

Análise espacial de dados composicionais 56^a RBRAS e 14^o SEAGRO

Ana Patricia Bastos Peixoto, ESALQ/USP e UEPB ¹

Tiago Viana Flor de Santana, IMECC - UNICAMP ²

Helga Restum Hissa, EMBRAPA SOLOS ³

Enio Fraga da Silva, EMBRAPA SOLOS⁴

José Ronaldo de Macedo, EMBRAPA SOLOS ⁵

Cláudio Lucas Capeche, EMBRAPA SOLOS ⁶

Maria Cristina Neves de Oliveira, EMBRAPA SOJA ⁷

1 Resumo

É usual expressar as frações granulométricas de solo como conteúdos de areia, silte e argila e usá-las para descrever e definir classes de solo. Em função disso, este trabalho se propõe a modelar o padrão espacial das frações que compõe o solo e descrever a distribuição espacial destas frações, validando suas características.

Palavras-chave: Dependência espacial, fração do solo, estrutura de correlação.

2 Introdução

Nas Ciências Agrárias, os diversos processos naturais relacionados ao desenvolvimento e à produção vegetal dependem, em grande extensão, de características do solo relacionadas com as capacidades de armazenar e infiltrar água, reter e disponibilizar nutrientes, com a resistência mecânica à penetração e com a capacidade de realizar trocas gasosas no seu espaço poroso devido à sua amplitude e geometria. Além dessas características fundamentais do solo, diversas outras podem exercer influência nesses processos, e todas estão intimamente relacionadas com a proporção de seus elementos constituintes. Essa composição caracteriza a textura do solo, expressa por meio da proporção relativa das frações granulométricas, tradicionalmente caracterizadas como areia, silte e argila (MARTINS, 2010)

Aitchison (1982 e 1986) em análise de dados composicionais apresentam uma metodologia adequada para analisar dados caracterizados por se apresentarem em forma de proporções complementares considerando a independência entre as observações. Diante do exposto, objetivo deste trabalho foi modelar o padrão espacial das frações que compõe o solo e descrever a distribuição espacial da composição do solo.

3 Material e Métodos

Os dados experimentais, que fazem parte deste trabalho, foram coletados do levantamento detalhado de solos da Estação Experimental de Campos, Fazenda Angra (CAPECHE et al., 1997), por pesquisadores da PESAGRO e Embrapa Solos. A área compreendida pelo levantamento situa-se ao Norte do estado do Rio de Janeiro, à margem esquerda do Rio Paraíba do Sul,

¹contato: apbpeixoto@usp.br

no município de Campos dos Goytacazes, entre os paralelos 21°44'47" e 41°18'24" WGr. Para o estudo geoestatístico foi considerada as frações do solo teor de silte (%), areia (%) e argila (%) na camada de 0-20cm em 273 amostras.

Dados composicionais consistem de vetores (composições), cujos componentes são frações e satisfazem a restrição de soma 1. Técnicas estatísticas convencionais podem levar a resultados inconsistentes na análise deste tipo de dados.

O simplex é um espaço amostral natural para dados composicionais, porém provou ser um espaço amostral complicado para ser tratado estatisticamente devido à ausência de classes paramétricas satisfatórias. A distribuição de Dirichlet pode ser usada para modelar dados no simplex, mas às vezes sua estrutura de correlação não é adequada aos dados composicionais (AITCHISON, 1982).

Qualquer vetor no simplex é chamado uma composição; cada elemento deste vetor é chamado de componente, e qualquer coleção de tais vetores, dados composicionais. Para a modelagem e análise deste tipo de dados Aitchison (1986) propôs, dentre outras, a transformação razão log-aditiva (ALR) que transforma o vetor $\underline{\mathbf{Y}}$ no espaço amostral simplex de dimensão B no vetor $ARL(\underline{\mathbf{Y}}) = (\ln(Y_1/Y_B), \dots, \ln(Y_{B-1}/Y_B))'$ no espaço R^{B-1} .

