

USE OF DATA MINING AND SPECTRAL PROFILES TO DIFFERENTIATE CONDITION AFTER HARVEST OF COFFEE PLANTS

RUBENS A. C. LAMPARELLI¹, JERRY A. JOHANN², ÉDER R. DOS SANTOS³,
JULIO C. D. M. ESQUERDO⁴, JANSLE V. ROCHA⁵

ABSTRACT: This study aimed at identifying different conditions of coffee plants after harvesting period, using data mining and spectral behavior profiles from Hyperion/EO1 sensor. The Hyperion image, with spatial resolution of 30 m, was acquired in August 28th, 2008, at the end of the coffee harvest season in the studied area. For pre-processing imaging, atmospheric and signal/noise effect corrections were carried out using Flaash and MNF (Minimum Noise Fraction Transform) algorithms, respectively. Spectral behavior profiles (38) of different coffee varieties were generated from 150 Hyperion bands. The spectral behavior profiles were analyzed by Expectation-Maximization (EM) algorithm considering 2; 3; 4 and 5 clusters. T-test with 5% of significance was used to verify the similarity among the wavelength cluster means. The results demonstrated that it is possible to separate five different clusters, which were comprised by different coffee crop conditions making possible to improve future intervention actions.

KEYWORDS: crop monitoring, spectral behavior, management, orbital remote sensing.

DISCRIMINAÇÃO DE DIFERENTES ESTADOS DE PLANTIOS DE CAFÉ PÓS-COLHEITA, POR MEIO DA TÉCNICA DE MINERAÇÃO DE DADOS E PERFIS ESPECTRAIS

RESUMO: Este trabalho teve o objetivo de identificar diferentes condições na cultura do café, após a colheita, utilizando mineração de dados e curvas espectrais obtidas do sensor Hyperion/EO1. A imagem Hyperion, com resolução espacial de 30 m, foi obtida em 28 de agosto de 2008, fim de safra do café na área de estudo. Como pré-processamentos, foram realizadas a correção atmosférica através do algoritmo Flaash e a correção do sinal/ruído pelo algoritmo MNF (*Minimum Noise Fraction Transform*). Posteriormente, foram geradas curvas espectrais (38) de 150 comprimentos de onda, que foram analisadas através do algoritmo EM (*Expectation-Maximization*), considerando simulações de 2; 3; 4 e 5 *clusters*. Para verificar a igualdade das médias dos comprimentos de onda dos clusters, foi utilizado o teste *t*, com 5% de significância. Os resultados mostraram que foi possível separar os cafés em cinco diferentes grupos, segundo sua condição pós-colheita, possibilitando melhorar ações futuras de intervenção.

PALAVRAS-CHAVE: monitoramento de cultura, comportamento espectral, manejo, sensoriamento remoto.

¹ Eng^o Agrícola, Pesquisador, CEPAGRI/UNICAMP, Av. André Tosello, 209, Cidade Universitária, Campinas - SP, rubens@cpa.unicamp.br.

² Eng^o Agrícola, Doutorando, FEAGRI/UNICAMP, Av. Candido Rondon, 501, Cidade Universitária, Campinas - SP, jerry.johann@feagri.unicamp.br.

³ Eng^o Agrônomo, Coordenador de Geoprocessamento, Cooxupé, Av. Dr. Antônio Pereira Lima, 762, Monte Santo de Minas - MG, edersantos@cooxupe.com.br.

⁴ Eng^o Agrônomo, Pesquisador, CNPTIA/EMBRAPA, Av. André Tosello, 209, Cidade Universitária, Campinas - SP, julio@cnptia.embrapa.br.

⁵ Eng^o Agrícola, Professor, FEAGRI/UNICAMP, Av. Candido Rondon, 501, Cidade Universitária, Campinas - SP, jansle.rocha@feagri.unicamp.br.

Recebido pelo Conselho Editorial em: 25-2-2011

Aprovado pelo Conselho Editorial em: 3-11-2011

INTRODUCTION

Spectral signatures describe the variation of the electromagnetic energy reflected by the targets along the electromagnetic spectrum, which behavior depends on optical characteristics that determine its capacity for absorption, transmission and reflection. In the case of plants, this spectral behavior is strongly influenced by the presence of leaf pigments, cell structure and presence of water in leaf tissues (GATES et al., 1965).

Several studies describe the spectral behavior typical of some plant species and the phenomena that cause changing in this pattern (JENSEN, 2009). In general, these studies were conducted in the laboratory with equipment called spectroradiometer capable of recording the energy reflected by the targets in hundreds of different wavelengths, resulting in very high resolution spectral signatures. Other studies have been conducted with multispectral satellite images to determine spectral signatures of plant populations. However, due to the low resolution of these sensors, the spectral curves generated have low level of details, reducing the capacity to assess the spectral characteristics of the targets under analysis.

Many studies show that certain wavelengths interact with plant pigments. The interaction in the visible wavelength, for example, is influenced by chlorophyll pigments (DENISE & BARANOSKI, 2007). Others have shown the relationship between pigment concentration and the optical properties of the leaf (PONZONI & SHIMABUKURO, 2007; PEÑA-BARRAGÁN et al., 2011). Most of these studies used data from satellites with sensors that operate in wide bands of the electromagnetic spectrum (MOREIRA et al., 2007; MOREIRA et al., 2010). However, with the improvement of sensors that record more detailed information from the surface, there is the possibility of conducting new studies involving the analysis of the relationship between plant components and their spectral response. Such studies rely on the fact that there is an innovation of orbital characteristics of the spatial, temporal, spectral, and radiometric sensor resolution.

Hyperspectral imager sensors have the advantage of instantly composing images acquired in hundreds of spectra with a level of resolution closer to that obtained with field or laboratory spectroradiometer (RUDORFF et al., 2007). The Hyperion, which was launched aboard the Earth Observing One satellite (EO-1) in November 2000, is the first orbital hyperspectral sensor that allows the acquisition of images in 220 spectral bands (10 nm wide each), with a spatial resolution of 30 m. Covering the spectrum from 400 nm to 2,500 nm, the bands are positioned in the visible, near infrared and short wave infrared, allowing more precise analyses of the relationships between plant components and spectral bands.

Remote sensing applications in agriculture are related, for example, to monitoring the coverage, vigor and type of existing vegetation. However, it is necessary to know the spectral behavior of these surfaces. Moreover, the same crop may have variable spectral behavior at different stages of development (PONZONI & SHIMABUKURO, 2007).

The evaluation of the spectral profile can allow not only crop differentiation (TISOT et al., 2007), but also bring inferences about their conditions. At this point, the analysis of spectral profiles can identify differences in plants and provide important insights on the spatial distribution of post-harvest damage, one of the critical steps of crop management.

