

GEOESTATÍSTICA APLICADA AO INVENTÁRIO FLORESTAL

JOSÉ MARCIO DE MELLO

Tese apresentada à Escola Superior de Agricultura Luiz de Queiroz”, Universidade de São Paulo, para obtenção do título de Doutor em Recursos Florestais, com opção em Silvicultura e Manejo Florestal.

PIRACICABA
Estado de São Paulo – Brasil
Setembro - 2004

GEOESTATÍSTICA APLICADA AO INVENTÁRIO FLORESTAL

JOSÉ MARCIO DE MELLO
Engenheiro Florestal

Orientador: Prof. Dr. **JOÃO LUÍS FERREIRA BATISTA**

Tese apresentada à Escola Superior de Agricultura Luiz de Queiroz”, Universidade de São Paulo, para obtenção do título de Doutor em Recursos Florestais, com opção em Silvicultura e Manejo Florestal.

PIRACICABA
Estado de São Paulo – Brasil
Setembro - 2004

Dados Internacionais de Catalogação na Publicação (CIP)
DIVISÃO DE BIBLIOTECA E DOCUMENTAÇÃO - ESALQ/USP

Mello, José Marcio de
Geoestatística aplicada ao inventário florestal / José Marcio de Mello. -- Piracicaba,
2004.
111 p.

Tese (doutorado) - - Escola Superior de Agricultura Luiz de Queiroz, 2004.
Bibliografia.

1. Geoestatística 2 . Inventário florestal I. Título

CDD 634.9285

"Permitida a cópia total ou parcial deste documento, desde que citada a fonte – O autor"

A Deus,
Autor da vida.

A todos os brasileiros que, através dos impostos,
ajudam a construir o conhecimento que promove,
ainda timidamente, melhorias na qualidade de vida
de todos nós.

A amada esposa, **Claudinelli**,
que pelo dom de gerar
vidas, proporcionou-me a beleza
de possuir dois lindos filhos,
Clara e Cláudio ;

À mamãe **Cidinha**, ao meu querido
irmão **Carlos Rogério** e sua esposa **Olívia**,
à minha sogra D. **Vilma**, que
juntamente com as tias **Vilgair** e
Valnete, nos apoiam constantemente.

Ao grande amigo, conselheiro, mestre e
colega, **José Roberto Soares Scolforo ;**

DEDICO.

AGRADECIMENTOS

À Universidade Federal de Lavras (UFLA), através do Departamento de Ciência Florestal (DCF), pela oportunidade desta qualificação.

À Escola Superior de Agricultura “Luiz de Queiroz”, da Universidade de São Paulo (ESALQ/USP), através do Departamento de Ciência Florestal (LCF), pela acolhida e apoio ao longo do curso.

Aos professores Dr. José Roberto Soares Scolforo, Dr. Antônio Donizette de Oliveira, Dr. Natalino Calegário, e Dr. José Luiz Pereira de Rezende e Fausto Acerbi, por assumirem as minhas tarefas no DCF/UFLA, possibilitando o meu afastamento.

Ao professor Dr. João Luis Ferreira Batista, pela orientação, amizade, compreensão e confiança depositada no meu trabalho.

Aos professores Dr. Marcelo Silva de Oliveira e Dr. Paulo Justiniano Ribeiro Júnior, pela co-orientação ativa, pela paciência, zelo científico e amizade.

Aos amigos Honório Kanegae Júnior e Sebastião Oswaldo Ferreira, pelo apoio, pelas valiosas sugestões, pelos atendimentos sempre solícitos e pela grande amizade.

Aos amigos professor Dr. Hilton Thadeu pela acolhida no curso, e Jefferson, do laboratório de Métodos Quantitativos, pelo apoio técnico nos recursos computacionais.

À Votorantim Celulose e Papel (VCP), através da pessoa do Honório Kanegae Júnior, pelo apoio financeiro, técnico e estrutural para realização deste trabalho.

A todos os amigos contemporâneos do curso, pela grata e prazerosa convivência. E a todos que colaboraram com o desenvolvimento deste trabalho, o meu muito obrigado.

Enfim, a Deus, por proporcionar-me tudo isso e permitir-me oferecer mais uma contribuição para o mundo científico.

SUMÁRIO

| | Página |
|--|--------|
| RESUMO..... | ix |
| SUMMARY..... | xi |
| 1 INTRODUÇÃO..... | 1 |
| 2 REVISÃO DE LITERATURA..... | 3 |
| 2.1 Teoria de Amostragem Clássica..... | 3 |
| 2.2 Teoria geoestatística..... | 4 |
| 2.2.1 Semivariograma..... | 6 |
| 2.2.2 Modelos de semivariograma e métodos de ajuste..... | 7 |
| 2.2.3 Inferência espacial – Krigagem..... | 9 |
| 2.3 Situações de dados dendrométricos com dependência espacial..... | 10 |
| 3 ESTUDO DA DEPENDÊNCIA ESPACIAL DE CARACTERÍSTICAS DENDROMÉTRICAS PARA <i>Eucalyptus grandis</i> | 13 |
| Resumo..... | 13 |
| Summary..... | 14 |
| 3.1 Introdução | 15 |
| 3.2 Material e Métodos..... | 17 |
| 3.2.1 Descrição da área e coleta de dados..... | 17 |
| 3.2.2 Semivariograma..... | 20 |
| 3.3 Resultados e Discussão..... | 23 |
| 3.3.1 Análise exploratória de dados..... | 23 |
| 3.3.2 Avaliação da continuidade espacial..... | 24 |
| 3.4 Conclusões..... | 33 |
| 4 AJUSTE E SELEÇÃO DE MODELOS ESPACIAIS DE SEMIVARIOGRAMA VISANDO À ESTIMATIVA VOLUMÉTRICA DE <i>Eucalyptus grandis</i> | 34 |
| Resumo..... | 34 |
| Summary..... | 35 |
| 4.1 Introdução..... | 35 |
| 4.2 Material e Métodos..... | 37 |
| 4.2.1 Descrição da área e obtenção dos dados..... | 37 |
| 4.2.2 Semivariograma..... | 38 |
| 4.2.3 Métodos de ajuste..... | 41 |
| 4.2.3.1 A sentimento..... | 41 |

