Assessment of the *Caatinga* phenological dynamics influence on digital classification

Verônica Fernandes Gama [1]
Cesar Edwin Garcia [1]
Flávio Jorge Ponzoni [1]

[1] Instituto Nacional de Pesquisas Espaciais - INPE
Caixa Postal 515 - 12227-010 - São José dos Campos - SP, Brasil
{veronica, cgarcia, flavio}@dsr.inpe.br

Abstract. In this study we objected to assess the phenological characteristics interference on differences in the spectral response and, consequently, on the classification results of the *Caatinga* vegetation. For this we compared the regions of interest (ROIs) collected for a supervised Maximum Likelihood classification of a *Caatinga* region multi-temporal set of images. In addition to the spectral bands of the TM and ETM+ sensors (Landsat satelittes 5 and 7 respectively), we used fraction images of these data obtained through Linear Spectral Unmixing as additional attributes for the classification. The ROIs on each image were analyzed using confusion matrixes, Kappa and Z Statistics. Z Test calculations with 95% of confidence showed that most of the matrixes of the ROIs are different from each other, although they represent same targets on the scene over the dates studied. Since each image had been captured during a different precipitation situation, which might have caused the vegetation to change its structure and physiology, the amount of radiation captured by the sensors from these targets varied. These results showed that phenological characteristics cause differences in the spectral response of the *caatinga* and, consequently, interfere on the classification of the vegetation, and causes the misclassifications of targets along the time.

Palavras-chave: Semi-arid ecosystems, phenology, regions of interest, confusion matrix, ecossistemas semi-áridos, fenologia, regiões de interesse, matriz de confusão.

1. Introduction

Caatinga is an ecosystem located on the northeastern part of Brazil. It is spread through over 1 million kilometers within the states of: Piauí, Maranhão, Ceará, Rio Grande do Norte, Paraíba, Pernambuco, Alagoas, Sergipe, Bahia and part of Minas Gerais (RESERVA DA BIOSFERA DA CAATINGA, 2009). Its vegetation has xerophytics characteristics, typical of semi-arid regions, with many spiny plants, cacti, bromeliads, among others. The stand stratification consists of grasses, shrubs and trees varying from 3 to 7 meters high (IBAMA, 2009).

Most of the vegetal species of the Caatinga is caducipholy and looses the leaves during dry periods. However, this vegetation reacts promptly to precipitation, becoming green and robust (IBAMA, 2009; KAZMIERCZAK, 1996). These phenological changes are associated with changes on the amount and type of electromagnetic energy that is transmitted, reflected and absorbed by the vegetation (ROBERTS *et al*, 1998). As shown by Kazmierczak (1996), throughout the year the normalized difference vegetation index (NDVI) - an index that indicates the greenness of the vegetation - values of the Caatinga vegetation vary considerably.

Remote sensing techniques can provide a viable and non-destructive approach to infer vegetation changes influence on the spectral response and, consequently, on digital processing techniques results. Satellite image spectral classifiers access each pixel's radiometric information in order to recognize patterns and homogenous objects, and they can be used to map areas on the Earth surface which correspond to the themes of interest (INPE, 2006). Maximum Likelihood (ML) is an example of parametric classifier which can be used in this case. It calculates the probabilities of a given pixel belonging to each class determined on the training phase, further associating it to the class it is more likely to occur (JENSEN, 2009).

In most cases, fraction images obtained through linear spectral unmixing (LSU) can be used, in addition to the spectral bands of a sensor's image, as attributes when mapping vegetation (PONZONI e SHIMABUKURU, 2007). LSU technique presumes that the spectral response of each pixel, at any given spectral band, is a linear combination of the components of the scene, and is modeled as a sum of *n* pure components in the IFOV, averaged by their proportional area (SHIMABUKURU e SMITH, 1991). It provides a good contribution for classification because it minimizes the problem of mixed pixels, which is a consequence of establishing the mean reflectance of the targets on the Earth surface on each pixel area as its absolute value (CAMPBELL, 1996).

In this study we compared the regions of interest (ROIs) collected for a supervised Maximum Likelihood (ML) classification of a *Caatinga* region multi-temporal set of images in order to assess the phenological characteristics interference on differences in the spectral response and, consequently, classification of the vegetation. In addition to the spectral bands of the TM and ETM+ sensors (Landsat satelittes 5 and 7 respectively), we also used fraction images of these data obtained through LSU as additional attributes for the classification.

