PERPENDICULAR CROP ENHANCEMENT INDEX: A NEW APPROACH TO SOYBEAN MONITORING USING TIME-SERIES

Carlos Antonio da Silva Junior[1], Paulo Eduardo Teodoro[2], João Lucas Della-Silva[1], Fernando Saragosa Rossi[3], and Marcos Rafael Nanni[4]

[1]Universidade do Estado de Mato Grosso (UNEMAT), Sinop, Mato Grosso, Brasil, carlosjr@unemat.br, joao.lucas1@unemat.br; [2]Universidade Federal de Mato Grosso do Sul (UFMS), Chapadão do Sul, Mato Grosso do Sul, Brasil, paulo.teodoro@ufms.br; [3]Universidade Estadual Paulista (UNESP), PPG-Ciência do Solo, Jaboticabal, São Paulo, Brasil, fernando.rossi@unesp.br; [4]Universidade Estadual de Maringá (UEM), Maringá, Paraná, Brasil. mrnanni@uem.br

ABSTRACT

In Brazil, despite the improvements with respect the technological knowledge, agricultural areas are often estimated in loco. Here, soybean areas in Paraná, Brazil, using MODIS imagery were mapped. We applied the vegetation index PCEI (Perpendicular Crop Enhancement Index) and threshold determination for the automation of soybean area discrimination by geo-object (GEOBIA). For this, vegetation indices (NDVI, EVI and CEI) and the development of the PCEI were used with the aid of time-series images from the TERRA/MODIS. By geo-objects and decision tree based on data mining support analysis, the new vegetation index was determined. Kappa and Overall Accuracy statistics were applied to evaluate classification precision. Regarding the ground line, R and R² were above 0.92 and 0.84, respectively ($p<0.01$). The test results indicate that the proposed methodology is efficient for mapping soybean distribution. Thus, this study allows automated mapping of with soybean crops areas at large scales.

Key words **—** Spatial distribution, vegetation index, PCEI, digital image processing, data mining.

1. INTRODUCTION

Remote sensing is an advanced technology that has been constantly improved, mainly with regard to its application to agriculture. Through this technique, it is possible to determine several crop traits, such as leaf area index, estimating areas, production, and vegetative vigor [1].

In Brazil, despite the technological knowledge improvements, agricultural areas are often estimated in loco. One of the greatest difficulties in estimating these areas by means of low spatial resolution images is that mixed pixels do not correspond to a unique crop, mainly for small cultivated areas (called “small farmers”), thus hindering classification and requiring additional information from the field and/or provided by the same professionals [2].

The development of agricultural crops identification and monitoring with remote sensing is growing [3]. During certain development stages, agricultural crops present phenological patterns that spectrally distinguish them from other vegetable classes. Thus, satellite time-series images made headlines in agricultural crops identifying [4], [5].

Furthermore, the use of remote sensing to study soybean distribution has been confirmed worldwide and is used for discriminating and quantifying areas, phenological stage identifying [6] and estimating yield by chlorophyll content [7]. Thus, to overcome problems due to heterogeneity of pixels and, for example, variability of agricultural crops in remote sensing images, techniques based on objects have been increasingly used in remote sensing [8]. The idea of analysis based on an image's objects (GEOBIA) is segmenting and building a hierarchical network of homogeneous objects that, e.g., chart limits of crop ratings.

Here, were estimated and mapped areas with soybean crop in the state of Paraná, Brazil using mono and multi-temporal MODIS images. We develop the Perpendicular Crop Enhancement Index (PCEI), and apply threshold determination for soybean area detection automation by GEographic Objects-based Image Analysis (GEOBIA).

2. MATERIAL AND METHODS

2.1. Subheadings

The study area comprises the state of Paraná in southern Brazil, among the geographic coordinates 22°29’ to 26°43’S and 48°20’ to 54°38’W (Fig. 1), with a total approximate area of 556,439.812 km². The altitude varies from 300 to 600 m. Three predominant climate types occur, according to Köppen-Geiger’s classification [9]: Cfa, Cfb and Cwa.

