006. Gene selection and classification of microarray data using random forest. *BMC Bioinformatics* 7. <http://dx.doi.org/10.1186/1471-2105-7-3> (Accessed 17 February, 2013).
- Dobyn, N., 2009. Environmental Management Plan: Sappi Forests, Hodgsons. Unpublished Report.
- Dye, M., Mutanga, O., Ismail, R., 2008. Detecting the severity of woodwasp, *Sirex noctilio*, infestation in a pine plantation in KwaZulu-Natal, South Africa, using texture measures calculated from high spatial resolution imagery. *Afr. Entomol.* 16, 263–275.
- Dye, M., Mutanga, O., Ismail, R., 2011. Examining the utility of random forest and AISA Eagle hyperspectral image data to predict *Pinus patula* age in KwaZulu-Natal, South Africa. *Geocarto Int.* 26, 275–289.
- ENVI, 2006. *Environment for Visualising Images*. ITT industries Inc., Boulder, USA.
- Everingham, Y.L., Lowe, K.H., Donald, D.A., Coomans, D.H., Markley, J., 2007. Advanced satellite imagery to classify sugarcane crop characteristics. *Agron. Sustain. Dev.* 27, 111–117.
- Ferencz, C., Bognár, P., Lichtenberger, J., Hamar, D., Tarcsai, G., Timár, G., Molnár, G., Pásztor, S., Steinbach, P., Székely, B., Ferencz, O.E., Ferencz-Árkos, I., 2004. Crop yield estimation by satellite remote sensing. *Int. J. Remote Sens.* 25, 4113–4149.
- Foody, G.M., 2004. Thematic map comparison: evaluating the statistical significance of differences in classification accuracy. *Photogramm. Eng. Remote Sens.* 70, 627–633.
- Goetz, A.F.H., 2009. Three decades of hyperspectral remote sensing of the Earth: A personal view. *Remote Sens. Environ.* 113, S5–S16.
- Goldstein, B.A., Hubbard, A.E., Cutler, A., Barcellos, L.F., 2010. An application of random forests to a genome-wide association dataset: methodological considerations & new findings. *BMC Genet.* 11, 1–19.
- Held, M. et al., 2012. *EnMAP-Box Manual, Version 1.4*, Humboldt-Universität zu Berlin, Germany.
- Hsu, C.-W., Lin, C.-J., 2002. A comparison of methods for multiclass support vector machines. *IEEE Trans. Neural Networks* 13, 415–425.
- Hsu, C., Chang, C., Lin, C., 2010. *A practical guide to support vector classification*. Department of Computer Science and Information Engineering, National Taiwan University, Taipei 106, Taiwan.
- Hsu, P.-H., 2007. Feature extraction of hyperspectral images using wavelet and matching pursuit. *ISPRS J. Photogramm. Remote Sens.* 62, 78–92.
- Huang, C., Davis, L.S., Townshend, J.R.G., 2002. An assessment of support vector machines for land cover classification. *Int. J. Remote Sens.* 23, 725–749.
- Huang, C., Song, K., Kim, S., Townshend, J.R.G., Davis, P., Masek, J.G., Goward, S.N., 2008. Use of dark object concept and support vector machine to automate forest cover change analysis. *Remote Sens. Environ.* 112, 970–985.
- Hughes, G.F., 1968. On the mean accuracy of statistical pattern recognition. *IEEE Trans. Inf. Theory* 14, 55–63.
- Hurley, B.P., Slippers, B., Wingfield, M.J., 2007. A comparison of control results for the alien invasive woodwasp, *Sirex noctilio*, in the southern hemisphere. *Agric. For. Entomol.* 9, 159–171.
- Hurley, B.P., Slippers, B., Croft, P.K., Hatting, H.J., van der Linde, M., Morris, A.R., Dyer, C., Wingfield, M.J., 2008. Factors influencing parasitism of *Sirex noctilio* (Hymenoptera:Siricidae) by the nematode *Deladenus siricidicola* (Nematoda:Neotylenchidae) in summer rainfall areas of South Africa. *Biol. Control* 45, 450–459.
- Ismail, R., Mutanga, O., Bob, U., 2006. The use of high resolution airborne imagery for the detection of forest canopy damage by *Sirex noctilio*. In: Langin, P.A., Antonides MC (Eds.), *Proc. The International Precision Forestry Symposium, University of Stellenbosch, Stellenbosch*, 5–10 March, pp. 119–134.
- Ismail, R., Mutanga, O., Bob, U., 2007. Forest health and vitality: The detection and monitoring of *Pinus patula* trees infected by *Sirex noctilio* using digital multispectral imagery (DMSI). *Southern Hemisphere For. J.* 69, 39–47.
