



Revista Árvore

ISSN: 0100-6762

r.arvore@ufv.br

Universidade Federal de Viçosa

Brasil

Oda-Souza, Melissa; Barbin, Décio; Justiniano Ribeiro Júnior, Paulo; Stape, José Luiz
Aplicação de métodos geoestatísticos para identificação de dependência espacial na análise de dados
de um ensaio de espaçamento florestal em delineamento sistemático tipo leque

Revista Árvore, vol. 32, núm. 3, mayo-junio, 2008, pp. 499-509

Universidade Federal de Viçosa

Viçosa, Brasil

Disponível em: <http://www.redalyc.org/articulo.oa?id=48813382011>

- ▶ Como citar este artigo
- ▶ Número completo
- ▶ Mais artigos
- ▶ Home da revista no Redalyc

redalyc.org

Sistema de Informação Científica

Rede de Revistas Científicas da América Latina, Caribe , Espanha e Portugal
Projeto acadêmico sem fins lucrativos desenvolvido no âmbito da iniciativa Acesso Aberto

APLICAÇÃO DE MÉTODOS GEOESTATÍSTICOS PARA IDENTIFICAÇÃO DE DEPENDÊNCIA ESPACIAL NA ANÁLISE DE DADOS DE UM ENSAIO DE ESPAÇAMENTO FLORESTAL EM DELINEAMENTO SISTEMÁTICO TIPO LEQUE¹

Melissa Oda-Souza², Décio Barbin³, Paulo Justiniano Ribeiro Júnior⁴, José Luiz Stape³

RESUMO – Os delineamentos sistemáticos se destacam pela sua compacidade e abrangência e por permitir testar maior número de espaçamentos possíveis. No entanto, não é utilizado devido ao arranjo sistemático (não casualizado) das plantas e à alta sensibilidade para valores perdidos. O objetivo deste trabalho foi descrever o modelo geoestatístico e métodos associados de inferência no contexto de análise de experimentos não aleatorizados, reportando resultados aplicados para identificar a dependência espacial em um particular experimento em delineamento sistemático tipo leque de *Eucalyptus dunnii*. Também foram propostas, analisadas e comparadas diferentes alternativas para tratar dados faltantes que pudessem advir de falhas e, ou, mortalidade de plantas. Os dados foram analisados seguindo-se três modelos que diferiram, com co-variáveis, na forma de tratar os dados faltantes. Para cada um destes foi construído um semivariograma, com o ajuste de três modelos de função de correlação, sendo os parâmetros estimados pelo método de máxima verossimilhança e selecionados pelo critério de Akaike. Esses modelos, com e sem o componente espacial, foram comparados pelo teste da razão de verossimilhança. De acordo com os resultados, verificou-se que: (1) as co-variáveis interagiram positivamente com a variável de resposta, evitando que dados coletados sejam desperdiçados; (2) a comparação dos modelos, com e sem o componente espacial, não confirmou a existência de dependência; (3) a incorporação da estrutura de dependência espacial aos modelos observacionais recuperou a capacidade de fazer inferências válidas na ausência de aleatorização, permitindo contornar problemas operacionais e, assim, garantindo que os dados possam ser submetidos a uma análise clássica.

Palavras-chave: *Eucalyptus dunnii*, semivariograma, máxima verossimilhança, critério de informação de Akaike e teste de razão de verossimilhança.

APPLICATION OF GEOSTATISTICAL METHODS TO IDENTIFY SPATIAL DEPENDENCE IN THE DATA ANALYSIS OF A FOREST SPACING EXPERIMENT WITH A FAN SYSTEMATIC DESIGN

ABSTRACT – The systematic design stands out for its compactness, broadness and for allowing the testing of a larger number of spacings. However, it is not used due to the systematic arrangement (non-randomized) of the plants and the high sensibility for missing values. The aim of this work was to describe the geostatistic model and associated methods of inference in the analysis context of non-randomized experiment, reporting applied results to identify the spatial dependence in a fan systematic design of *Eucalyptus dunnii*. Furthermore, different alternatives for treating missing values that can occur from flaws and/or mortality of plants were

¹ Recebido em 30.05.2007 e aceito para publicação em 19.05.2008.

² Programa de Pós-Graduação em Ciências Florestais da Escola Superior de Agricultura Luiz de Queiroz (ESALQ), Piracicaba-SP. E-mail:<melissa@esalq.usp.br>.

³ Departamento de Ciências Florestais da ESALQ. E-mail: <

⁴ Departamento de Estatística da Universidade Federal do Paraná (UFPR), Curitiba-PR. E-mail:<stape@pq.cnpq.br>.

proposed, analyzed and compared. Data were analyzed by three models that differed, with covariates, in the form of modeling missing data values. A semivariogram was built for each model, adjusting three correlation function models, being the parameters estimated through the maximum likelihood method and selected by the Akaike's criterion. These models, with and without the spatial component, were compared by the likelihood ratio test. The results showed that: (1) the covariates interacted positively with the response variable, avoiding data to be discarded; (2) the model comparison, with and without the spatial component, did not confirm the existence of dependence; (3) the incorporation of the spatial dependence structure into the observational models recovered the capacity to make valid inferences in the absence of randomization, overcoming operational problems and guaranteeing that the data can be subjected to classic analysis.

