

ANÁLISE DA VARIABILIDADE ESPACIAL DA PRODUTIVIDADE DE SOJA UTILIZANDO RECURSOS DO SOFTWARE R

FLÁVIA RONCATO FRASSON¹, JOSÉ PAULO MOLIN²

Resumo

O mapa de produtividade pode indicar a localização de áreas críticas em termos de produtividade da cultura, ajudando assim o produtor fazer intervenções localizadas, aumentando sua competitividade no mercado e conseqüentemente sua rentabilidade. Primeiramente realizou-se uma análise exploratória geral dos dados, seguida da análise exploratória espacial e diagnóstico sob a ótica da geoestatística utilizando o software livre R. Através de técnicas de geoestatística foi possível avaliar a variabilidade espacial da produtividade de soja, num talhão no município de Campos Novos Paulista, SP, que apresentou moderada variabilidade espacial. Também foi confeccionado o mapa de produtividade utilizando a interpolação através da krigagem.

Palavras-Chave: Geoestatística, mapa de produtividade, software livre

SPATIAL VARIABILITY ANALYSIS OF SOYBEAN YIELD USING R SOFTWARE RESOURCES

Summary

Yield maps can indicate location of critical areas within a field, thus helping the producer to make local interventions, increasing competitiveness and consequently profitability. A general statistical analysis of the data was conducted, followed by a geostatistical study using the free software R. Through geostatistics techniques it was possible to evaluate soybean yield spatial variability, in a field located in Campos Novos Paulista, SP, that presented moderate spatial variability. The yield map was generated using kriging interpolation.

Key Words: Geostatistics, yield maps, free software

Introdução

O estudo da relação entre produtividade, atributos do solo e do relevo, ou seja, das principais limitações à produção de uma determinada área ou região, torna-se essencial quando se tem como objetivo o manejo racional do cultivo e do solo, evitando a exaustão química e a degradação de seus atributos físicos, visando à máxima produtividade sustentável, assim como sua predição.

O rendimento das culturas é resultado da interação entre fatores do ambiente, a exemplo de radiação e precipitação, da cultura, a exemplo da partição de fotoassimilados e população de plantas, e do solo, a exemplo de fertilidade e

componentes bióticos (LOVENSTEIN et al, 1995). É reconhecido que os aumentos de rendimentos verificados nas principais culturas, a exemplo do aumento de 600 kg.ha^{-1} no rendimento médio de soja durante a década de 1990 no Brasil (EMBRAPA 2003), não são decorrentes de aumentos na produção biológica, mas sim, da melhor adaptação ecológica dos cultivares e do uso intensivo de tecnologia, principalmente a fertilização, filosofias da chamada Revolução Verde (HORTON, 2000).

Uma abordagem do gerenciamento localizado das atividades agroflorestais é a Agricultura de Precisão que, segundo Vettorazzi & Ferraz (2000), baseia-se na coleta e análise de dados geoespaciais, viabilizando intervenções localizadas no cultivo, com a exatidão e a precisão adequadas. O conceito de agricultura de precisão contempla o manejo de regiões distintas da gleba segundo seus atributos, notadamente do solo, o que é possível com a evolução e popularização de técnicas de georeferenciamento, a exemplo do Global Positioning System (GPS), e de coleta de dados georreferenciados, a exemplo dos monitores de rendimento (MOLIN, 2001). Com relação à distribuição espacial dos atributos, a geoestatística possibilita, a partir de dados amostrados, inferir valores a pontos intermediários utilizando as técnicas do semivariograma e da interpolação.

Os mapas de variabilidade espacial da produtividade são gerados a partir de um conjunto de sensores (ex: de fluxo de massa e de posicionamento espacial), permitindo uma cobertura plena da área. Para que se possa estudar esta variabilidade espacial, baseada em amostras, é conveniente utilizar-se técnicas como a geoestatística. A geoestatística permite estabelecer um modelo de semivariograma que melhor descreva a variabilidade espacial dos dados, o qual será utilizado no processo de interpolação digital pelo método da krigagem.

