

SELEÇÃO DE MODELOS DE DEPENDÊNCIA ESPACIAL COM PADRÃO-OURO

MARCIO PAULO DE OLIVEIRA¹, MIGUEL ANGEL URIBE-OPAZO², GUSTAVO HENRIQUE DALPOSSO³

¹ Prof. Dr. UTFPR, Toledo, (45) 3379-6800, marcioliveira@utfpr.edu.br

² Prof. Dr. UNIOESTE, Cascavel, (45) 32203228, miguel.opazo@unioeste.br

³ Prof. Dr. UTFPR, Toledo, (45) 3379-6800, gustavodalposso@utfpr.edu.br

Apresentado no
L Congresso Brasileiro de Engenharia Agrícola - CONBEA 2021
08 a 10 de novembro de 2021 - Congresso On-line

RESUMO: A produtividade de culturas agrícolas, como o trigo, possui relevância em ciências agrárias é dependente de propriedades do solo e um monitoramento adequado da produtividade permite que ações de correção sejam realizadas a fim de incrementar a produtividade da cultura. O monitoramento de culturas agrícolas pode ser realizado por mapas temáticos obtidos por interpolação, como a krigagem. Para que a interpolação por krigagem possa ser realizada é necessário que um modelo de dependência espacial seja selecionado. O objetivo deste trabalho foi propor um método de seleção do modelo de dependência espacial com o uso de modelagem de concordância com o padrão-ouro e comparar a seleção com aquela realizada com a metodologia de validação cruzada geralmente utilizada. Os dados utilizados são de produtividade trigo de um talhão na mesorregião Oeste do Paraná no ano-safra de 2019-2020. Os resultados mostraram que o modelo selecionado com o padrão-ouro foi coincidente com aquele selecionado com a validação cruzada. A aplicação da modelagem de concordância com padrão-ouro, para seleção de modelos para interpolação por krigagem, indica uma alternativa ao método de validação cruzada.

PALAVRAS-CHAVE: concordância, geoestatística, variabilidade espacial

SELECTION OF SPATIAL DEPENDENCE MODELS WITH GOLD STANDARD

ABSTRACT: The productivity of agricultural crops such as wheat is relevant in agricultural sciences and depends of the soil properties. The adequate monitoring of productivity allows corrective actions to be carried out, in order to increase the productivity. Monitoring of agricultural crops can be carried out by thematic maps obtained by interpolation such as kriging. In order for the interpolation by kriging to be performed, it is necessary to select a spatial dependence model. The objective of this work was to propose a method of selection of the spatial dependence model using the agreement modeling with gold standard and to compare the selection with that performed by cross validation methodology, which is generally used. The data set of the wheat yield was obtained in the 2019-2020 agricultural year and comes from an area located in the West mesoregion of Paraná-Brazil. The results showed that the model selected with the gold standard was the same which was selected with cross-validation. The application of the agreement modeling with gold standard indicates an alternative to the cross-validation for the selection of models used in interpolation by kriging.

KEYWORDS: agreement, geostatistical, spatial variability.

INTRODUÇÃO: O trigo é uma cultura de inverno cuja produtividade possui importância no cenário econômico, pois é utilizado em diversos produtos que contribuem para a nutrição humana (VIEIRA et al., 2019). O estado do Paraná-Brasil é um dos principais produtores de trigo com cerca de 50% da produção brasileira (CONAB, 2017). As áreas de cultivo de trigo podem ser monitoradas por meio de mapas temáticos de propriedades do solo e da planta que permitem ao produtor acompanhar o desenvolvimento da cultura e se necessário realizar ações com vistas a melhoria das condições da área para o incremento de produtividade. A obtenção de um mapa temático com interpolação por krigagem é realizada por meio de um ajuste de modelo de dependência espacial que geralmente é selecionado considerando o método de Validação Cruzada, o critério de informação de Akaike e o critério Bayesiano (FARACO, et al., 2008), sendo que outras estatísticas podem também ser utilizadas (DE BASTIANI et al., 2015). Assim, este trabalho teve por objetivo propor a aplicação da metodologia de seleção com um padrão-ouro (OLIVEIRA et al., 2020) como um método alternativo para a seleção de modelos de dependência espacial para dados de produtividade de trigo do ano-safra 2019/2020 em uma área localizada na mesorregião Oeste do Paraná. Esta metodologia permite avaliar a relação da variabilidade dos valores observados com os valores preditos pelo modelo e indicar por meio de um coeficiente de concordância o modelo de dependência espacial a ser escolhido para a criação de um mapa da produtividade do trigo utilizando interpolação por krigagem.