In this sense, data mining (DM) is placed as a tool to analyze large volumes of data. According to FAYYAD et al. (1996), DM can be defined as the extraction of knowledge from a database, by identifying standards that are valid, original, potentially useful and understandable.

For LAXMAN & SASTRY (2006) and MILLER & HAN (2009), the DM process is focused on the interaction between the various classes of users (domain expert, analyst, end user) and involves knowledge of the domain, problem identification, pre-processing, pattern extraction, post-processing, and use of the knowledge gained. During pre-processing, domain knowledge, and problem identification help select the data sets, to which treatment methods are applied, such as extraction, integration, processing, cleaning, selection of attributes, and data reduction, so that the

goals are achieved during the phase of standard extraction. In this phase, the choice of MD task, algorithm, and extraction of pattern to be used are defined. In the task choice, one must decide between a descriptive (association rules, summarization, clustering or grouping) and predictive (classification, regression) activity, according to the desired objectives and then define the algorithm to be used for this task. Finally, in post-processing, after selection of the most important or relevant patterns, the gained knowledge should be used to solve the identified problem.

DM predictions involve the use of attributes of a data set to predict the future value of the variable target, i.e., aiming to decision-making process. Cluster generation is a descriptive task that seeks to target a data set in a number of classes which the intra-class similarities and between classes are respectively minimized and maximized (MILLER & HAN, 2009).

RIE & OSAMU (2001) emphasized the importance of discovering new knowledge from large amounts of data, such as those derived from meteorological satellite images. These authors addressed the extraction of information on long time series of cloud images, which were analyzed in the form of clusters (groups) identification. Clusters were entered into a relational database that allowed queries to evaluate its usefulness. In the same line and corroborating the results found by RIE & OSAMU (2001), ZHANG et al. (2008) described almost the same procedures for analyzing time series of meteorological satellite images using DM to improve weather forecasting.

Among the main clustering methods, it is highlighted the partition, the density and the probabilistic (GUIDINI & RIBEIRO, 2006). In the partitioning method, the algorithm that is the most widely used is the *K-Means*, which identifies classes of objects with similar characteristics that are closest to a given centroid, often determined by the Euclidean distance or Manhattan distance metrics. However, the number of clusters must be defined in advance by the analyst, who chooses the best set of clusters after the event, which is a disadvantage of the method. Furthermore, this method is sensitive to noise or outliers in the data set.

Among the probabilistic methods, the EM (Expectation-Maximization) algorithm, also known as Gaussian Mixture, is more widely used in data clustering. It is based on the maximum likelihood statistics to estimate the parameters of normal distribution. The data are a mixture of n univariate normal distributions with the same variance σ^2 and estimate the mean of each normal distribution, i.e., the hypothesis that maximizes the likelihood of these means and, through an iterative process, to form clusters.

In this sense, considering the potential of DM tools, the objective of this study was to group profiles of hyperspectral sensor Hyperion/EO1 under several conditions during post-harvest phase of coffee plants in order to classify them according to their conditions.

MATERIAL AND METHODS

The Hyperion image was collected on August 28, 2008 in 220 spectral bands, each 10 nm wide, covering the wavelengths between 400 nm and 2,500 nm. However, only 198 tracks are provided with radiometric calibration. The spatial resolution of the sensor is 30 meters.

The images were preprocessed using the MNF (Minimum Noise Fraction Transform) technique, which made it possible to eliminate bands with noise. Afterwards, the atmospheric correction of the image was performed. For this, was applied the Flaash algorithm that transforms gray level values to radiance, using scale factors, and later to the surface reflectance, which were established within the following parameters for the correction: spectral model of tropical type, aerosol model of rural type, option for "water retrieval" (estimative of water amount) and 1,135 nm value in the "water absorption" feature (electromagnetic spectrum characterized by absorption of water). After the pre-processing activities, only 150 of 220 bands were effectively used.

To perform the task, it was built a database with information on the plots studied, such as plant variety, age, area, slope and spacing. The database was structured in a Geographic Information System (GIS).

The area under study is located on a farm in the municipality of Montes Claros, in the south of the state of Minas Gerais, as shown in the map in Figure 1. The region lacks a detailed survey of the soil classes, although there is predominance of more sandy types (<15% clay). The figure also illustrates the location of sampling points used to generate the spectral curves.