Hoje, os profissionais têm uma grande variedade de ferramentas para auxiliá-los na coleta e análise de dados. Esta segunda etapa é de extrema importância na qualidade dos resultados obtidos; assim a escolha do software que melhor atenda as necessidades é um ponto primordial. A escolha de softwares livres poderá reduzir sensivelmente os gastos com licenças de softwares proprietários, possibilitando a aplicação destes recursos em outras áreas de interesse nacional, como a pesquisa agropecuária.

O software livre R fornece uma variedade larga de técnicas estatísticas (modelar linear e não-linear, testes estatísticos clássicos, análise de séries, classificação, aglomerados) e gráficas, fornecendo ferramentas à pesquisa na metodologia estatística. O R pode ser prolongado (facilmente) através dos pacotes, que cobrem em larga escala os métodos estatísticos modernos. Dentro destes pacotes tem-se o geoR que contempla a área da geoestatística.

Assim, o objetivo do trabalho foi estudar a variação espacial da produtividade da cultura de soja, utilizando o software livre R (2005) com o propósito de identificar a dependência espacial desta na área estudada.

Material e Métodos

Os dados de produtividade analisados foram obtidos na propriedade Agropecuária Dois Irmãos, no município de Campos Novos Paulista, SP, na Fazenda Velha Lagoa, em 2005, em talhão com 22,8 ha, com altitude média de 450 m, clima subtropical úmido e sistema de semeadura direta.

A produtividade da soja foi mensurada com emprego de monitor de colheita marca New Holland®, produzido pela empresa AgLeader®, funcionando com detecção de massa por placa de impacto, instalado em uma colhedora New Holland®, modelo TC59 e equipado com um receptor de GPS Trimble Ag132 com correção diferencial por algoritmo interno.

Das produtividades amostradas, foram eliminados os rendimentos e teores de água nulos, os rendimentos de pontos localizados fora do perímetro da gleba, ou com distância nula e os rendimentos localizados na borda da gleba correspondentes ao tempo de enchimento, conforme descritos em Menegatti & Molin (2004).

Inicialmente, realizou-se uma análise exploratória geral, não considerando a posição geográfica de cada amostra coletada, sendo que nesta etapa foi realizada uma análise estatística descritiva do conjunto de dados, através do software livre R (2005).

Em seguida os dados foram analisados sob a ótica da geoestatística, na qual as posições de cada amostra são consideradas no estudo. Nesta etapa utilizou-se o pacote geoR do programa computacional R. Como este recurso não permite trabalhar com mais de 1000 dados foi feita uma amostragem aleatória de 400 pontos, em seguida realizou-se uma análise exploratória dos dados, sendo que nesta etapa foram determinados a média, mediana, quartis, desvio padrão, coeficiente de variação, assimetria e curtose, através do software estatístico R, com o propósito de verificar a normalidade dos dados, pois, a estimação por krigagem apresenta melhores resultados quando a normalidade dos dados é satisfeita (PAZ-GONZALEZ et al., 2001).

Para identificação da estrutura de dependência espacial entre as amostras, utilizou-se o semivariograma. Basicamente, o processo consistiu em se ajustar o melhor modelo teórico ao semivariograma experimental encontrado através do software R, pacote geoR, utilizando as funções “eyefit” e “variog”. O próximo passo foi estimar os valores em posições não amostradas no campo através da técnica de krigagem gerando, desta forma, o mapa interpolado para a variável produtividade.

Uma avaliação quantitativa da variabilidade espacial também foi realizada; para tanto, utilizou-se o “coeficiente efeito pepita” (E_0), como sendo a relação percentual entre o efeito pepita e o patamar, ou seja, $C_0/(C_1+C_0)$. Quanto maior este coeficiente, menor a variabilidade espacial. O “coeficiente de efeito pepita” de até 25% do patamar foi classificado como forte dependência espacial; entre 25 e 75%, como moderado e, acima de 75%, como fraca dependência espacial (SOUZA et al., 1999, FIETZ et al., 2000).

Resultados e Discussão

Na Tabela 1 apresenta-se a análise estatística descritiva geral dos valores da população e dos valores amostrados aleatoriamente (400) para a variável produtividade da soja.

