

ANÁLISE EXPLORATÓRIA DE DADOS ESPACIAIS APLICADA A PRODUTIVIDADE DE SOJA NA REGIÃO OESTE DO PARANÁ

**RODOLFO SEFFRIN 1¹, EVERTON COIMBRA DE ARAÚJO 2², CLAUDIO
LEONES BAZZI 3³, PAULO SERIGO LOGO RODRIGUES 4⁴**

¹ Discente do Programa de Mestrado em Tecnologias Computacionais para o Agronegócio, UTFPR/Medianeira

² Docente Dr. do Programa de Mestrado em Tecnologia Computacionais para o Agronegócio - UTFPR/Medianeira

³ Docente Dr. do Programa de Mestrado em Tecnologia Computacionais para o Agronegócio - UTFPR/Medianeira

⁴ Discente do Programa de Mestrado em Tecnologias Computacionais para o Agronegócio, UTFPR/Medianeira

Apresentado no

XLV Congresso Brasileiro de Engenharia Agrícola - CONBEA 2016
24 a 28 de julho de 2016 - Florianópolis - SC, Brasil

RESUMO: Compreender fenômenos resultantes da distribuição espacial de dados no espaço é um grande desafio para diversas áreas do conhecimento, em consequência disso, pesquisas estão sendo realizadas para descobrir algum padrão que esteja associado à distribuição espacial. Portanto, este estudo tem como objetivo analisar a autocorrelação espacial da produtividade de soja dos anos de 2007/2008 para 48 municípios da região oeste do Paraná, sendo utilizado técnicas de análise espacial de área, como o índice de Moran global e local. Os procedimentos aplicados no trabalho foram executados pelo software R. A escolha referente a cultura da soja se deve ao fato da sua importância no cenário brasileiro. Desde 1976, ela vem se destacando no país, tornando o Brasil o segundo maior produtor mundial de soja. Aplicando o índice de Moran global e local, concluiu-se a existência de dependência espacial para a produtividade de soja, formando-se clusters nessas áreas, aonde municípios que obtiveram alta produtividade estão perto de outros que também apresentaram uma alta produtividade, o mesmo ocorreu para os municípios que tiveram baixa produtividade.

PALAVRAS-CHAVE: Autocorrelação Espacial, Matriz de Vizinhança, Análise Espacial de Área

SPATIAL EXPLORATORY DATA ANALYSIS APPLIED TO SOYBEAN PRODUCTIVITY IN THE WEST REGION OF PARANÁ

ABSTRACT: To understand the phenomena resulting from the spatial distribution of data in space is a major challenge for many researchers, as a consequence of that, studies are being conducted out to find some standard that is associated to the space distribution. So, this study has as I aim to analyse the space autocorrelation of the productivity of soy of the years of 2007/2008 for 48 local cities of the western region of the Paraná, being used techniques of spatial analysis of area, like the Moran's index global and local . The proceedings applied in the work were executed by the software R. The choice referring to culture of the soy is due to the fact of his importance in the Brazilian scenery. Since 1976, it comes standing out in the country, making Brazil the world's second largest producer of soybeans. Applying the Moran's index global and local, it was concluded the existence of spatial dependence for soy productivity in the near cities, being formed clusters in these areas, where municipalities that obtained high productivity are near to others who also presented a high productivity, the same thing took place for the municipalities that had low productivity.

KEYWORDS: Spatial Autocorrelation, Neighborhood Matrix, Area Spatial Statistics

INTRODUÇÃO: A cultura da soja tem uma grande importância na economia brasileira, em 2010, o Brasil foi responsável por cerca de 26,3% da produção mundial de soja, com um valor de 67,5 milhões toneladas, sendo cultivada em uma área referente a 24,3 milhões de hectares. Em relação à economia, a soja propiciou aproximadamente 9% das exportações brasileiras, girando um valor de R\$ 17,5 bilhões. O commodity respondeu por 5,6% de um total de R\$ 821,8 Bilhões, correspondendo a porcentagem de 1,25% de todo PIB nacional. (CONAB, 2010). No cenário paranaense, de acordo com SEAB (2015), a soja é a cultura com maior importância, por exemplo, no ano de 2014 o Valor Bruto da Produção foi de 15,04 bilhões de reais ou 21% dos 70,6 bilhões de reais produzidos nos campos do Paraná. Na safra dos anos de 2014 e 2015 a produção de soja foi de 16,95 milhões de toneladas. E nesses anos, os produtores plantaram uma área recorde no Estado e com a contribuição do clima obtiveram a maior produção de soja da história paranaense. Portanto, a área semeada foi de 5,1 milhões de hectares o que apontou um acréscimo de 4% em relação à safra anterior (2013/14). Em relação à produção o aumento foi de 16% (SEAB, 2015). Devido à esta importância, o objetivo do trabalho é aplicar a estatística espacial de área para verificar a autocorrelação espacial da produtividade de soja (t ha⁻¹) na região oeste do Paraná, para tentar descobrir a existência de algum padrão espacial nos valores que estão sendo estudados, se apresentam algum agrupamento ou é aleatório, ou também se a distribuição está associada a causas mensuráveis.

