

Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Evaluation of Random Forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery

Jonathan Cheung-Wai Chan ^{a,*}, Desiré Paelinckx ^b^a Geography Department, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium^b Research Institute for Nature and Forest (INBO), Kliniekstraat 25, 1070 Brussels, Belgium

ARTICLE INFO

Article history:

Received 1 March 2007

Received in revised form 14 February 2008

Accepted 17 February 2008

Keywords:

Ecotope mapping

Ensemble classification

Adaboost

Random Forest

Airborne hyperspectral

Band selection

ABSTRACT

Detailed land use/land cover classification at ecotope level is important for environmental evaluation. In this study, we investigate the possibility of using airborne hyperspectral imagery for the classification of ecotopes. In particular, we assess two tree-based ensemble classification algorithms: Adaboost and Random Forest, based on standard classification accuracy, training time and classification stability. Our results show that Adaboost and Random Forest attain almost the same overall accuracy (close to 70%) with less than 1% difference, and both outperform a neural network classifier (63.7%). Random Forest, however, is faster in training and more stable. Both ensemble classifiers are considered effective in dealing with hyperspectral data. Furthermore, two feature selection methods, the out-of-bag strategy and a wrapper approach feature subset selection using the best-first search method are applied. A majority of bands chosen by both methods concentrate between 1.4 and 1.8 μm at the early shortwave infrared region. Our band subset analyses also include the 22 optimal bands between 0.4 and 2.5 μm suggested in Thenkabail et al. [Thenkabail, P.S., Enclona, E.A., Ashton, M.S., and Van Der Meer, B. (2004). Accuracy assessments of hyperspectral waveband performance for vegetation analysis applications. *Remote Sensing of Environment*, 91, 354–376.] due to similarity of the target classes. All of the three band subsets considered in this study work well with both classifiers as in most cases the overall accuracy dropped only by less than 1%. A subset of 53 bands is created by combining all feature subsets and comparing to using the entire set the overall accuracy is the same with Adaboost, and with Random Forest, a 0.2% improvement. The strategy to use a basket of band selection methods works better. Ecotopes belonging to the tree classes are in general classified better than the grass classes. Small adaptations of the classification scheme are recommended to improve the applicability of remote sensing method for detailed ecotope mapping.

© 2008 Elsevier Inc. All rights reserved.

1. Introduction

Land use/land cover classification is a generic tool for environmental monitoring. To measure subtle changes in the ecosystem, a land use/land cover classification at ecotope level with definitive biological and ecological characteristics is needed. Ecotopes are distinct ecological landscape features that provide information concerning biodiversity and environmental quality. In order to have adequate reference to the large variety of ecological conditions, the classification scheme of ecotopes is usually detailed and elaborate. While for decades remote sensing imagery has been used to generate land use/land cover maps, it will be a challenge to map ecotopes using conventional broadband multi-spectral data with coarse spatial resolution (Ellis et al., 2006). A better option will be hyperspectral images which provide a fuller spectral description with up to hundreds of narrow bands. Hyperspectral data are reportedly

competent in achieving substantially higher classification accuracies in the recognition of vegetation and agricultural crops (Bork et al., 1999, Thenkabail et al., 2004), and might have the potential to separate vegetation into taxonomic levels (Ustin et al., 2004). To classify ecotopes using hyperspectral data, however, we are faced with a classical problem of high dimensional inputs (hundred of bands) and outputs (many classes). For this complex problem, we need powerful algorithms for the classification and band selection methods that can reduce input dimension.

Recent studies on difficult classification problems using multi-source and hyperspectral data have pointed to the superiority of Random Forest, a type of ensemble classification that uses decision tree as the base classifier (Crawford et al., 2003; Gislason et al., 2006; Ham et al., 2005, Lawrence et al., 2006). Ensemble classification represents a new approach that uses not one, but many classifiers. Hundreds of classifiers are built and their decisions combined usually by plurality vote. The premise is that combining ensemble classifiers is often more accurate than any one from the ensemble. Ensemble classifications can be categorized into those based on several different learning algorithms and those based on just one. In the former case, several classifiers can be

* Corresponding author. Tel.: +32 2629 3556; fax: +32 2629 3378.

E-mail addresses: jonathan_cw_chan@yahoo.com, Cheung.Wai.Chan@vub.ac.be (J.C.-W. Chan).

generated using the same training set, each from a different learning algorithm and their decisions are combined by simple voting, or more sophisticated methods like consensus theory (Benediktsson & Swain, 1992) and stacking (Džeroski & Ženko, 2004). The drawback of this ensemble method is to have to handle different learning algorithms and that increases the complexity of processing. Combining methods can sometimes be complicated too. Finally, its effectiveness relies very much on the combining technique and an ensemble does not always give a more accurate classification (Foody et al., 2007).

Another form of ensemble classification based on only one learning algorithm and the ensemble is created by changing the training set. Bagging (Breiman, 1996), or bootstrap aggregating, and Adaboost (Freund & Schapire, 1996), or adaptive reweighting boosting, are two popular methods to generate new training sets. Both methods have been thoroughly studied (Bauer & Kohavi, 1998; Dietterich, 2000) and have been assessed with remote sensing data (Chan et al., 2001; DeFries & Chan, 2000; Pal & Mather, 2003). While theoretically Adaboost and bagging can be applied to any learning algorithms, they are originally designed to boost the accuracy of “weak” learners such as decision trees (Breiman, 2001). Implementations of Adaboost using decision trees as the base classifier showed that the results are comparable to support vector machines (Chan et al., 2001) and a neural network (Pal & Mather, 2003). Comparisons between bagging and Adaboost showed that Adaboost had higher accuracies but bagging was more stable (Chan et al., 2001; DeFries & Chan, 2000). But Adaboost was also described as susceptible to noise and more costly in processing time (Briem et al., 2002).

Random Forest is a tree-based ensemble classifier that uses the bagging technique to create new training sets. It includes two important methods: random feature subspace and out-of-bag estimates. The former enables a much faster construction of trees and the latter the possibility of evaluating the relative importance of each input feature. Breiman (2001) ran Random Forest on 20 data sets from different data domains, and his results showed that the performance of Random Forest is superior to other learning algorithms. When compared with Adaboost, he concluded that Random Forest is “favorably” comparable, only that Random Forest is cheaper in terms of computing time.

Ensemble classifications that build numerous classifiers are not an intuitive choice for analyses of hyperspectral data since they naturally add more computational burden to a procedure already complicated by high dimensional inputs. However, since decision trees are extremely fast to build, the costs of training hundreds of them will still be cheaper than training one artificial neural network, as demonstrated in Pal and Mather (2003). While Random Forest, a tree-based ensemble classifier, has been shown effective when applied to classification of hyperspectral data (Crawford et al., 2003; Ham et al., 2005; Lawrence et al., 2006), there has not been adequate discussion on Adaboost. Hence, it is natural to explore more about the potential of these two powerful algorithms in our specific application.

While hyperspectral data provides more spectral information, it also brings new challenges such as the curse of dimensionality. Since neighboring hyperspectral channels carry highly correlated information, some spectral bands are redundant and some contain possible noise. The importance of feature (band) selection is not only to reduce input dimensionality which will subsequently alleviate processing burden, but also to deepen our understanding of which spectral ranges are most suitable for ecotope mapping. This knowledge of the best wavebands might also aid future sensor designs dedicated for specific applications. Many methods have been proposed for band selection (Chang et al., 1999; Guo et al., 2006; Jia & Richards, 1994; Kaewpajit et al., 2003; Lee & Landgrebe, 1997; Melgani & Bruzzone, 2004) but some suggested using a basket of different methods instead of using one method. In Bajcsy and Groves (2004), seven techniques for band ranking were included in their regression model for soil electrical conductivity. In Thenkabail et al. (2004), four assessment techniques namely principal component analysis, lambda-lambda R2 models, stepwise discriminant analysis, and derivative greenness vegetation

indices were employed to find the optimal wavebands for the classification of vegetation. The reason for using a basket of methods is attributed to the no-free-lunch theorem which argues against the possibility of a universally superior algorithm for all problems. This echoes with a machine learning concept that for any specific problems, as many techniques and algorithms should be investigated as manageable possible because there is no one method which has been demonstrated to be superior for all problems (Kohavi et al., 1996).

In this paper, we assess two benchmark tree-based ensemble classification algorithms: Adaboost and Random Forest, using the criteria of standard classification accuracy, computing time and stability (DeFries & Chan, 2000). Two band selection algorithms are tested: the out-of-bag method (Breiman, 2001) and a wrapper approach feature subset selection (Kohavi & John, 1997) using the best-first search method (Ginsberg, 1993). In Thenkabail et al. (2004), hyperspectral data were gathered for shrubs, grasses, weeds, and agricultural crop species from four ecoregions of African savannas; after rigorous investigation with four assessment methods, they recommended 22 optimal bands. Due to the similarities of the target classes and the studied spectral range (0.4–2.5 μm) between Thenkabail et al.'s (2004) study and ours, we think there are reference values to investigate the optimal bands suggested in their article. Following the no-free-lunch theorem, we combine Thenkabail's selection with the outcomes from the other two band selection methods. All band subset selections are tested using both Adaboost and Random Forest.

2. Study area and data

For this study, we have used the ecotope classification in Belgium, locally known as the Biological Valuation Map (BVM), as a case study (see www.inbo.be/bvm). Under the BVM, the entire land surface is divided into 162 detailed classes according to soil conditions or agricultural practices. The BVM is produced by field surveys. Experts including biologists, agronomists and forest ecologists are sent to the field and visually identify features that are ecologically distinct. The map produced by the experts is treated as the groundtruth without further accuracy assessment. Using four criteria namely rarity, vulnerability, naturalness, and replaceability, each BVM class is assigned a qualitative biological value: very high biotic value, high biotic value and little biotic value (De Blust et al., 1994). This information is important to assess the state of the environment. Our study area is situated east of the town Geraardsbergen in the province of East-Flanders within the so called “South Flemish loamy hill district” of Belgium (Fig. 1). The valley of the river Dender crosses the northern part. Species rich improved grasslands dominate this valley, but also semi-natural grasslands (called Calthion-grasslands), poplar plantations, alluvial and oak forests are present. The central and southern parts are also crossed by some other small valleys of rivulets and wet depressions dominated by the same type of vegetation. The forests area in the central part (the Moerbeke forests area) consists mostly of oak forests, and the rest belongs to those of beech forests and woodland of alluvial soils. The interfluvies mostly consist of arable land, with a few of species poor to species rich improved grasslands. The area is particularly suitable for our study because of a recent and detailed survey done by only one field researcher. This leads to a high level of uniformity of the mapping.

