Journal of
Applied Remote Sensing

Classification of soybean varieties using different techniques: case study with Hyperion and sensor spectral resolution simulations

Fábio M. Breunig
Lênio S. Galvão
Antônio R. Formaggio
José C. N. Epiphanio

SPIE

Classification of soybean varieties using different techniques: case study with Hyperion and sensor spectral resolution simulations

Fábio M. Breunig, Lênio S. Galvão, Antônio R. Formaggio, and José C. N. Epiphanio
Instituto Nacional de Pesquisas Espaciais (INPE), Divisão de Sensoriamento Remoto, Caixa Postal 515, 12245-970, São José dos Campos, São Paulo, Brazil
breunig@dsr.inpe.br

Abstract. Next generation imaging spectrometers with higher signal-to-noise ratio and broader swath-width bring new perspectives for crop classification over large areas. Here, we used Hyperion/Earth Observing-One data collected over Brazilian soybean fields to evaluate the performance of four classification techniques (maximum likelihood — ML; spectral angle mapper — SAM; spectral information divergence — SID; support vector machine — SVM) to discriminate five soybean varieties. The spectral resolution influence on classifying them was analyzed by simulating the spectral bands of seven multispectral sensors using Hyperion data. Before classification, the Waikato environment for knowledge analysis was used for feature selection. Results showed the importance of the green, red-edge, near-infrared, and shortwave infrared to discriminate the soybean varieties. Because the soybean variety Monsoy 8411 was sensed by Hyperion in a later reproductive stage, it was more easily discriminated than the other varieties. The best classification techniques were ML and SVM with overall accuracy of 89.80% and 81.76%, respectively. The accuracy of spectral matching techniques was lower (70.84% for SAM and 72.20% for SID). When ML was applied to the simulated spectral resolution of the multispectral sensors, moderate resolution imaging spectroradiometer and enhanced thematic mapper plus presented the highest accuracy, whereas advanced very high resolution radiometer showed the lowest one. [DOI: 10.1117/1.3604787]

Keywords: hyperspectral; soybean; classification; Hyperion; sensor simulation; agriculture.

Paper 10192RR received Dec. 2, 2010; revised manuscript received May 11, 2011; accepted for publication Jun. 7, 2011; published online Jun. 29, 2011.

1 Introduction

Imaging spectrometry, also known as hyperspectral remote sensing, can be potentially used for the monitoring, classification, and grain yield estimation of perennial,[1,2] semiperennial,[3,4] and annual crops.[5–8] It can be also used for cultivar discrimination, which is important for inventory, crop certification, and for improving the performance of crop yield models.[9–11] For crop type discrimination, the possibilities in the near future of having orbital imaging spectrometers with better signal-to-noise ratio (SNR) and larger swath width than the available hyperspectral sensors introduce new perspectives to improve classification accuracy. For example, when compared to the available Hyperion/Earth Observing-One (EO-1),[12] which acquires data on demand over narrow swath width (7.7 km) and with low SNR [40 in the shortwave infrared (SWIR)], NASA's planned hyperspectral infrared imager will acquire images in more than 200 bands with swath width of 150 km and SNR of 400 in the SWIR.[13] Therefore, some limitations of multispectral

1931-3195/2011/$25.00 © 2011 SPIE

data for crop type discrimination may be overcome with a hyperspectral Landsat/TM alike sensor.[14–17]

Despite the low SNR and narrow swath width, Hyperion acquired data over Brazilian agricultural fields providing an opportunity to test different classification techniques and to simulate the spectral resolution of distinct multispectral sensors. Classification results can be then compared between the Hyperion and simulated multispectral sensors for cultivar discrimination. Maximum likelihood (ML), spectral angle mapper (SAM), spectral information divergence (SID), and support vector machine (SVM) are four examples of very distinct classification techniques usually used in hyperspectral analyses. ML assumes that the statistics for each class in each band are normally distributed and calculates the probability that a given pixel belongs to a specific class.[18] SAM calculates the angular similarity between a reference spectra and each pixel spectrum comparing the shape between the reflectance curves.[19,20] SID differs from SAM by using a measure of divergence to match pixel to reference spectra.[21–23] Finally, SVM is especially useful for classification of high dimensional and noisy data discriminating classes with a decision surface that maximizes the margin between them.[24,25] In general, classification of hyperspectral data using ML requires feature selection to avoid the Hughes phenomenon, that is, the loss of accuracy when data dimensionality increases while the training sample size remains fixed.[26,27] However, even for SVM, feature selection is recommended, as demonstrated recently by Ref. 28. The objective of feature selection is to identify a subset of original features (variables) that preserve the essential information, excluding the highly correlated and redundant features from classification.[28]

Soybean is an important agricultural commodity, which has presented an expressive expansion in the last decades in Brazil.[29,30] On February 8, 2005, Hyperion acquired data over a soybean farm located in the Brazilian state of Mato Grosso. Five soybean varieties were sensed by Hyperion at different reproductive stages. Thus, in this work we test the hypothesis that different classification techniques applied to hyperspectral data can improve cultivar discrimination when compared to simulated spectral resolution data of selected multispectral sensors. Furthermore, we consider the hypothesis that the discrimination of some varieties is much more related to small differences in phenological stages rather than to differences in canopy characteristics between them.

Thus, the objectives of this work were: a. to evaluate the performance of different classification techniques for the discrimination of soybean varieties, when using Hyperion data, and b. to analyze the spectral resolution influence on discriminating them by simulating band positioning of selected multispectral sensors from Hyperion.

2 Methodology

2.1 Study Area and Soybean Agronomic Data

The study area is the Tanguro farm located in Querência municipality, Mato Grosso state, central Brazil (Fig. 1). This area is characterized by a flat topography (350 m of altitude) with soil predominance of Latossolo Vermelho-Amarelo distrófico (Typic Acrustox, in the USA Soil Taxonomy).[31] The climate type is designated as tropical, with mean temperature of 26°C and annual rainfall of 1850 mm, with a well-defined dry season from May to September and a rainy season from October to April. The native vegetation comprises the transition between savannas (Cerrado) and Amazonian tropical rainforest.

