

HOLE EFFECT NA VARIABILIDADE ESPACIAL DA PRODUTIVIDADE DA SOJA

GUSTAVO HENRIQUE DALPOSSO¹, MIGUEL ANGEL URIBE-OPAZO², MARCIO PAULO DE OLIVEIRA³

¹ Matemático, Prof. Doutor, PPGbio, UTFPR, Toledo-PR, Fone: (45) 3379-6800, gustavodalposso@utfpr.edu.br

² Estatístico, Prof. Doutor, PGEAGRI, UNIOESTE, Cascavel-PR

³ Matemático, Prof. Doutor, COMAT, UTFPR, Toledo-PR

Apresentado no
LI Congresso Brasileiro de Engenharia Agrícola - CONBEA 2022
27 a 29 de outubro de 2022 - Pelotas - RS, Brasil

RESUMO: O *hole effect* é um comportamento da função semivariância que não apresenta um crescimento monotônico como acontece tradicionalmente, deixando o semivariograma experimental com uma aparência senoidal. Ao trabalhar com a função de semivariância de Genton, que é considerado um estimador robusto ante a presença de *outliers*, percebeu-se que o semivariograma experimental da produtividade da soja na safra 2021/22, em uma área agrícola no oeste paranaense, apresentou o fenômeno *hole effect*. Ajustou-se o modelo wave e verificou-se por meio de intervalos de confiança *bootstrap* que as semivariâncias experimentais realmente seguiam um comportamento cíclico. O modelo wave se destacou ao ser comparado com os modelos exponencial e gaussiano, sendo escolhido como o melhor modelo pelo critério da validação cruzada. Outros destaques deste modelo foram o efeito pepita, o que fornece as equações de krigagem uma estabilidade e robustez e um considerável alcance, um indicativo de que alta precisão pode ser obtida por krigagem. O mapa obtido por krigagem utilizando o modelo wave indicou que a produtividade da soja foi baixa, em decorrência da escassez hídrica que atingiu a região.

PALAVRAS-CHAVE: semivariograma de Genton, modelo wave, *bootstrap* espacial.

HOLE EFFECT IN THE SPATIAL VARIABILITY OF SOYBEAN YIELD

ABSTRACT: The hole effect is a behavior of the semivariance function that does not present a monotonic growth as usually happens, leaving the experimental semivariogram with a sinusoidal appearance. When working with Genton's semivariance function, which is considered a robust estimator in the presence of outliers, it was noticed that the experimental semivariogram of soybean yield in the 2021/22 crop, in an agricultural area in western Paraná, showed the phenomenon hole effect. The wave model was adjusted and it was verified through bootstrap confidence intervals that the experimental semivariations followed a cyclical behavior. The wave model stood out when compared with the exponential and gaussian models, being chosen as the best model by cross-validation criterion. Other highlights of this model were the nugget effect, which provides the kriging equations with stability and robustness and a considerable range, an indication that high precision can be obtained by kriging. The map obtained by kriging using the wave model indicated that soybean yield was low, due to the water scarcity that affected the region.

KEYWORDS: Genton's estimator, wave model, spatial bootstrap.

INTRODUÇÃO: Na safra 2021/22 da produtividade da soja, as áreas da região oeste do estado do Paraná foram impactadas pela estiagem, o que restringiu o desenvolvimento das lavouras e a formação dos grãos da soja. A qualidade do produto também acabou sendo

reduzida, com uma proporção expressiva de grãos miúdos (CONAB, 2022). Ao realizar uma análise geoestatística com dados da produtividade da soja de uma área localizada na região afetada, percebeu-se que o semivariograma experimental não apresentava um comportamento monotônico. Quando um semivariograma não tem um crescimento monotônico e apresenta um fenômeno senoidal (picos e vales), diz-se que este semivariograma apresenta *hole effects* (JOURNAL & FROIDEVAUX, 1982). A presença desta estrutura fornece valiosas informações sobre a variabilidade espacial e ignorá-las pode resultar em modelos de heterogeneidade irreais que não reproduzem os padrões de variabilidade observados (PYRCZ & DEUTSCH, 2003). Este comportamento do semivariograma já foi observado em trabalhos relacionados a agricultura de precisão (SANTRA et al., 2008) e os melhores ajustes foram obtidos com o modelo wave. Neste contexto, este trabalho teve o objetivo de realizar uma análise geoestatística considerando a existência de *hole effects* no semivariograma experimental da produtividade da soja na safra 2021/22.

