

AVALIAÇÃO DE METODOLOGIA DE ESTIMATIVA DE PRODUTIVIDADE DE SOJA POR MEIO DE AERONAVE NÃO TRIPULADA E TÉCNICA DE APRENDIZADO DE MÁQUINA BASEADA EM REGRESSÃO

EVALUATION OF SOYBEAN PRODUCTIVITY ESTIMATION METHODOLOGY BY UNMANNED AERIAL VEHICLE AND REGRESSION- BASED MACHINE LEARNING TECHNIQUE

Christopher Djonny Pereira Prestes¹ 

Alaine Margarete Guimarães² 

David Patrick Zampier³ 

Gislaine Gabardo⁴ 

Eduardo Fávero Caires⁵ 

Bruno Ricardo Asato⁶ 

Resumo: O uso de índices de vegetação (VI) em um ambiente adequado pode favorecer a identificação de diferentes tipos de vegetação e a detecção de informações relacionadas a uma cultura, como a estimativa de produtividade. Esta estimativa é especialmente importante quando realizada nos estágios iniciais da cultura permitindo obter informações relevantes que contribuam para aumentar a eficiência em seu desenvolvimento. Considerando a complexidade de cada ambiente e os desafios de determinar o IV mais adequado, este trabalho teve como objetivo avaliar uma nova metodologia para estimar os dados de produtividade da soja. A metodologia proposta baseia-se em imagens obtidas por meio de veículos aéreos não tripulados (VANT) e técnicas de aprendizado de máquina. Foram aplicados algoritmos de regressão nos dados espectrais extraídos de imagens com resolução de 15 cm / pixel obtidas com câmeras RGB e NIR acopladas no VANT. Como resultado, o Índice de Resistência Atmosférica Visível (VARI) baseado em RGB apresentou os melhores valores de correlação

¹ Mestrando em Computação Aplicada, UEPG, Ponta Grossa-PR, cdpprestes@gmail.com.

² Doutora em Computação Aplicada, UEPG, Ponta Grossa-PR, alainemg@uepg.com.

³ Engenheiro de Computação, UEPG, Ponta Grossa-PR, davidzampier@gmail.com.

⁴ Doutora em Agronomia, UEPG, Ponta Grossa-PR, gislainegabardo2007@yahoo.com.br.

⁵ Doutor em Agronomia, UEPG, Ponta Grossa-PR, efcaires@uepg.com.

⁶ Mestrando em Computação Aplicada, UEPG, Ponta Grossa-PR, asatobruno@gmail.com.

entre os cinco IVs avaliados. No entanto, o uso dos canais RGB com seus valores originais sem a realização de combinações entre eles apresentou o melhor resultado em relação a todos os índices avaliados. O fato de todos os testes mostrarem que os valores das bandas RGB apresentam maior correlação com a produtividade da soja apresentam uma importante contribuição, demonstrando que uma câmera de menor custo pode ser usada para fornecer maior eficiência.

Palavras-chave: Estimativa de produtividade. VANT. Aprendizado de máquina. Índice de Vegetação. Dados Espectrais.

Abstract: Using Vegetation Indices (VI) in a suitable environment may favor to identify different kinds of vegetation and to detection of information related to a crop, like the Yield estimative. This estimation is especially important when carried out in the growth crop initial stages allowing relevant informations to be obtained contributing to increase the efficiency in its development. Considering the complexity of each environment and the challenges of determine the most appropriate VI, this work aimed to evaluate a new methodology to estimate soybean yield data. The proposed methodology is based on images obtained from Unmanned Aerial Vehicle (UAV) and machine learning techniques. Regression algorithms were applied in the spectral data extracted from images with a 15 cm/pixel resolution obtained with RGB and NIR cameras coupled in the UAV. As result, the RGB-based Visible Atmospherically Resistant Index (VARI) presented the best correlation values among the five evaluated IVs. However, the use of the RGB channels with their original values without the realization of combinations between them presented the best result in relation to all indexes evaluated. The fact that all the tests show that the values of the RGB bands present a greater correlation with soybean yields ia an important contribution, demonstrating that the lower cost camera can be used to provide higher efficiency.

Keywords: Yield estimative. UAV. Machine learning. Vegetation Index. Spectral data

1 INTRODUÇÃO

O Sensoriamento Remoto (SR) de uma área permite que sejam obtidas informações sobre objetos na superfície terrestre sem que haja contato físico entre eles e o sensor (Figueiredo, 2007). Na agricultura, esse processo pode ser realizado por meio da correlação entre os dados obtidos em campo de uma determinada cultura e os valores de refletância de bandas espectrais extraídos da imagem do sensor.

