EVALUATION OF THE PERFORMANCE OF IMAGE CLASSIFICATION METHODS IN THE IDENTIFICATION OF VEGETATION

Marcio Augusto Reolon Schmidt[*], Jaciane Xavier Bressiani, Patrícia Antunes Dos Reis and Marcio Ricardo Salla

Department of Civil Engineering, Federal University of Uberlândia, Brazil

Abstract: Orbital imaging techniques offer comprehensive coverage of different regions for numerous environmental and socioeconomic applications, revealing the spatial characteristics and land use of those regions. The advantages of remote sensing include its ability to record spatial distribution patterns, and spectral and temporal data over large regions. The objective of this research is to evaluate the performance of different multispectral image classification methods in the selection of general vegetation, based on a set of samples taken from a Landsat 8 image. The quality of multispectral images and their final classification is usually evaluated based on the Kappa index, which is used as the quality standard in many remote sensing software programs. The classification methods chosen for this study were Parallelepiped, Maximum Likelihood, Mahalanobis Distance, and neural networks. The most suitable classification was used as standard and the other images were compared with it to determine the degree of similarity ranking (IS_{1x}), defined as the percentage of pixels classified differently from those of the standard image. The IS_{1x} was determined using a Matlab routine involving pixel subtraction between images. The results indicate that probability distribution methods are more suitable for discriminating vegetation types than other methods, and that some band combinations should be chosen.

---

[*] Correspondence to: Marcio A.R. Schmidt, Tel.: +55 34 3230-9430; Fax: +55 34 3239 4155.
E-mail: marcioschmidt@feciv.ufu.br

INTRODUCTION

Environmental analysis involves phenomena widely distributed in the landscape. Some examples of the wide spatial occurrence of phenomena include the dispersion of pollutants along coastlines, land use and occupation of watersheds, and growth vectors of medium and large cities. It is essential to select the correct tools that allow for this type of analysis with the necessary scale and precision, and which provide a continuous broad coverage in order to obtain data for environmental analysis by experts. Remote sensing images stand out in this context.

Orbital imaging techniques stand out because they provide a broad coverage of the study area, highlighting characteristics of land cover and land use, i.e., vegetation, soil, water and forests. In addition, this data collection technique can be used in different contexts such as mapping for military and civilian applications, environmental damage assessment, land use monitoring, urban planning and urban growth trends, agricultural productivity, etc. One of the main applications of remote sensing is to create thematic maps from significant land cover classes in a given setting as a way to reduce the dimensions of the analysis.

The remote sensing community is interested in research on remote sensing image classification because this information collection technique is applicable in environmental and socioeconomic studies (Perumal & Baskaran, 2010). According to Al-Ahmadi and Hames (2009), one of the reasons why researchers and users prioritize the use of satellite images over other survey methods is their ability to determine spatial, spectral and temporal data about large regions, instead of having to obtain pointwise data, information from hard to reach places, and to create spatial patterns.

According to Perumal & Baskaran (2010), image classification has become an important tool in the analysis of digital images in various applications involving spatial phenomena. However, the selection of the most suitable classification technique can have a significant impact on the results. If the classification is used as a final product or as an intermediate analytical process in a more complex analysis, it is essential to study the problem, select the images, collect samples, and choose the appropriate classification method.

SUPERVISED CLASSIFICATION

According to Richards & Jia (2006), supervised classification is the procedure most commonly used in the quantitative data analysis of remote sensing images. This technique involves labeling the pixels in an image to represent certain types or classes of land cover. A variety of algorithms is available for this classification, ranging from algorithms based on probability distribution models, in which the multispectral space is divided into specific regions, to segmentation, object orientation and other algorithms. However, despite the existence of these techniques, most image classification software programs are based on pixel-to-pixel classification techniques, considering multidimensional analyses with as many variables as dimensions of space. In these techniques, the distribution determines the chance of finding a pixel belonging to this class in any given location in the multispectral space. These techniques include probability distribution techniques such as Minimum Distance, Maximum Likelihood, and Mahalanobis Distance.

The probability distribution is determined based on an analysis of the samples selected by the user for each class in the image. By providing a group of pixels as a sample of a class, each pixel is considered a vector with each field corresponding to the value of reflectance of the class in a given bandwidth of the image. By determining the distance between the average pixel values of the class in each band, the evaluation can be performed based on a normal or Gaussian distribution, which measures the likelihood of the pixel belonging to one class or the other.

Richards & Jia (2006) state that this distribution is robust in the sense that classification accuracy is not overly sensitive to the violation of the assumption that the classes meet normal distributions. However, the use other techniques such as the Mahalanobis distance, which includes internal variance of the classes, and artificial neural networks (ANN), enables different rankings for the same group of pixels.

Classification procedures generally involve the definition of samples and their classification to evaluate their consistency, and the final classification is based on the limits set in the previous step. Regardless of the selected method, the quality of a supervised classification initially depends on the quality of the samples in terms of number of pixels and spectral representability. Subsequently, the results are dependent on the performance of the classification methods.

