

# AVALIAÇÃO DE MODELOS DE REGRESSÃO LINEAR MÚLTIPLA PARA ESTIMATIVA DA PRODUTIVIDADE DE CANA-DE-AÇÚCAR A PARTIR DE PARÂMETROS BIOMÉTRICOS E IMAGENS DO SENSOR MSI/SENTINEL-2

João Pedro de Sousa Costa <sup>1</sup> e Max Well de Oliveira Rabelo <sup>2</sup>

<sup>1</sup>Programa de Pós-Graduação em Sensoriamento Remoto - PGSER, Divisão de Observação da Terra e Geoinformática - DI-OTG, Instituto Nacional de Pesquisas Espaciais - INPE Av. dos Astronautas, 1758, Jardim da Granja, São José dos Campos/SP, Brasil; joao.costa@inpe.br;

<sup>2</sup> Instituto Federal de Goiás/Campus Goiânia/Departamento III-Coordenação de Geomatica, maxwell.rabelo@ifg.edu.br

## RESUMO

O Brasil é o maior produtor mundial de cana-de-açúcar, devido ao elevado valor agregado desta cultura, sendo necessário compreender sua produção para tomada decisão de produtores e órgãos governamentais. O presente trabalho teve como objetivo, integrar dados biométricos obtidos em campo e imagens orbitais do sensor MSI/Sentinel 2 e avaliar se é possível obter modelos mais precisos para a estimativa da produtividade da cana de açúcar a partir da combinação destas variáveis. Os melhores modelos avaliados pelo  $R^2$  ajustado obtiveram valor de 0,76, e o mesmo cenário apresentou o menor Erro Médio Absoluto de 2,94 ton.ha<sup>-1</sup>. Os modelos com biometria apresentaram melhores resultados nas métricas estatísticas da validação cruzada, possibilitando melhoria no desempenho dos modelos. Destaca-se a necessidade da obtenção de variáveis biométricas em campo com perspectiva de espacialização compatível com dados de sensoriamento remoto.

**Palavras-chave** — Sensoriamento remoto, cana-de-açúcar, produtividade, estimativa.

## ABSTRACT

*Brazil is the world's largest sugarcane producer, due to the high added value of this crop, making it necessary to understand its production for production by producers and government agencies. The objective, to integrate biometric data obtained in the field and orbital images from the MSISentinel 2 sensor and to evaluate it is possible to obtain more accurate models present a work of estimating the productivity of the sugar capacity from the variable variables. The best models evaluated by  $R^2$  obtained a value of 0.76, and the same environment presented the lowest Absolute Mean Error of 2.94 ton.ha<sup>-1</sup>. Models with biometrics showed better results in cross-validation measurements, which could improve the performance of the models. The need to adapt biometric variables in field of spatialization perspective with remote sensor data is highlighted.*

**Key words** — Remote sensing, sugarcane, productivity, estimation..

A cultura da cana-de-açúcar tem grande expressividade na produção agrícola brasileira, representando grande parcela da produção mundial, apenas em Julho de 2020 o Brasil exportou um volume de 3,487 milhões de toneladas [1]. A safra referente ao período de 2020/2021, representa uma área estimada a ser colhida de 8.409,8 mil hectares, com uma projeção de produtividade média de 76.348 Mg.ha<sup>-1</sup> [1]. Essa cultura movimenta anualmente no Brasil um valor bruto de mais de US\$ 100 bilhões, representando cerca de 2% do PIB, ou aproximadamente US\$ 40 bilhões [2]. Dessa forma, devido as extensas áreas de plantio e o elevado valor agregado na produção da cana-de-açúcar ao longo de toda safra, se faz necessário investigar em toda lavoura os resultados do manejo aplicado e identificar os problemas presentes na cultura [3].