An adapted classification scheme with thirty-two classes was used (Wils et al., 2004). To allow a more realistic assessment of the classification on vegetation types, water surface and built-up areas (houses and roads) which are known to be spectrally distinct from vegetation were masked out in our experiments. Sixteen classes were found to be present in our study area (Table 1). The tree classes include scrubs, orchard and woods. The grassland classes represent different degrees of species richness. Higher biological values are attached to grassland types that are comparatively richer in species. From high to

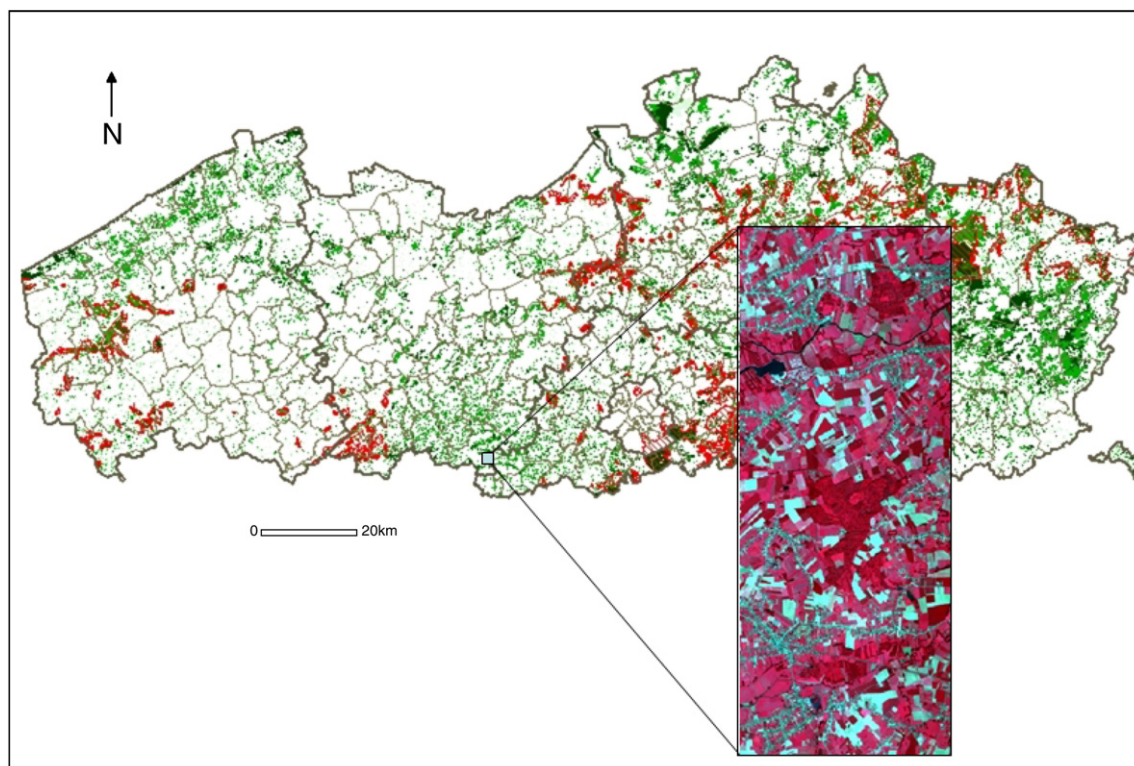


Fig. 1. Study area. The background is an abstracted draft Biological Valuation Map covering the entire Flanders in Belgium. Ecotopes are assigned different degrees of biological values. Locations of medium to high biological values are in green. White color represents less biological values and red areas are important for rare animal species (this layer is still under construction). The zoom shows the coverage of the airborne HyMap image.

low biological values, the grassland classes rank in the following order: semi-natural grasslands (*hpgh*), species rich improved grassland (*hpgs*), grasslands with patches *hp* and either patches *hpgs* or *hpgh* (*hpv*), species poor improved grassland (*hp*) and grass monocultures (*hx*).

Airborne HyMap data was acquired on 8th June, 2004. Pixel size is 4×4 m with 126 bands distributed between 0.4 and $2.5 \mu\text{m}$. A level 2 product was generated after radiometric correction, geometric correction (ortho-rectified using bilinear resampling), atmospheric correction and calibration. The image is geo-referenced in UTM with WGS-84 Geodetic datum. Fig. 2 shows the overlay of the ecotope classes over the airborne HyMap data. The dimension of the two mosaic strips is 953×2330 pixels covering an area of approximately 35.5 km^2 . Field works were carried out during the day of the overpass and the day after. A spectroradiometer ASD Fieldspec Pro FR (350 nm–2500 nm) was used for ground measurements. Those measurements were gathered to calibrate the airborne data as well as to assess the data quality. To prepare the training and validation samples, the groundtruth vector layer was first filtered by excluding the following land covers: fallowed land, built-up area and water surface. Then, samples were extracted manually and randomly divided into two equal halves for training and validation (Table 2). The mean spectrum of the grassland and the tree classes are presented in Fig. 3a and b, respectively. The spectral profiles are artificially linked at the absorption regions around $1.35\text{--}1.4 \mu\text{m}$ and $1.8\text{--}1.95 \mu\text{m}$, hence the strange shapes.

3. Methods

In this section, we recapture some important properties of the ensemble classification algorithms and band selection methods. But first we will briefly describe the basic attributes of a decision tree classifier. By applying certain statistical measures, a decision tree

recursively partitions the training data into more homogenous subsets. A decision tree is composed of nodes at different levels: a root node, a set of internal nodes (splits) and a set of terminal nodes (leaves) (Fig. 4). The root node is formed from all of the training samples. After a selected statistical measure is applied, the training samples are split into two descendent nodes. The same procedure is

Table 1
Description of the ecotope classes

ID	Category	Class code	Definitions
1	grassland	<i>b</i>	Arable land
2	grassland	<i>hp</i>	Species poor improved grassland (normally more homogenous for the whole parcel)
3	grassland	<i>hpgh</i>	Semi-natural grassland
4	grassland	<i>hpgs</i>	Species rich improved grassland (between <i>hpgh</i> and <i>hp</i>)
5	grassland	<i>hpv</i>	Grassland with patches <i>hp</i> and either patches <i>hpgs</i> or <i>hpgh</i>
6	grassland	<i>hx</i>	Grass monocultures (equal to arable land sown with grasses of one or more years)
7	tree/tall_veg	<i>f</i>	Deciduous forest, dominated by beech (<i>Fagus</i> sp.)
8	tree/tall_veg	<i>gml</i>	Plantation of deciduous tree species other than beech, oak, alder and poplar
9	tree/tall_veg	<i>kj</i>	Tall tree orchard
10	tree/tall_veg	<i>kl</i>	Low tree orchard
11	tree/tall_veg	<i>lh</i>	Poplar plantation
12	tree/tall_veg	<i>p</i>	Conifer plantation
13	tree/tall_veg	<i>q</i>	Deciduous forest, dominated by oak trees (<i>Quercus</i> sp.)
14	tree/tall_veg	<i>sc</i>	Scrubs of clearings and scrubs on abandoned land
15	tree/tall_veg	<i>sp</i>	Thorn ticket
16	tree/tall_veg	<i>v</i>	Woodland of alluvial soil, fens and bogs (mostly dominated by alder = <i>Alnus</i> sp.)

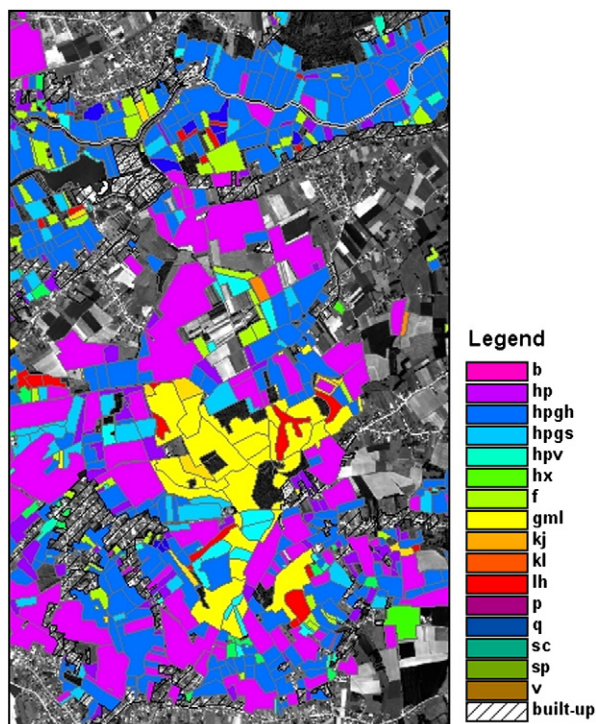


Fig. 2. Overlay of the recently updated Biological Valuation Map on top of the HyMap image at the area of Dender.

then applied to the descendent nodes. This sequential applications of the measure will form a tree-structured decision space with the terminal nodes representing the assignment of classes. An unknown pixel can be labeled by following the decision framework and is assigned to the class the same as the leaf node which it finally falls into. Different approaches can be used for tree construction. The more frequently used measures include the gini index (Breiman et al., 1984), the information gain ratio (Quinlan, 1993) and the chi-square measure (Mingers, 1989a). These measures dictate how the data is split at each node and also find the best thresholds for the splits. Furthermore, at each node, the split can be based on either several bands or just only one band. The tree is multivariate if each split is based on several bands, or univariate if each split uses only one band. Multivariate trees can have better performance when decision boundaries are better described by a combination of features (bands), but they also introduce complexity in calculations (Friedl & Brodley, 1997). In this study, a univariate decision tree is used. Finally, a key element in tree construction is to correct overfitting by pruning back the tree. Since the tree is grown in a way such that all the training data are correctly classified, if there are errors in the training set, an overfitting tree will lead to poor performance in classifying data not used in the training. Pruning can be done according to certain cost-complexity measures which allow a split only if the contribution to accuracy improvement has reached a predefined threshold. Some common pruning methods are discussed in Quinlan (1987) and Mingers (1989b). As this paper focuses more on ensemble classification, we refer the reader who is interested in more detailed explanation about the construction of decision trees to Pal and Mather (2003), Friedl and Brodley (1997), or Quinlan (1993).

