

ANÁLISE ESPACIAL DA PRODUTIVIDADE DA SOJA EM AMOSTRA REDUZIDA

Gustavo H. Dalposso¹, Fernando L. Incerti¹, Javier Kolodziej¹, Jones E. Schmitz¹

¹PPGBIO – Programa de Pós-Graduação em Tecnologias em Biociências, UTFPR,
Toledo-PR, Brasil
(gustavodalposso@utfpr.edu.br)

Resumo: Avaliou-se o desempenho do índice de Moran global clássico versus o modificado na análise da autocorrelação espacial da produtividade da soja (safra 2023/24) na Região Geográfica Imediata de Toledo-PR (n=14). Enquanto o índice global indicou autocorrelação positiva ($I=0,499$), o modificado ($I^*=-0,0546$) e os indicadores LISA não significativos evidenciaram distribuição aleatória. Conclui-se que o estimador modificado reduz interpretações equivocadas em amostras pequenas.

Palavras-chave: agricultura de precisão; Moran modificado; soja; LISA

INTRODUÇÃO

A soja consolidou-se como a principal *commodity* agrícola do agronegócio brasileiro, desempenhando um papel estratégico na inserção do país no mercado global (Darros; Ramos, 2026). No Paraná, a expansão desse cultivo não se limita à ocupação do solo, mas reflete as pressões por exportações que reconfiguram o território, sendo a região Oeste um polo central de alta tecnologia e produtividade (Lima; Silva; Freitas, 2025). Nesse contexto, a análise da eficiência e da previsibilidade desses mercados torna-se indispensável para compreender a resiliência da produção frente às oscilações internacionais e crises alimentares (Araujo *et al.*, 2020).

A agricultura não se desenvolve de forma isolada, mas está imersa em um espaço geográfico onde fatores ambientais e socioeconômicos interagem para criar padrões de dependência espacial (Rosano-Peña *et al.*, 2021). A estatística clássica, ao assumir a independência das observações, torna-se insuficiente para capturar a complexidade dessas interações, visto que o fenômeno da autocorrelação espacial frequentemente invalida as premissas de independência dos métodos tradicionais (Griffith, 1992). Nesse cenário, a Análise Exploratória de Dados Espaciais (AEDE) consolida-se como uma ferramenta fundamental, permitindo não apenas descrever a estrutura espacial dos fenômenos agrícolas, mas também superar os paradigmas da estatística convencional ao tratar a dependência não como um erro, mas como informação latente no conjunto de dados (Legendre, 1993).

A aplicação de indicadores clássicos de dependência espacial, como o Índice de Moran tradicional, em estudos geográficos com um número reduzido de unidades amostrais, impõe desafios severos à validade

inferencial, dada a tendência desses estimadores em subestimar a autocorrelação real e apresentar instabilidades estatísticas (Carrijo; Silva, 2017). A negligência quanto à estrutura de dependência ao trabalhar com séries espaciais curtas frequentemente resulta em estimativas enviesadas, comprometendo a precisão de modelos de risco agrícola e a interpretação da dinâmica produtiva (Hungerford; Goodwin, 2014). Esse cenário é agravado pela ineficiência dos testes convencionais de autocorrelação, que demonstram inconsistências marcantes entre as expectativas teóricas e empíricas quando o tamanho da amostra é insuficiente (Huitema; McKean, 1991). Desta forma, torna-se imperativa a adoção de metodologias robustas, capazes de garantir a confiabilidade da inferência estatística mesmo sob as restrições inerentes a amostras reduzidas.

Nesse contexto, o presente trabalho objetiva avaliar, de forma comparativa, o desempenho do índice I de Moran global frente ao Índice de Moran modificado, I^* , na análise da autocorrelação espacial da produtividade da soja. Utilizando como estudo de caso a Região Geográfica Imediata de Toledo, composta por catorze municípios, a pesquisa procura demonstrar como o estimador modificado minimiza as instabilidades estatísticas inerentes ao uso de amostras reduzidas, conferindo maior robustez e confiabilidade aos resultados obtidos em comparação aos métodos clássicos.

MATERIAL E MÉTODOS

A área analisada situa-se no oeste paranaense e compreende a Região Geográfica Imediata de Toledo, conforme ilustrado na Figura 1. Esse limite territorial, estabelecido na atualização das divisões regionais do

país realizada pelo IBGE em 2017, está inserido na Região Geográfica Intermediária de Cascavel.