| | |
|--|-----|
| 4.2.3.2 Métodos dos Quadrados Mínimos Ordinários e Quadrados Mínimos Ponderados..... | 42 |
| 4.2.3.3 Método da Máxima Verossimilhança..... | 42 |
| 4.2.3.4 Seleção de Modelos..... | 44 |
| 4.3 Resultados e Discussão..... | 46 |
| 4.3.1 Análise Exploratória de Dados..... | 46 |
| 4.3.2 Avaliação do Critério de Informação de Akaike (AIC)..... | 49 |
| 4.3.3 Avaliação dos modelos e métodos de ajuste através das técnicas de validação | 54 |
| 4.4 Conclusão..... | 56 |
| 5 MÉDIA ESPACIAL PARA VOLUME DE <i>EUCALYPTUS GRANDIS</i> ESTIMADA POR GEOESTATÍSTICA BASEADA EM MODELO..... | 57 |
| Resumo..... | 57 |
| Summary..... | 58 |
| 5.1 Introdução..... | 59 |
| 5.2 Material e Métodos | 63 |
| 5.2.1 Descrição da área e obtenção dos dados..... | 63 |
| 5.2.2 Modelo Geoestatístico..... | 64 |
| 5.2.3 Simulação Estocástica..... | 66 |
| 5.2.4 Processamento dos dados simulados..... | 68 |
| 5.3 Resultados e Discussão..... | 70 |
| 5.3.1 Simulação com base em população estocástica..... | 70 |
| 5.3.2 Simulação com base em população fixa..... | 74 |
| 5.3.3 Simulação estocástica no contexto do inventário florestal..... | 76 |
| 5.3.4 Correção empírica para o intervalo de confiança do estimador geoestatístico | 79 |
| 5.4 Conclusões..... | 80 |
| 6 USO DO ESTIMADOR GEOESTATÍSTICO PARA PREDIÇÃO VOLUMÉTRICA POR TALHÃO..... | 82 |
| Resumo..... | 82 |
| Summary..... | 83 |
| 6.1 Introdução..... | 84 |
| 6.2 Material e Métodos..... | 86 |
| 6.2.1 Descrição geral..... | 86 |
| 6.2.2 Inventário pré-corte..... | 87 |
| 6.2.3 Estudo variográfico..... | 87 |
| 6.2.4 Modelo geoestatístico..... | 89 |
| 6.2.5 Krigagem com simulação..... | 90 |
| 6.2.6 Predição espacial ou krigagem de bloco..... | 91 |
| 6.3 Resultados e Discussão..... | 91 |
| 6.3.1 Análise variográfica..... | 91 |
| 6.3.2 Predição espacial..... | 93 |
| 6.4 Conclusões..... | 101 |
| 7 CONCLUSÕES GERAIS..... | 102 |
| REFERÊNCIAS BIBLIOGRÁFICAS..... | 104 |

GEOESTATÍSTICA APLICADA AO INVENTÁRIO FLORESTAL

Autor: JOSÉ MARCIO DE MELLO

Orientador: Prof. Dr. JOÃO LUÍS FERREIRA BATISTA

RESUMO

Este trabalho teve como objetivo geral avaliar o uso da geoestatística aplicada ao inventário florestal. Especificamente avaliaram-se: a estrutura de continuidade espacial de quatro características dendrométricas, os métodos de ajuste e seleção de modelos da função de semivariância, o comportamento dos intervalos de confiança clássico e geoestatístico e o potencial da krigagem de bloco para a predição volumétrica por talhão. Os dados para realização do presente estudo foram obtidos num povoamento de 987 hectares de *Eucalyptus grandis*, com sete anos de idade, localizados no município de Itapetininga, São Paulo, pertencente à Votorantin Celulose e Papel (VCP). Pelos semivariogramas anisotrópico e isotrópico avaliou-se a estrutura de continuidade espacial das características: volume, diâmetro médio quadrático, área basal e altura média dominante. Dois modelos da função de semivariância foram ajustados para o volume, por meio de quatro métodos de ajuste e a seleção do melhor modelo foi efetuada a partir do critério de Akaike (AIC) e por meio das técnicas de validação cruzada e preditiva. Quanto às simulações estocásticas com o modelo geoestatístico, verificou-se o desempenho da cobertura do intervalo de confiança clássico e geoestatístico em relação à cobertura nominal. A predição volumétrica por talhão foi obtida da krigagem com simulação em cada um dos talhões do povoamento. Verificou-se que as características dendrométricas mensuradas no inventário florestal apresentaram-se estruturadas espacialmente. Este resultado sugeriu que no

processamento do inventário florestal deve-se considerar a componente espacial, ou seja, as parcelas não devem ser tratadas de forma independente. Sugere-se o uso conjugado do critério de Akaike e da validação cruzada para seleção de modelos da função de semivariância. Detectou-se que os intervalos de confiança clássico e geoestatístico apresentaram cobertura inferior à cobertura nominal, para uma população espacialmente contínua. Porém, a cobertura do intervalo de confiança do estimador clássico foi bem inferior à cobertura nominal, mostrando viés substancial na estimativa da variância da média quando há estrutura de continuidade espacial. Para populações fixas o estimador clássico proporcionou cobertura observada igual à cobertura nominal. Este resultado comprova a validade da aleatorização para uso do Teorema Central do Limite, na definição do intervalo de confiança, independente da estrutura de continuidade espacial. Na presença de continuidade espacial, o uso do estimador geoestatístico (krigagem de bloco), para estimativa do volume por talhão, é preferido por duas razões: proporciona estimativa da precisão sem viés e é de baixo custo quando comparado ao estimador clássico.

GEOSTATISTIC APPLICATION TO FORESTRY INVENTORY

Author: JOSÉ MARCIO DE MELLO

Adviser: Prof. Dr. JOÃO LUÍS FERREIRA BATISTA

SUMMARY

This research evaluated the use of applied geostatistic forestry inventory. Specifically, it evaluated: the continuity of spatial structure of four dendrometric characteristics, the adjustment methods and model selection of the semi variance function, the behavior of classic and geostatistic confidence intervals and the potential of the block krigage for the stand volumetric prediction. Data were obtained in a 987 ha of *Eucalyptus grandis* forest plantation aged 7 years, located in Itapetininga county- São Paulo State, belonging to Votorantim Pulp and Paper (VCP). The spatial continuity structure of volume, quadratic average diameter, basal area and dominant average height was evaluated through the anisotropic and isotropic semi variogrammes. Two semi variance function models were adjusted for volume, through four adjustment methods. The selection of the best model was made starting from the Akaike information criterion (AIC) and through the crossed validation and prediction techniques. Departing from stochastic simulations, using geostatistic model the performance of the classic and geostatistic confidence interval covering was verified, in relation to the nominal covering. The stands volumetric prediction was obtained from the krigage with simulation in each plantation stand. It was verified that the dendrometric characteristics measured in the forestry inventory presented spatially structured. This result suggested that, in the forestry inventory processing, spatial component should be considered, i.e.,

the plots should not be treated separately. It seems that the AIC and the cross validation for semi variance function model selection must be associated. It was detected that both classic and geostatistic confidence intervals presented lower covering than that of the nominal covering, for a spatial continuous population. However, the confidence covering interval of the classic estimator was much lower than the nominal covering, showing substantial bias in the estimate of the average variance, when there is structure of spatial continuity. For fixed populations, the classic estimator provided observed covering equal to the nominal covering. Thus, proving that the validity of the randomization for use of the Central Limit Theorem, in the definition of the confidence interval, independently of the structure of space continuity. The use of the geostatistic estimator (block krigage) for estimating stand volume, in the presence of spatial continuity, is preferred because it provides unbiased and precise estimate at lower cost as compared to the classic estimator.

1 INTRODUÇÃO

A dinâmica de vida das florestas sofrem mudanças rapidamente devido a fatores que mudam constantemente no espaço e no tempo. Fatores inerentes ao solo, ao clima e a própria planta, influenciam acentuadamente esta dinâmica, além das ações antrópicas (adubação, melhoramento genético, dentre outros). Toda esta dinâmica deve ser acompanhada por meio de técnicas de amostragem, que sejam eficientes na representação do estado atual e temporal da floresta, a fim de gerar informações precisas e com menor custo possível, para subsidiar as ações do manejo e planejamento florestal.

