



## USO DE DADOS GEOESPACIAIS PARA ESTIMATIVA DA PRODUTIVIDADE DO MILHO

Filipe Vieira da Silva<sup>1</sup>, Maria Cecília Lemes Santos<sup>1</sup>, Rayssa Santos Barbosa<sup>1</sup>, George Deroco Martins<sup>1</sup>

<sup>1</sup> Universidade Federal de Uberlândia, Campus Monte Carmelo, Monte Carmelo, Minas Gerais ([lipesilva@ufu.br](mailto:lipesilva@ufu.br));

**RESUMO:** Atualmente, a produtividade do milho pode ser estimada a partir de interpoladores como Krigagem e por imagens multiespectrais associadas a algoritmos de aprendizado de máquinas. Logo, o objetivo deste trabalho foi comparar diferentes modelos de estimativa frente ao seu uso para a análise espacial da produtividade, com o uso de softwares livres. Os dados foram separados em 2 conjuntos: um para treinamento do modelo preditivo e outro para a validação. A manipulação dos dados e a Krigagem foi realizada no programa QGIS, enquanto para a geração e análise do modelo multiespectral a partir de imagens Sentinel 2, foi empregado o software Weka. Para a Krigagem o modelo Linear ajustou-se melhor aos dados e trouxe à estimativa um RMSE de 15,86% e um  $R^2$  de 0,643. Enquanto isso, a Neural Network teve destaque para a geração do modelo multiespectral com um RMSE de 14,46% e  $R^2$  igual a 0,457. Com isso, foi possível deduzir que a Krigagem ainda é a melhor opção para a estimativa e análise espacial da produtividade da cultura do milho, embora a estimativa com o uso de imagens tenha se mostrado bastante válida e promissora para um monitoramento em larga escala.

**Palavras-chave:** agricultura de precisão, sensoriamento remoto, monitoramento agrícola.

### INTRODUÇÃO

O milho é o cereal mais produzido no mundo, tendo como principais produtores os Estados Unidos, a China e o Brasil. No Brasil, a cultura pode ser produzida em até 3 safras, sendo a primeira safra forte em Minas Gerais, a safrinha (segunda safra) no estado do Mato Grosso e a terceira safra com destaque no estado de Sergipe (OLIVEIRA, 2023; CONAB 2019).

Levando em consideração a importância que o cereal tem, tanto a nível nacional como em uma escala global, é preciso estimar a produção de milho para gerar informações essenciais ao manejo, comercialização, armazenamento e escoamento da cultura (LETRAS AMBIENTAIS, 2021). Além disso, a estimativa de produtividade pode ser uma ferramenta que auxilia na comparação da viabilidade de um produto, bem como na avaliação dos híbridos/cultivares plantados. (PANIAGO, 202?).

Utilizando amostras de produtividade coletada de forma pontual e posteriormente estimando para toda a área plantada por meio da interpolação, é possível se estimar a produção da lavoura (PANIAGO, 202?). Já Abreu Júnior (2022) explicita em seu trabalho outra forma



de se estimar a produtividade, através do uso de imagens multiespectrais, e nele determinou que a melhor época para se estimar a produtividade do milho com o auxílio das imagens se dá no estágio fenológico R2, cerca de 2 meses antes da colheita. Neste estudo, a estimativa foi realizada através de algoritmos de aprendizados de máquina considerando os valores de reflectância das bandas originais da imagem Planet e índices de vegetação derivados.

Dado que pode ser gerada uma estimativa de produtividade do milho por meio da Krigagem com o uso dos pontos amostrais de produtividade e através do uso de imagens multiespectrais, o seguinte trabalho tem como objetivo aplicar os dois métodos de estimativa de produtividade do milho e avaliar qual deles é mais acurado.

## MATERIAL E MÉTODOS

Para realizar o trabalho foi utilizado um pacote de imagens do satélite Sentinel 2, com nível de processamento 2A que conta com a correção de reflectância na base da atmosfera. O pacote possui 13 bandas espectrais com alta resolução espacial, chegando a 10 m, e uma resolução espectral que varia entre 443 e 2190 nanômetros (EMBRAPA, 2018)

Além das imagens, foram necessários pontos amostrais de produtividade e o limite da lavoura. Foram computados 3019 pontos de produtividade, contando com uma média de 2,65 T/ha. Já para a geração do modelo de estimativa com os dados espectrais, foram coletados 316 pixels, somente os pixels que continham dados de produtividade foram contabilizados.