MATERIAL E MÉTODOS: Os 92 elementos amostrais (Figura 1) foram coletados no ano agrícola 2021/22 em uma área de 167,35 hectares, localizada na região Oeste do Paraná – Brasil.

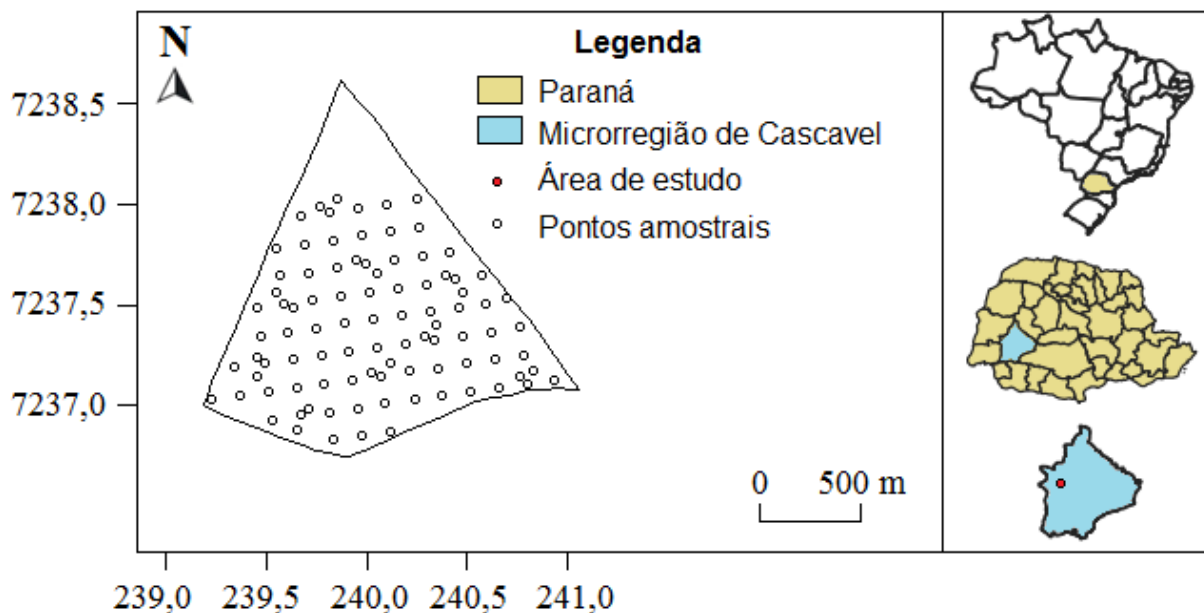


Figura 1 - Localização da área monitorada e posicionamento dos pontos amostrais.

Realizou-se uma análise exploratória espacial e utilizou-se o estimador da função semivariância de Genton (GENTON, 1998) (Equação 1) para identificar a estrutura de dependência espacial.

$$2\hat{\gamma}(h) = (Q_{N(h)})^2, \quad (1)$$

em que,

$\hat{\gamma}(h)$ é o estimador da semivariância de Genton;

$$Q_{N(h)} = d \left\{ |V_i(h) - V_j(h)|; i < j \right\}_{(k)};$$

$$V_i(h) = Z(s_i + h) - Z(s_i) \text{ e } V_j(h) = Z(s_j + h) - Z(s_j);$$

$d = 2,2191$ é a constante de consistência da distribuição normal (GENTON, 1998);

$$k = \left(\frac{\left\lfloor \frac{N(h)}{2} \right\rfloor + 1}{2} \right);$$

$\left\lfloor \frac{N(h)}{2} \right\rfloor$ denota a parte inteira de $\frac{N(h)}{2}$.