O modelo geoestatístico bivariado composicional proposto é dado por:

$$\begin{cases} Y_1(\underline{\mathbf{x}}_i) = \mu_1(\underline{\mathbf{x}}_i) + S_1(\underline{\mathbf{x}}_i) + Z_1(\underline{\mathbf{x}}_i) \\ Y_2(\underline{\mathbf{x}}_{i'}) = \mu_2(\underline{\mathbf{x}}_{i'}) + S_2(\underline{\mathbf{x}}_{i'}) + Z_2(\underline{\mathbf{x}}_{i'}) \end{cases},$$

em que $Y_1 = \ln(\text{areia}/\text{argila})$, $Y_2 = \ln(\text{silte}/\text{argila})$, $i, i' : 1, \dots, 76$, $S_j(\underline{\mathbf{x}})N(0; \sigma_j^2)$ e $Z_j(\underline{\mathbf{x}})N(0; \tau_j^2)$, $j = 1, 2$.

Os termos S_j modelam os efeitos espaciais, possuem uma função de correlação comum $\rho_U(\underline{\mathbf{x}}; \phi)$ e são escalonados por σ_j definindo o termo comum padronizado U com vetor de médias iguais a zero e matriz de variâncias/covariâncias induzida pela função de correlação exponencial. Por outro lado, os efeitos aleatórios Z_1 e Z_2 representam os efeitos composicionais no modelo. Desta forma o modelo pode ser reescrito como:

$$\begin{cases} Y_1(\underline{\mathbf{x}}_i) = \mu_1(\underline{\mathbf{x}}_i) + \sigma_1 U(\underline{\mathbf{x}}_i; \phi) + Z_1(\underline{\mathbf{x}}_i) \\ Y_2(\underline{\mathbf{x}}_{i'}) = \mu_2(\underline{\mathbf{x}}_{i'}) + \sigma_2 U(\underline{\mathbf{x}}_{i'}; \phi) + Z_2(\underline{\mathbf{x}}_{i'}) \end{cases},$$

e os termos da matriz de covariâncias Σ são:

$$Cov(Y_j(\underline{\mathbf{x}}_i); Y_j(\underline{\mathbf{x}}_{i'})) = \sigma_j^2 + \tau_j^2, \quad Cov(Y_j(\underline{\mathbf{x}}_i); Y_j(\underline{\mathbf{x}}_{i'})) = \sigma_j^2 \rho_U(\underline{\mathbf{x}}_i; \underline{\mathbf{x}}_{i'})$$

e

$$Cov(Y_1(\underline{\mathbf{x}}_i); Y_2(\underline{\mathbf{x}}_{i'})) = \sigma_1 \sigma_2 I_2(i, i') + \tau_1 \tau_2 I_3(i, i'),$$

$$I_2 = \begin{cases} 1 & , \quad sei = i' \\ \rho_U(\underline{\mathbf{x}}_i; \underline{\mathbf{x}}_{i'}) & , \quad sei \neq i' \end{cases} \quad I_3 = \begin{cases} \rho_U(\underline{\mathbf{x}}_i; \underline{\mathbf{x}}_{i'}) & , \quad sei = i' \\ 0 & , \quad sei \neq i' \end{cases}$$

A inferência sobre o vetor de parâmetros $\underline{\theta} = (\mu_1, \mu_2, \sigma_1, \sigma_2, \tau_1, \tau_2, \phi, \rho)'$ é feita usando a teoria da verossimilhança cuja função de verossimilhança é

$$L(\underline{\mathbf{Y}}) = (2\pi)^{-n/2} |\Sigma|^{-1/2} \exp\{-0,5(\underline{\mathbf{Y}} - \underline{\mu}_{\underline{\mathbf{Y}}})' \Sigma^{-1} (\underline{\mathbf{Y}} - \underline{\mu}_{\underline{\mathbf{Y}}})\}$$

Fazendo $\Sigma = \sigma_1^2 \mathbf{R} + \tau_1^2 \mathbf{I}_b = \sigma_1^2 \mathbf{V}$, em que \mathbf{R} é uma matriz de covariâncias relacionada aos efeitos espaciais; \mathbf{I}_b é a matriz bloco diagonal com elementos relacionados às covariâncias entre as composições tem-se a função de log-verossimilhança:

$$l(\underline{\theta}) = -0,5[n \log(2\pi) + 2n \log(\sigma_1) + \log(|\mathbf{V}|) + (1/\sigma_1^2)Qe], \quad (1)$$

em que $Qe = (\underline{Y} - \underline{\mu}_Y)' \Sigma^{-1} (\underline{Y} - \underline{\mu}_Y)$, $\underline{\mu}_Y = \mathbf{X}\underline{\mu}$ e X é a matriz associada às médias das variáveis.

