Experimental comparison of support vector machines with random forests for hyperspectral image land cover classification

B T Abe[1,*], O O Olugbara[2] and T Marwala[3]

[1]*School of Electrical and Information Engineering, University of Witwatersrand, Johannesburg, South Africa.*
[2]*Department of Information Technology, Durban University of Technology, Durban, South Africa.*
[3]*Department of Engineering and the Built Environment, University of Johannesburg, Johannesburg, South Africa.*
*Corresponding author. e-mail: abe_tolulope@yahoo.com

The performances of regular support vector machines and random forests are experimentally compared for hyperspectral imaging land cover classification. Special characteristics of hyperspectral imaging dataset present diverse processing problems to be resolved under robust mathematical formalisms such as image classification. As a result, pixel purity index algorithm is used to obtain endmember spectral responses from Indiana pine hyperspectral image dataset. The generalized reduced gradient optimization algorithm is thereafter executed on the research data to estimate fractional abundances in the hyperspectral image and thereby obtain the numeric values for land cover classification. The Waikato environment for knowledge analysis (WEKA) data mining framework is selected as a tool to carry out the classification process by using support vector machines and random forests classifiers. Results show that performance of support vector machines is comparable to that of random forests. This study makes a positive contribution to the problem of land cover classification by exploring generalized reduced gradient method, support vector machines, and random forests to improve producer accuracy and overall classification accuracy. The performance comparison of these classifiers is valuable for a decision maker to consider tradeoffs in method accuracy *versus* method complexity.

1. Introduction

The mathematical analysis of hyperspectral images has been the subject of interest for many researchers (Plaza *et al.* 2003, 2008; Sanchez *et al.* 2010; Iordache *et al.* 2011; Zhang *et al.* 2011). In general, image processing is of great value because it enables abundant data to be translated into useful information (Su *et al.* 2008; Zhang *et al.* 2011). The ability to retrieve information from datasets has motivated researchers to explore methods of data mining for identifying valid, novel, potentially useful, and ultimately understandable patterns in data (Xie *et al.* 2008; Liu *et al.* 2009). Earth observation has increasingly become a prime source of data in geosciences and related disciplines, thereby permitting research into distant past, present and future. One area of research interest has always been how to relate earth observation output such as aerial photographs and satellite images to known features (Xie *et al.* 2008). The preprocessing of satellite images prior to features extraction is essential to remove noise and increase the ability to interpret image data more accurately. The final product of image preprocessing is that all images should appear

as if they were acquired from the same sensor (Li 2011).

The state of the earth surface evaluation and monitoring is a major requirement for global change research. The conservation, preservation, and sustainable yield of natural resources are increasingly dependent upon remotely sensed data for inventory and monitoring of changes. A collection of digital data such as high resolution satellite images is currently available for this purpose. Generally speaking, in remote sensing, varieties of earth objects are present in the direct view of sensors because of the complexity of the target objects and the limited spatial resolution of remote sensors. The information contained in a particular pixel of a remote sensing image is a mixture of information on various ground objects, resulting in mixed pixels (Plaza *et al.* 2003, 2008; Sanchez *et al.* 2010; Iordache *et al.* 2011; Zhang *et al.* 2011). The presence of mixed pixels has a significant impact on some practical applications of remote sensing images such as information extraction, image classification, and object detection. It is therefore, an important task in remote sensing to discover objects and corresponding quantity of information present in the mixed pixel. This has led to the invention of hyperspectral remote sensing techniques to proffer solutions to the mixed pixel problem in remotely sensed imagery. Hyperspectral images have been used in many real applications because of their rich sources of information. Examples of the useful applications of hyperspectral imaging include mineral exploration, urban processes, agriculture, risk prevention, land cover mapping, surveillance system, resource management, tracking wildfires, detecting biological threats, and chemical contamination (Hall *et al.* 1991; Ellis 2001; Lacar *et al.* 2001; Zhang *et al.* 2011). These images provide abundant spectral information to identify and differentiate between spectrally similar, but unique materials. They provide potential, detailed, and accurate information extraction as compared to other remotely sensed data (Karaska *et al.* 2004). In addition, hyperspectral images provide a high-resolution reflectance spectrum for each pixel in the image (Boardman *et al.* 1995). As a result, large scale land cover maps constructed from remotely sensed data have become important information sources (Boardman *et al.* 1995).

This paper addresses the problem of land cover classification by using (i) a linear spectral mixture analysis technique, which is commonly accepted for mixed-pixel classification and (ii) Generalized Reduced Gradient (GRG) algorithm to estimate fractional abundance in the research data and thereby obtain the estimated numeric values for image classification. Land cover refers to the physical surface of the earth, including various combinations of vegetation types, soils, exposed rocks, water bodies, and anthropogenic elements, such as agriculture and built environments (Udelhoven *et al.* 2009; Sanchez *et al.* 2010). The purpose of this study is three-fold: The first, is to identify a collection of pure constituent spectral, called endmember spectral responses (Heinz and Chang 2001; Plaza *et al.* 2004; Martinez *et al.* 2006; Plaza *et al.* 2008; Dobigeon *et al.* 2009; Sanchez *et al.* 2010; Iordache *et al.* 2011; Zhang *et al.* 2011). In literature, endmember is also known as class label, class type, component (Gong and Zhang 1999). The measured spectrum of each mixed pixel is expressed as a linear combination of endmember spectral responses weighed by fractional abundances, which indicate the proportion of each endmember spectral response present in a pixel (Heinz and Chang 2001; Plaza *et al.* 2004; Sanchez *et al.* 2010). The second is to explore the GRG optimization algorithm (Abadie and Carpentier 1969; Lasdon *et al.* 1974) to estimate the fractional abundances of each pixel in some selected region of interest and thereby obtain the numeric values for land cover classification. The third is to experimentally compare performances of generalized Support Vector Machines (SVM) and Random Forests (RF) classifiers to examine the suitability of the GRG algorithm for solving land cover classification problems. The GRG method was initially employed for estimating fractional abundance (Abe *et al.* 2012). Machine learning techniques play an important role in map construction development management (Lacar *et al.* 2001; Bruzzone and Cossu 2002; Wolter *et al.* 2005; Chen *et al.* 2007).

