Sentinel-2B and Random Forest algorithm potential for sugarcane varieties identification

Daniel G. Duft[1], Ana C. S. Luciano[2], Peterson R. Fiorio[1]

[1]Department of Biosystems Engineering, College of Agriculture Luiz de Queiroz (ESALQ), University of São Paulo, Avenida Padua Dias, 11. P.O. Box 9, Piracicaba 13418-900, São Paulo, Brazil

[2]School of Agriculture Engineering (Feagri/Unicamp), Av. Cândido Rondon 501, Campinas 13083-875, Brazil

{daniel.duft,fiorio}@usp.br,ana.luciano@feagri.unicamp.br

Abstract. *The use of remote sensing for sugarcane varietal discrimination is based on fact that varieties have its own spectral pattern due to physical and morphological characteristics. The identification of sugarcane varieties using remote sensing helps to reduce the time taken to identify varieties in the field and non-certified varieties, moreover also to monitor the adoption of new varieties. Due to this scenario, the main objective of this paper is assessing the capability of Sentinel-2B satellite to identify sugarcane varieties in different dates of the year.*

Resumo. *O uso de sensoriamento remoto para diferenciar variedades de cana-de-açúcar é baseado no fato de que as variedades têm seu próprio padrão espectral devido as suas características físicas e morfológicas. A identificação de variedades de cana-de-açúcar através do sensoriamento remoto auxilia na redução do tempo de identificação em campo, na identificação de variedades não contratadas, além de monitorar a adoção de novas variedades. Devido a esse cenário, o principal objetivo deste artigo é avaliar a capacidade do satélite Sentinel-2B para identificar variedades de cana-de-açúcar em diferentes datas do ano.*

1. Introduction

Brazil is the world's largest sugarcane producer, with a cultivated area of 8.73 million of hectares in 2017/2018 [CONAB 2019] but it is estimated that less than 30% of this total area is declared and licensed. In this context, the knowledge of sugarcane varieties is important to improve the monitoring of sugarcane areas in Brazil.

The use of remote sensing for sugarcane varietal discrimination is based on that the variety need to have its own spectral pattern due to physical and morphological characteristics [Sandoval et al. 2011]. The varieties identification has expanded due to the availability of new satellite systems capable to record many bands of the spectrum, potentially identifying subtle changes in chlorophyll, water content, lignin/cellulose, nitrogen, and others [Galvão et al. 2005]. The identification of sugarcane varieties using remote sensing is needed to reduce the time taken to identify in the field and identify non-certified varieties also to monitor the adoption of new varieties [Fortes and Demate, 2005].

While remote sensing have primarily been used for identification of sugarcane vegetation status [Atzberger, 2013], for mapping sugarcane areas [Luciano et al., 2019; Rudorff et al., 2010], it could be also used in discrimination of sugarcane varieties [Abdel-Rahman and Ahmed, 2008]. Recently, the large range of satellite sensors, which many different spatial and temporal resolutions, such as Sentinel-2, help for monitoring of cropping practices, vegetation status, biomass and showed potential to classify crop varieties and to improve yield forecasting [Bégué et al., 2018].

Due to this scenario, the main objective of this paper is assessing the capability of Sentinel-2B satellite identify sugarcane varieties in different dates of the year.

2. Methodology

2.1. Study Area

The study area is a square inside the Sentinel-2B tile 22KHA. This square contains 9,070ha of sugarcane and 25 sugarcane varieties. All the areas are commercial plots and the varieties were determined by local inspection done by varieties specialists. Figure 1 shows the study area. The varieties inside this area are represented in 92% of Centre-South sugarcane total area. For this paper the varieties will be shown as capital letters because the target is the Sentinel-2B capability, not the variety factor.

2.2. Satellite Images

All images came from Sentinel-2B Level-2A, with radiometric and geometric corrections. It was used the reflectance from10 bands (Table 1) and 3 vegetation indices (NDBI, RENDWI and RENDVI – Table 2). It was used 3 images set from: 09/02/2019, 09/06/2019 and 13/08/2019. All reflectance data from sugarcane pixels were extracted and tabulated for statistical analysis. Inside the area there are all phenological phases of sugarcane represented.

