

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

Adjustment and selection of spatial models of semivariogram
envisaging *Eucalyptus grandis* volumetric estimates

José Marcio de Mello
João Luís Ferreira Batista
Paulo Justiniano Ribeiro Júnior
Marcelo Silva de Oliveira

RESUMO: O semivariograma é a parte central dos estudos geoestatísticos, sendo capaz de descrever tanto qualitativa quanto quantitativamente a variação espacial, além de ser o ponto chave na determinação do preditor geoestatístico-krigagem. Daí a importância do ajuste e seleção dos modelos da função de semivariância. O presente trabalho teve como objetivo avaliar quatro métodos de ajuste aplicados a dois modelos de função de semivariância, a partir das técnicas de validação cruzada e preditiva e do critério de Akaike (Akaike's Information Criterion - AIC). Os dados utilizados para ajuste dos modelos foram coletados em 130 parcelas de 525 m², mensuradas num povoamento de *Eucalyptus grandis*. Foram ajustados dois modelos da função de semivariância, por meio de quatro métodos de ajuste. A seleção do modelo e método de ajuste foi efetuada quanto ao critério de AIC e por meio das técnicas de validações cruzada e preditiva. Segundo as técnicas de seleção avaliadas, o modelo exponencial e o método da Máxima Verossimilhança tiveram melhor desempenho. A validação cruzada e a preditiva indicaram o mesmo modelo e o mesmo método de ajuste com melhor desempenho. Este resultado é interessante, uma vez que nem sempre se tem uma nova base de dados para efetuar a validação preditiva. Sugere-se portanto, o uso conjugado do critério de AIC e da validação cruzada para a seleção de modelos da função de semivariância.

PALAVRAS-CHAVE: Modelos espaciais, Semivariograma, Validação, Critério de AIC

ABSTRACT: The semivariogram is the central part of the geoestatistic studies, describing qualitative and quantitatively the spatial variation, and can be considered the key point in the determination geostatistic predictor-kriging, indicating the importance of the adjustment and selection of the models of the semi variance function. This research evaluated four adjustment methods applied to two semi variance function models, departing from the cross validation and predictive techniques and from Akaike information criterion (AIC). Data for adjustment of the models were collected in 130 plots of 525 m² each, measured in a *Eucalyptus grandis* stand. Two models of the semi variance function were adjusted, using four adjustment methods. Model selection and adjustment method were done using the AIC, through the crossed and predictive validation techniques. According to the appraised selection techniques, the exponential model and Maximum Likelihood method presented better performance. The cross and predictive validation techniques indicated the same model and the same adjustment method as presenting the best performance. This result is interesting, because sometimes a new data base is not available for the predictive validation. It is suggested therefore, that conjugating AIC and cross validation criterion for selecting semi variance function models is recommended.

KEYWORDS: Spatial models, Semivariogram, Validation, AIC criterion

INTRODUÇÃO

Estimativa de volume em florestas é um problema relevante na Biometria Florestal. Estas estimativas são de fundamental importância, pois são o suporte para tomadas de decisões no manejo e planejamento florestal. Assim, técnicas utilizadas na estimativa volumétrica devem ser alvo de estudo e aprimoramento dos pesquisadores da área de ciências florestais.

O uso de métodos geoestatísticos, em estudos envolvendo características dendrométricas, aumentou significativamente nas duas últimas décadas (SAMRA *et al.*, 1989; COHEN *et al.*, 1990; AUBRY e DEBOUZIE, 2000; BIONDI *et al.* 1994). Como não deveria ser diferente, estes estudos tiveram como ponto de partida a avaliação da dependência espacial por meio do semivariograma. O semivariograma é a parte central da Geoestatística, segundo McBratney e Webster (1986), capaz de descrever a estrutura de dependência espacial, além de ser o ponto chave na determinação do preditor geoestatístico (krigagem). Estes fatos demonstram a importância de se tratar adequadamente a questão da seleção e ajuste de modelos espaciais para semivariogramas experimentais.

Com relação ao semivariograma, o pesquisador é capaz de definir o modelo que melhor descreve o comportamento dos dados no espaço (JOURNAL e HUIJBREGTS, 1978). Em seguida, sua preocupação se volta para o ajuste da função matemática ao semivariograma experimental ou aos dados. Segundo McBratney e Webster (1986), em vários trabalhos na área de ciência do solo, a função matemática e o procedimento de ajuste dos parâmetros, se mencionados, foram simplesmente declarados de forma subjetiva, em que, praticamente 100% dos casos, o semivariograma foi estimado de forma visual.

Até a década de 80, o ajuste do modelo espacial ao semivariograma experimental, era usualmente feito de forma visual (neste trabalho foi denominado de ajuste “a sentimento”), sem nenhum procedimento matemático. Posteriormente, com o aumento da capacidade dos recursos computacionais, outros métodos de ajuste, sem subjetividade, foram e estão sendo estudados. Dentre estes métodos, destacam-se os Métodos dos Quadrados Mínimos Ordinários, Ponderados e o Método da Máxima Verossimilhança. Pode-se dizer que estes métodos tiveram o intuito de

retirar e/ou atenuar o caráter de subjetividade na estimação dos parâmetros do semivariograma. As particularidades e as pressuposições demandadas por cada método de ajuste, além daquelas descritas no presente trabalho, podem ser encontradas em Cressie (1985 e 1993) e Diggle e Ribeiro Jr. (2000).

Após a escolha do modelo e a estimativa dos parâmetros, a etapa seguinte é estudar o melhor modelo de semivariograma ajustado. Conforme McBratney e Webster (1986), a seleção se dá através de técnicas quantitativas como, por exemplo, o critério de Informação de Akaike, que em inglês é designado pela sigla AIC – Akaike’s Information Criterion (AKAIKE, 1983) e através das técnicas de validações cruzada e preditiva. Segundo Nielsen e Wendroth (2003), a validação é um meio efetivo para avaliar a qualidade das estimativas de Krigagem. Ela também é útil para selecionar modelos de semivariograma (MELLO *et al.*, 2003).

Dentro do exposto, o presente trabalho teve como objetivo ajustar e selecionar modelos espaciais de semivariograma para estimativa volumétrica de *Eucalyptus grandis*. Especificamente pretendeu-se, a partir da validação e do critério de AIC, avaliar o desempenho de quatro métodos de ajuste, aplicados a dois modelos de função de semivariância.

MATERIAL E MÉTODOS

Descrição da área e obtenção dos dados

Os dados para realização do trabalho foram coletados num povoamento de 986,78 hectares de *Eucalyptus grandis*, pertencente à Votorantim Celulose e Papel (VCP). O povoamento localizou-se no município de Itapetininga na região Sul do estado de São Paulo, com coordenadas geográficas 23°56’S de latitude e 42°21’W de longitude, a uma altitude de 655 metros. A precipitação e a temperatura média anual, respectivamente foram de 1600 mm e 20°C. O solo predominante na área foi LATOSSOLO VERMELHO Distrófico.