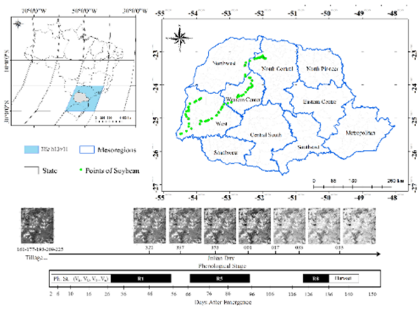

Figure 1. Study area (Paraná state) with its mesoregions, and temporal evolution of soybean stages compared to the composition time-series with vegetation indices.

2.2. Images and vegetation indices

We used images from the MODIS sensor, in which the values of the EVI (Enhanced Vegetation Index) and NDVI (Normalized Difference Vegetation Index) of the MOD13Q1 product (http://explorer.usgs.gov), tile H13V11, collection 5.0, were established from 16-day MODIS images (Huete et al., 1997) of the TERRA satellite, with spatial resolution of 250 m, downloaded from USGS LP-DAAC. The dates corresponding to the time-series used are presented in Table 1. We applied the orthogonal vegetation index PVI (Perpendicular Vegetation Index), which cancels the ground reflectance. To implement the PVI, a soil line regression based on red and near infrared spectral bands [10].

Julian Day	Date	Year
161	6/10	2011
177	6/26	2011
193	7/12	2011
209	7/28	2011
225	8/13	2011
321	11/17	2011
337	12/3	2011
353	12/19	2011
001	1/1	2012
017	1/17	2012
033	2/2	2012
049	2/18	2012

Table 1. Dates for composing the time-series used, from soil preparation until the final stage of the soybean crop.

For the development of an index that automatically identifies the soybean crop with a standard threshold slicing, the time-series over the phenological cycle and ground deletion for spectral interference were considered. In addition, the full exclusion of the soil by means of ground line or tasseled cap avoids the attenuations caused by no-tillage, where the biomass presents in this environment may influence data of vegetation indices mainly with regard the ratio between bands. With these values, PCEI (Perpendicular Crop Enhancement Index) were calculated [1].

The PCEI values range between -1 to 1, which provides a mechanism to verify greater positive differences between maximums and minimums of PVI and EVI observed over the soybean cycle. High values of PCEI indicate the soybean crop reflectance probability of the pixel. For PCEI values, oriented analysis in geo-object were used to indicate polygons of soybean crops. The slicing threshold was defined by data mining. For subsequent confirmation of "soybean" class, stipulated by PCEI, a comparison was performed between terrestrial analyses using soybean reflectance obtained by spectroradiometer and that obtained from MODIS images.

2.3. Oriented analysis and decision tree

GEOBIA and data mining and its integration with time-series MODIS images were applied to this analysis. The computing environments [eCognition 8.0 - Definiens Developer and WEKA the platform (Waikato Environment for Knowledge Analysis)] for each stage of GEOBIA were added to the data mining approach. eCognition 8.0 is software for image analysis based on objects [11], while WEKA is equipped with learning algorithms for data mining tasks [12]. To ensure homogeneity of the objects among the four indices utilized (NDVI [13], EVI [14], CEI [15] and PCEI [1].

Both classes of objects in the sample, soybean and non-soybean, were selected by checking the reflectance obtained using laboratory remote sensors and the threshold index CEI (≥ 0.28). In training, 260 objects were selected, of which 130 were soybean class and 130 to the non-soybean class. After the training set built up, the eCognition 8.0 platform was used to extract most representative attributes, which were spectral, spatial and textural [16].

2.4. Data analysis

To generate an independent sample set, we traveled through different regions of the state identifying areas that had soybean crops growing during the chosen period of MODIS image evaluation. A total of 172 crop reference spots were demarcated (Fig. 1). For this, we used a Trimble brand GPS receiver, GeoExplorer 2008 Series model, with L1 carrier and accuracy better than five meters after data correction.

Non-soybean spots were collected through visual interpretation of time-series MODIS images [17]. These 346 spots were distributed throughout the area and were randomly and independently generated [18]. The classification quality was quantitatively assessed using the Overall Accuracy coefficients (OA) and Kappa (κ); both extracted from a confusion matrix [18]. In addition, we extracted errors and accuracy under the producer and the user viewpoints [19].

Finally, the cultivated area estimated by remote sensing was compared to that provided by the Sistema IBGE de Recuperação Automática (SIDRA), from Instituto Brasileiro de Geografia e Estatística (IBGE) to verify the assessment of mapping for the 2011/2012 harvest. Currently, estimates carried out by IBGE are subjective, based on farmers and agriculture technicians' interviews.