Perumal & Baskaran (2010) argue that attention has only recently focused on comparative studies of classification methods that correlate their performance with certain types of data. Successful classification requires experience and experimentation by the user, who should select a classification method that best performs a specific task. Even if the best method for each type of image cannot be determined in every possible situation, it is important to be familiar with the characteristics of the methods and how to perform sampling correctly to ensure the highest possible accuracy in the classification of images. Another important point to keep in mind is the distribution of the samples. Samples concentrated in a small portion of the image tend to exhibit similar responses and increase the bias of the classification. Distributed samples show greater variability, and thus tend to be more representative of the class to which they belong. In

general, the sampling quality and final classification are assessed based on the Kappa index.

Kappa analysis is a discrete multivariate technique employed for the evaluation of thematic accuracy, which uses all the elements of a confusion matrix in its calculation. This robust technique is used as a standard to check the quality of the classical classification methods (Ben-David, 2008). The Kappa coefficient (K) is a measure of the true concordance (indicated by the diagonal elements of the confusion matrix) minus the chance concordance (indicated by the total product of the row and column, which does not include unrecognized entries). In other words, it is a measure of the degree to which the classification is in line with the reference data (Figueiredo & Vieira, 2007). This index therefore expresses the consistency of the sample itself, considering the pixels listed for each class, and how much the pixels in the image are consistent with the sampling of each class.

Therefore, this paper presents a variation of classification methods starting from the same sample set over a range of spectral band combinations in a Landsat 8 image to identify areas of vegetation in a small watershed. The classification methods employed in this study were Parallelepiped, Maximum Likelihood, Mahalanobis Distance and neural networks, the latter because of its emphasis in the literature and the number and versatility of its applications.

MATERIALS AND METHODS

The main objective of this research is to evaluate the performance of different multispectral image classification methods in the discretization of vegetation in general, using the same samples taken from the same image. The methods selected here are part of the computational tools of ENVI 4.5 software, and their selection was based on their image resolution compatibility with the Landsat 8 series so as to exclude the methods developed for high-resolution images such as object orientation, quad-tree, segmentation and others.

A Landsat 8 satellite image was used in this research. The sensor has eight spectral bands, with the multispectral bands ranging from blue to mid-infrared (bands 1–8), two thermal bands, and a panchromatic band (USGS, 2014). Three different combinations were used for the tests, based on the image taken at orbit 73, point 221, recorded on 04 Oct 2014. The first band combination was with bands 5, 4, 3, the second with bands 6, 5, 4, and the third with bands 4, 3, 2, corresponding, respectively, to false-color, infrared and visible. These combinations were chosen because of the behavior of the spectral response of the vegetation, which is the target of this research, and presented in the literature by Jensen (2009).

To assess the sensitivity of the classification methods, samples were selected containing 1, 3 and 5% of the total number of pixels in the image. This choice is explained by the fact that choosing samples in certain land cover classes presents a certain degree of difficulty, e.g., in cases in which untrained human operators may adopt an excessively small or large number of samples. In the former case, the samples would be insufficient to train the classification algorithm, and in the latter, a large number of samples could influence the responses of the algorithms, inducing unrealistic limits, particularly in the simpler classification methods. By default, eight areas of mixed vegetation were selected (pastureland, scrubland and forests) in the images. In ENVI software, the sample areas selected by the user are called ROI (Regions of Interest) and can be saved in a file. This feature of the program enabled the same samples to be used in each test and in each variation of the classification methods and band combination. A description of the tested methods is given below.

(a) Parallelepiped Method
This is the simplest method, in which each band that makes up an image is represented as an axis in a tri-orthogonal Cartesian system. Thus, a given class is represented by the minimum and maximum values of its projection on each axis. The class can be understood as a triaxial ellipsoid in this space, and the values of the vectors that define the characteristics of each pixel must be contained in the image in order to be allocated to this class.

However, this technique poses a problem when samples are not spectrally pure, i.e., they do not contain only pixels of the same class, which may cause them to overlap the boundaries of another class. This may lead some pixels to be incorrectly attributed to one of the conflicting classes. Moreover, some pixels may fail to be classified because the limits adopted by the sample may not be representative of all the pixels in the image.

(b) Maximum Likelihood Method (MLM)
This is one of the most popular methods of supervised classification used with remote sensing data (Richard & Jia, 2006; Jensen, 2009). This method involves the determination of the probability of a pixel belonging to a given class, based on normal probability distribution. The probability is estimated from the trained samples, which are used to calculate the means and variances of the classes to which they are associated, and also considers the variability of the brightness values in each class. This classifier is based on Bayesian probability theory and is one of the most powerful classification methods when accurate sampling data are provided (Perumal & Bhaskaran, 2010). This method depends strongly on a normal distribution of data in each spectral band of the image, and tends to over-classify spectral signatures with relatively high values in the covariance

matrix (Ahmadi & Quegan, 2012) when samples are contaminated with pixels from other classes. **Equation 1** shows the Bayesian classification (Richards & Jia, 2006):

$$p(x) = \sum_{i=1}^{n} p(x|w_i) . p(w_i) \qquad (1)$$

where:

x = intensity vector of the sampled pixel in each band of the image

w_i = class i of the set of classes selected by the user

$p(x)$ = probability of vector x

$p(x/w_i)$ = conditional probability vector of the pixel in each band for class w_i

$p(w_i)$ = probability of the class occurring in the image compared to that of the other classes

$i = 1 ... n$ = number of classes

(c) Mahalanobis Distance

The Mahalanobis distance method considers that the correlations between samples represent the distance from the average of the class, scaled by the class covariance (Li & Fox, 2011). Thus, the method considers both inter- and intra-class spectral variability. The Mahalanobis distance is a useful way to determine the similarity of an unknown sample to a set of known samples, as in the case of supervised classification. However, like the MLM, this method tends to over-classify pixels with spectral responses of relatively high values in the covariance matrix, and is also strongly dependent on the normal distribution of data in each band of the image (Li & Fox, 2011; Fauvel *et al.*, 2010).