Keywords: *Eucalyptus dunnii, semivariogram, maximum likelihood, Akaike's information criterion and likelihood ratio test.*

1. INTRODUÇÃO

A definição do espaçamento de plantio, para determinada espécie ou clone, é de alta importância em silvicultura, pois influencia as taxas de crescimento, sobrevivência e produtividade de madeira por hectare, afetando as práticas de manejo, colheita e, consequentemente, os custos de produção florestal (STAPE, 1995; RONDON, 2002). Para determinação do espaçamento ótimo de plantio, os ensaios experimentais devem conter um número suficiente de indivíduos mensuráveis, implicando grandes áreas para suas instalações (SILVA, 2005). Outro complicador é a necessidade de bordadura entre os tratamentos, visando evitar a influência exercida pelas árvores de uma parcela sobre as das parcelas vizinhas. Dessa forma, a instalação de experimentos tradicionais com repetição, aleatorização dos tratamentos, controle local e bordadura, têm restringido o número de espaçamentos testados nos ensaios de campo, dadas às dificuldades relativas às disponibilidades de área e recursos para instalação, manutenção e avaliações (STAPE, 1995).

Visando superar tais restrições, foram propostos na década de 1960 alguns delineamentos alternativos, entre eles os delineamentos sistemáticos de Nelder (1962). Esses delineamentos, inicialmente utilizados na área agrícola, foram usados na área florestal para estudos preliminares (FREEMAN, 1964; BLEASDALE, 1966; PANETSOS, 1980; HUXLEY, 1985; IMADA et al., 1997). Além de permitir avaliar maior número de espaçamentos possíveis, os delineamentos sistemáticos também se destacam pela sua compacidade, necessitando de pequenas áreas experimentais que facilitam o manejo e abrangência. No entanto, dois fatores são limitantes na utilização desse delineamento. O primeiro é que o arranjo sistemático (não casualizado) das plantas

não permite o uso das análises convencionais por ferir o princípio básico da casualização na alocação dos tratamentos. A falta da casualização não garante a independência dos erros associados a cada unidade experimental, tornando as unidades experimentais correlacionadas. O segundo é a alta sensibilidade para valores perdidos. Segundo Nelder (1962), devido à inexistência de bordadura entre plantas e entre linhas, a ocorrência de falhas e mortalidade de algumas árvores faz que as plantas vizinhas tenham área disponível de crescimento diferente daquela inicialmente estabelecida para os tratamentos. Assim, quando uma planta é perdida, o espaçamento das plantas vizinhas é alterado, e com isso esses valores não podem ser incluídos no conjunto de dados, e informações consideráveis são excluídas das análises. Mas se as ocorrências forem em pequeno número e aleatórias dentro do experimento, Gomez e De Datta (1972) propuseram eliminar da análise experimental dados de todas as plantas na vizinhança das falhas ou de mortes.

A análise usual de experimentos se baseia nos princípios de aleatorização para justificar as inferências estatísticas. Entretanto, em face das restrições para condução de um experimento aleatorizado, propõe-se um modelo geoestatístico para o mecanismo gerador dos dados, adotando métodos de inferência baseados na função de verossimilhança.

Na ausência de aleatorização, não há como evitar que as condições ambientais induzam dependência entre as unidades mensuradas. Dessa forma, o modelo especificado inclui explicitamente uma estrutura de dependência. Especificamente na situação considerada aqui modela-se a dependência através de um modelo geoestatístico como descrito, por exemplo, por Diggle e Ribeiro Jr. (2007).

O objetivo deste trabalho foi descrever o modelo geoestatístico e métodos associados de inferência no contexto de análise de experimentos não aleatorizados, reportando resultados aplicados para identificar a dependência espacial em um particular experimento em delineamento sistemático tipo *leque* de *Eucalyptus dunnii* com idade de 6 anos. Adicionalmente são propostas, analisadas e comparadas diferentes alternativas para tratar dados faltantes que podem advir de falhas e, ou, mortalidade de plantas.

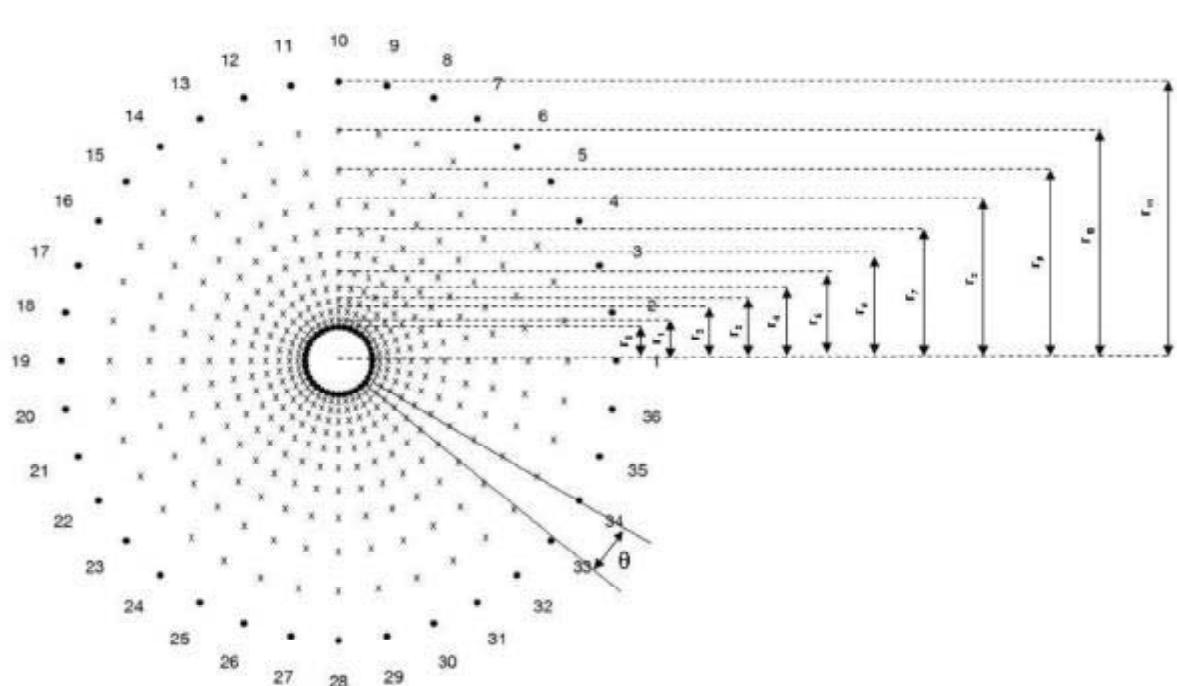
2. MATERIAL E MÉTODOS

2.1. Descrição da Área e Obtenção dos Dados

O trabalho foi conduzido na fazenda Ibiti, situada

no Município de Itararé, Estado de São Paulo, com coordenadas geográficas de 24° 11' S de latitude, 49° 08' W de longitude e 980 m de altitude.