Conforme explica GENTON (1998), $Q_{N(h)}$ é o simples, explícito e altamente robusto estimador proposto por ROUSSEEUW & CROUX (1992), obtido pela multiplicação do fator d pelo k -ésimo quantil ($k \approx \frac{1}{4}$ para $N(h)$ grande) calculado do conjunto de todas as diferenças absolutas ordenadas $|V_i(h) - V_j(h)|$ para $i < j$.

Considerando o método dos mínimos quadrados ordinários, foram ajustados os modelos wave (DEUTSCH & JOURNAL, 1998) (Equação 2), exponencial (Equação 3) e gaussiano (Equação 4) e estimados seus parâmetros.

$$\gamma(h) = \begin{cases} \varphi_1 + \varphi_2 \left(1 - \frac{\varphi_3}{h} \operatorname{sen} \left(\frac{h}{\varphi_3} \right) \right), & h > 0 \\ 0, & h = 0 \end{cases} \quad (2)$$

$$\gamma(h) = \begin{cases} 0, & h = 0 \\ \varphi_1 + \varphi_2 \left[1 - \exp \left(-\frac{h}{\varphi_3} \right) \right], & 0 < h \leq \varphi_3, \\ \varphi_1 + \varphi_2, & h > \varphi_3 \end{cases} \quad (3)$$

$$\gamma(h) = \begin{cases} 0, & h = 0 \\ \varphi_1 + \varphi_2 \left[1 - \exp \left(-\frac{h}{\varphi_3} \right)^2 \right], & 0 < h \leq \varphi_3, \\ \varphi_1 + \varphi_2, & h > \varphi_3 \end{cases} \quad (4)$$

em que,

φ_1 é o efeito pepita;

φ_2 é a contribuição;

φ_3 é o parâmetro de alcance ($a = g(\varphi_3)$, em que a é o alcance).

O melhor ajuste foi selecionado pelo critério da validação cruzada (FARACO et al., 2008). Utilizou-se o método *bootstrap* espacial (SOLOW, 1985) (Algoritmo 1) para elaborar intervalos de confiança para as semivariâncias experimentais estimadas.

Algoritmo 1: Intervalos de confiança para as semivariâncias utilizando *bootstrap* espacial

- Considerando o conjunto de dados espaciais original $\{Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n)\}$, determine o vetor dos resíduos $\hat{\boldsymbol{\varepsilon}} = (Z(\mathbf{s}_1) - \hat{\mu}, \dots, Z(\mathbf{s}_n) - \hat{\mu})^T$ sendo $\hat{\mu} = (1^T \hat{\boldsymbol{\Sigma}}^{-1} 1)^{-1} 1^T \hat{\boldsymbol{\Sigma}}^{-1} Z$ o estimador de mínimos quadrados de μ e $\hat{\boldsymbol{\Sigma}}$ a matriz de covariância estimada;
- Considerando a matriz $\hat{\boldsymbol{\Sigma}}$, utilize o método de decomposição de Cholesky para obter e $\hat{\boldsymbol{\Sigma}} = \hat{\mathbf{L}} \hat{\mathbf{L}}^T$, em que $\hat{\mathbf{L}}$ é uma matriz triangular inferior de ordem n ;

- c) Utilizando a matriz $\hat{\mathbf{L}}^{-1}$, determine $\hat{\boldsymbol{\varepsilon}}_{dec} = \hat{\mathbf{L}}^{-1}\hat{\boldsymbol{\varepsilon}}$, o vetor dos resíduos descorrelacionados e centralize seus valores, obtendo $\tilde{\boldsymbol{\varepsilon}} = \hat{\boldsymbol{\varepsilon}}_{dec} - \left(\frac{1}{n}\right)\sum \hat{\boldsymbol{\varepsilon}}_{dec}$;
- d) Considerando o conjunto dos resíduos descorrelacionados e centralizados $\tilde{\varepsilon}_1, \dots, \tilde{\varepsilon}_n$, realize uma amostragem com reposição, obtendo o vetor $\boldsymbol{\varepsilon}_{SB}^* = (\varepsilon_1^*, \dots, \varepsilon_n^*)^T$;
- e) Obtenha a amostra bootstrap espacial $\mathbf{Z}^* = \hat{\boldsymbol{\mu}} + \hat{\mathbf{L}}\boldsymbol{\varepsilon}_{SB}^*$;
- f) Utilizando \mathbf{Z}^* , determine as semivariâncias experimentais;
- g) Repita os passos d) até f) mil vezes e armazene os resultados, para determinar as distribuições empíricas das semivariâncias;
- h) Utilize o método percentil (EFRON, 1982) para determinar os intervalos de confiança.