Li & Fox (2011) argue that, because it is a statistical method, it can be used to produce the measure known as typicality, which is the probability of any site to have a Mahalanobis distance greater than or equal to the one observed for the site of interest. Thus, the pixel is classified according to the shortest Euclidean distance between the pixel and the class, considering the typicality of the sample (**Eq. 2**):

$$d(x, m_i)^2 = (x - m_i)^t . \Sigma^{-1}(x - m_i) \qquad (2)$$

where:

x = intensity vector of the sampled pixel in each band of the image

m_i = vector of mathematical expectation of each element of x in class i

$\Sigma^{-1} (x - m_i)$ = covariance matrix of sampled pixels for class i

$d(x, m_i)$ = Euclidean distance between pixel x and class m_i

(d) Artificial Neural Networks (ANN)

Artificial Neural Network (ANN) models are based on machine knowledge and learning, and according to Mohammady *et al.* (2014), they are powerful tools to quantify and model complex data distribution patterns whose function is to handle imprecise training data. In these models, the ANN is arranged in an input layer, at least one intermediate layer, and an output layer, and each layer can contain from one to several neurons, called perceptrons. The network learns by adjusting the weight of each link between layers to minimize the difference between the output and input. The input layer receives the vectors from the pixels used as samples. The intermediate layers combine the links between the perceptrons of the input layer to create a set of discriminant functions that can limit the characteristics of the classes in each band (Richards & Jia, 2006).

After the network has been trained with the samples taken from the images, the recursive method introduced by backpropagation enables the automated adjustment of the weights defined for the various layers of the network, which is understood as the network learning process. Forward and backpropagation is applied successively until the network has learned the characteristics of all the classes, and the root-mean-square error (RMSE) between the embedded and ideal activation levels reaches an acceptable value (Richards & Jia, 2006).

For the tests, the classification methods were set up according to a few minor user-defined adjustments of the ENVI settings. For the MLM, Mahalanobis and Parallelepiped methods, a single value of no less than 90% was set for the standard deviation in order to control the classification parameters. In the ANN method, the activation function was set at a value of 0.9 for the contribution of the training threshold and 0.2 for the training rate. The "Training RMS Exit Criteria" was set to 1.0 as the limit value in which the training should stop, with 1 being the number of internal layers for nonlinear classification and 1000 for the desired number of interactions in the implementation of training.
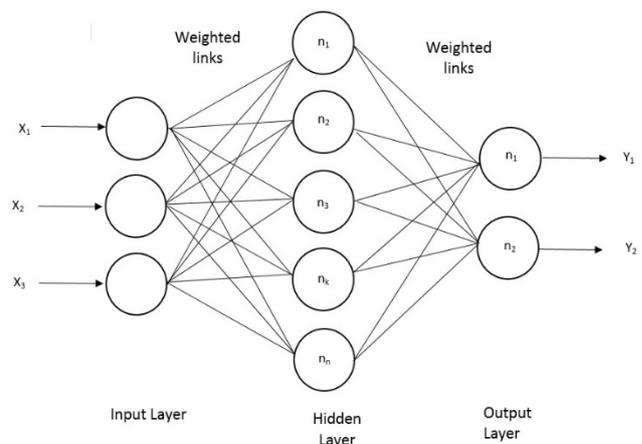
Fig. 1 Configuration diagram of a neural network. Source: Adapted from Richards & Jia (2006)

Table 1. Kappa index obtained in the classification of the samples in RGB Colorspace

COMBINATION	RGB Colorspace								
	False-color band combination (bands 6,5,4)			Infrared band combination (bands 5,4,3)			Visible band combination (bands 4,3,2)		
METHOD	Sample size								
	1%	3%	5%	1%	3%	5%	1%	3%	5%
Parallelepiped	0.998	0.964	0.965	0.147	0.152	0.214	0.927	0.907	0.879
Maximal Likelihood	1.000	1.000	1.000	1.000	0.915	0.943	0.998	0.919	0.925
Mahalanobis Distance	1.000	1.000	0.999	0.925	0.391	0.504	0.976	0.833	0.843
Artificial Neural Network	1.000	0.853	0.851	0.971	0.516	0.566	0.992	0.441	0.148

Finally, the classification results of each method were transformed into binary images. In these images, the pixel classified as vegetation was set to a value of 1 (white) and any other areas were classified as 0 (black). Based on photo-interpretation, the image that most closely resembled the original image was used as a standard against which the other binary images were compared in order determine the degree of similarity. The degree of similarity was defined as the percentage (IS_{1x}) of pixels classified differently from those of the standard image (**Eq. 3**),

$$IS_{1x} = \frac{\sum_{i=1}^{n}(pw_{iIx}-pw_{iI1})}{\sum_{i=1}^{n}pw_{iI1}} \qquad (3)$$

where:
pw_{iIx} = number of standard image pixels classified as vegetation
pw_{iI1} = number of pixels of the other images classified as vegetation.