Os dados para realização do estudo foram coletados em um experimento de espaçamento de *Eucalyptus dunnii* em delineamento sistemático tipo *leque*, conforme o modelo proposto por Nelder (1962) e denotado por 1a. Na Figura 1, tem-se o croqui do experimento que foi instalado com 10 tratamentos (arcos) e 36 repetições (raios) contando com bordadura interna com raio de 5,30 m e com uma bordadura externa com raio de 43,60 m. Em cada tratamento foi associada uma distância radial, área por planta e densidade conforme o Quadro 1a, de forma a manter uma retangularidade, isto é, garantindo distâncias constantes entre plantas no raio e nos arcos (STAPE, 1995).



×- plantas úteis; ● -bordadura; 1:36 - número dos raios; $\theta = 10^\circ$. $r_0 = 5,30$ m, $r_1 = 6,42$ m, $r_2 = 7,78$ m, $r_3 = 9,42$ m, $r_4 = 11,41$ m, $r_5 = 13,82$ m, $r_6 = 16,74$ m, $r_7 = 20,27$ m, $r_8 = 24,55$ m, $r_9 = 29,73$, $r_{10} = 36,00$ m, $r_{11} = 43,60$ m, calculado pela por: $m = r - r_0 \cdot a^n$, em que $a=1,21$ e $r_0 = 5,30$ m a distância radial do primeiro raio.

Figura 1 – Croqui de instalação do experimento sistemático leque com 10 tratamentos e 36 repetições com *Eucalyptus dunnii*.

Figure 1 – Installation outline of a fan systematic design experiment with 10 treatments and 36 repetitions with *Eucalyptus dunnii*.

Quadro 1 – Valores da distância radial, área por planta e densidade de plantas de 10 tratamentos (espaçamentos) no delineamento sistemático tipo *leque*

Table 1 – Radial distance values, area per plant and plant density for 10 treatments (spacings) in the fan systematic design

Tratamentos	Distância Radial (m)	Área (m^2 planta $^{-1}$)	Densidade (plantas ha $^{-1}$)
1	6,42	1,40	7143
2	7,78	2,05	4878
3	9,42	3,01	3322
4	11,41	4,41	2268
5	13,82	6,47	1546
6	16,74	9,50	1053
7	20,27	13,93	718
8	24,55	20,43	489
9	29,73	29,97	334
10	36,00	43,95	228

Foram realizadas avaliações aos 26, 42, 48, 60, 71 meses de idade, coletando-se dados individuais referentes à CAP (circunferência à altura do peito, 1,30 m em relação ao solo) e altura total das plantas úteis, sendo utilizados os dados da última avaliação. Para o cálculo do volume, utilizou-se o seguinte modelo de equação volumétrica:

$$V = \exp[-10,257052 + 0,9964961 \times \ln(DAP^2 \times H)] \quad (1)$$

em que, V é o volume individual da árvore (m^3) - DAP (diâmetro à altura do peito) = CAP/ (cm) e H é a altura (m).

A mortalidade no ensaio foi inferior a 2%, e, embora baixa, essa ocorrência foi utilizada como informação complementar para diferir os modelos, permitindo comparar diferentes estratégias de tratamento dos dados faltantes. Os resultados complementares das avaliações dendrométricas, assim como a caracterização do material genético, solo, bioclimática e práticas silviculturais podem ser encontrados em Stape (1995).

2.2. Análise Espacial do Experimento

2.2.1 Modelo Geoestatístico

O modelo geoestatístico descreve a dependência espacial através de um componente de erro aleatório e espacialmente correlacionado, assumindo-se que os dados $y = y_1, \dots, y_n$ seguem o modelo linear Gaussiano dado por:

$$Y_i = \mu(x_i) + S(x_i) + e_i \quad (2)$$

em que, Y_i é o valor observado na posição x_i , e x_i é a posição espacial da i-ésima parcela, $\mu(x_i)$ é

determinada a partir de um modelo de regressão linear dado por:

$$\mu(x_i) = \sum_{k=1}^p f_k(x_i) \beta_k \quad (3)$$

em que, $f_k(\cdot)$ são os valores observados de variáveis explanatórias. O termo $S(x_i)$ denota o valor na posição de um processo estacionário gaussiano, com média zero, variância σ^2 e função de correlação $\rho(u; \phi)$ com argumento dado pela distância u de separação entre observações e parametrizada por ϕ , que descreve o grau de associação espacial. Decorre, então, que para um conjunto finito de observações $S \sim NMV(0, \sigma^2 R(\phi))$, em que $R(\phi)$ é a matriz de co-variância $n \times n$ com (i,j) -ésimo elemento, $\rho(u)$, sendo $u = \|x_i - x_j\|$ a distância euclidiana entre x_i e x_j . O termo e denota variáveis aleatórias, independentes e normalmente distribuídas com média zero e variância τ^2 . Matricialmente, esse modelo é expresso por:

$$Y = F\beta + S(x) + e \quad (4)$$

Tem-se, então, que, para o modelo assumido, um vetor de variáveis observáveis Y de tamanho n possui distribuição normal multivariada, isto é, $Y \sim NMV(X\mu, \sigma^2 R(\phi) + \tau^2 I_n)$, em que I_n é uma matriz-identidade de dimensão n .