Uma das soluções para o levantamento de informações espaciais dos cultivos nas áreas de cana-de-açúcar é a utilização das técnicas de Sensoriamento Remoto (SR) como os Índices de Vegetação (IV) para estimativa da produtividade das culturas agrícolas em geral. De acordo com [3], o uso de IV de forma isolada não permite modelar precisamente informações como produtividade. Para tanto faz-se necessário correlacionar esses índices com outros parâmetros da cultura.

Nos dias atuais, diversos trabalhos têm sido desenvolvidos com o objetivo de correlacionar as informações biofísicas da cultura com a produtividade utilizando imagens de SR. Pode-se citar como exemplos, [4], que relacionou NDVI obtidos por imagens do Sensor OLI do Landsat 8 para estimar a biomassa de cana-de-açúcar, obtendo um modelo com maior significância estatística aos 8 meses de idade após a colheita. [5], que utilizaram o sensor de reflectância de dossel para estimar a produtividade na cana-de-açúcar, mostrando que essa pode ser uma ferramenta útil para estimar esses dados, além de apresentar um modelo generalizado baseado no NDVI para estimar a produtividade. Neste contexto, com uso dos dados biométricos da cana-de-açúcar adquiridos em campo e com os dados adquiridos a partir de imagens de Sensoriamento remoto em nível orbital, o presente trabalho revê por objetivo obter modelos de regressão para estimativa da produtividade da cana de açúcar.

## 1. INTRODUÇÃO

## 2. MATERIAL E MÉTODOS

O presente trabalho foi realizado no Talhão 10, da Fazenda Miss Brasil, localizado no Município de Nova Londrina – PR, como mostra a Figura 1. O talhão possui uma área de 15,578 hectares cultivados com cana-de-açúcar implantada em espaçamento duplo-alternado de 90x120 cm, com cana soca de quarto corte da variedade RB86-7515 implantada em 21/11/2016.

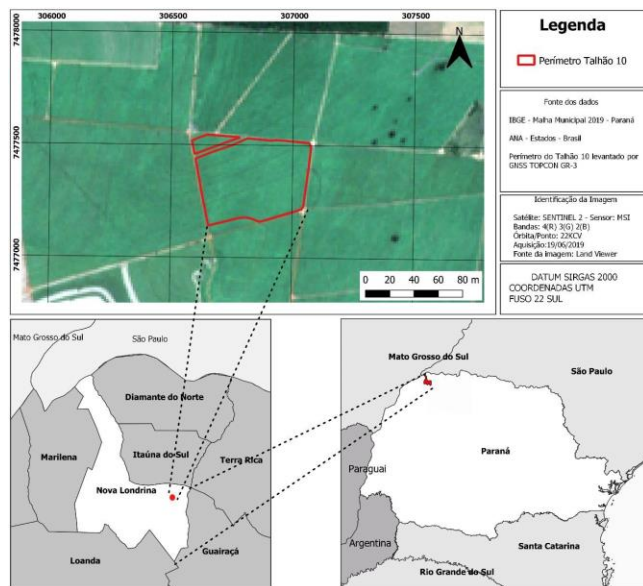


Figura 1. Localização da Área de Estudo - Talhão 10.

Para a execução desse estudo foram utilizados os dados de colheita de cana de açúcar adquiridos pela colhedora CH670 com monitor de colheita Harvest Monitor e o monitor GS3 2630, equipamentos da marca John Deere. A partir desses dados brutos de produtividade, foi aplicado um pré-processamento de filtragem para remoção de outliers utilizando do software *Mapfilter*.

Foram utilizados levantamento de campo de 12 pontos amostrais com dados sobre os parâmetros biométricos da cana-de-açúcar referente ao dia 19/06/2020, elaborado pela administração da Fazenda Miss Brasil. No referido relatório constam os seguintes dados: o diâmetro médio do colmo (DMC), o número de perfis industrializáveis (NPI), estatura média dos colmos (EMC) e Tonelada de Cana por Hectares (TCH), como apresenta a Figura 2.