- Ismail, R., Mutanga, O., Ahmed, F., 2008. Discriminating *Sirex noctilio* attack in pine forest plantations in South Africa using high spectral resolution data. In: Kalacska Margaret, and Sanchez-Azofeifa G. Arturo (Eds.), *Hyperspectral Remote Sensing of Tropical and Sub-Tropical Forests*. Taylor and Francis, London, pp. 161–74.
- Karatzoglou, A., Meyer, D., Hornik, K., 2006. Support Vector Machines in R. *Journal of Statistical Software* 15, 1–28.
- Kavzoglu, T., Colkesen, I., 2009. A kernel functions analysis for support vector machines for land cover classification. *Int. J. Appl. Earth Obs. Geoinf.* 11, 352–359.
- Kohavi, R., John, G.H., 1997. Wrappers for feature subset selection. *Artif. Intell.* 97, 273–324.
- Kumar, L., Dury, S.J., Schmidt, K., Skidmore, A., 2003. Imaging spectrometry and vegetation science. In: Meer, F.D.V.D., Jong, S.M.D. (Eds.), *Image Spectrometry*. Kluwer Academic Publishers, London, pp. 111–156.
- Leeuw, J.D., Jia, H., Yang, L., Liu, X., Schmidt, K., Skidmore, A.K., 2006. Comparing accuracy assessments to infer superiority of image classification methods. *Int. J. Remote Sens.* 27, 223–323.
- Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. *R News* 2, 18–22.
- Lillesand, T.M., Kiefer, R.W., 2001. *Remote Sensing and Image Interpretation*, fourth ed. John Wiley & Sons Inc., New York.
- Marcus, W.A., Marston, R.A., Colvard, C.R., Gray, R.D., 2002. Mapping the spatial and temporal distributions of woody debris in streams of the Greater Yellowstone Ecosystem. *USA. Geomorphology* 44, 323–335.
- Mathur, A., Foody, G.M., 2008. Multiclass and binary SVM classification: Implications for training and classification users. *IEEE Geosci. Remote Sens. Lett.* 5, 241–245.
- Mazzoni, D., Garay, M.J., Davies, R., Nelson, D., 2007. An operational MISR pixel classifier using support vector machines. *Remote Sens. Environ.* 107, 149–158.
- Meddens, A.J.H., Hicke, J.A., Vierling, L.A., 2011. Evaluating the potential of multispectral imagery to map multiple stages of tree mortality. *Remote Sens. Environ.* 115, 1632–1642.
- Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F., 2012. Package 'e1071'. <<http://cran.rproject.org/web/packages/e1071/index.html>>. (accessed 15.02.13).
- Meyer, D., 2001. Support vector machines. *R News* 1/3, 23–26.
- Mountrakis, G., Im, J., Ogole, C., 2011. Support vector machines in remote sensing: a review. *ISPRS J. Photogramm. Remote Sens.* 66, 247–259.
- Neumann, F.G., Minko, G., 1981. The siren wood wasp in Australian *radiata* pine plantations. *Aust. Forestry* 44, 46–63.
- Ogutu, J.O., Piepho, H.-P., Schulz-Streeck, T., 2011. A comparison of random forests, boosting and support vector machines for genomic selection, *BMC Proc.* 5, doi:10.1186/1753-6561-5-S3-S11 (accessed 15.01.13).

- Oommen, T., Misra, D., Twarakavi, N.K.C., Prakash, A., Sahoo, B., Bandopadhyay, S., 2008. An objective analysis of support vector machine based classification for remote sensing. *Math. Geosci.* 40, 409–424.
- Pal, M., 2005. Random forest classifier for remote sensing classification. *Int. J. Remote Sens.* 26, 217–222.
- Pasher, J., King, D.J., 2009. Mapping dead wood distribution in a temperate hardwood forest using high resolution airborne imagery. *For. Ecol. Manage.* 258, 1536–1548.
- Petropoulos, G.P., Kalaitzidis, C., Vadrevu, K.P., 2012. Support vector machines and object-based classification for obtaining land-use/cover cartography from Hyperion hyperspectral imagery. *Comput. Geosci.* 41, 99–107.
- Petropoulos, G.P., Kontoes, C., Keramitsoglou, I., 2011. Burnt area delineation from a uni-temporal perspective based on Landsat TM imagery classification using support vector machines. *Int. J. Appl. Earth Observ. Geoinform.* 13, 70–80.
- Plaza, A., Benediktsson, J.A., Boardman, J.W., Brazile, J., Bruzzone, L., Camps-Valls, G., Chanussot, J., Fauvel, M., Gamba, P., Gualtieri, A., 2009. Recent advances in techniques for hyperspectral image processing. *Remote Sens. Environ.* 113, S110–S122.
- Pontius, R.G., Millones, M., 2011. Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. *Int. J. Remote Sens.* 32, 4407–4429.