A preocupação em amostrar adequadamente florestas não é assunto novo. Desde o início da ciência florestal pesquisadores se preocupam com métodos de amostragem eficiente e de baixo custo. O inventário florestal tem por objetivo aplicar e avaliar métodos de amostragem que geram resultados mais precisos. O desejável para todo empreendimento florestal é obter informações precisas e a baixo custo sobre a produção florestal. A redução do custo no inventário florestal está alicerçada basicamente na redução da intensidade amostral. Isto nos remete a seguinte pergunta: como reduzir custos sem perder precisão? Um dos caminhos possíveis é o uso de um método estatístico que explore adequadamente as relações existentes entre as unidades amostrais. Os métodos tradicionais de estatística usados para análise dos inventários florestais, utilizam-se de uma medida central (média) e uma de dispersão (variância) para descrever um determinado fenômeno, sem levar em consideração, as possíveis correlações entre observações vizinhas. Portanto, eles não exploram suficientemente as relações que possam existir entre as unidades amostrais.

Mediante o exposto, o objetivo geral desta pesquisa foi avaliar o potencial de um método estatístico que seja capaz de explorar as relações espaciais existentes entre as amostras. No item 3, intitulado “*Estudo da dependência espacial de características dendrométricas para Eucalyptus grandis*”, teve como objetivo avaliar a estrutura de continuidade espacial de quatro características dendrométricas. Estas características foram obtidas a partir das informações de dois inventários florestais, cujas parcelas foram lançadas conforme dois procedimentos de amostragem diferentes.

No item 4, intitulado “*Ajuste e seleção de modelos espaciais de semivariograma visando à estimativa volumétrica de Eucalyptus grandis*”, teve como objetivo ajustar e selecionar modelos da função de semivariância para estimativa volumétrica de *Eucalyptus grandis*. Especificamente pretendeu-se, a partir da validação cruzada e do critério de AIC, avaliar e selecionar o desempenho de quatro métodos de ajuste, aplicados a dois modelos espaciais.

No item 5, “*Média espacial para volume de Eucalyptus grandis estimada por Geoestatística baseada em modelo*”, procurou-se desenvolver um método para análise de inventário florestal, a partir da fórmula que gera o intervalo de confiança para observações independentes (“Estatística Clássica”), mas apresentando melhorias na estimativa da média, da variância da média e na estatística t de Student, considerando a estrutura de dependência espacial da característica volume.

E, no item 6, “*Uso do estimador Geoestatístico para predição volumétrica por talhão*”, comparou-se a estimativa volumétrica por talhão, da Geoestatística baseada em modelo, com a estimativa usando o estimador clássico através do inventário pré-corte.

2 REVISÃO DE LITERATURA

2.1 Teoria de Amostragem Clássica

O inventário florestal é uma atividade que faz uso dos fundamentos da teoria da amostragem, a fim de determinar ou estimar uma dada característica, seja quantitativa ou qualitativa, de uma amostra (Péllico Netto & Brena,1997). As unidades amostrais são selecionadas conforme o procedimento de um delineamento amostral. Existem diversos estudos na área florestal visando avaliar o desempenho de diferentes procedimentos de amostragem (Higuchi,1986/87; Korning et al. 1991; Machado, 1988; Mello & Scolforo 2000). Estes estudos foram embasados na comparação entre precisão e custo dos procedimentos de amostragem. Os métodos tradicionais de estatística usados para análise desses procedimentos foram construídos com base na Teoria Clássica de Amostragem (Cochran, 1977). Por esta teoria, o valor médio de uma dada característica, em uma região amostrada, é igual ao valor da esperança desta característica em qualquer ponto dentro da região, com um erro de estimativa correspondente à variância dos dados amostrados (Trangmar et al. 1987). Um dos princípios que a Teoria de Amostragem Clássica se baseia é a *casualização*. Segundo Yates citado por Duarte (2000), cabe a este princípio neutralizar os efeitos da correlação espacial. Assim, as posições das unidades amostrais no campo podem ser ignoradas nesta abordagem. A pressuposição desses métodos é que as variações numa dada característica, de um local para outro, são aleatórias (independentes). A Teoria de Amostragem Clássica consiste em dois componentes, conforme a equação apresentada a seguir

$$Z(x_i) = M(x) + e_i$$

onde $Z(x_i)$ é a variável avaliada, $M(x)$ é uma função determinística que descreve a componente estrutural de Z e e_i é o erro aleatório. A parte e_i é tão importante quanto $M(x)$, e a análise destes resíduos é fundamental em todo trabalho estatístico. O que se deseja é que a parte e_i não contenha nenhuma “suavidade” ou comportamento estruturado. É fundamental que ela seja aleatória. É importante ressaltar que a Teoria de Amostragem Clássica desperdiça a estrutura de dependência espacial, quando ela está presente na característica avaliada. Deve-se observar que a formulação desta teoria discerne somente o tamanho da variabilidade, através do desvio padrão, e não discerne a estrutura da variabilidade.

O desenvolvimento teórico da amostragem clássica foi efetuado com base no conceito de populações fixas, ou seja, aquelas cujo conjunto de unidades amostrais é finito e que, para cada unidade amostral, o valor da variável avaliada é fixada (Thompson & Seber, 1996 e Hoef, 2002). Os procedimentos de amostragem mais utilizados nos inventários florestais são: casual simples, casual estratificado e sistemático. Segundo Matérn (1960) e Cressie (1993), estes procedimentos também são considerados na amostragem espacial. Os estimadores dos procedimentos de amostragem casualizados foram construídos com base nas probabilidades geradas pela aleatorização dos mesmos sobre a população fixa (Brus & Gruijter, 1997). A descrição detalhada de cada um destes procedimentos pode ser encontrada em Cochran (1977), Thompson (1992), Péllico Netto & Brena (1997) e Scolforo & Mello (1997), dentre outros.

2.2 Teoria geoestatística

Trabalhando com dados de mineração de ouro, Krige (1951), concluiu que, as variâncias obtidas através da abordagem clássica de amostragem, não faziam sentido se não considerasse as distâncias entre as amostras. Foi então, que Matheron (1963),

baseado nestas observações, desenvolveu a Teoria das Variáveis Regionalizadas. Ela foi definida como uma função espacial numérica, variando de um local para outro, apresentando continuidade aparente e cuja variação não pode ser representada por uma simples função matemática. A aplicação dessa teoria a problemas voltados para a geologia e mineração recebeu o nome de geoestatística.

A teoria das variáveis regionalizadas pressupõe que cada dado $z(x_i)$ é modelado como uma variável aleatória que pode ser expressa pela soma de três componentes: um componente estrutural, associado a um valor médio constante ou a uma tendência constante; uma componente aleatória, espacialmente correlacionada; e um erro aleatório.