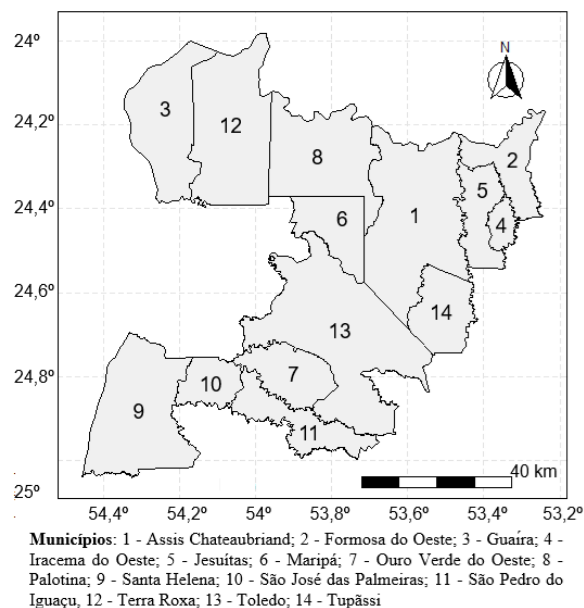


Figura 1. Região Geográfica Imediata de Toledo.

Para esta análise, utilizaram-se as estatísticas oficiais de produtividade da soja ($t\ ha^{-1}$) referentes à safra 2023/2024, disponibilizadas pelo Departamento de Economia Rural (Seab, 2025), órgão integrante da Secretaria de Estado da Agricultura e do Abastecimento do Paraná.

Análise exploratória

A análise exploratória inicial fundamentou-se na aplicação de estatísticas descritivas, como média, mediana, medidas de variabilidade e assimetria, para estruturar a interpretação dos dados de produtividade da soja (Bussab; Morettin, 2017). Complementarmente, utilizou-se o gráfico *boxplot* como uma ferramenta robusta para sintetizar a distribuição e a dispersão dos dados, facilitando a visualização da posição central, da assimetria e da presença de *outliers* (Martinez et al., 2017).

A espacialização da produtividade foi realizada por meio do método de classificação por desvio padrão, que estabelece os intervalos de classe a partir da média aritmética e seus múltiplos para mensurar a dispersão dos dados.

Análise da autocorrelação espacial

A análise da autocorrelação espacial foi iniciada com a definição da matriz de pesos espaciais (W),

adotando-se o critério de contiguidade do tipo *Queen* (Rainha) para estabelecer a conectividade entre as unidades de estudo. Essa etapa é fundamental para a modelagem estatística, visto que a literatura aponta que a escolha do critério de vizinhança deve ser teoricamente fundamentada, dado que os resultados dos testes subsequentes de autocorrelação global e local são sensíveis a essa especificação (Cliff; Ord, 1981).

Para a análise da autocorrelação espacial em nível global, utilizou-se o índice I de Moran, apresentado na Equação 1.

$$I = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (z_i - \bar{z})(z_j - \bar{z})}{\sum_{i=1}^n (z_i - \bar{z})^2} \quad [1]$$

Na equação acima, n é o número de áreas, z_i representa valor da variável na área i , $\bar{z}$ representa a média da variável na região monitorada e w_{ij} são os elementos da matriz de proximidade espacial. A significância estatística do Índice I de Moran global foi aferida por meio de um teste de permutação, que gera a distribuição empírica sob a hipótese nula de aleatoriedade espacial a partir de múltiplas reamostragens dos dados observados (Anselin, 1995).

O índice I de Moran global atua como uma estatística de síntese, capturando a estrutura de dependência espacial de todo o conjunto de dados em um único escalar. Contudo, para decompor essa medida e compreender como as interações locais contribuem para o padrão global, utilizou-se o diagrama de espalhamento de Moran, formalizado por Anselin (1996). Este diagrama consiste em um gráfico bidimensional que permite visualizar e interpretar a relação linear entre os valores observados e sua respectiva defasagem espacial.