O modelo selecionado foi utilizado na krigagem para elaborar o mapa da produtividade da soja (t ha^{-1}) o que permitiu estimar a produção total na área monitorada.

RESULTADOS E DISCUSSÃO: Durante a safra 2021/22, os períodos prolongados de estiagem na região Oeste do Paraná acabaram atingindo as lavouras em fases críticas no ciclo fenológico da cultura, como floração e frutificação, corroborando para uma baixa produtividade. Este fato é observado nas estatísticas descritivas (Tabela 1), pois a produtividade média ficou abaixo do esperado.

Tabela 1 – Estatísticas da produtividade da soja (t ha^{-1}).

n	Min	Máx	Q ₁	\bar{x}	\tilde{x}	Q ₃	sd	\tilde{u}_3	CV(%)
92	0,670	1,830	0,950	1,103	1,055	1,260	0,231	0,774	21,0

n: Número de observações; Min: Mínimo, Máx: Máximo, Q₁: Primeiro quartil, \bar{x} : Média, \tilde{x} : Mediana, Q₃: Terceiro quartil, sd: desvio padrão; \tilde{u}_3 : Coeficiente de Assimetria, CV: Coeficiente de variação.

A maior concentração de valores na zona de valores mais reduzidos da amostra ocasionou uma assimetria positiva, classificando os dois maiores valores de produtividade encontrados ($1,800$ e $1,830 \text{ t ha}^{-1}$) como *outliers* (Figura 2).

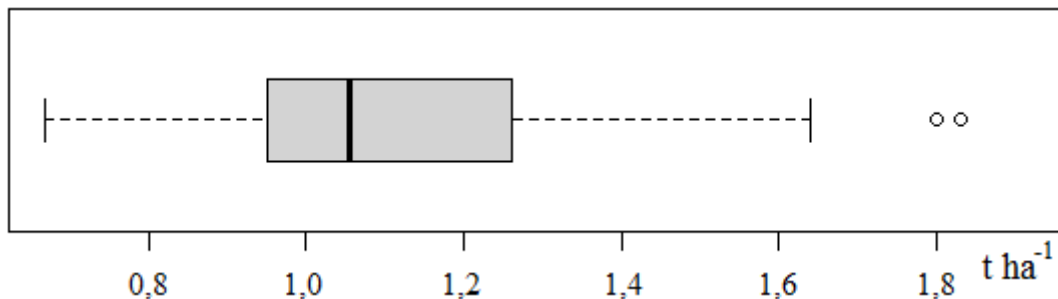


Figura 2 – Boxplot da produtividade da soja (t ha^{-1}) na área monitorada.

Como a presença de *outliers* altera o formato do semivariograma, exibindo estimativas não fidedignas, foi utilizado o estimador da função semivariância robusta de Genton para elaborar o semivariograma experimental da produtividade da soja (Figura 3).

