

PREDIÇÃO DE PRODUTIVIDADE DE CAFEEIRO A PARTIR DO COMPORTAMENTO ESPECTRAL

Afonso Vizula de Castro Júnior¹, João Mendes Cicarini Hott², Amanda Teixeira
Moreira Capacia², Thiago Alexandre da Silva Oliveira², Maísa Martins de
Oliveira², Samuel de Assis Silva²

¹Engenheiro Agrônomo, afonsovizula@gmail.com.

²Universidade Federal do Espírito Santo/Departamento de Engenharia Rural, Alto Universitário, s/n,
Guararema - 29500-000 – Alegre-ES, Brasil, joaoagro18@hotmail.com, amandacapacia@gmail.com,
thiagoalexandre984@gmail.com, maisamartins20@gmail.com,
samuel.assilva@gmail.com.

Resumo

Objetivou-se com este trabalho estimar a produtividade do cafeeiro conilon através de modelo simples de regressão usando como input a reflectância das plantas de café conillon nas bandas do infravermelho-próximo (NIR) e vermelho (Red) obtidas através de sensoriamento remoto orbital. Para isso foram obtidas cenas do satélite Sentinel-2, no período entre outubro de 2016 a junho de 2017, e dados de produtividade de uma lavoura cafeeira. Tanto o NIR quanto o Red apresentaram correlação com a produtividade. Os modelos para predição de produtividade apresentaram R^2 de 0,7889 e 0,7546 tendo como input as variáveis NIR e Red, sucessivamente. Foi possível estimar a produtividade das plantas de café conilon a partir de modelos tendo como base a reflectância nas bandas NIR e Red.

Palavras-chave: Agricultura de Precisão e Digital. Cafeicultura. Sentinel-2. Sensoriamento Remoto.

Área do Conhecimento: Engenharia Agrônômica – Engenharia Agrícola.

Introdução

O Brasil se destaca como produtor mundial de café, sendo uma das principais culturas do Espírito Santo. Estimar a produtividade de uma lavoura é essencial para o manejo do cultivos (Carneiro et al., 2023). Além disso, essa informação é importante nas etapas de colheita e pós-colheita, como a contratação de mão de obra e alocação de espaço para beneficiamento e armazenamento (Liu et al., 2023).

Plantas saudáveis refletem grande parte radiação eletromagnética (REM) da região do infravermelho-próximo (NIR) e absorve na região do vermelho (Red) (Fontana et al., 2019). Dessa forma, através do sensoriamento remoto orbital é possível obter respostas rápidas e econômica e não destrutiva a partir da reflectância da radiação eletromagnética dos alvos (Carneiro et al., 2023).

Diante disso, este trabalho teve como objetivo estimar a produtividade do cafeeiro conilon através de modelo simples de regressão usando como input a reflectância das plantas de café conillon nas bandas NIR e Red obtidas através de sensoriamento remoto orbital.

Metodologia

O estudo foi realizado em uma lavoura, com cerca de 2ha cultivada com café conilon (*Coffea canephora* Pierre ex Froehner), variedade "EMCAPER 8151", localizada no município de Alegre-ES. Foi estabelecido um grid amostral irregular contendo 120 pontos amostrais. Cada ponto amostral foi composto por três plantas de café localizadas lado a lado na mesma linha. Os pontos amostrais foram georreferenciados com auxílio de uma estação total.

Foram utilizadas imagens do sensor MSI embarcado na plataforma orbital Sentinel-2. Foram utilizadas as bandas espectrais do vermelho (Red) e infravermelho próximo (NIR) com resolução espacial de 10m. Os dados do Sentinel-2 são disponibilizados na plataforma online Earth Explore. Utilizou-se cenas no período compreendido entre outubro/2016 e junho/2017, com a menor interferência de nuvens.

A era digital e suas implicações sociais: Desafios e contribuições

As imagens têm nível de correção Level-2A, portanto, são ortorretificadas, georreferenciadas com projeção geográfica WGS84 e com correção radiométrica no topo da atmosfera-TOA. Posteriormente, foi realizado o pré-processamento das imagens no software QGIS utilizando o complemento Semi-Automatic Classification Plugin. Esta etapa consiste na correção radiométrica ortorretificada em superfície-BOA e na conversão dos números digitais em números de reflectância (RT) da superfície dos alvos.

Os valores de reflectância para os 120 pontos amostrais foram extraídos utilizando o complemento Point Sampling Tools através do software QGIS. O complemento é capaz de extrair os valores de um arquivo raster (bandas NIR e Red) com base em um arquivo vetorial (pontos amostrais).