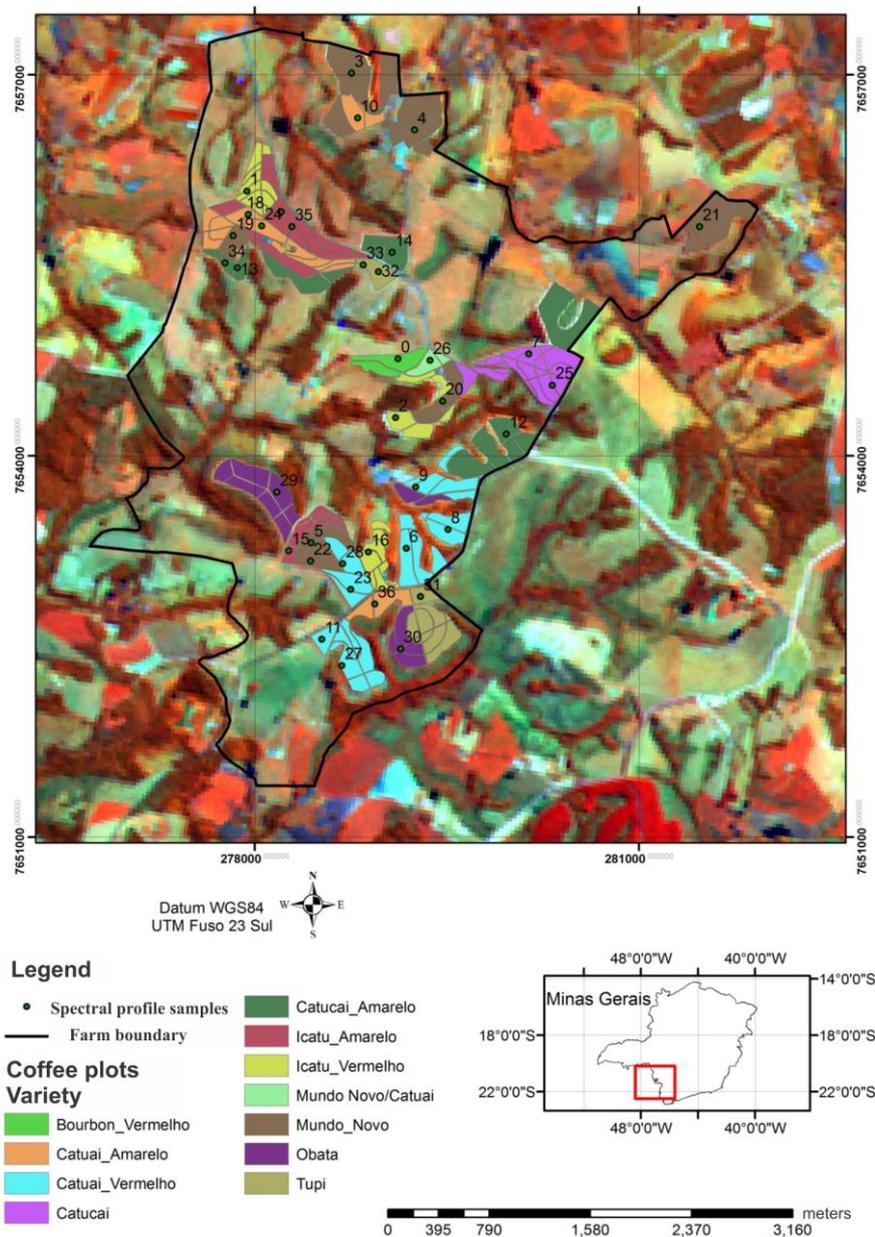


FIGURE 1. Study area showing plots and locations of sampled spectral profiles.

The farm is located on a plateau between 1,000 and 1,110 meters of altitude. Since eight varieties of coffee (Catuaí Amarelo and Vermelho, Catucaí Amarelo, Icatú Amarelo and Vermelho, Mundo Novo, Obatã andTupi) are found in the study area, and this simplification was considered in the analysis. First, all varieties have approximately the same spacing (1 x 4 m) and number of plants per hectare (3,200); the analysis was conducted considering the spectral image for crops or sections with more foliage, and another for crops or sections with less foliage. In case there was no way to identify differences in the image, the plot was considered homogeneous and a single sample was collected. It must consider that the age of the plants, which varies widely in the areas studied (2 to 10 years), influenced the spectral response. Figure 2 shows these variations in the image (a) and field (b), and the differentiation was carried out visually, by analyzing the Hyperion image in color composition (R 833 nm - 1,215 nm G - B 2,304 nm).

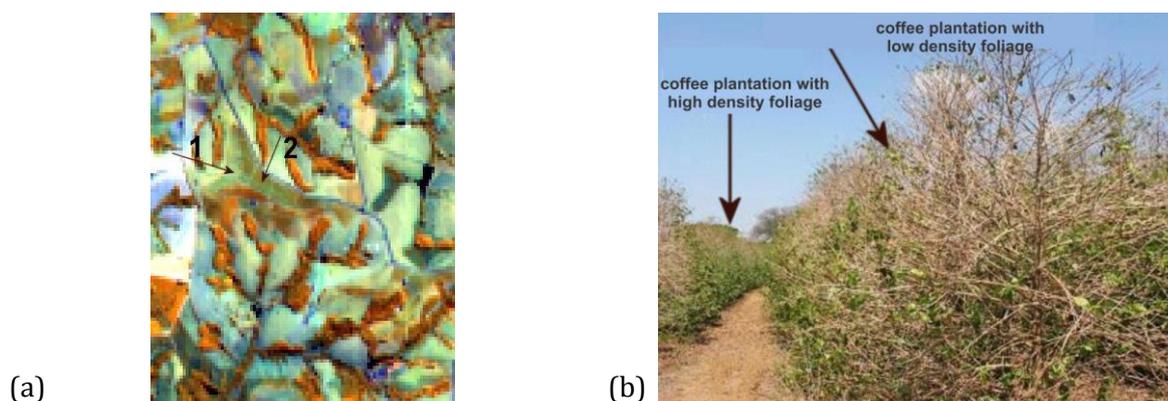


FIGURE 2. Color composite of a Hyperion image from August 28th, 2008 (a) and panoramic field photo (b) highlighting different coffee foliage conditions in the same parcel (Ex. parcel 211). 1 = coffee plantation with low density foliage; 2 = coffee plantation with high density foliage.

Table 1 presents a description of some after-harvest characteristics of coffee plants that were considered for cluster analysis. It should be noted that the descriptive characteristics were indicative of the condition of the plant, and the information was the result of harvest (management), which varied from plot to plot, depending on the type of harvesting, mechanized or manual, and in more detail, manpower ability or machine settings. This assessment was carried out in the field and during technical meetings with the producer and technicians at the Regional Cooperative of Coffee Growers in Coaxupé (“Cooperativa Regional de Cafeicultores de Guaxupé Ltda.”)

TABLE 1. Plots and its characteristics.

Plots sample number	Variety	Age (years)	Height (m)	Prod. (sc ha ⁻¹)	Description
329_0	Bourbon	4	1.8	56.8 ^(a)	By the image had little leaf.
212_1	Icatu Vermelho	8	1.8	47.8 ^(m)	Less leaf
372_2	Icatu Vermelho	3	1.8	61.5 ^(a)	Seated in a sandy soil with irrigation finalized in May, harvest in June, image in August. Malnourished plant with no rain in the period. 100% harvested.
428_3	Mundo Novo	3	1.8	8.1 ^(b)	Seated in a sandy soil with irrigation finalized in May, harvest in June, image in August. Malnourished plant with no rain in the period. 100% harvested.
458_4	Mundo Novo	2	-	-	No production. Practically soil.
1381_5	Mundo Novo	13	4.0	97.5 ^(a)	Tip cutting. No rain.
164_6	Catuaí Vermelho	12	2.3 a 2.4	90.0 ^(a)	Tip cutting. By the image had little leaf.
315_7	Catuaí	9	2.1 a 2.2	81.1 ^(a)	By the image had little leaf. 100% harvested.
354_8	Catuaí Vermelho	5	1.8 a 1.9	73.7 ^(a)	Low height. Relatively new crop. 100% harvested.
364_9	Catuaí Vermelho	7	1.8 a 2.0	63.7 ^(a)	By the image had little leaf. 100% harvested.
443_10	Catuaí Amarelo	2	-	-	Soil 100%.
1104_11	Catuaí Vermelho	10	-	60.5 ^(a)	Less leaves. 77% harvested.
355_12	Catuaí	5	1.8	66.6 ^(a)	Usually lose leaves after harvest. 100% harvested.
2651_13	Catuaí	4	1.7	82.8 ^(a)	Less leaves. Stain with less leaves. 0% harvested.
2652_14	Catuaí	4	1.7	81.7 ^(a)	Less leaves. Stain with less leaves. 0% harvested.