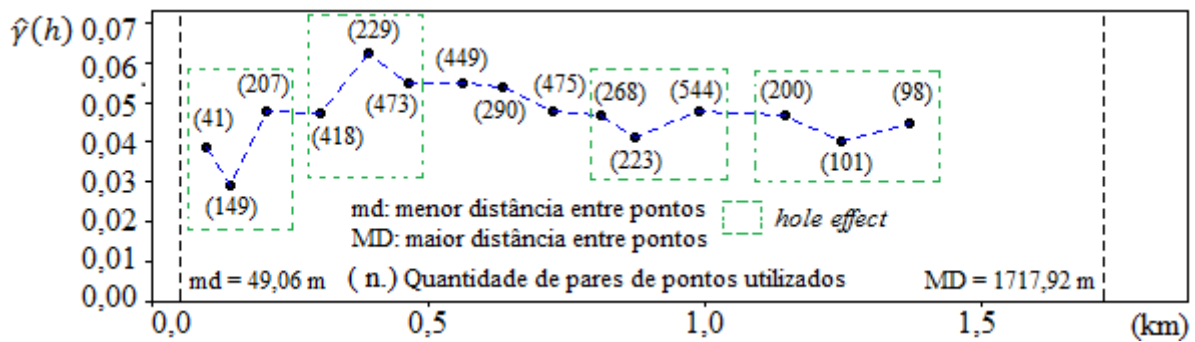


Figura 3 – Semivariograma experimental da produtividade da soja ($t\ ha^{-1}$).

No semivariograma experimental, observam-se regiões de picos e vales (buracos), conhecidas como *hole effect*. Embora algumas destas regiões estejam localizadas em distâncias superiores ao limite sugerido que é a metade da distância máxima (OLIVER & WEBSTER, 2015), destaca-se que são provenientes de semivariâncias calculadas com uma quantidade considerável de pares de pontos amostrais (maior que 30). Diante desta constatação, ajustou-se o modelo wave e, após a estimação dos parâmetros, utilizou-se o método *bootstrap* espacial para determinar intervalos de 95% de confiança para as semivariâncias (Figura 4). Observa-se uma dispersão nas semivariâncias usando intervalos de 95% de confiança, porém, destaca-se que a escala do eixo das ordenadas é pequena e a disposição dos intervalos de confiança reforça o comportamento cíclico das semivariâncias. A título de comparação, além do modelo wave, os usuais modelos exponencial e gaussiano também foram ajustados, e seus comportamentos podem ser observados na Figura 4.

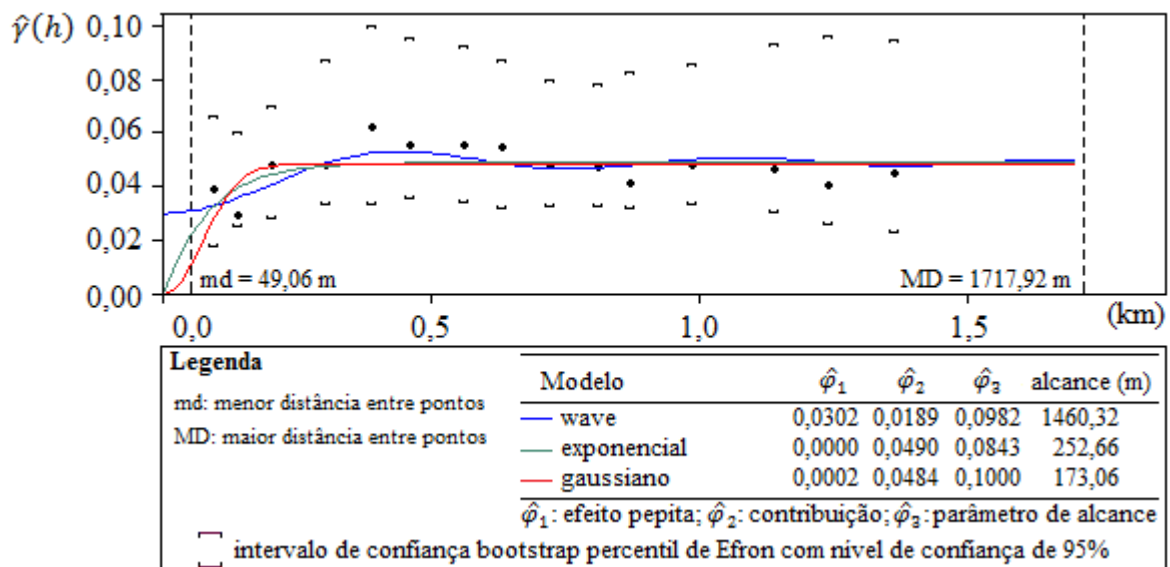


Figura 4 – Modelos ajustados ao semivariograma experimental.

