Geoestatística

Natalia da Silva Martins[1]

Artigo 1

Ajuste e seleção de modelos espaciais de semivariograma visando à estimativa volumétrica de *Eucalyptus grandis*.

Autores

José Marcio de Mello - Departamento de Ciências Florestais da UFLA.
João Luís Ferreira Batista - Departamento de Ciências Florestais da ESALQ.
Paulo justiniano Ribeiro Júnior - Departamento de Estatística da UFPR.
Marcelo Silva de Oliveira - Departamento de Ciências Exatas da UFLA.

Artigo 2

A Composite Likelihood Approach to Semivariogram Estimation.

Autores

Frank C. Curriero - Departament of Bioestatistics, The Johns Hopkins University.
Subhash Lele - Departament of Bioestatistics, The Johns Hopkins University.

Resenha

Nos artigos "Ajuste e seleção de modelos espaciais de semivariograma visando à estimativa volumétrica de *Eucalyptus grandis*"e "A Composite Likelihood Approach to Semivariogram Estimation"o objetivo dos autores foi avaliar métodos de ajuste do semivariograma, que além de descrever a estrutura de dependência espacial é o ponto chave do preditor krigagem.

O objetivo dessa resenha é apresentar, baseado nos artigos citados, uma breve descrição dos métodos de ajuste do semivariograma, sendo eles: "A sentimento", Momentos,

[1]ESALQ-USP, naty-martins@hotmail.com

Quadrados Mínimos Ordinários, Ponderados, Máxima Verossimilhança, Máxima Verossimilhança Restrita e Máxima Verossimilhança Composta.

O método "A sentimento"é um método sem nenhum procedimento matemático, o qual consiste em um ajuste visual do modelo selecionado aos pontos do semivariograma experimental.

O método dos Quadrados Mínimos Ordinários e Quadrados Mínimos Ponderados são baseados no método dos Mínimos Quadrados que visa obter os valores dos parâmetros de um modelo que minimize a soma dos quadrados da diferença entre os valores observados e os valores estimados. A estimativa dos parâmetros do semivariograma pelo método dos Quadrados mínimos é obtida pela minimização da equação 1.

$$Q(\theta) = \sum_{i=1}^{k}[g(h_i) - \gamma(h_i, \theta)]^2 \tag{1}$$

Em que:

θ é o vetor de parâmetros estimados que define o semivariograma, sendo cada estimativa denotada por $g(h_i)$;
k o numero de *lags* do semivariograma experimental;
$\gamma(h_i, \theta)$ é a semivariância calculada pelo modelo.

Dividindo-se a equação 1 pelo número de pares de pontos em cada distância do semivariograma $m(h_i)$ temos o método dos Quadrados Mínimos Ponderados, sendo este uma tentativa de ponderar a semivariância.

O Método dos Momentos (MM) nada mais é do que supor que os momentos da distribuição populacional coincidem com os da amostra, podendo-se assim, ao igualar-se os momentos de menor ordem, obter-se um sistema de equações que forneça as estimativas desejadas, sendo o método mais popular para se estimar os parâmetros do semivariograma. Seja $z(s_1), ..., z(s_n)$ observações espaciais, o estimador do MM para o semivariograma é dado por:

$$\hat{\gamma}(h) = \frac{1}{2|N(h)|} \sum_{N(h)} (z(s_i) - z(s_j))^2 \tag{2}$$

em que:
$N(h)$ são os números de pares de valores medidos.
$(z(s_i) - z(s_j))$ são os pares de valores da variável resposta separada pela distância h.

O Método da Máxima Verossimilhança (ML) consiste em maximizar a função de densidade de probabilidade, ou seja, obter, a partir de uma amostra, o estimador mais verossímel dos parâmetros de um certo modelo probabilístico. Sendo uma técnica de estimação bastante utilizada devido as suas propriedades assintóticas, tais como eficiência e consistência. Na estimação por ML assume-se que os dados seguem uma distribuição multivariada Gaussiana, isto é, considerando um conjunto de dados observados $Z(s_i), ..., Z(s_n)$ temos que $Z \sim MVN(F\beta, \sigma^2\Sigma + \tau^2 I)$, em que:

F é a matriz com os valores das funções $f_k, k = 1, ..., p$, que descrevem deterministicamente a variável Z a partir das coordenadas (x_i, x_n).
β é vetor de parâmetros do modelo linear.
Σ é a matriz de covariâncias baseada no modelo de dependência espacial.
τ^2 é o chamado efeito pepita.
I a matriz identidade.

Considerando $K = \sigma^2\Sigma + \tau^2 I$, temos que a função a ser maximizada tem a seguinte forma:

$$L(\beta, \theta) = -\frac{1}{2} log \mid K \mid + (Z - F\beta)^t K^{-1} (Z - F\beta) \tag{3}$$

A maximização da equação 3 é obtida por meio de processos iterativos, no qual a matriz de covariância é atualizada e invertida a cada iteração.

Método da Máxima Verossimilhança Restrita (REML): é uma variante do processo de Máxima Verossimilhança, em que cada observação é dividida em duas partes independentes, a primeira referente aos efeitos fixos e outra aos efeitos aleatórios, de forma que a função densidade de probabilidade das observações é dada pela soma das funções densidade de probabilidade de cada parte. A maximização da função densidade de probabilidade da parte referente aos efeitos aleatórios, em relação aos componentes de variância, elimina o viés resultante da perda de graus de liberdade na estimação dos efeitos fixos do modelo.

Método da Máxima Verossimilhança Composta (CL): a idéia do método de Verossimilhança Composta (conhecida também como pseudo-verossimilhança ou verossimilhança parcial) é ajustar as log-verossimilhanças para cada evento condicionais ou marginais, que conseguirmos escrever. O log da verossimilhança composta é portanto formada pela adição das log- verossimilhanças desses componentes individuais. Portanto, o log da verossimilhança composta refere-se simplesmente conjunto das contribuições das log-verossimilhanças em uma circunstância de arranjo aditivo onde os componentes não necessáriamente representam réplicas independentes.

Existem duas motivações para a "construção"da verossimilhança composta. Em primeiro lugar, elas possibilitam substituir métodos de estimação quando a máxima verossimilhança é difícil de ser calculada. Segundo, elas, as vezes, representam aquela porção do modelo que estamos mais "confortáveis"em modelar, e o estimador resultante pode ser consistente mesmo quando o estimador de máxima verossimilhança não o é, sendo uma forma de consistência robusta.