Decision tree classifiers have many advantages. Unlike most of the conventional classifiers which derive a complex membership decision using all the input bands at once, decision trees modulate a classification problem into a hierarchy of simple decisions with each decision based on only one or several of the input bands. This strategy makes the computing burden particularly light. A decision tree clas-

sifier is also non-parametric, so there is no requirement on data distribution. The structure of the tree provides information about which of the input bands have been used for classification. This helps us to understand which bands are more important for our applications. However, decision trees are considered as ‘weak’ learners meaning it is not the most accurate classification algorithm. Next we introduce the two benchmark ensemble algorithms that are used for boosting the accuracy of decision trees.

3.1. Random Forest and Adaboost

In Random Forest, the method to build an ensemble of classifiers is to change the training set using the same strategy as bagging (Breiman, 1996). Bagging creates new training sets by resampling from the original data set n times, n being the number of samples in the original training set, randomly with replacement. This means the sample just being chosen will not be removed from the data set in the next draw. Hence, some of the training samples will be chosen more than once while some others will not be chosen at all in a new set. Bagging helps classification accuracy by decreasing the variance of the classification errors. In another words, it taps on the instability of a classifier. ‘Instability’ of a classifier means that a small change in the training samples will result in comparatively big changes in accuracy. The classifiers are combined by a majority vote and the vote of each classifier carries the same weight. In the case of a tie, the decision can be made randomly or by prescribed rules.

Random Forest creates multiple trees using the impurity gini index (Breiman et al., 1984). However, when constructing a tree, Random Forest searches for only a random subset of the input features (bands) at each splitting node and the tree is allowed to grow fully without pruning. Since only a portion of the input features is used and no pruning, the computational load of Random Forest is comparatively light. The computing time is in the order of $T\sqrt{MN}\log(N)$ where T is the number of trees, M is the number of bands used in each split, and N is the number of training samples (Breiman, 2001). In addition, in case a separate test set is not available, an out-of-bag method can be used. For each new training set that is generated, one-third of the samples are randomly left out, called the out-of-bag (OOB) samples. The remaining (in-the-bag) samples are used for building a tree. For accuracy estimation, votes for each sample are counted every time when it belongs to OOB samples. A majority vote will determine the final label. Only approximately one-third of the trees built will vote for each case. These OOB error estimates are unbiased in many tests (Breiman, 2001). The number of features for each split has to be defined by the user, but it is insensitive to the algorithm. Majority vote is used to combine the

Table 2
Number of samples in training and testing

ID	Classes	Number of training	Number of testing
1	<i>b</i>	3712	3711
2	<i>hp</i>	928	928
3	<i>hpgh</i>	939	939
4	<i>hpgs</i>	6277	6277
5	<i>hpv</i>	910	909
6	<i>hx</i>	286	285
7	<i>f</i>	1663	1662
8	<i>gml</i>	852	851
9	<i>kj</i>	369	369
10	<i>kl</i>	607	607
11	<i>lh</i>	2802	2802
12	<i>p</i>	320	319
13	<i>q</i>	4907	4906
14	<i>sc</i>	675	675
15	<i>sp</i>	361	360
16	<i>v</i>	1259	1259
	<i>Total</i>	26,867	26,859

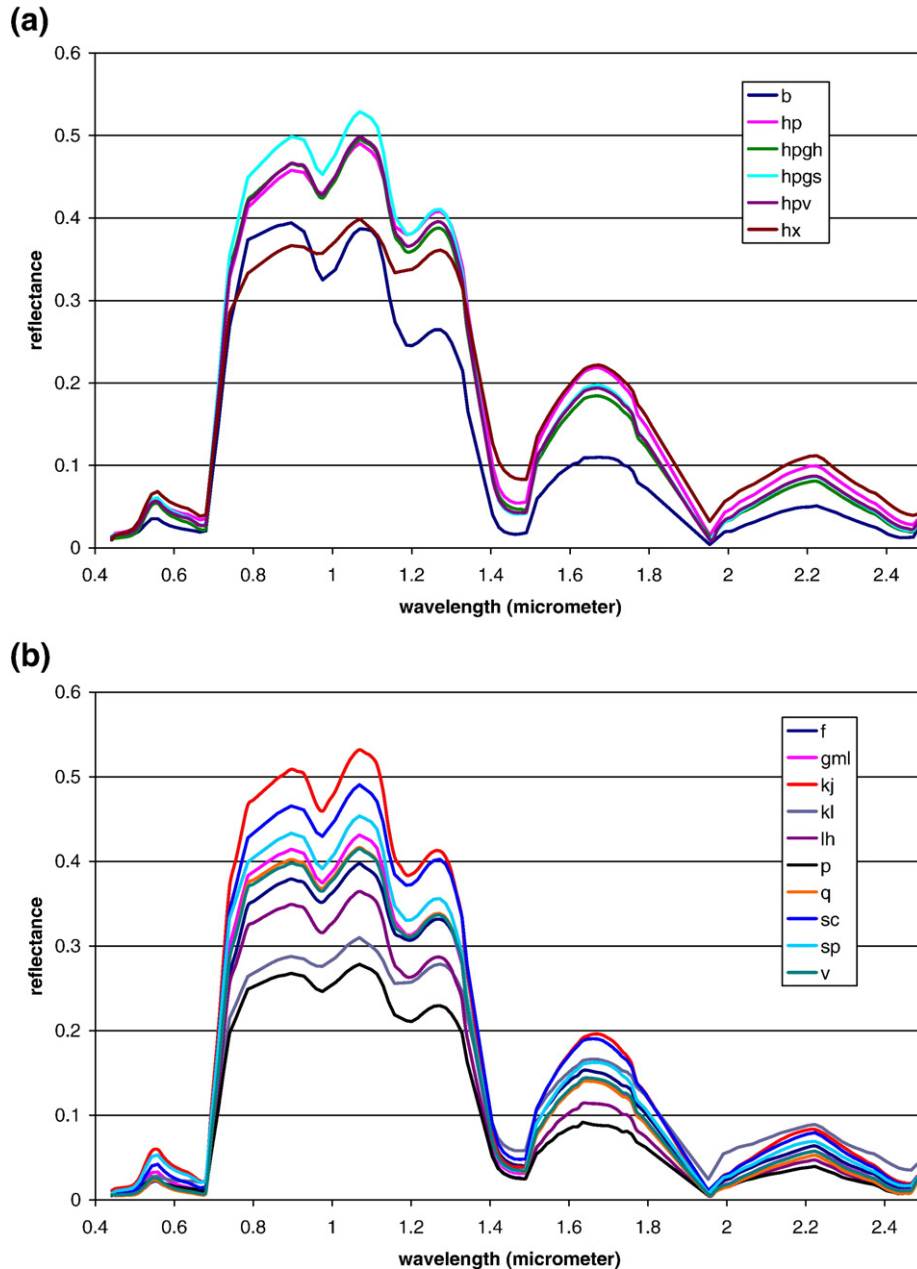


Fig. 3. Mean spectrum of the (a) six grassland classes and (b) ten tree classes.

decisions of the ensemble classifiers. For a full detail of Random Forest, the reader is referred to the manual (http://oz.berkeley.edu/~breiman/RandomForests/cc_home.htm) or Breiman (2001). A very good summary can also be found in Gislason et al. (2006).

Adaboost (Freund & Schapire, 1996) changes the distribution of the original training set in a different manner. If ϵ_t is the misclassification rate at trial t , then the weights of misclassified samples in the training set will be changed by the factor $\beta_t = (1 - \epsilon_t) / \epsilon_t$. The total sum of the adjusted weights is normalized to 1. The classifiers C_1, \dots, C_t are combined according to the weighted voting where C_t is weighted by $\log(\beta_t)$. The trials will be terminated if $\epsilon_t > 0.5$ (which means the classifier's accuracy is worse than a random guess) and trial T is altered to $t - 1$. If $\epsilon_t = 0$ (i.e. 100% accuracy), then trial T becomes t . In so doing, Adaboost forces the newly constructed ensemble classifiers to focus on 'harder' cases. The voting for the final labels is weighted by the accuracy of each classifier. In this study, Adaboost is implemented using C5.0 where the trees are constructed using the information gain ratio (Quinlan, 1993).

3.2. Band selection methods

The first method is the OOB method implemented in Random Forest (Breiman, 2001). Though the structure of a decision tree provides information concerning important features, such interpretation is rather impossible for hundreds of trees in an ensemble. One additional feature of Random Forest is its ability to evaluate the importance of each input feature by the internal OOB estimates. To evaluate the importance of each feature (band), the values of the m feature (band) of the OOB samples would be allowed to permute. The perturbed OOB samples will run down on each tree again. The differences in accuracy between the original and the perturbed OOB samples over all the trees grown in the Random Forest are averaged. This number will become the importance score of the m feature (band) and is used as a ranking index. We will use the 25 highest ranking bands from the OOB method.

The second band selection method we used is a wrapper method using the best-first search as the search algorithm (Kohavi & John, 1997).