Utilizaram-se dados coletados em parcelas de 525 m² (25 X 21m), sistematizadas em toda a área, distantes entre si 500 metros. Foram selecionados três pontos na área, para realização de uma amostragem mais intensa (Figura 1). Este desenho amostral teve como objetivo obter melhores estimativas de parâmetros dos modelos

avaliados. Ao todo foram estimados os volumes por hectare de 130 parcelas. No contexto de análise exploratória, avaliou-se a presença de possíveis dados discrepantes através do gráfico de Box Plot e a normalidade dos dados originais, bem como a normalidade das diferenças em cada distância do semivariograma experimental. O teste utilizado para avaliar a normalidade, foi o Jarque-Bera conforme descrito por (JARQUE e BERA, 1980). A verificação da normalidade das diferenças justificou-se pelo fato de que a Geostatística trabalha com diferenças entre pares de pontos. Esta avaliação serviu como indicativo para admitir um Modelo Multivariado Normal para o processo estocástico subjacente ao fenômeno.

Semivariograma

Considera-se uma característica dendrométrica Z , tal como o volume, que pode variar continuamente num espaço geográfico, em função das coordenadas geográficas. Cada valor observado $z(x_i)$ nos locais $x_i, i=1,2,\dots,n$, em que x_i denota coordenada geográfica em duas dimensões, é considerado uma realização da variável aleatória $Z(x)$. O conjunto das variáveis aleatórias $Z(x_1) \dots Z(x_n)$ constitui uma função aleatória ou processo estocástico. A representação da estrutura de variação destas variáveis aleatórias no espaço, quando x varia, é dada pelo semivariograma.

Consideram-se dois pontos separados por uma distância h . A variação entre eles é caracterizada por uma função variograma (McBRATNEY e WEBSTER, 1986). Esta função é definida como

$$2\gamma(x,h) = E[Z(x) - Z(x+h)]^2 \quad (1)$$

Nota-se que a função variograma como representada na equação (1), depende da localização (x) e do deslocamento (h). Para que o variograma seja função apenas de h , é necessário adotar a Hipótese Intrínseca, ou seja, a variância das diferenças entre dois pontos amostrais só depende desta diferença, sendo a mesma para toda a área avaliada (estacionaridade). É importante ressaltar que, na presença de estacionaridade de segunda ordem, há uma relação direta entre a correlação e a semivariância. Esta relação é dada pela seguinte expressão:

$$\gamma(h) = c(0) - \rho(h) \quad (2)$$

em que $c(0)$ é a variância dos dados "sill total", igual à $\tau^2 + \sigma^2$, τ^2 efeito pepita, σ^2 contribuição e $\rho(h)$ é a correlação entre dois pontos separados por uma distância h .

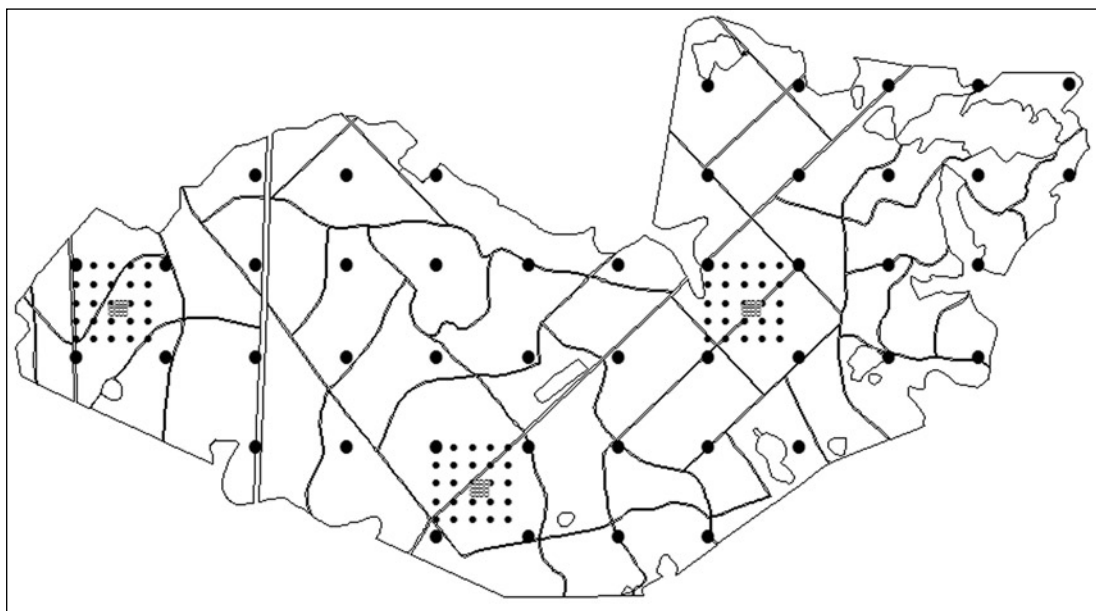


Figura 1. Croqui da área experimental com as parcelas lançadas sistematicamente. (Experimental design area with plots launched systematically distributed).

Na Teoria das Variáveis Regionalizadas, outro nome dado à Geoestatística (JOURNEL e HUIJBREGTS, 1978), a quantidade $\gamma(h)$ é conhecida como semivariância e é, como definido acima, a metade da esperança da variância entre pares de pontos separados por uma distância (h). A função $\gamma()$ é o semivariograma. Obedecidas às condições da hipótese intrínseca, a semivariância pode ser estimada simplesmente através de uma única realização do processo estocástico por:

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z(x_i) - z(x_i+h)]^2 \quad (3)$$

onde $\hat{\gamma}(h)$ é a semivariância estimada entre pares de pontos, $N(h)$ é o número de pares de valores medidos $z(x_i)$, $z(x_i+h)$, separados pela distância h . Esta função permite gerar o semivariograma experimental. A partir deste semivariograma é possível ajustar uma função matemática que expresse a estrutura de dependência espacial da característica avaliada.

No presente estudo ajustou-se o modelo de Matérn com kappa 0.3 e 0.5 cujo modelo com kappa 0.5 refere-se ao modelo exponencial e sua função de correlação é dada pela equação 4. Para os demais valores de kappa, a função de correlação é dada pela equação 5.

$$\rho(h) = \exp(-h/\theta) \quad (4)$$

$$\rho(h) = \{2^{k-1} \Gamma(k)^{-1}\}^{-1} (-h/\theta)^k K_k(-h/\theta) \quad (5)$$

onde $\rho(h)$ é a correlação entre pares de pontos separados pela distância h , θ é o alcance do modelo, K é o parâmetro ordem que determina a suavidade do sinal, $\Gamma(k)$ é a função Gama e K_k é a função Bessel. Doravante, o modelo de Matérn com kappa 0.5 será denominado de modelo exponencial e kappa 0.3, simplesmente de Matérn.