MATERIAL E MÉTODOS: A área de estudo abrange 48 municípios da região oeste do Estado do Paraná. Sendo utilizada a produtividade de soja dos anos safra 2007/2008 como variável para aplicar no trabalho. As informações referentes à produtividade de soja em (t ha⁻¹) foram fornecidas pela SEAB (2010). Para calcular a autocorrelação espacial foram utilizados os índices de Moran Global e Local utilizando o software R na versão 3.1.1 (R, 2013). De acordo com Araújo 2013 e Anselin 2007, O índice I de Moran é o método mais usado para o cálculo de autocorrelação espacial. A equação 1 é a fórmula que calcula a autocorrelação espacial global das variáveis, verificando se os dados estão autocorrelacionados no espaço.

$$I = \frac{n \sum_{i=1}^n \sum_{j=1}^n z_i z_j w_{ij}}{S_0 \sum_{i=1}^n z_i^2} \quad (1)$$

Em que,

n - é o número de populações (n polígonos);

$z_i = (x_i - \bar{x})$ e $z_j = (x_j - \bar{x})$ para $i \neq j = 1 \dots n$, que se referem aos valores analisados das populações i e j centradas na média \bar{x} da variável de interesse;

w_{ij} - é o elemento da matriz de proximidade W, n x n, pelo qual indica a relação espacial para as n populações;

S₀ - é definida pela Equação (2):

$$S_0 = \sum_{i=1}^n \sum_{j=1}^n w_{ij} \quad (2)$$

Quando se estuda um grande número de áreas como neste trabalho, é necessário aplicar os indicadores locais que permitem um maior detalhamento para cada região, tornando possível, a identificação de diferentes padrões espaciais, como agrupamentos com valores semelhantes (*clusters*) ou divergentes (*outliers*) (DRUCK et al. 2004).

RESULTADOS E DISCUSSÃO: Na Tabela 1, é apresentado o índice de Moran Global para a variável produtividade de soja, e o nível de significância, segundo os critérios de contingência *queen*, distância entre o centroide e vizinho mais próximo para a matriz de proximidade W.

Pode-se verificar que os níveis descritivos (p-valor) são menores que 0,05 (nível de significância), concluindo que todas as variáveis têm autocorrelação espacial significativa a 5% de probabilidade.

Tabela 1 - Índice Global de Moran e o nível descritivo (p-value) com 999 permutações

	Contiguidade (queen)	Distância entre centroides	Vizinhos mais próximos
Índice de Moran	I: 0.3112 P: 0.001	I: 0.2200 P: 0.001	I: 0.3051 P: 0.001

não significativo ($P > 0,05$); *: significativo ($P < 0,05$); I: Índice de Moran Global

O mapa da figura 2 apresentando as conexões das matrizes de vizinhança utilizado no cálculo do índice de Moran global.

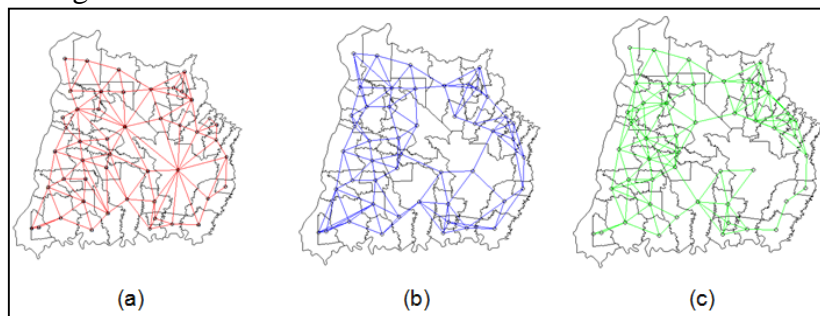


Figura 1 - Contiguidade *queen* (a), Vizinho mais próximo(b), Distância entre centroide(c)