A combinação dos dados dessas bandas pode proporcionar uma melhor avaliação das propriedades de refletância obtidas, ajudando, por exemplo, a realçar a vegetação em uma determinada área. As bandas RGB (*Red, Green e Blue*) e NIR (*Near Infrared*) correspondem a alguns dos intervalos do espectro eletromagnético que comumente são utilizados nessas combinações, definidas como índices de vegetação (IV) (Bernardi et al., 2014).

Segundo (Xue e Su, 2017), diferentes ambientes têm suas próprias características variáveis e complexas, que precisam ser consideradas quando se usam diferentes IVs, e essa escolha precisa ser feita com cautela. Em vista disso, pesquisas são realizadas para avaliar o desempenho de IVs para determinadas condições e objetivos. Wu (2014) avaliou o desempenho de quinze IVs em diversos tipos de biomas utilizando imagens de satélite. Foi constatado que alguns índices são mais aplicáveis ao estudo de áreas densamente vegetadas, por exemplo, florestas e áreas de cultivo, ao invés de áreas secas.

Liu (2006) destaca que a utilização de um IV, quando aplicado a um ambiente adequado, favorece a distinção de diferentes tipos de vegetação e detecção de dados e informações relacionados aquela cultura. A estimativa de produtividade de uma cultura é um exemplo de informação que pode ser obtida por meio de dados espectrais, uma vez que tem forte relação com o vigor das plantas. Essa estimativa é importante principalmente quando realizada nos estádios iniciais de crescimento da planta, porque permite que se obtenha informações antecipadas que podem resultar em maior segurança e eficiência em seu desenvolvimento. Os estudos de Panda et al. (2010) no

milho, e Rodrigues (2015) no trigo, são exemplos que utilizaram IVs para correlacionar com valores de produtividade de grãos.

Considerando a complexidade de cada ambiente e os desafios da escolha de um IV apropriado, esse estudo tem como objetivo principal testar e avaliar o desempenho de cinco IVs quando correlacionados com dados de produtividade de soja, produto que possui grande importância na agricultura brasileira, sobretudo no Estado do Paraná (IBGE, 2014). Também serão avaliados o desempenho das bandas RGB e NIR e a influência do período em que foi coletado os dados em relação a semeadura.

2 DESENVOLVIMENTO

2.1 Bases de Dados

Para esse estudo foram utilizadas bases contendo dados extraídos de imagens com resolução de 15 cm/pixel de uma área experimental de cultivo de soja situada na Fazenda Escola da Universidade Estadual de Ponta Grossa (UEPG) localizada na cidade de Ponta Grossa – PR. As imagens foram obtidas por meio de um VANT modelo eBee em doze voos entre os meses de dezembro a julho de 2017. As câmeras utilizadas foram uma Canon S110 NIR (12 megapixels) e uma Sony Cybershot WX RGB (18.2 megapixels). O processo de obtenção e processamento dessas imagens estão descritos no estudo de Zampier et al. (2017), que teve como objetivo analisar a influência da resolução espacial de imagens com resoluções de 3,4 e 15 cm/pixel, na estimativa da ferrugem asiática e do oídio na soja. A Figura 1 apresenta o Mapa de Refletância da área do estudo gerado a partir das imagens da camera RGB.

Figura 1 – Mapa de Refletância da área de estudo no dia 21/02/2017 gerado a partir das imagens da câmera RGB



Fonte: Zampier et al., 2017

Foram selecionadas quatro bases com dados correspondentes a duas parcelas da área total desse estudo, ambas semeadas no dia 24/10/2016 e contendo dados espectrais das bandas RGB e NIR extraídos em diferentes dias durante o período de 27/01/2017 a 13/03/2017. As bases também possuem valores de produtividade coletados em campo contabilizados por tratamento da área. A Tabela 1 apresenta a descrição de cada uma das bases utilizadas nesse estudo e o valor de DAS (dias após a semeadura), que é o tempo decorrido em dias desde a semeadura até a execução do voo. Quanto menor for o DAS mais tempo o agricultor terá para corrigir um eventual problema identificado antes da colheita.

Tabela 1 – Descrição das bases de dados utilizadas nesse estudo.

