

Cokrigagem de Partículas de Solo

R. M. LARK & T. F. A. BISHOP

Ana Beatriz Tozzo Martins - LEG/UFPR; DES/UEM
Paulo Justiniano Ribeiro Junior - LEG/UFPR

14 de dezembro de 2009

Resumo

1. Predição de partículas de solo \Rightarrow **cokrigagem ordinária (CO)** não deve ser usada;
2. alr \Rightarrow variável composicional como uma realização de uma v.a. ilimitada;
3. agl da estimativa ótima das v.a.'s alr transformadas **não produz estimativa ótima**;
4. **Aproximação numérica** que não tem sido usada em ciências do solo;
5. **Estudos de casos:**
 - a CO alr com volta: **agl (viesada)** e **não-viesada**;
 - b CO alr de dados não transformados - ignora restrição soma 1;
5. **Resultados:**
 - I Benefícios no uso de transformação alr / vantagem não consistente no uso de agl;
 - II Vantagem não consistente no uso de transformação alr / viés da transformação agl;

Introdução

- Composição v.a. (3) não é selecionada do \mathbb{R}^3 , mas de um simplex bidimensional plano \mathbb{S}^2 ;
- Simplex \Rightarrow **diagrama ternário triangular** no qual as classes texturais do solo são dispostos;
- **Não existe garantia** que estimativas somem 1 por **krigagem ordinária** (KO) (Odeh et. al. (2003));
- **Krigagem composicional** (KC):
 - a Condições impostas no sistema de KO em acréscimo à não-tendenciosidade;
 - b Resolução numérica - lenta;
- Chang (2002) propos **Cokrigagem Composicional** (CC):
 - a Composição é estimada multivariavelmente ao invés de uma coleção de estimadores de KO;
 - b Procedimento alternativo: **Cokrigagem log-razão** (CO alr) (Pawlowsky-Glahn e Olea(2004));

Introdução

Odeh et. al.(2003) aplicaram KO para dados alr.

- **Resultado:**

predições com menor viés e raiz de erro quadrático médio (EQM)
do que KC e KO das composições não transformadas.

- **Problema:**

- a agl das composições alr é VIESADA;
- b Transformação de volta NÃO-VIESADA desconhecida (Gruitjer (2004) desenvolveu KC)

- **Solução:**

- a Método numérico no caso de krigagem (Pawlowsky-Glahn e Olea(2004) segundo Aitchison(1986))
- b Transformação de volta NÃO-VIESADA desconhecida (Gruitjer (2004) desenvolveu KC)

Introdução (cont.)

- **Proposta:** Demonstrar e avaliar o procedimento numérico para a predição espacial de partículas de solo.
- **Um "achado":**

Estrutura de covariância cruzada da variável alr transformada contém toda a informação da dependência espacial da composição que é fornecida por outras transformações (a menos da log-razão centrada), as matrizes de covariância são **NÃO SINGULARES**.
- Pawlowsky-Glahn e Olea(2004) mostram que CO alr - invariância da permutação:

Estimativas não afetadas se a ordem dos eltos for alterada.
- Pawlowsky-Glahn e Olea(2004) e Odeh et. al. (2003) aplicaram CO alr e KO univariada e agl (viesada), mas Pawlowsky-Glahn e Olea(2004) mostram que **estimativas** podem ser **obtidas numericamente** e não usaram em estudo de caso.

Introdução (cont.)

- **Quadratura Gaussiana** - Método de integração numérica

$$\int_{\mathbb{R}^{B-1}} g(\underline{Z}) f(-\underline{Z}'\underline{Z}) d\underline{Z} \approx \sum_{i_1=1}^k \cdots \sum_{i_{B-1}=1}^k \omega_{i_1} \cdots \omega_{i_{B-1}} g(Z_{i_1}, \dots, Z_{i_{B-1}})$$

$\omega_{i_1} \cdots \omega_{i_{B-1}}$ - pesos derivados da polinomial de Hermite;

$Z_{i_1}, \dots, Z_{i_{B-1}}$ - zeros, abscissas, nós (Abramowitz e Stegun (1972)):
7 são suficientes.