Pode-se perceber a diferença entre as conexões dos polígonos nos três mapas, o primeiro utilizando a regra *queens*, considerou como vizinho apenas as áreas que fazem fronteira com os municípios, desconsiderando qualquer critério de distância. Já a regra do vizinho mais próximo, foi considerada os quatro polígonos mais próximos de cada município. O último critério foi à distância entre os centroides, neste cenário utilizou-se um peso nas distâncias entre as duas áreas mais próximas, caso a distância dos polígonos sejam menores que o peso permitido eles são considerados vizinhos. Conforme DRUCK et al. (2004), para um grande número de áreas, como neste caso com 48 municípios, é importante utilizar indicadores de associação espacial local que permitam identificar melhor os agrupamentos. Para o Índice de Moran Local (LISA) foram classificados os municípios em função dos seus níveis de significância, conforme Figura 3.

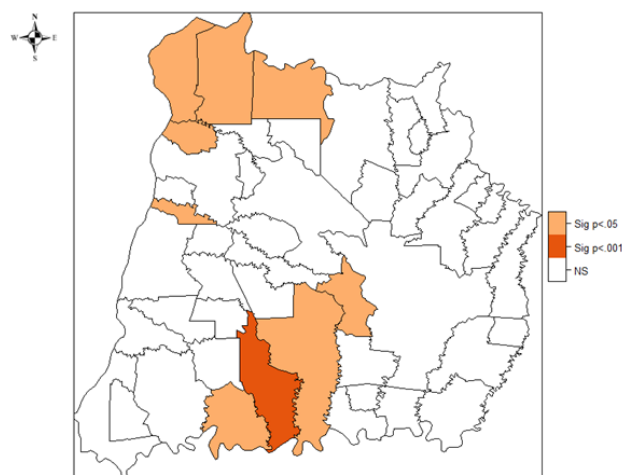


Figura 2 - Indicador local de autocorrelação espacial (LISA) para a variável Produtividade de Soja.

Os municípios que não apresentam valores significativos ao nível de 5% de probabilidade estão representados no mapa com a cor branca. Já os municípios que apresentaram LISA significativo são exibidos com diferentes tons de marrom, tanto a 5% significância como a 1% de significância, e são os que apresentam uma característica própria, que merecem uma análise detalhada. Para calcular o Índice de Moran Local foi utilizado a matriz de vizinhança baseada no vizinho mais próximo.

CONCLUSÕES: A partir deste estudo concluiu-se que houve autocorrelação espacial da produtividade da soja nas safras de 2007/2008 em meio da análise exploratória espacial por áreas, usando técnicas estatísticas como o índice I de Moran Global e Local Univariado. Portanto, os 48 municípios da região oeste do Paraná possuem dependência espacial, constatando áreas com agrupamentos e também de transição. Na análise local, utilizando o índice LISA, áreas com características individuais correlacionadas também foram possíveis de identificação.

REFERÊNCIAS:

ANSELIN L.; SRIDHARAN S.; GHOLSTON S. **Using exploratory spatial data analysis to leverage social indicator databases: the discovery of interesting patterns. Social Indicators Research**, [s. l.], v. 82, n. 2, p.287-309, 2007.

ARAUJO, Everton C. de; URIBE-OPAZO, Miguel A.; JOHANN, Jerry A.. **Análise de agrupamento da variabilidade espacial da produtividade da soja e variáveis agrometeorológicas na região oeste do Paraná**. Eng. Agríc. Jaboticabal, v. 33, n. 4, p. 782-795, Aug. 2013 .

CONAB. **Companhia Nacional de Abastecimento. Soja-Brasil: série histórica de produção:** safras 1976/77 a 2009/10. 2010. Disponível em: <http://www.conab.gov.br/conteudos.php?a=1252&t=2&Pagina_objcmsconteudos=2#A_objcmsconteudos>. Acesso em 20 fev. de 2016.

DRUCK, S.; CARVALHO, M.S.; CÂMARA, G.; MONTEIRO, A.V.M. (ed). **Análise Espacial de Dados Geográficos**. Brasília: EMBRAPA, 2004. 209p.

SEAB - Secretaria de Estado da Agricultura e do Abastecimento. **Soja análise da conjuntura agropecuária**. 2015. Disponível em <<http://www.agricultura.pr.gov.br/modules/conteudo/conteudo.php?conteudo=32>>. Acesso em: 21 fev. de 2016.

SEAB. **Secretaria da Agricultura e do Abastecimento do Paraná**. 2010.

R. (2013). Acesso em 21 de 02 de 2016, disponível em <http://www.r-project.org/>