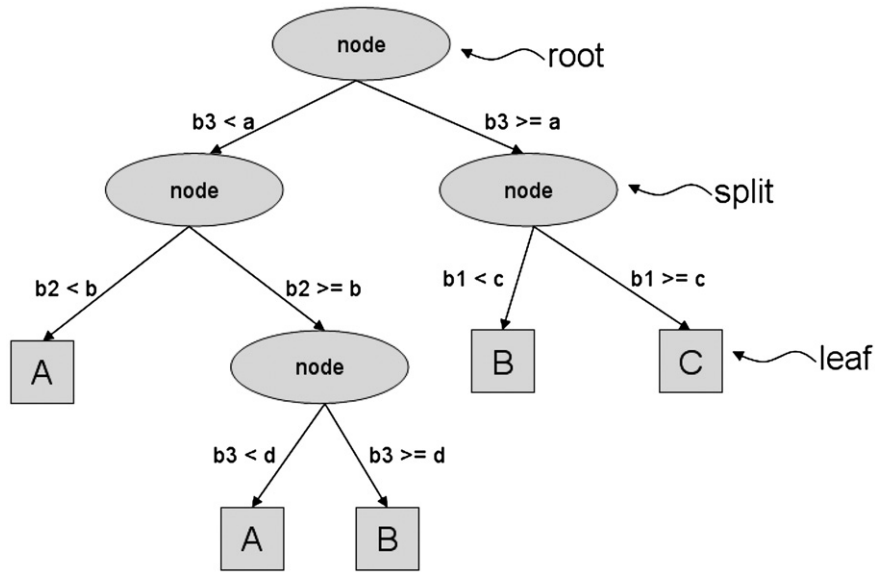


Fig. 4. A hypothetical decision tree classifier with three input bands (b1, b2 and b3). At each of the root and internal nodes (splits), a statistical measure is applied. The values a , b , c and d are thresholds for splitting. A data set is subdivided into smaller groups until the terminal nodes (leaves) which contain the class labels (A, B and C). This figure is adapted from Fig. 1 in Friedl and Brodley (1997).

Current feature selection techniques can be classified as either filter approaches or wrapper approaches. A filter approach defines relevancy of features based on the training data alone, and the feature subsets are chosen independently of the learning algorithm. It is like a preprocessing step. Wrapper approaches, however, search for the best feature subset using the learning algorithm as part of the evaluation function. The feature subset with the highest accuracy is selected. Wrapper approaches are reportedly producing superior results than filter approaches (Kohavi & John, 1997). While wrapper approaches are producing better accuracy, it is computationally heavy because it has to call on the learning algorithm for the evaluation of each subset. Hence, a faster search engine is needed. Conventional methods that exhaustively search the entire search space (all feature combinations) are impractical unless the number of feature is small. The best-first search is a method that avoids searching the entire search space (Ginsberg, 1993). The idea of best-first search is to jump to the most promising band subset which is defined as the subset with the highest accuracy. Fig. 5 illustrates the paths of the best-first search algorithm using a scenario with four bands. Each bit within the square brackets shows the presence (1) or absence (0) of a band. For example, [0, 1, 0, 1] represents a subset with the second and the fourth bands. The number next to each bracket is the accuracy of that subset evaluated by the learning algorithm. The search starts with no feature, hence [0, 0, 0, 0]. Then, it proceeds to turn on consecutively each of the four bands. Since the first band attained the highest accuracy (65%), the search continues to expand on its children which represent new subsets that always have the first band present. The same logic applied to the next expansion where adding the fourth band to the first gives the highest accuracy. The search stops when the highest attainable accuracy was achieved. Since it is not always possible to achieve 100% accuracy in reality, the search is stopped when the estimated accuracy no longer improves, and the feature subset at that search state represents the optimal subset.

This method has been tested on 18 real-world datasets from the U.C. Irvine machine learning repository in Kohavi and Sommerfield (1995), and 14 datasets in Kohavi and John (1997). Their results showed that for most data sets, the feature subsets increased accuracy. In rare cases, the feature subsets would lower accuracy. When applied in remote sensing data, the method generated a substantially smaller subset but the accuracies were comparable with the entire feature set (Chan et al., 2000). We applied this method using the Feature Subset Selection

routine within the Machine Learning Library in C++ (Kohavi et al., 1996). The search is terminated if the last 5 expansions see no increase of accuracy that is more than 0.001%.

4. Experiments and results

We implemented the same procedures for Adaboost and Random Forest. A hundred trees were built and the accuracy snapshots of each additional ten trees were shown. Adaboost was run using C5.0 with a default pruning of 25%. The only parameter with Random Forest is the number of features (bands) used for each split. Following the suggestion in the manual, this parameter was tried from 1 to 12 (square root of the number of input features) to build 10 and 20 trees. The number that has the highest accuracy with the test set is chosen (Table 3). In the case of using all 126 bands, this number is twelve. The accuracy of Adaboost after 99 trials is 69.5%, a 9% increase compared to a single C5.0 tree (60.2%). However, accuracy already reached 69% at 29 trials (Fig. 6). In the case of Random Forest, one tree produced an accuracy of 57.1%, 3% lower than a C5.0 tree. Generating 70 or 90 trees increased the accuracy to 68.6% which is comparable to Adaboost. The accuracies of Random Forest and Adaboost are both higher than conventional classifiers. With the same training and test samples, a

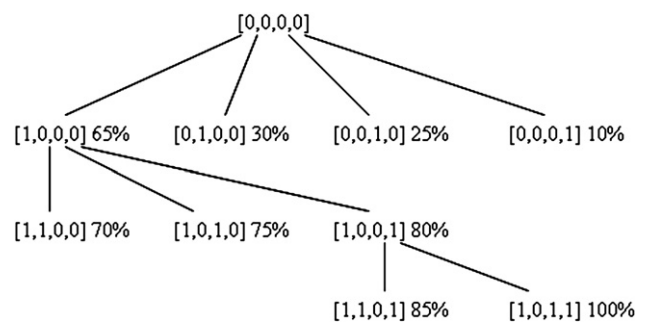


Fig. 5. An example showing the paths of the best-first search in a case with four bands. Zeros and ones within the square brackets represent the absence and the presence, respectively, of each band. Instead of exhaustively looking into all possible combinations, the best-first search jumps to the most promising subset – the subset that has achieved the highest accuracy so far, and expands on their children nodes.

Table 3
Accuracies for the out-of-bag estimates and the test set with five different band subsets

Trees	ALL				First-Best				Thenkabail				RF_25				Combine_53			
	10	20	10	20	10	20	10	20	10	20	10	20	10	20	10	20	10	20	10	20
Split	Out-of-bag acc. (%)	Test set acc. (%)	Out-of-bag acc. (%)	Test set acc. (%)	Out-of-bag acc. (%)	Test set acc. (%)	Out-of-bag acc. (%)	Test set acc. (%)	Out-of-bag acc. (%)	Test set acc. (%)	Out-of-bag acc. (%)	Test set acc. (%)	Out-of-bag acc. (%)	Test set acc. (%)	Out-of-bag acc. (%)	Test set acc. (%)	Out-of-bag acc. (%)	Test set acc. (%)	Out-of-bag acc. (%)	Test set acc. (%)
1	72.40	65.52	77.20	67.05	74.02	63.73	78.78	65.76	73.18	66.01	78.10	67.12	71.29	65.72	76.43	67.02	73.79	65.75	78.32	67.59
2	74.40	65.72	79.01	67.38	76.07	65.64	80.22	66.85	75.16	65.76	79.65	67.07	73.15	65.39	77.84	66.73	75.50	66.53	80.08	68.03
3	75.21	65.62	79.78	67.01	76.20	65.30	80.34	66.44	76.11	66.42	80.33	67.60	74.07	65.56	78.38	66.57	76.22	66.98	80.68	67.75
4	76.07	65.89	80.35	67.06	76.99	65.65	80.85	66.83	76.41	66.27	80.78	67.30	74.08	65.42	78.61	66.57	76.71	66.62	81.06	67.71
5	76.44	66.43	80.70	67.57	76.85	65.87	80.93	67.29	76.66	66.27	80.88	67.64	74.45	65.62	78.96	66.71	76.85	66.51	81.42	67.44
6	76.57	66.03	80.74	67.65	77.24	65.71	81.19	66.72	76.78	66.72	81.08	67.69	74.41	64.93	78.62	66.04	77.21	66.83	81.45	67.95
7	76.70	66.64	81.11	67.58	77.07	65.62	81.32	66.62	77.23	66.79	67.69	67.57	74.68	64.85	78.81	66.12	77.27	66.93	81.46	68.02
8	77.19	66.62	81.55	67.71	77.13	65.38	81.42	66.46	77.06	66.93	81.39	67.87	74.79	64.70	79.19	65.80	77.37	66.46	81.49	67.86
9	77.00	66.27	81.33	67.39	77.26	65.78	81.36	66.73	77.50	66.94	81.26	68.05	74.85	64.99	79.08	66.00	77.94	66.68	81.88	67.94
10	77.50	66.44	81.41	67.67	77.79	65.60	81.75	66.63	77.38	66.62	81.26	68.05	74.79	64.94	79.32	66.17	77.46	67.02	81.79	68.14
11	77.20	66.53	81.30	67.63	77.73	65.63	81.62	66.43	77.63	66.75	81.45	67.82	74.68	64.60	78.77	66.01	77.62	67.01	81.80	67.82
12	76.99	66.65	81.46	67.71	77.66	65.79	81.51	66.23	77.79	66.21	81.71	67.38	74.62	64.84	78.96	65.91	77.81	67.25	81.80	68.46

The numbers for the splitting is tried from 1 to 12 and 10 and 20 trees are built.

multi-layer perceptron produced 63.7% accuracy and a Spectral Angle Mapper is even lower, under 30%.

The selections from the OOB method under the Random Forest, and the wrapper approach with the best-first search algorithm (hereafter we use 'best-first' to represent this method) are listed in Table 4. In order to have a comparison of the band distribution of different selection methods, the OOB, the best-first, and the Thenkabail's subsets are plotted in Fig. 7. The waveband regions are named following the convention in Thenkabail et al. (2004): VIS=visible (0.450–0.750 μm); NIR=near infrared (0.751–1.050 μm); FNIR=far near-infrared (1.051–1.300 μm); ESWIR=early short-wave infrared (1.301–1.900 μm); FSWIR=far short-wave infrared (1.901–2.500 μm). The best-first method selected 21 out of 126 bands. The OOB method in Random Forest has ranked each feature according to their importance in accuracy. The first 25 bands are used and their scores are plotted in Fig. 7. The bands selected by the best-first and the OOB methods have two similarities in terms of the waveband regions. First, both have skipped almost the entire spectral region at 0.8–1.2 μm which belongs to the NIR and the FNIR. Second, both methods have a large number of bands located within the SWIR, in particular the ESWIR. The best-first and the OOB methods have selected respectively thirteen and seventeen bands from

the SWIR. A clear concentration is found in the ESWIR with nine of the 25 OOB bands and eight of the 21 best-first bands.