- **Aplicação:** Encontrar uma função $g(\cdot)$ tq a expressão a ser integrada possa ser escrita como o lado esquerdo da eq.

Introdução (cont.)

- Para $\underline{\mu}_{\underline{Z}}$:

$$g_1(\underline{Z}) = \pi^{-\frac{B-1}{2}} \text{agl} \left(\underline{\mu}_{\underline{Y}} + \sqrt{2}R'\underline{Z} \right)$$

$\underline{\mu}_{\underline{Y}}$ -estimativa cokrigada de X

R -matriz de Cholesky da matriz de covariância dos erros cokrigados;

- Para $\underline{\Sigma}_{\underline{Z}}$:

$$g_2(\underline{Z}) = \pi^{-\frac{B-1}{2}} \left(\text{agl}(\underline{\mu}_{\underline{Y}} + \sqrt{2}R'\underline{Z}) - \underline{\mu}_{\underline{X}} \right) \left(\text{agl}(\underline{\mu}_{\underline{Y}} + \sqrt{2}R'\underline{Z}) - \underline{\mu}_{\underline{X}} \right)'$$

Estudo de Caso 1

- Dados de uma transeção regular de espaçamento 10m perto de Sandford, Inglaterra (Webster & Cuanalo(1975));
- 321 localizações, solo amostrado em 3 layers, de 5 a 6 cm de grossura cada, centrado nas profundidades 8, 30 e 65 cm;
- Amostras de areia, silte e argila na profundidade 30 cm;
- Removeram de 3 em 3 dados para validação e predição do conjunto de dados restante foi usado na análise geoestatística;
- **Areia** como **denominador**;
- Calcularam auto e variograma cruzado e ajustaram modelo de correionalização linear usando **simulated annealing** (LMCRs) : **dados alr** e **sem transformação**;

Estudo de Caso 1 (cont.)

- **Procedimentos** de predição em localizações não amostradas:
 - 1 Argila e silte por CO de dados em linha - Areia por complementar;
 - 2 Cokrigagem alr - agl;
 - 3 Cokrigagem alr - GH;
- **Comparação das estimativas:**
 - 1 EQM - Erro quadrático médio;
 - 2 STRESS - Soma de quadrados residual padronizada;
- **Distância de Aitchison:** distância Euclidiana entre a transformação clr pois como usa todos os eltos, a distância de Aitchison é invariante à permutação. **Reflete as restrições em sua variação conjunta.**
- **Resultados:**
 - 1 EQM e STRESS \Rightarrow CO alr $<$ CO ;
 - 2 GH **não** apresentou evidências de **melhoria** nas predições **do que** agl;

Estudo de Caso 2

- Dados de uma porção de um “paddock”, Austrália em 2 ocasiões:
 - 1 04/1996: 110 a.a. **estratificadas** (Shatar e McBratney(1999));
 - 2 09/1999: 109 a.a. **simples** (Bishop e McBratney(2001))frações de partículas de solo em layer na profundidade 15-30cm;

- Tab.1: **Assimetria** aumenta para dados alr - Nota-se **outliers** de areia no **diagrama ternário**;

- **Octile skew**: medida de assimetria insensível a outliers;

$$O_s = \frac{(P_{0,875} - P_{0,5}) - (P_{0,5} - P_{0,125})}{P_{0,875} - P_{0,125}}$$

Se $O_s = 0 \Rightarrow$ o 1o e 7o octile são simétricos sobre a mediana;

Dados com $A_s > 1 \Rightarrow$ transformação;

Dados com $O_s > 0,2 \Rightarrow$ transformação;

$$(y_1 = -0.2522547 ; y_2 = -0.2696939)$$

- $O_s \text{ alr} < O_s \text{ dados}$ não transf. \Rightarrow dados normal/te distribuídos com alguns outliers;

Estudo de Caso 2

- **Resultados:**

a) LMCRs ajustado para dados transformados **difere** para não transformados, **embora as correlações** estruturais dos componentes de variação espacialmente dependentes são todas **fracas**:

- -0,35 p/ dados não transformados;
- 0,35 p/ dados alr com agl;
- 0,30 p/ dados alr com GH;

b) Estatística de validação:

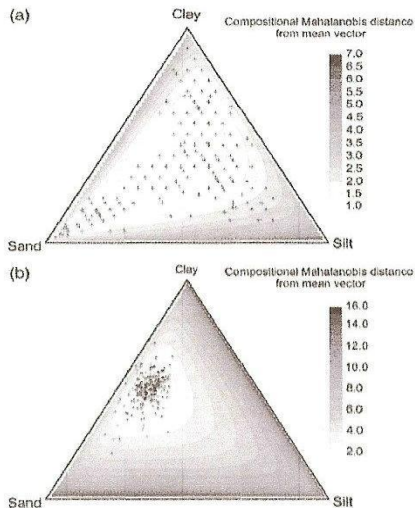
- STREES: $CO = CO\ alr/agl = CO\ alr/GH = CO\ alr/GH\ LCMR$ robusto;
- EQM **menor** p/ CO/GH (robusto e padrão);
- Entretanto, **Não Existe** vantagem consistente destes resultados de volta NÃO-VIESADA **sobre** CO p/ dados não transformados.

- **Caso 1:**
 1. EQM: Vantagem CO alr sobre CO dados não transformados;
 2. Não existe diferença consistente entre agl e GH.
- **Caso 2:** GH **melhor** agl, mas não consistente sobre CO das variáveis não transformadas.
- **Diagrama ternário:** Plotou as partículas superimpondo nos contornos os ptos unidos(join) no simplex onde as distâncias de Mahalanobis composicional do vetor de médias dos dados alr transformados são iguais.

$$\delta_m(Z, \bar{X}) = \{[alr(Z) - \bar{X}]' S^{-1} [alr(Z) - \bar{X}]\}^{1/2}$$

S_X - matriz de covariância dos dados após transformação alr;
 \bar{X} vetor de médias dos dados alr

Discussões (cont.)



Discussões (cont.)

- Se $\text{alr}(Z)$ é v.a. NM \Rightarrow 2 obs que correspondem a mesma densidade de probabilidade desta variável estará a mesma distância de Mahalanobis da média.

Quando estes contornos são projetados no plano real, terão forma elipsoidal refletindo a correlação entre as variáveis no plano.

Entretanto, a projeção no simplex mostra distorções devido às restrições na distribuição dos dados perto dos eixos e vértices.

- Pelos desvios da forma elipsoidal pode-se verificar que no:
 - TRI-1 que os dados não são normais multivariados e por isso a CO de dados não transformados é melhor.
 - TRI-2 os dados são normais multivariados e a CO alr é melhor.

Conclusões práticas

- Vantagem CO alr **sobre** CO dados não transformados: esperado na teoria mas não visto consistentemente.

Fato: CO alr $c/$ GH tem forte base teórica \Rightarrow **confiança**
como método geral.

- **Diagrama ternário:** ferramenta de diagnóstico p/ decidir qdo a análise complexa é necessária;
- Na CO alr, as **estimativas não** são **afetadas** pela escolha do **denominador**;
- Nos estudos de caso **não houve grandes melhorias** no uso de GH p/ aproximar a esperança condicional dos valores alr cokrigados, **MAS existe vantagem** sobre agl.
Vantagem: Obtenção da variância de estimação nas unidades originais com cautela na interpretação.

Pesquisas e Conclusões

Pesquisas:

- incluir dados de sensoriamento remoto, condutividade elétrica do solo no sistema de cokrigagem p/ melhorar as previsões. Poderia usar modelo misto incluindo tendência.

Complicação: Calcular matriz de covariância dos erros de co-predição de termos alr, exigidos para a transformação de volta.

Conclusões:

- A extensão das vantagens da CO alr sobre CO parecem depender de quão longe a distribuição dos dados no espaço real é realmente restrito pelo simplex;
- Existem vantagens se a transformação de volta das estimativas alr cokrigadas é calculada por GH para aproximar a esperança condicional