Expressões analíticas fechadas podem ser obtidas para os estimadores de máxima verossimilhança de σ_1 e $\underline{\mu}$

$$\hat{\sigma}_1 = \sqrt{Qe/n}$$

e

$$\hat{\underline{\mu}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{V}^{-1}\underline{Y})$$

Substituindo estes estimadores na expressão (1) e adotando-se a reparametrização $\eta = \sigma_2/\sigma_1$, $\nu_1 = \tau_1/\sigma_1$, e $\nu_2 = \tau_2/\sigma_1$, obtém-se a função de log-verossimilhança concentrada

$$l(\underline{\theta}') = -0,5[\log(|V|) + n(\log(2\pi) + \log(\hat{Q}e) - \log(n) + 1)],$$

em que $\underline{\theta}' = (\eta, \nu_1, \nu_2, \phi, \rho)'$ e que será maximizada numericamente.

Para a obtenção dos erros padrão na escala original dos parâmetros é utilizado o método delta e a variância dos estimadores é dada por

$$Var_{\underline{\theta}}(g(\underline{\theta})) \geq \left(\frac{\partial g(\underline{\theta})}{\partial \underline{\theta}} \right)' \mathbf{I}_E(\underline{\theta})^{-1} \frac{\partial g(\underline{\theta})}{\partial \underline{\theta}},$$

em que $\partial g(\underline{\theta})/\partial \underline{\theta}$ é a matriz diagonal com elementos $\partial l(\underline{\theta})/\partial \theta_{i^*}$, $i^* = 1, \dots, 5$, e $\mathbf{I}_E(\underline{\theta})$ é a matriz informação de Fisher para $\underline{\theta}$ baseada nos dados \underline{Y} que é substituída pela matriz Hessiana obtida no processo de maximização. Todos os resultados foram implementados no ambiente estatístico R (R development Core Team, 2010).

4 Resultados e Discussão

Os valores das distribuições de areia, silte e argila podem ser observados na Figura 1, através dos histogramas e por meio do diagrama ternário é possível verificar que nas composições amostradas, o silte foi o componente que apresentou o menor valor devido o afastamento dos pontos destes vértices. Neste diagrama, pontos próximos a um vértice tem alta proporção do componente representado naquele vértice e baixas proporções dos outros dois componentes. Desta forma, pode-se observar uma maior proporção de argila seguida de areia em relação ao silte. Nota-se, também, que a variabilidade da razão argila/areia é maior que a razão silte/areia pela disposição dos pontos ao longo dos vértices areia-argila e areia-silte.

Utilizando a transformação razão log-aditiva (ALR) que transforma o vetor Y como é observado na Figura 2 no espaço amostral simplex de dimensão B. Percebe-se que a relação silte/argila após transformados apresentam uma distribuição simétrica nos dados.

A análise composicional revela que as amostras de solo são mais arenosas do que argilosas. Os métodos de otimização considerados convergiram e as estimativas pelos quatro métodos são

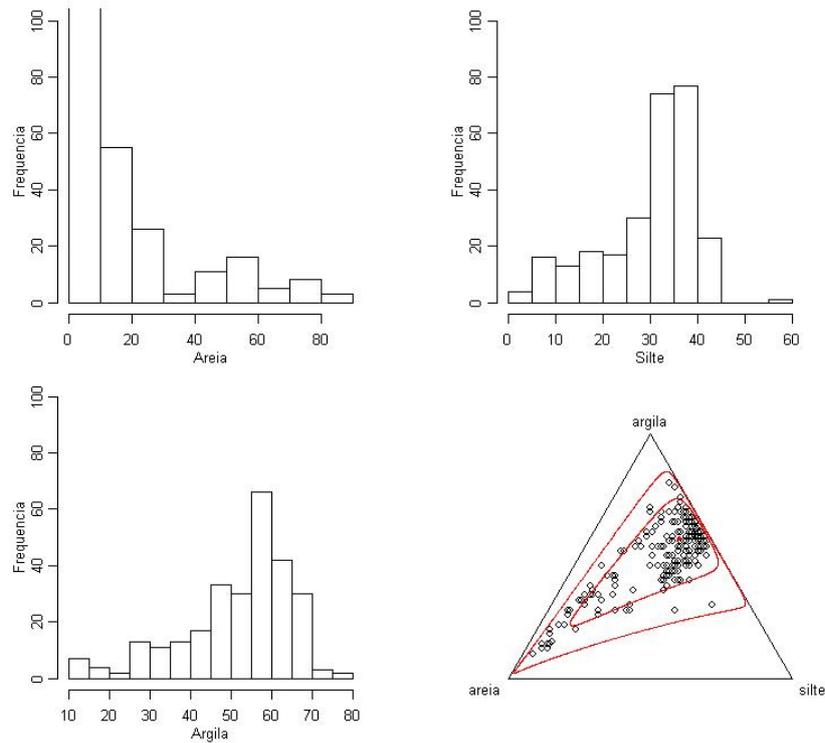


Figura 1: Distribuição de areia, silte e argila e o diagrama ternário das composições.