RESULTS AND DISCUSSION

Table 1 describes the results of the classification of the samples by combination, sampling size and classification method. It is noteworthy that, as the sample size increases, the Kappa index decreases to less than 1% of the sample. The values in **Table 1** suggest that the more complex the method the lower the resulting Kappa index.

The infrared combination presented the worst results, i.e., significantly lower quality than that of the other methods. This may indicate that this combination does not allow for a clear definition of the limits of vegetation classes based on the samples in the simpler methods or not based on probability distribution. This is partially due to the spectral response of vegetation in the red and infrared bands of the Landsat satellite, which show similar responses of chlorophyll, although these responses present seasonal variations.

False-color band combination (bands 6,5,4)
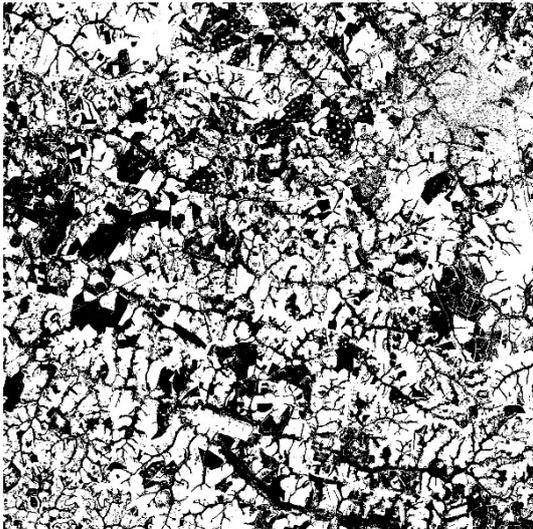
a – Parallelepiped method

Visible band combination (bands 4,3,2)
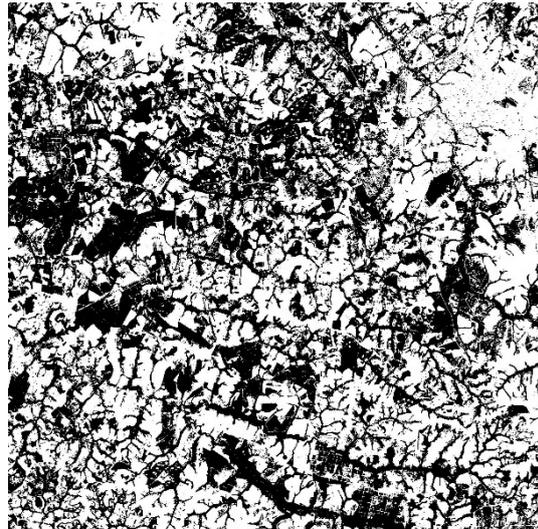
b – Parallelepiped method
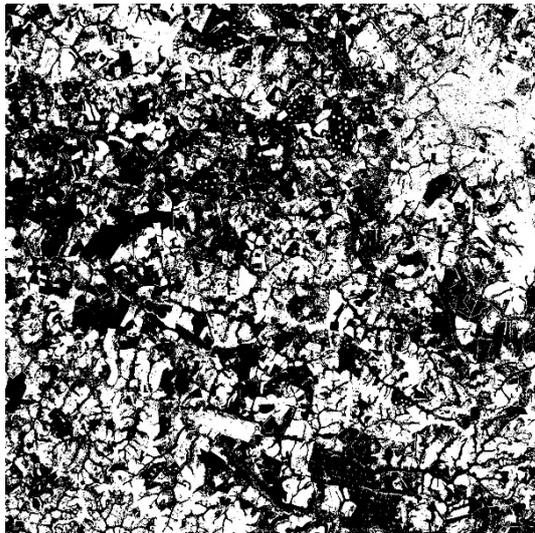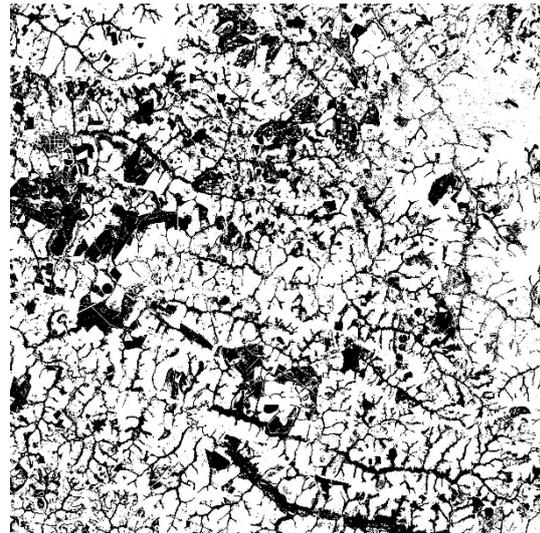
c – Maximum Likelihood method   d – Maximum Likelihood method
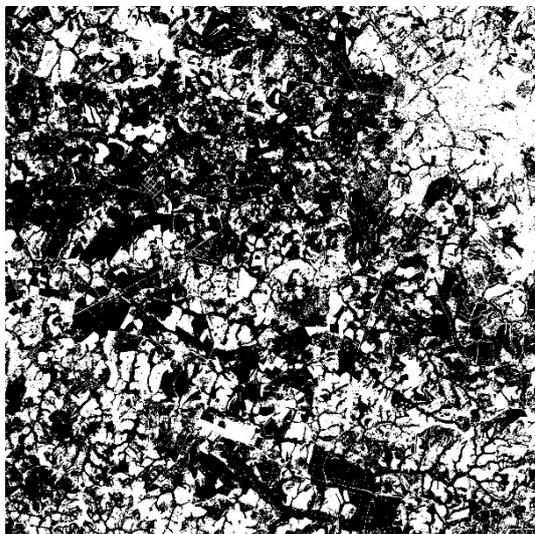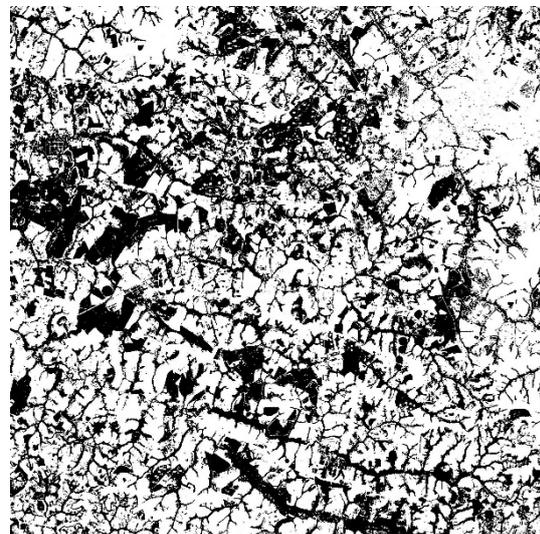
e – Mahalanobis Distance method   f – Mahalanobis Distance method
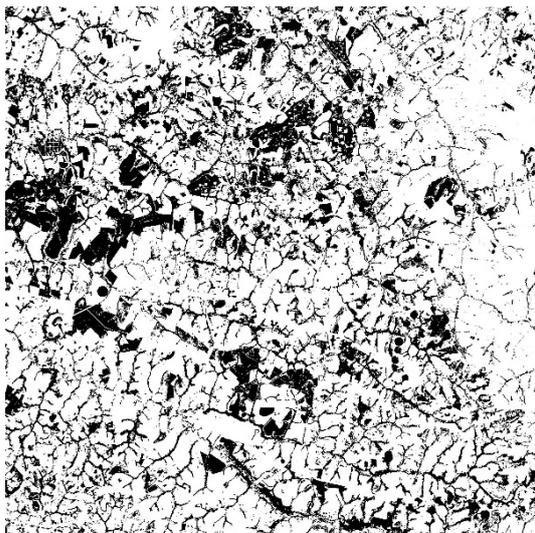
g – Artificial Neural Network   h – Artificial Neural Network