2. Problem description

The task of land cover classification can be generally formulated as a linear spectral unmixing problem. The linear spectral unmixing is a subpixel classification process that decomposes mixed pixels and determines the combination of fractional abundances. The basic idea behind linear spectral mixture analysis is that every image pixel is a mixture of different endmember spectral responses. The spectrum recorded by the sensor is a linear combination of endmember spectral responses (Kärdi 2007). Linear unmixing model can be mathematically expressed as (Sanchez *et al.* 2010; Zhang *et al.* 2011):

$$x(i, j) = \sum_{k=1}^{P} a_k(i, j) e_k + n(i, j). \qquad (1)$$

The component $x(i, j)$ is spectral response vector, $a_k(i, j)$ is a scalar value representing fractional abundance of endmember vector e_k at pixel, is a vector that denotes the spectral band error and P is the total number of endmembers. Equation (1) operates under two physical constraints on fractional abundances to account for the full composition of a mixed pixel. These constraints are non-negativity of all fractional abundance values and fractional abundance values must sum to unity (Heinz and Chang 2001; Sanchez *et al.* 2010). The purpose of the land cover classification is to evaluate the performance of the RF and SVM classifiers per class basis. The GRG algorithm is used to obtain the estimated numeric values of the endmembers' fractional abundance.

3. Design methodology

The study design methodology consists of the sequence of steps that the input hyperspectral image undergoes for its land cover to be classified into one of the desired multiple classes. The input hyperspectral image has to be taken through four essential steps of image dimension reduction, endmember spectral response determination, fractional abundance estimation, and land cover classification. The experimental dataset used is first introduced to begin the discussion on the essential steps of the design methodology.

3.1 *Experimentation dataset*

Figure 1 shows the Indiana pine test site in northwestern Indiana collected using the AVIRIS sensor in 1992. The dataset, which serves as a benchmark has been extensively researched in recent times (Landgrebe 1998; Tarabalka *et al.* 2009; Ul Haq *et al.* 2010). The dataset designated 92AV3C is
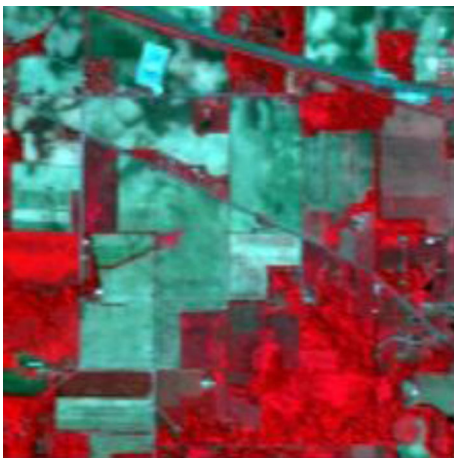
descriptive of the problem of hyperspectral image analysis to determine land cover use. Each band in the dataset has equal dimensions of 145 × 145 (21,025 pixels) of mixed agriculture/forestry lands. The dataset contains 224 spectral reflectance bands within a wavelength of 0.4–2.5 μm and nominal spectral of 10 nm, a 16-bit radiometric resolution, and a 20 m spatial resolution. The total number of the spectral bands was reduced to 220 because four of the spectral bands contain no signal. The dataset also contains highways and railways, which were ignored because they are not properly discernible (Chakrabarty *et al.* 2012). The processed image is free of path radiance including the light scattered by the interaction between surface and atmosphere (Landgrebe and Biehl 2001). Noise and water absorption bands (1–3, 104–112, 148–165 and 217–224) were also removed for the purpose of this study because they can have negative effect on the classification results (Nascimento and Bioucas-Dias 2005). The remaining 186 bands (4–103, 113–147, and 166–216) were used for the experiment.

The extensive ground-truth data, which is available online (Landgrebe 1998) contain 16 classes of vegetation and background as shown in figure 2. As pointed out by Landgrebe (1998) and Congalton (1991), no reference dataset may be completely accurate, but it is important that the reference data have high accuracy or else it is not a fair assessment. Therefore, it is important that reference data be carefully considered in any accuracy determination. In this study, 10 land endmembers were used for the classifiers' investigation while the remaining were discarded because of insufficient number of pixels available (Melgani and Bruzzone 2004; Camps-Valls and Bruzzone 2005; Nascimento and Bioucas-Dias 2005). The endmembers used are corn-notill, corn-minimum till, grass/pasture, grass/tree, hay-windowed, soybeans-notill, soybean-clean, soybeans-minimum till, woods, and the background.
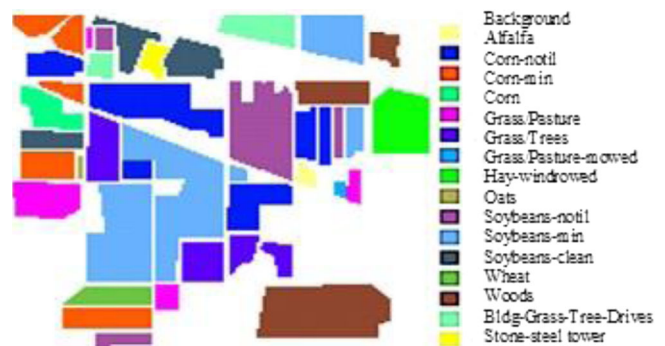

Figure 1. Indiana pine hyperspectral image.

Figure 2. Indiana pine ground truth and land cover classes.

3.2 *Experiment 1*

3.2.1 *Image dimension reduction*

The method of dimension reduction is aimed at reducing the number of spectral bands in a hyperspectral image and to reduce the time taken to process the image. This is usually accomplished by mapping the data from a higher dimensional space onto a lower dimensional space at the same time preserving the main features of the original data. This method does not generate an image that is different from the original image. Instead, it was designed to reduce error by finding minimum representation of the original image that adequately keeps the original information for successful unmixing in the lower dimension (Keshava and Mustard 2002). Among the various algorithms normally used for dimension reduction are Principal Component Analysis (PCA) and Maximum Noise Fraction (MNF) transform. This study applies the MNF algorithm because it is more computationally effective than PCA algorithm (Chaudhry *et al.* 2006).

3.2.2 *Endmember spectral response determination*

The method of endmember spectral response determination involves identifying the most discriminative measurements from a set of potentially useful measurements. An endmember spectral response is a spectrally pure pixel that portrays various mixed pixels in a hyperspectral image (Plaza *et al.* 2004). The endmember spectral response determination has been widely used in hyperspectral image analysis because it significantly improves spatial and spectral resolutions that are provided by hyperspectral imaging sensors (Keshava and Mustard 2002; Chaudhry *et al.* 2006). The determination of image endmember spectral response is an important task in hyperspectral image data exploitation such as image classification (Heinz and Chang 2001). The endmember spectral responses e_k, $k = 1, 2, \ldots, P$ in equation (1) can be determined by using various algorithms such as Manual Endmember Selection Tool (MEST) (Bateson and Curtiss 1996), Pixel Purity Index (PPI) (Gonzalez *et al.* 2010) and Automated Morphological Endmember Extraction (AMEE) (Sanchez *et al.* 2010), N-FINDER (Winter 1999), Optical Real-time Adaptive Spectral Identification System (ORASIS) (Bowles *et al.* 1995) and Simulated Annealing Algorithm (SAA) (Bateson *et al.* 2000).