A produtividade média de cada ponto amostral foi mensurada no momento em que cerca de 90% dos grãos estavam maduros, sendo a colheita feita por derriça manual. A massa de café úmido colhida foi separada em cereja, verde e seco. Após a separação, retirou-se uma amostra de 1,0 Kg a qual foi colocada em estufa a 40°C até atingir $\pm 12\%$ de umidade.

O café beneficiado foi determinado retirando-se a casca dos grãos secos, sendo esses submetidos à pesagem, obtendo-se a massa de café beneficiado. Para cada ponto amostral da área, os valores de café beneficiado foram convertidos em produtividade na unidade Mg.ha^{-1} .

Os valores levantados pelas variáveis em estudo foram submetidos à análise de estatística descritiva e exploratória dos dados, por meio das medidas de posição e dispersão, bem como analisou-se a normalidade dos dados, verificada pelo teste de Shapiro-Wilk ($p < 0,05$), utilizando o software Statistica 7.0.

Em seguida, os dados de produtividade e das bandas NIR e Red foram submetidos à análise geoestatística para verificar a existência e quantificar a dependência espacial, a partir do ajuste das funções teóricas aos modelos de variogramas experimentais com base na pressuposição de estacionaridade da hipótese intrínseca dos dados. Os variogramas experimentais foram estimados pela Equação 1:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z_1(x_i) - Z_2(x_i + h)]^2 \quad (1)$$

Em que: $N(h)$ é o número de pares de valores $[Z(x_i), Z(x_i + h)]$ separados por um vetor h e, x_i , a posição espacial da variável Z .

No ajuste dos modelos teóricos aos variogramas experimentais, assumiu-se comportamento isotrópico. Os modelos testados para ajuste foram o esférico, exponencial, gaussiano e linear com patamar. A escolha dos modelos foi feita com base no critério dos mínimos quadrados, optando-se na seleção pelos modelos com maior valor de R^2 (coeficiente de determinação), menor SQR (soma de quadrado dos resíduos) e maior valor do coeficiente de correlação obtido pelo método de validação cruzada.

Após a escolha dos modelos teóricos que melhor se ajustaram aos experimentais foram determinados os coeficientes: efeito pepita (C_0), patamar (C_0+C) e alcance (a). Os parâmetros (C_0 e C_0+C) dos variogramas foram escalonados pela variância.

Comprovada a dependência espacial, as variáveis foram interpoladas por krigagem ordinária, em uma matriz de pontos de 150×150 . Esse interpolador geoestatístico utiliza um estimador linear não viciado com mínima variância que leva em consideração a estrutura de variabilidade espacial encontrada para a variável.

Com finalidade de testar a correlação espacial entre as variáveis, foi feita a análise de correlação linear de Pearson ao nível de 5% de probabilidade, a partir dos dados da matriz de proximidade espacial. Confirmada a correlação espacial da produtividade com o NDVI em todos os meses analisados, os dados foram submetidos à análise de regressão linear múltipla, sendo essas as variáveis independentes e a produtividade a variável dependente.

Para a elaboração dos mapas temáticos, foram utilizadas todas as variáveis em todos os períodos que apresentaram significância ($p < 0,05$) nas análises de regressão linear múltipla. Confirmada a significância, foi possível estimar a produtividade, a partir da Equação 2:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_k X_k \quad (2)$$

A era digital e suas implicações sociais: Desafios e contribuições

Em que: Y é a variável dependente, ou seja, a estimativa de produtividade; β_0 são os valores do intercepto; β_i ($i=1...k$) são os valores dos coeficientes de regressão parciais e os valores de $X_{i(i=1 \text{ a } k)}$ são as variáveis independentes, obtidas pela interpolação dos dados por krigagem ordinária.

De posse das estimativas de produtividade, tornou-se possível calcular o erro em relação a produtividade real, convertido em sacas.ha⁻¹, conforme Equação 3:

$$Erro = \frac{(Y_{ij} - X_{ij}) \times 1000}{60} \quad (3)$$

Em que: Erro é corresponde ao erro em sacas por hectare (sc.ha⁻¹); Y_i é a produtividade estimada a partir das variáveis para a posição i e j ; X_{ij} é a produtividade real interpolada a partir dos dados de campo para a posição i e j .

Resultados

A estatística descritiva das bandas do infravermelho-próximo (NIR) e vermelho (Red) e da produtividade avaliados no período entre outubro/2016 e junho/2017, encontra-se nas Tabelas 1 e 2.