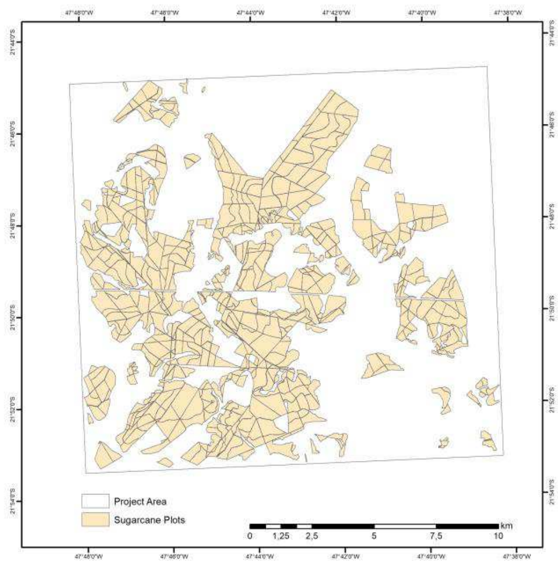

Figure 1. Study area

Table 1. Sentinel-2B bands specification

Sentinel-2B Bands	Central Wavelength (nm)	Resolution (m)
B2	490	10
B3	560	10
B4	665	10
B5	705	20
B6	740	20
B7	783	20
B8	842	10
B8A	865	20
B11	1610	20
B12	2190	20

Table 2. Vegetation Indices

Index	Algorithm	Source
NDBI	(B11-B8)/(B11+B8)	Zha et al (2001)
RENDVI	(B8-B6)/(B8+B6)	Gitelson et al (1994)
RENDWI	(B3-B5)/(B3+B5)	Gitelson et al (1994)

2.3. Statistical Learning

Were created three different regression models, for each year of available data (09/02/2019, 09/06/2019 and 13/08/2019). All models were calibrated and tested for sugarcane varieties prediction, using a Random Forest classification algorithm (RF)

[Breiman, 2001], under R software. RF is an ensemble learning algorithm method based on decision trees for classification. The RF parameter mtry (the number of variables randomly sampled as candidates at each split) was defined as default and the number of trees (ntrees) was defined after an evaluation of the statistical performances to classify sugarcane varieties equal to 500.

The training and testing process were made on independent datasets. Sampling was done using 70% of the dataset for training and 30% for testing, which represented approximately 569,488 samples for training and 244,082 samples for testing, at each model. The samples were selected randomly by varieties. The models were evaluated based on overall accuracy and kappa index.. The relative importance of the predictor variables was obtained based on Gini Index of the RF algorithm [Breiman, 2001].

3. Results and Discussion

Evaluating the results was possible to see that all three models showed good performance to classify sugarcane variety using Sentinel-2B images. Figure 2 shows the accuracy and kappa index of each model. It was possible to see that the best fit was the 13/08/2019 model with accuracy equal 0.86 and kappa index of 0.81. The Sentinel-2B and RF algorithm showed better results than previous studies with multispectral images to classify sugarcane varieties [Sandoval et al., 2005; Galvão et al., 2005; Fortes and Demate, 2005].

Looking the importance of each band and indices for the model's performance, it was possible to see that although Sentinel-2B has more bands between red and NIR (near infrared), the ones that best fit to the models was the SWIR (short-wave infrared) B11 and B12. Figure 3 shows the importance of variables to classify the sugarcane varieties and it's possible to conclude that for all models B11 and B12 were the most important bands.

For the best model, 13/08/2019, all varieties performed an accuracy over 70% but four of them more than 90% (Figure 4). Looking for relations between these varieties it was possible to conclude that varieties, classes A, R, O and D in Figure 4, are good to close canopy earlier than the others, but none of them belong to same family, unless three belongs to the same company owner. Figure 4 shows the accuracy for all 25 varieties (classes).

