

INTERPOLAÇÃO POR MACHINE LEARNING E KRIGAGEM ORDINÁRIA NO QGIS UTILIZANDO O PLUGIN SMART-MAP

GUSTAVO WILLAM PEREIRA¹, DOMINGOS SÁRVIO MAGALHÃES VALENTE²,
DANIEL MARÇAL DE QUEIROZ³, ANDRÉ LUIZ DE FREITAS COELHO⁴

¹ Ciência da Computação, Professor Adjunto. Doutor, Instituto Federal de Educação, Ciência e Tecnologia do Sudeste de Minas Gerais, Muriaé - MG, gustavo.willam@ifsudestemg.edu.br

² Engenharia Agrícola, Professor Associado. Doutor. Departamento de Engenharia Agrícola, UFV, Viçosa - MG

³ Engenharia Agrícola, Professor Titular. Doutor. Departamento de Engenharia Agrícola, UFV, Viçosa - MG

⁴ Engenharia Agrícola, Professor Adjunto. Doutor. Departamento de Engenharia Agrícola, UFV, Viçosa - MG

Apresentado no
LIII Congresso Brasileiro de Engenharia Agrícola - CONBEA 2024
6 a 8 de agosto de 2024 – Natal – RN, Brasil

RESUMO: A agricultura de precisão (AP) é uma técnica de gestão agrícola baseada na observação, medição e resposta às variabilidades espaciais e temporais que ocorrem nas áreas de produção agrícola. Em AP, o mapeamento das características físicas e químicas do solo permite estimar com maior precisão a variabilidade espacial do solo. Dado um conjunto de pontos amostrados deseja-se obter mapas dos atributos de solo ou inferir valores em lugares específicos em locais não amostrados, para isso técnicas de interpolação são utilizadas. Objetivou-se neste trabalho implementar um software que permitisse a interpolação de atributos do solo através de técnicas de Machine Learning (ML) e Krigagem Ordinária (OK). O software desenvolvido consistiu em um plugin (complemento) para o QGIS. O Coeficiente de Determinação (R^2) e a Raiz Quadrada do Erro Quadrático Médio (RMSE) foram utilizados para avaliar o desempenho dos métodos. O Índice de Correlação Espacial de Moran (ICM) foi utilizado para medir a dependência espacial. A validação dos modelos foi realizada em uma área de 204 ha. Para os atributos de solo analisados, o ICM variou de 0,52 a 0,84. Os resultados demonstraram que, para o ICM inferior a 0,76 a interpolação por ML foi superior ao OK.

PALAVRAS-CHAVE: agricultura de precisão, mapeamento de solos, inteligência artificial.

INTERPOLATION BY MACHINE LEARNING AND ORDINARY KRIGING IN QGIS USING THE SMART-MAP PLUGIN

ABSTRACT: Precision agriculture (PA) is an agricultural management technique based on observation, measurement and response to spatial and temporal variability that occurs in agricultural production areas. In PA, mapping the physical and chemical characteristics of the soil makes it possible to estimate the spatial variability of the soil with greater precision. Given a set of sampled points, it is desired to obtain maps of soil attributes or infer values at specific locations in non-sampled locations, for which interpolation techniques are used. The objective of this work was to implement software that allowed the interpolation of soil attributes through Machine Learning (ML) and Ordinary Kriging (OK) techniques. The software developed consisted of a plugin (add-on) for QGIS. The Coefficient of Determination (R^2) and the Root Mean Square Error (RMSE) were used to evaluate the performance of the methods. Moran's Spatial Correlation Index (ICM) was used to measure

spatial dependence. Validation of the models was carried out in an area of 204 ha. For the analyzed soil attributes, the ICM ranged from 0.52 to 0.84. The results demonstrated that, for ICM below 0.76, interpolation by ML was superior to OK.

KEYWORDS: precision agriculture, soil mapping, artificial intelligence.

INTRODUÇÃO: O mapeamento digital dos atributos do solo e das plantas fornecem informações para aplicação de insumos agrícolas a taxas variadas desde o desenvolvimento da agricultura de precisão (AP) (MALLA et al., 2020). O mapeamento do solo frequentemente envolve a interpolação de valores desconhecidos de pontos vizinhos que foram amostrados (HEDGE et al., 2017; HENGL; HEUVELINK; STEIN, 2004). Existem diversos métodos interpoladores com o objetivo de prever valores em pontos não amostrados. Inverso da Distância Ponderada (IDW) e Krigagem Ordinária (OK) tem sido as técnicas de interpolação espacial mais populares utilizadas em AP (SEKULIĆ et al., 2020). Embora o método de Krigagem seja um dos melhores interpoladores, ele requer alta densidade de amostragem e conhecimento do usuário para modelar o semivariograma. Por causa disso, esse método pode se tornar caro e não adequado para certas aplicações. Por outro lado, técnicas de *Machine Learning* (ML) têm sido utilizadas como alternativa aos métodos convencionais e geoestatísticos no mapeamento digital de atributos do solo. No entanto, os algoritmos de ML apresentam muitas variantes que podem dificultar a sua aplicação por usuários finais. Para preencher essa lacuna, o objetivo desse trabalho foi desenvolver o plugin *Smart-Map*, utilizando modernas ferramentas de inteligência artificial (IA). O plugin foi desenvolvido para ser utilizado como complemento no software de Sistema de Informação Geográfica (SIG) QGIS versão 3. Foram implementados no plugin o método geoestatístico OK e o modelo de ML *Support Vector Machine* (SVM), para geração de mapas interpolados com base em aprendizado de máquina e OK.