Os resultados indicam que a média de produtividade da área estudada está um pouco abaixo dos valores da média brasileira (2231 kg.ha⁻¹) para o ano de 2005, segundo o IBGE, sendo que a amostra não apresenta valores discrepantes. O coeficiente de variação, observado, da produtividade amostrada foi de 14,36%. PIERCE et al. (1995) afirmam que a produtividade das culturas apresentam moderada variação espacial, com valores de coeficiente de variação variando de 8 a 29%.

Tabela 1: Resultados estatísticos da população e da amostra da variável produtividade de soja.

Estatísticas	Prod. População (kg.ha ⁻¹)	Prod. Amostra (kg.ha ⁻¹)
Média	2108	2093
Erro padrão	3,89	15,03
Mediana	2132	2114
Desvio padrão	307,27	300,62
Coeficiente de variação (%)	14,6	14,4
Variância da amostra	94414,05	90374,14
Curtose	-0,357	-0,433
Assimetria	-0,264	-0,184
Mínimo	1303	1311
Máximo	2967	2896
Número de observações	6228	400

A análise de “boxcox” confirmou a normalidade dos dados de produtividade de soja, sendo desnecessárias transformações para realização de análises subsequentes (Figura 1).

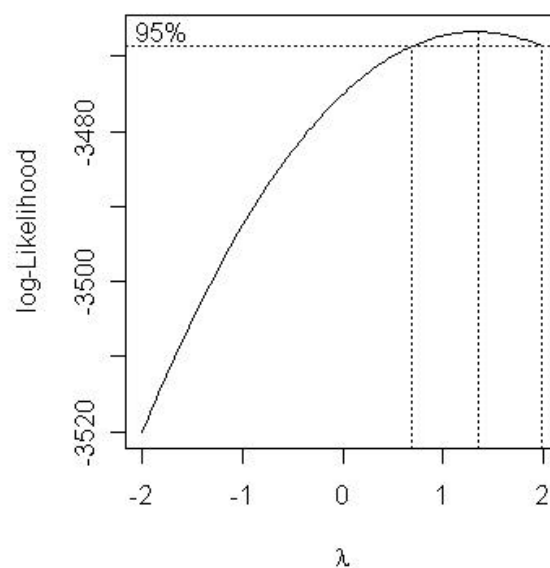


Figura 1: Análise de “Boxcox” para a variável produtividade de soja

O “postplot” gerado associa cores a cada valor indicando a qual quartil ele pertence. Concentrações de cores iguais podem indicar sub-regiões e variações gradativas ao longo de alguma direção indicam existência de tendências. Portanto para atender à hipótese intrínseca não deve ser possível identificar padrões neste gráfico.

A distribuição espacial dos dados, agrupados pelos seus quartis, através da dispersão dos valores da produtividade em relação aos eixos coordenados pode ser visualizado na Figura 2, o qual sinaliza a provável existência de dependência espacial da produtividade de soja, à medida que há zonas de concentração de cores no mapa.