Enquanto os modelos exponencial e gaussiano iniciaram na origem, o modelo wave apresentou um efeito pepita. Este fato é interessante pois além de não se ter conhecimento da variabilidade existente em distâncias inferiores a 49,06 m (menor distância entre pontos), em um sentido prático, o efeito pepita fornece às equações de krigagem uma estabilidade e robustez (SAMUI & SITHARAM, 2011). Sem um efeito pepita, a inversão das matrizes de krigagem pode levar a erros computacionais de arredondamento.

Outra característica destacada no modelo wave foi o seu raio de dependência espacial (1460,32 m), consideravelmente superior ao dos modelos exponencial e gaussiano. Este fato pode ser explicado pelo baixo coeficiente de variação dos dados amostrais (mesmo com a

existência de *outliers*) e é um indicativo de que alta precisão pode ser obtida por krigagem. O modelo wave foi escolhido como melhor modelo pelos critérios de validação cruzada.

A Figura 5 apresenta o mapa da produtividade da soja obtido por krigagem, juntamente com uma estimativa da área (ha) e da produção (ton) correspondente a cada classe. Os maiores valores de produtividade (última classe da legenda) foram detectados na região noroeste, parte da região central e em uma pequena faixa localizada na região sudoeste, totalizando uma área de 12,90 hectares (7,7% da área total) com uma produção estimada de 17,28 toneladas. Em 46% da área monitorada, a produtividade da soja manteve-se na segunda classe da legenda (valores variando de 1,000 t ha⁻¹ à 1,135 t ha⁻¹), proporcionando uma produção estimada de 82,33 toneladas.

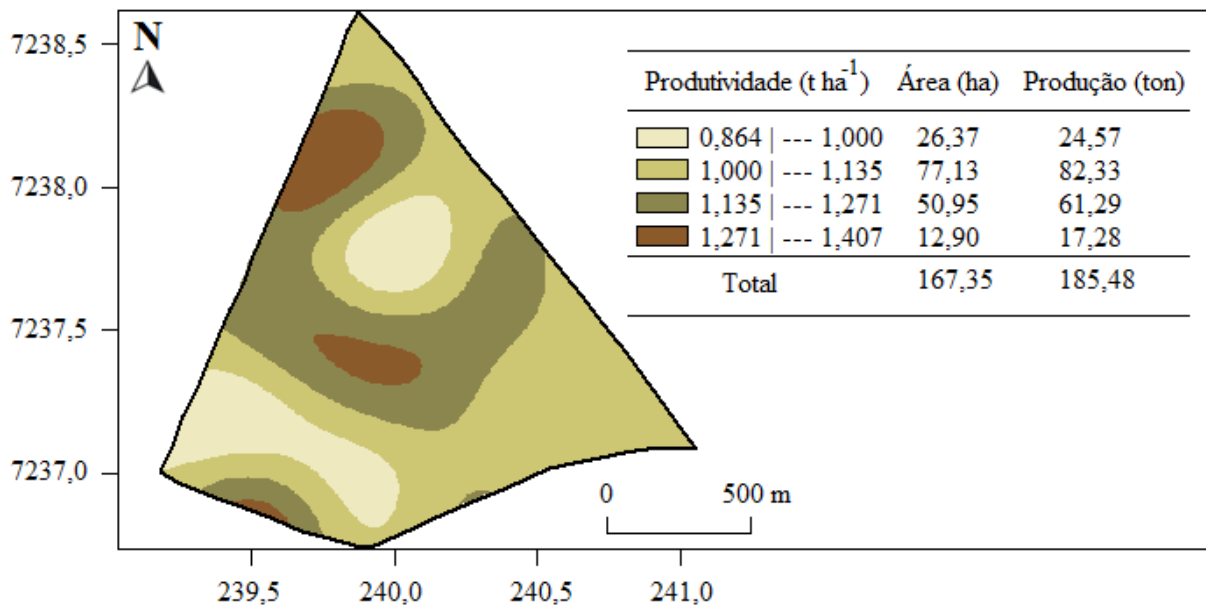


Figura 5 – Mapa da produtividade da soja (t ha⁻¹) com estimativa de área e produção.