Plots sample number	Variety	Age (years)	Height (m)	Prod. (sc ha ⁻¹)	Description
121_15	Icatu Amarelo	7	3.0	72.9 ^(a)	Soil completely closed. 100% harvested.
152_16	Icatu Vermelho	8	3.5	61.7 ^(a)	96% harvested.
211_17	Icatu Amarelo	8	-	77.9 ^(a)	More leaves. 80% harvested.
212_18	Icatu Vermelho	8	-	47.8 ^(m)	More leaves. 70% harvested.
248_19	Mundo Novo	13	-	22.9 ^(b)	22% harvested.
338_20	Mundo Novo	9	-	80.6 ^(a)	45% harvested.
468_21	Mundo Novo	14	3.4	102.9 ^(a)	More leaves by the image. 100% harvested.
1381_22	Mundo Novo	3	-	97.5 ^(a)	More leaves. 98% harvested.
144_23	Catuaí Vermelho	9	2.0 a 2.4	29.7 ^(b)	100% harvested.
233_24	Catuaí Amarelo/Vermelho	11	-	91.0 ^(a)	More leaves. 5% harvested.
315_25	Catuaí	9	-	81.3 ^(a)	More leaves. 100% harvested.
348_26	Catuaí Mundo Novo	11	-	60.0 ^(a)	More leaves. 100% harvested.
1104_27	Catuaí Vermelho	10	-	60.3 ^(a)	More leaves. 77% harvested.
1382_28	Catuaí Vermelho	13	2.0 a 2.4	73.8 ^(a)	With leaves. 100% harvested.
116_29	Obatã	6	2.0 a 2.3	61.4 ^(a)	With leaves. 35% harvested.
196_30	Obatã	10	-	63.9 ^(a)	With leaves. 0% harvested.
177_31	Tupi	10	-	52.6 ^(a)	More leaves. 100% harvested.
257_32	Tupi	6	-	78.9 ^(a)	More leaves. 100% harvested.
257_33	Tupi	6	-	78.9 ^(a)	Less leaves. 100% harvested.
2651_34	Catucui Amarelo	4	-	82.8 ^(a)	0% harvested.
211_35	Icatu Amarelo	8	-	77.9 ^(a)	Less leaves. 80% harvested.
183_36	Catuaí Amarelo	10	2.0 a 2.4	52.6 ^(a)	100% soil.

Note: in the first column Plots sample numbers, the first number is the plot and the second is the place where the sample was removed for generating the spectral profile: (a) - high; (m) - medium and (b) - low productivity.

For cluster generation, based on behavior of the spectral profile of each plot, the software WEKA (Waikato Environment for Knowledge Analysis) was used (WITTEN&FRANK, 2005). This software aggregates algorithms from different paradigm methods, and carries out statistical and computational analyses of the data provided by using data mining techniques in order to acquire new knowledge, either inductively or deductively. Thus, for the generation of clusters, it was used the EM algorithm, which allowed more user interaction.

Aiming at differentiating coffee crop conditions after harvest, a spectral profile for each plot was generated by selecting a point close to the center of the field in order to avoid the influence of adjacent targets. To group the 37 plot profiles according to the spectral profile, simulations were performed considering two, three, four and five clusters.

Student's t-test with a significance level of 5% was applied to check the equality of the means of wavelengths in all groups performed.

RESULTS AND DISCUSSION

Cluster categorization is relative, that is, for every set number of classes, grading was performed using spectral curves as basis. The worst the coffee in terms of structure and biomass of the plant, the more the spectral curve tends to approximate to the soil relative curve, otherwise the spectral curve approximated to the vegetation curve pattern. Therefore, categorization changes from grouping to grouping. Because of this, categories were created for each cluster division, as shown in Table 2.

TABLE 2. Spectral profiles characterization by clustering.

Plots / sample number	2 clusters division: C1>C0 C0 – worse C1 - better	3 clusters division: C2>C0>C1 C1 – worse C2 - better	4 clusters division: C2>C1>C3>C0 C0 – worse C2 -better	5 clusters division: C4>C0>C2>C3>C1 C1 – worse C4 - better
329_0	C0	C0	C3	C3
212_1 less leaves	C0	C0	C3	C3
372_2	C0	C0	C3	C3
428_3	C0	C0	C3	C3
458_4	C0	C1	C0	C1
1381_5	C0	C0	C3	C3
164_6	C0	C0	C3	C3
315_7	C0	C1	C0	C1
354_8	C0	C1	C0	C1
364_9	C0	C0	C0	C3
443_10	C0	C0	C0	C3
1104_11 less leaves	C0	C0	C3	C3
355_12	C0	C1	C0	C1
2651_13 less leaves	C0	C2	C3	C3
2652_14	C0	C2	C3	C3
121_15	C1	C2	C1	C2
152_16	C1	C2	C2	C0
211_17 more leaves	C1	C2	C1	C0
212_18 more leaves	C1	C2	C2	C0
248_19	C1	C2	C1	C0
338_20	C1	C2	C1	C0
468_21	C1	C2	C2	C0
1381_22 more leaves	C1	C2	C2	C0
144_23	C1	C2	C1	C0
233_24	C1	C2	C2	C4
315_25	C1	C2	C2	C4
348_26	C1	C2	C1	C0
1104_27 more leaves	C1	C2	C2	C4
1382_28	C1	C2	C1	C0
116_29	C1	C2	C1	C2
196_30	C1	C2	C2	C4
177_31	C1	C2	C1	C2
257_32 more leaves	C1	C2	C2	C0
257_33 less leaves	C1	C2	C1	C2
2651_34	C1	C2	C2	C4
211_35 less leaves	C1	C2	C1	C2
183_36	C1	C2	C1	C0

Note: C0 - cluster0; C1 - cluster1; C2 - cluster2; C3 - cluster3; C4 - cluster4; C5 - cluster5

Figure 3 shows spectral profiles of the division into two clusters (C0 and C1). It can be clearly seen that the spectral behavior placed into cluster zero (C0), which describes the worst coffee crop condition, has fundamental differences from those presented in Cluster 1 (C1), which describes the best condition. The differentiation occurs in the visible range (400-720 nm) where the peaks of radiation for blue (440-485 nm), green (500-565 nm) and red (625-740 nm) are more pronounced in C1, as did the water absorption peaks that are more pronounced than those observed in the spectral curves for the coffee plants with more leaves. Figure 3c shows more clearly this condition through the mean curves and standard deviations for each of the clusters generated. In general, the standard deviation of the spectral profile of the cluster 0 (worst crop) was greater than that of Cluster 1 (best crop), showing the worse condition of cluster 0 in terms of variability, which was expected. The application of Student's t-test showed that there was a 5% difference between the average spectral profiles (compared to all wavelengths) in cluster 0 (15 varieties) and cluster 1 (22 varieties). In order to verify that the variances between these two clusters were significant at 5%, the Snedecor's

F hypothesis test was also applied for comparison of variances, and found that there was significant difference between the average spectral profile of the cluster 0 and 1 in the following wavelengths (671-742 nm; 905-932 nm; 1,305-1,336 nm and 1,749-1,780 nm), as identified in yellow in Figure 3.