Entretanto, uma limitação do índice de Moran global reside no fato de ser uma única medida para toda a área em estudo, o que pode ocultar heterogeneidades locais, como a coexistência de *clusters* e *outliers* em diferentes partes do território (Anselin, 1995). Para mitigar esse viés e identificar instabilidades locais, utilizou-se o Indicador Local de Associação Espacial (LISA, do inglês *Local Indicators of Spatial Association*), apresentado na forma do Índice de Moran Local, na Equação 2:

$$I_i = \frac{z_i \sum_{j=1}^n w_{ij} z_j}{\sum_{j=1}^n z_j^2} \quad [2]$$

Este índice decompõe o indicador global e permite avaliar a significância da dependência espacial em cada unidade geográfica individualmente (Bivand; Wong, 2018). A aferição da significância estatística do Índice de Moran Local foi realizada de maneira

semelhando ao caso global, utilizando permutações aleatórias para gerar uma distribuição de referência sob a hipótese nula de aleatoriedade espacial (Anselin, 1995).

Índice de Moran Modificado

Para minimizar a instabilidade estimatória do Índice de Moran clássico frente ao tamanho reduzido do conjunto de dados, adotou-se o Índice de Moran Modificado, apresentado na Equação 3.

$$I^* = \frac{\sum_{i=1}^n (x_i - \bar{x})(\sum_j w_{ij} x_j - \bar{x})}{[\sum_{i=1}^n (x_i - \bar{x})^2]^{1/2} [\sum_{i=1}^n (\sum_j w_{ij} x_j - \bar{x})^2]^{1/2}} \quad [3]$$

Proposta por Carrijo e Silva (2017), esta abordagem adequa a análise estatística a cenários de amostras limitadas, oferecendo um estimador mais robusto e estatisticamente apropriado para a avaliação da autocorrelação espacial.

Recursos computacionais

A análise dos dados foi conduzida no ambiente computacional R (R Core Team, 2025). As estatísticas descritivas e a construção do *boxplot* foram obtidas por meio de funções nativas do programa, enquanto as classes do mapa coroplético foram elaboradas com o auxílio do pacote *classInt* (Bivand, 2024). Para a definição da matriz de pesos espaciais, o cálculo do Índice de Moran global e a determinação dos indicadores LISA, utilizaram-se as funções do pacote *sdep* (BIVAND, 2022). Adicionalmente, o Índice de Moran Modificado foi implementado manualmente.

RESULTADOS E DISCUSSÃO

A Tabela 1 apresenta as estatísticas descritivas da produtividade da soja ($t \cdot ha^{-1}$) na safra 2023/24. Embora a produtividade média ($3,174 t \cdot ha^{-1}$) tenha sido inferior às médias do Paraná ($3,307 t \cdot ha^{-1}$) e do Brasil ($3,282 t \cdot ha^{-1}$) reportadas pela CONAB (2024), a expressiva variação observada de $2,000$ a $3,800 t \cdot ha^{-1}$ evidencia uma desigualdade local da produtividade.

A produtividade da soja na região apresentou uma mediana de $3,288 t \cdot ha^{-1}$, o que indica que exatamente metade dos municípios monitorados apresentaram valores de produtividade iguais ou superiores a este valor. A assimetria registrada foi de $-0,848$, caracterizando uma distribuição assimétrica negativa (ou à esquerda). Esse valor sugere uma maior concentração de municípios apresentando altas

produtividades, com a presença de alguns valores mais baixos que acabam puxando a média para um patamar inferior ao da mediana.

Tabela 1. Estatísticas descritivas da produtividade da soja ($t \cdot ha^{-1}$) na safra 2023/24.

Mínimo	2,000
Máximo	3,800
Primeiro quartil	3,025
Média	3,174
Mediana	3,288
Terceiro quartil	3,546
Desvio padrão	0,506
Assimetria	-0,848
Coeficiente de variação (CV%)	15,94

Por fim, o desvio padrão de $0,506 t \cdot ha^{-1}$ reflete a dispersão absoluta dos dados em torno dessa média, quantificando a variabilidade e confirmando a desigualdade local. A fim de permitir a observação gráfica do formato da distribuição dos dados e caracterizar visualmente o comportamento da cauda que gerou essa assimetria negativa, o *boxplot* da produtividade da soja é apresentado na Figura 2.