Soybean varieties are continuously developed to improve the yield, the resistance to natural diseases, and environmental stresses (soil adaptation; heavy rain or water stress), and the adaptation to the local photoperiod.[32] As shown in the reference map (Fig. 1), five soybean varieties were planted at the Tanguro farm in the 2004 to 2005 growing season, covering approximately 8500 ha: Perdiz, Monsoy 8411, Monsoy 9010, Kaiabi, and Tabarana. Perdiz was predominant in the study area and Monsoy 8411 presented the shortest life cycle (116 days). In general, the varieties were planted on different dates along October to November and harvested in the period

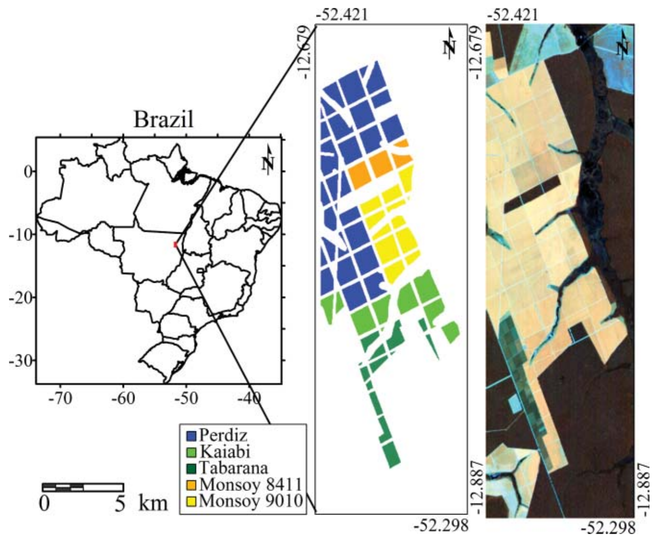

Fig. 1 Location of the Tanguro farm in central Brazil. The reference map provided by the farm of the soybean varieties planted in the 2004 to 2005 growing season is indicated and other land cover types are masked in white. The false color composite includes the Hyperion bands positioned at 884 nm, 1659 nm and 2203 nm in RGB.

of March to April (Table 1). Better grain yields were obtained for Monsoy 9010 and Monsoy 8411. The average yield of all varieties was 2942 kg/ha. All soybean varieties showed complete canopy closure but different reproductive stages, that ranged from R5 (beginning seed) to R6 (full seed) and R7 (beginning maturity), at the acquisition time of the Hyperion image (February 8, 2005). The row spacing was constant (45 cm).

2.2 Pre-Processing of Hyperion Data

Hyperion acquires images in 196 radiometrically calibrated narrow and continuous bands (10 nm of width) in the 400 to 2400 nm spectral region. The swath width is 7.7 km and the spatial resolution is 30 m. The 16-days temporal resolution can be reduced by side looking or sensor pointing.[12,34] However, the resultant off-nadir viewing usually enhances directional effects due

Table 1 Agronomic data for the five soybean varieties planted at the Tanguro farm in the 2004 to 2005 growing season.

Varieties	Planting date MM/DD,YY	Harvesting date MM/DD,YY	Local cycle (Days)	DAP[a] (Days)	Yield (kg/ha)	R[b]
Perdiz	Oct. 30 to Nov. 8, 2004	March 7 to April 9, 2005	138	96	2926.70	5
Monsoy 8411	Oct. 29 to Nov. 1, 2004	Feb. 20 to 28, 2005	116	100	3190.23	6 to 7
Monsoy 9010	Nov. 12 to 17, 2004	March 24 to April 1, 2005	131	85	3160.81	5
Kaiabi	Nov. 8 to 11, 2004	March 31 to April 10, 2005	147	91	2706.06	5
Tabarana	Nov. 10 to 13, 2004	March 31 to April 10, 2005	145	89	2919.26	5

[a]Days after planting (DAP), considering the mean planting date and the Hyperion image acquisition on February 8, 2005.
[b]Reproductive stage ranging from R5 (beginning seed) to R6 (full seed) and R7 (beginning maturity), defined according to Ref. 33.

to crop anisotropy in spectral response to changes in view-illumination geometry.[7,35,36] The impact of directional effects on the reflectance of different soybean varieties was discussed in Ref. 7 when using Moderate Resolution Imaging Spectroradiometer (MODIS)/Terra data acquired at close dates (fixed phenological stages) and in opposite view angles (+44 deg and −42 deg). For a given soybean variety, reflectance differences up to 20% in the 250-m MODIS near-infrared band were observed between the forward scattering and backscattering directions.

In this study, Hyperion acquired the image on February 8, 2005, with 21 deg view angle in the forward scattering direction, in which shade effects predominated for the sensor. Image preprocessing involved the application of an algorithm for striping removal[37] and another one for wavelength recalibration by spectral matching of both the MODTRAN-modeled and observed radiance spectra.[38] This spectral matching procedure is also useful to reduce the smile effect, which may produce variations between 2.6 and 3.6 nm in the visible and near-infrared (VNIR) and less than 1 nm in the SWIR.[39] Such variations have a small impact on crop classification considering the 10-nm bandwidth of Hyperion. Keystone effects were not corrected. The fast line-of-sight atmospheric analysis of spectral hypercubes was used to convert radiance data into atmospherically corrected surface reflectance.[40,41] A tropical-rural atmospheric model was adopted with water absorption estimates from the 1130 nm spectral feature. We used the 2-band (K-T) method[42] to estimate the amount of aerosols and the scene average visibility (63 km). Unfortunately, there were no weather stations in the region that could be used for atmospheric parameters extraction. To smooth data for artifacts resultant from the atmospheric correction and poor SNR of Hyperion, an inverse minimum noise fraction transformation was applied over different spectral regions.[43]

After preprocessing the data, the following 150 Hyperion bands were used in data analysis: bands 6 to 50 (477 to 915 nm), 51 to 91 (932 to 1336 nm), 105 to 136 (1477 to 1790 nm), and 160 to 192 (2032 to 2355 nm). Noisy Hyperion bands were excluded, as well as those positioned around 1400 and 1900 nm, which are wavelengths of strong water vapor absorption. The final step was the geometric correction using a polynomial function and the nearest neighbor resampling method.