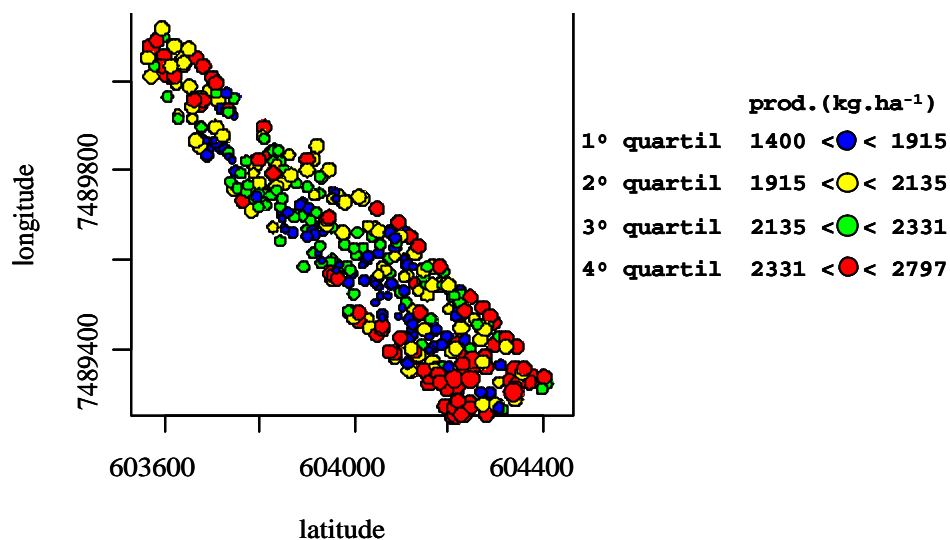


Figura 2 – Distribuição espacial da produtividade de soja

Os valores atípicos de forma geral podem ser identificados pelo gráfico da Figura 3. São definidos limites subtraindo o primeiro quartil e somando ao terceiro uma vez e meia o valor da amplitude interquartílica, sendo os valores situados fora destes limites são destacados. Nota-se a ausência de pontos discrepantes devido à normalidade apresentada pelos dados.

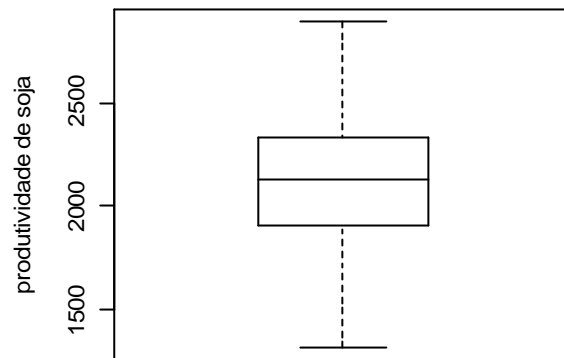


Figura 3. Visualização de dados discrepantes para produtividade de soja

A análise exploratória espacial teve como objetivo analisar a existência de tendências direcionais nos dados, necessária ao emprego de métodos de geoestatística. A verificação de tal tendência foi avaliada pelo gráfico da observação versus às coordenadas (Figura 4). Não houve correlações significativas entre a variável produtividade e suas respectivas coordenadas, sendo os coeficientes de correlação lineares de Pearson, para longitude e latitude de 0,109 e -0,0971 respectivamente, o que não justificou nenhuma transformação dos dados nem a eliminação de tendências.

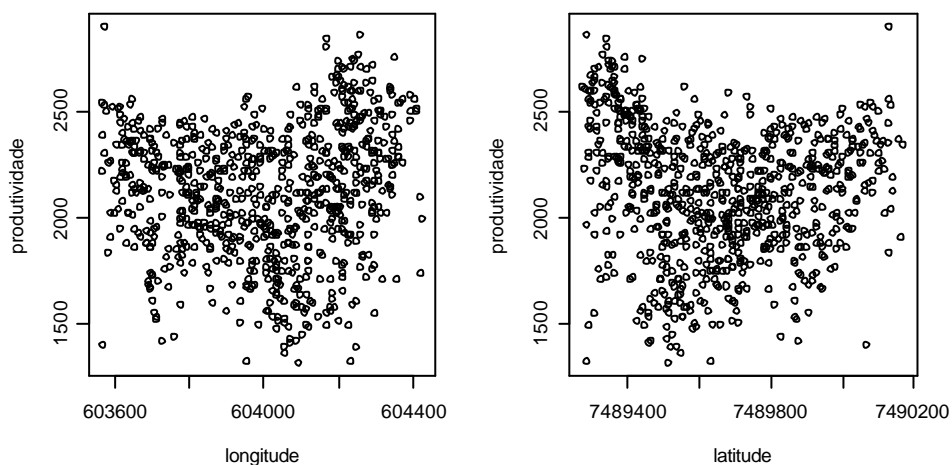


Figura 4 – Diagramas de dispersão de dados de produtividade versus coordenadas

Este gráfico para avaliação de tendências é uma importante ferramenta na avaliação da estacionariedade ou tendência direcional, no qual cada parcela do experimento é classificada segundo os seus quartis; assim, os dados que pertençam ao mesmo intervalo quartílico são classificados e representados como iguais e a concentração de cores iguais no mapa pode indicar a existência de uma sub-região. Desta forma, se o mapa apresentar uma variação gradativa numa região ou direção, poderá estar indicando a existência de uma tendência ou falta de estacionariedade (RIBEIRO JUNIOR, 1995).