Alem das imagens orbitais do sensor MSI do satélite Sentinel 2, bandas 490 nm (B2), 560 nm (B3), 665 nm (B4) e 842 nm (B8), com resolução espacial de 10 metros; e a banda 705 nm (B5), com resolução espacial de 20 metros, todas imagens em valores de reflectância de superfície. Todas as imagens foram obtidas na data de 19/06/2020, ou seja, a data do levantamento de campo que contém os parâmetros biométricos.

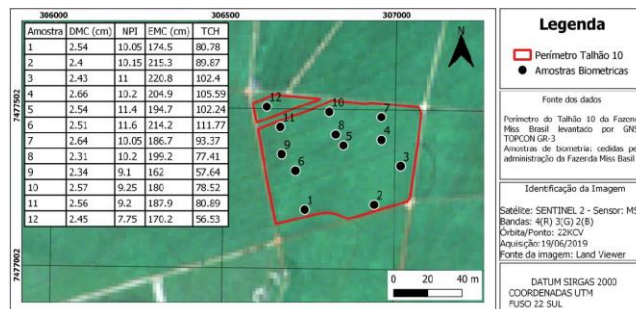


Figura 2. Distribuição das amostras no talhão 10.

Após a obtenção das imagens foi realizado o cálculo dos 15 Índices de Vegetação (IV) mostrados na Tabela 1, utilizando o *software Qgis* aplicando a técnica de álgebra de mapas.

Índice de vegetação	Modelo	Referência
Canopy chlorophyll content index (CCCI)	NDRE NDVI	[6]
Enhanced Vegetation Index (EVI)	$2.5 \frac{NIR - RED}{NIR + 6RED - 7.5BLUE + 1}$	[7]
Green Chlorophyll Index (GCI)	$\frac{NIR}{GREEN} - 1$	[8]
Green-Red Vegetation Index (GRVI)	$\frac{GREEN - RED}{GREEN + RED}$	[9]
First Modified Chlorophyll Absorption Ratio Index (MCARI1)	$1.2(2.5(NIR - GREEN) - 1.3(NIR - GREEN))$	[10]
Second Modified Chlorophyll Absorption Ratio Index (MCARI2)	$1.5 \frac{2.5 NIR - RED - 1.3(NIR - GREEN)}{2NIR + 1 - 6NIR - 5RED - 0.5}$	[11]
Modified SAVI (MSAVI)	$\frac{2NIR + 1 - \sqrt{2NIR + 1^2 - 8(NIR - RED)}}{2}$	[12]
Modified SR (MSR)	$\frac{\frac{NIR}{RED} - 1}{\frac{NIR}{RED} + 1}$	
Normalized difference red-edge index (NDRE)	$\frac{NIR - rededge}{NIR + rededge}$	[13]
Normalized Difference Vegetation Index (NDVI)	$\frac{NIR - RED}{NIR + RED}$	[14]
Optimized SAVI (OSAVI)	$1 + 0.16 \frac{NIR - RED}{NIR + RED + 0.16}$	[15]
Renormalized Difference Vegetation Index (RDVI)	$\frac{NIR - RED}{NIR + RED}$	[16]
Soil Adjusted Difference Vegetation Index (SAVI)	$1 + L \frac{NIR - RED}{NIR + RED + L}$	[17]
Structural Insensitive Pigment Index (SIPI)	$\frac{NIR - BLUE}{NIR + RED}$	[18]
Simple Ratio (SR)	$\frac{NIR}{RED}$	[19]

Em que: BLUE, valores de reflectância em 490 nm; GREEN, valores de reflectância em 560 nm; RED, valores de reflectância em 665 nm; NIR, valores de reflectância em 842 nm; rededge, valores de reflectância em 705 nm; L (SAVI), valor de 0,5.