Ademais, foi utilizado o software QGIS para manipulação dos dados, bem como para seleção dos dados de treinamento e validação, a estimativa por krigagem e o cálculo dos índices de vegetação. Já o software WEKA foi necessário para a estimativa através dos algoritmos de aprendizado de máquina.

O trabalho foi dividido em duas partes, uma que realizou a estimativa considerando o uso da imagem orbital e outra etapa considerando somente a Krigagem. Mas para iniciar ao processo de estimativa, antes é preciso selecionar e separar os arquivos de treinamento e validação, método conhecido como Holdout.

Para dar início à separação dos dados, fez-se necessário a seleção dos pixels que continham dados de produtividade, transformando o raster para polígono e unindo os atributos que intersectam essa feição. Este processo pode ser feito com a ferramenta “unir atributos pela localização (sumário)” onde selecionamos a produtividade como variável a ser unida e a média como valor a ser computado e registrado.

Através da seleção por localização, foram selecionados somente os polígonos que intersectavam o limite da área de estudo e salvos como uma camada vetorial. Este procedimento



garante que somente os pixels que possuem informações de reflectância do dossel influencie na geração do modelo, eliminando então todos os pixels que extrapolam a região de estudo.

Em seguida foram calculados alguns Índices de Vegetação (IV) a fim de ressaltar características biofísicas da planta, como a concentração de clorofila e a densidade foliar da cultura. Índices como NDVI (Normalized Difference Vegetation Index), GNDVI (Green Normalized Difference Vegetation Index), RVI (Ratio Vegetation Index), TVI (Triangular Vegetation Index) e CIG (Chlorophyll Index Green) foram computados.

Extraíu-se os centroides da camada vetorial e por fim, fez-se a extração das informações espectrais de cada banda e dos valores dos IVs calculados através da ferramenta “Add raster values to points”.

Neste momento os dados estão prontos para serem separados em dois conjuntos: um para a geração e outro para a validação do modelo de estimativa. Logo, foram selecionados aleatoriamente 80% dos pontos para treinamento do modelo e os 20% restante para a validação. A tabela de atributos foi exportada e o arquivo ‘.arff’ (padrão do Weka) foi montado para iniciar o estágio de processamento no software Weka.

Vale ressaltar que os mesmos pontos utilizados para a avaliação do desempenho do modelo multiespectral, devem ser utilizados para a estimativa através da Krigagem, para permitir a comparação entre os métodos.

No Weka, basta abrir o acervo de 80% na aba “Preprocess”, em “Classify” selecionar o dado de validação através do “Supplied test set” e testar os modelos disponíveis. Através da ferramenta “Select attributes” é possível identificar quais atributos possuem uma baixa correlação com a produtividade, em seguida removê-los e assim melhorar a estimativa.

Para gerar a interpolação por Krigagem, foi utilizado a camada vetorial contendo os 80% dos pontos e o *plug-in* Smart-Map. Dentro do *plug-in* é possível selecionar a variável a ser interpolada, o limite da área de estudo garantindo que toda a região seja estimada, a dimensão do pixel e o modelo ajustado.

Para finalizar, com o arquivo de validação fez-se a extração dos valores estimados pelo *plug-in*, computou a diferença entre o valor real e estimado e por fim foi calculado o RMSE (Root Mean Square Error) em porcentagem de acordo com a fórmula 1. Já para o modelo multiespectral bastou dividir o RMSE disponibilizado pelo software pelo quociente do somatório dos valores medidos pelo número de pontos.

$$RMSE\% (y, \hat{y}) = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}}{\frac{1}{n} \sum_{i=1}^n y_i} * 100 \quad (1)$$

Onde:  $y$  são os valores medidos e  $\hat{y}$  são os valores estimados.

Uma vez que os dados foram estimados, é possível espacializar toda a produtividade da lavoura, tornando possível uma análise com relação às áreas de maior e menor produtividade.

## RESULTADOS E DISCUSSÃO

Dentre os modelos testados o que melhor se ajustou aos dados foi o Neural Network. Além disso, 7 atributos foram utilizados na geração da melhor estimativa, sendo eles: a produtividade, os índices CIG e TVI, e as bandas espectrais 5, 6, 11 e 12. Já para a Krigagem, o modelo Linear obteve melhor desempenho.

