

PREDIÇÃO DA PRODUTIVIDADE MÉDIA DE CANA-DE-AÇÚCAR A PARTIR DE IMAGENS ORBITAIS DO SENSOR MSI/SENTINEL-2

João Pedro de Sousa Costa¹, Luiz Gabriel da Silva¹, Marcos Adami¹ e Ieda Del'Arco Sanches¹

¹ Programa de Pós-Graduação em Sensoriamento Remoto - PGSER, Divisão de Observação da Terra e Geoinformática - DIOTG, Instituto Nacional de Pesquisas Espaciais - INPE
Av. dos Astronautas, 1758, Jardim da Granja, São José dos Campos/SP, Brasil {joao.costa, luiz.silva, marcos.adami, ieda.sanches}@inpe.br

1. INTRODUÇÃO

Estimar a produtividade de culturas agrícolas é fundamental por diversos motivos, desde a definição de políticas por órgãos governamentais ao planejamento de ações de colheita e pós colheita por produtores agrícolas, cooperativas e *tradings*. A cana-de-açúcar é uma cultura agrícola de larga escala, sendo o Brasil o maior produtor mundial. O objetivo deste trabalho foi estimar a produtividade média de talhões de cana-de-açúcar utilizando séries temporais de imagens do sensor MSI/Sentinel-2. Foram utilizadas estimativas de produtividade geradas por um sensor instalado em uma colhedora em Nova Londrina-PR para construção da variável resposta. Ao total foram utilizadas informações de 42 talhões. As bandas das imagens foram utilizadas para construção de atributos, que foram utilizados para treinamento de um estimador de aprendizado de máquina, Gradient Boosting Regressor. Foram utilizados 10 talhões para validação, e em sete desses talhões o nível de erro relativo foi inferior a 5%.

Palavras-chave — *Sensoriamento remoto, cana-de-açúcar, aprendizado de máquina, gradient boosting*

ABSTRACT

Estimating the yield of agricultural crops is essential for several reasons, from the definition of policies by government agencies to the planning of harvest and post-harvest actions by farmers, cooperatives and trading companies. Sugarcane is a large-scale agricultural crop, with Brazil being the world's largest producer. The objective of this work was to estimate the average yield of sugarcane using a time series of images from the MSI/Sentinel-2. Yield estimates generated by a sensor installed in a harvester in Nova Londrina-PR were used for the construction of the target. Information from 42 areas was used. The bands of the images were used to construct attributes, which were used to train a machine learning estimator, Gradient Boosting Regressor. Ten areas were used for validation, and in seven of them the relative error level was less than 5%.

Keywords — *Remote Sensing, Sugar Cane, Machine Learning, Gradient Boosting*

Estimar a produtividade de uma cultura agrícola é de grande importância para previsão de safra, pois fornece informações essenciais para diferentes escalas de atuação, como ações de forma regional do governo em criação de políticas públicas de incentivo à produção, assim como a logística de escoamento, armazenamento e investimento. A previsão de safras é de interesse para os setores público e privado, dado que permite planejamento das exportações e importações, vendas futuras, abastecimento e distribuição, entre outras ações [1].

As estimativas de produtividade de cana-de-açúcar, por propriedade ou talhão, podem ser utilizadas para auxiliar o planejamento de operações das usinas, subsidiando tomadas de decisão ao longo da safra [2]. Algumas das principais operações são as operações de corte, carregamento e transporte (CCT), que garantem o abastecimento de cana nas moendas durante a maior parte do ano [3]. Os números de colhedoras, transbordos e caminhões necessários para cada frente de trabalho são determinados pelo responsável do transporte de forma empírica, geralmente baseado em conhecimentos práticos e subjetivos [3].