Fig 2. Images resulting from each classification method using samples equivalent to 3%.

Another classification that stands out is the results of the neural networks in the infrared (bands 6, 5, 4) and visible (bands 4, 3, 2) band combinations of RGB Colorspace. **Table 1** reveals a significant decrease in the

Kappa index in response to increasing sample size. This appears to indicate that the network configuration used here may be sensitive to noise in the training data in these combinations. In the ENVI program, the network configuration options for the user are limited. This also holds true for the other classification methods that are less robust than the MLM.

The three original combinations were then classified within the limits identified by each method. This study assessed only the vegetation class in general. Therefore, the image resulting from each method is binary, i.e., 1 (white) corresponds vegetation and 0 (black) to non-vegetation. **Figure 1** illustrates the results of two combinations in the RGB Colorspace applied to the Parallelepiped (**Figs 1a** and **1b**), MLM (**Figs 1c** and **1d**), Mahalanobis Distance (**Figs 1e** and **1f**) and ANN (**Figs 1g** and **1h**) methods with a 3% sample.

Note that the variations in the areas obtained with the 3% sample were visually similar in the false-color band combination (**Fig. 1**, left-hand column), and significantly greater the visible band combination.

A visual analysis revealed that the MLM presented better results than the other methods in the false-color band combination. The Mahalanobis method also presented a good resolution of the vegetation polygons in the visible and false-color band combinations, but poor resolution for the area of scrubland vegetation, where the pixels merged with those of exposed soil. The Parallelepiped method contains information not identified by the sample, known as *no data*, which appeared in the image of the false-color band combination, directly influencing the quality of the final image. The best classification obtained by the ANN method was with the false-color band combination, which presented highly detailed vegetation polygons.

To calculate the relative degree of similarity between the classification methods (IS_{1x}), a routine was developed in Matlab to compare each of the classified images against the standard image by means of pixel-by-pixel subtraction. As a result, new images were obtained whose pixel values express the differences from the standard image. These differences are expressed mainly at the edges of vegetation polygons. The IS_{1x} were then determined, as shown in **Table 2**.

The correlation established between the Kappa index of the samples and the resulting classification by each method are shown in **Fig. 2**. In **Fig. 2**, note that the highest Kappa values are lower in the (1%) samples. In the false-color band combination (**Fig. 2a**), all the methods presented a classification index above 0.8, but they tended to stabilize at lower values when the sample size was increased. In **Fig. 2b**, the infrared band combination showed the worst results in response to variations in sample size. The parallelepiped method was significantly affected by this combination. As can be seen in **Fig. 2c**, visible band combination, only the ANN method was affected by the change in sample size, possibly indicating that the method was unable to discretize the new samples. However, the values obtained by other methods were higher than 0.8, with values close to the false-color ones.