Some clear differences can be found between the band distributions of the OOB and the best-first subsets and that of the Thenkabail's subset. Thenkabail's bands are more evenly distributed and have higher concentrations in the VIS, the NIR and the FNIR (0.4 to 1.3 μm). The OOB and the best-first subsets, however, focus more on the ESWIR and the FSWIR, especially between 1.3 and 1.9 μm. Previous studies have shown that the SWIR has high correlation with water thickness and plant moisture content (Hardisky et al., 1983; Yilmaz et al., 2008). The importance of the SWIR for vegetation and cropland is well-known (Thenkabail et al., 1994) and actually it was also the goal in Thenkabail et al.'s (2004) study to explore the potential role of the SWIR. Their results showed that FNIR bands were more effective than SWIR bands in increasing accuracies. The importance of SWIR in our study area is more apparent as illustrated by the results of the band selection algorithms. It should be noted that climatic conditions in the African savanna ecoregions are quite different from our study area due to the latitudinal and geographical differences. For example, precipitation contrast between dry and wet seasons is high in African savanna. Comparatively, precipitation in our study area of Belgium is

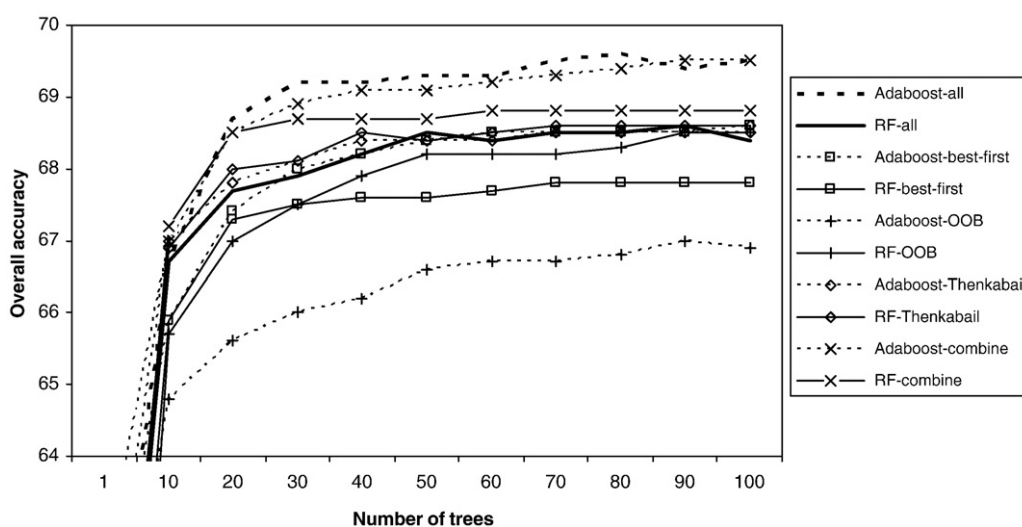


Fig. 6. Snapshots of overall accuracies at 10 trees interval for Adaboost and Random Forest using different band subsets. Their results are compared to using all the bands (Adaboost-all and RF-all). The accuracies with only one tree are not shown in order to have a better visualization of different input scenarios.

Table 4
Selected bands from different methods

Number of bands	Best-first search (21 features)		Thenkabail (2004) (22 features)		Out-of-bag (25 features)	
	Band ID	Band centers μm	Band ID	Band centers μm	Band ID	Band centers μm
1	5	0.497	5	0.497	2	0.452
2	7	0.528	9	0.558	4	0.482
3	9	0.558	15	0.650	19	0.711
4	14	0.635	17	0.680	20	0.725
5	15	0.650	19	0.711	21	0.740
6	18	0.695	21	0.740	22	0.756
7	19	0.711	31	0.879	24	0.786
8	55	1.244	33	0.912	25	0.801
9	66	1.448	38	0.991	63	1.405
10	67	1.462	44	1.084	64	1.420
11	70	1.504	47	1.129	65	1.434
12	74	1.558	53	1.216	66	1.448
13	77	1.597	55	1.244	67	1.462
14	78	1.610	57	1.273	71	1.517
15	89	1.747	66	1.448	74	1.558
16	94	1.806	84	1.685	82	1.661
17	100	2.047	87	1.722	87	1.722
18	108	2.188	97	1.991	96	1.972
19	112	2.257	99	2.028	97	1.991
20	118	2.356	111	2.24	109	2.205
21	119	2.372	113	2.274	110	2.223
22			116	2.323	112	2.257
23					113	2.274
24					124	2.451
25					126	2.467

constant and remains high throughout the year. These differences will bring different phenologies and could contribute to the differences in band selections.

We applied Adaboost and Random Forest using different feature subsets. In the case of Random Forest, the number of features (bands) for each split for the best-first subset and the Thenkabail's subset are 5 and 9, respectively (Table 3). For the OOB subset, there was a disagreement between the results from 10 and 20 trees. An average was taken and the

Table 5
Overall accuracy of using different band subsets

	Number of bands	Adaboost	Random Forest
All bands	126	69.5%	68.6%
Thenkabail	22	68.5%	68.6%
Best-first	21	68.6%	67.8%
Out-of-bag	25	66.9%	68.5%
Combine all subsets	53	69.5%	68.8%

number with the highest average accuracy rate from the test set was chosen. The number for the split is one. The accuracies of Adaboost and Random Forest on all different subsets are very comparable at around 68.5% (Table 5). The two worst cases are found to be Adaboost with the OOB subset where the accuracy is 2.6% lower and Random Forest with the best-first search subset, the accuracy is 0.8% lower. Comparatively, Random Forest is slightly more robust than Adaboost with different inputs because its accuracy almost does not change. Apparently, all feature subset selection methods produce rather good results even though the selected bands have rather different distributions.

The OOB is a data-driven method. It is like a *leave-one-out* method, where each band is left out consecutively to check its influence on accuracy, cleverly implemented in an ensemble scenario. It is much faster than the best-first method which is heuristic and does not search the entire search state, or all the combinations. The result of the best-first is not considered optimal, or unique. However, it has a better average accuracy over the two classifiers (Table 5).

A last experiment was performed combining all feature subsets. A combined subset was formed with 53 bands (see its distribution in Fig. 7). The number of split for Random Forest is found to be 12 (see Table 3). The accuracies of Random Forest and Adaboost are 68.8% and 69.5%, respectively. Comparing to the entire set, the combined set has attained the same accuracy with Adaboost and even a little higher accuracy (+0.2%) with Random Forest.

Figs. 8 and 9 show the accuracy performance of different inputs at class level for Random Forest and Adaboost. Random Forest has higher accuracies with *v*, *hpg* and *hpgs*, while Adaboost does better with *hx*, *kj*, *kl*, *lh*, *q* and *sc*. In terms of input subsets, there is no evidence that any feature subsets has absolute advantage over any individual classes.

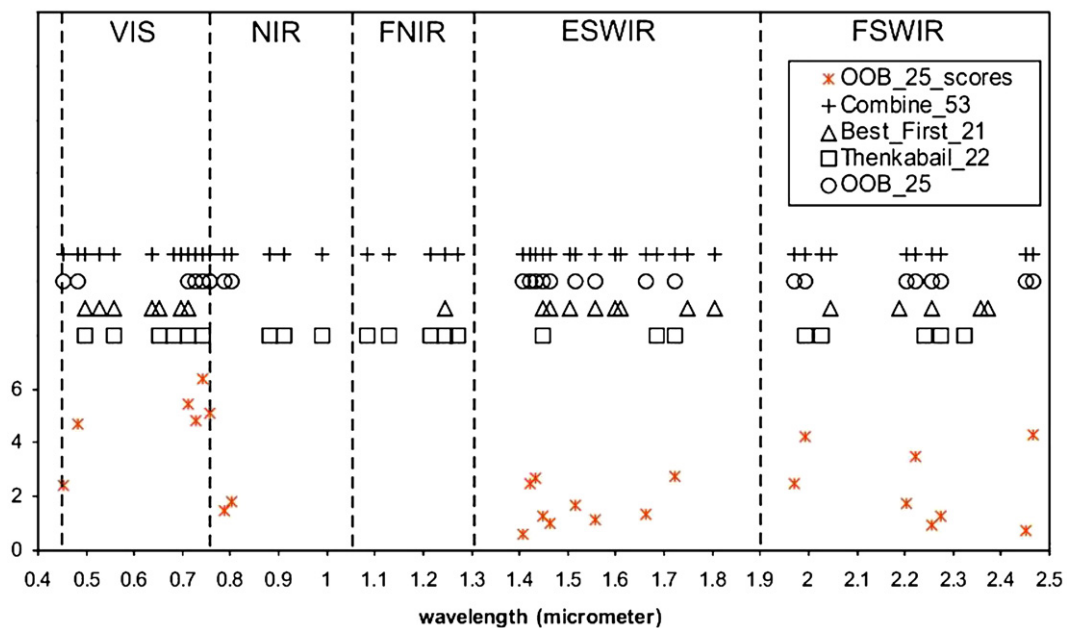


Fig. 7. The distributions of selected bands from the best-first method (Best_First_21), the out-of-bag method (OOB_25) and the optimal wavebands from Thenkabail et al. (2004) (Thenkabail_22). The scores of the 25 highest ranked features using OOB estimates are also shown.