1.1 Study Area

The study area is located in Riachão do Jacuípe, a city along Jacuípe River in Bahia State. Its economy is based on cattle and agriculture, specially bovines, swines and extraction of *sisal* for exportation (Figure 1).

Figure 1. Location of Bahia state on the detail and location of Riachão do Jacuípe city on the main map.

During a field campaign, it was possible to notice that the image covers an area where three main *Caatinga* phytophysiognomies prevail: riparian - located along river margins; dense - on the north, with a continuous canopy; and open - on the south, constituted of isolated individuals and small patches of bare ground or rocks (Figure 2).
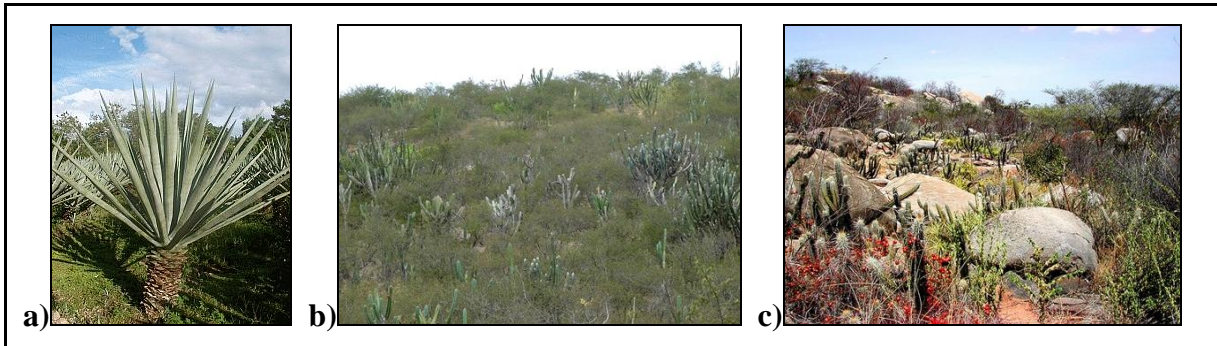
Figure 2. Photographs of: a *sisal* individual (a), dense *caatinga* (b) and open *caatinga* (c).

2. Methodology

We used two images of the sensor TM (onboard Landsat-5) and three of the sensor ETM+ (onboard Landsat-7) that had been previously pre-processed and converted to surface reflectance. The path of the images was 216 and row 68, and the time span of the data ranged from the year 1997 to 2008. The total precipitation information was obtained at CPTEC-INPE (2010) for the exact month when each image was collected. Because some images were acquired in the beginning of a month (e.g. 03/02/2008) and could have its spectral response influenced by the climate of the previous month, we got the total precipitation data of the previous months of each image as well (Table 1).

Table 1. Details of the images and precipitation data.

IMAGES		TOTAL PRECIPITATION (MM)	
DATE	SENSOR	ON THE MONTH	PREVIOUS MONTH
16/09/1997	TM	1-25	25-50
07/02/2001	ETM+	25-50	25-50
05/10/2001	ETM+	25-50	50-100
12/01/2003	ETM+	25-50	50-100
03/02/2008	TM	150-200	1-25

We used the software ENVI 4.7 to digitally process the images. First we collected endmembers of the following targets: *caatinga*, soil and water for each image, which were further used for the application of the LSU, with the mean square root estimator, and generate their respective fraction images.

Furthermore, we collected ten ROIs for each of the categories: (1) dense *caatinga*, (2) open *caatinga*, (2) riparian *caatinga*, (3) *sisal*, and (4) others. This last category grouped all the elements of the scene except for *caatinga* and *sisais*: like water, bare ground, urban areas and pastures. These ROIs where chosen on the image of the last date (03/02/2008) and represented regions where the targets had been the same over the time analyzed (e.g. each dense *caatinga* ROIs had been dense *caatinga* in all images, as each *sisal* ROI had been *sisal* since the first date either).

For each date, we used its six spectral bands and the 3 corresponding fraction images (*caatinga*, soil and water) and the ROIs (dense *caatinga*, open *caatinga*, riparian *caatinga*, *sisal*, and others) as subside for the supervised ML classification. Finally, we analyzed the ROIs on each image using confusion matrixes, Kappa and Z Statistics.