- Popescu, C.-M., Popescu, M.-C., Singurel, G., Vasile, C., Argyropoulos, D.S., Willfor, S., 2007. Spectral characterization of *Eucalyptus* wood. *Appl. Spectrosc.* 61, 1168–1177.
- Prasad, A.M., Iverson, L.R., Liaw, A., 2006. Newer classification and regression tree techniques: bagging and random forests for ecological prediction. *Ecosystems* 9, 181–199.
- R Development Core Team. 2012. R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, <<http://www.R-project.org>>, Vienna. (accessed 5.01.13).
- Rabe, A., van der Linden, S., Hostert, P., 2010. imageSVM, Version 2.1, Software. <www.hu-geomatics.de>. (accessed 18.11.13).
- Rakotomamonjy, A., 2003. Variable selection using SVM-based criteria. *J. Mach. Learn. Res.* 3, 1357–1370.
- Rutherford, M.C., Mucina, L., Powrie, L.W., 2006. Biomes and bioregions of southern Africa. In: Mucina, L., Rutherford, M.C. (Eds.), *The vegetation of South Africa, Lesotho and Swaziland*. Strelitzia 19. South African National Biodiversity Institute, Pretoria, pp. 30–51.
- Sappi, 1993. *Forest Land Types of the Natal Region Sappi Forests Research*, Howick, KwaZulu Natal, South Africa.
- Schulze, R.E., Maharaj, M., 1997. South African Atlas of Agrohydrology and Climatology, Report TT82/96. Water Research Commission.
- Schulze, R.E., Maharaj, M., Lynch, S.D., Howe, B.J., Melvil-Thomson, B., 1997. South African atlas of agrohydrology and climatology. Water Research Commission, Report, TT82/96.
- Sesnie, S.E., Finegan, B., Gessler, P.E., Thessler, S., Bendana, Z.R., Alistair, M.S.S., 2010. The multispectral separability of Costa Rican rainforest types with support vector machines and random forest decision trees. *Int. J. Remote Sens.* 31, 2885–2909.
- So, C.-L., Groom, L.H., Rials, T.G., Snell, R., Kelley, S.S., Meglen, R., 2002. Rapid assessment of the fundamental property variation of wood. In: Outcalt, K.W. (Ed.), *Proc the Eleventh Biennial Southern Silvicultural Research Conference*. General Technical Report SRS-48. Asheville, NC. US Department of Agriculture, Forest Service, Southern Research Station. pp. 176–180.
- Spradbery, J.P., 1973. A comparative study of the phytotoxic effects of siricid woodwasps on conifers. *Ann. Appl. Biol.* 75, 309–320.
- Statnikov, A., Wang, L., Aliferis, C.F., 2008. A comprehensive comparison of random forests and support vector machines for microarray-based cancer classification. *BMC Bioinformatics* 9, 1–10.
- Touw, W.G., Bayjanov, J.R., Overmars, L., Backus, L., Boekhorst, J., Wels, M., Hijum, S.A.F.T.V., 2012. Data mining in the life sciences with random forest: a walk in the park or lost in the jungle? Briefings in Bioinformatics, doi:10.1093/bib/bbs034 (accessed 15.01.13).
- Tribe, G.D., Cillié, J.J., 2004. The spread of *Sirex noctilio* Fabricius (Hymenoptera: Siricidae) in South African pine plantations and the introduction and establishment of its biological control agents. *Afr. Entomol.* 12, 9–17.
- Tso, B., Mather, P.M., 2009. *Classification Methods for Remotely Sensed Data*, second ed. CRC Press/ Taylor & Francis Group, London.
- Vapnik, V.N., 1995. *The Nature of Statistical Learning Theory*. Springer-Verlag, New York.
- Verikas, A., Gelzinis, A., Bacauskiene, M., 2011. Mining data with random forests: a survey and results of new tests. *Pattern Recogn.* 44, 330–349.
- Waske, B., Benediktsson, J.A., 2010. Sensitivity of support vector machines to random feature selection in classification of hyperspectral data. *IEEE Trans. Geosci. Remote Sens.* 48, 2880–2889.
- Waske, B., Benediktsson, J.A., Arnason, K., Sveinsson, J.R., 2009. Mapping of hyperspectral AVIRIS data using machine-learning algorithms. *Can. J. Remote Sens.* 35, 106–116.
- Waske, B. et al., 2012. ImageRF – a user-oriented implementation for remote sensing image analysis with Random Forests. *Environ. Modell. Softw.* 35, 192–193.
- Yu, L., Porwal, A., Holden, E.-J., Dentith, M.C., 2012. Towards automatic lithological classification from remote sensing data using support vector machines. *Comput. Geosci.* 45, 229–239.