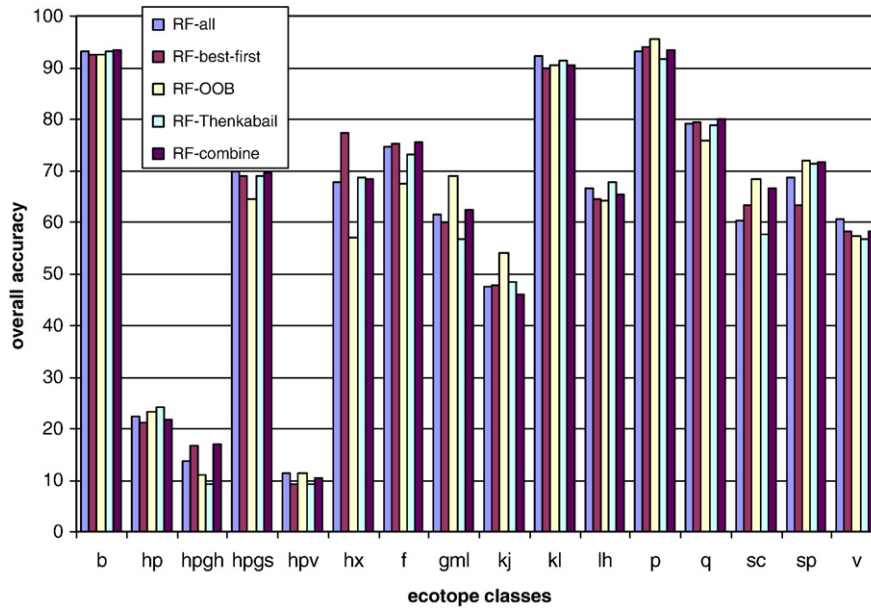


Fig. 8. Class accuracies from Random Forest using different band subsets. RF-all represents the accuracies with all the 126 bands.

As mentioned above, Random Forest uses only a portion of the input features for each split which makes it computationally lighter. Fig. 10 shows the differences in time between these two algorithms. It is based on a Pentium IV machine at 3.4 GHz. For Adaboost, 100 trees (or with 99 boosting trials) with 126 bands takes a bit more than 35 minutes, compared to 5 minutes with Random Forest. Using 22 bands to build a hundred trees, Random Forest will take one and a half minutes compared to almost 5 minutes with Adaboost. This difference in computational time will grow significantly larger with higher number of bands. Naturally, this will become important only when the classification has to be finished in a very short period of time, as in an emergency situation, or it has to be repeated hundreds and thousands of times. For the application in this study, the difference in elapsed time of the two classifiers is insignificant.

Another important criterion in the operational context, other than accuracy and computing time, is stability. An algorithm gives stable results implies that it is not over-sensitive to variability in reflectances

caused by bidirectional effects, solar zenith angle, or other factors unrelated to a change of land covers (DeFries & Chan, 2000). A minor change in the training data does not alter greatly the classification result. We assess the stability of the two ensemble classifications using the κ -error diagram (Margineantu & Dieterich, 1997) which has been applied in classification algorithms using remotely sensed data (Chan et al., 2001, DeFries & Chan, 2000). This diagram helps to visualize the relationship between stability and accuracy of an algorithm. The original training set was randomly sampled 10% ten times, creating 10 new training sets. This resampling approximates a minor change in the training set. A degree-of-agreement statistic κ is calculated for each pair of classifications generated from each of the 10 training sets. A scatter plot can then be created with each point representing a pair of classifiers. The x coordinates is the stability value (κ) and the y coordinates is the average accuracy of the pair of classifiers. When $\kappa=0$, the agreement of the two classifications is entirely by chance; when $\kappa=1$, the pair agree on every sample. For a stable algorithm, the plot will produce a compact

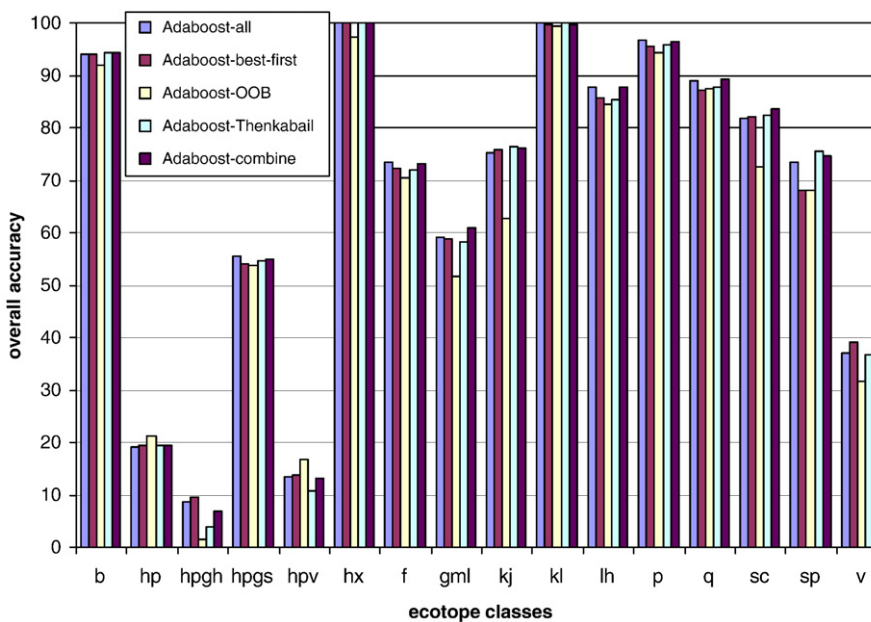


Fig. 9. Class accuracies from Adaboost using different band subsets. Adaboost-all represents the accuracies with all the 126 bands.

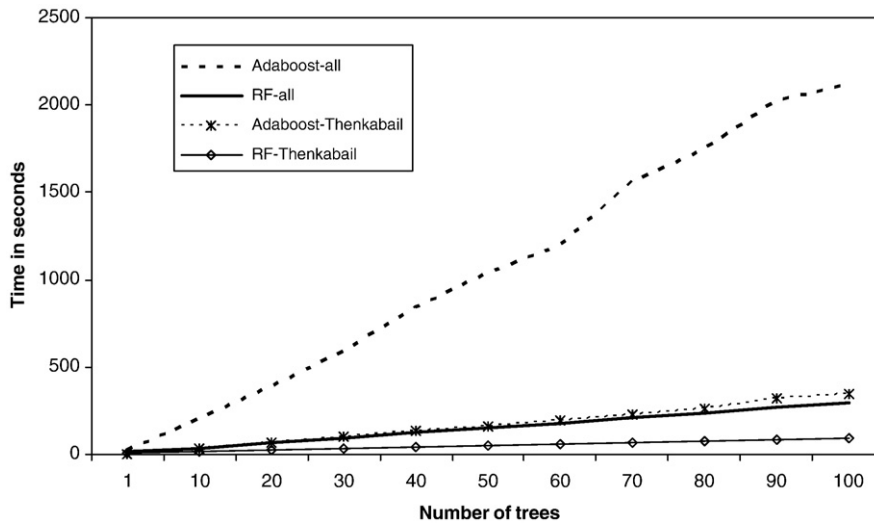


Fig. 10. A comparison of computing burden between Adaboost and Random Forest using 126 and 22 bands.

cloud with similar values close to 1. If on a contrary the values of κ are spread out and farther away from the value 1, the algorithm is considered less stable. Since the formula of the κ statistic is quite well-known within the remote sensing communities, we do not repeat here. Instead the reader is referred to Margineantu and Dietterich (1997) or DeFries and Chan (2000). Only to note that the conventional use of the kappa values in accuracy assessment is to compare the classification result with the groundtruth labels. For the purpose of stability assessment, it compares the results from each pair of the classifiers. From 10 training sets, a total of ${}_{10}C_2=45$ pair values are created. We have plotted the κ -error diagram for each of the five input scenarios for both Random Forest and Adaboost (Fig. 11). In all input scenarios, the κ values (internal agreements) of Random Forest are higher than those of Adaboost. This shows that the former is a more stable algorithm. The coefficient of variation (c.v.) is given for each plot to measure the degree of variation for each spread. In all scenarios except the best-first subset, Adaboost has a higher c.v. value than Random Forest, though the differences are small. In terms of stability, Random Forest has outperformed Adaboost.

In summary, both Adaboost and Random Forest are powerful learning algorithms. Though Adaboost has marginally higher accuracy (<1%) when using all the bands, it is slightly less robust to different

feature subsets as compared to Random Forest. The latter is also more stable and cheaper in terms of running time. All of the subsets worked well. Comparable accuracies are obtained using roughly 17% (21 bands) to 20% (25 bands) of the original feature set. Selections from the two band selection methods, the OOB and the best-first, agree on the importance of the SWIR, especially the ESWIR. The combination of different subset selection results has produced the best accuracies. However, it should be noted that, like bagging, Adaboost is only an algorithm to change the training set. It is not equipped with the features used in Random Forest such as the random feature subset for splitting and the OOB estimates. Both features in theory can also be implemented with Adaboost. Hence, a direct comparison between Adaboost and Random Forest should be handled with care.

5. Applicability of hyperspectral data for ecotope mapping

To date, most ecotope mapping tasks are done manually. This survey-driven approach is labor-intensive and time-consuming. If the task can be partially replaced by automated procedure using remote sensing methods, it will save a lot of resources. A classification of the Biological Valuation Map based on the Adaboost classifier is shown in

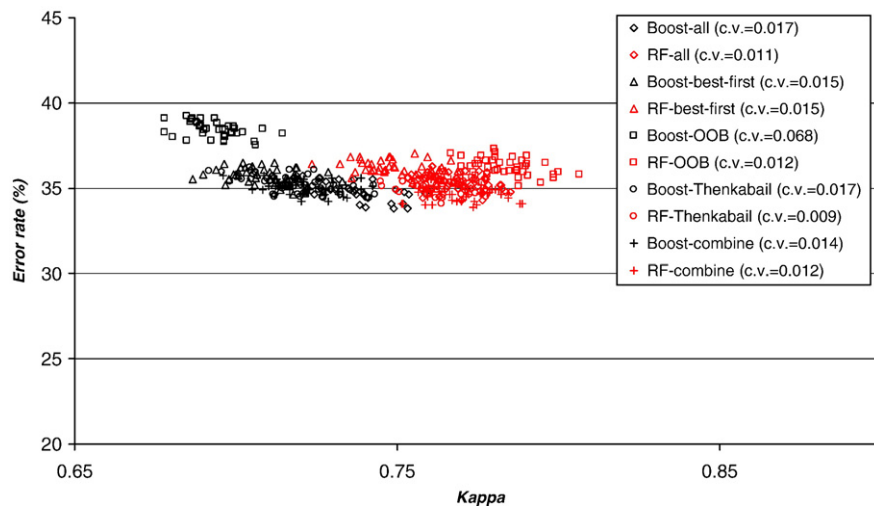


Fig. 11. κ -error diagram for all different input subsets with Adaboost and Random Forest (RF). Coefficient of variation (c.v.) for each band subset is shown in the legend.