2.3. Especificação dos Modelos

Considerando que as observações medidas apresentam dependência espacial, os dados foram analisados segundo três modelos, que incluíram a estrutura de dependência segundo descrito anteriormente e

diferiram na forma de tratar os dados faltantes, conforme se descreve a seguir. Dessa maneira, no modelo I eliminaram-se os dados das plantas vizinhas às plantas mortas, conforme proposto por Gomez e De Datta (1972) e pode ser representado por:

$$Y_{ij} = \mu + t_i + S(x_i) + e_{ij} \quad i = 1, \dots, I \text{ e } j = 1, \dots, n_i \quad (5)$$

em que Y_{ij} é o valor observado que recebeu o i -ésimo tratamento na j -ésima repetição, μ é a média geral, e t_i é o efeito do i -ésimo tratamento.

No modelo II, as informações de plantas mortas foram consideradas como uma co-variável denotada por f_1 e no modelo III, além da co-variável *planta morta*, foi inserido o *tempo* da ocorrência da morte f_2 , como outra co-variável, conforme expresso nas equações (6) e (7), respectivamente.

$$\begin{aligned} Y_{ij} &= \mu + t_i + \beta_1 f_1(x_i) + S(x_i) + e_{ij} \quad i = 1, \dots, I \\ &\quad \text{e } j = 1, \dots, n_i \end{aligned} \quad (6)$$

em que, além dos efeitos considerados na equação (5), tem-se que β_1 é o parâmetro de regressão desconhecido associado à co-variável *planta morta*.

$$\begin{aligned} Y_{ij} &= \mu + t_i + \beta_1 f_1(x_i) + \beta_2 f_2(x_i) + S(x_i) + e_{ij} \\ &\quad i = 1, \dots, I \quad j = 1, \dots, n_i \quad k = 1, \dots, K \end{aligned} \quad (7)$$

em que, além dos efeitos considerados nas equações (5) e (6), tem-se que β_2 é o parâmetro de regressão desconhecido associado à co-variável tempo.

Para a construção da co-variável f_1 , considerou-se o número de plantas mortas, em que zero foi atribuído às plantas que não perderam vizinhas e um às que perderam uma planta vizinha, e no modelo III, na co-variável f_2 , foram levado em consideração os anos de avaliação. Dessa forma, uma planta morta no segundo ano receberá o valor 5 na co-variável no último ano de avaliação, no caso de sexto ano de avaliação; se a morte da planta ocorrer somente no sexto ano o valor na co-variável será 1.

Antecedendo a análise geoestatística, foi realizada uma análise exploratória descritiva dos dados com o objetivo de observar o seu comportamento quanto à normalidade, distribuição, homogeneidade. Embora não se considerem a estrutura de dependência espacial e a posição geográfica de cada observação, essa análise

é importante, pois permite a identificação de dados discrepantes (*outliers*) que podem exercer algum tipo de influência. Para testar a hipótese de normalidade, foi utilizada a família de transformações Box-Cox (BOX e COX, 1964). Essa família depende diretamente de um parâmetro λ , que é o coeficiente de transformação dos dados. Na prática para um conjunto de dados, obtém-se um intervalo de confiança para λ que, caso inclua o valor 1, indica que a transformação dos dados não é necessária. Esse tipo de transformação tem sido usado com sucesso em pesquisas florestais (SILVA e BAILEY, 1991) e é discutido no contexto de modelagem geoestatística por Christensen et al. (2001).

2.4. Semivariograma

A estrutura de variâncias e co-variâncias dos modelos considerados aqui pode ser resumida pelo semivariograma que é definido por $\gamma(u) = t^2 + s^2(1 - ?(u))$. Dessa forma alguns métodos geoestatísticos utilizam estimadores dos parâmetros que descrevam a dependência como o variograma (ou semivariograma), o co-variograma e o correlograma empíricos. Esses métodos têm sido aplicados e desenvolvidos em várias áreas como mineração (JOURNEL e HUIJBREGTS, 1978), solos (MCBRATNEY e WEBSTER, 1986) e inventários florestais (MANDALLAZ, 1991; SCHREUDER e WILLIAMS, 1995). O semivariograma deve o seu nome a Matheron (1963), e foi definido como uma função matemática que representa o nível de dependência entre valores de uma variável observada em duas posições na mesma região de estudo. O estimador de métodos dos momentos do semivariograma é dado por:

$$\hat{\gamma}(u) = \frac{1}{2N(u)} \sum_{a=1}^{N(u)} (z(x_a) - z(x_a + u))^2 \quad (8)$$

em que, $\hat{\gamma}(u)$ é a semivariância estimada para cada distância ou classe de distância u , $N(u)$ é o número de pares de pontos separados por u , e $z(x_a)$ e $z(x_a + u)$ são os valores das variáveis nos ponto x_a e $x_a + u$, respectivamente.

Na presença de dependência espacial, em geral, duas parcelas próximas são mais semelhantes do que duas mais afastadas, permitindo afirmar que $\hat{\gamma}(u)$ aumenta à medida que u também aumenta até estabilizar. O ponto em que os dados deixam de apresentar dependência espacial é chamado de patamar (*sill*) (σ^2), e a distância

a partir da origem até o patamar é chamada de alcance prático (*range*), que se denota aqui por ϕ , que é função do parâmetro ϕ . Assim, o alcance prático determina a distância a partir da qual a variável não tem mais influência relevante sobre a vizinhança. Pela definição, $\gamma(u)$ para $u=0$ deverá ser zero, mas na prática se observa que há casos em que $u=0$ e $\gamma(u)$ se aproxima de um valor positivo, denominado efeito-pepita (*nugget*), que nos modelos considerados aqui correspondem à variância τ^2 . Na Figura 2, esses parâmetros são apresentados em uma representação esquemática do semivariograma.