Métodos de Ajuste

Os métodos de ajuste podem ser divididos em dois grupos:

- Ajuste dos modelos ao semivariograma experimental – os métodos de ajuste deste grupo são: Método dos Quadrados Mínimos Ordinários (Ordinary Least Squares - OLS), Método dos Quadrados Mínimos Ponderados (Weight Least Squares - WLS) e método de ajuste denominado de “a sentimento”.

- Método de ajuste de um modelo direto aos dados - Método da Máxima Verossimilhança (Maximum Likelihood - ML)

A sentimento

Escolher o modelo mais adequado não é um procedimento automático. Em Geoestatística é comum o ajuste visual do modelo selecionado aos pontos do semivariograma experimental, o que carece de sustentação estatística. Este tipo de ajuste é bastante subjetivo e depende muito da experiência do pesquisador. Foi amplamente utilizado nos trabalhos cujo objetivo era apenas descrever a estrutura de dependência espacial. Neste trabalho, foi solicitado a outro pesquisador da área de Geoestatística que ajustasse “a sentimento” os parâmetros do semivariograma experimental. Estes parâmetros foram então considerados no método de ajuste “a sentimento”.

Métodos dos Quadrados Mínimos Ordinários e Quadrados Mínimos Ponderados

Segundo Bussab e Moretin (2002), este é um dos estimadores mais utilizados na Estatística. Ele se baseia no princípio dos quadrados mínimos, introduzido por Gauss em 1809. O método consiste em obter os valores dos parâmetros de um modelo que minimizam a soma do quadrado da diferença entre os valores observados e os estimados. A estimativa dos parâmetros do semivariograma pelo Método dos Quadrados Mínimos Ordinário é obtida pela minimização da seguinte expressão:

$$Q(\theta) = \sum_{i=1}^k [g(h_i) - \gamma(h_i; \theta)]^2 \quad (6)$$

em que θ representa o vetor de parâmetros estimados que definem completamente o semivariograma, sendo cada estimativa denotada por $g(h_i)$ e k refere-se ao número de “lags” do semivariograma experimental. A quantidade $\gamma(h_i; \theta)$ é a semivariância calculada pelo modelo e depende dos parâmetros estimados θ e após minimização, a função $g(h_i)$ se torna a estimativa $\hat{\gamma}(h_i)$ de quadrados mínimos. David (1977) sugere dividir o erro quadrático médio pelo número de pares de pontos em cada distância no semivariograma $m(h_i)$, conforme a expressão a seguir:

$$Q(\theta) = \frac{\sum_{i=1}^k [g(h_i) - \gamma(h_i; \theta)]^2}{m(h)} \quad (7)$$

Esta é uma tentativa de ponderar a semivariância, uma vez que a ponderação pela variância das estimativas das semivariâncias não é conhecida. Este método é conhecido como o Método dos Quadrados Mínimos Ponderados. O ajuste de modelos espaciais por estes métodos ocorre a partir dos pontos do semivariograma experimental.

Método da Máxima Verossimilhança

O método da Máxima Verossimilhança é uma técnica de estimação usada com frequência porque suas propriedades assintóticas são interessantes, como por exemplo, eficiência e consistência (WONNACOTT e WONNACOTT, 1972). A idéia do método é obter, a partir de uma amostra, o estimador “mais verossímil” dos parâmetros de um certo modelo probabilístico. Considera-se um conjunto de observações de volume (Z_1, \dots, Z_n) gerados a partir de um modelo linear gaussiano, volume este que pode ser explicado pela soma de três componentes:

$$Z_i = \mu(x_i) + S(x_i) + \varepsilon_i \quad i = 1, \dots, n \quad (8)$$

Na equação 8, $\mu(x_i)$ é uma componente completamente determinística e é determinado por um modelo linear

$$\mu(x_i) = \sum_{k=1}^p f_k(x_i) \beta_k \quad (9)$$

onde, (f_1, \dots, f_p) é um conjunto de p funções que descrevem deterministicamente a variável Z a partir das coordenadas espaciais (x_1, \dots, x_n) . Nesta mesma equação, $S(x_i)$ é uma componente aleatória entendida como um processo estocástico gaussiano autocorrelacionado. A esperança de $S(x_i)$ é igual a zero e a variância constante (σ^2). Esta componente é que confere continuidade espacial aleatória à variável Z , isto é, $S(x)$ é que possui a correlação espacial. A correlação espacial entre dois pontos segue uma função autorizada, para a qual no presente trabalho, foram avaliados os modelos exponencial e Matérn. A componente ε_i é a parte totalmente aleatória do modelo, sendo $\varepsilon_1, \dots, \varepsilon_n$ independentes, identicamente distribuídos, com

média zero e variância τ^2 , denominada de nugget ou efeito pepita (DIGGLE e RIBEIRO JR., 2000). É conveniente escrever o modelo em forma de matriz, a fim de derivar a função de verossimilhança, para obter sua maximização. Considera-se que $\mu(x) = F\beta$, em que F é a matriz com os valores das funções f_k e β o vetor de parâmetros do modelo linear. Na definição da matriz de covariância dos dados observados V , utilizou-se o modelo de dependência espacial (equações 4 e 5) com vetor de parâmetros $\theta = (\tau^2, \sigma^2, \rho)$, descrita matricialmente como $VAR(Z) = K = \tau^2 I + \sigma^2 \Sigma$, onde I é a matriz de identidade e Σ é a matriz de covariâncias baseada no modelo de dependência espacial.

Finalmente, dota-se o modelo definido na equação 8 para o volume de madeira com uma distribuição Multivariada Normal, que pode se representada da seguinte forma: $Z \sim MVN(F\beta, \sigma^2 \Sigma + \tau^2 I)$. Logo, a função de verossimilhança a ser maximizada tem a seguinte forma:

$$L(\beta, \theta) = -1/2 \{ \log |K| + (Z - F\beta)' [K]^{-1} (Z - F\beta) \} \quad (10)$$

A maximização da equação 10 é obtida por procedimentos numéricos iterativos, na qual a matriz de covariância é atualizada e invertida a cada iteração. Estes procedimentos requerem bastante tempo, o qual é proporcional ao cubo do número de observações (RIBEIRO JR., 1995).