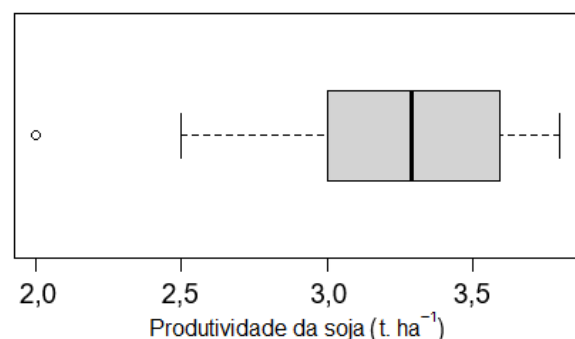


Figura 2. Boxplot da produtividade de soja ($t \cdot ha^{-1}$).

A análise do *boxplot* revela que a produtividade da soja apresenta estabilidade na maior parte da área de estudo, com 50% dos dados centrais concentrados entre o primeiro ($Q_1 = 3,025 t \cdot ha^{-1}$) e o terceiro quartil ($Q_3 = 3,546 t \cdot ha^{-1}$). Nota-se, contudo, a ocorrência de um valor atípico (*outlier*) posicionado abaixo do limite inferior da distribuição, correspondente a $2,000 t \cdot ha^{-1}$. Para espacializar essa variação e melhor compreender a distribuição territorial da produtividade, recorreu-se a técnicas de Análise Exploratória de Dados Espaciais (AEDE/ESDA), resultando no mapa coroplético classificado pelo método do desvio padrão (Figura 3)

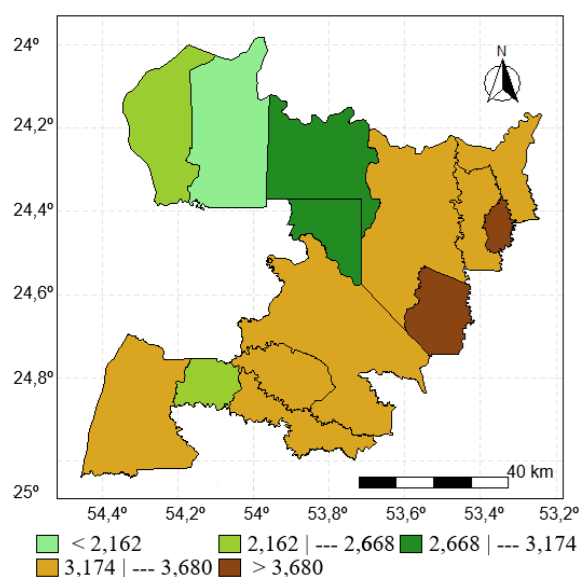


Figura 3. Mapa coroplético da produtividade da soja ($t\ ha^{-1}$) classificado pelo método do desvio padrão.

A visualização da variabilidade espacial da produtividade da soja, por meio do mapa coroplético classificado pelo método do desvio padrão (Figura 3), permite identificar padrões claros de heterogeneidade na região estudada. Observa-se que o município de Terra Roxa (código 12) configura-se isoladamente na primeira classe do mapeamento, o que evidencia uma acentuada restrição produtiva local. Segundo Pinto et al. (2025), o solo de Terra Roxa é predominantemente um argissolo vermelho-amarelo distrófico com textura arenosa na superfície, o que proporciona menor produtividade em comparação com os latossolos vermelhos predominantes em grande parte da região.

Em contrapartida, constata-se uma expressiva concentração espacial em níveis intermediários de rendimento, com 50% dos municípios classificados na faixa correspondente à classe da média mais um desvio padrão. Por fim, a última classe do mapeamento, caracterizada por valores acima de $3,680\ t\ ha^{-1}$, é composta exclusivamente pelos municípios de Tupãssi (código 14) e Iracema do Oeste (código 4), que registraram as maiores produtividades na área em estudo.

Após a identificação dos padrões visuais na distribuição dos dados, avança-se para a investigação da autocorrelação espacial em nível global por meio do diagrama de espalhamento de Moran (Figura 4). O gráfico, construído com base nos valores normalizados da produtividade em relação à média dos seus vizinhos, serve como base para quantificar a estrutura de dependência espacial da região.

