Relationship between multi-spectral data and sugarcane crop yield

Fernando Benvenuti[A] and **Mara Weill**[B]

[A]Cosan Group, Piracicaba, SP, Brazil, Email Fernando.Benvenuti@cosan.com.br
[B]College of Agricultural Engineering, State University of Campinas, Campinas, SP, Brazil, Email mweill@agr.unicamp.br

Abstract
In Sao Paulo State, the extent of agricultural land under sugarcane production encompasses 3.2 million ha which produce more than 300 million Mg/y. In 2006, this corresponded to 60% of the national yield. Due to these huge numbers reliable yield estimations are vital for adequate crop production planning and monitoring at production units. The purpose of this work was to assess the relationship between spectral data and sugarcane yield in a commercial production area in Sao Paulo State to forecast harvest. Three Landsat 7/ ETM+ images were used to extract spectral variables from sugarcane plantations, two of them during the period of maximum vegetative crop growth and the third at the beginning of the maturation stage. Before analyses, the images were processed to perform geometric and radiometric restorations. Three spectral variables were extracted: reflectance of the band 4 (0.76-0.90um), NDVI and GVI indices. Simple linear regression analyses were performed between sugarcane yield (dependent variable) and spectral variables (independent variables). The average values of NDVI, GVI, and reflectance of band 4 declined with time. The temporal differences were significant for GVI index and reflectance of band 4. However, they were not significant for the NDVI index, which is indicative of a less sensitivity of this index to detect multi-temporal differences in sugarcane spectral response during its lifecycle.

Key Words
Yield prediction, vegetation indices, multi-spectral data, remote sensing.

Introduction
In Brazil, sugarcane plantations cover more than 6.2 millions of hectares (CONAB). From this total, 3.2 millions correspond to the planted area in Sao Paulo State, where a hundred of productive units produce up to 300 million Mg/y that represent almost 60% of the national sugarcane yield. Given these huge numbers, methods to forecast harvest are vital not only for adequate monitoring of the production units but also to anticipate the negotiation of sugarcane industrial products, sugar and ethanol. In this context, multi-temporal remote sensing data makes possible monitoring the crop growth during its lifecycle, and also to get anticipated information to forecast harvest. This is due to the fact that healthy canopies of green vegetation have a very distinct interaction with energy in the visible and near-infrared regions of the electromagnetic spectrum that can be detected and quantitatively assessed (Thiam and Eastman 1999). Spectral vegetation indices are models designed to provide a quantitative assessment of green vegetation biomass. Miura *et al*. (2001) had mentioned their application to monitor the vigor of the vegetation cover on global and regional scales. Other authors had already mentioned the correlation between spectral vegetation indices and bio-physical parameters of the vegetation, including leaf area index, biomass, yield, leaf photosynthetic activity, photosynthetic active absorbed radiation (PAR), fraction of photosynthetic active radiation absorbed by the plantation (*f*APAR), and green cover percentage (Huete 1988; Epiphanio and Huete 1994; Elvidge and Chen 1995). In Brazil, the relationship between spectral response of sugarcane and yield was studied by Fortes (2003), Machado (2003), and Simões (2005). The purpose of this work was to assess the relationship between spectral data and sugarcane yield in a commercial area in Sao Paulo State.

Methods
Study area
The studied area encompassed 15 plots occupying 150ha cultivated with a medium maturation sugarcane variety. General topography is gentle (0.05m/m), with increasing slopes until 0.13m/m at the lower part of the hillsides. Main soil types occurring in the area include Typic Hapludox (Red Yellow Latosols medium textured) and Typic Quartzipsamments, both distrofic or with base saturation under 50%. Sugarcane yield records per plot (n=15) for the agricultural year 2002/03 were extracted from the agricultural data base available at the production unit.

Remote sensing data

Three Landsat 7/ ETM+ images, scene 220-076, were used to extract spectral variables from sugarcane plantations in the study area. The first two (dated of Jan 08 and Feb 25, 2003) were chosen to represent the period of maximum vegetative crop growth, and the third (dated of May 16, 2003) to represent the beginning of the maturation stage. Prior to analyses, geometric and radiometric restorations were performed. The atmospheric interferences were corrected employing the software Scoradis (Zullo Junior, 1994), using entry atmospheric parameters extracted from MODIS (*Moderate Resolution Imaging Spectroradiometer)* images, designed by *Level 2 Aerosol over Land and Ocean Product* (MOD04_L2). Three spectral variables were extracted from the images: the reflectance of band 4 (b4, 0.76 – 0.90 µm), corresponding to near-infrared (NIR); the Normalized Difference Vegetation Index (NDVI), introduced by Rouse *et al.* (1974) and calculated by equation 1, and, the Green Vegetation Index (GVI) of the Tasselled Cap, originally described by Kauth and Thomas (1976), and calculated by equation 2 as described by Crist (1985). The image processing and analyses were performed in a GIS environment.