MATERIAL E MÉTODOS: O estado do Paraná possui extensão territorial de 199.709 km e localiza-se na região sul do Brasil, sendo atravessado pelo trópico de Capricórnio (IBGE, 2010). A área em estudo da mesorregião Oeste do Paraná, é uma das produtoras de trigo e possui uma extensão territorial de 22.864,70 km². Os climas predominantes no estado são: temperado úmido com verão quente, subúmido com pequena deficiência hídrica, megatérmico e subtropical úmido com inverno seco (APARECIDO et al., 2016).

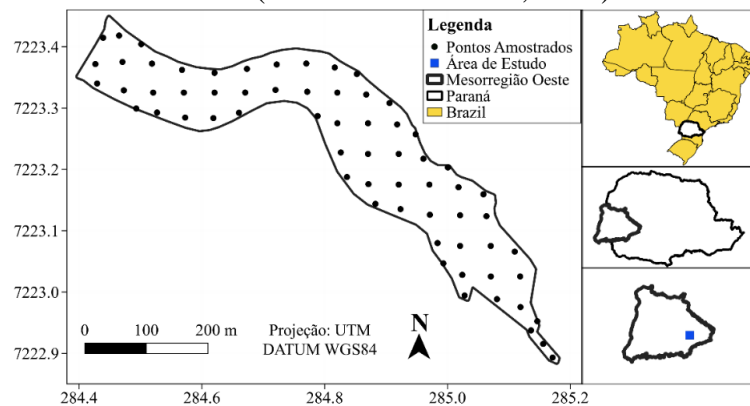


FIGURA 1. Mapa de localização da área em estudo na mesorregião Oeste do Paraná, contendo os pontos amostrais.

Os dados de produtividade de trigo ($t\ ha^{-1}$) foram obtidos no ano-safra de 2019/2020 em uma área com 9,73 ha. O modelo de comparação utilizado para avaliar o grau de concordância entre dois ou mais instrumentos de medição sob presença de um padrão-ouro, ou medida referência foi proposto por Laurent (1998) e pode ser escrito em notação matricial como

$$\mathbf{Y}_i = x_i \mathbf{1}_p + \boldsymbol{\epsilon}_i, \quad (1)$$

em que $\mathbf{Y}_i = (y_{i1}, \dots, y_{ip})^T$ é o vetor de medidas $p \times 1$ dos p métodos aproximados na unidade i , $\mathbf{1}_p$ é o vetor $p \times 1$ de elementos iguais 1 e $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \dots, \epsilon_{ip})^T$ é o vetor dos erros aleatórios $p \times 1$ dos p métodos na unidade i , para $i = 1, \dots, n$, sendo o padrão-ouro com média $E(x_i) = \mu$ e variância $Var(x_i) = \phi$, com ϵ_{ij} independente de x_i , com média $E(\epsilon_{ij}) = 0$ e variância $Var(\epsilon_{ij}) = \sigma^2$. As medidas padrão-ouro são as observações de produtividade do trigo nos

pontos amostrado e as medidas aproximadas são obtidas por krigagem por meio dos modelos de dependência espacial. Os estimadores de máxima verossimilhança (EML) dos parâmetros μ , ϕ e σ são dados, respectivamente, por $\hat{\mu} = \bar{x}$, $\hat{\phi} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$ e $\hat{\sigma} = \mathbf{S}_d = \frac{1}{n} \sum_{i=1}^n \mathbf{D}_i \mathbf{D}_i^T$, em que $\mathbf{D}_i = \mathbf{Y}_i - x_i \mathbf{1}_p$ para $i = 1, \dots, n$. O EML do coeficiente de concordância $\rho_j = \phi / (\phi + \sigma_{jj})$ para a concordância entre os métodos $j = 1, \dots, p$ sob teste e o padrão-ouro são dados por $\hat{\rho}_j = \hat{\phi} / (\hat{\phi} + \hat{\sigma}_{jj})$, sendo $\hat{\sigma}_{jj}$ os elementos da diagonal da matriz \mathbf{S}_d (GALEA, 2013). Na comparação com padrão-ouro o modelo a ser selecionado é deve possui maior valor de $\hat{\rho}_j$. Distribuições assimétricas podem também ser consideradas (OLIVEIRA et al., 2020).