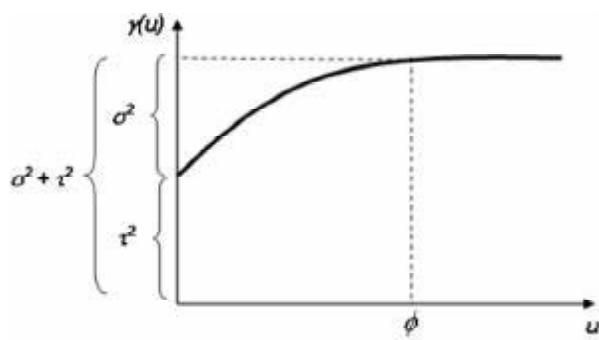


Figura 2 – Semivariograma com efeito pepita.
Figure 2 – Semivariogram with nugget effect

A fim de obter o conjunto de parâmetros para as respectivas funções, para cada modelo avaliado ajustou-se a função de correlação Matérn (DIGGLE e RIBEIRO JR., 2007):

$$\gamma(u) = \left\{ 2^{k-1} G(k) \right\}^{-1} (u/\phi)^k K_k(u/\phi) \quad (9)$$

com $k=0,5, 1,0$ e $2,0$, em que o primeiro caso corresponde ao modelo exponencial:

$$\gamma(u) = \exp\{-(u/\phi)\} \quad (10)$$

em que $\gamma(u)$ é a correlação entre pares de pontos separados pela distância u , ϕ é o parâmetro de alcance do modelo, k é o parâmetro que determina a suavidade do processo $S(x)$, $G(k)$ é a função Gama e K_k é a função Bessel modificada de terceiro tipo de ordem k .

2.5. Ajuste dos Modelos

Existem vários modelos de ajuste aplicáveis aos diferentes fenômenos com continuidade espacial. A sensibilidade dos semivariogramas para detectar a

variabilidade espacial das amostras está diretamente ligada ao melhor ajuste entre os dados experimentais e os modelos teóricos do semivariograma.

Na literatura, mencionam-se alguns métodos de ajuste de modelo de semivariograma, e um, freqüentemente utilizado, é chamado de ajuste a *sentimento*. Por meio deste, escolhe-se um modelo que será sobreposto aos pontos estimados, e a partir de uma nuvem de pontos obtém-se a estimativa dos parâmetros do semivariograma, por meio de uma avaliação visual. O processo deve ser repetido até que se obtenha um modelo adequado. Esse ajuste subjetivo é questionável e tende a ser cada vez menos utilizado (RIBEIRO JR., 1995).

Outros métodos utilizados são os de mínimos quadrados ordinários, ponderados ou generalizados. McBratney e Webster (1986) consideraram o método de mínimos quadrados ponderados o mais realista e computacionalmente eficiente. No entanto, com o desenvolvimento de recursos computacionais, o método de máxima verossimilhança tem sido uma técnica de estimação muito utilizada com maior freqüência, pois tem propriedades da teoria das grandes amostras que torna o seu resultado mais atrativo. O estimador é assintoticamente consistente, de forma que, quanto maior for o tamanho da amostra, mais próximos os valores das estimativas estarão dos verdadeiros valores; não-viesado, isto é, sua esperança é igual ao valor estimado; e também é assintoticamente eficiente, pois, quanto maior a amostra, maior a precisão das estimativas.

Considerando o modelo apresentado em (2) com estrutura de co-variancia da família Matérn e assumindo que $y=(y_1, \dots, y_n)$ segue uma distribuição multivariada Gaussiana, tem-se que a função de verossimilhança é expressa por:

$$L(\beta; \theta; y) = (2\pi)^{-\frac{n}{2}} |G(\theta)|^{-\frac{1}{2}} \exp \left[-\frac{1}{2}(y - F\beta)' \{G(\theta)\}^{-1} (y - F\beta) \right] \quad (11)$$

em que y é o vetor de observações experimentais; F é a matriz do delineamento, β é o vetor de parâmetros desconhecidos associados aos efeitos fixos; e θ são os parâmetros da matriz de co-variancia, $G(\theta) = \sigma^2 R(\phi) + \tau^2 I_n$ com $\theta = (\tau^2, \sigma^2, \phi)$.

A função de verossimilhança é uma função dos parâmetros desconhecidos, portanto, a estimativa de máxima verossimilhança consiste em encontrar os valores $(\hat{\beta}; \hat{\theta})$ estimativas de $(\beta; \theta)$ com máxima plausibilidade de terem produzido os valores observados y , através da maximização do logaritmo da função de verossimilhança dado por:

$$L(\mu; ?) = -\frac{1}{2}(n \log(2p) + \log|G(?)| + \\ y - F\beta)' \{G(?)\}^{-1} (y - F\beta)) \quad (11)$$

A maximização da equação (12) é obtida por procedimentos numéricos iterativos. Estes procedimentos envolvem computações que são de ordem proporcional ao cubo do número de observações (RIBEIRO JR., 1995).

2.6. Seleção dos Modelos

Para a avaliação da adequação dos modelos espaciais foi utilizado o critério de informação de Akaike - AIC (AKAIKE, 1983). Esse critério, em geral, tem sido utilizado para comparação de modelos com diferentes estruturas de co-variâncias. Primeiramente, comparou-se o AIC dos modelos para cada função de correlação e, em seguida, entre as diferentes funções de correlação. O AIC é baseado na teoria de decisão e é definido como a quantidade:

$$AIC = -2l + 2p \quad (12)$$

em que l é o logaritmo neperiano do máximo da função de verossimilhança, p é o número de parâmetros do modelo considerado. De acordo com esse critério, o melhor modelo é aquele que apresenta o menor valor de AIC.