Se x representa uma posição em uma, duas ou três dimensões, então o valor da variável Z , em x , é dado por:

$$Z(x_i) = M(x) + e'_x + e_i$$

onde $M(x)$ é uma função determinística que descreve a componente estrutural Z em x ; e'_x é um termo estocástico, que varia localmente e depende espacialmente de $M(x)$; e e_i é um ruído aleatório não correlacionado, com distribuição normal média zero e variância σ^2 . A geoestatística atua no termo estocástico e'_x . A Teoria das Variáveis Regionalizadas, outro nome dado a geoestatística, pode ser vista como um grande ramo da estatística espacial (Cressie, 1993), que estuda a relação de uma variável regionalizada com ela mesma numa outra posição.

Conforme Oliveira (1991), há essencialmente duas propostas estatísticas para modelar populações do tipo $\{Z(x) : x \in \mathfrak{R} \subset \mathbb{R}^2\}$:

i. considere $\{Z(x) : x \in \mathfrak{R} \subset \mathbb{R}^2\}$ como uma população no sentido da teoria clássica da amostragem, ou seja, as unidades amostrais são independentes. Desta forma, não toma a vantagem de considerar a estrutura de dependência que possivelmente possa existir naquela população;

ii. considere $\{Z(x): x \in \mathfrak{R} \subset \mathbb{R}^2\}$ como uma amostra retirada de uma superpopulação, que é uma “população” de populações do tipo $\{Z(x): x \in \mathfrak{R} \subset \mathbb{R}^2\}$. Nesta situação, a geoestatística necessita de supostos para ser válida (a hipótese intrínseca) porém, ela considera a correlação espacial entre os dados.

Segundo Vieira (2000) a hipótese intrínseca requer apenas a estacionaridade do variograma. Esta é a principal hipótese sobre a qual a geoestatística está alicerçada. Uma variável regionalizada é intrínseca quando:

- existir a esperança matemática para a variável regionalizada e ela não depende da posição x ;

$$E[Z(x)] = m, \quad \forall x; \quad (01)$$

- para todo vetor h , a variância da diferença $[V(x)-V(x+h)]$ é finita e independe da posição x , dependendo apenas do vetor de h .

$$VAR[Z(x) - Z(x+h)] = E[Z(x) - Z(x+h)]^2 = 2g(h) \quad (02)$$

2.2.1 Semivariograma

O semivariograma ou simplesmente variograma é o método geoestatístico para diagnosticar a presença da correlação entre as unidades amostradas (Duarte, 2000). Segundo Ribeiro Júnior. (1995), os semivariogramas são preferidos para caracterizar a estrutura de continuidade espacial da característica avaliada, por exigirem hipóteses de estacionaridade menos restritivas (Hipótese Intrínseca). O semivariograma representa uma função de semivariâncias em relação às respectivas distâncias. A semivariância é definida como a metade da variância de diferenças entre observações de uma variável aleatória Z , separadas por uma distância h . Assim, valores baixos indicam menor variabilidade (maior similaridade). Segundo Vieira (2000), dentre os vários estimadores de semivariâncias, o mais utilizado é o método dos momentos.

Segundo Journel e Huijbregts (1978), a função variograma, denominada de $2g(h)$, é a esperança matemática do quadrado da diferença entre pares de pontos separados por uma distância (h)

$$2g(h) = E\{[Z(x) - Z(x+h)]^2\} \quad (03)$$

A metade da função variograma é denominada de função semivariograma. O estimador da semivariância $[\hat{g}(h)]$ é igual a média aritmética das diferenças ao quadrado entre pares de valores experimentais, em todos os pontos separados pela distância h .

$$\hat{g}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x) - Z(x+h)]^2 \quad (04)$$

onde $[\hat{g}(h)]$ é a semivariância estimada para cada distância (h), $N(h)$ é o número de pares de pontos separados por uma distância h , $Z(x)$ é o valor da variável regionalizada no ponto x e $Z(x+h)$ é o valor no ponto $x+h$. A função semivariograma permite gerar o gráfico da semivariância em função da distância (h), denominado de semivariograma experimental, o qual permite interpretar a continuidade espacial da variável regionalizada.

2.2.2 Modelos de semivariograma e métodos de ajuste

Existem vários modelos de ajuste aplicáveis a diferentes fenômenos com continuidade espacial. Estes modelos, chamados de modelos teóricos, devem fornecer soluções estáveis para o estimador linear – krigagem. Isto quer dizer que as covariâncias têm de ser definidas positivamente, ou seja, a matriz de covariância utilizada na krigagem pode ser invertida. A condição de positividade do modelo, limita o conjunto de funções usadas na modelagem do semivariograma experimental. Os principais modelos de correlação são: Gaussiano, Esférico e a família de funções de Matérn. A função

Matérn com kappa igual a 0,5 é a função exponencial. A seguir serão apresentadas as expressões destes modelos de correlação:

$$\text{- Gaussiano} \quad r(h) = \exp\left[\left(-\frac{h}{\mathbf{f}}\right)^2\right] \quad (05)$$

$$\text{- Esférico} \quad r(h) = \begin{cases} 1 - 1,5\left(\frac{h}{\mathbf{f}}\right) + 0,5\left(\frac{h}{\mathbf{f}}\right)^3, & \text{se } h < 0 \\ 0 & \text{se } h > 0 \end{cases} \quad (06)$$

$$\text{- Exponencial} \quad r(h) = \exp\left(-\frac{h}{\mathbf{f}}\right) \quad (07)$$

$$\text{- Matérn} \quad r(h) = \left\{2^{k-1} \Gamma(K)\right\}^{-1} \left(\frac{h}{\mathbf{f}}\right)^k K_k\left(\frac{h}{\mathbf{f}}\right) \quad (08)$$

onde:

h – distância (m)

ϕ - parâmetro alcance

$\Gamma(k)$ – função Gama

$K_{(k)}$ – função Bessel de ordem K

Através do semivariograma experimental, o pesquisador é capaz de definir o modelo que melhor descreve o comportamento dos dados no espaço (Journel & Huijbregts, 1978). Em seguida, sua preocupação se volta para o ajuste da função matemática ao semivariograma experimental ou aos dados. Segundo Mc Bratney & 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, em 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 (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 & Ribeiro Júnior (2000).

Após a escolha do modelo e a estimativa dos parâmetros, a etapa seguinte é estudar o melhor modelo de semivariograma ajustado. Conforme Mc Bratney & 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 & Wendroth (2003), a validação é um meio efetivo para avaliar a qualidade das estimativas de Krigagem, bem como para selecionar modelos de semivariograma (Mello et al 2003).

2.2.3 Inferência espacial – Krigagem

O estimador espacial denominado de Krigagem, tem como base os dados amostrados da variável regionalizada e as propriedades estruturais do semivariograma obtido a partir destes dados. Krigagem é um método de inferência espacial, o qual estima dados em pontos não amostrados a partir de pontos amostrados, considerando a estrutura de dependência espacial do fenômeno. Este método foi desenvolvido por Matheron na década de 60, e recebeu o nome de Krigagem em homenagem ao engenheiro de minas sul-africano Daniel G. Krige, que primeiro desenvolveu e implementou esta fórmula de inferência em 1951. Conforme descrito por Isaak e Serivastava (1989), existem vários tipos de Krigagem: Krigagem pontual, ordinária, universal, de blocos e outras. Krigagem de bloco envolve estimativas de valores da variável regionalizada para uma sub-área da área total. A krigagem de bloco é

apropriada (em relação a krigagem pontual), para situações em que a informação média é mais útil do que o valor exato no ponto (Burgess & Webster, 1980).