MATERIAL E MÉTODOS: Para apresentar a metodologia de OK e ML utilizada pelo *Smart-Map* foi realizado um estudo de caso conduzido em uma área de 204 ha localizada no município de São Desidério (12°25'12" S e 45°29'46" W), altitude média de 493 m, região oeste do estado da Bahia, Brasil (Figura 1). Nesta área foram coletados 204 pontos amostrais. No presente trabalho, para efeito de análise da performance dos métodos de interpolação, foram utilizados os seguintes atributos de solos: Areia Total (AT), Argila (ARG), Matéria Orgânica (MO) e Silte (SIL), totalizando assim 4 atributos de solos. A dependência espacial dos 4 atributos de solo analisados em cada área foi calculada de acordo com o Índice de Correlação de Moran (ICM). O ICM foi calculado de acordo com a equação 1 (LEGENDRE; FORTIN, 1989).

$$I = \frac{n}{W} \cdot \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x}) \cdot (x_j - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (1)$$

em que: n é o número de pontos amostrados; x_i, x_j representam os valores observados dos atributos de solo nos pontos i, j ; \bar{x} a média dos n pontos observados do atributo de solo; w_{ij} é a matriz de pesos espaciais com valor 0 na diagonal ($w_{ii} = 0$); e W representa a soma de todos os w_{ij} .

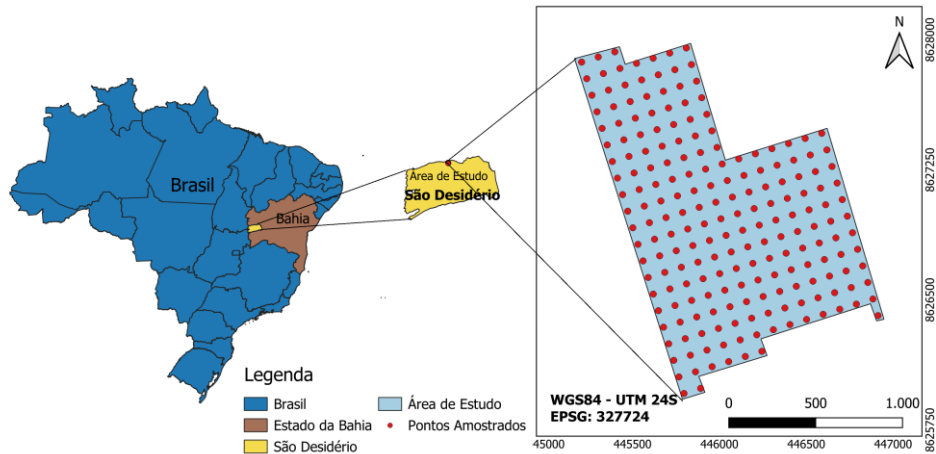


FIGURA 1. Localização geográfica da área de estudo e distribuição dos pontos amostrais em São Desidério, Oeste da Bahia, Brasil.

Os critérios de comparação utilizados para medir o desempenho entre os métodos foram o Coeficiente de Determinação (R^2) e a Raiz Quadrada do Erro Quadrático Médio (RMSE) da validação cruzada *leave-one-out cross-validation (LOOCV)* (CELISSE; ROBIN, 2008). O R^2 e o RMSE foram calculados de acordo com as Equações 2 e 3 respectivamente.

$$R^2 = 1 - \frac{\sum_{i=1}^n (x_i - \hat{x}_i)^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (2)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \hat{x}_i)^2} \quad (3)$$

em que: \hat{x}_i representa o valor estimado do atributo de solo no ponto i ; \bar{x} a média dos n pontos amostrados do atributo de solo; x_i o valor observado do atributo de solo no ponto i ; e n , o número de pontos amostrados.

RESULTADOS E DISCUSSÃO: *Smart-Map* foi registrado no Instituto Nacional de Propriedade Industrial (INPI, Ministério da Economia, Brasil, BR 51 2021 000002-1). A última versão pode ser instalada a partir do repositório de plugins do QGIS (https://plugins.qgis.org/plugins/Smart_Map). Os valores de ICM, p-valor, R^2 e RMSE para os 4 atributos de solo analisados separadamente, para os métodos OK e SVM podem ser visualizados na Tabela 1.

Tabela 1: Valores de Índice de Moran com pseudo p-valor, R^2 e RMSE calculados para os métodos OK e SVM para os quatro atributos de solo na área de estudo.