3. Results and Discussion

The fraction images subsided the classification with additional information making the resultant maps visually more accurate (Figures 3 and 4).
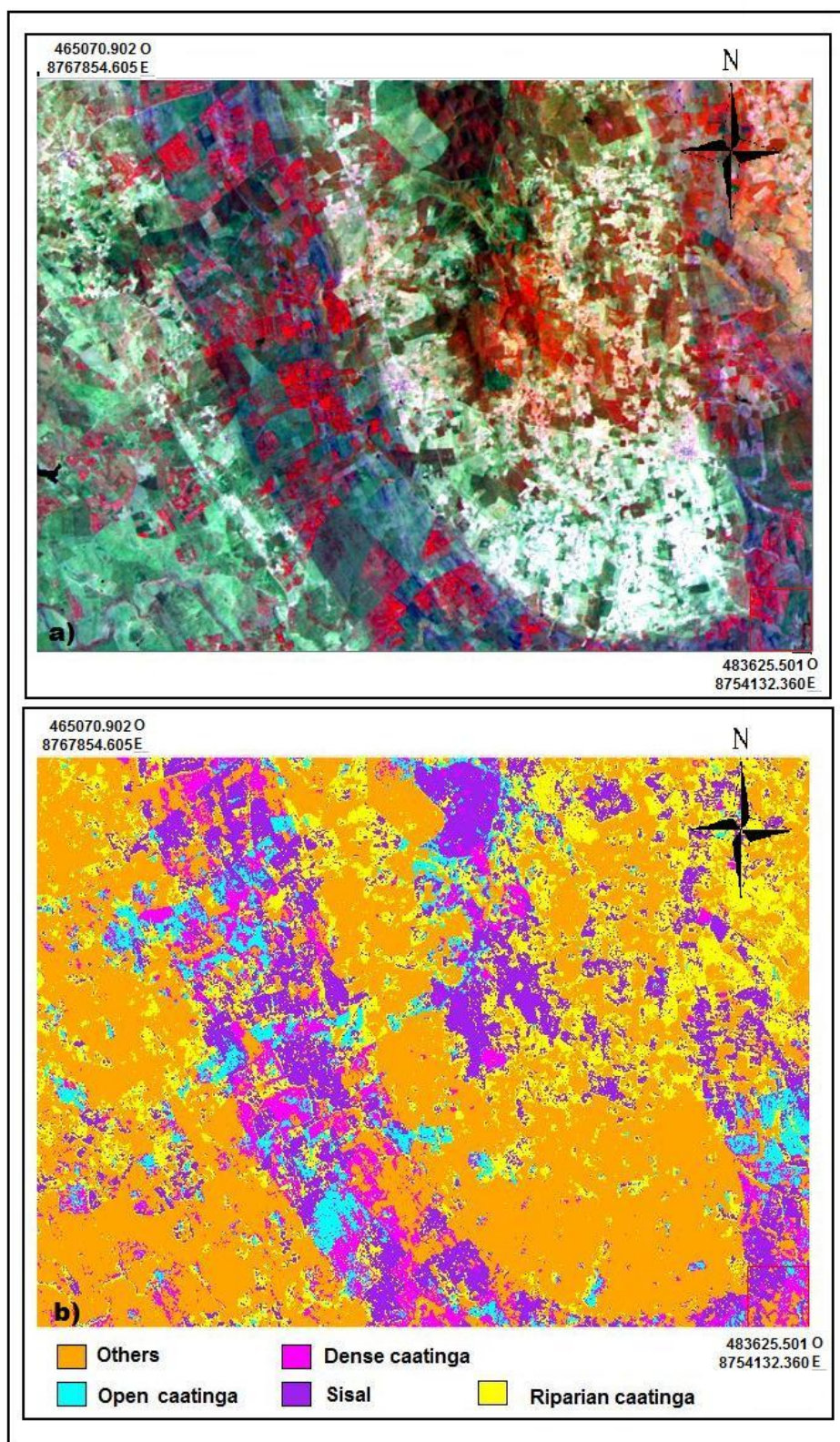
Figure 3.   Subsets of the 03/02/2008 Landsat image: B3(B), B4(R) and B5(G), and its resultant classification map.
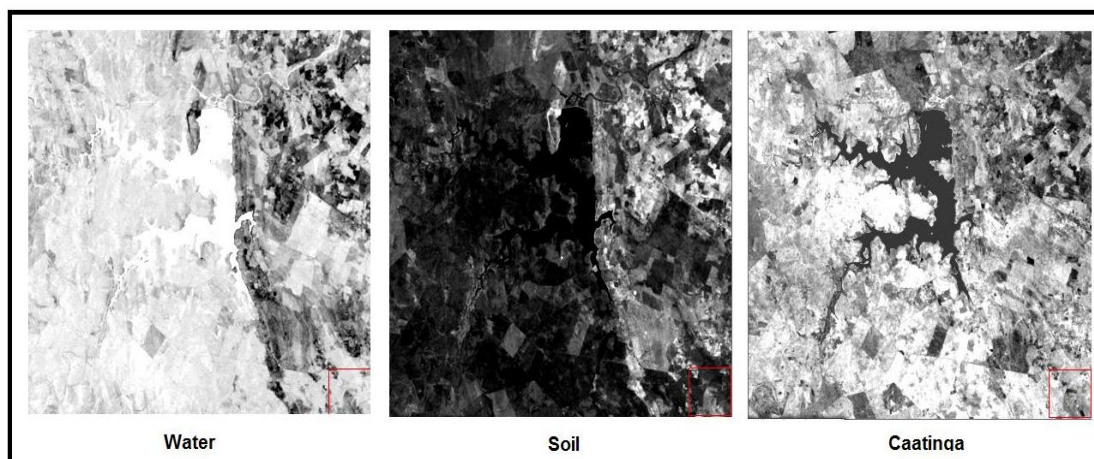
Figure 4.   Subsets of the 03/02/2008 Landsat fraction images obtained through Linear Spectral Unmixing.