3. RESULTS AND DISCUSSION

Owing to reflectance factor, soil lines were extracted at various times through MODIS time-series images, as presents different forms for the 2011/2012 harvest (Fig. 2). The difference between soil lines relies on two different harvests studied, as can be related to the time of imaging, soil tillage, cultivation and harvesting, or to other issues related to the spectral response composition received by the sensor, such as forests, urban centers, and various cultures not observed in the line of 45° to the x-axis. Soil influences the vegetation index determination, and this influence depends on its spectral properties and type and amount of vegetation present [20]. Thus, any type of soil in the canopy background

ISBN: 978-65-89159-04-9

will detract the index for certain vegetation [21]. All soil lines obtained were built from more than 65,000 random spots.

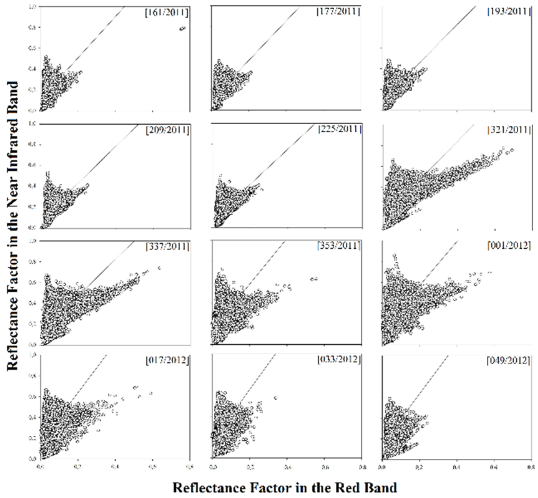

Figure 2. Soil line obtained from the ratio between bands of red and near infrared using TERRA/MODIS data on the 2011/12 harvest (Julian days).

The similarity between them is expected given that no abrupt large-scale soil occupation changes happen. The model for defining soybeans in images and acquiring a threshold for the PCEI is presented in the form of a decision tree for 2011/2012 harvest (Fig. 3).

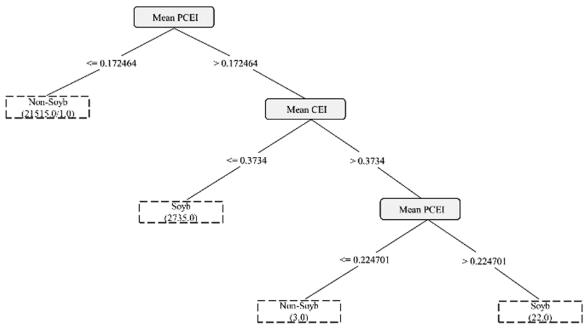

Figure 3. Decision tree generated by J48 algorithm for the 2011/2012 harvest.

It is noted that the first branch of the decision tree identified this process by setting a threshold for the PCEI vegetation index (Fig. 2), wherein the average of >0.172464 for values that vary from -1 to 1 can be considered soybean. This threshold is due to overlapping samples by the vegetation indices [15], in which a cut-off threshold of 0.28 was assigned to the last index for soybean. Dates on which summer crops are concentrated, i.e., Julian days 321, 001, 017, 033 and 049, the data trend toward the ground line is spaced and presents as vegetation behavior. This fact was confirmed by trend equations (Table 3); for vegetation in images following day 321 the R2 decreases, demonstrating that soil and vegetation are present.

Julian Day	Year	Trend Equation	R	R^2
161	2011	$\hat{y} = 2.1700x + 0.1282$	0.95	0.91^{**}
177	2011	$\hat{y} = 2.0449x + 0.1172$	0.94	0.89^{**}
193	2011	$\hat{y} = 1.9607x + 0.1093$	0.94	0.89^{**}
209	2011	$\hat{y} = 1.9533x + 0.1099$	0.93	0.88^{**}
225	2011	$\hat{y} = 1.9310x + 0,1088$	0.93	0.87^{**}
321	2011	$\hat{y} = 2.1764x + 0.1444$	0.92	0.84^{**}
337	2011	$\hat{y} = 2.3036x + 0.1534$	0.92	0.85^{**}
353	2011	$\hat{y} = 2.4553x + 0.1601$	0.94	0.88^{**}
001	2012	$\hat{y} = 2.5267x + 0.1702$	0.93	0.87^{**}
017	2012	$\hat{y} = 2.5658x + 0.1732$	0.93	0.86^{**}
033	2012	$\hat{y} = 2.5959x + 0.1730$	0.94	0.88^{**}
049	2012	$\hat{y} = 2.6111x + 0.1754$	0.94	0.88^{**}

Table 2. Julian days from the 2011/2012 harvest, trend equations, coefficient R and R2 for the relationship between bands of red and near infrared, as obtained by MODIS. *, ** significant at 0.05 and 0.01 probability, respectively, by t-test.