2.3 Data Analysis

A reference map of the 2004 to 2005 growing season was provided by the Tanguro farm showing the precise location per plot of the five planted soybean varieties (Fig. 1). We tested their discrimination using four different classification techniques (ML, SAM, SID, and SVM) applied to Hyperion data. Before classifying the Hyperion data, we used the Waikato Environment for Knowledge Analysis (WEKA) data mining algorithm for feature selection.[44,45] WEKA is a collection of machine learning algorithms for data mining tasks, which contains several modules. Feature selection is one of the modules, which consisted of selecting the best subset of Hyperion bands from the set of 150 preprocessed bands. It was based on the greedy hill climbing stepwise search method with rank generation.[27,45] The method was applied to the training set of 1500 pixels (300 pixels per soybean variety). The correlation-based feature selection subset evaluator,[46] which considers the individual predictive ability of each band and the redundancy between the bands, was adopted in data analysis.[27,46,47] The ranked bands from feature selection were also used to evaluate the Hughes effect on each classification technique. The Hughes effect has been observed in many remote sensing studies for a range of classifiers. For example, parametric techniques such as ML may not be able to classify a dataset accurately if the ratio of the sample size to number of features is small.[28]

To allow comparison of classification accuracy between these techniques from a common baseline, we fixed the number of training (300 randomly selected per variety) and validation (total of 46,875 of the reference map) pixels. The same procedure was adopted for the number of input bands for classification, which was defined from feature selection. The classification was performed exclusively over the soybean farm and other land cover types were masked.

Table 2 Nominal band positioning of selected multispectral sensors. Spectral resolution simulations from Hyperion data used the actual spectral band shapes (filter functions) of the sensors. Variations in SNR between them were not evaluated.

Sensor	Blue band (nm)	Green band (nm)	Red band (nm)	Red edge band (nm)	NIR-1 band (nm)	NIR-2 band (nm)	SWIR-1 band (nm)	SWIR-2 band (nm)
AVHRR	–	–	580 to 680	–	725 to 1000	–	1580 to 1640	–
ETM+	450 to 515	525 to 605	630 to 690	–	775 to 900	–	1550 to 1750	2090 to 2350
MODIS	438 to 448 459 to 479 483 to 493	526 to 536 545 to 565 546 to 556	620 to 670 662 to 672 673 to 683	743 to 753	841 to 876 862 to 877 890 to 920	1230 to 1250 931 to 941 915 to 965	1628 to 1652	2105 to 2155
ASTER	–	520 to 600	630 to 690	–	760 to 860	–	1600 to 1700	2145 to 2185 2185 to 2225 2235 to 2285 2295 to 2365
HRG	–	500 to 590	610 to 680	–	780 to 890	–	1580 to 1750	–
IKONOS	445 to 516	506 to 595	632 to 698	–	757 to 853	–	–	–
RapidEye	440 to 510	520 to 590	630 to 685	690 to 730	760 to 850	–	–	–

Using the reference map (Fig. 1) and masking the training pixels, confusion matrices, which are used to assess classification accuracy and misclassification between classes, were generated to obtain the producer's and user's accuracies, the overall classification accuracy, and the Kappa values. Discussion on the discrimination of the soybean varieties considered also the differences in reproductive stages between the varieties.

To analyze the spectral resolution influence for discriminating the studied five soybean varieties, we simulated band positioning of seven multispectral sensors using their respective filter functions (Table 2): 1. advanced spaceborne thermal emission and reflection radiometer (ASTER)/Terra; 2. advanced very high resolution radiometer (AVHRR)/NOAA-17; 3. IKONOS; 4. enhanced thematic mapper plus (ETM+)/Landsat-7; 5. MODIS/Terra; 6. rapideye; and 7. high resolution geometric (HRG)/SPOT-5. ASTER band 9 (2360 to 2430 nm) and MODIS bands 8 (405 to 420 nm) and 20 to 36 (3660 to 14,385 nm) were not simulated because they were out of the spectral range of Hyperion operability. The sensors listed in Table 2 were selected to represent distinct nominal band positioning and bandwidths. In Table 2, AVHRR, RapidEye, and IKONOS represent data acquisition only in the visible and near-infrared, whereas the remaining sensors represent an extended operability into the SWIR. The actual SNR of the simulated sensors was not evaluated in this study. All the analyses were done keeping the Hyperion spatial resolution constant (30 m) and using the spectral bands representative of different satellite sensors.

The classification technique of best performance with Hyperion was applied to images of the spectral resolution simulated sensors and results were compared between them. Training (300 per variety) and validation pixels (reference map) were kept constant between the simulated sensors. After masking the training pixels, confusion matrices were generated for each sensor classification.

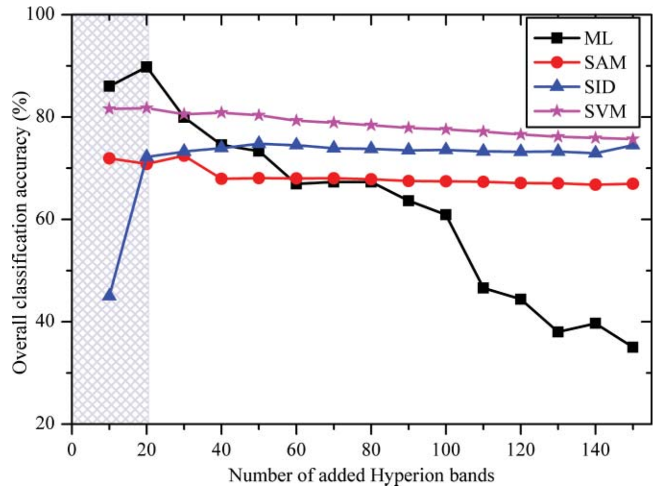

Fig. 2 Effects of the inclusion of the Hyperion ranked bands (feature selection) on the overall classification accuracy of the 1500 training pixels by the four classification techniques. The first 20 bands (dashed area) used for subsequent Hyperion image classification are discussed in the text.