RESULTADOS E DISCUSSÃO: As estatísticas descritivas para os dados da produtividade de trigo (Prod) em t ha⁻¹ na área em estudo são apresentadas na Tabela 1. Nessas estatísticas é possível observar pelo coeficiente de variação uma alta variação nos dados (30.17%) com uma produtividade média de 2.152 (t ha⁻¹).

TABELA 1. Estatísticas descritivas para os dados da produtividade de trigo (t ha⁻¹).

Variavel	n	Min	Q1	Med	Média	Q3	Max	DP	CV (%)	Ass	k
Prod	65	0,701	1,869	2,152	2,219	2,659	4,146	0,6694	30,17	0,3747	0,5348

Nota: Var: Variável, Prod: produtividade do trigo, n: número de observações, Min: valor mínimo, Q1: primeiro quartil, Med: mediana, Q3: terceiro quartil, Max: valor máximo, DP: desvio padrão, CV: coeficiente de variação, Ass: coeficiente de assimetria, k: coeficiente de curtose.

Os parâmetros dos modelos ajustados e o efeito pepita relativo são apresentados na Tabela 2. Por meio do efeito pepita relativo é possível observar dependência espacial moderada nos modelos estimados CAMBARDELLA et al. (1994).

TABELA 2. Modelos ajustados e parâmetros estimados por máxima verossimilhança.

Modelo	$\hat{\phi}_1$	$\hat{\phi}_2$	\hat{a}	\hat{E} (%)
Exponencial	0,238	0,200	0,122	0,543
Gaussiano	0,331	0,109	0,131	0,752
Matérn (k=1)	0,296	0,143	0,138	0,673
Matérn (k=1,5)	0,312	0,127	0,142	0,711

Nota: $\hat{\phi}_1$: efeito pepita estimado, $\hat{\phi}_2$: contribuição estimada, \hat{a} : alcance, $\hat{E} = 100 \cdot \hat{\phi}_1 / (\hat{\phi}_1 + \hat{\phi}_2)$ efeito pepita relativo.

Para a seleção dos modelos foi utilizado a validação cruzada, juntamente com os critérios de informação de Akaike-AIC e critério bayesiano-BIC. As estatísticas da validação cruzada e dos critérios de informação estão apresentadas na Tabela 3. Segundo Faraco et al. (2008) o melhor modelo é o que fornece o valor do erro médio - \widehat{EM} e erro médio relativo - \widehat{ER} mais próximo de zero, o valor do desvio padrão do erro médio - \widehat{DP}_{EM} menor e o valor do erro médio relativo - \widehat{DP}_{ER} mais próximo de um. Para os critérios de informação Akaike-AIC e Bayesiano-BIC quanto menor o valor melhor será o modelo ajustado. Assim, o modelo a ser selecionado como melhor ajuste é o modelo Gaussiano, obtido por máxima verossimilhança, seguindo os critérios da Tabela 3. A seleção de modelo foi também realizada com o uso da modelagem de concordância com padrão-ouro. O modelo Gaussiano apresentou o valor da medida de dependência espacial (SDM) igual a 6,618, o que indica dependência espacial fraca (Neto, et al., 2020). Neste caso as medidas padrão-ouro são aquelas observadas nos pontos amostrados e as medidas aproximadas são os valores preditos pelos modelos em estudo. As estimativas para os parâmetros são $\hat{\mu} = 2,219$, $\hat{\phi} = 0,4481$, $\hat{\sigma}_{11} = 0,4055$, $\hat{\sigma}_{22} = 0,4017$, $\hat{\sigma}_{33} = 0,4052$, $\hat{\sigma}_{44} = 0,4047$ e $\hat{\sigma}_{55} = 0,4054$ são utilizadas para o cálculo de $\hat{\rho}_j$ (Tabela 3). Dessa forma o modelo a ser escolhido é o modelo Gaussiano que é aquele que possui maior concordância com

o padrão-ouro. Esta seleção coincidiu com a seleção realizada pelo método de validação cruzada, AIC e BIC.

TABELA 3. Resultados da Validação cruzada e dos critérios de informação de Akaike-AIC e Bayrsiano- BIC e a concordância com padrão-ouro para seleção do melhor modelo.