Seleção de Modelos

A avaliação do desempenho de cada modelo na estimativa volumétrica se deu através de dois critérios. No primeiro, avaliou-se o AIC gerado para cada modelo dentro dos respectivos métodos de ajuste. O AIC foi obtido através da função de Máxima Verossimilhança, a partir dos parâmetros ajustados para os modelos conforme os métodos. O AIC para os métodos “a sentimento” e quadrados mínimos, foram obtidos a partir dos respectivos parâmetros obtidos por estes métodos, através da função likGRF do geoR. O critério de AIC foi desenvolvido a partir da distância ou informação de Kulback e Leibler (1951). Esta distância é uma medida de discrepância entre as linhas do modelo verdadeiro e o modelo aproximado. Akaike (1983) relacionou a distância de Kulback e Leibler com a Máxima Verossimilhança, surgindo o AIC. O AIC é dado pela seguinte expressão:

$$AIC = -2 \log L + 2K \quad (11)$$

onde L é a Verossimilhança Maximizada do modelo candidato e K é o número de parâmetros deste modelo. O critério de seleção pelo AIC foi baseado no trabalho de Burnhan e Anderson (2002). Neste trabalho os autores propuseram uma alternativa para a comparação de modelos, baseada na diferença de AIC (Δ_i). A diferença de AIC foi obtida pela seguinte expressão:

$$\Delta_i = AIC_i - AIC_{min} \quad (12)$$

onde, AIC_i é o valor do AIC para o modelo (i) e AIC_{min} é o valor do AIC para o modelo suporte, ou seja, o modelo que apresentou menor AIC para um dado conjunto de dados. Para valores de Δ_i maior do que dois, o modelo (i) difere consideravelmente do modelo suporte.

O segundo critério de comparação do desempenho utilizou as técnicas de validações cruzada e preditiva. Para a realização da validação preditiva foram utilizadas 124 parcelas de igual tamanho, georeferenciadas, localizadas na mesma área, que não fizeram parte do ajuste do modelo espacial. Esta técnica consistiu em prever o valor conhecido Z de volume de madeira nas novas parcelas, comparando com o valor observado. Neste trabalho o estimador usado foi o de krigagem, cuja expressão do estimador é:

$$\hat{z}(x_0) = \sum_{i=1}^n \lambda_i z(x_i) \quad (13)$$

em que $\hat{z}(x_0)$ é o valor estimado para o ponto x_0 ; λ_i são os pesos de krigagem definidos conforme os parâmetros do semivariograma e $z(x_i)$ são os valores observados nos pontos amostrados (vizinhança de krigagem).

A comparação entre os modelos, em cada método de ajuste, foi efetuada através do erro médio reduzido (\overline{ER}), desvio padrão dos erros reduzidos (S_{ER}) e do erro absoluto (EA). Segundo McBratney e Webster (1986) e Cressie (1993), o erro reduzido é definido como:

$$\overline{ER} = \frac{1}{n} \sum_{i=1}^n \frac{z(x_{i0}) - \hat{z}(x_{i0})}{\sigma(x_{i0})} \quad (14)$$

onde $z(x_{i0})$ é o valor observado no ponto $i0$; $\hat{z}(x_{i0})$ é o valor estimado para o ponto $i0$ e $\sigma(x_{i0})$ é o desvio padrão da krigagem no ponto $i0$. O desvio padrão dos erros reduzidos foi obtido a partir da seguinte expressão:

$$S_{ER} = \sqrt{\frac{1}{n} \sum_{i=1}^n \frac{z(x_{i0}) - \hat{z}(x_{i0})}{\sigma(x_{i0})}} \quad (15)$$

O erro reduzido médio mais próximo de zero e o desvio padrão próximo de um são os critérios para escolha do melhor modelo. O erro absoluto é uma medida da magnitude dos erros na unidade da variável regionalizada. Tendo o conjunto de valores medidos e estimados $z(x_{i0})$ e $\hat{z}(x_{i0})$, então pôde-se definir o erro absoluto na unidade da variável regionalizada estudada, como:

$$EA = \hat{z}(x_{i0}) - z(x_{i0}) \quad (16)$$

Todas as análises efetuadas no presente trabalho foram realizadas pelo programa R Development Core Team (2003), utilizando-se pacote geoR (RIBEIRO JR. e DIGGLE, 2001).

RESULTADOS E DISCUSSÃO

Análise Exploratória de Dados

Na Figura 2 pode-se verificar os quatros principais gráficos a serem observados numa análise exploratória de dados. Nas letras (a) e (b) estão apresentados os gráficos que mostram o comportamento da variável regionalizada (volume por hectare) em relação aos sentidos NS e WE. Verificou-se a inexistência de tendência nos valores de volume por hectare nestas duas direções. Esta é uma característica importante em estudos geoestatísticos, mostrando a plausibilidade de se aplicar a pressuposição da Hipótese Intrínseca para a variável estudada.

O gráfico "box plot" (c) permitiu verificar a presença de dados candidatos a "outliers" ou discrepantes. Dados discrepantes afetam substancialmente o comportamento do semivariograma, sobretudo na parte inicial do mesmo (DIGGLE e RIBEIRO JR., 2000). Na base de dados avaliada verificou-se um valor candidato a "outliers". Este número foi removido da base de dados para, posteriormente, efetuar as análises geoestatísticas. Na letra (d) da Figura 2, observou-se que a variável regionalizada em estudo, desconsiderando a relevância de uma possível estrutura de dependência espacial, tem distribuição aproximadamente normal. O teste de Jarque-Bera foi não significativo ($P\text{-value} = 0,3512$), ou seja, aceitou-se a hipótese de que os dados têm distribuição aproximadamente normal.

Buscando refinar melhor a pesquisa da normalidade para o processo de inferência no campo da Estatística Espacial, avaliou-se a normalidade das diferenças em cada distância do semivariograma experimental, uma vez que a Geoestatística trabalha com diferenças entre pares de pontos. Foi verificado pelo teste de Jarque-Bera a 5% que, das 14 distâncias do semivariograma experimental, apenas duas apresentaram-se significativas, ou seja, para estas duas classes de distâncias, as diferenças não apresentaram dis-

tribuição Normal. Nas outras doze classes de distâncias, as diferenças entre pares de pontos apresentaram distribuição Normal como pode ser visto na Tabela 1. Estes resultados indicam a plausibilidade de assumir um modelo multivariado Normal para o processo estocástico do fenômeno em estudo, já que, se os dados obedecem a uma Normal multivariada e, se a Hipótese Intrínseca é válida, espera-se que as diferenças apresentem uma distribuição de frequência também Normal.