The PPI algorithm generates from the image a large number of n-dimensional scattered plots on random vectors that are called skewer (Chaudhry *et al.* 2006; Plaza *et al.* 2008). The N-FINDR fully automated a method that locates a set of

pixels with the largest possible volume by inflating a simplex within the image data (Winter 1999; Plaza *et al.* 2004). The AMEE uses the morphological method where spectral and spatial information are equally required to derive endmember spectral responses (Heinz and Chang 2001; Keshava and Mustard 2002; Plaza *et al.* 2004; Sanchez *et al.* 2010).

This study applies the PPI algorithm (Boardman 1993; Gonzalez *et al.* 2010), which is available in the environment for visualizing images (ENVI) to determine endmember spectral responses from hyperspectral dataset. The choice of the PPI algorithm is motivated by its publicity in ITTVIS (http://www.ittvis.com/) ENVI software that was originally developed by Analytical Imaging and Geophysics (AIG) (Gonzalez *et al.* 2010). The PPI algorithm searches through a set of the vertices of a convex geometry in a certain dataset to present pure signatures in the data (Keshava and Mustard 2002; Sanchez *et al.* 2010). In order to generate the endmember spectral responses from the study site, noise whitening and dimensionality reduction was first performed using the MNF transform (Karaska *et al.* 2004; Plaza *et al.* 2004) before the pixel purity score is calculated for each point in the image cube. This was accomplished by randomly generating lines in the N-dimensional space (an ENVI's visualizer that provides an interactive tool for finding endmember spectral responses) containing the MNF transformed data. In an N-dimensional component space, it is assumed that endmember (P) spectral responses will occur at the vertices of the hyper-solid or geometric shape bounding the pixel values in that space. The spectral points were projected on lines and the points at extremes of each line were selected by drawing a polygon around few of the extreme data points to create the endmember (P) spectral responses. The highest-valued of these pixels were input to the ENVI visualizer for the clustering process that developed the individual endmember $(1, 2, 3, \ldots, P)$ spectral response. The result is not a single map representation as in thematic image classification but a series of images, each having the size of the original image (Settle and Drake 1993; Adams *et al.* 1995; Van der Meer and Jia 2012).

3.3 *Experiment 2*

3.3.1 *Fractional abundance estimation*

The GRG optimization algorithm was executed to estimate per pixel fractional abundances by using spectral responses results obtained in experiment 1. The previous efforts on linear spectral unmixing problem (Sanchez *et al.* 2010; Iordache *et al.* 2011; Zhang *et al.* 2011) have investigated the least

square (LSU) method (Heinz and Chang 2001) to estimate a set of fractional abundances as follows.

$$a(i,j) = (e^T e)^{-1} e^T x(i,j). \qquad (2)$$

In equation (2), e^T is the transpose of the matrix e of endmember spectral responses and e^{-1} represents the inverse matrix of e matrix.

In general, the fractional abundances obtained from equation (2) can only satisfy the sum to unity constraint, but the non-negativity of fractional abundances cannot always be guaranteed. The solutions obtained by the LSU method are therefore generally not optimal in terms of material quantification (Heinz and Chang 2001; Du *et al.* 2008; Sanchez *et al.* 2010). The linear spectral unmixing problem has to be formulated as an optimization problem that minimizes the spectral band error. The reason for the minimization is to obtain optimal fractional abundances that simultaneously satisfy the two changing constraints Δ_1 and Δ_2 with respect to the spectral coordinate (i,j). The following fully constrained linear spectral unmixing optimization problem has to be solved by finding a set of fractional abundances $a(i,j)$ that minimize the spectral band error in equation (1).

Minimize

$$J(a) = \Delta \left\{ (x(i,j) - a(i,j)e)^T (x(i,j) - a(i,j)e) \right\} \qquad (3)$$

subject to

$$\Delta_1 = \left\{ a(i,j) | \sum_{k=1}^{P} a_k(i,j) - 1 = 0 \right\}, \qquad (4)$$

$$\Delta_2 = \{ a(i,j) | 0 \le a_k(i,j) \le 1 \}. \qquad (5)$$

In equation (3), $J(a)$ is the objective function to be minimized, Δ_1 is the abundance sum-to-one constraint and Δ_2 is the abundance non-negativity constraint with respect to the spatial coordinate (i,j). The PPI algorithm is first applied to extract endmember spectral responses from the hyperspectral image before equation (3) can be solved for optimal fractional abundances using the GRG algorithm (Abadie and Carpentier 1969; Lasdon *et al.* 1974; Su and Lii 1995). The PPI method efficiently handles hyperspectral images as it provides a convenient and physically motivated decomposition of an image in terms of relatively few components (Theiler *et al.* 2000). Once a set of endmember spectral responses $e = \{e_k\}_{k=1}^{p}$ (where e is a square matrix having 10×10 dimensionality) is determined, the corresponding fractional abundances $a(i,j) = \{a_k(i,j)\}_{k=1}^{p}$ in a specific pixel vector $x(i,j)$ of the hyperspectral image can be estimated by using the GRG algorithm.

The GRG algorithm uses the endmember set produced by the PPI to produce a set of endmember numeric values as follows:

- The first step is to calculate the compute matrix $(e^T e)^{-1} e^T$, where $e = \{e_k\}_{k=1}^{p}$ is formed by the P endmember extracted from the PPI. Using excel solver for calculation, the compute matrix is multiplied by all the pixel vectors $x(i,j)$.
- The compute matrix calculated in step 1 is multiplied by each pixel $x(i,j)$ from the region of interest, thus obtaining a set of vectors $a(i,j)$, each containing the fractional abundances of the P endmembers in each pixel.

The new values obtained were used to train and test SVM and Random Forest classifiers for the land cover classification procedure. Using the mixture model, the spectral vectors were converted to abundance vectors and were classified by the classifiers. Implementation of the classifiers on the numeric values was to assess in a benchmark exercise, the computational efficiency and the speed up between the renowned classifiers in remote sensing classification.

3.3.2 *Land cover classification*

The purpose of thematic image classification is to represent the land cover in terms of a number of fixed classes where each image pixel represents a unique endmember which in turn is used to produce a single map representation. Table 1 shows the selected region of interest (ROI) from the Indiana pine dataset used for the classification procedure. The newly generated numeric values obtained from the GRG algorithm in the two steps mentioned above were used for the classification procedure. The WEKA (Garner 1995) data mining software is selected to build SVM and RF classifiers. The two classifiers are experimentally compared for land cover classification.