The matrixes were organized for each image and showed the similarity of each group of ROIs for the same class with the ROIs of the other classes (Table 2). We could notice a high confusion between the classes Open and Dense *Caatinga*, compared to *Sisal* and Riparian *Caatinga*, which are very stable. Since the former ones have highly dynamic phenological characteristics, loosing the leaves during drier periods and becoming green and robust when there is precipitation (IBAMA, 2009), and the latter ones are not supposed to change much over the time, it's reasonable to infer that in this case the phenological factor influenced the spectral response of the *caatinga* phytophysiognomies.

Table 2. Similarity (%) of the ROIs of the classes of each image (on the lines) with the all the classes (on the columns).

			Open Caatinga	Dense Caatinga	Riparian Caatinga	Sisal	Others
MATRIX 1	16/09/1997	**Open Caatinga**	62,92	30,36	2,63	2,42	1,68
		Dense Caatinga	15,12	79,96	0,9	2,58	1,46
		Riparian Caatinga	5,71	5,71	44,76	39,05	4,76
		Sisal	3	12,1	4,26	79,15	1,5
		Others	2,37	0,51	2,2	0,14	94,78
MATRIX 2	07/02/2001	**Open Caatinga**	29,79	42,11	1,23	25,42	1,46
		Dense Caatinga	1,05	92,23	4,1	1,89	0,74
		Riparian Caatinga	0,95	13,33	50,48	26,67	8,57
		Sisal	1,73	19,24	4,72	72,12	2,19
		Others	0,07	3,46	3,51	0,44	92,52
MATRIX 3	05/10/2001	**Open Caatinga**	81,41	11,87	0,9	5,49	0,34
		Dense Caatinga	15,76	77,42	2,94	1,68	2,21
		Riparian Caatinga	0	19,05	43,31	30,48	6,67
		Sisal	12,1	3,23	3	78,11	3,57
		Others	1,42	1,48	4,73	1,37	91
MATRIX 4	12/01/2003	**Open Caatinga**	83,65	11,42	2,37	2,58	0
		Dense Caatinga	21,53	70,38	5,04	2,21	0,84
		Riparian Caatinga	1,9	14,29	60,95	10,48	12,38
		Sisal	19,12	17,74	4,84	54,38	3,92
		Others	0,35	0,08	3,66	0,34	95,56
MATRIX 5	03/02/2008	**Open Caatinga**	94,18	3,47	0,78	0,78	0,78
		Dense Caatinga	15,23	80,67	0,11	0,11	3,89
		Riparian Caatinga	2,86	0	83,81	13,33	0
		Sisal	0,58	0	10,02	89,29	0,12
		Others	0,71	4,65	1,12	0,02	93,5

According to Kazmierczak (1996), during the maximum photosynthetic activity period the *caatinga* phytophysionomies have NDVI values close to those of dense humid forests. Adversely, in rainy periods the small variation causes uniformity in the spectral responses and, consequently, complicates this discrimination. This results in high degrees of confusion between Open and Dense *Caatinga* on the matrixes. Controversially, the Sisal and Riparian *Caatinga* vegetations are stable throughout the year, throughout drier or rainier periods they don't loose leaves or change greenness, and such a characteristic can be noticed on low confusion between these classes on the matrixes.