Figure 3 illustrates the behavior of the similarity index, which follows the trend described by the Kappa index. The classifiers based on probability distribution yielded better results when the neural networks showed a higher variation in response to the increase in samples. **Figure 3b** shows that the best results for the infrared band combination were obtained with the classification method based on probability distribution. Moreover, the similarity values and these methods show similar and constant results, while the parallelepiped method shows significant variations in this combination. Overall, the neural networks showed the poorest classification results, indicating the need for further tests using new settings to achieve more consistent results.

Table 2. Similarity index of the classification methods (in %)

COMBINATION	RGB Colorspace								
	False-color band combination (bands 6,5,4)			Infrared band combination (bands 5,4,3)			Visible band combination (bands 4,3,2)		
METHOD	Sample size								
	1%	3%	5%	1%	3%	5%	1%	3%	5%
Parallelepiped	0.7902	0.83928	0.8945	0.8506	0.6046	0.6163	0.8347	0.8498	0.8591
Maximal Likelihood	0.9099	1	0.9303	0.8669	0.8169	0.8299	0.6208	0.7424	0.8002
Mahalanobis Distance	0.8855	0.8872	0.8896	0.7794	0.8472	0.8438	0.8143	0.8083	0.8336
Artificial Neural Network	0.8316	0.7022	0.2950	0.6437	0.4866	0.5060	0.6977	0.4889	0.4482
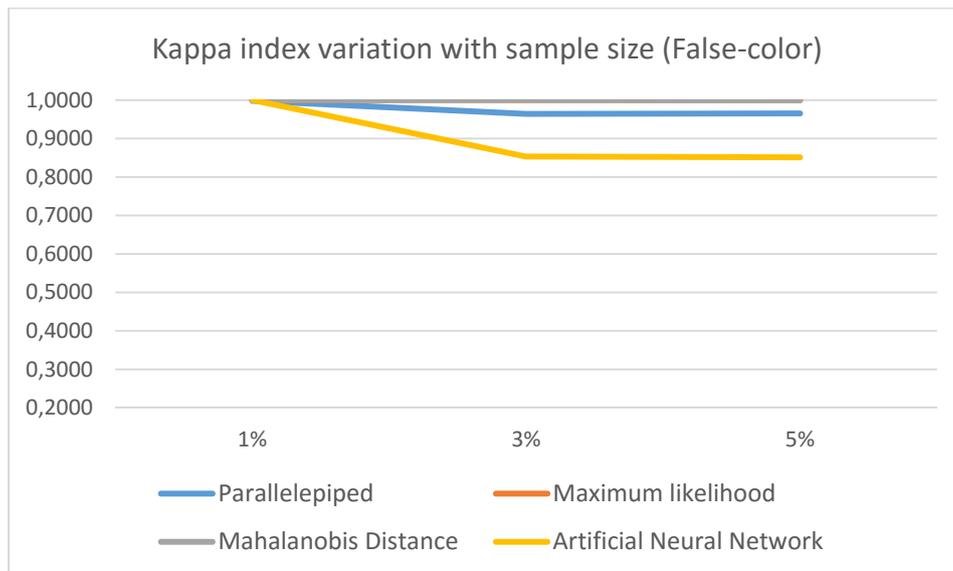
(a)
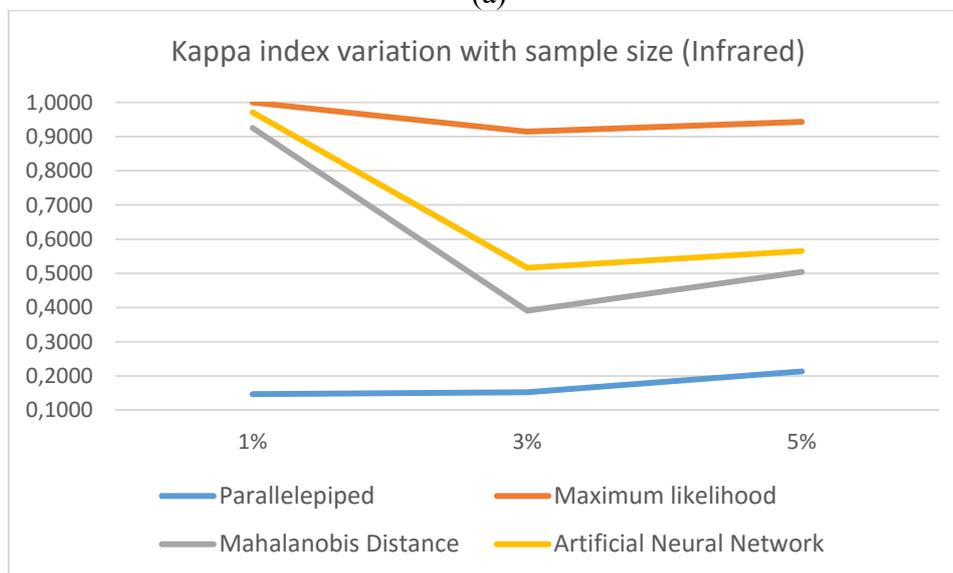
(b)
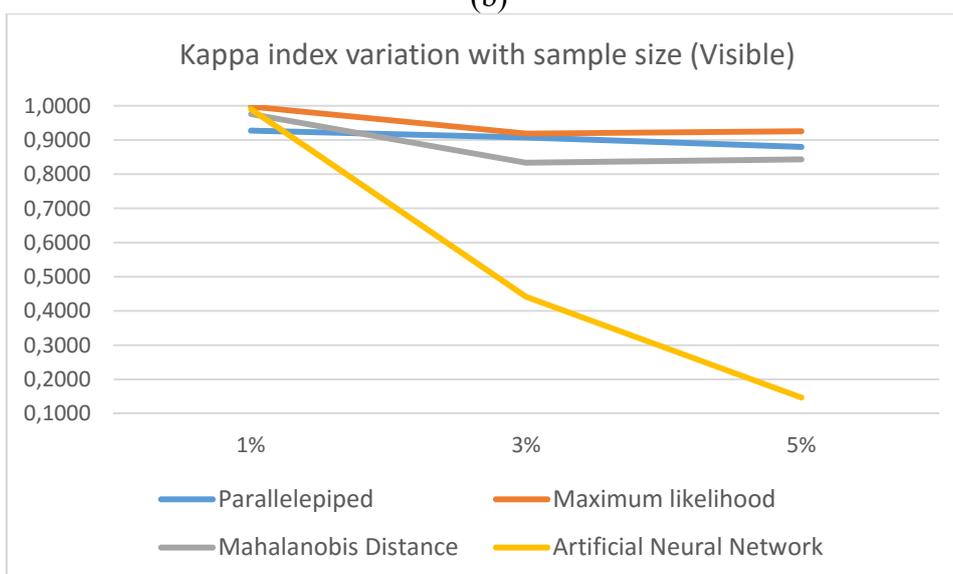
(c)