Table 1. *Number of pixels extracted from the ROI.*

Endmembers	Number of pixels
Corn-notill (cnt)	359
Corn-minimum till (cmn)	305
Grass/pasture (gp)	264
Grass/trees (gt)	339
Hay-windowed (hw)	279
Soybeans-notill (snt)	350
Soybean-clean (scl)	203
Soybeans-minimum till (smn)	425
Woods	400
Background (bg)	300
Total number of pixels	3224

The SVM developed by Cortes and Vapnik (1995) is a non-parametric supervised learning classifier that takes root from statistical learning theory (Winter 1999; Gonzalez *et al.* 2010). The training procedure of SVM is aimed at discovering a hyperplane that separates a dataset into a distinct predefined number of classes in a way that it is consistent with the training examples. The classifier works in the sense that it generalizes a training dataset that is not necessarily linearly separable (Lennon *et al.* 2002). It has been established to have a high generalized ability to solve classification problems and for this purpose, it is extensively used in supervised classification of hyperspectral images (Winter 1999; Martinez *et al.* 2006). The other application domains of SVM include object detection, and text categorization and it has outperformed the traditional neural network technique in terms of generalization capacity. SVM is relatively not sensitive to training sample size for it can be applied to either small or large training datasets (Lennon *et al.* 2002; Mountrakis *et al.* 2011). The algorithm was originally designed for binary classification, but the usual method of extending it to multi-class classification is to decompose a multi-class problem into a series of two-class problems. Though the one-against-all method experiences output scaling problems and suffers from training set inconsistency, it is currently the most common SVM approach due to its simplicity and efficiency (Bishop 2006; Shao and Lunetta 2012). This study applies one-against-all, which is one of the most widely used methods of solving multi-class classification problems (Lennon *et al.* 2002; Liu and Zhen 2005; Mountrakis *et al.* 2011).

The random forests classifier is an ensemble that builds several decision trees randomly as proposed (Breiman 2001; Pal 2005; Rodriguez-Galiano *et al.* 2011) for the classification of multisource remote sensing, geographic data, and hyperspectral imaging. There are various ensemble classification methods proposed in recent times that have been proven to considerably improve classification accuracy. The most famous and widely used ensemble methods are boosting and bagging (Breiman 1996, 2001). The boosting method is based on sample re-weighting technique, but a bagging method uses bootstrapping. RF classifier uses bagging or bootstrap aggregating to yield an ensemble of classification and regression trees. The method works by searching only a random subset of the features for a split at each node to minimize the correlation between the classifiers in the ensemble. The method selects a set of features randomly and creates an algorithm with a bootstrapped sample of the training dataset (Breiman 2001; Pal 2005; Rodriguez-Galiano *et al.* 2011).

This method provides a potential benefit that it is computationally lighter than methods based on boosting or bagging and often produces excellent results (Breiman 2001; Pal 2005; Rodriguez-Galiano *et al.* 2011). The RF algorithm uses Gini index (Breiman 2001) as a feature selection measure. In this case, the impurity of a feature is measured against the classes. For our experiment, 10 trees were constructed and the out of bag error was 0.5471 while considering 186 random features and 10 output targets.

4. Results and discussion

The discussion on the results of our research on endmember spectral response determination is presented. The session also discusses the performance of the SVM and RF classifiers investigated.

4.1 *Results of endmember spectral response determination*

The first experiment performed was aimed at obtaining endmember spectral responses from the image dataset by using the ENVI software application. The MNF transformation of the input hyperspectral image was performed for dimension reduction. In the next stage, a set of (P) endmember spectral responses is selected by applying the PPI algorithm. Figure 3 shows the result obtained, wherein the extreme pixels correspond to endmember spectral responses in each projection that is recorded. The total number of times that each pixel is marked as extreme is noted. A threshold value of 65% is used to define how many pixels are marked as extreme at the ends of the projected vector.
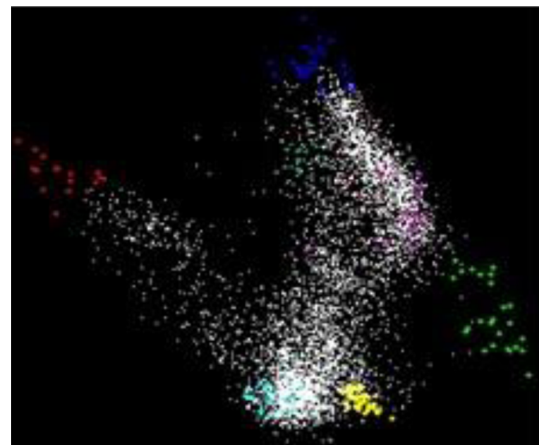

Figure 3. Purest pixels occur at edges of the projected vector.

The estimated number of endmember spectral responses and the corresponding spectral signatures were obtained by using ENVI visualizer. At the completion of specified iterations, PPI 10 images (P) were created in which the value of each pixel corresponds to the number of times that a pixel was recorded as extreme. The bright pixels in the PPI image are generally the image endmember spectral responses to characterize the vegetation structure. Figure 4 shows the generated images and the Root Mean Square (RMS) error of the image.

4.2 *Results of land cover classification*

The SVM and RF classifiers were evaluated by using error confusion matrix method, which is a representation of the entire thematic classification result. According to Congalton (1991), the error confusion matrix can be used to compute the overall accuracy and the individual endmember accuracy. The error confusion matrix is a widely accepted method to report error of raster data and to assess the classification accuracy of a classifier. The matrix expresses the number of sample units allocated to each endmember as compared to the reference data. The diagonal of the matrix designates agreement between the reference data and the interpreted endmember (Nascimento and Bioucas-Dias 2005). Table 2 shows the result of the error confusion matrix for the performance of SVM classifier. This result shows that scl, snt, cnt have 100% classification accuracy because none of their pixel members is misclassified while others have some of their pixels misclassified and therefore are not 100% accurate.

Table 3 shows the result of the error confusion matrix for the performance of RF classifier. Similarly, it can be observed that using the RF classifier, all the endmembers have a few of the endmembers misclassified.