Thus, the angular coefficient values of the trend equations for Julian days 225, 321 and 017 are 1.9310, 2.1764 and 2.5658, respectively. This shows that there is a high probability of the presence of exposed soil on day 225. With the gradual increase in angular coefficient starting on day 017, there was full vegetative crop development until its decrease from drying on day 049, close to harvest (angular coefficient 2.6111). In 2011/2012, intense rainfall in the same period, causing harvest delay, providing greater angular coefficients and a high probability of vegetation presence, thereby confirming the findings of Goulart et al. [22].

However, objective methods also carry uncertainties, being confirmed only with reliable data that are represented in reality. Normally, there are over- or underestimates when compared to official data [2], [7], based on remote sensing techniques in the agricultural areas estimation. Fig. 4 shows the spatial distribution of soybean in the state of Paraná (Brazil) obtained by different classification systems. There is similarity between CEI and GEOBIA indices classification.

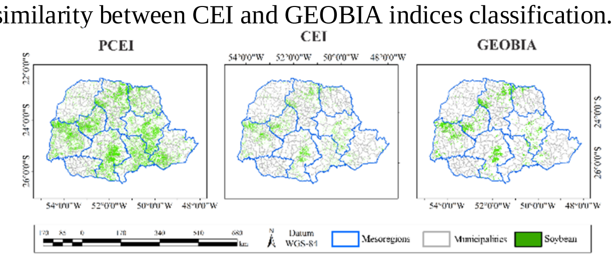

Figure 4. Spatial distribution of soybean-growing areas in the 2010/2011 harvest, according to various classification techniques.

Classifiers that underestimated the areas (Table 4) were of lower color intensity than the estimates of possible soybean areas. The methodological techniques demonstrated and assessed here present significant potential for soybean mapping, thus having the potential for use as a complement to official state and federal agencies in crop management, specializing in details by municipality and obtaining faster objective agricultural statistics. Kappa index, which evaluates the agreement or disagreement between classifications, varied between 0.40 (GEOBIA), 0.59 (PCEI), and 0.66 (CEI), which is from reasonably good to good quality (κ>0.21 and 0.81), respectively [23].

5. CONCLUSIONS

The mapping, discrimination and quantification of soybean areas in the state of Paraná, Brazil, is possible with the use of MODIS classifiers and images, in which the systematization presented satisfactory results. The development of the Perpendicular Crop Enhancement Index (PCEI) obtained satisfactory results in comparison to other vegetation indices used in the literature, mainly when minimizing soil reflectance supported by time-series, with a cutting threshold of 0.172464 obtained by means of a decision tree. Thus, values higher than this threshold represent the soybean crop when composing the images proposed here. This new approach can be routinely used by soybean producers.

6. REFERENCES

[1] C. A. da Silva Junior, M. R. Nanni, P. E. Teodoro, and G. F. C. Silva. Vegetation Indices for Discrimination of Soybean Areas: A New Approach. *agronomy journal*, v. 109: pp. 1331–1343, Jul. 2017.

[2] B. Wu and Q. Li. Crop planting and type proportion method for crop acreage estimation of complex agricultural landscapes. *international journal of applied earth observation and geoinformation*, v. 16: pp. 101–112, 2012.

[3] D. Peng, A. R. Huete, J. Huang, F. Wang, and H. Sun. Detection and estimation of mixed paddy rice cropping patterns with MODIS data. *international journal of applied earth observation and geoinformation*, v. 13: pp. 13–23, 2011.

[4] T. Bernardes, M. Adami, A. R. Formaggio, M. A. Moreira, D. de A. França, and M. R. de Novaes. Imagens mono e multitemporais Modis para estimativa da área com soja no estado de Mato Grosso. *pesquisa agropecuária brasileira*, v. 46: pp. 1530–1537, Nov. 2011.