Modelo	Método	\widehat{EM}	\widehat{ER}	\widehat{DP}_{EM}	\widehat{DP}_{ER}	\widehat{AIC}	\widehat{BIC}	$\hat{\rho}_j$
Exponencial	ML	0,0017	0,0014	0,6418	1,0100	135,7000	144,4000	0,5249
Gaussiano	ML	0,0007	0,0006	0,6387	1,0110	135,2000	143,9000	0,5273
Matérn (k=1)	ML	0,0013	0,0010	0,6415	1,0100	135,6000	144,3000	0,2520
Matérn (k=1,5)	ML	0,0011	0,0009	0,6411	1,0100	135,6000	144,3000	0,5255

Nota: ML: máxima verossimilhança, \widehat{EM} : erro médio estimado, \widehat{ER} : erro médio reduzido estimado, \widehat{DP}_{EM} : desvio padrão do erro médio estimado, \widehat{DP}_{ER} : desvio padrão dos erros reduzidos estimados, \widehat{AIC} : critério da informação de Akaike, \widehat{BIC} : critério de informação bayesiano.

CONCLUSÕES: A seleção dos modelos para interpolação por krigagem obtido com os métodos de validação cruzada e de concordância com padrão-ouro foi coincidente. A aplicação da modelagem de concordância com padrão-ouro, para seleção de modelos para interpolação por krigagem, indica uma alternativa ao método de validação cruzada.

AGRADECIMENTOS: Os autores agradecem a UTFPR, UNIOESTE, CAPES – Código de Financiamento 001, CNPq e Laboratório de Estatística Espacial (LEE).

REFERÊNCIAS:

- APARECIDO, L. E. DE O.; ROLIM, G. DE S.; RICHETTI, J.; SOUZA, P. S. DE; JOHANN, J. A. Köppen, Thornthwaite and Camargo climate classifications for climatic zoning in the State of Paraná, Brazil. **Ciência e Agrotecnologia**, 40, 405–417, 2016.
- APPEL, E.; SEIDEL, E.; OLIVEIRA, M. S. DE. Geostatistical-based index for spatial variability in soil properties. *Revista Brasileira de Ciência Do Solo*, 44, 2020.
- DE BASTIANI, F., DE AQUINO CYSNEIROS, A. H. M., URIBE-OPAZO, M. A., & GALEA M. Influence diagnostics in elliptical spatial linear models. *Test*, 24, 322-340. 2015. <https://doi.org/10.1007/s11749-014-0409-z>
- CAMBARDELLA, C. A.; MOORMAN, T. B.; NOVAK, J. M.; PARKIN, T. B.; KARLEN, D. L.; TURCO, R. F.; KONOPKA, A. E. Field-scale variability of soil properties in Central Iowa Soils. **Soil Science Society of American Journal**, v.58, p.1501-1511, 1994.
- CONAB. COMPANHIA NACIONAL DE ABASTECIMENTO. (2017). Acompanhamento da safra brasileira: grãos. V.5 - safra 2017/18 - N.3 - terceiro levantamento - dezembro 2017.
- DE BASTIANI, F.; MARIZ DE AQUINO CYSNEIROS, A. H.; URIBE-OPAZO, M. A.; GALEA, M. Influence diagnostics in elliptical spatial linear models. *TEST*, 24(2), 322–340, 2015.
- FARACO, M. A.; URIBE-OPAZO, M. A.; SILVA, E. A. A. DA; JOHANN, J. A.; BORSSOI, J. A. Seleção de modelos de variabilidade espacial para elaboração de mapas temáticos de atributos físicos do solo e produtividade da soja. *Revista Brasileira de Ciência Do Solo*, 32(2), 463–476, 2008.
- GALEA, M. Comparación de métodos de medición en presencia de un gold standar. **Estadística**, 65(185), 1–17, 2013.
- IBGE. INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. (2010). **Censo Demográfico**. Disponível em: <http://www.ibge.gov.br/estadosat/perfil.php?sigla=pr#>
- LAURENT, R. T. St. Evaluating Agreement with a Gold Standard in Method Comparison Studies. **Biometrics**, v. 54, n. 2, p. 537–545, 1998.
- OLIVEIRA, M. P.; URIBE-OPAZO, M. A.; GALEA, M.; JOHANN, J. A. Concordance Modeling With a Gold Standard for Variables From the Three-Parameter Gamma Distribution. **Journal of Agricultural Studies**, v. 8, n. 2, p. 284-305, 2020.
- VIEIRA, V. M.; OLIVEIRA, R. A. DE; DAROS, E. Factors related to the economic performance of wheat commercial fields. **Revista Ceres**, 66(5), 333–340, 2019.