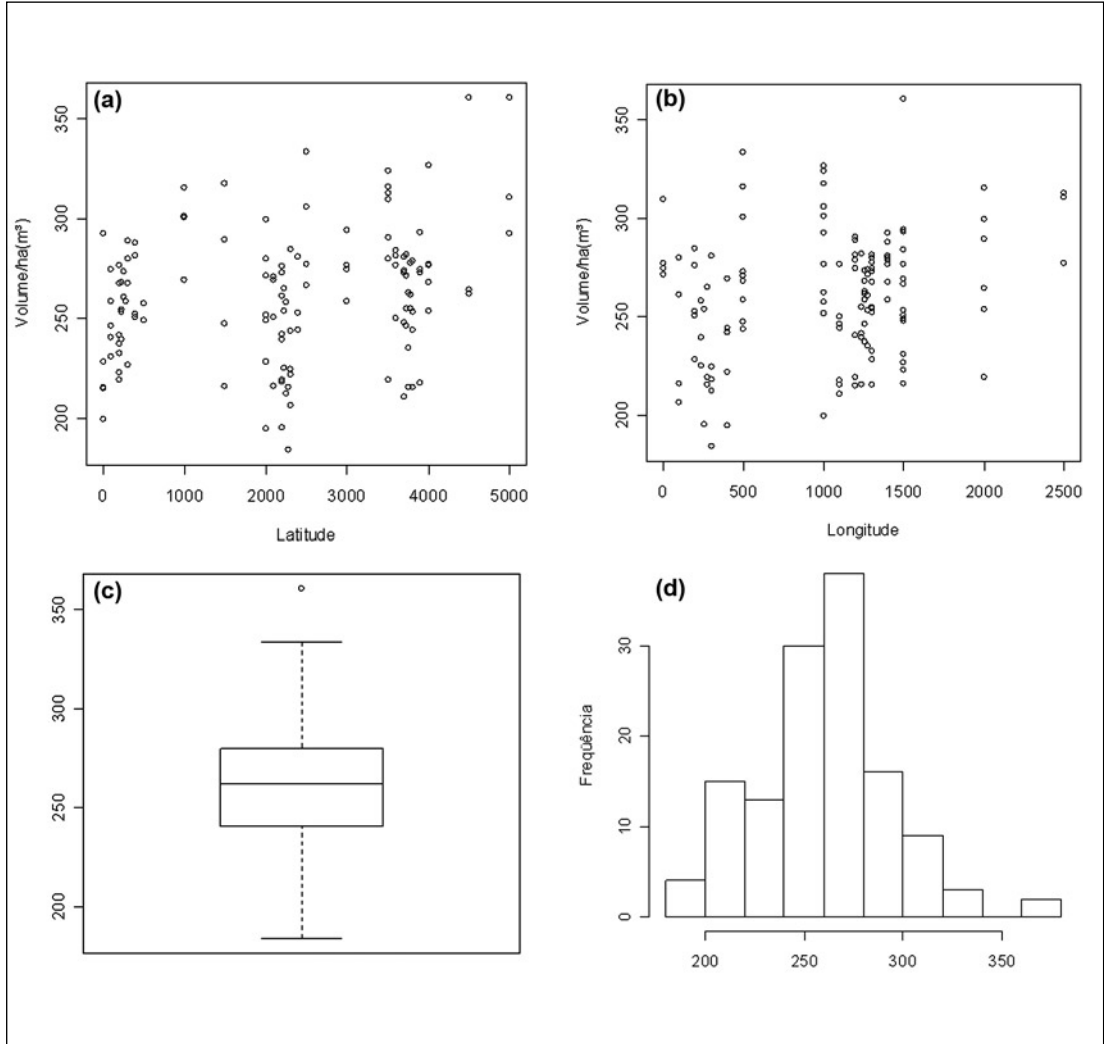


Figura 2.

Gráficos exploratórios do volume por hectare; (a) relação do volume com a latitude; (b) relação do volume com a longitude; (c) histograma de frequência e (d) gráfico de Box Plot.

(Exploratory graph of volume per hectare; (a) relation between volume and latitude; (b) relation between volume and longitude; (c) histogram of frequency and (d) Box Plot graph).

Tabela 1.

Teste de Jarque.Bera para as diferenças nas respectivas distâncias do semivariograma empírico. (Jarque.Bera test for the differences on respective distances of the empiric semivariogram).

DISTÂNCIA (m)	D	P – value	N
20	3,5916	0,1660	53
40	6,9086	0,0316	108
60	16,0339	0,0003*	85
80	13,0135	0,0015*	65
100	4,0572	0,1315	176
200	7,5561	0,0229	807
300	5,9096	0,0521	575
400	3,7663	0,1521	440
500	0,7767	0,6782	155
800	1,0530	0,5907	878
1100	0,7344	0,6927	914
1400	0,2283	0,8921	1133
1700	7,6848	0,0214	1673
2000	0,1707	0,9182	826

D – estatística de Jarque.Bera; n – número de pares de pontos em cada distância do semivariograma.

Avaliação do Critério de Informação de Akaike (AIC)

Nas Figuras 3 e 4 estão apresentados os semivariogramas experimentais e as linhas referentes aos modelos testados conforme cada método de ajuste. As linhas dos dois modelos, exponencial e Matérn, praticamente confundiram-se nos métodos dos quadrados mínimos, havendo diferenças na estimativa do nugget. No método da Máxima Verossimilhança, até 500 metros de distância, os dois modelos tiveram comportamento semelhante. A seleção de um modelo em qualquer método, analisando apenas o comportamento dos gráficos pode ser questionável. Portanto, uma definição mais objetiva pode ser realizada através do AIC de cada modelo nos respectivos métodos de ajuste.

A seleção do modelo correto, que representa adequadamente as semivariâncias, principalmente nas pequenas distâncias, é altamente desejável no processo de krigagem (NIELSEN e WENDROTH, 2003). Segundo estes autores, um bom ajuste de semivariâncias a pequenas distâncias é desejável para se ter um modelo de semivariograma de boa qualidade. Visualmente, em todos os modelos dos métodos avaliados no presente estudo tiveram desempenho semelhante nas pequenas distâncias. Assim, qualquer um deles pode possuir um bom desempenho no processo de krigagem. Especificamente, ao se analisar as curvas geradas pelos quatro méto-

dos de ajuste, para o modelo exponencial (linha fina), verificou-se similaridade entre as mesmas. Portanto, visualmente o modelo exponencial não diferiu entre os métodos de ajuste. Porém, esta é uma avaliação subjetiva, logo, o uso de critérios quantitativos, como o AIC e as técnicas de validações, são mais aplicáveis para a seleção do melhor modelo.