Tabela 1- Estatística descritiva e distribuição de frequência para as análises mensais reflectância da banda do Infravermelho próximo.

Variável	Média	Mediana	Mínimo	Máximo	CV (%)	Cs	Ck	SW
NIR_Out_16	0,20	0,20	0,18	0,22	4,12	0,22	-0,37	ns
NIR_Nov_16	0,39	0,40	0,34	0,44	5,83	-0,33	-0,36	*
NIR_Dez_16	0,36	0,36	0,30	0,42	7,93	0,28	-0,72	*
NIR_Jan_16	0,45	0,45	0,40	0,49	4,97	-0,09	-0,89	*
NIR_Fev_17	0,37	0,37	0,33	0,41	5,26	0,09	-1,06	*
NIR_Mar_17	0,41	0,41	0,34	0,45	6,93	-0,33	-0,71	*
NIR_Abr_17	0,43	0,43	0,37	0,47	5,30	-0,28	-0,63	*
NIR_Maio_17	0,33	0,33	0,28	0,39	7,33	0,66	-0,20	*
NIR_Jun_17	0,36	0,36	0,31	0,41	5,98	-0,06	-0,31	ns
Produtividade	0,76	0,18	0	5,94	152,93	-2,07	-4,55	*

CV% - coeficiente de variação; Cs – coeficiente de assimetria; Ck – coeficiente de curtose; SW– Shapiro-Wilk (5% de probabilidade); ns – não significativo; * significativo ($p < 0,05$).

Fonte: Os autores.

Tabela 2- Estatística descritiva e distribuição de frequência para as análises mensais da reflectância da banda do vermelho.

Variável	Média	Mediana	Mínimo	Máximo	CV (%)	Cs	Ck	SW
Red_Out_16	0,07	0,07	0,05	0,10	13,97	0,29	-0,29	ns
Red_Nov_16	0,03	0,03	0,03	0,04	7,53	0,79	-0,10	*
Red_Dez_16	0,03	0,03	0,03	0,04	11,42	0,92	-0,09	*
Red_Jan_17	0,04	0,03	0,03	0,05	11,49	0,93	-0,03	*
Red_Fev_17	0,05	0,05	0,04	0,06	11,97	0,43	-0,44	*
Red_Mar_17	0,03	0,03	0,02	0,03	6,46	0,63	-0,19	*
Red_Abr_17	0,03	0,03	0,03	0,03	6,51	0,52	-0,40	*
Red_Maio_17	0,03	0,03	0,03	0,05	10,89	0,54	-0,30	*
Red_Jun_17	0,04	0,04	0,03	0,05	12,62	0,35	-0,58	*

CV% - coeficiente de variação; Cs – coeficiente de assimetria; Ck – coeficiente de curtose; SW– Shapiro-Wilk (5% de probabilidade); ns – não significativo; * significativo ($p < 0,05$).

Fonte: Os autores.

A produtividade média foi de 0,76 Mg.ha⁻¹, correspondente a cerca de 13 sacas.ha⁻¹, o CV foi de 152,93% devido a maior amplitude de variação dos dados. Para o NIR e Red apresentaram baixa variabilidade (CV<15%). Os coeficientes de assimetria e curtose que apresentaram valores próximos

A era digital e suas implicações sociais: Desafios e contribuições

de zero, indicando baixa variação em relação à média. Ademais, com exceção do NIR e Red para o mês de outubro/2016 e do NIR para o mês de junho/2017.

O resultado da análise espacial para os dados de produtividade da banda NIR e Red são apresentados nas Tabelas 3 e 4.

Tabela 3- Modelos e parâmetros escalonados dos variogramas simples para a reflectância do NIR e produtividade.

Variável	Modelo	C ₀	C ₀ +C	a(m)	R ² (%)	IDE(%)
NIR_Out_16	GAU	0,06	1,09	20,26	84	94,9
NIR_Nov_16	GAU	0,12	1,33	59,58	90,1	91,3
NIR_Dez_16	GAU	0,04	0,51	29,27	86,2	91,7
NIR_Jan_16	EXP	0,02	0,59	54,90	90,2	95,9
NIR_Fev_17	GAU	0,09	0,75	40,87	98,1	87,8
NIR_Mar_17	ESF	0,03	1,20	57,10	99,5	97,6
NIR_Abr_17	GAU	0,23	1,48	89,20	97,4	84,7
NIR_Maio_17	GAU	0,09	0,71	29,96	99,1	87,3
NIR_Jun_17	GAU	0,20	1,49	88,85	93,8	86,9
NDVI_Jun_17	ESF	0,16	1,11	61,20	96,5	85,9
Produtividade	EXP	0,27	1,06	25,5	81	75

NIR – Infravermelho próximo; GAU – Gaussiano; ESF – Esférico; EXP – Exponencial; C₀ – efeito pepita; C₀+C – patamar; a – alcance; IDE – índice de dependência espacial; R² – coeficiente de determinação do modelo.