Um segundo critério de avaliação foi através do teste de razão de verossimilhanças, comparando os modelos com e sem a estrutura de dependência espacial, em que razão de verossimilhanças somente é válida para modelos seqüencialmente reduzidos (RAO, 1973). Se denotarmos por L_1 a verossimilhança do modelo mais geral e L_2 a verossimilhança do modelo mais restrito, tem-se que $L_1 > L_2$ e, consequentemente, $\ln(L_1) > \ln(L_2)$. A estatística do teste da razão de verossimilhança (LR) foi calculada para testar a significância dos modelos com o componente espacial, em comparação com os modelos sem esse componente. A estatística do teste da razão de verossimilhança é definida como:

$$LR = -2 \log(L_2/L_1) = 2[\log(L_1) - \log(L_2)] \quad (14)$$

em que L_2 é a máxima verossimilhança para o modelo sem o componente espacial e L_1 a máxima verossimilhança para o modelo com o componente espacial.

Se a diferença do logaritmo das funções de verossimilhanças do modelo 2 *versus* o modelo 1, considerando-se a diferença entre os números de parâmetros dos modelos, for menor que o valor do χ^2 tabelado, considera-se o modelo com menor número de parâmetros como o mais adequado.

Neste estudo, as análises foram realizadas utilizando-se o pacote geoR (RIBEIRO JR. e DIGGLE, 2001) do projeto R de computação estatística (R DEVELOPMENT CORE TEAM, 2005).

3. RESULTADOS E DISCUSSÃO

Os resultados das estimativas do parâmetro da família de transformações Box-Cox com os respectivos intervalos de confiança dos modelos I, II e III foram, respectivamente, $0,4363 (0,3771 \leq \lambda \leq 0,5091)$, $0,4393 (0,3745 \leq \lambda \leq 0,5091)$ e $0,4454 (0,3770 \leq \lambda \leq 0,5091)$, verificando-se que os modelos não tiveram distribuição aproximadamente normal. Como os intervalos de confiança incluem o valor 0,5, adotou-se em todos os casos a transformação \sqrt{Y} .

A análise variográfica exploratória (Figura 3) mostrou indicação de dados estruturados espacialmente. Verificou-se também que os modelos ajustados com diferentes estruturas de co-variância ficaram muito próximos, sendo os modelos I e III com estrutura de co-variância Matérn com $k=1,0$ e $2,0$ quase coincidentes.

No Quadro 2 são apresentados os resultados das estimativas dos parâmetros $(\hat{\sigma}^2, \hat{\tau}^2, \hat{\phi})$ logaritmo da função de verossimilhança (Log), número de parâmetros do modelo (Npars) e valor do critério de Akaike (AIC).

Os resultados indicaram que o valor estimado para efeito-pepita (τ^2) em todos os modelos avaliados foram aproximadamente zero. Isso pode estar associado à alta homogeneidade de solo e relevo em que o ensaio foi instalado, associada a erros de medida desprezíveis. No caso das estimativas do patamar (τ^2), os valores quantificam a variabilidade do processo, independentemente da estrutura de co-variância e do número de parâmetros do modelo. Para os resultados do alcance (ϕ), os valores foram sempre menores nos modelos com estrutura de co-variância Matérn com $k=2,0$, seguido pelos valores com $k=1,0$ (Quadro 2).

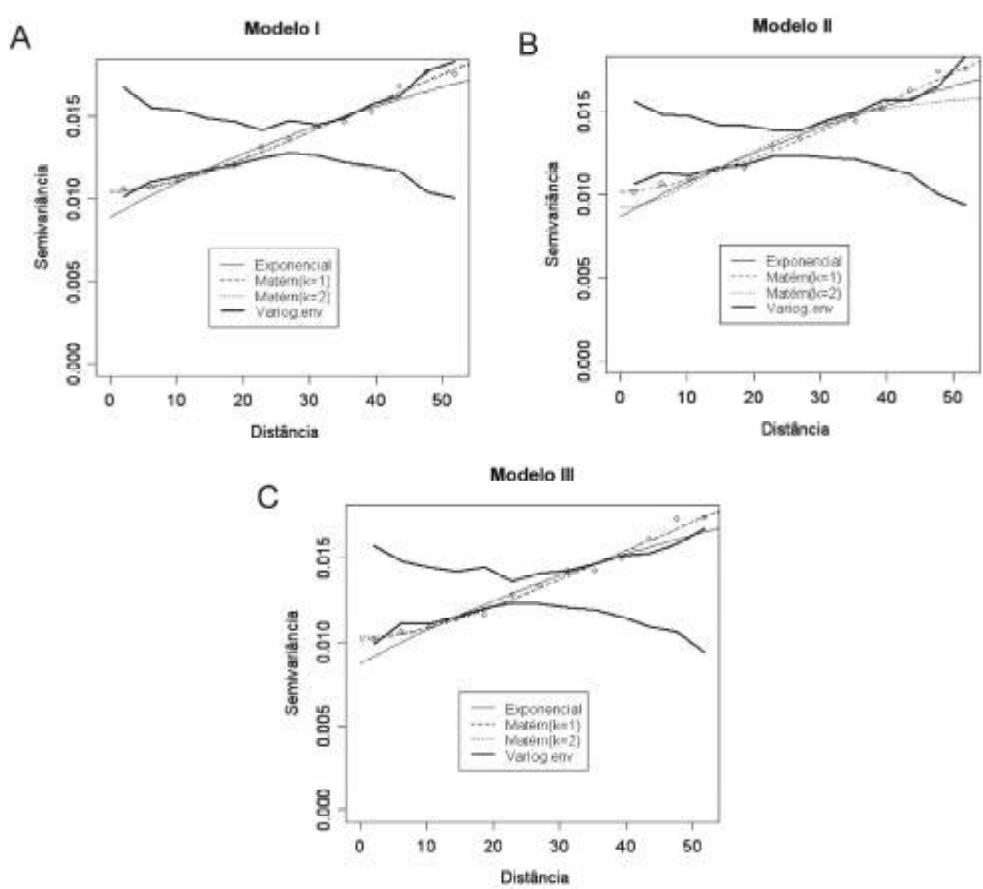


Figura 3 – Ajuste dos semivariogramas teóricos aos valores do semivariograma empírico e envelope simulado dos modelos A - I, B - II e C - III. Parâmetros ajustados por máxima verossimilhança considerando as estruturas de co-variação: Exponencial, Matérn ($k=1$) e Matérn ($k=2$).