Estimar a safra de uma cultura é uma tarefa complexa, pelo fato de que muitos modelos convencionais de predição utilizam um grande número de parâmetros agrônômicos, relacionados à vegetação, ao solo, a fatores climáticos etc [1]. Devido à complexidade de obtenção desses dados para grandes extensões de área, torna-se complicada a implementação desses modelos. Uma das alternativas para estimar a produtividade de culturas agrícolas consiste na utilização das técnicas e dados de sensoriamento remoto (SR). Usualmente, dados de SR são obtidos pelo registro da reflectância, resultado da interação da radiação eletromagnética com a matéria. Espera-se que estes dados apresentem correlação com as informações agrônômicas, podendo ser utilizados em modelos de predição. Portanto, o objetivo deste trabalho foi estimar a produtividade média de talhões de cana-de-açúcar utilizando séries temporais de imagens do sensor MSI/Sentinel-2.

2. MATERIAL E MÉTODOS

O presente trabalho teve como área de estudo 42 talhões localizados no Município de Nova Londrina – PR (Figura 1). Todos os talhões foram cultivados com cana-de-açúcar

implantada em espaçamento duplo-alternado de 90x120 cm e com a variedade RB86-7515.

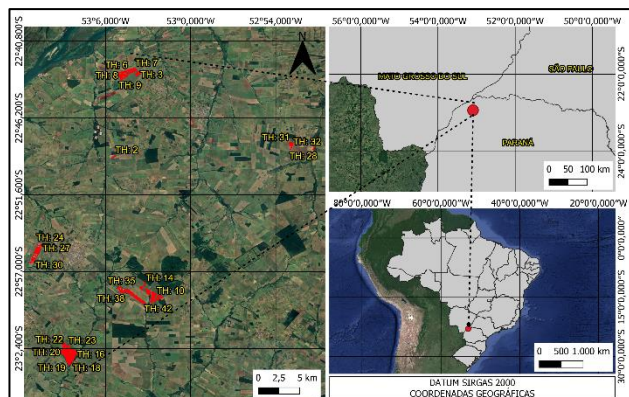


Figura 1. Área de estudo

As estimativas de produtividade de cana-de-açúcar foram feitas por um sensor instalado em uma colhedora CH670 com monitor de colheita *Harvest Monitor* e o monitor GS3 2630. Equipamentos da marca John Deere (Figura 2).



A) colhedora

B) monitor



C) dados brutos de produtividade

Figura 2. A) Colhedora CH670, B) monitor GS3 2630 e C) dados brutos de produtividade.

A metodologia para estimativa de produtividade consistiu das seguintes etapas:

- tratamentos dos dados de campo;
- aquisição das imagens orbitais;
- cálculo de índices espectrais;
- extração de atributos;
- parametrização do estimador.

2.1 Tratamento dos dados de campo

A partir dos dados brutos de produtividade, foi aplicado um pré-processamento de filtragem para remoção de outliers. A filtragem foi realizada em duas partes, na primeira foram

removidas estimativas evidentemente errôneas, por exemplo produtividades nulas e negativas, na segunda foi calculado limite inferior e superior de corte para remoção de produtividades abaixo e acima da média (Equação 1).

$$\text{Limite} = \text{Média} \pm 2.5 \text{ Desvio Padrão} \quad (\text{Equação 1})$$

Para compatibilizar a resolução espacial dos dados de produtividade coletados pela colhedora com a resolução espacial das imagens, e permitir avaliar a aplicação dos modelos, foi realizada a generalização dos dados de produtividade. Inicialmente foi gerada uma grade com resolução de 10x10 metros, compatível com a resolução espacial das imagens do sensor MSI, e para cada célula foi calculado o valor médio dos pontos de produtividade ali contidos (Figura 3).

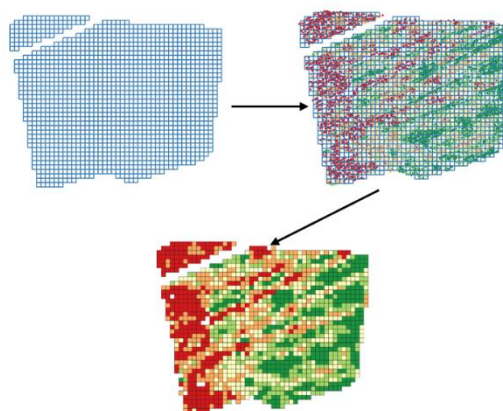


Figura 3. Generalização dos dados de produtividade.