3 Results and Discussion

3.1 Feature Selection

From the use of the WEKA data mining algorithm, we selected 20 Hyperion bands centered at the following wavelengths as the best subset of bands to discriminate the five soybean varieties: 508; 538; 548; 559; 569; 589; 701; 711; 721; 732; 752; 874; 932; 1124; 1628; 1659; 1749; 2102; 2163; 2203 nm. Classification accuracy of the training pixels using ML, SAM, SID, and SVM did not improve significantly with the inclusion of additional bands in the analysis (Fig. 2). However, the Hughes effect, decreasing the classification accuracy, was evident on the ML classification of the training pixels when more than 100 bands were included in data analysis. Even for SVM, classification accuracy decreased as more bands were added from 82% (20 bands) to 76% (150 bands). This result is in agreement with that found by Ref. 28, who demonstrated that SVM is still sensitive to the Hughes phenomena, especially when a small number of training samples is used.

The first ranked bands by WEKA included the green (508, 538, 548, 559, 569, and 589 nm), red edge (701, 711, 721, 732 nm), and NIR (752, 874, 932, and 1124 nm) spectral regions, closely followed by the SWIR-1 (1628, 1659, and 1749 nm). The last ranked bands were placed in the SWIR-2 (2102; 2163; 2203 nm). Green and red edge bands are mainly sensitive to leaf pigments such as chlorophyll, whereas NIR bands are more sensitive to biophysical attributes such as leaf area index. However, NIR bands placed at 932 and 1124 nm comprise the limits of well-defined canopy/leaf water spectral features centered at 980 and 1200 nm. Canopy moisture is the most important spectral factor to reduce the reflectance in the SWIR-1 and SWIR-2 spectral regions, but the selected 2102 nm Hyperion band absorbs radiation due to lignin-cellulose.

Relationships between the reflectance of the green (559 nm) and NIR (874 nm) Hyperion bands showed a clear discrimination of Monsoy 8411, which presented higher reflectance values in the green and lower in the NIR than the other soybean varieties [Fig. 3(a)]. As discussed before, when compared to the other varieties, Monsoy 8411, which has the shortest life cycle (116 days), was sensed by Hyperion in a more advanced reproductive stage (Table 1). Hyperion bands positioned in the red edge spectral region (690 to 750 nm) could also differentiate Monsoy 8411, whereas the SWIR-1 range (1500 to 1750 nm) was especially useful to discriminate Monsoy 9010 from the other varieties [Fig. 3(b)].

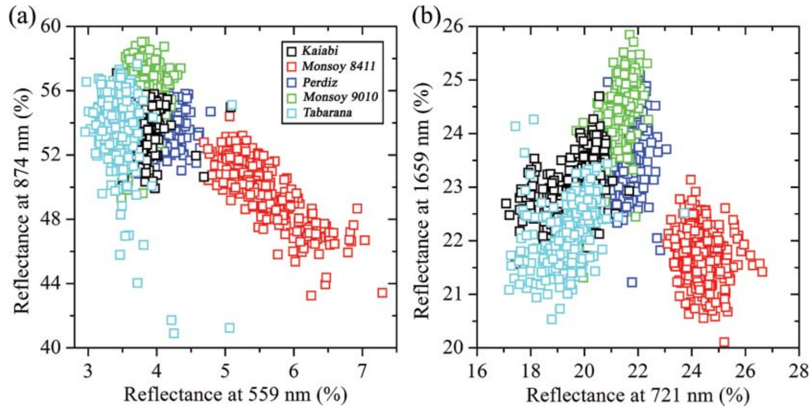

Fig. 3 Relationships between the reflectance of the Hyperion bands positioned at: (a) 559 nm and 874 nm; and (b) 721 nm and 1659 nm. A total of 300 pixels per variety were used.

3.2 Classification Techniques Applied to Hyperion Data

When compared to the reference map, the best classification results with the selected 20 Hyperion bands were obtained using ML, whereas the worst results were obtained using SAM (Fig. 4). Overall classification accuracy and Kappa values ranged from 89.80% and 0.85 for ML to 70.84% and 0.59 for SAM, respectively (Table 3). Inspection of the confusion matrices showed that all of the studied techniques discriminated adequately the variety Monsoy 8411 which presented producer's accuracy values close or equal to 100%.

The good discrimination of Monsoy 8411 was associated to its more advanced reproductive stage (R6 to R7) at the time of Hyperion image acquisition. In relation to the other varieties, Monsoy 8411 presented larger reflectance in the visible and lower reflectance in the NIR and SWIR (Fig. 5), as well as it also showed larger variability (standard deviation bars; results not shown). As demonstrated by Ref. 7, soybean development stages impact classification results,

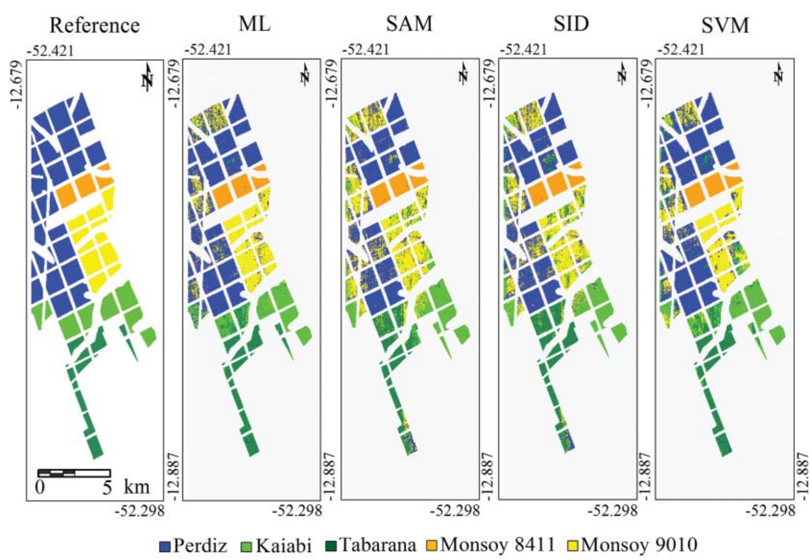

Fig. 4 Reference map of the soybean varieties provided by the farm and Hyperion image classification by the four techniques. ML and SAM presented the highest and lowest overall classification accuracy, respectively.

Table 3 Confusion matrices for the classification of soybean varieties using four different classification techniques applied to Hyperion data (reference map) using a fixed number of 20 bands chosen by the feature selection algorithm.