Fig. 12. In this section, we discuss the possibility of utilizing hyperspectral data for ecotope mapping by analyzing the confusion matrix. We will focus on the confusion matrix of Adaboost using all the bands (Table 6) which gives the highest overall accuracy at 69.5%.

5.1. Grasslands and arable land

Even though some classes are poorly separated, it is observed that some regrouping or merging of classes will immediately improve the applicability of the method. It is clear that arable land (*b*) and grass monocultures (*hx*) can be distinguished at high accuracies and that both differ clearly from the permanent grasslands group (*hp*–*hpv*–*hpgs*–*hpgh*). Though there are special interests in separating the permanent grasslands group into more specific subgroups, our results suggest that this will be rather difficult even with hyperspectral data. The major confusions within the group can be attributed to the class descriptions that are by definition rather mixed. For example, a lot of the semi-natural grassland (*hpgh*) pixels are assigned to species rich improved grasslands (*hpgs*) which is actually a transitional class between species poor improved grassland (*hp*) and *hpgh*. A land parcel labeled as *hpgs* can in fact consist of different proportion of *hpgs*, *hpgh* and *hp*. Furthermore, *hpv* are grasslands with scattered nature values and within the study area most of these grasslands are in fact a mosaic of *hp* and *hpgs*. Apparently, they can not be separated. However, it would be already a valuable application if we could produce maps representing three classes: arable land (*b*), grass monocultures (*hx*) and permanent grassland (*hp*–*hpv*–*hpgs*–*hpgh*). This is important for environmental policies as there are some legislation rules concerning permanent grasslands (Wils & Paelinckx, 2004). The remote sensing techniques can be used to reduce the amount of field work to search and map these grasslands.

5.2. Tree classes

The results for deciduous forests (*f* and *q* class, respectively dominated by beech and oak) and for poplar and conifer plantations are promising (Table 6). Also the accuracy for plantations of other deciduous trees (*gml*) is not bad. In all cases it is possible that the degree of accuracy is even higher because:

- Within the class of deciduous forests dominated by beech (*f*), most “misclassifications” are assigned to the class of deciduous forests dominated by oak (*q*) and vice versa. In fact the BVM legends *f* and *q* reflect the dominant tree species, so it is possible that beech occurs in an oak forests and vice versa.
- In the case of the class poplar plantation (*lh*), there is a high commission error of *lh* to *q*. In reality, oak trees (*q*) can occur in poplar plantation (*lh*). In such cases, this is not an error.
- For the class of plantations of deciduous tree species other than beech, oak, alder and poplar (*gml*), there is a large amount of misclassification to poplar plantation (*lh*) indicating real misclassifications. However, the misclassification of *gml* to oak trees (*q*) might in fact be oak trees within *gml*.

The reached accuracy to distinguish scrubs (*sc*) from other ecotopes, and even to distinguish them from thorn tickets (*sp*) is promising. The accuracies for orchards *kj* and *kl* are good. Interestingly, almost all “misclassifications” for the tall tree orchard (*kj*) are within the class species rich improved grasslands (*hpgs*). This could be explained by the fact that the legend unit *kj* of the BVM represents a tree coverage ranging from 20% to 100% and within the study area many of these orchards have a *hpgs* vegetation underneath the trees.

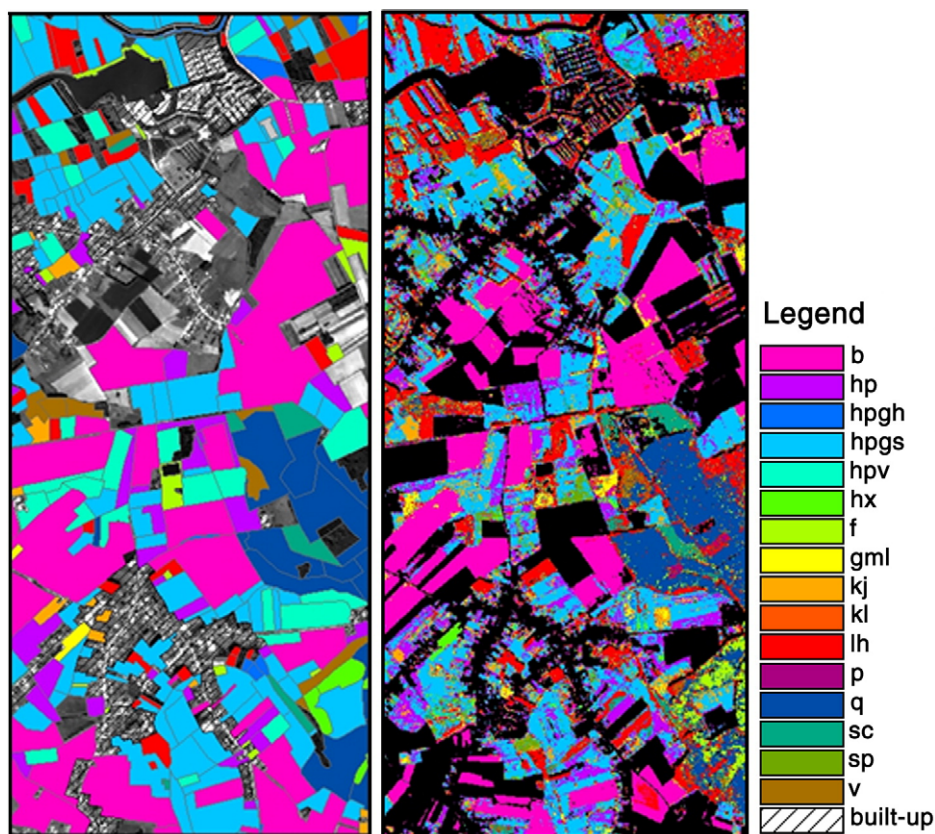


Fig. 12. On the left is the ground truth image of the Biological Valuation Map. On the right is the Biological Valuation Map classification based on airborne hyperspectral data using 99 trials of Adaboost with 21 bands selected by the best-first search method. The black areas represent unclassified land covers that have been masked out.

Table 6
The confusion matrix of 99 Adaboost trials

	<i>b</i>	<i>hp</i>	<i>hpgh</i>	<i>hpgs</i>	<i>hvp</i>	<i>hx</i>	<i>f</i>	<i>gml</i>	<i>kj</i>	<i>kl</i>	<i>lh</i>	<i>p</i>	<i>q</i>	<i>sc</i>	<i>sp</i>	<i>v</i>	%
<i>b</i>	3488	19	12	54	12	29	0	10	4	10	59	0	0	14	0	0	94.0
<i>hp</i>	77	176	34	279	217	15	0	1	50	1	31	0	5	33	9	0	19.0
<i>hpgh</i>	2	25	82	633	39	5	1	4	4	3	26	0	6	100	2	7	8.7
<i>hpgs</i>	141	535	345	3487	776	77	3	142	178	20	326	9	45	126	56	11	55.6
<i>hvp</i>	30	96	123	346	121	11	0	60	54	0	27	0	2	14	25	0	13.3
<i>hx</i>	0	0	0	0	0	285	0	0	0	0	0	0	0	0	0	0	100.0
<i>f</i>	0	1	1	4	0	0	1222	1	0	0	46	0	353	3	1	30	73.5
<i>gml</i>	9	2	7	22	0	0	12	504	0	0	158	1	102	12	0	22	59.2
<i>kj</i>	3	2	5	70	1	0	0	0	278	0	1	0	0	4	5	0	75.3
<i>kl</i>	0	0	0	0	0	0	0	0	0	607	0	0	0	0	0	0	100.0
<i>lh</i>	6	3	6	9	14	0	38	22	1	6	2458	12	128	22	4	73	87.7
<i>p</i>	0	0	0	0	0	0	1	0	0	0	7	308	2	0	0	1	96.6
<i>q</i>	0	0	6	2	2	0	246	24	1	1	146	0	4365	16	2	95	89.0
<i>sc</i>	0	0	2	31	1	1	2	6	0	4	20	0	25	553	13	17	81.9
<i>sp</i>	1	1	3	14	2	4	0	3	4	0	51	0	4	5	264	4	73.3
<i>v</i>	1	2	3	8	2	0	60	24	1	0	231	4	430	28	0	465	36.9

Overall accuracy=69.5%.

The accuracy of the woodland of alluvial soil, fens and bogs (*v*) is poor at only 37% with Adaboost, but it is much higher with Random Forest at 60% (compare Figs. 8 and 9). This higher accuracy with Random Forest, though, implies tradeoffs in other classes. The reason for the low accuracy of this class with Adaboost is related to the algorithmic formation of C5.0 rather than the ineffectiveness of the spectral information. More training pixels from more test sites might improve the accuracy of class *v*. Nevertheless, as most misclassifications of *v* are placed in the classes of poplar (*lh*) and oak (*q*), these can be also due to the fact that the dominant tree species (*Alnus*) are shadowed by taller poplar and oak trees. The mixed nature of these ecotopes makes them a challenge to classify.

6. Conclusions

In this study, we assess the utility of two tree-based ensemble classifiers, Adaboost and Random Forest, for the classification of detailed ecotopes using hyperspectral data. Two feature subset selection algorithms are also examined, together with the optimal wavebands suggested in Thenkabail et al. (2004). Our evaluation criteria include accuracy assessment, training time and also stability. In terms of accuracy performance, Random Forest and Adaboost are almost the same and both have outperformed a neural network classifier. Training time is shorter for Random Forest which is expected because it used only a random subset for each split and applied no pruning. Random Forest generated comparatively more stable results with higher internal agreements as shown by the *k*-diagram plots. When provided with different feature subsets, Random Forest is more robust than Adaboost, though the differences are considered marginal. The highest overall accuracy, nevertheless, is obtained from Adaboost. Due to relatively few requirements for parameter settings, both algorithms are easy to use and their results are highly reproducible. We note that direct comparison between these two classifiers needs to be careful as some features such as the random feature subset and OOB strategy can also be implemented with Adaboost. This study concludes that both classifiers are efficient in handling hyperspectral data and either of them can be a good choice for our application.