Este estimador foi utilizado pela primeira vez por Matheron (1963), e este nome foi dado em homenagem aos trabalhos pioneiros de dependência espacial na mineração, (Soares, 2000). Ele é obtido pela seguinte expressão:

$$\hat{z}(x_0) = \sum_{i=1}^n \lambda_i z(x_i) \quad (09)$$

onde λ_i são os pesos de krigagem definidos conforme o semivariograma e $z(x_i)$ são os pontos amostrados.

2.3 Situações de dados dendrométricos com dependência espacial

A preocupação com a estrutura espacial de características dendrométricas não é assunto novo. Osborne (1942) e Matérn (1960) apresentaram a função exponencial para caracterizar a continuidade espacial de volumes de madeira obtidos de amostras sistemáticas. Matérn (1960) avaliou o desempenho da amostragem sistemática com grade quadrada e casual simples. Usando a função exponencial para obtenção da correlação, o autor verificou que a grade sistemática quadrada foi superior à amostragem casual simples no controle da variação do volume de madeira. Para estimar a área coberta com floresta, Matérn (1960) avaliou dois procedimentos de amostragem: sistemático quadrado e sistemático desencontrado versus os procedimentos casualizados para uma área coberta com floresta. Segundo Matérn (1960), o procedimento sistemático, quadrado ou desencontrado, foi superior aos procedimentos acidentais ou aleatórios na estimativa da variância. Quenouille (1949), em estudos teóricos, verificou que a amostra sistemática desencontrada foi frequentemente superior à sistemática quadrada e à casual estratificada. A amostragem sistemática desencontrada é o procedimento que mais se identifica com aqueles frequentemente praticados pelas empresas de reflorestamento. Este procedimento é caracterizado por não apresentar

distâncias fixas entre parcelas, sendo as parcelas distribuídas de forma a cobrir toda a floresta.

Nos anos 80, o tema dependência espacial voltou a ser alvo de estudo na ciência florestal. Samra et al. (1989) procuraram modelar o crescimento de altura para *Melia azedarach* na Índia fazendo uso de técnicas da Geoestatística. A associação de técnicas geoestatísticas com sensoriamento remoto foram trabalhadas por Cochen et al. (1990) e Wulder et al. (1996). Estes autores procuraram caracterizar a estrutura de copa de coníferas a partir de imagens de satélite e parâmetros biofísicos, tais como índice de área foliar e área basal. Eles concluíram ter sido possível usar métodos geoestatísticos de krigagem para mapeamento da altura dominante e área basal na área estudada. Outras aplicações da Geoestatística na área florestal são, na estimativa de índices de sítio (Höck et al. 1998), estimativa do volume total de madeira num povoamento de *Pinus* sp. (Gunnarson et al. 1998) e na determinação do grau de dependência espacial para área basal de *Pinus ponderosa* (Biondi et al. 1994). Thompson (1992) mostrou a importância da krigagem em estudos ecológicos, utilizando informações de uma área amostrada para prever determinada característica em toda a região. A krigagem de bloco foi utilizada por Biondi et al. (1994) para mapear o diâmetro e a área basal de *Pinus ponderosa*.

Segundo Aubry & Debouzie (2000), o uso do estimador de variância da média da amostragem casual simples para a amostragem sistemática aumenta os problemas estatísticos, uma vez que na amostragem sistemática percebe-se melhor a dependência entre unidades amostrais (correlação espacial). A estimativa de \bar{z} da média populacional, tanto na ausência quanto na presença de correlação, não tem tendência (viés). Porém, o estimador clássico para o cálculo da variância da média em dados com correlação conduz a um viés substancial (Koch & Gillings, 1983). Segundo estes autores, na ausência de correlação, os dois estimadores proporcionam valores iguais de variância da média. A fórmula apresentada por Cochran (1977) e Thompson (1992) para estimativa da variância da média na amostragem sistemática também deixa implícita a presença de viés no cálculo da variância da média, ao considerar a correlação entre unidades amostrais:

$$s_{\bar{x}}^2 = \left[\frac{s^2}{n} \left(1 - \frac{n}{N} \right) \right] \cdot [1 + (n-1) \cdot r] \quad (10)$$

em que σ^2 é a variância populacional; n tamanho da amostra; N é o número total de parcelas cabíveis na área (população fixa) e r é o coeficiente de correlação entre as parcelas. Conforme descrito por Cochran (1977), a presença de correlação positiva entre unidades da amostra inflaciona a variância da média. Mesmo uma pequena correlação positiva pode ter um grande efeito por causa do multiplicador $(n-1)$.

3 ESTUDO DA DEPENDÊNCIA ESPACIAL DE CARACTERÍSTICAS DENDROMÉTRICAS PARA *Eucalyptus grandis*

Resumo

O conhecimento da estrutura de continuidade espacial das características dendrométricas é de fundamental importância para o inventário, manejo e planejamento florestal. A inclusão do efeito da continuidade espacial nas análises do inventário proporciona resultados estáveis e seguros que poderão ser utilizados no manejo e no planejamento florestal. Objetiva-se com este capítulo avaliar a estrutura de continuidade espacial de quatro características dendrométricas obtidas por dois diferentes procedimentos de amostragem. Os dados foram coletados em 987 hectares, plantada com *Eucalyptus grandis*, localizada na região sul do estado de São Paulo. Estes dados foram obtidos a partir de duas amostras lançadas por dois procedimentos de amostragem diferentes, sistemático quadrado e sistemático desenhado. Para cada parcela das duas amostras, obtiveram-se as seguintes características dendrométricas: volume, diâmetro médio quadrático, área basal e altura média das dominantes. Para cada uma destas características construíram-se semivariogramas anisotrópicos e isotrópicos, a fim de avaliar a estrutura de continuidade espacial. Os semivariogramas foram construídos com o estimador do método dos momentos. Ajustaram-se aos semivariogramas experimentais as principais funções autorizadas pelo Método dos Quadrados Mínimos. Através do grau de dependência espacial e dos semivariogramas escalonados, avaliou-se o

comportamento da continuidade espacial para as quatro características, nos respectivos métodos de amostragem avaliados. Todas as características avaliadas apresentaram-se estruturadas espacialmente, independente do procedimento de amostragem avaliado. Verificou-se que a estrutura de continuidade espacial das quatro características foi isotrópica, ou seja, a variância entre pares de pontos depende apenas do vetor h de separação. Assim, pode-se construir semivariogramas omnidirecionais para as quatro características avaliadas. Os resultados sugerem que, no processamento do inventário florestal, deve-se considerar a componente espacial, ou seja, as parcelas não devem ser tratadas de forma independente.