Fig 2. Kappa index of each classification method for 1, 3 and 5% of the samples.

Similarity index variation with sample size (False-color)

(a)

Similarity index variation with sample size (Infrared)

(b)

Similarity index variation with sample size (Visible)
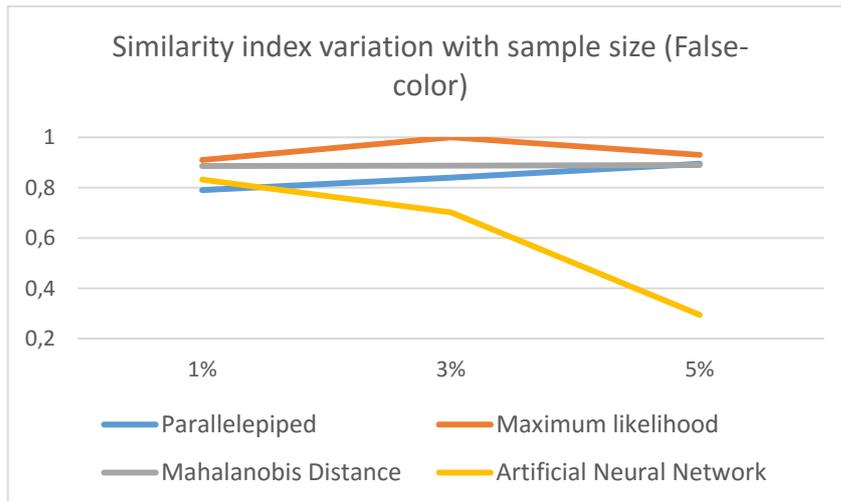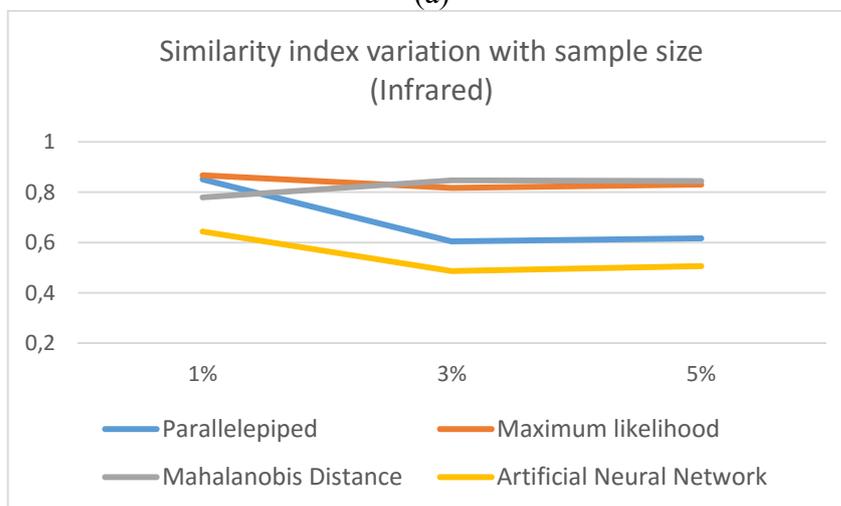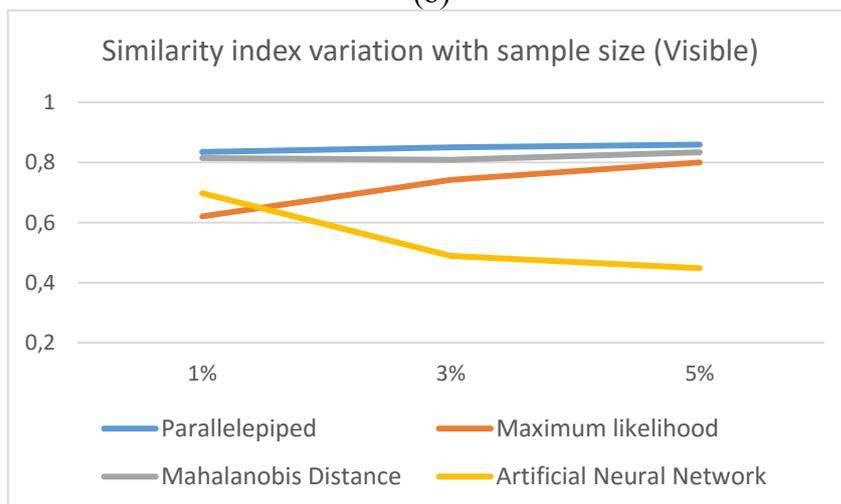
(c)