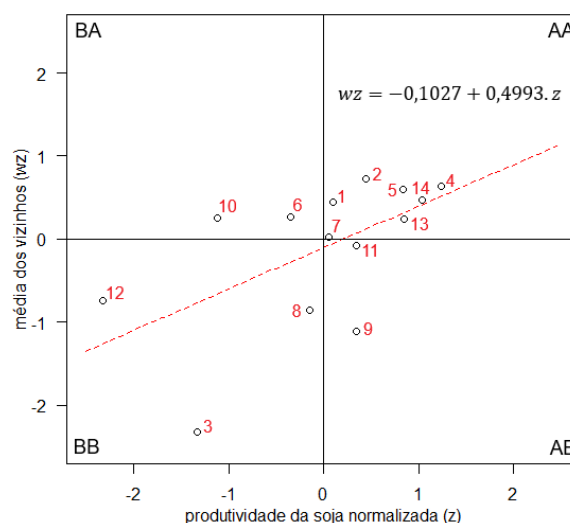


Figura 4. Diagrama de espalhamento de Moran da produtividade da soja ($t\ ha^{-1}$).

O diagrama de espalhamento de Moran (Figura 4) revela um índice I de Moran global de 0,4993, o que confirma a existência de autocorrelação espacial positiva na produtividade da soja. Este resultado ratifica a presença de *clusters* espaciais caracterizados pela similaridade de atributos entre localizações vizinhas (Anselin, 2024). Identifica-se uma concentração significativa de municípios no quadrante AA (Alto-Alto), denotando um padrão espacial de alta produtividade, enquanto o quadrante BB (Baixo-Baixo) agrupa os municípios de Guaíra (código 3), Palotina (código 8) e Terra Roxa (código 12), que apresentam desempenho inferior circundados por vizinhos em condições análogas.

As observações situadas nos quadrantes AB (valor alto cercado por vizinhos com valores baixos) e BA (valor baixo cercado por vizinhos com valores altos) podem ser caracterizadas como *outliers* espaciais (Zafeiris et al., 2022). A baixa incidência destas ocorrências corrobora a tendência de autocorrelação espacial positiva, afastando a hipótese de um padrão de autocorrelação espacial global negativa, conforme discutido por Dubé e Legros (2014). Para validar a significância estatística desse padrão, adotou-se um teste de pseudo-significância baseado em permutações, abordagem robusta que dispensa pressupostos *a priori* sobre a distribuição dos dados. Conforme ilustrado na Figura 5, o histograma das simulações permite confrontar o valor observado do índice I de Moran global com a distribuição empírica gerada, evidenciando sua relevância estatística.

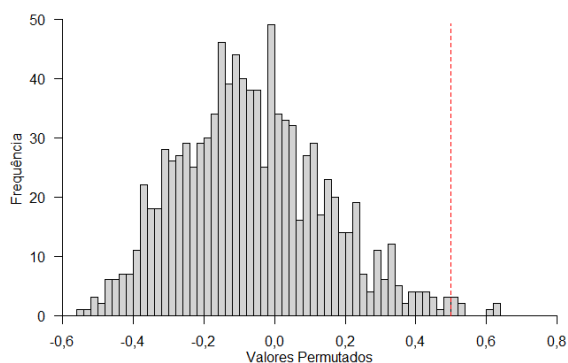


Figura 5. Histograma das simulações do índice I de Moran global via teste de permutação. Em vermelho, o verdadeiro valor do índice.

Observamos que o valor do índice I de Moran global localiza-se no extremo da distribuição simulada e conforme explicam Druck et al. (2004), trata-se de valor com significância estatística. Entretanto, a significância do índice global não permite identificar quais unidades espaciais contribuem para o padrão observado de autocorrelação espacial. Dessa forma, torna-se necessária a análise dos indicadores locais de associação espacial (LISA), que possibilitam identificar a localização de agrupamentos espaciais e verificar se a autocorrelação espacial global observada se manifesta efetivamente em nível local. A Tabela 2 apresenta os valores do Índice de Moran Local para a produtividade da soja.

Tabela 2. Índice de Moran Local da produtividade da soja ($t. ha^{-1}$).