Na Tabela 2 está apresentado o AIC dos ajustes para cada um dos modelos testados nos respectivos métodos de ajuste. Nesta etapa de avaliação, procurou-se comparar os dois modelos em cada método de ajuste. O AIC para cada método de ajuste através de modelos avaliados, apontou como o melhor modelo o exponencial, por apresentar o menor AIC em todos os métodos de ajuste. Portanto, ele foi denominado modelo suporte para cálculo da diferença de AIC conforme Burnhan e Anderson (2002). A diferença de AIC (Δ_i) entre os modelos exponencial e Matérn, em cada método de ajuste foi respectivamente: 17,34; 109,3 e 0,37. Para os métodos dos Quadrados Mínimos, o modelo exponencial teve melhor desempenho ($\Delta_i > 2$). No método da Máxima Verossimilhança, o modelo Matérn foi substancialmente tão bom quanto o modelo exponencial para esta base de dados ($\Delta_i = 0,37$). A igualdade entre estes dois modelos para o método da Máxima Verossimilhança também, pode ser observada na semelhança entre os parâmetros es-

timados para cada um dos modelos (Tabela 3). Estes resultados contrariam a análise da Figura 3(a), que revelou diferenças substanciais entre os

modelos no método da Máxima Verossimilhança em que se evidenciou a importância de selecionar modelos por métodos quantitativos.

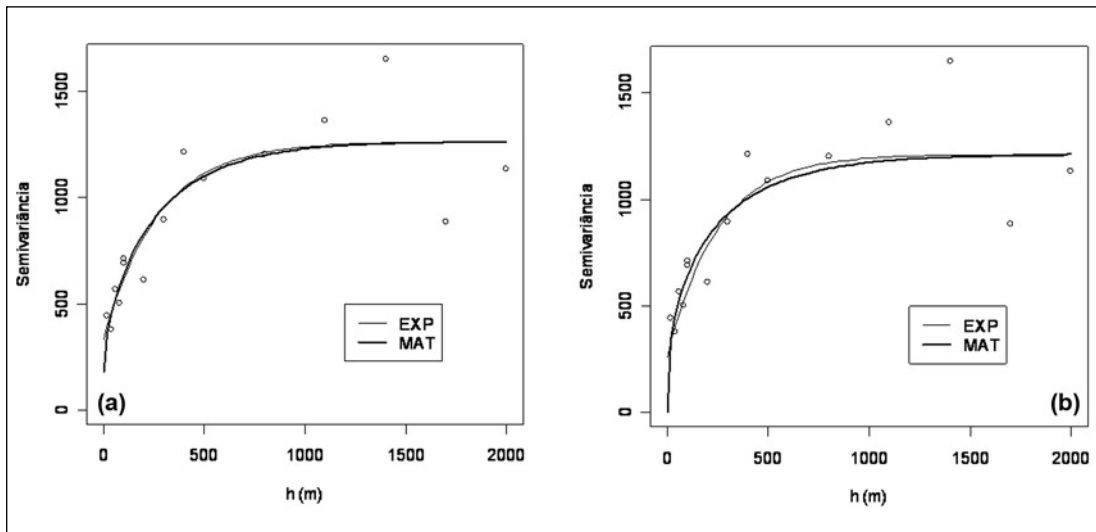


Figura 3. Modelos ajustados ao semivariograma conforme os Métodos dos Quadrados Mínimos Ordinários (a) e Quadrados Mínimos Ponderados (b). (Models fitted to the semivariogram based on Ordinary Minimum Square (a) and Weighted Minimum Square methods (b)).

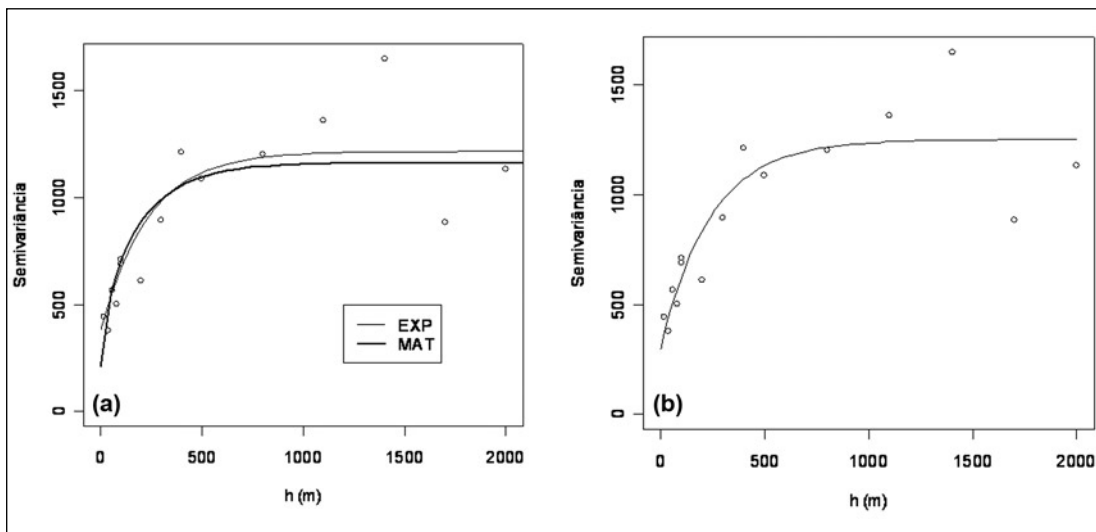


Figura 4. Modelos ajustados ao semivariograma conforme o Método da Máxima Verossimilhança (a) e “a sentimento” para o modelo exponencial (b). (Models adjusted at the semivariogram based on Maximum Likelihood (a) and “at feelings” (b) methods to the exponential model).

Tabela 2.

Critério de Akaike (AIC), para cada ajuste conforme o método e o modelo avaliado. (Akaike Criteria (AIC) for each adjusting according to method and model evaluated).

MÉTODO DE AJUSTE	AIC	
	Exponencial	Matérn
OLS	1257,39	1274,73
WLS	1259,78	1369,09
ML	1256,02	1256,39
“a sentimento”	1258,87	-

Tabela 3.

Parâmetros estimados para cada modelo através dos respectivos métodos de ajustes. (Parameters estimated for each model through of respective adjusting methods).

MÉTODO DE AJUSTE	Exponencial			Matérn		
	τ^2	σ^2	Φ	τ^2	σ^2	Φ
OLS	312,5	956,86	274,4	153,4	1119,84	362,1
WLS	218,6	991,8	240,5	6,5	1203,86	295,3
ML	383,0	831,6	233,4	214,9	949,0	240,0
“a sentimento”	278,8	925,0	291,7	-	-	-