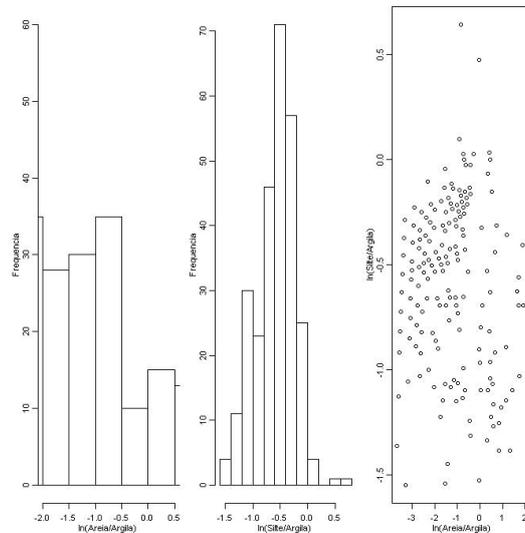


Figura 2: Distribuição das log-razão e correspondente diagrama de dispersão.

coincidentes com precisão de uma casa decimal (Tabela 1). A análise do modelo indica que existe dependência espacial.

A variável argila mostra uma fração acentuada, esse alto percentual para algumas regiões e tipos de solo mostra-se prejudicial na obtenção da produtividade das culturas implantadas no local. O ideal é que haja uma proporção razoável de cada uma das frações afim de aumentar a produtividade. Muitos fatores são responsáveis para a mudança da composição granulométrica do solo o que afeta diretamente no desenvolvimento e produtividade das culturas.

Tabela 1: Estimativas dos parâmetros de acordo com os métodos de otimização e intervalo de confiança

Parâmetros	Estimativas	Erro padrão	LI.Delta	LS.Delta
μ_1	1.151901e+04	1.764928e+05	-1.649738e+05	1.880118e+05
μ_2	-5.676068e-01	.707101e+02	-1.712777e+02	1.701425e+02
σ_1	2.010736e+05	2.171546e+01	2.010519e+05	2.010953e+05
σ_2	292.010736e-01	6.806976e-75	2.010736e-01	2.010736e-01
τ_2	2.010736e-01	6.806976e-75	2.010736e-01	2.010736e-01
ϕ	2.716060e+02	4.305735e+01	2.285486e+02	3.146633e+02
ρ	-8.648320e-01	5.143200e-02	-9.162640e-01	-8.134000e-01

5 Conclusão

Há uma maior proporção de argila seguida de areia em relação ao silte, no solo estudado, do qual podemos dizer que solos com teores de argila superiores a 35%, possuem baixa permeabilidade e alta capacidade de retenção de água.

Referências

- [1] AITCHISON, J. The statistical analysis of compositional data. Royal Statistical Society, Series B, v.44, 1982.
- [2] AITCHISON, J. (Ed.). The statistical analysis of compositional data. New Jersey: The Blackburn Press, 1986.
- [3] CAPECHE, C.L.; MACEDO, J.R.; MANZATTO, H.R.H.; SILVA, E.F. Caracterização pedológica da fazenda Angra - PESAGRO/RIO - Estação experimental de Campos (RJ). (compact disc). In: CONGRESSO BRASILEIRO DE CIÊNCIA DO SOLO. 26., Informação, globalização, uso do solo; Rio de Janeiro, 1997. trabalhos. Rio de Janeiro: Embrapa/SBCS, 1997.
- [4] MARTINS, A.B.T. Análise geoestatística para dados composicionais. Tese (Doutorado), Universidade Federal do Paraná, Curitiba, 2010
- [5] R DEVELOPMENT CORE TEAM. R: A language and environment for statistical computing. Vienna, Austria, 2010. (<http://www.R-project.org>)