Comparing the performance of each classifier on individual endmember, SVM produces a higher level of classification accuracy per endmember as compared to RF. The entire accuracy assessment procedure is that the error confusion matrix must be a representative of the entire area mapped from the remotely sensed data (Congalton *et al.* 1983; Story and Congalton 1986). The overall accuracy for correctly classified instances, incorrectly classified instances, unclassified instances and Kappa statistics are identified from the error confusion
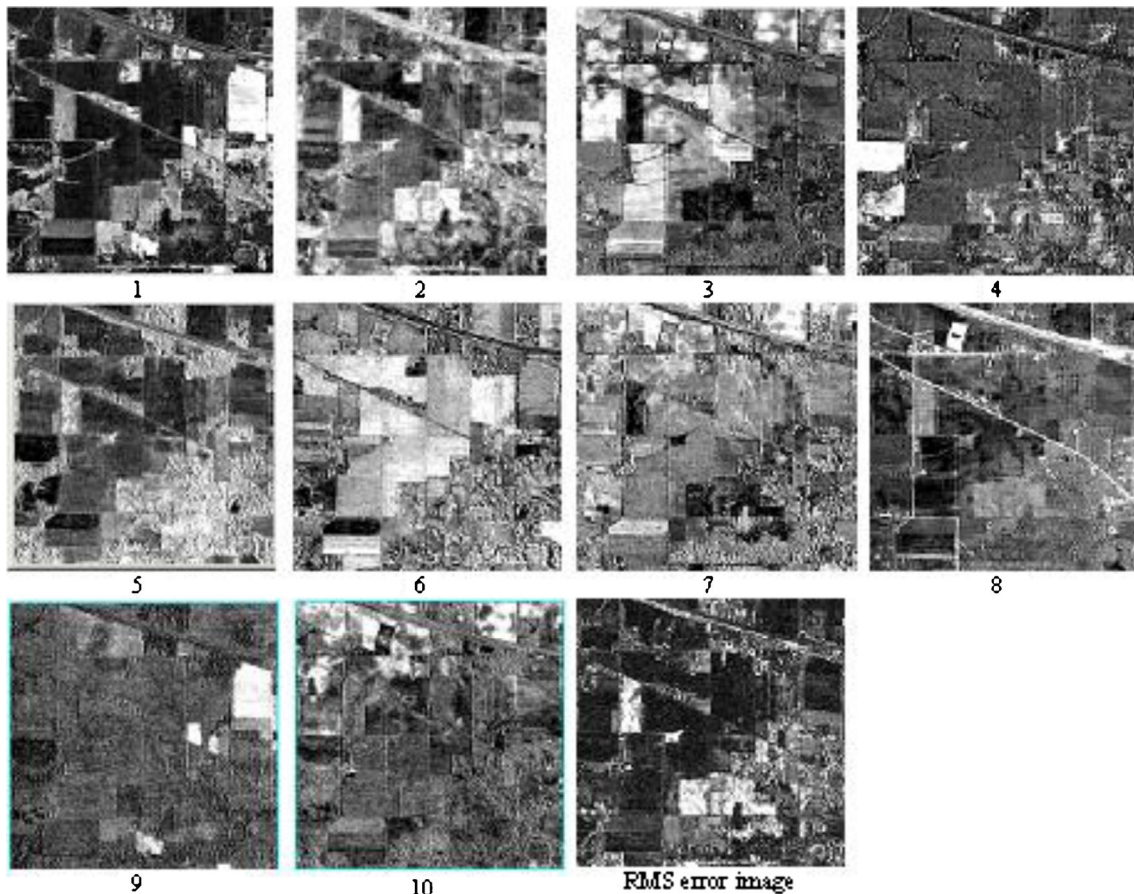

Figure 4. Generated images and RMS error from PPI method.

Table 2. *Support vector machines' error confusion matrix.*

a	b	c	d	e	f	g	h	i	j	<− Classified as
246	0	0	0	15	0	0	0	1	2	| a = gp
0	203	0	0	0	0	0	0	0	0	| b = scl
0	0	350	0	0	0	0	0	0	0	| c = snt
0	1	0	424	0	0	0	0	0	0	| d = smn
0	0	0	0	333	6	0	0	0	0	| e = gt
0	0	0	0	0	359	0	0	0	0	| f = cnt
0	0	0	0	1	0	278	0	0	0	| g = hw
0	0	0	0	0	4	0	301	0	0	| h = cmn
8	0	0	0	0	0	0	0	391	1	| i = woods
5	0	0	0	7	2	0	0	0	286	| j = bg

Table 3. *Random forests error confusion matrix.*

a	b	c	d	e	f	g	h	i	j	<− Classified as
247	0	0	0	15	0	1	0	1	0	| a = gp
0	203	0	0	0	0	0	0	0	0	| b = scl
0	0	347	0	0	2	0	1	0	0	| c = snt
0	0	3	420	0	1	0	1	0	0	| d = smn
1	0	0	0	332	6	0	0	0	0	| e = gt
0	1	0	2	0	356	0	0	0	0	| f = cnt
0	0	0	0	1	0	278	0	0	0	| g = hw
0	0	1	0	0	4	0	300	0	0	| h = cmn
2	0	0	0	0	0	0	0	398	0	| i = woods
0	0	0	0	9	2	1	0	3	285	| j = bg

matrices (Congalton 1988, 1991). The overall accuracy is calculated by dividing the total correct (which is the sum of the major diagonal) by the total number of pixels in the error confusion matrix (Story and Congalton 1986; Congalton 1991; Gómez and Montero 2011; Zhou *et al.* 2011). If all the non-major diagonal elements of the error confusion matrix are zero, then no area in the map has been misclassified and the classifier accuracy is 100%. Otherwise, there are certain percentages of misclassified instances (Congalton 1988; Nascimento and Bioucas-Dias 2005). In our experiment, 53 instances = 1.66% and 58 instances = 1.8% were misclassified by SVM and RF classifiers, respectively.

The Kappa statistics (Story and Congalton 1986; Congalton 1991) is a discrete multivariate technique for determining whether the remotely sensed classification is better than a random classification and whether two or more error confusion matrices are significantly different from each other. The outcome of performing Kappa analysis is a KHAT statistics, which can be viewed as a measure of accuracy or agreement (Congalton *et al.* 1983; Story and Congalton 1986; Landgrebe 1998; Nascimento and Bioucas-Dias 2005; Gómez and Montero 2011; Zhou *et al.* 2011). The Kappa coefficient of agreement is a measure of how well the

accuracy of the classifier compares to the reference or ground truth data (Congalton and Mead 1983). It ranges from –1 to 1, with negative values meaning agreement worse than expected, low negative values (0 to –0.10) generally implying no agreement between the classified land cover and ground truth, while 1 indicates complete agreement. Table 4 shows the result of error, Kappa statistics, and overall accuracy of the classification. According to this result, no unclassified instance during the SVM and RF classification procedures and the overall classification accuracies of the classifiers are seen to be comparable. The difference between the performance results is 0.14%, in favour of SVM, but this difference is insignificant. However, the RF classifier has more endmembers that are incorrectly classified as compared to SVM classifier.