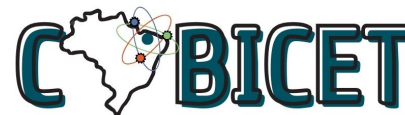
Municípios	I_i	p-valor
1 - Assis Chateaubriand	0,04878	0,2577
2 - Formosa do Oeste	0,34883	0,1399
3 - Guaíra	3,32339	0,0739
4 - Iracema do Oeste	0,85312	0,2418
5 - Jesuítas	0,53407	0,1958
6 - Maripá	-0,09834	0,7752
7 - Ouro Verde do Oeste	0,00146	0,9491
8 - Palotina	0,13359	0,1858
9 - Santa Helena	-0,41803	0,3536
10 - São José das Palmeiras	-0,29956	0,8691
11 - São Pedro do Iguaçu	-0,02737	0,8771
12 - Terra Roxa	1,84319	0,0959
13 - Toledo	0,21770	0,5554
14 - Tupãssi	0,52880	0,3796

Embora o índice de Moran global tenha resultado em $I = 0,4993$, a análise da Tabela 2 evidencia que os indicadores locais não alcançaram significância estatística, ou seja todos tiveram p-valores maiores do que 0,05. Esse padrão divergente é um indicativo clássico das limitações do estimador de Moran em cenários de pequena amostragem ($n=14$), onde a escassez de vizinhos para a matriz de pesos pode resultar em estimativas instáveis e enviesadas. Conforme demonstrado por Carrijo e Silva (2017), o índice clássico tende a subestimar ou produzir resultados inconsistentes em amostras reduzidas, o que justifica a aplicação de metodologias alternativas.

Ao calcular o índice de Moran modificado, obteve-se $I^* = -0,0546$ e, como este valor é extremamente próximo de zero, indica a inexistência de autocorrelação espacial significativa. Segundo Zhang et al. (2019), um índice nulo demonstra que não há agregação espacial em torno da área de estudo, o que implica uma distribuição aleatória da produtividade da soja na região investigada. Portanto, a ausência de significância nos indicadores locais mostra-se consistente com o valor do índice de Moran modificado. Esse achado reforça a necessidade de priorizar a métrica modificada para a inferência da autocorrelação espacial em amostras reduzidas, evitando interpretações equivocadas de *clusters* que, na realidade, decorrem da instabilidade do estimador clássico e não de uma estrutura espacial genuína.

CONCLUSÃO

Em síntese, este estudo reforça a necessidade de cautela na aplicação de técnicas de estatística espacial em amostras reduzidas. Os resultados mostraram elevada variabilidade na produtividade da soja na Região Geográfica Imediata de Toledo, com distribuição assimétrica dos dados. Embora o índice de Moran global tenha indicado autocorrelação espacial positiva, essa evidência não foi confirmada pelos indicadores locais, o que sugere limitações desse método em amostras pequenas. Por outro lado, o Índice de Moran modificado indicou ausência de autocorrelação espacial, apontando para um padrão de distribuição aleatório. Dessa forma, os resultados demonstram que o uso do índice de Moran modificado pode reduzir interpretações equivocadas sobre a existência de agrupamentos espaciais, contribuindo para análises mais confiáveis e para a tomada de decisões.



AGRADECIMENTOS

Os autores agradecem ao Programa de Pós-Graduação em Tecnologias em Biociências (PPGBio) da Universidade Tecnológica Federal do Paraná (UTFPR), pelo suporte acadêmico e institucional que viabilizou a realização desta pesquisa.

REFERÊNCIAS

ANSELIN, L. **An Introduction to Spatial Data Science with GeoDa Volume 1: Exploring Spatial Data**. Boca Raton: CRC Press, 2024. 444 p.

ANSELIN, L. The Moran Scatterplot as an ESDA Tool to Assess Local Instability in Spatial Association. In: FISCHER, M. M.; SCHOLTEN, H. J.; UNWIN, D. (Ed.). **Spatial Analytical Perspectives on GIS**. London: Taylor & Francis, 1996. p. 111-125.

ANSELIN, L. Local Indicators of Spatial Association—LISA. *Geographical Analysis*, v. 27, n. 2, p. 93–115, 1995

ARAUJO, F. H. A. de; BEJAN, L.; STOSIC, B.; STOSIC, T. An analysis of Brazilian agricultural commodities using permutation information theory quantifiers: The influence of food crisis. **Chaos, Solitons and Fractals**, v. 139, 110081, 2020

ASWI, A.; CRAMB, S.; DUNCAN, E.; MENERSEN, K. Detecting spatial autocorrelation for a small number of areas: a practical example. **Journal of Physics: Conference Series**, v. 1899, n. 1, 012098, 2021.

BIVAND R (2024) classInt: Choose Univariate Class Intervals. R package version 0.4-11.