Figure 3 – Adjustment of theoretical semivariograms to empiric semivariograms and simulated envelope for the models A - I, B - II and C - III. Adjusted parameters by maximum likelihood considering the covariance structure: Exponential, Matérn ($k=1$) and Matérn ($k=2$).

Quadro 2 – Estimativas dos parâmetros ($\hat{\sigma}^2, \hat{\tau}^2, \hat{\phi}$), logaritmo da função de verossimilhança (Log), número de parâmetros do modelo (Npars) e critério de akaike (AIC) dos modelos I, II e III

Table 2 – Parameters estimate ($\hat{\sigma}^2, \hat{\tau}^2, \hat{\phi}$), likelihood function logarithmic (Log), model's parameter number (Npars) and Akaike's criterion (AIC) for the models I, II and III

Modelo	Função de Correlação	\hat{s}^2	$\hat{\tau}^2$	$\hat{\phi}$	Log	Npars	AIC
I	Exponencial	0,0007	0,0128	14,7680	248,8784	13	- 471,7569
	Matérn ($k=1$)	0,0007	0,0128	10,8015	249,1508	13	- 472,3016
	Matérn ($k=2$)	0,0008	0,0128	7,4183	249,4095	13	- 472,8189
II	Exponencial	0,0009	0,0124	15,7573	267,1691	14	- 506,3381
	Matérn ($k=1$)	0,0009	0,0125	11,4108	267,4671	14	- 506,9342
	Matérn ($k=2$)	0,0009	0,0125	7,8339	267,7342	14	- 507,4685
III	Exponencial	0,0008	0,0124	14,7581	267,7417	15	- 505,4835
	Matérn ($k=1$)	0,0008	0,0125	10,8284	268,0239	15	- 506,0477
	Matérn ($k=2$)	0,0008	0,0125	7,5082	268,2830	15	- 506,5661

Verificou-se superioridade do modelo II com função de co-variância Matérn com $k = 2,0$, pois apresentou menor valor de AIC (Quadro 2). A comparação individual dos modelos com correlação de Matérn mostrou valores próximos, sendo mais adequados do que os modelos com correlação exponencial ($k = 0,5$), ressaltando-se que o critério de Akaike também leva em consideração o número de parâmetros de cada modelo.

Para avaliar a evidência da presença de dependência espacial em cada um dos modelos, utilizou-se primeiramente, uma forma não-paramétrica, através de envelopes de simulação. Observando os gráficos ajustados (Figura 3), os semivariogramas encontram-se no limite do intervalo do envelope simulado, apresentando alguns pontos fora, sugerindo a presença de dependência espacial em todos os modelos avaliados. Mas para confirmar a existência dessa dependência, aplicou-se o teste da razão de verossimilhança (Quadro 3). Os resultados quando comparados com os modelos com e sem o componente espacial, foram inferiores ao valor crítico de $\chi^2(2;0,05)=5,99$, considerando-se que a inclusão do efeito espacial adiciona dois parâmetros (σ^2, ϕ) ao modelo. Assim, no nível de significância e 5%, conclui-se que os modelos sem o componente espacial são mais adequados aos dados. Os valores de AIC nos modelos I, II e III sem o componente espacial foram -473,8708, -507,0241 e -507,0187, respectivamente, em que o modelo II obteve o menor valor de AIC.

Considerando os modelos com e sem a estrutura de dependência espacial, observou-se que os modelos II e III, que contêm as co-variáveis f_1 (planta morta) e f_2 (tempo), foram superiores aos modelos que se excluíram as plantas mortas. Assim os resultados apresentados até o presente momento foram contrários à proposta de Gomez e De Datta (1972), que sugeriram eliminar da análise experimental todas as plantas vizinhas às falhas ou mortes. Verificou-se que nesses casos as co-variáveis interagiram positivamente com a variável resposta, evitando-se que dados coletados sejam desperdiçados.

A incorporação da estrutura de dependência espacial aos modelos observacionais recuperou a capacidade de fazer inferências válidas na ausência de aleatorização, permitindo contornar problemas operacionais e, assim, garantindo que os dados possam ser submetidos a uma análise clássica. É importante ressaltar que esse processo não substitui a recomendação da aleatorização, quando esta é possível. Entretanto, fornece uma alternativa da análise para condições experimentais que impõe restrições aos procedimentos de aleatorização. Nessa abordagem e na ausência ou restrição de aleatorização, a inferência e resultados obtidos são fortemente baseados na correção do modelo adotado. Dessa forma, ferramentas exploratórias para avaliar a adequação de pressupostos, como as descritas aqui, são fundamentais para guiar a escolha adequada da forma do modelo a ser utilizado.