2.2 Aquisição das imagens orbitais

A data das séries de imagens a serem adquiridas foi determinada a partir da data de colheita de cada talhão, sendo que foram adquiridos os dados referentes aos 12 meses retroativos da colheita. As bandas foram adquiridas em reflectância de superfície e as imagens com presença de nuvem foram removidas da série. Os dados espectrais foram processados pela plataforma *Google Earth Engine* (Tabela 1).

2.3 Cálculo de índices espectrais

Foram utilizados cinco índices espectrais, o Enhanced Vegetation Index (EVI) utilizado para reduzir as influências atmosféricas [4]. O First Modified Chlorophyll Absorption Ratio Index (MCARI1) foi desenvolvido para estimar o Índice de Área Foliar (IAF), seu modelo matemático foi elaborado para minimizar os efeitos da clorofila e com a integração do infravermelho próximo para aumentar a sensibilidade às mudanças de IAF [5]. O Normalized Difference Red-Edge index (NDRE) é um índice com relação com fotossíntese [6] e o Normalized Difference Vegetation Index (NDVI) com a parte estrutural das folhas [7]. O

Normalized Difference Water Index (NDWI) é utilizado para estimar o conteúdo de água [8] (Tabela 1).

Número da banda	Banda	Resolução espacial (m)	Comprimento de onda central (nm)
2	Blue (Azul)	10	490
3	Green (Verde)		560
4	Red (Vermelho)		665
8	NIR (infravermelho Próximo)		842
5	Red Edge 1	20	705
6	Red Edge 2		740
7	Red Edge 3		783
8a	Red Edge 4		865
11	SWIR 1		1610
12	SWIR 2		2190

Índice de vegetação	Modelo	Referência
Enhanced Vegetation Index (EVI)	$2.5 \frac{NIR - RED}{NIR + 6RED - 7.5BLUE + 1}$	[4]
First Modified Chlorophyll Absorption Ratio Index (MCARI1)	$1.2(2.5(NIR - GREEN) - 1,3(NIR - GREEN))$	[5]
Normalized difference red-edge index (NDRE)	$\frac{NIR - rededge}{NIR + rededge}$	[6]
Normalized Difference Vegetation Index (NDVI)	$\frac{NIR - RED}{NIR + RED}$	[7]
Normalized Difference Water Index (NDWI)	$\frac{NIR - SWIR}{NIR + SWIR}$	[8]

Tabela 1. Bandas do sensor MSI e Índices espectrais utilizados.

2.4 Extração de atributos

Como as datas de colheita dos talhões utilizados no estudo eram distintas, os atributos foram construídos de modo relativo. Dada uma data de colheita, foram descartadas as imagens contidas no intervalo de quatro semanas antecedente a data de colheita. Ou seja, foram utilizadas informações com até 28 dias anteriores à data de colheita. Foram utilizadas imagens das 40 semanas anteriores a essa data limite de corte. Esse processo é ilustrado na Figura 4. Essas 40 semanas foram divididas em grupos de quatro semanas, e os atributos, bandas e índices foram obtidos a partir da média das bandas e dos índices das imagens contidas em cada grupo. As bandas e índices médios de cada grupo foram utilizados como variáveis preditivas para o estimador.

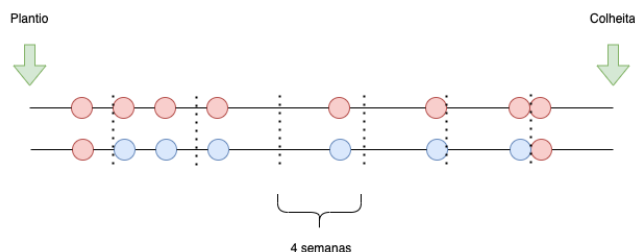


Figura 4. Ilustração do processo de extração de atributos. As bolas vermelhas representam as imagens coletadas e as bolas azuis representam as imagens efetivamente utilizadas.