A produção total da soja estimada de 185,48 toneladas em uma área de 167,35 hectares indica uma produtividade média de 1,10 t ha⁻¹, um valor considerado baixo. Esta baixa produtividade da soja identificada resulta da escassez hídrica que atingiu a região pois estudos realizados na mesma área em safras passadas indicaram valores mais elevados. Por exemplo, na safra 2012/13, 84% da área apresentou valores variando de 3,175 t ha⁻¹ à 3,592 t ha⁻¹ (DALPOSSO et al., 2018) e na safra 2017/18, a região central e o sudoeste da área apresentaram valores variando de 3,502 t ha⁻¹ à 4,215 t ha⁻¹ (DALPOSSO et al., 2021).

CONCLUSÕES: O estimador da função semivariância de Genton possibilitou trabalhar com *outlier* e identificar *hole effects* no semivariograma experimental. O modelo wave se mostrou eficaz, permitindo tanto ratificar a eficiência das semivariâncias estimadas quando utilizado com o método *bootstrap* espacial quanto estimar a produção total quando utilizado com krigagem. É evidente que a estiagem provocou uma quebra considerável na produtividade da soja na área.

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

REFERÊNCIAS

CONAB, 2022. Boletim de Monitoramento Agrícola, Brasília, v.11, n.02, p.1-20.

DALPOSSO, G.H.; URIBE-OPAZO, M.A.; DE BASTIANI, F. Spatial-temporal analysis of soybean productivity using geostatistical methods. *Journal of Agricultural Studies*, 9(2):283-303.

DALPOSSO, G.H.; URIBE-OPAZO, M.A.; JOHANN, J.A.; GALEA, M.; DE BASTIANI, F. 2018. Gaussian spatial linear model of soybean yield using bootstrap methods. *Engenharia Agrícola*, 38(1):110-116.

DEUTSCH, C. V., JOURNAL, A. G. 1998. *GSLIB: geostatistical software library and user's guide* 2nd ed.: Oxford University Press, New York, 369 pp.

EFRON, B. 1982. *The jackknife, the bootstrap, and other resampling plans*. Society for Industrial and Applied Mathematics, Philadelphia, PA.

FARACO, M.A., URIBE-OPAZO, M.A., SILVA, E.A.A., JOHANN, J., BORSSOI, J.A. 2008. Seleção de modelos de variabilidade espacial para elaboração de mapas temáticos de atributos físicos do solo e produtividade da soja. *Rev. Bras. Ciênc. Solo* 32, 463-476.

GENTON, M.G. 1998. Highly Robust Variogram Estimation. *Mathematical Geology*, v.30, n.2, p.213-221.

JOURNAL, A. G.; FROIDEVAUX, R.1982. Anisotropic Hole-Effect Modeling. *Mathematical Geology*, v.14, n.3, p.217-239.

OLIVER, M.A.; WEBSTER, R. 2015. *Basic Steps in Geostatistics: The Variogram and Kriging*. SpringerBriefs in Agriculture, 100p.

PYRCZ, M.; DEUTSCH, C. 2003. The whole story on the hole effect. *Geostatistical Association of Australasia Newsletter*, 18: 3–5.

ROUSSEEUW, P.J.; CROUX, C. 1992. Explicit scale estimators with high breakdown point, in DODGE, Y., ed., *LI Statistical analyses and related methods*: North-Holland, Amsterdam, p.77-92.

SAMUI, P.; SITHARAM, T.G. 2011. Application of Geostatistical Models for Estimating Spatial Variability of Rock Depth. *Engineering*, v.3, p.886-894.

SANTRA, P.; CHOPRA, U.K.; CHAKRABORTY, D. 2008. Spatial variability of soil properties and its application in predicting surface map of hydraulic parameters in an agricultural farm. *Current Science*, v.95, n.7, p.937-945.

SOLOW, A. 1985. Bootstrapping correlated data. *Math. Geol.*, v.17, n.7, p.769-775.