Data	DAS	Instâncias
27/01/2017	96	7838
08/02/2017	108	7838
21/02/2017	121	7838
13/03/2017	141	7838

2.2 Índices de Vegetação

Cinco IVs foram selecionados e calculados para esse estudo. Essa seleção foi realizada por meio de pesquisas em trabalhos correlatos que tinham como objetivo a estimativa da produtividade da soja ou culturas semelhantes. Também foi selecionado apenas IVs multiespectrais com combinações possíveis de se calcular com as bandas disponíveis para o estudo.

Segundo Karnieli et al. (2010), um dos índices obtidos por meio de informações multiespectrais mais usados e calculados é o Normalized Difference Vegetation Index (NDVI). Esse índice é calculado através de um procedimento de normalização entre 0 e 1, tendo uma resposta sensível à vegetação verde mesmo para áreas cobertas com vegetação baixa (Xue e Su, 2017). Tendo em vista que a produtividade de grãos está fortemente associada ao teor de verde da planta, este IV é bastante utilizado com o objetivo de prever produtividade de colheita, como demonstrado nos trabalhos de Panda et al. (2010) e Gerke (2017). A fórmula do NDVI consiste na razão entre a diferença entre as bandas NIR e Red.

Outro IV que combina dados multiespectrais utilizando a NIR é o Wide Dynamic Range Vegetation Index (WDRVI). Esse IV utiliza em sua fórmula a coeficiente de ponderação a de 0,1-0,2, que lineariza os valores do índice e aumenta a correlação com a fração da vegetação (Huang et al., 2015). Para esse estudo foi atribuído a ele o valor de 0.1.

Os índices Modified Chlorophyll Absorption Ratio Index 2 (MCARI2) e Green Vegetation Index (GVI) foram selecionados para esse estudo para testar o desempenho de IVs que utilizam a banda G em suas fórmulas. Rodrigues (2015) destaca que o MCARI2 obteve desempenho consistente em seu estudo que consistia em avaliar diferentes IVs para a estimativa de produtividade de grãos de trigo, ficando entre os cinco primeiros IVs avaliados.

Também foi calculado o Visible Atmospherically Resistant Index (VARI), com o objetivo de avaliar a diferença entre os IVs que utilizam a banda NIR e um IV que combina apenas dados das bandas do espectro visível (RGB). O Quadro 1 apresenta os IVs calculados nesse estudo e suas respectivas fórmulas.

Quadro 1 - Índices de Vegetação utilizados nesse estudo.

Índice	Fórmula
<i>Normalized Difference Vegetation Index (NDVI)</i>	$(NIR - RED) / (NIR + RED)$
<i>Wide Dynamic Range Vegetation Index (WDRVI)</i>	$(a * NIR - RED) / (a * NIR + RED)$
<i>Modified Chlorophyll Absorption in Reflectance Index 2 (MCARI2)</i>	$1,5 * \frac{[2,5 * (NIR - RED) - 1,3 * (NIR - GREEN)]}{\sqrt{(2 * NIR + 1)^2 - (6 * NIR - 5 * VR) - 0,5}}$
<i>Green Vegetation Index (GVI)</i>	$(NIR + GREEN) / (NIR - GREEN)$
<i>Visible Atmospherically Resistant Index (VARI),</i>	$(GREEN - RED) * (GREEN + RED + BLUE)$

2.3 Análise dos Dados

As bases utilizadas nesse estudo foram submetidas primeiramente a testes estatísticos por meio de regressão. A ferramenta usada para os testes foi o software Weka (versão 3.8). A metodologia utilizada foi de dez modelos para validação cruzada e a métrica adotada foi o coeficiente de correlação (r).

Foram utilizados modelos de regressão para correlacionar o valor espectral das bandas e dos IVs com o valor da produtividade. Primeiramente foram realizados testes preliminares utilizando apenas os atributos RGB em uma das bases com o objetivo de comparar o desempenho de três algoritmos previamente selecionados e disponíveis na ferramenta, sendo eles Multilayer perceptron, M5Rules e SMOReg.

O SMOReg, um algoritmo baseado em SVR (Support Vector Regression), foi selecionado para as próximas etapas de testes por apresentar resultados superiores aos do Multilayer perceptron e similares aos do M5Rules, porém em menor tempo de execução.