BIVAND, R. S. (2022). spdep: Spatial Dependence: Weighting Schemes, Statistics. R Package Version 1.2-1

BIVAND, R. S.; WONG, D. W. S. Comparing implementations of global and local indicators of spatial association. *TEST*, v. 27, n. 3, p. 716–748, 2018.

BUSSAB, W.; MORETTIN, P. A. **Estatística básica**. 9. ed. São Paulo: Saraiva, 2017.

CARRIJO, T. B.; SILVA, A. R. da. Modified Moran's I for Small Samples. **Geographical Analysis**, v. 49, n. 4, p. 451-467, 2017.

CLIFF, A. D.; ORD, J. K. **Spatial processes: models & applications**. London: Pion, 1981. 266 p.

DARROS, R. de A.; RAMOS, L. P. dos S. Concentração de commodities na pauta exportadora brasileira, no comércio bilateral com a China: uma análise sob a ótica da deterioração dos termos de troca. **Revista RGE Interdisciplinar**, v. 17, n. 5, p. 1-25, 2026

DRUCK, S. et al. (org.). *Análise espacial de dados geográficos*. Brasília, DF: EMBRAPA, 2004. 209 p.

DUBÉ, J.; LEGROS, D. **Spatial Econometrics Using Microdata**. Reino Unido: Wiley, 2014. 242 p.

COMPANHIA NACIONAL DE ABASTECIMENTO (Brasil). **Séries históricas de safras: Soja**. Brasília: CONAB, 2024. Disponível em: <https://www.gov.br/conab/pt-br/atuacao/informacoes-agropecuarias/safras/series-historicas/gaos/soja/sojaseriehist.xls/view>. Acesso em: 5 jun. 2026.

GRIFFITH, D. A. What is spatial autocorrelation? Reflections on the past 25 years of spatial statistics. **L'Espace géographique**, v. 21, n. 3, p. 265-280, 1992.

HUITEMA, B.E.; MCKEAN, J.W. Autocorrelation Estimation and Inference With Small Samples. **Psychological Bulletin**, v. 110, n. 2, p. 291-304, 1991

HUNGERFORD, A. E.; GOODWIN, B. Big assumptions for small samples in crop insurance. **Agricultural Finance Review**, v. 74, n. 4, p. 477-491, 2014

LEGENDRE, P. Spatial autocorrelation: trouble or new paradigm? **Ecology**, v. 74, n. 6, p. 1659-1673, 1993

LIMA, J. M. A. de; SILVA, M.; FREITAS, E. P. de. Geopolitics and the territorialization of soybean production in the State of Paraná. **Revista Cerrados**, v. 23, n. 01, p. 101-131, 2025

MARTINEZ, W. L.; MARTINEZ, A. R.; SOLKA, JEFFREY. **Exploratory data analysis with MATLAB**. Chapman and Hall/CRC, 2017.

PINTO, L.A.S.R. *et al.* Management systems in the subtropical region of Brazil: aggregation in the assessment of the chemical quality of fragile soils. **Experimental Agriculture**, v. 61, p. 1-20, 2025



R CORE TEAM. **R: A Language and Environment for Statistical Computing**. Vienna, Austria: R Foundation for Statistical Computing, 2025.

ROSANO-PENÑA, C.; ALMEIDA, C. A. R.; RODRIGUES, E. C. C.; SERRANO, A. L. M. Dependência Espacial da Ecoeficiência da Agricultura em São Paulo. **BBR - Brazilian Business Review**, v. 17, n. 3, p. 329-343, 2020

SEAB (2025) Secretaria de Estado da Agricultura e do Abastecimento. Departamento de Economia Rural. **Diagnóstico Agropecuário Paranaense: Safra 2023/2024**. Disponível em: <https://www.agricultura.pr.gov.br>

ZAFEIRIS, Konstantinos N. et al. (org.). **Data analysis and related applications, volume 1: computational, algorithmic and applied economic data analysis**. Hoboken: John Wiley & Sons, 2022. 480 p.

ZHANG, H.; YANG, L.; LI, L.; XU, G.; ZHANG, X. The epidemic characteristics and spatial autocorrelation analysis of hand, foot and mouth disease from 2010 to 2015 in Shantou, Guangdong, China. **BMC Public Health**, v. 19, n. 998, 2019.