Soybean variety	Perdiz (%)	Monsoy 8411 (%)	Monsoy 9010 (%)	Kaiabi (%)	Tabarana (%)	Total (pixels)	Producer's Accuracy (%)
ML							
Perdiz (%)	87.36	0.00	6.73	2.15	0.51	23605	87.36
Monsoy 8411 (%)	0.33	100.00	0.03	0.04	0.04	3242	100.00
Monsoy 9010 (%)	8.68	0.00	89.20	1.09	0.02	7418	89.20
Kaiabi (%)	3.38	0.00	3.85	92.42	7.89	7769	92.42
Tabarana (%)	0.24	0.00	0.19	4.30	91.54	4841	91.54
User's accuracy (%)	96.76	97.42	75.60	83.00	91.72	46875	
Overall accuracy (%)	89.80						
Kappa coefficient	0.85						
SAM							
Perdiz (%)	74.34	0.03	11.39	2.82	6.69	23605	74.34
Monsoy 8411 (%)	0.34	99.97	0.00	0.00	0.02	3242	99.97
Monsoy 9010 (%)	18.80	0.00	62.69	16.80	0.23	7418	62.69
Kaiabi (%)	1.65	0.00	2.32	47.81	9.32	7769	47.81
Tabarana (%)	4.87	0.00	23.59	32.57	83.74	4841	83.74
User's accuracy (%)	92.67	97.56	44.69	78.59	42.74	46875	
Overall accuracy (%)	70.84						
Kappa coefficient	0.59						
SID							
Perdiz (%)	69.73	0.06	4.11	2.19	6.61	23605	69.73
Monsoy 8411 (%)	0.17	99.94	0.01	0.03	0.06	3242	99.94
Monsoy 9010 (%)	24.89	0.00	81.28	12.59	5.40	7418	81.28
Kaiabi (%)	3.64	0.00	3.67	55.02	8.73	7769	55.02
Tabarana (%)	1.58	0.00	10.93	30.18	79.20	4841	79.20
User's accuracy (%)	95.36	98.60	45.86	73.28	52.40	46875	
Overall accuracy (%)	72.20						
Kappa coefficient	0.62						
SVM							
Perdiz (%)	79.42	0.00	8.76	3.53	1.31	23605	81.63
Monsoy 8411 (%)	0.19	100.00	0.00	0.00	0.02	3242	100.00
Monsoy 9010 (%)	15.19	0.00	81.63	6.24	0.10	7418	81.63
Kaiabi (%)	3.71	0.00	8.48	76.65	9.26	7769	76.65
Tabarana (%)	1.49	0.00	1.13	13.58	89.31	4841	89.31
User's accuracy (%)	95.00	98.57	59.77	75.25	74.60	46875	
Overall accuracy (%)	81.76						
Kappa coefficient	0.74						

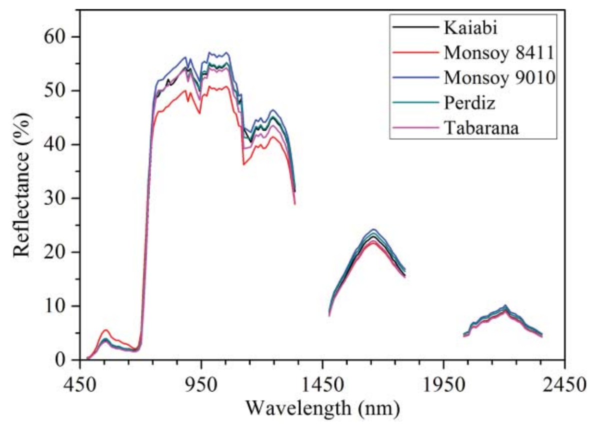

Fig. 5 Hyperion reference spectra of the five studied soybean varieties.

because the varieties have different life cycles (growth rates) and they are planted at slightly different periods from October to November in central Brazil.

Continuing in the analysis of Table 3, the most important difference in classification performance between the techniques used was associated with the better ability of ML and SVM to differentiate the Kaiabi variety. Producer's accuracy for Kaiabi ranged from 92.42% (ML) and 76.65% (SVM) to 55.02% (SID) and 47.81% (SAM). Kaiabi was generally misclassified as Perdiz or Tabarana by SAM and SID (Fig. 4; Table 3), which are techniques that compare each pixel spectrum with reference spectra of the varieties. Results were consistent with Fig. 5 that showed similarity in spectra shape between the Kaiabi and Perdiz-Tabarana varieties.

3.3 Spectral Resolution Influence on the Discrimination of Soybean Varieties

When the spectral resolution simulation of selected multispectral sensors was performed, MODIS/Terra and ETM+/Landsat-7 presented the highest classification accuracy values (86.72% and 85.94%, respectively), whereas the AVHRR/NOAA-17 showed the lowest value (68.85% in Table 4). Classification derived from AVHRR/NOAA-17 was affected by its poor spectral resolution and the absence of bands in the SWIR. Kappa values ranged from 0.57 (AVHRR) to 0.81 (MODIS). ML MODIS classification accuracy and Kappa values (Table 4) were comparatively lower than Hyperion (Table 3). In this context, spectral resolution simulation from Hyperion to multispectral sensors may impact positively on classification results because of the noise attenuation due to spectral resampling from narrow to broad bands.

In agreement with previous results for Hyperion, inspection of the ML classified images (Fig. 6) and of the resultant confusion matrices (Table 5) for MODIS and AVHRR showed

Table 4 Overall classification accuracy and Kappa coefficient values derived from ML classification of the five soybean varieties after the spectral resolution simulation of seven multispectral sensors.

Sensor	Overall accuracy (%)	Kappa coefficient
AVHRR/NOAA-17	68.85	0.57
ETM+/Landsat-7	85.94	0.80
MODIS/Terra	86.72	0.81
ASTER/Terra	80.47	0.72
HRG/SPOT-5	76.46	0.67
IKONOS	82.87	0.76
RapidEye	82.97	0.76

Table 5 Confusion matrices for the classification of soybean varieties using the ML technique applied to MODIS/Terra and NOAA/AVHRR-17 multispectral resolution simulated sensors from Hyperion/EO-1.