Analisando o comportamento dos métodos para os dois modelos, verificou-se que os menores valores de AIC foram obtidos pelo Método da Máxima Verossimilhança. Este resultado já era esperado, uma vez que o princípio do método consiste em maximizar o melhor conjunto de parâmetros dentre os possíveis, para uma dada base de dados. Assim, este método foi considerado o método suporte para comparação com os outros métodos. As diferenças de AIC entre o Método da Máxima Verossimilhança e os métodos dos Quadrados Mínimos (Ordinários e Ponderados) e “a sentimento”, para o modelo exponencial, foram respectivamente 1,4; 3,76; e 2,85. Conforme Burnhan e Anderson (2002), se a diferença de AIC for superior a dois, indica que o modelo (i) difere consideravelmente do modelo suporte (< AIC). Assim, o método da Máxima Verossimilhança, para o modelo exponencial, teve melhor desempenho quando comparado com o método “a sentimento” e o método dos Quadrados Mínimos Ponderados. A diferença de AIC do método da Máxima Verossimilhança para o método dos Quadrados Mínimos Ordinários, revelou que este último foi tão bom quanto o método da Máxima Verossimilhança. Porém, a diferença do AIC de 1,4 está no limiar do valor proposto por Burnhan e Anderson (2002). O ideal seria que esta diferença fosse a mais próxima de zero. Assim, pode-se considerar que nenhum método testado para o modelo exponencial foi tão bom quanto o método da Máxima

Verossimilhança. Estas diferenças foram maiores quando se avaliou o modelo Matérn perante os métodos de ajuste. Quanto maior essas diferenças, mais distante estarão as curvas dos outros métodos em relação à curva gerada pelo método da Máxima Verossimilhança.

O método denominado de “a sentimento” (Figura 4b), modelo exponencial, apresentou uma diferença de AIC para o método da Máxima Verossimilhança de 2,85. Este método de ajuste foi amplamente utilizado nos anos 80 e 90, sobretudo na área de Ciência do Solo, como se pode observar em trabalhos de Vieira (2000), dentre outros. Segundo Soares (2000), na prática de modelização do semivariograma experimental por uma curva atenuada, deve-se conjugar todo conhecimento pericial e interdisciplinar que se tem do fenômeno, a fim de se estabelecer uma função que represente bem o padrão de continuidade espacial. No presente estudo, o critério de AIC revelou que este tipo de ajuste pode proporcionar erros substanciais no processo de inferência estatística.

A superioridade do Método da Máxima Verossimilhança em relação aos demais métodos de ajuste, ocorre devido às fontes de subjetividade que envolvem os outros métodos. No método “a sentimento” estão envolvidas duas fontes de subjetividade no processo de ajuste: o número de classes de distâncias definidas para o semivariograma experimental e a incerteza por onde passa a cur-

va do modelo entre os pontos do semivariograma experimental. O uso dos métodos dos Quadrados Mínimos anula a segunda fonte de subjetividade. Porém, a primeira ainda permanece envolvendo os dois métodos de ajuste, ou seja, ao se alterar o número de classes de distância no semivariograma experimental, tem-se um novo conjunto de parâmetros (CRESSIE, 1993). No presente estudo, fixou-se o número de classes de distâncias para os métodos de ajuste, possibilitando uma melhor comparação do AIC entre os métodos.

No método da Máxima Verossimilhança não há nenhuma das duas fontes de subjetividade descritas anteriormente. O método consiste em estimar os parâmetros quanto ao conjunto original dos dados e não quanto aos pontos do semivariograma experimental. Desta forma, ao se aplicar o método para um mesmo conjunto de dados, serão sempre obtidos os mesmos parâmetros, eliminando assim, as fontes de subjetividade. Na condição de normalidade, o método da Máxima Verossimilhança proporcionará o melhor conjunto de parâmetros para um dado conjunto de dados.

Avaliação dos modelos e métodos de ajuste através das técnicas de validação

Esta avaliação consistiu em efetuar a técnica de validação cruzada e preditiva para cada modelo através do respectivo método de ajuste, podendo auxiliar na seleção do melhor modelo e método de ajuste. A seguir estão apresentados nas Tabelas 4 e 5 as estatísticas referentes à qualidade dos ajustes mediante a técnica de validações cruzada e preditiva, respectivamente (\overline{ER}) - erro médio reduzido; (S_{ER}) - desvio padrão do erro reduzido; (EA) - erro absoluto.

Conforme McBratney e Webster (1986), Cressie (1993) e Vieira (2000), aplicando-se a condição de não tendenciosidade, o valor populacional para o erro médio reduzido deve ser zero e do desvio padrão do erro reduzido deve ser igual a

um. Verificou-se pelas tabelas 4 e 5, que a condição de não tendenciosidade e de variância mínima foi mais pronunciada na validação cruzada do que na preditiva. Este resultado foi esperado, uma vez que a validação cruzada ocorreu no contexto da base de dados que geraram o modelo, enquanto que a validação preditiva ocorre no contexto de novos dados do mesmo povoamento. Verificou-se pelas estatísticas das duas técnicas de validação, que os modelos ajustados pelo método da Máxima Verossimilhança apresentaram melhor desempenho do que os demais métodos de ajuste. Este resultado foi semelhante àquele gerado pelo AIC. Todavia, as diferenças destas estatísticas em cada validação foram relativamente pequenas.

Os valores apresentados nas Tabelas 4 e 5 levaram à conclusão de que houve uma similaridade acentuada entre os modelos exponencial e Matérn ajustados pelo método da Máxima Verossimilhança. Assim, no presente estudo optou-se por trabalhar com o modelo exponencial, uma vez que sua estrutura matemática é menos complexa do que a do modelo Matérn. As estatísticas das validações cruzada e preditiva para o modelo exponencial revelaram que o método da Máxima Verossimilhança foi superior aos demais métodos avaliados. A partir desta análise, para todo processo de estimação, seja krigagem ou geração de intervalos de confiança, recomenda-se o modelo exponencial ajustado pela Máxima Verossimilhança.

É interessante ressaltar que as duas técnicas de validação indicaram o mesmo modelo e o mesmo método de ajuste com melhor desempenho. Isto é importante, uma vez que nem sempre é possível contar com uma nova base de dados para efetuar a validação preditiva, ou seja, pode-se fazer o uso da validação cruzada como técnica auxiliar de seleção de modelos. Com estes resultados, o presente trabalho propõe que se utilize o critério de AIC e a validação cruzada como medidas a serem avaliadas na seleção de modelos.

Tabela 4.

Estatísticas da validação cruzada para cada modelo nos respectivos métodos de ajuste. (Cross validation statistics for each model on respective adjusting methods).

MODELO	MÉTODO DE AJUSTE	(\overline{ER})	(S_{ER})	EA (m ³ /ha)
Exponencial	“a sentimento”	0,0204	1,1974	23,9473
Exponencial	OLS	0,0203	1,1375	23,8037
Exponencial	WLS	0,0203	1,1375	23,8037
Exponencial	ML	0,0157	1,0216	23,6540
Matérn	OLS	0,0184	1,0765	23,8475
Matérn	WLS	0,0236	1,1776	24,0457
Matérn	ML	0,0169	1,0189	23,5173

Tabela 5.

Estatísticas da validação preditiva para cada modelo nos respectivos métodos de ajuste. (Predictive validation statistics for each model on respective adjusting methods).