In order to access the accuracy of the ROIs and determine statistically whether they where alike or different over the images we obtained Kappa values (HUDSON e RAMM, 1987) for each matrix and calculated the Z Statistics for each pair of Kappa. Through these procedures we could notice that matrix 5 had a higher Kappa value (0,8536) when comparing to the other images, and it's reasonable to say that this happens because this is the image from which we collected the ROIs used on the classification procedures (Table 3).

Table 3. Kappa values for each matrix shown on Table 2.

Matrizes	1	2	3	4	5
KAPPA	0,6539	0,5928	0,6789	0,6623	0,8536

Z Test calculations with 95% of confidence showed that most of the matrixes are different from each other, although the ROIs represent same targets on the scene (Table 4). As can be noticed on Table 1, some images were acquired after drier periods, as 1997, 02/2001 and 2008 images, and some after periods with more precipitation, as 10/2001 and 2003. This change in precipitation might have caused the vegetation to change its structure (falling or sprouting of leaves) and physiology, which affects the amount and type of radiation they absorb, transmit and reflect thus affecting the radiation captured by the sensors from these targets, and causes the classifications to differ from each other. This means that same targets were different throughout the years analyzed, which might be a consequence of phenological variations of the *Caatinga* ecosystems varying along the year.

Table 4. Results of the Z Test with 95% of confidence, where "different" means Kappa values are significantly different and "similar" means Kappa values are not significantly different.

	Z TEST			
MATRIXES	2	3	4	5
1	different	similar	similar	different
2	x	different	different	different
3	x	x	similar	different
4	x	x	x	different

4. Conclusion

The comparison of the regions of interest (ROIs) collected for a supervised Maximum Likelihood (ML) classification of a *Caatinga* region multi-temporal set of images enabled us to conclude that the highly dynamic phenological characteristics of the *Caatinga* phytophysionomies resulted in variations of the radiation they reflected thought the time analyzed and, consequently, interfere on the classification of the vegetation. The use of linear spectral umixing (LSU) fraction images as additional attributes for the classification made the resultant maps more accurate.

References

CAMPBELL, J. B. **Introduction to remote sensing.** New York: The Guilford Press, 1996. 622p.

CPTEC- INPE. Centro de Previsão de Tempo e Estudos Climáticos. **Previsão Climática.** 2010. Available at: http://clima1.cptec.inpe.br/~rclima1/monitoramento_brasil.shtml. Visited on: 01/11/2010.

HUDSON, W.D.; RAMM C.W. Correct formulation of the Kappa coefficient of agreement, **Photogrammetric Engineering**. Remote Sensing, v.6, n. 53, p. 421-422, 1987.

IBAMA. **Ecossistemas brasileiros.** 2009. Available at: http://www.ibama.gov.br/ecossistemas/caatinga.htm. Visited on: 01/11/2010.

INPE. **Tutorial de Geoprocessamento.** 2006. Available at: http://www.dpi.inpe.br/spring/portugues/tutorial/classific.html. Visited on: 01/11/2010.

JENSEN, J. R. **Introductory Digital Image Processing: a Remote Sensing Perspective.** Prentice-Hall, New Jersey, 2009, 379 p.

KAZMIERCZAK, M. L.; "Uso de Dados AVHRR/NOAA GAC para Análise da Sazonalidade Da Caatinga". **Anais VIII Simpósio Brasileiro de Sensoriamento Remoto**, Salvador, Brasil, INPE.   14-19, p. 513-518, 1996.

PONZONI, F. J.; SHIMABUKURU Y. E. **Sensoriamento remoto no estudo da vegetação.** São Jose dos Campos, SP: Parêntese, 2007 135 ISBN 978-85-60507-02-3.

RESERVA DA BIOSFERA DA CAATINGA. **O Bioma Caatinga.** 2009. Available at: http://www.biosferadacaatinga.org.br/o_bioma_caatinga.html. Visited on: 01/11/2010.

ROBERTS, D.A.; NELSON, B.W.; ADAMS, J.B.; PALMER F. Spectral changes with leaf aging in Amazon caatinga". **Trees.** V. 12., p. 315–325, 1998.

SHIMABUKURU, Y. E.; SMITH, J. A. "The least-squares mixing models to generate fraction images derived from remote sensing multispectral data". **IEEE Transactions on Geoscience and Remote Sensing**, v. 29, p. 16−20, 1991.