Soybean variety	Perdiz (%)	Monsoy 8411 (%)	Monsoy 9010 (%)	Kaiabi (%)	Tabarana (%)	Total (pixels)	Producer's Accuracy (%)
MODIS/Terra							
Perdiz (%)	86.33	0.00	12.12	3.77	0.86	23605	86.33
Monsoy 8411 (%)	0.39	100.00	0.30	0.05	0.06	3242	100.00
Monsoy 9010 (%)	11.33	0.00	83.15	3.06	0.18	7418	83.15
Kaiabi (%)	1.58	0.00	4.23	82.97	7.75	7769	82.97
Tabarana (%)	0.37	0.00	0.20	10.14	91.15	4841	91.15
User's accuracy (%)	94.29	96.38	67.86	85.80	83.38	46875	
Overall accuracy (%)	86.72						
Kappa coefficient	0.81						
NOAA/AVHRR-17							
Perdiz (%)	64.15	0.00	16.59	19.56	0.43	23605	64.15
Monsoy 8411 (%)	0.39	95.38	0.04	1.03	1.90	3242	95.38
Monsoy 9010 (%)	11.95	0.00	75.58	8.13	0.96	7418	75.58
Kaiabi (%)	19.24	3.48	5.08	56.49	13.36	7769	56.49
Tabarana (%)	4.27	1.14	2.71	14.79	83.36	4841	83.36
User's accuracy (%)	84.53	92.00	61.57	43.55	63.06	46875	
Overall accuracy (%)	68.85						
Kappa coefficient	0.57						

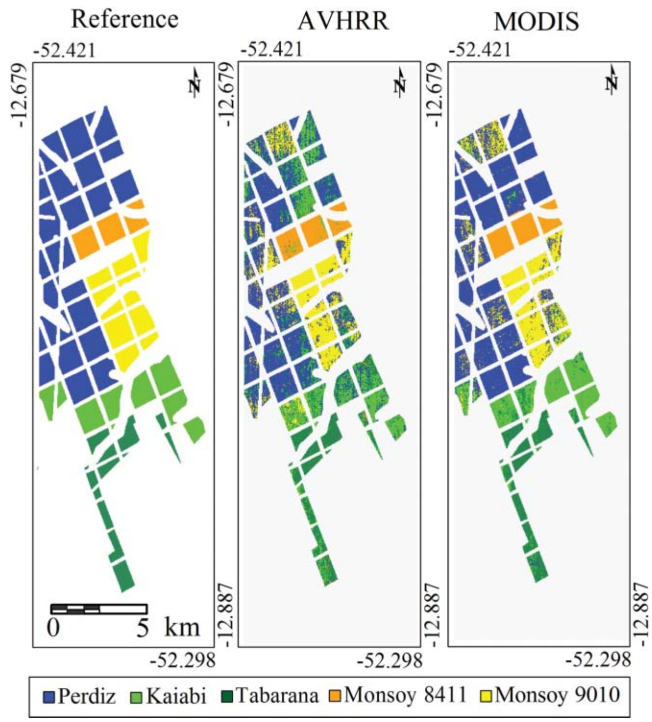

Fig. 6 Reference map of the soybean varieties and classification results for the simulated spectral response of the AVHRR/NOAA-17 and MODIS/Terra bands.

that both simulated sensors adequately discriminated Monsoy 8411. This result indicates the importance of the information on soybean planting date and phenological stage for the correct interpretation of remote sensing, as already reported by Refs. 48 and 49. Kaiabi presented the lowest producer's accuracy values for both sensors (Table 5), which was consistent with its reflectance spectrum of Fig. 5 that showed confusion with spectra of other varieties.

4 Conclusions

The Hughes effect on the ML classification was observed, especially when more than 100 bands were included in the analysis, decreasing classification accuracy. Results from feature selection indicated the importance of the Hyperion bands positioned in the green, red edge, NIR, and SWIR-1 to discriminate the five studied soybean varieties.

From the four tested classification techniques, ML and SVM presented the best results with overall classification and Kappa values of 89.80% and 0.85 and of 81.76% and 0.74, respectively. The worst performance was obtained by SAM (70.84% and 0.59). The most easily discriminated soybean variety was Monsoy 8411, because it was sensed by Hyperion at a later reproductive stage, presenting comparatively higher reflectance in the visible and lower in the NIR-SWIR-1 than the other varieties. Differences in classification performance between the techniques were associated with the better ability of ML and SVM to differentiate the Kaiabi variety when compared to the matching spectra techniques. Kaiabi presented reflectance spectra shape similarity with Tabarana and Perdiz. Thus, techniques that compared the similarity between the shape of the spectra (SAM and SID) were less effective to discriminate varieties than techniques based on differences between soybean absolute reflectance values (ML and SVM).

Classification performance decreased from Hyperion to the simulated spectral resolution data of the studied multispectral sensors. Among the simulated sensors from Hyperion, MODIS and ETM+ presented the highest classification accuracy, whereas the AVHRR showed the lowest one due its poor spectral resolution and the absence of bands in the SWIR. Kappa values ranged from 0.57 (AVHRR) to 0.81 (MODIS).

Results from Hyperion and the spectral resolution simulation of seven multispectral sensors demonstrated the importance of taking into account the differences in phenological stages between the soybean varieties in discrimination studies.

Acknowledgments

The authors are very grateful to Grupo André Maggi for providing agronomic information for this research. Thanks are also due to FAPESP (Fundação de Amparo à Pesquisa do Estado de São Paulo; Grant No. 2008/11499-8), CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), and to CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior). Comments by the reviewers were really appreciated.