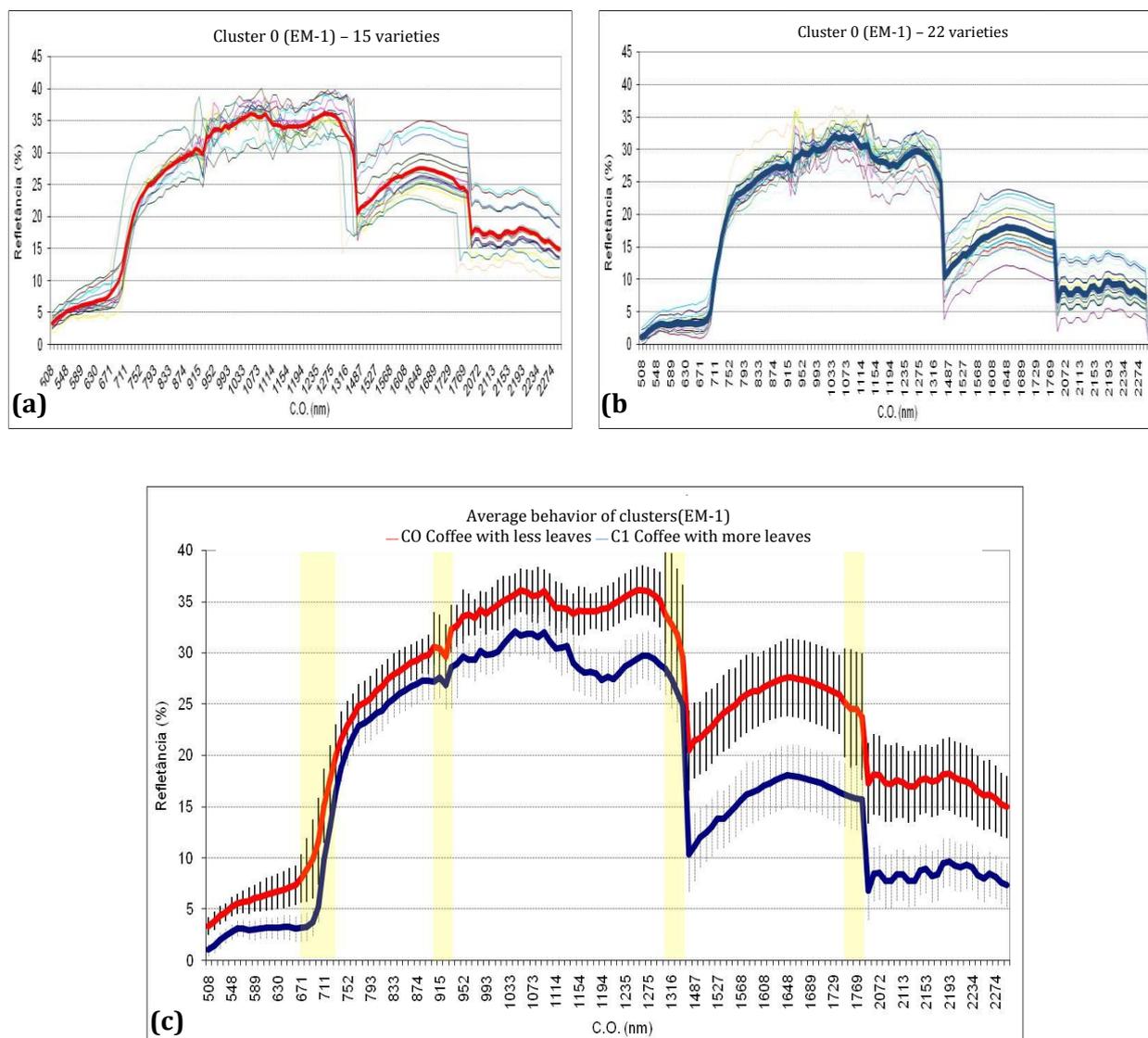


FIGURE 3. Grouping of parcels spectral profile in 2 clusters, a) C0 (worse), b) C1 (better) and c) mean and standard deviation of each cluster.

To the extent that defined most of the clusters (3; 4 and 5), there was a new subdivision of the spectral profiles between those for the best and worst coffee. This can be seen in Figures 4a, 4b and 4c with the average curves of these new groupings. It can be observed in all three cases, that the greatest differences occurred at the mid-infrared wavelength (>1,300 nm) in water absorption peaks. This was because the coffee fields in this case are all irrigated, and irrigation was terminated in May, before the harvest. The curves showed the plant conditions not only with regard to its structure and amount of photosynthetic pigments, but also in relation to its water content. Thus, the most disturbed plants are those with less intense water absorption peaks. Since these plants are in sandy soils following the withdrawal of irrigation, the results suggest that the greater the water absorption peaks, the better the conditions of plant hydration.