2.4 Resultados e Discussão

Nos modelos de regressão as bandas da câmera RGB apresentaram o maior coeficiente de correlação (r) com produtividade em todos os testes

realizados, obtendo valores entre 0,63 (96 DAS) e 0,74 (121 DAS). Entre os IVs avaliados, os melhores resultados foram obtidos pelo VARI, que também utiliza as bandas RGB, com valores de r entre 0,48 (96 DAS) e 0,64 (121 DAS). A Tabela 2 apresenta o valor de r obtido para cada um dos IVs e bandas nas quatro bases utilizadas.

Tabela 2 - Coeficiente de correlação obtido para cada conjunto de atributos nos diferentes períodos de dias após a semeadura (DAS).

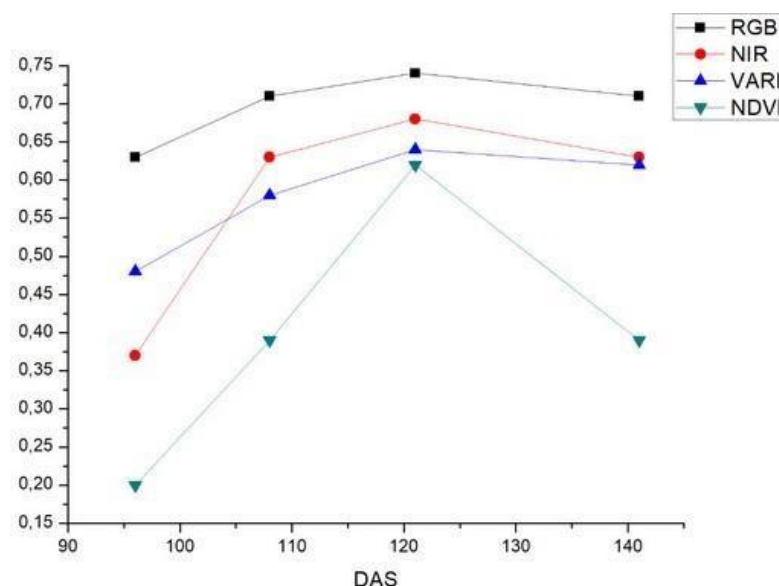
Atributos	DAS			
	96	108	121	141
RGB	0,63	0,71	0,74	0,71
NIR	0,37	0,63	0,68	0,63
VARI	0,48	0,58	0,64	0,62
NDVI	0,20	0,39	0,62	0,39
WDRVI	0,20	0,38	0,60	0,39
GVI	0,19	0,49	0,62	0,23
MCARI2	0,16	0,13	0,61	0,38

Considerando que a câmera RGB tem um custo correspondente em média a 1/10 do custo de uma câmera NIR, os resultados obtidos trazem uma boa e importante perspectiva de que o uso da câmera RGB, mais barata, pode oferecer uma solução mais econômica e viável para estimativa de produtividade da soja.

Entre os IVs que combinam bandas NIR, os índices NDVI, WDRVI e GVI apresentaram desempenhos gerais semelhantes entre eles e superiores ao do MCARI2, que apresentou a menor consistência entre os períodos de DAS. Esse padrão foi diferente do apresentado por Rodrigues (2015), que apontou o MCARI2 como um dos IVs mais consistentes em testes com o objetivo de estimar produtividade de trigo. Essa diferença de consistência no desempenho pode ser explicada pelas diferenças entre as culturas de trigo e soja, sobretudo em estádios de desenvolvimento da planta, uma vez que esse IV apresentou menores correlações em períodos com menores quantidades de DAS. Vale ressaltar que embora o MCARI2 tenha apresentado resultados consistentes entre os períodos avaliados, nenhum dos IVs utilizados por Rodrigues (2015) apresentou correlação maior que 0,56. A Figura 2 apresenta um gráfico de

desempenho obtidos pelas duas câmeras e IVs com melhores resultados de acordo com o período de DAS.

Figura 2 – Valores de r obtidos pelas bandas e IVs em diferentes períodos de DAS.



Embora os IVs testados no atual estudo tenham tido um aumento significativo de correlação, sobretudo no período de 121 DAS (Figura 2), nenhuma de suas combinações entre bandas apresentou ganho de desempenho em relação aos valores originais das câmeras RGB e NIR. Esse padrão também foi apresentado por Gerke (2017) ao estimar a produtividade do trigo, no qual o autor concluiu que o IV utilizado no estudo (NDVI) não se mostrou adequado para a estimativa. Também foi destacado pelo autor o desempenho superior da câmera RGB em relação a NIR, o que também foi confirmado no atual estudo. Esses resultados indicam que, assim como no trigo, a câmera RGB também é mais adequada para a estimativa de produtividade de soja.