References

1. F. Lacar, M. Lewis, and I. Grierson, "Use of hyperspectral imagery for mapping grape varieties in the Barossa Valley, South Australia," in *Proceedings of IGARSS*, IEEE Computer Society, Sydney, Vol. 6, pp. 2875–2877 (2001).
2. P. J. Zarco-Tejada, A. Berjón, R. López-Lozano, J. R. Miller, P. Martín, V. Cachorro, M. R. González, and A. Frutos, "Assessing vineyard condition with hyperspectral indices: Leaf and canopy reflectance simulation in a row structured discontinuous canopy," *Remote Sens. Environ.* **99**, 271–287 (2005).
3. L. S. Galvão, A. R. Formaggio, and D. A. Tisot, "Discrimination of sugarcane varieties in Southeastern Brazil with EO-1 Hyperion data," *Remote Sens. Environ.* **94**, 523–534 (2005).
4. R. Johnson, R. Viator, J. Veremis, E. Richard, and P. Zimba, "Discrimination of sugarcane varieties with hyperspectral reflectance measurements and plant pigment analysis," *J. Am. Soc. Sugar Cane Technol. (JASSCT)* **28**, 63–75 (2008).
5. W. B. Henry, D. R. Shaw, K. R. Reddy, L. M. Bruce, and H. D. Tamhankar, "Spectral reflectance curves to distinguish soybean from common cocklebur (*Xanthium strumarium*) and sicklepod (*Cassia obtusifolia*) grown with varying soil moisture," *Weed Sci.* **52**, 788–796 (2004).
6. C. H. Koger, D. R. Shaw, K. N. Reddy, and L. M. Bruce, "Detection of pitted morning glory (*Ipomoea lacunosa*) with hyperspectral remote sensing. II. Effects of vegetation ground cover and reflectance properties," *Weed Sci.* **52**, 230–235 (2004).
7. L. S. Galvão, A. D. Roberts, A. R. Formaggio, I. Numata, and F. M. Breunig, "View angle effects on the discrimination of soybean varieties and on the relationships between vegetation indices and yield using off-nadir Hyperion data," *Remote Sens. Environ.* **113**, 846–856 (2009).
8. C. J. Gray, D. R. Shaw, and L. M. Bruce, "Utility of hyperspectral reflectance for differentiating soybean (*Glycine max*) and six weed species," *Weed Technol.* **23**, 108–119 (2009).
9. L. S. Galvão, A. R. Formaggio, and D. A. Tisot, "The influence of spectral resolution on discriminating Brazilian sugarcane varieties," *Int. J. Remote Sens.* **27**, 769–777 (2006).
10. M. Ferreiro-Armán, J. P. Costa, S. Homayouni, and J. Martín-Herrero, "Hyperspectral image analysis for precision viticulture," *Lecture Notes Comput. Sci.* **4142**, 730–741 (2006).
11. Y. Everingham, K. Lowe, D. Donald, D. Coomans, and J. Markley, "Advanced satellite imagery to classify sugarcane crop characteristics," *Agron. Sustain. Dev.* **27**, 111–117 (2007).
12. J. S. Pearlman, P. S. Barry, C. C. Segal, J. Shepanski, D. Beiso, and S. L. Carman, "Hyperion, a space-based imaging spectrometer," *IEEE Trans. Geosci. Remote Sens.* **41**, 1160–1173 (2003).

13. S. Chien, D. Silverman, A. G. Davies, and D. Mandl, “Onboard science processing concepts for the HyspIRI mission,” *IEEE Intell. Syst.* **24**, 12–19 (2009).
14. P. Thenkabail, R. B. Smith, and E. De Pauw, “Hyperspectral vegetation indices and their relationships with agricultural crop characteristics,”*Remote Sens. Environ.* **71**, 158–182 (2000).
15. P. Thenkabail, “Accuracy assessments of hyperspectral waveband performance for vegetation analysis applications,” *Remote Sens. Environ.* **91**, 354–376 (2004).
16. J. L. Hatfield, A. A. Gitelson, J. S. Schepers, and C. L. Walthall, “Application of spectral remote sensing for agronomic decisions,” *Agron. J.* **100**, S-117–S-131 (2008).
17. A. Miglani, S. S. Ray, R. Pandey, and J. S. Parihar, “Evaluation of EO-1 Hyperion data for agricultural applications,” *J. Indian Soc. Remote Sens.* **36**, 255–266 (2009).
18. J. A. Richards, *Remote Sensing Digital Image Analysis*, Springer Verlag, Berlin (1999).
19. F. Kruse, A. Lefkoff, J. Boardman, K. B. Heidebrecht, A. T. Shapiro, P. J. Barloon, and A. F. H. Goetz, “The spectral image processing system (SIPS)—interactive visualization and analysis of imaging spectrometer data,” *Remote Sens. Environ.* **44**, 145–163 (1993).
20. Y. Wang, L. Guo, and N. Liang, “Improving the classification precision of spectral angle mapper algorithm,” *Proc. SPIE* **7498**, 749827 (2009).
21. C. Chang, “Spectral information divergence for hyperspectral image analysis,” in *Proceedigs of the IGARSS*, IEEE Computer Society, Hamburg, Vol. 1, pp. 509–511 (1999).
22. C. Chang, *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*, Springer, New York (2003).
23. Y. Du, C. Chang, H. Ren, C. Chang, J. O. Jensen, and F. M. D’Amico, “New hyperspectral discrimination measure for spectral characterization,” *Opt. Eng.* **43**, 1777–1786 (2004).
24. C. Cortes and V. Vapnik, “Support-vector networks,” *Mach. Learn.* **20**, 273–297 (1995).
25. V. Vapnik, *The Nature of Statistical Learning Theory (Information Science and Statistics)*, 2nd ed., Springer, Berlin (1999).
26. G. Hughes, “On the mean accuracy of statistical pattern recognizers,” *IEEE Trans. Inf. Theor.* **14**, 55–63 (1968).
27. I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann, San Francisco, CA (2005).
28. M. Pal and G. M. Foody, “Feature selection for classification of hyperspectral data by SVM,” *IEEE Trans. Geosci. Remote Sens.* **48**, 2297–2307 (2010).
29. E. Ortega, O. Cavalett, R. Bonifácio, and M. Watanabe, “Brazilian soybean production: Energy analysis with an expanded scope,” *Bull. Sci. Technol. Soc.* **25**, 323–334 (2005).
30. R. D. V. Epiphanio, A. R. Formaggio, B. F. T. Rudorff, E. E. Maeda, and A. J. B. Luiz, “Estimating soybean crop areas using spectral-temporal surfaces derived from MODIS images in Mato Grosso, Brazil,” *Pesqui. Agropecu. Bras.* **45**, 72–80 (2010).
31. J. C. Corrêa, “Effect of cropping systems on the stability of Red-Yellow Latosol aggregates in Querência, Mato Grosso, Brazil,” (in Portuguese), *Pesqui. Agropecu. Bras.* **37**, 203–209 (2002).
32. Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA), “Soybean,” EMBRAPA, Brasilia (2010) (in Portuguese).
33. Ontario Ministry of Agriculture, Food and Rural Affairs (OMAFRA), Growth stages. Agronomy Guide. Pub. 811 (2001).
34. J. S. Pearlman, S. L. Carman, C. C. Segal, P. Jarecke, P. Clancy, and W. Browne, “Overview of the Hyperion imaging spectrometer for the NASA EO-1 mission,” in *Proceedings of IGARSS*, IEEE Computer Society, Sydney, Vol. 7, pp. 3036–3038 (2001).
35. J. C. Kollenkark, V. C. Vanderbilt, C. S. T. Daughtry, and M. E. Bauer, “Influence of solar illumination angle on soybean canopy reflectance,” *Appl. Opt.* **21**, 1179–1184 (1982).
36. K. J. Ranson, L. L. Biehl, and M. E. Bauer, “Variation in spectral response of soybean with respect to illumination, view and canopy geometry,” *Int. J. Remote Sens.* **6**, 1827–1842 (1985).