Fonte: Os autores.

Tabela 4- Modelos e parâmetros escalonados dos variogramas simples para a banda do vermelho.

Variável	Modelo	C ₀	C ₀ +C	a(m)	R ² (%)	IDE(%)
Red_Out_16	GAU	0,21	2,22	78,46	99,2	90,5
Red_Nov_16	GAU	0,33	1,78	78,00	98,1	81,8
Red_Dez_16	GAU	0,13	1,27	63,73	81,4	89,9
Red_Jan_17	GAU	0,05	1,52	78,00	85,7	96,7
Red_Fev_17	GAU	0,16	1,27	67,20	93	87,2
Red_Mar_17	ESF	0,27	1,78	83,20	88,9	85,2
Red_Abr_17	ESF	0,23	0,96	61,10	93,1	76,4
Red_Maio_17	ESF	0,08	1,27	40,40	96,6	93,8
Red_Jun_17	ESF	0,20	1,24	55,50	97,7	83,5

Red – Vermelho; GAU – Gaussiano; ESF – Esférico; EXP – Exponencial; C₀ – efeito pepita; C₀+C – patamar; a – alcance; IDE – índice de dependência espacial; R² – coeficiente de determinação do modelo.

Fonte: Os autores.

Os valores da correlação espacial entre a produtividade e as bandas NIR e Red são descritos na Tabela 5.

Tabela 5- Correlação espacial de dados mensais de índices foliares aparentes de clorofila com dados de NDVI e das reflectâncias do infravermelho próximo e do vermelho do caféiro conilon.

Clorofila	NIR	Red
Out_16	0,37	-0,07
Nov_16	0,07	-0,59
Dez_16	0,67	-0,21
Jan_17	0,54	-0,23
Fev_17	0,14	-0,24

A era digital e suas implicações sociais: Desafios e contribuições

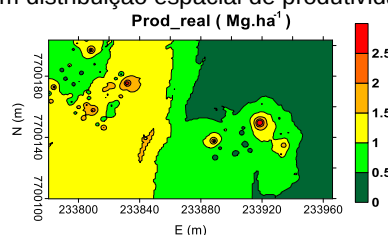
Mar_17	0,31	0,19
Abr_17	0,15	-0,17
Mai_17	-0,5	-0,07
Jun_17	0,14	-0,09

Negrito - significativo ao nível de 5% de probabilidade.

Fonte: Os autores.

A partir da Figura 1 é possível observar como a produtividade é espacialmente distribuída, com valores que variam de 0 e 2,5 Mg.ha⁻¹.

Figura 1. Mapas temáticos com distribuição espacial de produtividade do cafeeiro conilon.



Fonte: Autores (2023).

Na Tabela 6 é possível observar as estimativas de produtividade por meio da análise de regressão linear múltipla tendo como variável o NDVI avaliado durante o período de outubro de 2016 a junho de 2017.

Tabela 6- Equações de regressão múltipla com nível de significância de 5% para estimativa de produtividade.

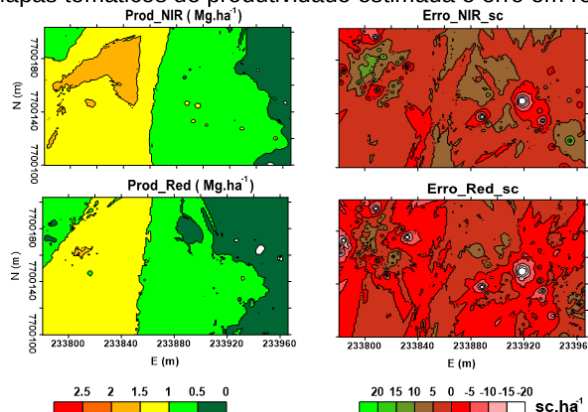
Variável	Estimativa	R ² (%)
NIR	$Y = -10,7516 + 22,4750 * \text{Out} + 11,0888 * \text{Nov} + 5,5725 * \text{Jan} + 5,9151 * \text{Fev} - 14,3841 * \text{Mar} - 5,1438 * \text{Abr} + 22,6890 * \text{Maio} - 3,9656 * \text{Jun}$	78,89
Red	$Y = 4,933 + 18,690 * \text{Out} + 11,816 * \text{Nov} + 34,360 * \text{Dez} - 34,433 * \text{Jan} - 46,104 * \text{Fev} + 3,644 * \text{Mar} - 101,037 * \text{Abr} - 69,512 * \text{Maio} + 50,912 * \text{Jun}$	75,46

NIR – infravermelho próximo; Red – vermelho; Y – produtividade estimada; R² – coeficiente de determinação.