Palavras-chave: *Eucalyptus grandis*, inventário florestal, geoestatística, características dendrométricas

Summary

The knowledge of spatial continuity structure of dendrometric characteristics is crucial in forestry inventory, management and planning. The inclusion of spatial continuity effect in inventory analyses provides stable and safe results that can be used in forestry management and planning. This chapter evaluated the structure of spatial continuity of four dendrometric characteristics obtained by two sampling procedures. Data were collected in 987 hectares of *Eucalyptus grandis*, located in the south of São Paulo state. Two sampling procedures were used, systematic grading and unencumbered systematic. For each plot of the two samples, the following dendrometric characteristics were measured: volume, quadratic average diameter, basal area and average height of the dominant trees. For each one of these characteristics, anisotropic and isotropic semi variogrammes were built, in order to evaluate the spatial continuity structure. The semi variogrammes were built using the moment estimator method. The main authorized functions were adjusted to the experimental semi variogrammes, by The Minimum Square Method. The behavior of the spatial continuity was evaluated through the degree of spatial dependence and of the assigned semi variogrammes for the four characteristics, in the appraised sampling methods respectively. All the appraised

characteristics presented spatially structured, independently of the appraised sampling procedure. The continuity structure of the four characteristics was isotropic, i.e., the variance among pairs of points depends on the separation vector h . Therefore, unidirectional semi variogrammes can be built for all appraised characteristics. The results suggest that, in forestry inventory, the spatial component should be considered, i.e., plots should not be treated separately.

Key-words: *Eucalyptus grandis*, forestry inventory, geostatistics, dendrometric characteristics

3.1 Introdução

Uma característica marcante em área reflorestada é a sua aparente homogeneidade, contudo, é freqüente a observação de variações significativas ao longo do plantio. Estas variações podem ser devido a vários fatores, dentre eles: solo, topografia, variações na prática silvicultural de plantio e outros. O conhecimento dessas variações é importante para o inventário e o manejo florestal. É usual a adoção de um valor médio, suposto representativo, obtido de uma amostra (conjunto de parcelas), para caracterizar o estado atual e temporal da variável de interesse dentro da floresta. Essa variabilidade espacial pode, muitas vezes, afetar de forma marcante a qualidade das estimativas feitas pelo inventário. Desse modo, uma análise criteriosa, visando conhecer o comportamento espacial das características dendrométricas, é de fundamental importância na escolha de um método estatístico que otimize recurso sem perda de precisão no processo de estimativa.

Embora a Geoestatística seja amplamente aplicada na Geologia e Ciência do Solo, apresenta também grande potencial para ser aplicada em outras áreas da ciência, podendo caracterizar o comportamento espacial de variáveis, tais como as características dendrométricas. A Teoria das Variáveis Regionalizadas, versão inicial da Geoestatística, foi desenvolvida por Matheron (1963), que a definiu a partir de uma função espacial numérica, variando de um local para outro, apresentando continuidade aparente, cuja variação não pode simplesmente ser representada por uma função matemática. Hoje, a

Geoestatística tem a sua fundamentação teórica totalmente inserida na teoria convencional da Estatística, como pode ser visto em Cressie (1993), Oliveira (1991) e Diggle & Ribeiro Júnior (2000).

A preocupação com a estrutura espacial de características dendrométricas não é assunto novo. Osborne (1942) e Matérn (1960) apresentaram a função exponencial para caracterizar a continuidade espacial de volumes de madeira obtidos de amostras sistemáticas. Matérn (1960) avaliou o desempenho da amostragem sistemática com grade quadrada e casual simples. Usando a função exponencial para obtenção da correlação, o autor verificou que a grade sistemática quadrada foi superior à amostragem casual simples no controle da variação do volume de madeira. Para estimar a área coberta com floresta, Matérn (1960) avaliou dois procedimentos de amostragem: sistemático quadrado e sistemático desencontrado versus os procedimentos casualizados para uma área coberta com floresta. Segundo Matérn (1960), o procedimento sistemático, quadrado ou desencontrado, foi superior aos procedimentos acidentais ou aleatórios na estimativa da variância. Quenouille (1949), em estudos teóricos, verificou que a amostra sistemática desencontrada foi freqüentemente superior à sistemática quadrada e à casual estratificada. A amostragem sistemática desencontrada é o procedimento que mais se identifica com aqueles freqüentemente praticados pelas empresas de reflorestamento. Este procedimento é caracterizado por não apresentar distâncias fixas entre parcelas, sendo as parcelas distribuídas de forma a cobrir toda a floresta.

Nos anos 80, o tema dependência espacial voltou a ser alvo de estudo na ciência florestal. Samra et al. (1989) procuraram modelar o crescimento de altura para *Melia azedarach* na Índia fazendo uso de técnicas da Geoestatística. A associação de técnicas geoestatísticas com sensoriamento remoto foram trabalhadas por Cohen et al. (1990) e Wulder et al. (1996). Estes autores procuraram caracterizar a estrutura de copa de coníferas a partir de imagens de satélite e parâmetros biofísicos, tais como índice de área foliar e área basal. Eles concluíram ter sido possível usar métodos geoestatísticos de krigagem para mapeamento da altura dominante e área basal na área estudada. Outras aplicações da Geoestatística na área florestal são, na estimativa de índices de sítio (Höck

et al., 1998), estimativa do volume total de madeira num povoamento de *Pinus* sp. (Gunnarson et al., 1998) e na determinação do grau de dependência espacial para área basal de *Pinus ponderosa* (Biondi et al., 1994). Thompson (1992) mostrou a importância da krigagem em estudos ecológicos, utilizando informações de uma área amostrada para prever determinada característica em toda a região.

Mediante o exposto, o presente trabalho teve como objetivo avaliar a estrutura de dependência espacial de quatro características dendrométricas, obtidas a partir das informações de dois inventários florestais, cujas parcelas foram lançadas conforme dois procedimentos de amostragem diferentes.

3.2 Material e Métodos

3.2.1 Descrição da área e coleta de dados

O trabalho foi conduzido numa área de 986,78 hectares de *Eucalyptus grandis* com 7 anos de idade, denominada “Projeto Ligiana”, pertencente à Votorantim Celulose e Papel (VCP). O plantio localizou-se no município de Itapetininga na região sul do estado de São Paulo. A área em estudo apresentou coordenadas geográficas 23°56’ de latitude sul e 42°21’ de longitude oeste, com uma altitude média de 655 metros. O clima da região classificou-se como mesotérmico úmido, sem estiagem, com temperatura do mês mais quente superior a 22°C, e com frequência média de geada de 5 a 10 dias por ano. A precipitação e a temperatura média anual, respectivamente, foi de 1600 mm e 20°C. O solo predominante na área foi LATOSSOLO VERMELHO Distrófico.

Os dados para realização do presente estudo foram obtidos a partir de duas amostras lançadas por dois procedimentos de amostragem diferentes. Foi conduzido um experimento com parcelas de 525 m² (25 X 21m), sistematizada em toda a área, distantes entre si de 500 metros. Cada parcela foi georeferenciada num ponto do vértice superior esquerdo, (Figura 1). Doravante, este procedimento de amostragem será denominado de procedimento 1.