3. CONSIDERAÇÕES FINAIS

O VARI (Visible Atmospherically Resistant Index), um índice que utiliza combinações das bandas RGB, apresentou o melhor desempenho em estimar

produtividade de soja entre os cinco IVs testados nesse estudo. No entanto, nenhuma das combinações das bandas utilizadas pelos IVs proporcionaram ganho de desempenho em relação aos dados originais das câmeras, demonstrando que a utilização dos valores originais das bandas RGB de imagens obtidas por VANT associados a técnica de aprendizado de máquina SMOReg se mostram como uma metodologia mais eficiente para estimativa de produtividade da soja do que o uso dos índices de vegetação já estabelecidos. O fato de todos os testes demonstrarem que os valores das bandas RGB apresentam maior correlação com produtividade de soja em relação aos valores da NIR, traz também uma importante contribuição, visto que demonstra que a câmera de menor custo pode ser utilizada proporcionando maior eficiência.

O VANT utilizado nesse estudo demonstrou ser uma eficiente ferramenta para se estimar produtividade de soja, sobretudo pela qualidade de suas imagens e praticidade de seu uso.

REFERÊNCIAS

ARUMUGAM, A. (2017). A predictive modeling approach for improving paddy crop productivity using data mining techniques. **Turkish Journal of Electrical Engineering and Computer Sciences**, p. 4777– 4787.

BERNARDI, A. C. et al. (2014). **Agricultura de Precisão: resultados de um novo olhar**. Empresa Brasileira de Pesquisa Agropecuária – EMBRAPA.

FERNANDES, J. et al. (2011). Sugarcane yields estimates using time series analysis of spot vegetation images. **Scientia Agricola**, p. 39–146.

FIGUEIREDO, D. (2007). **Conceitos básicos de sensoriamento remoto**. Companhia Nacional de Abastecimento – CONAB. 2007.

GERKE, T. (2017). **Mineração de dados de imagens obtidas com aeronave remotamente pilotada para estimativa de produtividade do trigo**. *Dissertação (Mestrado em Computação Aplicada) – Setor de Ciências*

Agrárias e de Tecnologia, Universidade Estadual de Ponta Grossa, Ponta Grossa, 2017.

HUANG, S. et al. (2015). Satellite remote sensing-based in-season diagnosis of rice nitrogen status in northeast china. **Remote Sens**, p. 10646–10667.

IBGE (2014). **Agencia IBGE Notícias** – Disponível em <https://agenciadenoticias.ibge.gov.br/agencia-noticias/2012-agencia-de-noticias/noticias/10167-records-de-soja-e-milho-levam-estimativa-da-safra-2017-a-29-2-acima-de-2016.html>. Acessado em :09 de abril de 2019.

KARNIELI, A. et al. (2010). Gutmanuse of ndvi and land surface temperature for drought assessment: Merits and limitations. **Journal of Climate**, p. 618– 633.

LIU, W. T. H. (2006). **Aplicações de sensoriamento remoto**. campo grande: Uniderp.

PANDA, S. S. et al. (2010). Application of vegetation indices for agricultural crop yield prediction using neural network techniques. **Remote Sens**, p. 673–696.

RODRIGUES, F. A. (2015). High resolution remote and proximal sensing to assess low and high yield areas in a wheat field. **Precision Agriculture'15**, p. 38–50.

WU, W. (2014). The generalized difference vegetation index (gdvi) for dryland characterization. **Remote Sens**, p. 1211–1233.

XUE, J. e SU, B. (2017). Significant remote sensing vegetation indices: A review of developments and applications. **Journal of Sensors**. p. 201-219.

ZAMPIER, D. et al. (2017). **Influência da resolução espacial em imagens obtidas por rpa na estimativa de doenças na soja**. *Anais do XI Congresso Brasileiro de Agroinformática (SBIAgro 2017), Campinas, São Paulo, 2 a 6 de outubro de 2017*, p.169–178.

Edição especial – I Encontro Nacional Interdisciplinar em Ciência, Tecnologia e Sociedade (ENICTS 2019)

Enviado em: 15 mai. 2020

Aceito em: 28 jun. 2020

Editor responsável: Mateus das Neves Gomes