Atributo	Unidade	Moran	p-valor	OK		SVM	
				R2	RMSE	R2	RMSE
ARG ⁽¹⁾	(g kg ⁻¹)	0,843	0,001*	0,682	1,095	0,625	1,195
AT ⁽²⁾	(g kg ⁻¹)	0,766	0,001*	0,521	1,858	0,514	1,884
MO ⁽³⁾	(g dm ⁻³)	0,565	0,051	0,029	0,371	0,042	0,367
SIL ⁽⁴⁾	(g kg ⁻¹)	0,528	0,133	0,014	1,333	0,075	1,264

^{1/} ARG, Argila; ^{2/} AT, Areia Total; ^{3/} MO, Matéria Orgânica; ^{4/} SIL, Silte;

Obs: *Significância ao nível de 5% de probabilidade.

Os valores de RMSE estão relacionados ao R^2 . Como esperado, os valores de RMSE tendem a ser menores para valores maiores de R^2 . Com valores de R^2 maiores, pode-se inferir que

OK foi superior ao SVM na predição de ARG e AT, enquanto o SVM obteve R^2 superior ao OK para os atributos MO e SIL. O ICM para MO e SIL foram baixos e não significativos para estes dois atributos de solo analisados (Tabela 1). Os atributos ARG, AT, MO e SIL foram interpolados utilizando os métodos OK e SVM, através de um grid com células de tamanho 10 x 10 m, totalizando 18.736 pontos interpolados. Para o mapa obtido por interpolação da Argila no grid de 204 pontos as maiores concentrações de ARG estão localizadas na parte leste do mapa (Figura 2).

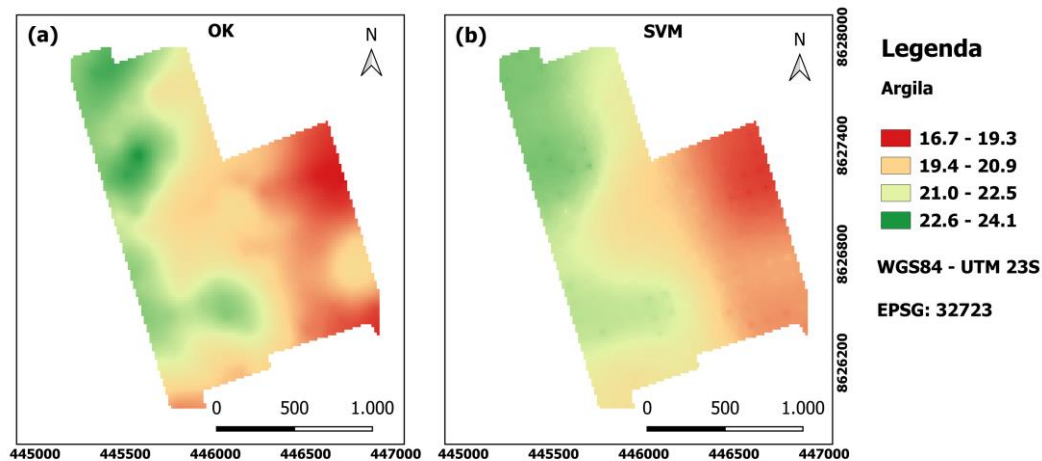


FIGURA 2. Mapa obtido por interpolação de Argila: (a) OK, (b) SVM

CONCLUSÕES: Os resultados apontaram que o método OK foi superior ao método SVM quando a correlação espacial medida pelo Índice de Moran foi superior a 0,76 significativo ao nível de 5% de probabilidade. Para valores de ICM inferior a 0,76 o método SVM foi superior ao método OK. Logo os resultados confirmaram que a técnica de *Machine Learning* se apresenta como uma alternativa para a interpolação de atributos do solo tanto em áreas com alta e baixa correlação espacial entre os pontos amostrados.

AGRADECIMENTOS: Esse estudo foi financiado em parte pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001 e Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

REFERÊNCIAS:

- CELISSE, A.; ROBIN, S. Nonparametric density estimation by exact leave-p-out cross validation. **Computational Statistics and Data Analysis**, v. 52, n. 5, p. 2350-2368, 2008.
- HEDGE, N. G. et al. Survey paper on Agriculture Yield Prediction Tool using Machine Learning. **International Journal of Advance Research in Computer Science and Management Studies**, v. 5, n. 11, p. 36-39, 2017.
- HENGL, T.; HEUVELINK, G. B. M.; STEIN, A. A generic framework for spatial prediction of soil variables based on regression-kriging. **Geoderma**, v. 120, n. 12, p. 75-93, 2004.
- LEGENDRE, P.; FORTIN, M.-J. Spatial pattern and ecological analysis. **Vegetatio**, v. 80, p. 107-138, 1989.
- MALLA, R. et al. Soil Fertility Mapping and Assessment of the Spatial Distribution of Sarlahi District, Nepal. **American Journal of Agricultural Science**, v. 7, n. 1, p. 8-16, 2020.
- SEKULIĆ, A. et al. Random Forest spatial interpolation. **Remote Sensing**, v. 12, n. 10, p. 1-29, 2020.