Com o objetivo de estimar melhor o efeito pepita (variação ao acaso), foram selecionados três pontos na área, em que se realizou uma amostragem intensa. Por motivo de melhor controle nas operações de campo, fez-se o ponto selecionado coincidir com um dos vértices da parcela da malha 500 X 500 m². Após cada ponto selecionado, estabeleceu-se uma segunda malha com parcelas sistematizadas de 100 em 100 m². Posteriormente, da quadrícula central da malha 100 X 100 m², lançou-se sistematicamente parcelas de 25 em 25 metros. Ao todo foram avaliadas 136 parcelas no delineamento, em três escalas diferentes: pequena (25 X 25 m²), média (100 X 100 m²) e grande (500 X 500 m²). Em cada uma das 136 parcelas, mediu-se o diâmetro de todas as árvores, a altura das 10 primeiras plantas e a altura dominante conforme o conceito de Assmam (Prodan et al. 1997). Com estas características medidas, foi possível obter o volume por hectare, o diâmetro médio quadrático, a altura média das dominantes e a área basal por hectare, as quais foram avaliadas espacialmente.

O segundo procedimento de amostragem, denominado de procedimento 2, foi aquele utilizado pela VCP para lançamento das parcelas permanentes, as quais permitem acompanhar o crescimento da floresta. Neste procedimento, a intensidade amostral foi de uma parcela a cada 10 hectares, sendo que todos os talhões foram amostrados. Esta amostragem fez com que houvesse uma boa distribuição das parcelas dentro de toda a área do projeto, proporcionando uma varredura quase sistemática sobre a mesma (Figura 2). Este procedimento de amostragem pode ser considerado sistemático, não regular ou desencontrado (Cochran, 1977). Avaliaram-se as mesmas características mensuradas no procedimento de amostragem 1.

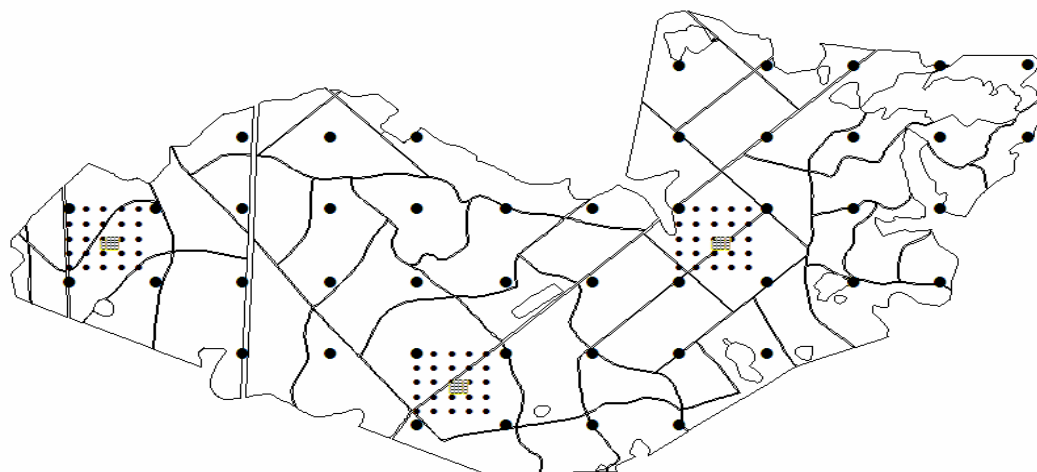


Figura 1 - Croquis da área experimental com as parcelas lançadas conforme procedimento de amostragem 1

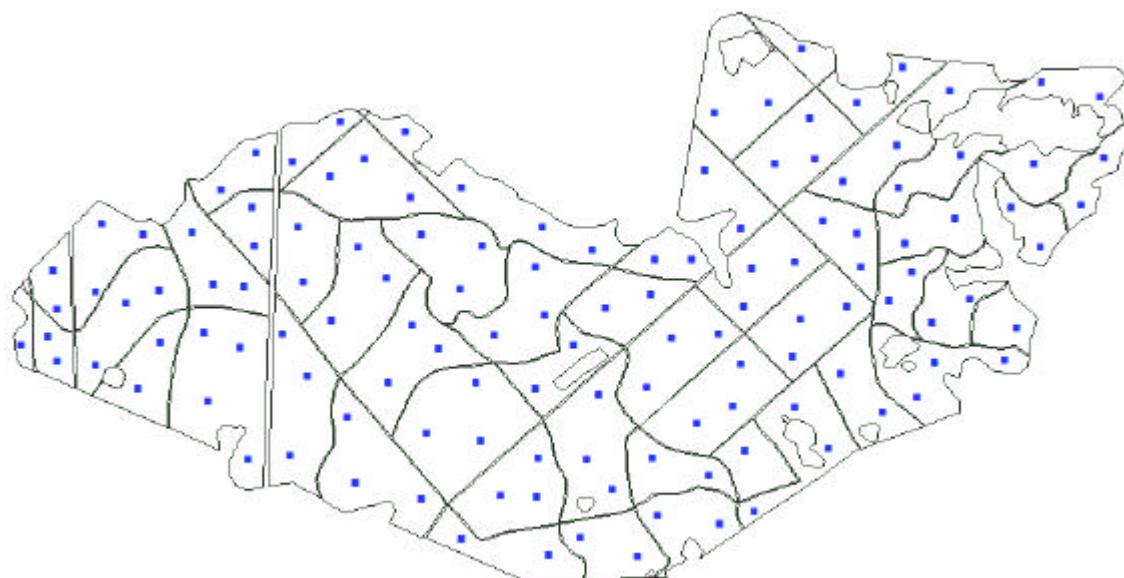


Figura 2 - Croquis da área experimental com as parcelas do procedimento de amostragem 2

A primeira etapa da análise dos dados consistiu na análise exploratória dos mesmos, realizada com o objetivo de observar o comportamento geral dos dados (forma, distribuição, tendência central). Este tipo de análise, embora desconsidere a estrutura de dependência espacial, é importante por permitir a identificação de dados atípicos que exercem algum tipo de influência nas análises geoestatísticas, como também para comparação entre resultados obtidos em outros trabalhos. A análise exploratória de dados constituiu na determinação das principais medidas de posição e de dispersão para as características avaliadas em cada procedimento de amostragem. Para testar a hipótese de normalidade dos dados de cada característica, utilizou-se o teste de Jarque.Bera (Jarque e Bera, 1980) a 5%.

3.2.2 Semivariograma

Considere 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 correlação destas variáveis aleatórias no espaço, quando x varia, é dado pelo semivariograma. Na Figura 3 está apresentado um semivariograma típico para dados sem tendência e com ruído. A linha tracejada refere-se a um semivariograma sem ruído. Neste semivariograma estão apresentados os parâmetros que caracterizam a estrutura de correlação: efeito pepita ou “nugget” (τ^2), contribuição ou “sill partial” (σ^2) e alcance ou “range” (ϕ).