Mean absolute error (MAE) is described as the sum of absolute errors divided by the number of predictions. This is measured set of predicted value to real value. Root mean square error (RMSE) is defined as square root of the sum of squares error divided by the number of predictions. This measures the difference between the values predicted by the classifier and the actual values observed. The smaller the values of RMSE and mean absolute error (MAE), the better the performance of the classifier. Relative absolute error (RAE) is the

Table 4. *Spectral unmixing classification accuracy.*

Classifier	CCI	ICI	UI	KS	MAE	RMSE	RAE (%)	RRSE (%)	Accuracy (%)
SVM	3171	53	0	0.9817	0.0033	0.0573	1.8343	19.1535	98.34
RF	3166	58	0	0.9799	0.0286	0.0875	15.954	29.2364	98.20

Note: CCI: correctly classified instances, ICI: incorrectly classified instances, UI: unclassified instances, KS: Kappa statistic, MAE: mean absolute error, RMSE: root mean squared error, RAE: relative absolute error, RRSE: root relative squared error, A: accuracy.
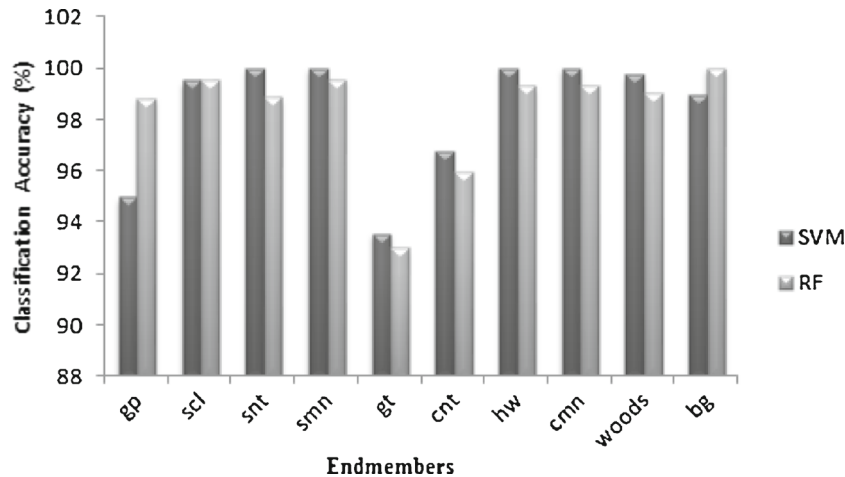

Figure 5. Producer accuracies of RF and SVM classifiers on land cover dataset.

total absolute error, using the same type of normalization (Witten and Frank 2005). Root relative squared error (RRSE) uses the total squared error and normalizes it by dividing by the total squared error of the default classifier (Witten and Frank 2005).

The two observers' error matrices were used to calculate an inter-rater agreement statistics (Kappa) to evaluate the agreement between the two classification results. The result of Kappa statistics agreement was 0.977035, which shows that we did not obtain 100% agreement.

In order to further evaluate the results of the classification accuracy so as to establish which of the two classifiers performed better, the McNemar's test, which is a non-parametric statistical test, was conducted on the generated results. According to Japkowicz and Shah (2011), the McNemar's test relies on the following four values observed on the testing set. The number of misclassified instances by both classifiers, the number of misclassified instances by SVM but correctly classified by RF, the number of misclassified instances by RF but correctly classified by SVM, and the number of correctly classified instances by both classifiers. Results obtained from McNemar's test revealed that the null hypothesis that SVM and RF performed equally with $1-\alpha$ confidence ($\alpha = 0.5$) was rejected.

The performances of SVM and RF classifiers were also examined by using the producer accuracy metrics. The producer accuracy is the accuracy of each endmember obtained by dividing the number of correctly classified pixels by the total number of pixels in the corresponding row. Figure 5 shows this result, wherein the performance of SVM classifier is comparable with that of RF classifier. The reason for the low producer classification accuracy obtained for grass/pasture (gp) and perennial vegetation (gt) classes is because they have less leaves cover that may not be distinguished in terms of endmembers (Landgrebe 1998).

The overall classification accuracy result shows that RF based classification algorithm can increase land cover classification accuracy (Congalton and Mead 1983; Rosenfield and Fitzpatrick-Lins 1986; Demir and Ertürk 2007). This result further shows that classification accuracy of RF classifier is comparable with that of SVM classifier, which performs slightly better than RF with a difference of 0.14%. In terms of the time taken by the classifiers to build their classification models, RF classifier took an average of 2.7 s, but SVM took an average of 26 s. This result directly implies that using RF classifier instead of regular SVM classifier is by far computationally efficient. Comparing the results obtained in this study with that which were obtained by previous authors (Palsson *et al.* 2012)

who used pansharpening technique on IKONO and QuickBird datasets, RF performance is comparable with SVM. The experiment conducted by Bosch *et al.* (2007) who used multiclass classification on general scene images reported that RF was much faster than SVM, which this study also confirmed.

5. Conclusion

This study implemented PPI algorithm for endmember spectral response determinations. The method generated a series of eight images, each having the size of the original image. The study also aimed at establishing performance comparison between SVM and RF classifiers for land cover classification. Spectral unmixing scheme for hyperspectral image procedure was used in the preprocessing of the dataset. Environment for visualizing image (ENVI) tool was used to extract endmember spectral response of the region of interest. The GRG optimization technique was used to estimate fractional abundance of each pixel in the region of interest after obtaining endmember spectral responses and pixel values of the endmembers. The normalized numeric values of the fractional abundances were then used for land cover classification.

The Kappa statistics was computed to evaluate the agreement between the classification results obtained by SVM and RF classifiers. We also explored the McNemar's test to verify the results of classification accuracy. In addition, performance assessment was done by using overall accuracy results and the error confusion matrix to compute producer accuracy. Experimental results demonstrate that the generation of regular SVM and RF based land cover classification systems significantly improve overall accuracy, producer accuracy, and user accuracy. The comparability and high accuracy performance of SVM and RF indicate that the GRG method is effective for solving a linear spectral unmixing problem of land cover classification.

Acknowledgements

References

Abadie J and Carpentier J 1969 Generalization of the Wolfe reduced gradient method in the case of non-linear constraints; In: *Optimization* (ed.) Fletcher R (London: Academic Press), pp. 37–47.

Abe B T, Olugbara O O and Marwala T 2012 Hyperspectral image classification using random forest and neural network; *Lecture Notes in Engineering and Computer Science: Proceedings of the World Congress on Engineering and Computer Science*, WCECS 2012, 24–26 October, San Francisco, USA, pp. 522–527.

Adams J B, Sabol D E, Kapos V, Filho R A, Roberts D A, Smith M O and Gillespie A R 1995 Classification of multispectral images based on fractions of endmembers: Application to land-cover change in the Brazilian Amazon; *Remote Sens. Environ.* **52(2)** 137–154.

Bateson C A and Curtiss B 1996 A method for manual endmember selection and spectral unmixing; *Remote Sens. Environ.* **55** 229–243.