[5] A. da S. J. Carlos, R. N. Marcos, E. T. Paulo, F. C. S. Guilherme, G. de L. Mendelson, and E. Marta. Comparison of mapping soybean areas in Brazil through perceptron neural networks and vegetation indices. *afr. j. agric. res.*, v. 11: pp. 4413–4424, 2016.

[6] 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 sensing of environment*, v. 114: pp. 2146–2159, 2010.

[7] Y. Peng and A. A. Gitelson. Remote estimation of gross primary productivity in soybean and maize based on total crop chlorophyll content. *remote sensing of environment*, v. 117: pp. 440–448, 2012.

[8] S. Lang and T. Blaschke. *Análise da paisagem com SIG*. 2009.

[9] C. A. Alvares, J. L. Stape, P. C. Sentelhas, J. L. de M. Gonçalves, G. Sparovek, and others. Köppen's climate classification map for Brazil. *meteorologische zeitschrift*, v. 22: pp. 711–728, 2013.

[10] M. R. Nanni and J. A. M. Demattê. Spectral Reflectance Methodology in Comparison to Traditional Soil Analysis. *soil science society of america journal*, v. 70: pp. 393–407, 2006.

[11] A. G. Definiens. *Definiens professional 5: reference book*, Definiens AG, Munich. Munich, 2006.

[12] J. Gao. *Digital analysis of remotely sensed imagery*. McGraw Hill, New York, 2009.

[13] Jr. Rouse J. W., R. H. Haas, J. A. Schell, and D. W. Deering. Monitoring Vegetation Systems in the Great Plains with ERTS, in *Proceeding of ERTS-1 Symposium*, 1974, v. 351: p. 309, p. 309.

[14] A. R. Huete, H. Q. Liu, K. Batchily, and W. van Leeuwen. A comparison of vegetation indices over a global set of TM images for EOS-MODIS. *remote sensing of environment*, v. 59: pp. 440–451, 1997.

[15] R. Rizzi, J. Risso, R. Dalla, V. Epiphanio, B. Friedrich, T. Rudorff, A. R. Formaggio, Y. E. Shimabukuro, and S. L. Fernandes. Estimativa da área de soja no Mato Grosso por meio de imagens MODIS, in *Anais XIV Simpósio Brasileiro de Sensoriamento Remoto*, 2009, pp. 387–394, pp. 387–394.

[16] T. Blaschke. Object based image analysis for remote sensing. *isprs journal of photogrammetry and remote sensing*, v. 65: pp. 2–16, 2010.

[17] R. Morais De Freitas, E. Arai, M. Adami, A. S. Ferreira, F. Y. Sato, Y. E. Shimabukuro, R. R. Rosa, L. O. Anderson, B. Friedrich, and T. Rudorff. Virtual laboratory of remote sensing time series: visualization of MODIS EVI2 data set over South America. *journal of computational interdisciplinary sciences*, v. 2: pp. 57–68, 2011.

[18] R. G. Congalton and K. Green. *Assessing the accuracy of remotely sensed data*, 2nd ed. CRC Press, Boca Raton, FL, 2008.

[19] J. Francisco Gonçalves Antunes, E. Mercante, J. César Dalla Mora Esquerdo, R. Augusto de Camargo Lamparelli, and J. Vieira Rocha. Estimativa de área de soja por classificação de imagens normalizada pela matriz de erros.

[20] L. S. Galvão and Í. Vitorello. Variability of Laboratory Measured Soil Lines of Soils from Southeastern Brazil. *remote sensing of environment*, v. 63: pp. 166–181, 1998.

[21] G. de Gois, R. C. Delgado, J. F. Oliveira-Junior, T. C. de O. Souza, and P. E. Teodoro. EVI2 index trend applied to the vegetation of the state of Rio de Janeiro based on non-parametric tests and Markov chain. *biosci. j.*, pp. 1049–1058, 2016.

[22] A. C. de O. Goulart, R. C. Delgado, J. F. de Oliveira Júnior, G. de Gois, and E. O. dos Santos. Vegetation indexes and rain time-spectrum relationship in the city of Rio de Janeiro. *revista de ciências agrárias*, v. 58: pp. 277–283, 2015.

[23] J. R. Landis and G. G. Koch. The Measurement of Observer Agreement for Categorical Data. *biometrics*, v. 33: pp. 159–174, 1977.