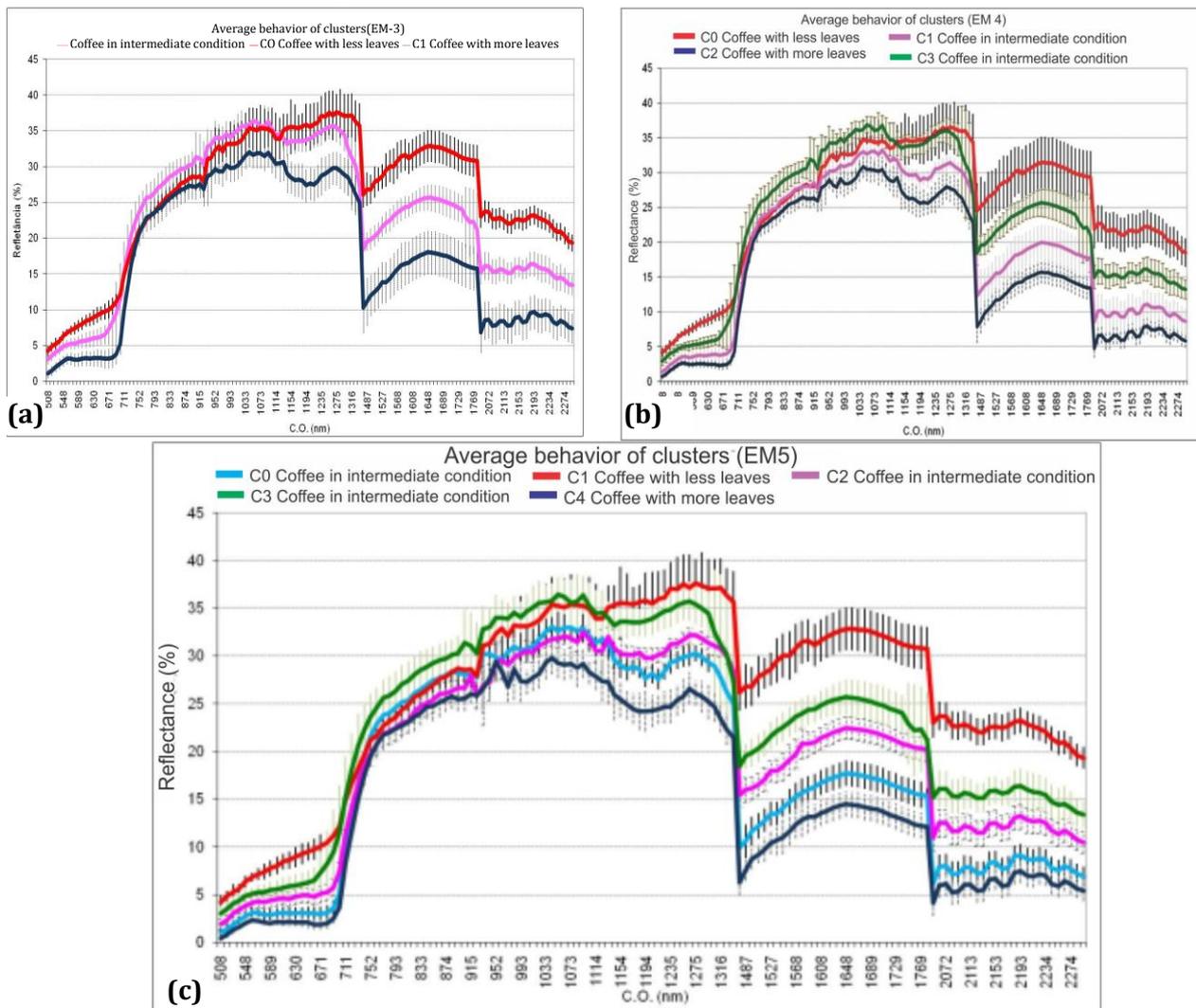


FIGURE 4. Grouping of mean spectral profile of the parcels in: a) 3 clusters, 4 clusters, 5 clusters.

Figure 5 (a, b, c) shows the division of the various spectral curves into three classes, where Cluster 0 represents spectral profiles of blocks with coffee in an intermediate situation, the coffee plants in Cluster 1 was worse and coffee plants in the Cluster 2 were in better condition. It may be noticed in C1 (Figure 5b) the spectral profile in blue (indicated by an arrow), which corresponds to a coffee plot of two years old with the Mundo Novo variety, indicating that the spectral curve refers to almost uncovered soil. The other curves within this class denote, from the spectral behavior, that the plots are under the same conditions, i.e., with a high degree of soil exposure. On the other hand, the curves shown in Cluster 2 (Figure 5c) are from plots where there coffee plants were better, which is explained well by the peaks of absorption and reflection of the visible, near infrared and mid-infrared.

The same happens when the varieties are grouped into clusters 4 and 5. To confirm this, it was followed each stand in relation to its position in the division by 2; 3; 4 and 5 clusters (Table 2). It can illustrate this analysis by monitoring of field 458 (two year-old plot with Novo Mundo variety), which in all divisions performed it is found in the vegetation behavior in the worst condition. The other extreme can be illustrated by the data of field 233 (Table 1), representing an eleven-year old plot with the Catuaí variety, which only 5% of its area had been harvested, therefore, presenting a better condition, with plants well covered with leaves. In this case, it always remained in the cluster with best plant conditions in all divisions performed, that is, C1 (2 classes), C2 (3 classes), C2 (4 classes) and C4 (5 classes).

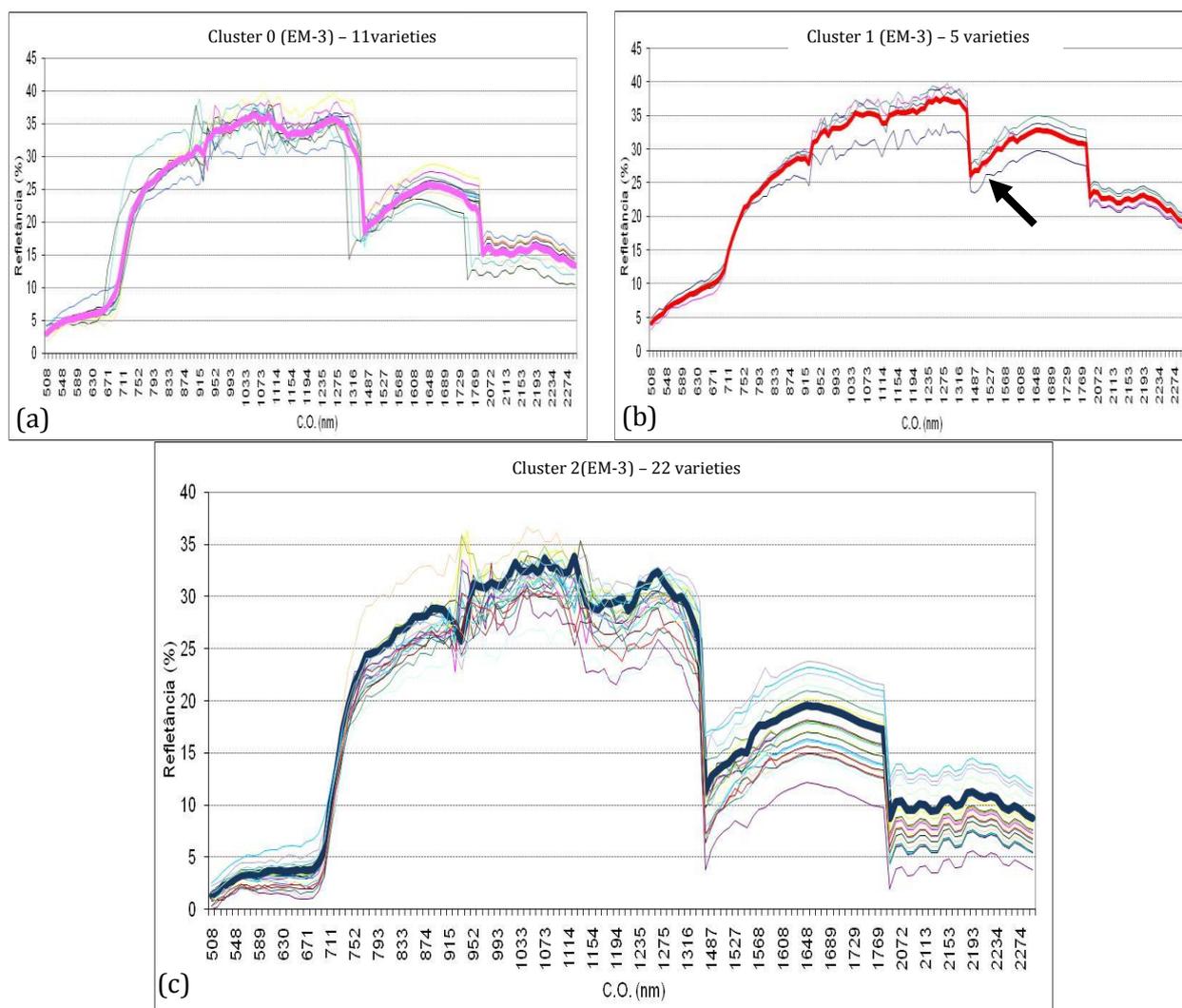
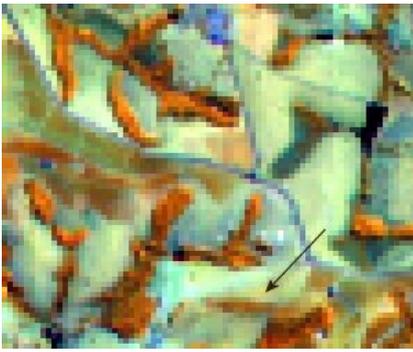