MODELO	MÉTODO DE AJUSTE	$(\bar{E}R)$	(S_{ER})	EA (m ³ /ha)
Exponencial	“a sentimento”	-0,2102	1,4134	32,5744
Exponencial	OLS	-0,2012	1,3905	32,5585
Exponencial	WLS	-0,2012	1,3905	32,5586
Exponencial	ML	-0,1847	1,2548	32,0578
Matérn	OLS	-0,2010	1,3423	32,2148
Matérn	WLS	-0,1611	1,5783	32,2809
Matérn	ML	-0,1858	1,2539	31,4685

CONCLUSÃO

Nas condições em que o presente trabalho foi desenvolvido, verificou-se que o teste de normalidade para as diferenças entre pares de pontos, em cada distância do semivariograma, foi importante por aferir maior segurança ao assumir um Modelo Multivariado Normal para o fenômeno estudado. Mediante a condição de normalidade, o critério de AIC e as técnicas de validações detectaram similaridade acentuada entre os modelos avaliados ajustados pelo método da Máxima Verossimilhança. Verificou-se que as diferenças nas estatísticas das validações foram pequenas. Assim, sugere-se o uso conjugado do critério de AIC e validação cruzada para a seleção de modelos espaciais. Apesar da igualdade entre os modelos testados quando ajustados pelo método da Máxima Verossimilhança, pode-se optar pelo modelo exponencial, uma vez que sua estrutura matemática é menos complexa.

AUTORES

José Marcio de Mello é Professor Doutor do Departamento de Ciências Florestais da UFLA – Universidade Federal de Lavras – Caixa Postal 3037 – Lavras, MG – 37200-000 – E-mail: josemarcio@ufla.br

João Luís Ferreira Batista é Professor Doutor do Departamento de Ciências Florestais da ESALQ – Escola Superior de Agricultura Luiz de Queiroz da Universidade de São Paulo - Av. Pádua Dias, 11 - Piracicaba, SP - 13418-900 – E-mail: parsilval@usp.br

Paulo Justiniano Ribeiro Júnior é Professor Doutor do Departamento de Estatística da UFPR – Universidade Federal do Paraná – Av. Lothário Meissner, 3400 – Jardim Botânico – Curitiba, PR – 80120 – E-mail: paulojus@est.ufpr.br

Marcelo Silva é Professor Doutor do Departamento de Ciências Exatas da UFLA – Universidade Federal de Lavras – Caixa Postal 3037 – Lavras, MG – 37200-000 – E-mail: marcelos@ufla.br

REFERÊNCIAS

- AKAIKE, H. Information measures and model selection. **International Statistical Institute**, Voorburg, v.44, p.277-291, 1983.
- AUBRY, P.; DEBOUZIE, D. Geostatistical estimation variance for the spatial mean in two-dimensional systematic sampling. **Ecology**, Ithaca, v. 81, n. 2, p. 543-553, 2000.
- BIONDI, F.; MYERS, D.E.; AVERY, C.C. Geostatistically modeling stem size and increment in an old-growth forest. **Canadian Journal of Forest Research**, Ottawa, v. 24, n. 7, p. 1354-1368, 1994.
- BURNHAN, K.P.; ANDERSON, D.R. **Model selection and multimodel inference**. 2.ed. New York: Springer, 2002. 350p.
- BUSSAB, W.O.; MORETTIN, P.A. **Estatística básica**. 5.ed. São Paulo: Saraiva, 2002. 526p.
- COHEN, A.G.; SPIES, T.; BRADSHAW, G. Semivariograms of digital imagery for analysis of conifer canopy structure. **Remote Sensing of Environment**, New York, v. 35, p.167-178, 1990.
- CRESSIE, A.G. Fitting variogram models using weighted least squares. **Journal of the International Association of Mathematical Geology**, Kingston, v.17, p.563-586, 1985.
- CRESSIE, A.G. **Statistics for spatial data**. New York: Wiley, 1993. 900p.
- DAVID, M. **Geostatistical Ore Reserve Estimation**. Amsterdam: Elsevier, 1977. 205p.

- DIGGLE, P.J.; RIBEIRO JÚNIOR, P.J. Model-based geostatistics. In: SIMPÓSIO NACIONAL DE PROBABILIDADE E ESTATÍSTICA, 14., 2000, Caxambu. **Resumos.....** 192p.
- JARQUE, C.M.; BERA, A.K. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. **Economics Letters**, v.6, p.255-59, 1980.
- JOURNEL, A.G.; HUIJBREGTS, C.J. **Mining geostatistics**. London: Academic Press, 1978. 600 p.
- KULBACK, S.; LEIBLER, R.A. On information and sufficiency. **Annals of Mathematical Statistics**, v.22, p.79-86, 1951.
- McBRATNEY, A.G.; WEBSTER, A.G. Choosing functions for semi-variograms and fitting them to sampling estimates. **Journal of Soil Science**, v.37, p.617-39, 1986.
- MELLO, C.R.; LIMA, J.M.; SILVA, A.M.; MELLO, J.M.; OLIVEIRA, M.S. Krigagem e inverso do quadrado da distância para interpolação dos parâmetros da Equação de chuvas intensas. **Revista Brasileira de Ciência do Solo**, Viçosa, v.27, p.925-933, 2003.
- NIELSEN, D.R.; WENDROTH, O. **Spatial and temporal statistics: sampling fields soils and their vegetation**. Cremlingen-Destedt: Catena Verlag, 2003. 398p.
- R DEVELOPMENT CORE TEAM. **R: A language and environment for statistical computing**. Vienna: R Foundation for Statistical Computing, 2003. Disponível em: www.R-project.org
- RIBEIRO JÚNIOR, P.J. **Métodos geoestatísticos no estudo da variabilidade espacial de parâmetros do solo**. 1995. 99p. Dissertação (Mestrado) – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 1995.
- RIBEIRO JÚNIOR., P.J.; DIGGLE, P.J. GeoR: a package for geostatistical analysis. **R-news**, v.1, n.2, p.15-18, 2001.
- SAMRA, J.S.; GILL, H.S.; BHATIA, V.K. Spatial stochastic modelling of growth and forest resource evaluation. **Forest Science**, Bethesda, v.35, n.3, p.663-676, 1989.
- SOARES, A.G. **Geoestatística para as ciências da terra e do ambiente**. Lisboa: ISTPress, 2000. 186p.
- VIEIRA, S.R. Geoestatística em estudos de variabilidade espacial do solo. In: UFV – UNIVERSIDADE FEDERAL DE VIÇOSA. **Tópicos em Ciências do Solo**. Viçosa: SBCS, 2000.
- WONNACOTT, T.H.; WONNACOTT, R.J. **Introductory statistics**. 2.ed. New York: John Wiley Sons, 1972. 510p.