NDVI = (NIR-RED)/ (NIR+RED) (1)

Where: NIR (near infrared, 0.76-0.90µm) and RED (b3, 0.60-0.70µm) of Landsat 7/ ETM+sensor.

GVI = - 0.1603*b1 – 0.2819*b2 – 0.4934*b3 + 0.7940*b4 – 0.0002*b5 – 0.1446*b7 (2)

Where: b1(blue, 0,45-0,52µm), b2 (green, 0,52-0,60µm), b3 (red, 0,60-0,70µm), b4 (NIR, 0,76-0,90µm), b5 (MIR, 1,55-1,75µm), and b7 (MIR, 2,08-2,35µm) of Landsat 7/ ETM+sensor.

Statistical analysis

The relationship between sugarcane yield (dependent variable) and spectral variables (independent variables) was assessed by simple linear regression analyses, employing the software Statistica 6.0. The average value per plot of each spectral variable (b4, NDVI, and GVI) was calculated. In this operation, to exclude the influence of the soil reflectance at the corridors, only the values of the pixels inside each plot were computed (only sugarcane). Then, for each plot (n=15) the values of the spectral variables were correlated with the correspondent sugarcane yield record. Before performing the regression analyses, the normality of yield and spectral data distributions was tested.

Results

Temporal variation of spectral data has shown that average values of NDVI, GVI, and reflectance of band 4 declined with time (Figure 1). For GVI and reflectance of b4, values in January were bigger and statistically different than in February and in May. For NDVI, no differences were found among average values for the three dates, showing the lower sensitivity of this index to detect multi-temporal differences on sugarcane spectral response during its lifecycle. About this point, Moreira (2000) had stated that after certain degree of crop development, NDVI values became invariable and insensitive to increases in green biomass. From the results of Table 1, it can be observed that independent of the crop age, the better performance was for the GVI index. The NDVI index was less efficient to explain sugarcane yield variation in the study area. Despite the fact that both indices, GVI and NDVI, are based on the contrast between the spectral responses of the green biomass in visible and in near infra-red regions (Crist and Cicone 1984), the GVI index has the advantage of considering not only the bands 3 (red) and 4 (near infra-red), like the NDVI index, but also by computing the responses in bands 1 (blue) and 2 (green), which incorporates the influence of photosynthetic pigments present in leaves, and in bands 5 and 7 (medium infra-red), both sensitive to the water content in leaves. The results indicate that the relationship between sugarcane yield data and spectral variables was influenced by crop age and phenological stage. For GVI index and reflectance of band 4, the best fit was found by employing the Feb 25, 2003 Landsat image, practically at the middle of the annual lifecycle of the crop, or when sugarcane plants had an age of 6.3 months (190 days). The best simple linear regression model is represented by equation 3. This model could explain 79% of the observed variation on sugar cane yield in the study area as a function of the GVI value at Feb 25, 2003. The model is significant with 95% confidence ($p < 0.0001$) and the residues show normal distribution according to the *Shapiro-Wilk* test (*p*-value= 0,597).

Sugarcane yield (Mg/ha) = -23.6820 + 5.9434 GVI_Feb 25 (R^2 =0.79) (3)

For practical application, the model must be validated with supplementary analyses.

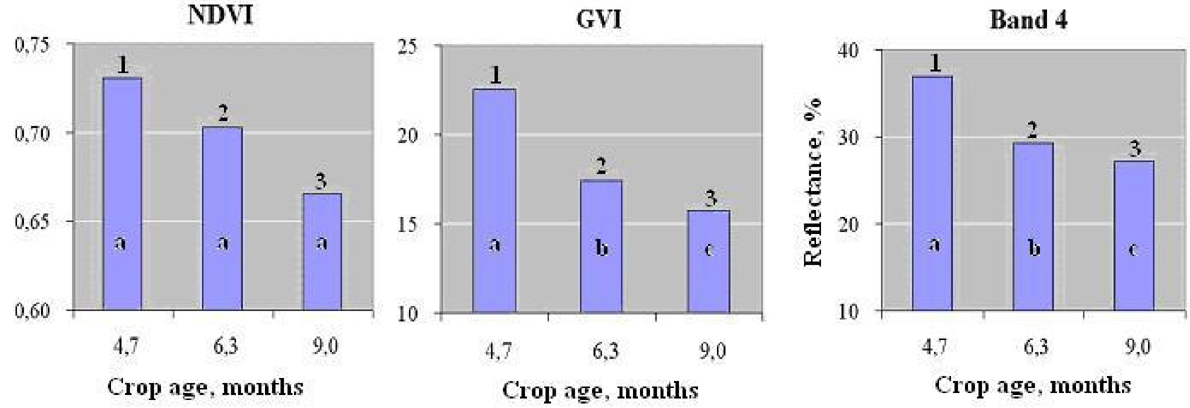

Figure 1. Average values of spectral variables NDVI, GVI, and reflectance of b4 calculated for three dates during sugarcane lifecycle (1- Jan 8, 2- Feb 25, 3- May 16; for each image n=15). For the same spectral variable, average value (represented by bar) assigned with the same case letter don't present statistical difference at the 0,05 confidence level by Tukey test.