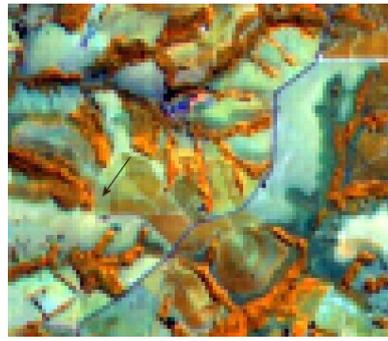
FIGURE 5. Grouping of spectral profile of parcels in 3 clusters, a) C0 (intermediate condition), b) C1 (less leaves) and c) C2 (more leaves).

To illustrate and corroborate the reasoning of analysis of the conditions of coffee plants, for each plot, the following two examples are shown: one with plot 329 and another with plot 121. Plot 329 consists of a 4 year Bourbon variety crop, that is tall, approximate height of 1.80 m, good foliage, high productivity ($56.75 \text{ bags ha}^{-1}$) and low biomass, as seen in the scene of the satellite Hyperion (Table 2). In this case, it was entered in C0 (worse) when the curves were divided into two classes. Later, with the division of three classes, it took the position C0 (intermediate). In the division into four classes, it took the position C3 (intermediate for worse). When considering the division into five classes, this plot was considered also in Class C3 (intermediate for worse) (Figure 6a).

On the other hand, plot 121 is a seven year old Icatu variety crop, 100% harvested, of approximate 3 meters height, with high percentage of ground coverage, and high yield ($72.85 \text{ bags ha}^{-1}$) (Table 2 and Figure 6b). This analysis was performed for all plots (Table 2) and found a strong relationship between plant condition, spectral response and the grouping of the plots. Thus, the DM technique was able to identify these variations, as can be seen on the spatial distribution of plots in each simulation (two, three, four and five clusters) (Figure 7 a, b, c, d). In these spatializations, spectral behavior was not considered in the plots with sections with less foliage and another section with more foliage.

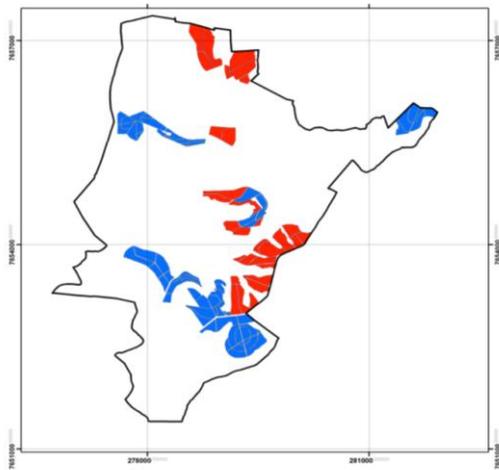


a) Plot 329



b) Plot 121

FIGURE 6. Images highlighting different coffee foliage conditions.



Legend

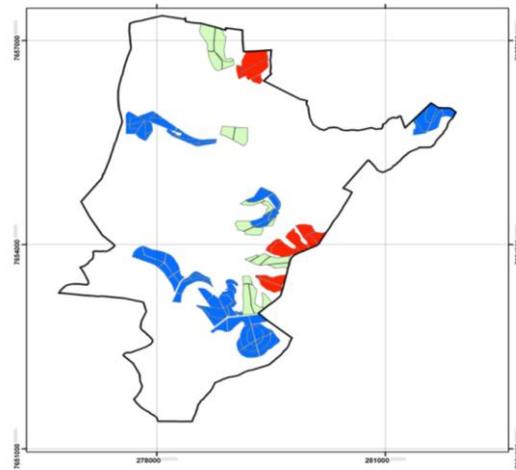
- C1_2 clusters (better)
- C0_2 clusters (worst)

(a)

Datum WGS84
UTM Fuso 23 Sul



0 400 800 1,600 2,400 3,200 meters



Legend

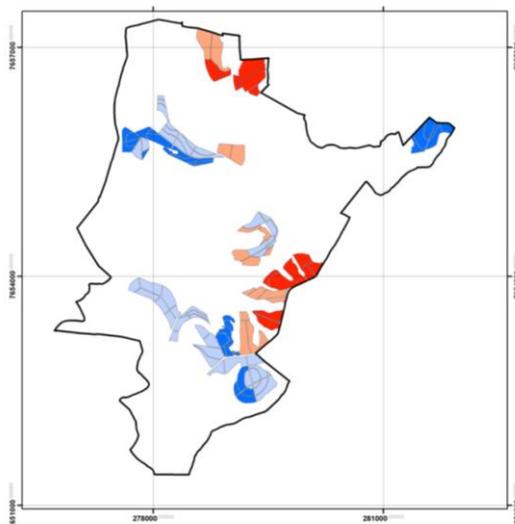
- farm boundary
- C2_3 clusters (better)
- C1_3 clusters (worst)
- C0_3 clusters (intermediate)