Fonte: Os autores.

Na Figura 2 são apresentados os mapas com as estimativas de produtividade a partir das variáveis NIR e Red e os dos erros calculados para cada uma das variáveis em relação à produtividade real.

Figura 2. Mapas temáticos de produtividade estimada e erro em relação a real.



Fonte: Autores (2023).

A era digital e suas implicações sociais: Desafios e contribuições

Discussão

A análise descritiva da produtividade e das bandas do infravermelho-próximo (NIR) e do vermelho (Red) indicam que as medidas de tendência centrais foram semelhantes para todas as variáveis analisadas, exceto para a produtividade. Dessa forma, a produtividade ao longo da área se comportou de forma assimétrica, diferente das bandas NIR e Red.

A partir dos resultados da análise espacial da produtividade e das bandas NIR e Red, observa-se que todas as variáveis apresentam dependência espacial para os modelos ajustados. A dependência espacial da produtividade do cafeeiro foi observada por Ferraz et al. (2019), para a espécie *C. arábica*.

Para a produtividade, o modelo que melhor ajustou foi exponencial. Para a banda NIR, o modelo foi o gaussiano, com exceção dos meses de janeiro/2016 e março e junho/2017 em que os foram o exponencial e esférico, respectivamente. Já para a banda Red, no período de outubro/2016 a fevereiro/2017, o modelo que mais se adequou foi o gaussiano, já para os meses seguintes o esférico.

O coeficiente de correlação linear de Pearson (r) da banda NIR indica que com o aumento da reflectância nessa banda há aumento da produtividade, com exceção para maio/2017. Já para a banda Red indica que o decréscimo da reflectância nessa banda há aumento da produtividade, com exceção de março/2017.

Os modelos para predição de produtividade apresentaram R^2 de 0,7889 e 0,7546 tendo como input as variáveis NIR e Red, sucessivamente, assim apresentando boa correlação com a produtividade. Carneiro et al. (2023) obtiveram boa correlação entre a produtividade e o modelo de predição de produtividade de algodoeiro a partir da banda Red utilizando Sentinel-2A. Os mapas de produtividade estimados pelo modelo e o real mostram a similaridade e a confiabilidade nos modelos gerados.

Conclusão

A produtividade do cafeeiro e a reflectância das plantas nas bandas NIR e Red possuem dependência espacial.

Foi possível realizar a estimativa de produtividade do cafeeiro conilon a partir do comportamento espectral das bandas NIR e Red.

A plataforma orbital Sentinel-2A se destaca como ferramenta gratuita para o fornecimento de dados utilizados na criação de modelos de regressão múltipla para predição de produtividade de cafeeiro conilon.

Referências

CARNEIRO, F.M.; BRITO, A.L.F.; FERREIRA, F.M.; SEBEN, G.F.J.; BRANDÃO, Z.N.; SILVA, R.P.; SHIRATSUCHI, L.S. Soil and satellite remote sensing variables importance using machine learning to predict cotton yield. **Smart Agricultural Technology**, vol 5, 2023.

FERRAZ, G.A.S.; AVELAR, R.C.; BENTO, N.L.; SOUZA, F.R.; FERRAZ, P.F.P.; DAMASCENO, F.A.; BARBARI, M. Spatial variability of soil fertility attributes and productivity in a coffee crop farm. **Agronomy Research**, vol. 17, n. 4, pg. 1630–1638, 2019.

FONTANA, D.C.; SANTOS, L.N.; DALMAGO, G.A.; SCHIRMBECK, J.; SCHIRMBECK, L. NDVI e alguns fatores de variabilidade. In: **Anais do XIX Simpósio Brasileiro de Sensoriamento Remoto**, 2019.

LIU, J.; SHANG, J.; QIAN, B.; HUFFMAN, T.; ZHANG, Y.; DONG, T.; JING, Q.; MARTIN, T. Crop Yield Estimation Using Time-Series MODIS Data and the Effects of Cropland Masks in Ontario, Canada. **Remote Sens**, v. 11, 2019.

Agradecimentos

A Fundação de Amparo à Pesquisa e Inovação do Espírito Santo (FAPES) e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo apoio a pesquisa.