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

Com os dados vetoriais dos parâmetros biométricos, foi criada uma zona de influência com um raio de 5 metros no entorno das amostras, na qual todos os dados de produtividade contidos foram associados aos respectivos valores de biometria. Ainda com a localização das amostras de biometria, foi gerada uma zona de influência com raio de 10 metros, e para cada dado de produtividade selecionado na etapa anterior foi associado o valor médio dos pixels de IV cujo centroide estava a menos de 10 metros da localização da amostra de biometria (Figura 3). O arquivo vetorial gerado com as informações extraídas foi exportado para o formato de tabela. Esta tabela por sua vez foi manipulada do *software* WEKA para geração dos modelos de regressão.

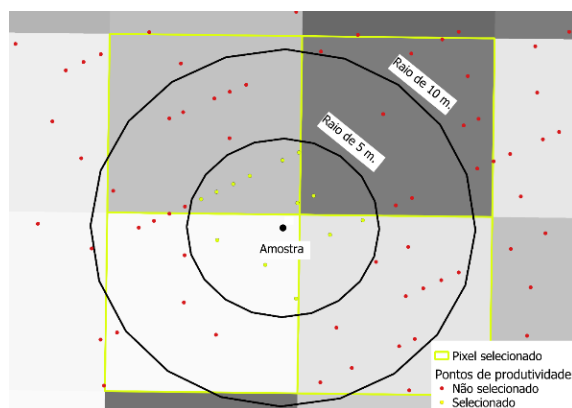


Figura 3 - Raios de influência.

No *software* Weka, para desenvolvimento dos modelos de Regressão Linear Múltipla foi adotado o algoritmo de *Greedy* para seleção dos atributos que serão associados às variáveis independentes do modelo. Esse algoritmo tem como objetivo reduzir os efeitos de colinearidade entre as biometrias e os IV, a fim de evitar redundância ou até mesmo à inserção de erros nos modelos de estimativa.

Para ajuste do modelo de regressão, foi realizada a *Leave-One-Out Cross-Validation* (LOOCV), uma validação cruzada na qual o modelo é recalculado removendo uma amostra por vez, obtendo um valor predito para a amostra removida, até que todas as possíveis combinações tenham sido realizadas. Foram considerados 16 cenários para criação dos modelos. Todos os cenários contêm como variáveis os 15 índices de vegetação mencionados anteriormente, e como fator de variação entre cenários, a presença de um, dois, três ou quadro variáveis biométricas, com suas respectivas possibilidades de combinação.

Para análise da qualidade dos modelos de regressão obtidos em cada cenário, foram utilizadas as seguintes métricas estatísticas: coeficiente de correlação (R), coeficiente de determinação (R<sup>2</sup>) e coeficiente de determinação ajustado para múltiplas regressões (R<sup>2</sup> ajustado). Além disso, através da LOOCV foram calculados o erro médio absoluto e a raiz do erro quadrático médio (RMSE). Foi utilizado o teste estatístico *Shapiro-Wilk* no

programa *Rstudio* para avaliação da normalidade dos resíduos.

### 3. RESULTADOS E DISCUSSÃO

A Tabela 2, apresenta as métricas estáticas da validação cruzada de cada cenário. Os melhores modelos avaliados pelo R<sup>2</sup> ajustado foram dos cenários DMC – NPI, DMC – TCH e EMC - TCH – DMC, com um R<sup>2</sup> ajustado de 0,76, sendo que esse último cenário apresentou o menor Erro Médio Absoluto de 2,94 ton.ha<sup>-1</sup>. O maior coeficiente de correlação de 0,88 foi do cenário DMC - NPI - EMC/DMC - NPI - EMC – TCH que também apresentou o menor RMSE de 3,62 ton.ha<sup>-1</sup>. Os melhores modelos apresentaram algum parâmetro biométrico, indo de encontro com os estudos realizados por Rocha (2018). Segundo o autor, a integração de parâmetros biométricos e IV pode ser utilizada para criação de modelos preditivos mais assertivos.