As análises efetuadas através dos gráficos direcionais da variável resposta confirmam a inexistência de tendências direcionais, parecendo assim não mostrar problemas que afrontem as hipóteses necessárias ao emprego da geoestatística no estudo da variabilidade espacial da produtividade de soja.

Visando melhores resultados preditivos, em relação ao estimador do semivariograma, utilizou-se da ferramenta “*Box-Cox*”, não sendo necessário a transformação dos dados. A próxima etapa foi estudar os estimadores das estruturas de dependência espacial através de semivariogramas da variável.

Para avaliar o comportamento espacial da produtividade, o mais indicado é a utilização de técnicas de geoestatística, que, através de semivariogramas, permitem definir o alcance e o grau da dependência espacial e fornecem parâmetros para a elaboração de mapas de variabilidade. Para a realização dos ajustes, testaram-se os métodos visuais, Mínimos Quadrados e Máxima Verossimilhança. O modelo que melhor se ajustou aos dados foi o exponencial, cujas estimativas dos parâmetros de cada um dos efeitos, de acordo com o processo usado, são apresentados na Tabela 2.

Tabela 2 – Estimativas dos parâmetros para o modelo exponencial.

Método	Efeito Pepita (C_0)	Patamar	Alcance (a)	Coefficiente E_0
Mínimos Quadrados	38790.00	55712.80	65.64	41,04 %
Máxima Verossimilhança	29247.62	54585.10	36.63	34,88 %

De acordo com os resultados, verifica-se pelo coeficiente E_0 a existência de moderada dependência espacial, sendo mais forte se adotada a estimativa de máxima verossimilhança, a qual foi adotada na estimativa dos parâmetros.

De acordo com a classificação de E_0 dada por Souza et al. (1999) e Fietz et al. (2000), a variável produtividade de soja apresentou moderada dependência espacial dos dados. Essa área vem sendo monitorada nos últimos seis anos, sem intervenções localizadas, fazendo com que a dependência espacial seja governada por efeitos do manejo convencional.

Com base nessas informações, foi possível a geração de um mapa representando uma superfície contínua para a variável em estudo, caracterizando o comportamento desta no campo.

A Figura 5 mostra o comportamento espacial da produtividade da soja para o ano de 2005. Percebe-se que as menores produtividades estão localizadas na parte central do talhão, enquanto as maiores produtividades estão localizadas a sudoeste. Nota-se uma homogeneidade maior na variância, sendo maior nas laterais do talhão.

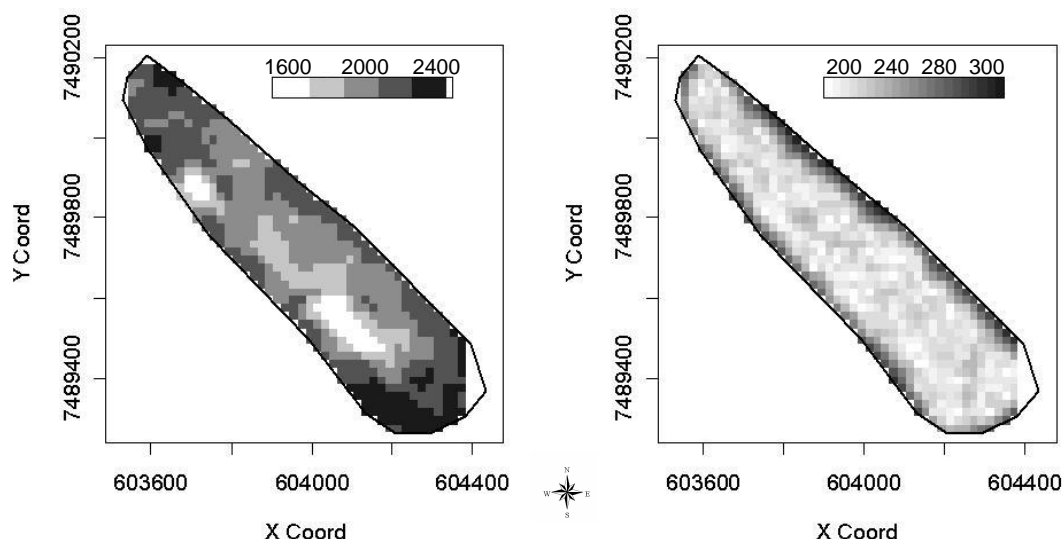


Figura 5. Mapa interpolado por krigagem para as variáveis produtividade ($\text{kg}\cdot\text{ha}^{-1}$) e variância.