Bateson C A, Asner G P and Wessman C A 2000 Endmember bundles: A new approach to incorporating endmember variability into spectral mixture analysis; *IEEE Trans. Geosci. Remote Sens.* **38** 1083–1094.

Bishop C M 2006 Pattern recognition and machine learning, 1st edn, Springer.

Boardman J 1993 Automating spectral unmixing of AVIRIS data using convex geometry concepts; In: *Summaries of airborne earth science workshop*, JPL Publication 93–26, pp. 111–114.

Boardman J W, Kruse F A and Green R O 1995 Mapping target signatures via partial unmixing of AVIRIS data; In: *Summaries of the VI JPL Airborn Earth Science Workshop*, pp. 23–26.

Bosch A, Zisserman A and Munoz X 2007 Image classification using random forests and ferns; In: *Proc. ICCV* 1–8.

Bowles J, Palmadesso P J, Antoniades J A, Baumback M M and Rickard L J 1995 Use of filter vectors in hyperspectral data analysis; *Proc. SPIE* **2553** 148–157.

Breiman L 1996 Bagging predictors; *Machine Learning* **24(2)** 123–140.

Breiman L 2001 Random forests; *Machine Learning* **45(1)** 5–32.

Bruzzone L and Cossu R 2002 A multiple-cascade-classifier system for a robust and partially unsupervised updating of land-cover maps; *IEEE Trans. Geosci. Remote Sens.* **40(9)** 1984–1996.

Camps-Valls G and Bruzzone L 2005 Kernel based methods for hyperspectral image classification; *IEEE Trans. Geosci. Remote Sens.* **43(6)** 1351–1362.

Chakrabarty A, Choudhury O, Sarkar P, Paul A and Sarkar D 2012 Hyperspectral image classification incorporating bacterial foraging-optimized spectral weighting; *Artificial Intelligence Res.* **1(1)** 63–83.

Chaudhry F, Wu C, Liu W, Chang C-I and Plaza A 2006 Pixel purity index-based algorithms for endmember extraction from hyperspectral imagery; In: *Recent Advances in Hyperspectral Signal and Image Processing* (ed.) Chang C-I, Trivandrum, India: Research Signpost 3, pp. 31–61.

Chen Y, Crawford M and Ghosh J 2007 Knowledge based stacking of hyperspectral data for land cover classification; *Proceedings of the IEEE Symposium on*

Computational Intelligence and Data Mining, pp. 316–322.

Congalton R 1988 A comparison of sampling schemes used in generating error matrices for assessing the accuracy of maps generated from remotely sensed data; *Photogramm. Eng. Remote Sens.* **54(5)** 593–600.

Congalton R G 1991 A review of assessing the accuracy of classifications of remotely sensed data; *Remote Sens. Environ.* **37(1)** 35–46.

Congalton R G and Mead R A 1983 A quantitative method to test for consistency and correctness in photo-interpretation; *Photogramm. Eng. Remote Sens.* **49** 69–74.

Congalton R G, Oderwald R G and Mead R A 1983 Assessing landsat classification accuracy using discrete multi-variate statistical techniques; *Photogramm. Eng. Remote Sens.* **49** 1671–1678.

Cortes C and Vapnik V 1995 Support-Vector Networks; *Machine Learning* **20(3)** 273–297.

Demir B and Ertürk S 2007 Hyperspectral image classification using relevance vector machines; *IEEE Geosci. Remote Sens. Lett.* **4(4)** 586–590.

Dobigeon N, Moussaoui S, Coulon M, Tourneret J-Y and Hero A O 2009 Joint Bayesian endmember extraction and linear unmixing for hyperspectral imagery; *IEEE Trans. Signal Process.* **57(11)** 4355–4368.

Du Q, Raksuntorn N, Younan N H and King R L 2008 End-member extraction for hyperspectral image analysis; *Appl. Opt.* **47** 77–84.

Ellis J M 2001 Searching for oil seeps and oil-impacted soil with hyperspectral imagery; *Earth Observation Magazine*, pp. 25–28.

Garner S R 1995 WEKA: The Waikato environment for knowledge analysis; *Proceedings of the New Zealand Computer Science Research Students Conference*, pp. 57–64.

Gómez D and Montero J 2011 Determining the accuracy in image supervised classification problems; *Proceedings of the 7th Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT-2011) and âœles rencontres francophones sur la Logique Floue et ses Applicationsâ (LFA-2011)*, Aix-les-Bains, France, pp. 342–349.

Gong P and Zhang A 1999 Noise effect on linear spectral unmixing; *Geographic Inform. Sci.* **5(1)** 52–57.

Gonzalez C, Resano J, Mozos D, Plaza A and Valencia D 2010 FPGA implementation of the pixel purity index algorithm for remotely sensed hyperspectral image analysis; EURASIP; *J. Adv. Signal Process.* **9,69,806** 1–13.

Hall F G, Strebel D E, Nickeson J E and Goets S J 1991 Radiometric rectification: Towards a common radiometric response among multidate, multisensor images; *Remote Sens. Environ.* **35** 11–27.

Heinz D C and Chang C 2001 Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery; *IEEE Trans. Geosci. Remote Sens.* **39(3)** 529–545.

Iordache M-D, Bioucas-Dias J M and Plaza A 2011 Sparse unmixing of hyperspectral data; *IEEE Trans. Geosci. Remote Sens.* **49(6)** 2014–2039.

Japkowicz N and Shah M 2011 *Evaluating Learning Algorithms: A Classification Perspective*; Cambridge University Press.

Karaska M A, Hugenin R L, Beacham J L, Wang M, Jenson J R and Kaufman R S 2004 AVIRIS measurements of chlorophyll, suspended minerals, dissolved organic carbon and turbidity in the Neuse River, North Calina; *Photogramm. Eng. Remote Sens.* **70(1)** 125–133.

Kärdi T 2007 Remote sensing of urban areas: Linear spectral unmixing of landsat thematic mapping images acquired over Tartu (Estonia); *Proc. Estonian Acad. Sci. Biol. Ecol.* **56(1)** 19–32.

Keshava N and Mustard J F 2002 Spectral unmixing; *IEEE Signal Processing Magazine* **19(1)** 44–57.

Lacar F M, Lewis M M and Grierson I T 2001 Use of hyperspectral imagery for mapping grape varieties in the Barossa Valley, South Australia; *Geosci. Remote Sens. Symp. IGARSS '01* **6** 2875–2877.

Landgrebe D 1998 *Multispectral data analysis: A signal theory perspective*; School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN.

Landgrebe D and Biehl L 2001 *An Introduction to MultiSpec. School of Electrical and Computer Engineering*, Purdue University.