Cenários	R <sup>2</sup>	R <sup>2</sup> a *	C*	EMA*	RMSE
Sem Biometrias/DMC	0,59	0,57	0,73	4,22	5,16
NPI	0,76	0,74	0,84	3,23	4,04
EMC	0,74	0,73	0,84	3,23	4,09
TCH	0,74	0,72	0,83	3,41	4,14
DMC - NPI	0,78	0,76	0,87	3,05	3,73
DMC - EMC	0,77	0,75	0,85	3,12	3,92
DMC - TCH	0,78	0,76	0,87	3	3,7
NPI - EMC	0,72	0,71	0,82	3,43	4,32
NPI - TCH	0,66	0,65	0,79	3,91	4,63
EMC - TCH	0,74	0,73	0,83	3,41	4,23
DMC - NPI - EMC/DMC - NPI - EMC - TCH	0,73	0,72	0,88	2,95	3,62
DMC - NPI - TCH	0,71	0,7	0,83	3,36	4,23
NPI - EMC - TCH	0,75	0,73	0,85	3,08	3,93
EMC - TCH - DMC	0,78	0,76	0,87	2,94	3,7

Tabela 2. Métricas estáticas dos dados de treinamento. C: coeficiente de correlação. R<sup>2</sup>a: R<sup>2</sup> ajustado. EMA: Erro médio absoluto

Os cenários Sem Biometrias e DMC obtiveram os piores resultados dentro da validação cruzada, o modelo gerado por esse cenário não incluiu o parâmetro DMC na estimativa tendo como variáveis somente os índices de vegetação. O cenário TCH também não apresentou biometrias como parâmetros para estimativa, contudo, seu modelo apresentou melhores resultados na validação cruzada, apresentando assim que somente uso de IV foi possível ter uma estimativa assertiva contradizendo o que foi apresentado anteriormente pelos resultados dos melhores modelos e por [3].

O teste de normalidade *Shapiro-Wilk*, que foi aplicado no conjunto dos valores dos resíduos dos modelos nos dados

de treinamento, exceto pelo cenário NPI – TCH, todos os modelos apresentaram normalidade em nível de significância de 0,05, sendo, portanto, modelos que atendem a premissa da regressão linear multivariada, que determina que os resíduos devem apresentar distribuição normal.

#### 4. CONCLUSÕES

Os parâmetros biométricos que mais contribuíram para determinação da produtividade foram a Estatura Média dos Colmos (EMC) e o Número de Perfilhos Industrializáveis (NPI). Entre os índices de vegetação, os mais importantes foram CCCI e MCARI1, mostraram-se ser índices bastante promissores não só para estimativa de produtividade mais também para estimativa da biometria EMC, portanto, para futuros estudos de estimativa da cultura da cana de açúcar esses IV podem contribuir nos modelos de predição.

Os melhores modelos avaliados pelo  $R^2$  ajustado obtiveram valor de 0,76, e o mesmo cenário apresentou o menor Erro Médio Absoluto de 2,94 ton.ha<sup>-1</sup>. O uso de dados biométricos possibilitou melhoria no desempenho dos modelos na validação cruzada. Portanto, destaca-se a necessidade de maiores estudos com relação à obtenção de variáveis biométricas em campo com perspectiva de espacialização compatível com a escala dos dados de sensoriamento remoto orbital.