37. D. G. Goodenough, A. Dyk, K. O. Niemann, J. S. Pearlman, Hao Chen, T. Han, M. Murdoch, and C. West, "Processing Hyperion and ALI for forest classification," *IEEE Trans. Geosci. Remote Sens.* **41**, 1321–1331 (2003).
38. T. Wang, G. Yan, H. Ren, and X. Mu, "Improved methods for spectral calibration of on-orbit imaging spectrometers," *IEEE Trans Geosci. Remote Sens.* **48**, 3924–3931 (2010).
39. L. Liao, P. Jarecke, D. Gleichauf, and T. Hedman, "Performance characterization of Hyperion imaging spectrometer instrument," *Proc. SPIE* **2000**, 264–275 (2000).
40. G. W. Felde, G. P. Anderson, T. W. Cooley, M. W. Matthew, S. M. Adler-Golden, A. Berk, and J. Lee, "Analysis of Hyperion data with the FLAASH atmospheric correction algorithm," in *Proceedings of IGARSS*, IEEE Computer Society, Toulouse, Vol. 1, pp. 90–92 (2003).
41. ITT—Visual Information Solutions, *ENVI*, Version 4.7. Boulder, Colorado, ITT Visual Information Solutions (2009).
42. Y. J. Kaufman, D. Tanré, L. A. Remer, E. F. Vermote, A. Chu, and B. N. Holben, "Operational remote sensing of tropospheric aerosol over land from EOS moderate resolution imaging spectroradiometer," *J. Geophys. Res.* **102**, 17051–17067 (1997).
43. A. A. Green, M. Berman, P. Switzer, and M. D. Craig, "A transformation for ordering multispectral data in terms of image quality with implications for noise removal," *IEEE Trans. Geosci. Remote Sens.* **26**, 65–74 (1988).
44. M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: An update," *ACM SIGKDD Explorations Newsletter* **11**, 10–18 (2009).
45. R. R. Bouckaert, E. Frank, M. Hall, R. Kirkby, P. Reutemann, A. Seewald, and D. Scuse, *WEKA Manual for Version 3.6.0*, Hamilton, New Zealand (2008).
46. M. A. Hall and L. A Smith, "Feature subset selection: A correlation based filter approach," in *Proceedings of the 4th International Conference on Neural Information Processing and Intelligent Information Systems*, pp. 855–858, Springer, Berlin (1997).
47. M. A. Hall, Correlation-based feature selection for machine learning, Doctoral Thesis, University of Waikato, Hamilton, New Zealand (1999).
48. C. P. Ferri, A. R. Formaggio, and M. A. Schiavinato, "Narrow band spectral indexes for chlorophyll determination in soybean canopies [Glycine max (L.) Merril]," *Braz. J. Plant Physiol.* **16**, 131–136 (2004).
49. T. Sakamoto, B. D. Wardlow, A. A. Gitelson, S. B. Verma, A. E. Suyker, and T. J. Arkebauer, "A two-step filtering approach for detecting maize and soybean phenology with time-series MODIS data," *Remote Sens. Environ.* **114**, 2146–2159 (2010).

Fábio M. Breunig is a PhD student at Brazilian National Institute for Space Research (INPE). He received his BS degree in geography from Federal University of Santa Maria (Brazil) in 2002 and his MS degree in remote sensing from the Brazilian National Institute for Space Research (INPE) in 2006. Currently he is working with agriculture applications of remote sensing, focusing on directional effects and radiative transfers modeling.

Lênio S. Galvão is a researcher at the Instituto Nacional de Pesquisas Espaciais (INPE) in Brazil. He is a geologist with a MS Degree in remote sensing (INPE) and a Doctor Degree in geophysics (University of São Paulo — USP). His research interest includes reflectance spectroscopy and hyperspectral remote sensing. He has performed several investigations using airborne (e.g., AVIRIS, HYMAP) and orbital (e.g., Hyperion/EO-1 and CHRIS/PROBA) imaging spectrometers.

Antonio R. Formaggio is an agronomist by Escola Superior de Agricultura Luiz de Queiroz (ESALQ/University of São Paulo), with a Masters degree in remote sensing of environment from the National Institute for Space Research (INPE/Ministry of Science and Technology) and with a Doctors degree in remote sensing and agriculture from ESALQ/University of São Paulo. He is a researcher and post-graduation teacher/advisor in remote sensing and environment and agriculture since 1989. His research interests include agricultural statistics, impacts of agriculture in the environment, hyperspectral remote sensing, and estimation of biophysical variables of crop canopies, environmental modeling and radiative transfer models for the retrieval of vegetation biophysical variables.

José C. N. Epiphanio is a senior researcher at the National Institute for Space Research (INPE), in Brazil. He received his BS in agronomy from University of Sao Paulo (USP) in 1978, his MS degree in remote sensing from INPE in 1982, and his PhD in agronomy/remote sensing from USP in 1988. He spent a sabbatical at University of Arizona in 1993 to 1995. He is actively involved in agriculture/remote sensing research and teaching, and coordinates the Brazilian Remote Sensing Symposia and the China-Brazil Earth Resources Satellite (CBERS) Application Program.