Lasdon L S, Fox R L and Ratner M W 1974 Nonlinear optimization using the generalized reduced gradient method; *Revue française d' automatique, d' informatique et de recherché* **3** 73–103.

Lennon M, Mercier G and Hubert-Moy L 2002 Classification of hyperspectral images with nonlinear filtering and support vector machines; In: *Proc. IEEE Trans. Geosci. Remote Sens. Symposium*, Toronto, ON, Canada, **3** 1670–1672.

Li J 2011 Remote sensing image information mining with HPC cluster and DryadLINQ; Proceedings of the 49th ACM Southeast Conference ACM SE'11, pp. 227–232.

Liu Y and Zhen Y F 2005 One-against-all multi-class SVM classification using reliability measures; *Proc. IEEE Int. Joint Conf. Neural Networks* **2** 849–854.

Liu H, Fan Y, Deng X and Ji S 2009 Parallel processing architecture of remotely sensed image processing system based on cluster; *Image and Signal Process.* CISP '09, 2nd Int. Congress, pp. 1–4.

Martinez P J, Pérez R M, Plaza A, Aguilar P L, Cantero M C and Plaza J 2006 Endmember extraction algorithms from hyperspectral images; *Ann. Geophys.* **49(1)** 93–101.

Melgani F and Bruzzone L 2004 Classification of hyperspectral remote sensing images with support vector machines; *IEEE Trans. Geosci. Remote Sens.* **42(8)** 1778–1790.

Mountrakis G, Im J and Ogole C 2011 Support vector machines in remote sensing: A review; *ISPRS J. Photogramm. Remote Sens.* **66** 247–259.

Nascimento J M P and Bioucas-Dias J M 2005 Does independent component analysis play a role in unmixing hyperspectral data? *IEEE Trans. Geosci. Remote Sens.* **43(1)** 175–187.

Pal M 2005 Random forest classifier for remote sensing classification; *Int. J. Remote Sens.* **26(1)** 217–222.

Palsson F, Sveinsson J R, Benediktsson J A and Aanaes H 2012 Classification of Pansharpened Urban Satellite Images; *Selected Topics in Appl. Earth Observations and Remote Sens. IEEE J.* **5(1)** 281–297.

Plaza A, Martínez P, Plaza J and Pérez R 2003 Spectral analysis of hyperspectral image data; *Advances in Technique for Analysis of Remotely Sensed Data*, IEEE Workshop, pp. 298–307.

Plaza A, Martinez P, Perez R and Plaza J 2004 A quantitative and comparative analysis of endmember extraction algorithms from hyperspectral data; *IEEE Trans. Geosci. Remote Sens.* **42(3)** 650–663.

Plaza A, Plaza J and Cristo A 2008 Morphological feature extraction and spectral unmixing of hyperspectral images; *IEEE Int. Symp. Signal Process. Information Technol.*, pp. 12–17.

Rodriguez-Galiano V F, Ghimire B, Rogan J, Chica-Olmo M and Rigol-Sanchez J P 2011 An assessment of the

effectiveness of a random forest classifier for land cover classification; *ISPRS J. Photogramm. Remote Sens.* **67(9)** 3–104.

Rosenfield G and Fitzpatrick-Lins K 1986 A coefficient of agreement as a measure of thematic classification accuracy; *Photogramm. Eng. Remote Sens.* **52** 223–227.

Sanchez S, Martin G, Plaza A and Chang C 2010 GPU implementation of fully constrained linear spectral unmixing for remotely sensed hyperspectral data exploitation; *Proceedings SPIE Satellite Data Compression, Communications, and Processing VI* **7810** 78100G-1–78100G-11.

Settle J J and Drake N A 1993 Linear mixing and the estimation of ground cover proportions; *Int. J. Remote Sens.* **14(6)** 1159–1177.

Shao Y and Lunetta R S 2012 Comparison of support vector machine, neural network, and CART algorithms for the land-cover classification using limited training data points; *ISPRS J. Photogramm. Remote Sens.* **70(0)** 78–87.

Story M and Congalton R G 1986 Accuracy assessment: A user's perspective; *Photogramm. Eng. Remote Sens.* **52** 397–399.

Su C-T and Lii G-R 1995 Reliability optimization design of distribution systems via multi-level hierarchical procedures and generalized reduced gradient method, energy management and power delivery; *Proceedings of EMPD '95, Int. Conference* **1** 180–185.

Su H, Sheng Y and Du P 2008 A new band selection algorithm for hyperspectral data based on fractal dimension; *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **37 Part B7** 279–283.

Tarabalka Y, Chanussot J and Benediktsson J A 2009 Classification based marker selection for watershed transform of hyperspectral images; *Geosci. Remote Sens. Symp. IEEE Int. IGARSS* **3** III–105–III–108.

Theiler J, Lavenier D, Harvey N, Perkins S and Szymanski J 2000 Using blocks of skewers for faster computation of pixel purity index; *Proc. SPIE Int. Conf. Optical Sci. Technol.* **4132** 61–71.

Udelhoven T, Waske B, Linden S and Heitz S 2009 Land-cover classification of hypertemporal data using ensemble systems; *IEEE Trans. Geosci. Remote Sens.*, pp. III-1012–III-1015.

Ul Haq Q S, Shi L, Tao L and Yang S 2010 Hyperspectral data classification via sparse representation in homotopy; *Inf. Sci. Eng. (ICISE) 2010, 2nd Int. Conf.*, pp. 3748–3752.

Van der Meer F D and Jia X 2012 Collinearity and orthogonality of endmembers in linear spectral unmixing; *Int. J. Appl. Earth Observ. Geoinform.* **18** 491–503.

Winter M E 1999 N-FINDR: An algorithm for fast autonomous spectral endmember determination in hyperspectral data; *Proc. SPIE Imaging Spectrometry V* **3753** 266–275.

Witten I H and Frank E 2005 Data mining: Practical machine learning tools and techniques, 2nd edn, Morgan Kaufmann, San Francisco, pp. 176–178.

Wolter P T, Johnston C A and Niemi G 2005 Mapping submergent aquatic vegetation in the US great lakes using quickbird satellite data; *Int. J. Remote Sens.* **26** 5255–5274.

Xie Y, Sha Z and Yu M 2008 Remote sensing imagery in vegetation mapping: A review; *J. Plant Ecol.* **1** 9–23.

Zhang B, Sun X, Gao L and Yang L 2011 Endmember extraction of hyperspectral remote sensing images based on the ant colony optimization (ACO) algorithm; *IEEE Trans. Geosci. Remote Sens.* **49(7)** 2635–2646.

Zhou G, Wu B and Li M 2011 Improved accuracy assessment indices for object-based high resolution remotely sensed imagery classification; *Image Analysis and Signal Processing (IASP), 2011 International Conference* **181–186** 21–23.