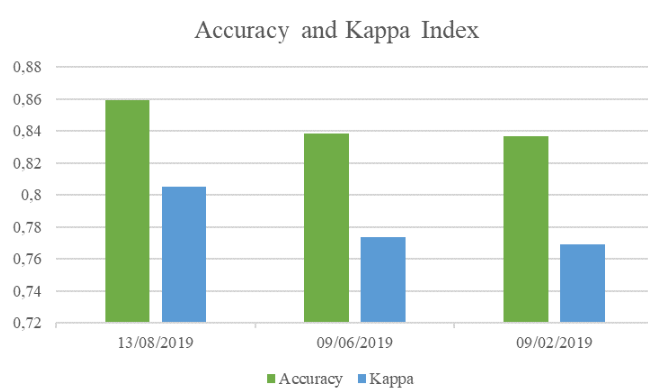

Figure 2. Models assessment

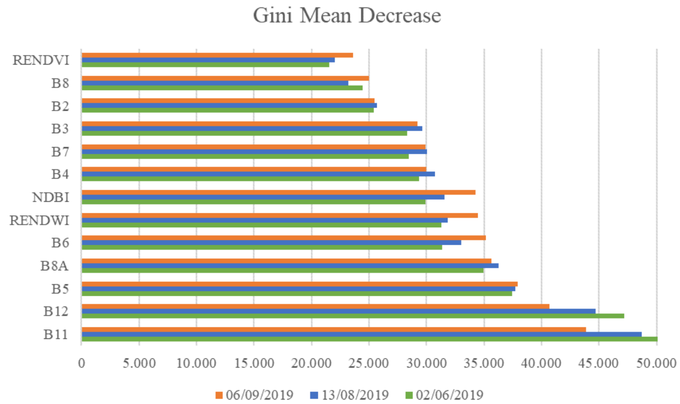

Figure 3. Importance of each factor for the models

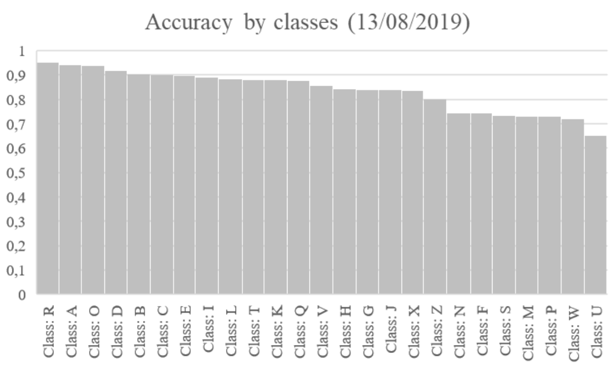

Figure 4. Accuracy by classes

4. Conclusion

The use of Sentinel-2B adn RF algorithm showed potential to classify sugarcane varieties. With a global accuracy of 86% and kappa index 81%, the 13/08/2109 model performed better than the others but an multitemporal approach maybe can bring even better results. Observing varieties, classes R, A, O and D had the accuracy over than 90% and it signalize that expanding study area looking for varieties that have big planted areas the global model accuracy can get higher. The last point is that Sentinel-2B bands 11 and 12 were the most important variables to the models, probably because in this wavelength most varieties have different reflectances.

References

Abdel-Rahman, E.M., Ahmed, F.B., (2008). The application of remote sensing techniques to sugarcane (Saccharum spp. hybrid) production: a review of the literature. Int. J. Remote Sens. 29, 3753–3767. https://doi.org/10.1080/01431160701874603

Atzberger, C., (2013). Advances in remote sensing of agriculture: Context description, existing operational monitoring systems and major information needs. Remote Sens. 5, 949–981.

Bégué et al., 2018. A. Bégué, D. Arvor, B. Bellon, J. Betbeder, D. de Abelleyra, R. Ferraz, V.Lebourgeois, C. Lelong, M. Simões, S.R. VerónRemote sensing and cropping practices: a review Remote Sens. (2018), p. 10

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

CONAB, C. N. de (2018) 'Acompanhamento da Safra Brasileira Cana-de- açúcar', V. 4 - SAFRA 2017/18 N.4 - Quarto levantamento | ABRIL 2018.

Fortes, C. and Demattê, J. (2006). Discrimination of sugarcane varieties using Landsat 7 ETM+ spectral data. International Journal of Remote Sensing, 27(7), pp.1395-1412.

Galvão, L., Formaggio, A. and Tisot, D. (2006). The influence of spectral resolution on discriminating Brazilian sugarcane varieties. International Journal of Remote Sensing, 27(4), pp.769-777.

Gitelson, A., and M. Merzlyak. "Spectral Reflectance Changes Associated with Autumn Senescence of Aesculus Hippocastanum L. and Acer Platanoides L. Leaves." Journal of Plant Physiology 143 (1994): 286-292.

Luciano, A.C. dos S., Picoli, M.C.A., Rocha, J.V., Duft, D.G., Lamparelli, R.A.C., Leal, M.R.L.V., Le Maire, G., 2019. A generalized space-time OBIA classification scheme to map sugarcane areas at regional scale, using Landsat images time-series and the random forest algorithm. Int. J. Appl. Earth Obs. Geoinf. 80, 127–136.

Rudorff, B.F.T., de Aguiar, D.A., da Silva, W.F., Sugawara, L.M., Adami, M., Moreira, M.A., 2010. Studies on the rapid expansion of sugarcane for ethanol production in São Paulo state (Brazil) using Landsat data. Remote Sens. 2, 1057–1076.

Sandoval, J.; Gonzales, C.; Murillo, A. (2011) Evaluation of Landsat 7 ETM+ Data for Spectral Discrimination and Classification of Sugarcane Varieties in Colombia. Journal of Agricultural Science and Technology: 101-107.

Zha, Y.; Gao, J.; Ni, S. Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. International Journal Of Remote Sensing, China, v. 24, n. 3, p.583-594, 23 out. 2001.