Fig 3. Variation of the similarity index among the classified images.

The final quality of the images was evaluated based on an ANOVA significance test. This test assesses the degree of significance of the means of two samples based on normal distribution. In this test, the pixels used to calculate IS_{1x} were written as vectors of pairs of values and were analyzed in Matlab. The results did not show a positive relationship between the samples, indicating that the results are uncorrelated, and hence, the Kappa values and IS_{1x} indices are independent of each other.

CONCLUSIONS

This study compared four of the multispectral classification methods most widely used in remote sensing. The image band combinations were found to affect class separation quality, and hence, the Kappa index values. This indicates that deterministic methods may produce a lower quality of pixel separability and class identification in the classification process. In addition, the Kappa index may not reflect the same degree of coherence in an evaluation of the sample and of the classified image. Based on the results, it can be stated that the more complex the method the lower the Kappa index obtained in the classification, and that a large number of samples representative of the classes does not always suffice to qualify the result.

The size, distribution and composition of samples should be analyzed carefully. The collection of samples with an acceptable level of reliability requires experience on the part of the user, since this process, as well as the visual checking of the results, are very time-consuming, but lead to positive outcomes in the definition of the results.

This research revealed that the same combination of classification method, image bands, sample size and distribution cannot always be applied for all land use identification purposes. In fact, each situation requires a separate analysis, applying band combinations that best meet the objectives of the study. The conclusion reached in this research is that the use of supervised classification should take into account not only band combinations but also the classifier method and the number of samples, aiming for a result that is most representative of the reality of the classes. As a continuation of this research, we plan to change the color space to HSV and carry out a new analysis.

REFERENCES

Ahmad, A. & Quegan, S. (2012) Analysis of Maximum Likelihood Classification on Multispectral Data. *Applied Mathematical Sciences* **6**(129), 6425 – 6436.

Al-Ahmadi, F.S. & Hames, A. S. (2009) Comparison of Four Classification Methods to Extract Land Use and Land Cover from Raw Satellite Images for Some Remote Arid Areas, Kingdom of Saudi Arabia. JKAU; *Earth Sci.* **20**(1), 167 – 191.

Ben-David, A. (2008) Comparison of classification accuracy using Cohen's Weighted Kappa. *Expert Systems with Applications* **34**,825–832.

Fauvel, M.,Villa, A., Chanussot, J. & Benediktsson, J.A. (2010) Mahalanobis Kernel for the Classification of Hyperspectral Images. In: *Geoscience and Remote Sensing Symp. (IGARSS)*, IEEE International, 3724 – 3727. DOI: 10.1109/IGARSS.2010.5651956

Figueiredo, G. C. & Vieira, C. A. O. (2007) Estudo do comportamento dos índices de Exatidão Global, Kappa e Tau, comumente usados para avaliar a classificação de imagens do sensoriamento remoto. In: *XIII Simpósio Brasileiro de Sensoriamento Remoto.* (Florianópoplis, INPE, 2007). Available at: http://marte.sid.inpe.br/col/dpi.inpe.br/sbsr@80/2006/11.13.17.35/doc/5755-5762.pdf Accessed on 20 Nov 2014.

Jensen, J.R. 2009. *Sensoriamento Remoto do Ambiente*. Parentese.

Li, Z. & Fox, J. M (2011) Integrating Mahalanobis typicalities with a neural network for rubber distribution mapping, Remote Sensing Letters, **2**(2),157 – 166. doi: 10.1080/01431161.2010.505589

Mohammady, S. Delavar, M. R. & Pahlavani, P. (2014) Urban growth modeling using an artificial neural network a case study of Sanandaj City, Iran. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume XL-2/W3, 2014 and The 1st ISPRS International Conference on Geospatial Information Research, (Tehran, Iran, November 2014), 15–17. doi:10.5194/isprsarchives-XL-2-W3-203-2014

Perumal, K. & Bhaskaran, R. (2010) Supervised classification performance of multispectral images. *J. Computing*. **2**, Issue 2, February 2010, ISSN 2151-9617 https://sites.google.com/site/journalofcomputing/

Richards, J.A. & Jia, X. (2006) *Remote Sensing Digital Image Analysis: An introduction*. Springer-Verlag Berlin Heidelberg.

USGS. Landsat 8. (2014) Available at: <http://LandSat.usgs.gov/landsat8.php>. Accessed on 28 Dec 2014.