#### 5. REFERÊNCIAS

[1] CONAB. Acompanhamento da safra brasileira de cana-de-açúcar - V. 7 - SAFRA 2020/21 N.2 - Segundo levantamento - Agosto 2020

[2] UNICA. União da Indústria de Cana-de-Açúcar. Balanços de atividades 2012/13 a 2018/10. Disponível em: <https://www.unica.com.br/wp-content/uploads/2019/06/Relatorio-Atividades-201213-a-201819.pdf>. Acesso: 18 set. 2020

[3] ROCHA, Murillo Grespan. Influência de variáveis agrônomicas na produtividade de cana-de-açúcar e potencial uso de sensoriamento proximal. 2018. 1 recurso online (93 p.). Dissertação (mestrado) - Universidade Estadual de Campinas, Faculdade de Engenharia Agrícola, Campinas, SP. Disponível em: <<http://www.repositorio.unicamp.br/handle/REPOSIP/331897>>. Acesso em: 05 set. 2020

[4] OLIVEIRA, Gildriano Soares de. Produtividade de biomassa de cana-de-açúcar em função dos índices de vegetação utilizando técnicas de sensoriamento remoto. 2015. vii, 65 p. Dissertação (mestrado) - Universidade Estadual Paulista Júlio de Mesquita Filho, Faculdade de Ciências Agrárias e Veterinárias de Jaboticabal, 2015. Disponível em: <<http://hdl.handle.net/11449/136733>>. Acesso: 03 set. 2020

[5] AMARAL, L. R.; PORTZ, G.; ROSA, H. J. A.; MOLIN, J. P. Estimativa de produtividade de cana-de-açúcar a partir de dados de sensor de refletância do dossel. In: CONGRESSO BRASILEIRO DE AGRICULTURA DE PRECISÃO, 5., 2012, Ribeirão Preto. Anais... Jaboticabal: SBEA, 2012.

[6] LONG, D. S., EITEL, J. U. H., and HUGGINS, D. R. Assessing nitrogen status of dryland wheat using the canopy chlorophyll content index. *Crop Management*. Online. doi:10.1094/CM-2009-1211-01-RS.

[7] 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).

[8] GITELSON, A. A.; GRITZ, Y.; MERZLYAK, M. N. Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *Journal of Plant Physiology*, v. 160, n. 3, p. 271-282, 2003.

[9] TUCKER, C. J. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, v. 8, n. 2, p. 127-150, 1979. [http://dx.doi.org/10.1016/0034-4257\(79\)90013-0](http://dx.doi.org/10.1016/0034-4257(79)90013-0)

[10] 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>

[11] QI, J.; CHEHBOUNI, A.; HUETE A.R.; KERR, Y. H.; SOROOSHIAN, S. A modified soil adjusted vegetation index. *Remote Sensing Environmental*, v. 48, p. 119-126, 1994.

[12] 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 ground-based multispectral data. In: Proceedings of the 5th International Conference on Precision Agriculture, 2000.

[13] 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).

[14] SANTI, A. L., Giotto, E., Sebem, E., & Amado, T. J. C. (2016). Agricultura de Precisão no Rio Grande do Sul. Santa Maria: CESPOL.

[15] RONDEAUX, G.; STEVEN, M.; BARET, F. Optimization of soil-adjusted vegetation indices. *Remote Sensing of Environment*, v. 55, n. 2, p. 95-107, 1996.

[16] ROUJEAN, J. L.; BREON, F. M. Estimating PAR absorbed by vegetation from bidirectional reflectance measurements. *Remote Sensing of Environment*, v. 51, n. 3, p. 375-384, 1995

[17] HUETE, A. R. A soil adjusted vegetation index (SAVI). *Remote Sensing of Environment*, v. 25, n. 3, p. 295-309, 1988. [http://dx.doi.org/10.1016/0034-4257\(88\)90106-X](http://dx.doi.org/10.1016/0034-4257(88)90106-X)

[18] PEÑUELAS, J.; BARET, F.; FILELLA, I. Semi-empirical indices to assess carotenoids/chlorophyll a ratio from leaf spectral reflectance. *Photosynthetica* v. 31, p. 221-230, 1995.

[19] BIRTH, G. S.; MCVEY, G. Measuring the colour of growing turf with a reflectance spectrophotometer. *Agronomy Journal*, v. 60, n. 6, p. 640-643, 1968. <http://dx.doi.org/10.2134/agronj1968.00021962006000060016x>