2.5 Parametrização do estimador

Foi utilizado um algoritmo de aprendizado de máquina, *Gradient Boosting Regressor*, implementação LightGBM. Foram utilizadas as amostras de 30 talhões para treinamento e amostras de dois talhões para calibração da quantidade de estimadores (*Early Stopping*). Os 10 talhões restantes foram utilizados para validação. A métrica utilizada para otimização foi o erro quadrático médio, com taxa de aprendizagem de 0,01. O modelo resultante possui 429 estimadores (árvores), com no mínimo 100 amostras por folha.

3. RESULTADOS E DISCUSSÃO

Os resultados obtidos podem ser observados na Figura 5 e na Tabela 2. Com exceção de 3 talhões o erro relativo obtido entre a observação e a estimativa foi abaixo de 5%.

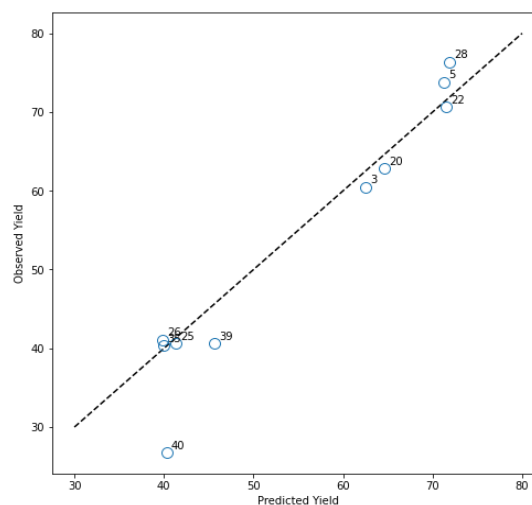


Figura 5. Comparação entre os valores preditos e observados de produtividade.

O talhão 40 foi o que apresentou maior discrepância. Não se descarta a possibilidade de erro de coleta do equipamento, pois ao se observar no mapa, esse talhão é o único de seu setor com esse nível de produtividade (aproximadamente 27 ton/ha). Os demais talhões do setor, por exemplo, os talhões 35 e 39 apresentam cerca de 40 ton/ha de produtividade.

Talhão	Produtividade Predita	Produtividade Observada	Erro (%)
3	62.5	60.4	3.4
5	71.3	73.8	3.4
20	64.6	62.8	2.8
22	71.5	70.7	1.2
25	41.3	40.6	1.8
26	39.8	41.1	3.0
28	71.9	76.3	5.9
35	39.9	40.3	1.1
39	45.6	40.6	12.3
40	40.3	26.8	50.5

Tabela 2. Estimativas de produtividade e seus respectivos erros relativos. Os maiores erros estão destacados em vermelho.

Spagnolo et al [9] obtiveram uma precisão satisfatória com diferenças com estimativa oficial do IBGE menores de 4% adaptando o modelo proposto por Doorenbos & Kassam [10] com o uso dos produtos MOD13Q1. Sugawara et al [11] utilizando da técnica de análise de regressão linear múltipla, apresentou em seu modelo uma concordância de 87% dos dados de validação para predição de produtividade de cana

Nesse estudo foram utilizados dados de SR de até um mês antes da colheita, com nível de erro aceitável, em 70% dos casos inferior ou compatível com os níveis de erro obtidos por métodos tradicionais, tal como pesagem de caminhões de transporte, após a colheita. O resultado é bastante satisfatório principalmente pelo fato de que não foram utilizadas informações específicas de cana-de-açúcar, tal como a quantidade de cortes, ou produtividade em anos anteriores ou produtividade regional, tampouco foram utilizados dados agrometeorológicos ou relativos ao solo.

4. CONCLUSÕES

O presente trabalho objetivou estimar a produtividade média de talhões de cana-de-açúcar utilizando somente dados espectrais de séries históricas de imagens do sensor MSI/Sentinel-2.