Both feature selection routines, the best-first search and the out-of-bag ranking index under Random Forest, are successful in identifying substantially smaller band subsets that attained almost the same accuracy as all the bands. Thenkabail's bands worked equally well. The use of a smaller feature set significantly reduces computing time. This is more beneficial to Adaboost since Random Forest has

already rather light computing burden. One interesting finding is the concentration of the selected bands, for both methods, in the early shortwave infrared region (1.3–1.9 μm), and both have excluded presumably important wavebands at the near infrared and far infrared regions between 0.5 and 1.3 μm . These results signal the importance of SWIR for ecotope classification. However, more studies are needed for more conclusive statements. The feature set which combined three individually derived feature subsets with 53 bands have produced the best accuracy. This supports the idea of using a basket of feature selection methods, instead of just one.

Though the best overall accuracy of the 16 ecotopes is only 69.5%, individual classes have much higher accuracies. Arable land (*b*) and grass monocultures (*hx*) have over 90% accuracies. Five out of the ten tree classes have attained accuracies over 80%. The mixed natural of some grassland classes, however, have made separation impossible. An adaptation of the classification scheme to create a class of "permanent grassland" would enhance the applicability of the remote sensing method.

Acknowledgements

This work was partly supported by the Belgian Science Policy Office under the framework of the STEREO programme – projects ECOMALT (contract SR/03/046) and HABISTAT (contract SR/00/103). We would like to thank four anonymous reviewers for their helpful comments and suggestions.

References

Bajcsy, P., & Groves, P. (2004). Methodology for hyperspectral band selection. *Photogrammetric Engineering and Remote Sensing*, 70, 793–802.

Bauer, E., & Kohavi, R. (1998). An empirical comparison of voting classification algorithms: Bagging, boosting, and variants. *Machine Learning*, 1–38.

Benediktsson, J. A., & Swain, P. H. (1992). Consensus theoretic classification methods. *IEEE Transactions on Systems, Man Cybernet*, 22, 688–704.

Bork, E. W., West, N. E., & Price, K. P. (1999). Calibration of broad- and narrow-band spectral variables for rangeland cover component quantification. *International Journal of Remote Sensing*, 20, 3641–3662.

Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24, 123–140.

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

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and Regression Trees*. Wadsworth.

Briem, G. J., Benediktsson, J. A., & Sveinsson, J. R. (2002). Multiple classifiers applied to multisource remotes sensing data. *IEEE Transactions on Geoscience and Remote Sensing*, 40, 2291–2299.

Chan, J. C. -W., DeFries, R. S., Zhan, X., Huang, C., & Townshend, J. R. G. (2000). Texture features for change detection at 250 m resolution – An application of machine

- learning feature subset selection. *Proceedings IGARSS (International Geoscience and Remote Sensing Symposium) 2000, July 24–28, Hawaii, Vol. VII*. (pp. 3060–3062).
- Chan, J. C. -W., Huang, C., & DeFries, R. S. (2001). Enhanced algorithm performance for land cover classification from remotely sensed data using bagging and boosting. *IEEE Transactions on Geoscience and Remote Sensing*, 39, 693–695.
- Chang, C. -I., Du, Q., Sun, T. -L., & Althouse, M. L. G. (1999). A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 37, 2631–2641.
- Crawford, M. M., Ham, J., Chen, Y., & Ghosh, J. (2003). Random forests of binary hierarchical classifiers for analysis of hyperspectral data. *IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data, 27–28 Oct. 2003* (pp. 337–345).
- De Blust, G., Paelinckx, D., & Kuijken, E. (1994). Up-to-date information on nature quality for environmental management in Flanders. In F. Klijn (Ed.), *Ecosystem Classification for Environmental Management* (pp. 223–249).
- DeFries, R. S., & Chan, J. C. -W. (2000). Multiple criteria for evaluating machine learning ensembles for land cover classification from satellite data. *Remote Sensing of Environment*, 74, 503–515.
- Dieterich, T. G. (2000). An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. *Machine Learning*, 40, 139–158.
- Džeroski, S., & Ženko, B. (2004). Is combining classifiers with stacking better than selecting the best one? *Machine Learning*, 54, 255–273.
- Ellis, E. C., Wang, H., Xiao, H., Peng, K., Liu, X. P., Li, S. C., Ouyang, H., Cheng, X., & Yang, L. Z. (2006). Measuring long-term ecological changes in densely populated landscapes using current and historical high resolution imagery. *Remote Sensing of Environment*, 100, 457–473.
- Footy, G. M., Boyd, D. S., & Sanchez-Hernandez, C. (2007). Mapping a specific class with an ensemble of classifiers. *International Journal of Remote Sensing*, 28, 1733–1746.
- Freund, Y., & Schapire, R. E. (1996). Experiments with a new boosting algorithm. *Machine Learning: Proceedings of the Thirteenth International Conference* (pp. 148–156).
- Friedl, M. A., & Brodley, C. E. (1997). Decision tree classification of land cover from remotely sensed data. *Remote Sensing of Environment*, 61, 399–409.
- Ginsberg, M. L. (1993). *Essentials of Artificial Intelligence*. Morgan Kaufmann.
- Gislason, P. O., Benediktsson, J. A., & Sveinsson, J. R. (2006). Random Forests for land cover classification. *Pattern Recognition Letters*, 27, 294–300.
- Guo, B., Gunn, S. R., & Damper, R. I. (2006). Band selection for hyperspectral image classification using mutual information. *IEEE Transactions on Geoscience and Remote Sensing Letters*, 3, 522–526.
- Ham, J., Chen, Y., Crawford, M. M., & Ghosh, J. (2005). Investigation of the Random Forest framework for classification of hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 43, 492–501.
- Hardisky, M. A., Klemas, V., & Smart, R. M. (1983). The influences of soil salinity, growth form, and leaf moisture on the spectral reflectance of *Spartina alterniflora* canopies. *Photogrammetric Engineering and Remote Sensing*, 49, 77–83.
- Jia, X., & Richards, J. A. (1994). Efficient maximum likelihood classification for imaging spectrometer data sets. *IEEE Transactions on Geoscience and Remote Sensing*, 32, 274–281.
- Kaewpajit, J., Moigne, L., & El-Ghazawi, T. (2003). Automatic reduction of hyperspectral imagery using wavelet spectral analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 41, 863–871.
- Kohavi, R., & John, G. H. (1997). Wrappers for feature subset selection. *Artificial Intelligence*, 97, 273–324.
- Kohavi, R., & Sommerfield, D. (1995). Feature subset selection using the wrapper model: Overfitting and dynamic search space topology. *Proceedings of the First International Conference on Knowledge Discovery and Data Mining* (pp. 192–197).
- Kohavi, R., Sommerfield, D., & Dougherty, J. (1996). *Data mining using MLC++: A machine learning library in C++*. Tools with Artificial Intelligence. IEEE Computer Society Press 234–245 pp.
- Lawrence, R. L., Wood, S. D., & Sheley, R. L. (2006). Mapping invasive plants using hyperspectral imagery and Breiman and Cutler classifications (RandomForest). *Remote Sensing of Environment*, 100, 356–362.
- Lee, C., & Landgrebe, D. A. (1997). Decision boundary feature extraction for neural networks. *IEEE Transactions on Neural Networks*, 8, 75–83.
- Margineantu, D. D., & Dieterich, T. G. (1997). Pruning adaptive boosting. *Proceedings of the 14th International Conference on Machine Learning* (pp. 211–218).
- Melgani, F., & Bruzzone, L. (2004). Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on Geoscience and Remote Sensing*, 42, 1778–1790.
- Mingers, J. (1989). An empirical comparison of selection measures for decision tree induction. *Machine Learning*, 3, 319–342.
- Mingers, J. (1989). An empirical comparison of pruning methods for decision tree induction. *Machine Learning*, 4, 227–243.
- Pal, M., & Mather, P. M. (2003). An assessment of the effectiveness of decision tree methods for land cover classification. *Remote Sensing of Environment*, 86, 554–565.
- Quinlan, J. R. (1987). Simplifying decision trees. *International Journal of Man-Machine Studies*, 27, 221–234.
- Quinlan, J. R. (1993). *C4.5: Programs for Machine Learning*. San Mateo, CA: Morgan Kaufmann Publishers, Inc.
- Thenkabail, P. S., Enclona, E. A., Ashton, M. S., & Van Der Meer, B. (2004). Accuracy assessments of hyperspectral waveband performance for vegetation analysis applications. *Remote Sensing of Environment*, 91, 354–376.
- Thenkabail, P. S., Ward, A. D., & Lyon, J. G. (1994). Landsat-5 Thematic Mapper models of soybean and corn crop characteristics. *International Journal of Remote Sensing*, 15, 49–61.
- Ustin, S. L., Roberts, D. A., Gamon, J. A., Asner, G. P., & Green, R. O. (2004). Using imaging spectroscopy to study ecosystem processes and properties. *BioScience*, 54, 523–534.
- Wils, C., & Paelinckx, D. (2004). Permanent grasland binnen VEN en binnen habitat- en vogelrichtlijngebied. *Testbestand. Verslag en digitaal bestand Instituut voor Natuurbehoud IN. 02004.14 Brussel* (pp. 11). (In Dutch). (Title translated in English: Permanent grassland in the Flemish ecological network and in the habitat and bird directive areas. Report and draft GIS databank of the Institute of Nature Conservation.).
- Wils, C., Paelinckx, D., Adams, Y., Berten, B., Bosch, H., De Knijf, G., et al. (2004). Biologische Waarderingskaart en natuurgerichte bodembedekkingkaart van het Vlaamse Gewest. *Rapport en digitaal bestand Instituut voor Natuurbehoud IN. R2004.08 Brussel* (pp. 39). (In Dutch). (Title translated in English: Biological Valuation Map and land cover map of Flanders. Report and draft GIS databank of the Institute of Nature Conservation).
- Yilmaz, M. T., Hunt, E. R., Jr., Goins, L. D., Ustin, S. L., Vanderbilt, V. C., & Jackson, T. J. (2008). Vegetation water content during SMEX04 from ground data and Landsat 5 Thematic Mapper imagery. *Remote Sensing of Environment*, 112, 350–362.