Conclusão

Os dados da amostra apresentam comportamento normal e não tendencioso, não precisando de transformação para análise geostatística. O modelo exponencial foi o que melhor se ajustou ao modelo experimental sendo que a variável produtividade de soja apresentou moderada dependência espacial, tanto na análise dos mínimos quadrados quando na máxima verossimilhança e as menores produtividades encontra-se localizadas mais ao centro do talhão analisado. O software livre R mostrou-se uma boa ferramenta na análise estatística dos dados. Entretanto, quando se trata da análise geostatística ele apresenta boas ferramentas, porém com a limitação de apenas 1000 pontos, o que faz com que informações originais sejam desperdiçadas; portanto, pode ser muito bem utilizado em análises com quantidade de dados menores, como amostragem de solo.

Referências Bibliográficas

- FIETZ, C.R.; FABRICIO, A.C.; SALTON, J.C. Mapa de aplicação localizada de calcário de uma área experimental. In: Balastreire, L.A. O estado-da-arte da agricultura de precisão no Brasil, Piracicaba. 2000. p.165-169,
- HORTON, P. Prospects for crops improvement through the genetic manipulation of photosynthesis: morphological and biochemical aspects of light capture. *Journal of Experimental Botany*, v.51, p. 475-485. 2000.
- IBGE. Instituto Brasileiro de Geografia e Estatística. <http://www.ibge.gov.br/home/estatistica/indicadores/agropecuaria/lspa/lspa02200606.sh> tm. Visitado em 28 de março de 2006.
- LOVENSTEIN, H.; LANTINGA, E.A.; RABBINGE, R.; KEULEN, H. Principles of production ecology. Wageningen: Agricultural University of Wageningen, 1995. 85p.
- MENEGATTI, L.A. A.; MOLIN, J. P. Metodologia para identificação e caracterização de erros em mapas de produtividade. *Revista Brasileira de Engenharia Agrícola e Ambiental*, v.7, n.2, p.367-374, 2003.
- MOLIN, J.P. Agricultura de precisão – o gerenciamento da variabilidade. Piracicaba: José Paulo Molin, 2001. 83p.
- PIERCE, F. J.; WARNCKE, D. D. & EVERETT, M. W. Yield and nutrient variability in glacial soils of Michigan. In: P. C. Robert et al. (ed) Site-specific management for agricultural systems. *assa misc. publ., ASA-CSSA-SSSA*, Madison, WI, p. 133-154, 1995.
- R Development Core Team (2005). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.

RIBEIRO JUNIOR, P.J. Métodos geoestatísticos no estudo da variabilidade espacial de parâmetros do solo. Piracicaba: ESALQ 1995, 99p. Dissertação Mestrado

RIBEIRO JR., P.J.; DIGGLE, P.J. GeoR: A package for geostatistical analysis. R-NEWS, v., n. 2. p.1609-3631, 2001.

SNEDECOR, G. W.; COCHRAN, W. G. Statistical methods. 6 ed. Ames: Iowa State University, 1967. 593 p.

SOUZA, E.G.; JOHANN, J.A.; ROCHA, J.V.; RIBEIRO, S.R.A.; SILVA, M.S.; URIBE-OPAZO, M.A.; MOLIN, J.P.; OLIVEIRA, E.F.; NÓBREGA, L.H.P. Variabilidade espacial dos atributos químicos do solo em um Latossolo Roxo distrófico na região de Cascavel, PR. Engenharia Agrícola, Jaboticabal, v.8, n.3, p.80-92, 1999.

VETTORAZZI, C.A.; FERRAZ, S.F.B. Silvicultura de precisão: uma nova perspectiva para o gerenciamento de atividades florestais. In: BORÉM, A.; GIUDICE, M.P.; QUEIRÓZ, D.M. de; et al. (Ed.). Agricultura de Precisão. Viçosa: Os autores, 2000. p.65-75.