Para a comparação entre os modelos, foram anotados os coeficientes de correlação e o RMSE (%), apresentados na tabela 1.

Tabela 1. Métricas para avaliação dos modelos

Método	RMSE (%)	R <sup>2</sup>
Krigagem	15,86	0,643
Modelo Multiespectral	14,46	0,457

O método de Krigagem deteve de um RMSE de 15,86% e o coeficiente de correlação igual a 0,643, enquanto o método multiespectral contou com 14,46% de RMSE e anotou 0,457 para o índice de correlação. Logo, fica evidente que a melhor forma de se estimar a produtividade do milho se dá através do método de Krigagem.

O valor de RMSE é menor para a estimativa através da imagem, mostrando que o modelo gerado é capaz de estimar melhor o valor médio absoluto de produtividade quando comparado ao modelo da Krigagem, porém o coeficiente de correlação deixa evidente que para a espacialização da produtividade o modelo multiespectral é inferior. Assim sendo, o modelo multiespectral pode não ser capaz de especializar as regiões com picos de produtividade, o que é de grande importância para o manejo da cultura.

É necessário evidenciar que se trata de uma análise local, ou seja, caso o trabalho seja avaliado em uma outra região e com dados diferentes, podem surgir outros resultados.

## CONCLUSÕES

A estimativa por imagem se mostrou bastante promissora frente ao procedimento da Krigagem. Visto que a resolução espacial e espectral vem aumentando ao longo dos tempos, pode chegar o momento em que a estimativa através das imagens possa superar o outro método.



Por meio dos resultados apresentados, fica evidente que é preciso que as métricas de avaliação da estimativa sejam inversamente proporcionais, onde o RMSE seja o mais baixo possível e o coeficiente de correlação mais próximo de 1.

Com isso, podemos concluir que o método de Krigagem ainda é a melhor forma de se estimar a produtividade espacialmente, apesar de apresentar um RMSE um pouco maior que o modelo multiespectral, o coeficiente de correlação torna o método a melhor opção.

## REFERÊNCIAS

ABREU JÚNIOR, C. A. B. M. de. **Estimativa de produtividade da cultura do milho a partir de imagens multiespectrais**. 2022. Dissertação (Mestrado em Agricultura e Informações Geoespaciais) – Universidade Federal de Uberlândia, Monte Carmelo, MG, 2022. Disponível em:

<https://repositorio.ufu.br/bitstream/123456789/36220/3/EstimativaProdutividadeCultura.pdf>.

Acesso em: 17 ago. 2023

CONAB. **Milho**: análise mensal – junho-julho 2019. Brasília, DF: CONAB, 2019. Disponível em: <https://www.conab.gov.br/info-agro/analises-do-mercado-agropecuario-e-extrativista/analises-do-mercado/historico-mensal-de-milho/item/12123-milho-analise-mensal-junho-julho-2019>. Acesso em: 17 ago. 2023.

EMBRAPA TERRITORIAL. **Sentinel**: missão Sentinel. Campinas, 2018. Disponível em:

<https://www.embrapa.br/satelites-de-monitoramento/missoes/sentinel#:~:text=Possui%20resoluçã%20de%20500m%20para,1%20km%20de%20resoluçã%20espacial>.

Acesso em: 17 ago. 2023.

LETRAS AMBIENTAIS. Os 9 passos para se estimar a produtividade agrícola por satélites. **Letras Ambientais**, [S.l.], 2021. Disponível em:

<https://www.letrasambientais.org.br/posts/os-9-passos-para-se-estimar-a-produtividade-agricola-por-satelites#>. Acesso em: 17 ago. 2023.

PANIAGO, B. Plantação de milho: 3 métodos para estimar a produtividade. **Agointeli**, [S.l.], 202?. Disponível em: <https://blog.aegro.com.br/producao-de-graos-no-brasil/>. Acesso em: 17 ago. 2023.

OLIVEIRA, C. Produção de grãos no Brasil: entenda o cenário e como ele afeta sua lavoura. **Aegro**, [S.l.], 2023. Disponível em: <https://blog.aegro.com.br/producao-de-graos-no-brasil/>. Acesso em: 17 ago. 2023.