Utilizando de um estimador de aprendizado de máquina, *Gradient Boosting Regressor*, resultou em valores de produtividade menores que 5% de erro nos talhões 3, 5, 20, 22, 25, 26 e 35, sete dos dez talhões usados para validação do modelo. Com erros aceitáveis, o modelo em 70% dos casos apresentou erros inferiores ou compatíveis por métodos tradicionais, tal como pesagem de caminhões de transporte, após a colheita.

Assim, os resultados deste trabalho demonstraram que a utilização de séries temporais de bandas e índices espectrais de sensoriamento remoto pode ser útil para estimar com eficiência a produtividade média de talhões de cana-de-açúcar.

5. REFERÊNCIAS

- [1] Ippoliti-Ramilo, G. A., Epiphanyo, J. C. N., Shimabukuro, Y. E., Formaggio, A. R. Sensoriamento remoto orbital como meio auxiliar na previsão de safras. *Agricultura em São Paulo*, 46, 89-101. 1999.
- [2] Oliveira, R. A., Santos, R. S. D., Ribeiro, A., Zolnier, S., & Barbosa, M. H. (2012). Estimativa da produtividade da cana-de-açúcar para as principais regiões produtoras de Minas Gerais usando-se o método ZAE I. *Revista Brasileira de Engenharia Agrícola e Ambiental*, 16(5), 549-557.
- [3] Mundim, J. U. C. Uso de simulação de eventos discretos para o dimensionamento de frota para colheita e transporte de cana-de-açúcar. 2009. Tese de Doutorado. Universidade de São Paulo.
- [4] Huete, A. R.; Liu, H. Q.; Batchily, K.; Van Leeuwen, W. A comparison of vegetation indices global set of TM images for EOS-MODIS. *Remote Sensing of Environment*, v. 59, n. 3, p. 440-451, 1997. [http://dx.doi.org/10.1016/S0034-4257\(96\)00112-5](http://dx.doi.org/10.1016/S0034-4257(96)00112-5).
- [5] Haboudane, D.; Miller, J. R.; Pattey E.; Zarcotejada, P. J.; Strach AN, I. B. Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: modeling and validation in the context of precision agriculture. *Remote Sensing of Environment*, v. 90, n. 3, p.337-352, 2004. <http://dx.doi.org/10.1016/j.rse.2003.12.013>
- [6] Barnes, E.M.; T.R. Clarke; S.E. Richards; P.D. Colaizzi; J. Haberland; M. Kostrzewski;. Coincident detection of crop water stress, nitrogen status and canopy density using groundbased multispectral data. In: *Proceedings of the 5th International Conference on Precision Agriculture*, 2000.
- [7] Rouse, J. W.; Haas, R. H.; Schell, J. A.; Deering, D. W.; Harlan, J. C. Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation. Texas: Texas A&M University, 1974. (Type III final rep. NASA/GSFC).
- [8] Gao, Bo-Cai (1996). NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment*. 58 (3): 257–266. Bibcode:1996RSEnv..58..257G. doi:10.1016/S0034-4257(96)00067-3
- [9] Spagnolo, T. F. S., Souza, A. L. F., Fernandes, L. B., Oliveira, T. R. P. Modelo agrometeorológico espectral adaptado para estimativa de produtividade da cana-de-açúcar ao longo da safra. *Anais XVII Simpósio Brasileiro de Sensoriamento Remoto – SBSR*. 2012.
- [10] Doorenbos, J., Kassam, A. H. Yield response to water. Rome, FAO, 1979. 197p. (FAO Irrigation and Drainage Paper, 33).
- [11] Sugawara, L. M., Rudorff, B. F. T., Freitas, C. C., Picoli, M. C. A., & Adami, M. (2007). Estimativa de produtividade de cana-de-açúcar (*Saccharum officinarum* L.) por meio de técnica de análise de regressão linear múltipla. *Simpósio Brasileiro de Sensoriamento Remoto*, 13, 435-442.