Table 1. Results for simple linear regression analysis between sugarcane yield and spectral variables.

Image date	Crop age (months)	Number of observations	NDVI		GVI		Band 4	
		n	R^2	*p-value*	R^2	*p-value*	R^2	*p-value*
Jan 08, 2003	4.7	15	0.65	0.0003	0.68	0.0001	0.34	0.0229
Feb 25, 2003	6.3	15	0.61	0.0005	0.79	<0.0001	0.76	<0.0001
Mai 05, 2003	9.0	15	0.62	0.0005	0.74	<0.0001	0.71	<0.000

Conclusion

Spectral data extracted from sugarcane plantations at the middle of surcarcane annual lifecycle has shown a direct and strong relationship with the final crop yield observed in the study area. The use of spectral variables seems to be a good alternative procedure for obtaining valuable information to enable forecast of the harvest.

References

Benvenuti FA (2005) Relação de índices espectrais de vegetação com a produtividade da cana-de-açúcar e atributos edáficos. Dissertação (Mestrado em Engenharia Agrícola), Faculdade de Engenharia Agrícola, Universidade Estadual de Campinas, Campinas.

Crist EPA (1985) TM tasseled cap equivalent transformation for reflectance factor data. *Remote Sensing of Environment* **17**, 301-306.

Crist EP, Cicone CC (1984) A Physically-Based Transformation of Thematic Mapper data – The TM Tasseled Cap. IEEE Transactions on Geosciences and Remote Sensing, v.GE-22, n.3, may.

Elvidge CD; CHEN Z (1995) Comparison of broad-band and narrow-band red and near-infrared vegetation indices. *Remote Sensing of Environment* 54, 38-48.

Epiphanio JCN; Huete AR (1994) Influence of sun-view geometries on the relationships among vegetation indices, LAI, and absorbed PAR. In 'International Geoscience and Remote Sensing Symposium (IGARSS'94). Surface and Atmospheric Remote Sensing, Pasadena'. *Proceedings. Piscataway: IEEE* **3**, 1455-1457.Fortes C. (2003) Discriminação varietal e estimativa de produtividade agroindustrial de cana-de-açúcar pelo sensor orbital ETM+/LANDSAT7. 131p. Dissertação (Mestrado em Agronomia), Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba.

Huete AR (1988) A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment* **25**, 295-309.

Kauth RJ; Thomas GS The tasseled cap – A graphic description of the spectral-temporal development of agricultural crops as seen by Landsat. In 'Proc. The Symposium on Machine Processing of Remotely Sensed Data, Purdue University, West Lafayette, Indiana' pp. 41-50, 1976.

Machado HM (2003) Determinação da biomassa da cana-de-açúcar considerando a variação espacial de dados espectrais do satélite Landsat 7 – ETM+. Dissertação (Mestrado em Engenharia Agrícola), Faculdade de Engenharia Agrícola, Universidade Estadual de Campinas, Campinas.

Moreira RC (200) Influência do posicionamento e da largura de bandas de sensores remotos e dos efeitos atmosféricos na determinação de índices de vegetação. Dissertação (Mestrado em Sensoriamento Remoto) Instituto de Pesquisas Espaciais, São José dos Campos-SP (INPE-7528-TDI/735).

Miura T, Huete AR, Yoshioka H, Holben BN (2001) An error and sensitivity analysis of atmospheric resistant vegetation indices derived from dark target-based atmospheric correction. *Remote Sensing of Environment* **78**, 284-298.

Simões MS, Rocha, JV, Lamparelli RAC (2005) Spectral variables, growth analysis and yield of sugarcane. *Sci. agric. (Piracicaba, Braz.)*.**62**, 199-207.

Thiam A, Eastman JR (1999) 'Vegetation Indices. In: Guide to GIS and Image Processing'. V.2. Idrisi 32. Clark Labs. (Clark University: Worcester, MA).

Zullo JJ (1994) Correção atmosférica de imagens de satélite e aplicações. 189p. Tese (Doutorado em Engenharia Elétrica), Faculdade de Engenharia Elétrica, Universidade Estadual de Campinas.