Quadro 3 – Logaritmo da função de verossimilhança (Log), número de parâmetros (Npars) dos modelos com e sem o componente espacial e valores obtidos no teste da razão de verossimilhança (LR) dos modelos I, II e III

Table 3 – Likelihood function logarithmic (Log), parameter number (Npars) for the models with and without the spatial component, and values for the likelihood ratio test (LR) for the models I, II and III

Modelo	Sem componente Espacial		Com Componente Espacial			LR
	Npars	Log	Npars	Função de correlação	Log	
I	11	247,9354	13	Exponencial	248,8784	1,8860 ^{ns}
	11	247,9354	13	Matérn ($k=1$)	249,1508	2,4308 ^{ns}
	11	247,9354	13	Matérn ($k=2$)	249,4095	2,9482 ^{ns}
II	12	265,5121	14	Exponencial	267,1691	3,3140 ^{ns}
	12	265,5121	14	Matérn ($k=1$)	267,4671	3,9100 ^{ns}
	12	265,5121	14	Matérn ($k=2$)	267,7342	4,4442 ^{ns}
III	13	266,5094	15	Exponencial	267,7417	2,4646 ^{ns}
	13	266,5094	15	Matérn ($k=1$)	268,0239	3,0290 ^{ns}
	13	266,5094	15	Matérn ($k=2$)	268,2830	3,5472 ^{ns}

4. CONLCUSÃO

Considerando as condições que o estudo foi desenvolvido, verificou-se que os métodos geoestatísticos foram adequados para verificar a existência de dependência espacial no delineamento sistemático tipo leque. A inserção da co-variável planta morta, para baixos índices de mortalidade, possibilitou a utilização dos dados de plantas vizinhas às falhas ou mortes.

5. REFERÊNCIAS

- AKAIKE, H. Information measures and model selection. **International Statistical Institute**, v.44, p.277-291, 1983.
- BLEASDALE, J. K. A. Systematic designs for spacing experiments. **Experimental Agriculture**, n.3, p.73-85, 1966.
- BOX, G. E. P.; COX, D. R. An Analysis of transformations. **Journal of the Royal Statistical Society**, v.26, n.2, p.211-252, 1964.
- CHRISTENSEN, O. F.; DIGGLE, P. J.; RIBEIRO JR, P. J. Analysing positive-valued spatial data: the transformed gaussian model. In: MONESTIEZ, P.; ALLARD, D.; FROIDEVAUX, R. (Eds.) **geoENV - Geostatistics for environmental applications**. Amsterdam, Kluwer, 2001. p.287-298.
- DIGGLE, P. J.; RIBEIRO JR, P. J. **Model based geostatistics**. New York: Springer, 2007. 230p.
- FREEMAN, G. H. The use of systematic design for a spacing trial with a tropical tree crop. **Biometrics**, n.20, p.200-203, 1964.
- GOMEZ, K. A.; DE DATTA, S. K. Missing hills in rice experimentation plots. **Agronomy Journal**, n.64, p.163-164, 1972.
- HUXLEY, P.A. Systematic designs for field experimentation with multipurpose trees. **Agroforestry Systems**, n.3, p.197-207, 1985.
- IMADA, M. et al. Optimum planting density for Japanese oak (*Quercus mongolica* var.*grosseserrata*) based on spacing experiment with systematic design. **Journal of Forest Research**, n.2, v.2, p.89-93, 1997.
- JURNEL, A. G.; HUIJBREGTS, C. J. **Mining geostatistics**. New York: Academic Press, 1978.
- MANDALLAZ, D. **A unified approach to sampling theory for forest inventory based on infinite population and superpopulation models**. 1991. 266f. Thesis (Doctoral Thesis and Habilitation) - Chair of Forest Inventory and Planning, Swiss Federal Institute of Technology, Zurich, 1991.
- MATHERON, G. Principles of geostatistics. **Economic Geology**, v.58, p.1246-1266, 1963.
- MCBRATNEY, A. B.; WEBSTER, R. Choosing functions for semi-variograms of soil properties and fitting them to sampling estimates. **Journal of Soil Science**, v.77, p.617-639, 1986.
- NELDER, J. A. New kinds of systematic designs for spacing experiments. **Biometrics**, n.18, p.283-307, 1962.
- PANETSOS, C. P. Selection of new poplar clones under various spacings. **Silvae Genetica**, n.29, p.130-135, 1980.
- R DEVELOPMENT CORE TEAM R. **A language and environment for statistical computing**. Vienna, R Foundation for Statistical Computing, 2005. URL <http://www.R-project.org>.
- RAO, C. R. **Linear statistical inference and its applications**. 2.ed. New York: John Wiley & Sons, 1973. p.417-420.
- RIBEIRO JÚNIOR, P. J. **Métodos geoestatísticos no estudo da variabilidade espacial de parâmetros do solo**. 1995. 99f. Dissertação (Mestrado) - Escola Superior de Agricultura "Luiz de Queiroz", Piracicaba, 1995.
- RIBEIRO JR., P. J.; DIGGLE, P. J. The geoR package functions for geostatistical data analysis: **R.News**, v.1, n.2, p.15-18, 2001.
- RONDON, E. V. Produção de biomassa e crescimento de árvores de *schizolobium mazonicum* (huber) ducke sob diferentes espaçamentos na região de mata. **Revista Árvore**, v.26, n.5, p.573-576, 2002.

SCHREUDER, H. T.; WILLIAMS, M. S. Design-based estimation of forest volume within a model-based sample selection framework. **Canadian Journal of Forest Research**, v.25, p.121-127, 1995.

SILVA, C. R. **Efeito do espaçamento e arranjo de plantio na produtividade e uniformidade de clones de Eucalyptus na região nordeste do Estado de São Paulo**. 2005. 50f. Dissertação (Mestrado) - Escola Superior de Agricultura “Luiz de Queiroz”, Piracicaba, 2005.

SILVA, J. A. A.; BAILEY, R. L. Uso de transformações normalizadoras no ajuste de modelos volumétricos. **Revista Árvore**, v.15, n.12, p.199-206, 1991.

STAPE, J. L. **Utilização de delineamento sistemático tipo leque no estudo de espaçamentos florestais**. 1995. 86f. Dissertação (Mestrado) - Escola Superior de Agricultura “Luiz de Queiroz”, Piracicaba, 1995.