(b)

Datum WGS84
UTM Fuso 23 Sul



0 400 800 1,600 2,400 3,200 meters



Legend

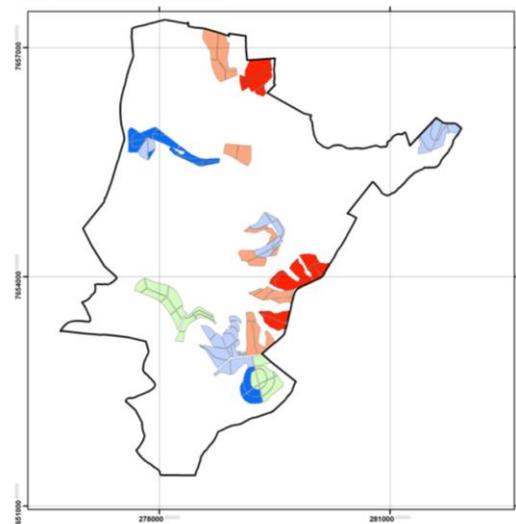
- farm boundary
- C3_4 clusters (interm. worse)
- C2_4 clusters (better)
- C1_4 clusters (interm. better)
- C0_4 clusters (worst)

(c)

Datum WGS84
UTM Fuso 23 Sul



0 400 800 1,600 2,400 3,200 meters



Legend

- farm boundary
- C4_5 clusters (better)
- C3_5 clusters (interm. worse)
- C2_5 clusters (intermediate)
- C1_5 clusters (worst)
- C0_5 clusters (interm. better)

(d)

Datum WGS84
UTM Fuso 23 Sul



0 400 800 1,600 2,400 3,200 meters

FIGURE 7. Spatial distribution of the parcels considering 2(a), 3(b), 4(c) and 5(d) clusters format.

CONCLUSION

Data mining technique, by applying grouping (clustering) in the spectral profiles, was able to distinguish plots of coffee crops under different conditions, showing that this technique can assist in agricultural planning of coffee plantations.

REFERENCES

- DENISE, E.; BARANOSKI, G.V.G. The application of photoacoustic absorption spectral data to the modeling of leaf optical properties in the visible range. *IEEE Transactions on Geoscience and Remote Sensing*, Piscataway, v.45, n.12, p.4077-4086, 2007.
- FAYYAD, U.; SHAPIRO, G.P.; SMYTH, P. Knowledge discovery and data mining: Towards a unifying framework. In: INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING. 2., 1996, Portland. *Proceedings...*
- GATES, D.M.; KEEGAN, J.; WEIDNER, V.R. Spectral properties of plants. *Applied Optics*, Washington, v.4, n.1, p.11-20, 1965.
- GUIDINI, M.; RIBEIRO, C. Utilização da biblioteca Terra Lib para algoritmos de agrupamento em Sistemas de Informações Geográficas. In: SIMPÓSIO BRASILEIRO DE GEOINFORMÁTICA, 8., 2006, Campos do Jordão. *Anais...*
- JENSEN, J. R. *Sensoriamento remoto do ambiente: uma perspectiva em recursos terrestres*. Tradução de José Carlos Epihânio et al. São José dos Campos: Parêntese, 2009. 598 p.
- LAXMAN, S.; SASTRY, P.S. A survey of temporal data mining. *Sādhanā*, Bangalore, v.31, n.2, p.173-198, 2006.
- MILLER, H.J.; HAN, J. *Geographic data mining and knowledge discovery: an overview*. 2nd. ed. New York: CRC Press, 2009. Cap. 1, p.2-26.
- MOREIRA, M.A.; BARROS, M.A.; DE FARIA, V.G.C.; ADAMI, M. Tecnologia de informação: imagens de satélite para o mapeamento de áreas de café de Minas Gerais. *Informe Agropecuário*, Belo Horizonte, v.28, n.241, p.27-37, 2007.
- MOREIRA, M.A.; RUDORFF, B.F.T.; BARROS, M.A.; DE FARIA, V.G.C.; ADAMI, M. Geotecnologias para mapear lavouras de café nos Estados de Minas Gerais e São Paulo. *Engenharia Agrícola*, Jaboticabal, v.30, n.6, p.1123-1135, 2010.
- PEÑA-BARRAGÁN, J.M.; NGUGI, M.K.; PLANT, R.E.; SIX, J. Object-based crop identification using multiple vegetation indices, textural features and crop phenology. *Remote Sensing of Environment*, New York, v.115, p.1301-1316, 2011.
- PONZONI, F.J.; SHIMABUKURO, Y. E. *Sensoriamento remoto no estudo da vegetação*. São Paulo: Parênteses Editora, 2007. Cap 1.p.7-40.
- RIE, H.; OSAMU, K. Temporal rule discovery for time-series satellite images and integration with RDB. Congrès PKKD 2001: principles of data mining and knowledge discovery. In: EUROPEAN CONFERENCE ON PRINCIPLES OF DATA MINING AND KNOWLEDGE DISCOVERY, 5., 2001, Freiburg. v.2168, p.204-215.
- RUDORFF, C.M.; NOVO, E.M.L.M.; GALVÃO, L.N.; PEREIRA FILHO, W. Análise derivativa de dados hiperespectrais medidos em nível de campo e orbital para caracterizar a composição de águas opticamente complexas na Amazônia. *Acta Amazônica*, Manaus v.37, n.2, p.269-280, 2007.
- TISOT, D.A.; FORMAGGIO, A.R.; RENNÓ, C.D.; GALVÃO, L.S. Eficácia de dados Hyperion/EO-1 para identificação de alvos agrícolas: comparação com dados ETM⁺/ Landsat. *Engenharia Agrícola*, Jaboticabal, v.27, n.2, p.511-519, 2007.

WITTEN, I.H.; FRANK, E. *Data mining: practical machine learning tools and techniques*. 2nd. ed. San Francisco: Morgan Kaufmann, 2005. 525 p.

ZHANG, Z.; WU, W.; HUANG, Y. Effective spatio-temporal analysis of remote sensing data. In: *Progress in WWW Research and Development*. Berlin, 2008. p.584-589. Disponível em: <www.springerlink.com/content/xm577273235k8661>. Acesso em: 16 mar. 2009.