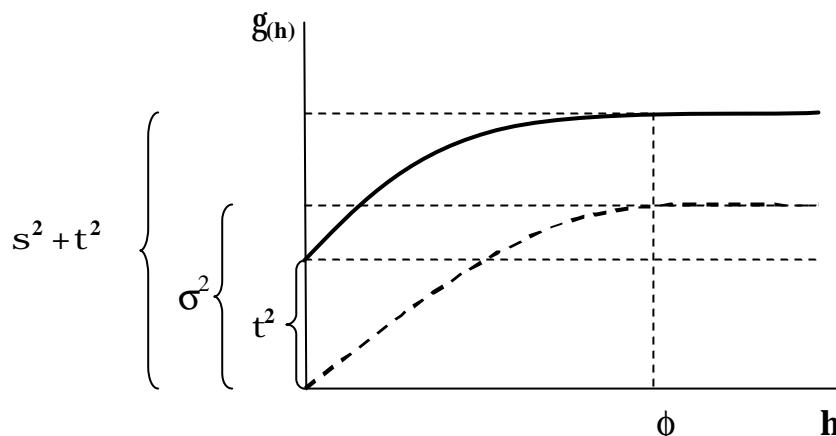


Figura 3 - Semivariograma típico para dados sem tendência e com ruído

Considere dois pontos separados por uma distância h . A variação entre eles é caracterizada por uma função variograma $2g(x,h)$ (Mc Bratney & Webster, 1986). Esta função é definida como

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

Note que a função variograma, como representado 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, há uma relação direta entre a correlação e a semivariância. Esta relação é dada pela seguinte expressão:

$$g(h) = c(0) - r(h) \quad (2)$$

em que $c(0)$ é a variância dos dados “sill total”, igual à $\tau^2 + \sigma^2$, $r(h)$ é a correlação entre dois pontos separados por uma distância h .

Na Teoria das Variáveis Regionalizadas, outro nome dado à Geoestatística (Journel & Huijbregts, 1978), a quantidade $\mathbf{g}(h)$ 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 $\mathbf{g}(\cdot)$ é o semivariograma. Se obedecerem-se às condições da hipótese intrínseca, a semivariância pode ser estimada simplesmente através de uma única realização do processo estocástico por:

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

onde $\hat{\mathbf{g}}(h)$ 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.

Para cada característica dendrométrica avaliada, ajustaram-se as funções autorizadas (modelos) apresentadas por Journel & Huijbregts (1978) e Mc Bratney & Webster (1986), a fim de se obter o conjunto de parâmetros para as respectivas funções. Os ajustes foram feitos pelo Método dos Quadrados Mínimos Ponderados, usando o programa R, através do pacote geoR (Ribeiro Júnior & Diglle, 2001).

A avaliação do grau de dependência espacial (DE) das características, após a definição do melhor modelo, foi realizada a partir do percentual de variação estruturada (σ^2) em relação ao patamar ($\tau^2 + \sigma^2$) apresentado por Biondi et al. (1994). Foram construídos os semivariogramas escalonados, para todas as características em cada procedimento de amostragem. Este semivariograma é útil para visualização da estrutura de dependência espacial entre características com diferentes unidades. Plotaram-se também, os semivariogramas experimentais de cada característica para os dois planos de amostragem. Estes gráficos permitiram avaliar o efeito dos planos de amostragem (1 e 2) sobre a estrutura de dependência espacial de cada característica dendrométrica.

3.3 Resultados e Discussão

3.3.1 Análise exploratória de dados

Na Tabela 1 estão apresentados os momentos estatísticos de cada característica avaliada nos respectivos procedimentos de amostragem. Verificou-se pelo teste de Jarque.Bera (JB) que todas as características nos respectivos procedimentos, tiveram distribuição aproximadamente normal. A estatística Jarque.Bera testa a hipótese nula de que a amostra é proveniente de uma população com distribuição normal, ou seja, que as características apresentaram distribuição unimodal, em geral simétrica, que se aproxima da distribuição Normal. A semelhança entre as medidas de posição (média e mediana) das características, também expressa a simetria das distribuições. Segundo Cressie (1993), um fundamento comum da modelagem estocástica é que os dados tenham distribuição Normal, sob a qual toda inferência geoestatística linear tem-se desenvolvido. Deve-se observar, porém, que a geoestatística não exige normalidade dos dados para ser aplicada. Se houver normalidade, as inferências realizadas ganharão outras propriedades estatísticas ótimas, tal como a Máxima Verossimilhança. Assim sendo, as características trabalhadas no presente estudo apresentaram condições ótimas para aplicações de técnicas geoestatísticas que dependem de um modelo previamente estabelecido.

Em geral, os resultados da Tabela 1 apresentaram, para as características avaliadas, pequena magnitude de variação. Independente do procedimento de amostragem, o coeficiente de variação variou de aproximadamente 5 a 13%. A menor medida de dispersão em torno da média foi para o diâmetro médio quadrático e a altura dominante média.

Tabela 1. Momentos estatísticos média (\bar{X}), mediana (med), desvio padrão (s), coeficiente de variação (CV%) e estatística de Jarque.Bera (JB) das características dendrométricas (CAR.DEND.): volume total (VT); diâmetro médio quadrático (DG); altura média das dominantes (HD) e área basal (G) avaliadas nos dois procedimentos de amostragem

| CAR. DEND. | PROCEDIMENTO 1 | | | | | PROCEDIMENTO 2 | | | | |
|---------------|----------------|-------|------|-------|--------|----------------|-------|------|-------|--------|
| | \bar{X} | Med | S | CV(%) | JB | \bar{X} | Med | S | CV(%) | JB |
| VT | 260,9 | 261,9 | 32,9 | 12,6 | 0,9842 | 270,5 | 268,0 | 34,7 | 12,8 | 0,9853 |
| DG | 15,5 | 15,4 | 1,1 | 7,1 | 0,9862 | 15,4 | 15,4 | 0,9 | 5,8 | 0,9931 |
| HD | 27,6 | 27,5 | 1,9 | 6,9 | 0,9884 | 27,9 | 27,9 | 1,8 | 6,5 | 0,9904 |
| G | 22,6 | 22,6 | 2,3 | 10,2 | 0,9923 | 23,3 | 23,0 | 2,5 | 10,7 | 0,9838 |

Também foi possível detectar através da análise exploratória de dados, que houve valores candidatos a “outliers” ou discrepantes. Valores discrepantes são valores que têm comportamento diferente do apresentado pela maioria dos dados (Hoaglin, 1992). Segundo Diggle & Ribeiro Júnior (2000), estes valores provocam um forte impacto principalmente na parte inicial do variograma. Este fato pode conduzir a conclusões erradas sobre a variância do ruído (efeito pepita) e ou a existência de correlação espacial. Diante do exposto, os valores candidatos a “outliers”, de cada variável estudada, foram removidos da base de dados, a fim de que o estudo variográfico não fosse comprometido.

3.3.2 Avaliação da continuidade espacial

A análise variográfica mostrou que todas as características estudadas nos dois procedimentos apresentaram-se estruturadas espacialmente, ou seja, existiu uma função estrutural, com semivariância de comportamento modelável. Para todas as características avaliadas, o modelo exponencial foi o que melhor se ajustou ao semivariograma experimental. Os ajustes dos parâmetros estão apresentados na Tabela 2. Nas Figuras 4 e

5 encontram-se os semivariogramas escalonados para as quatro características avaliadas nos dois procedimentos de amostragem.

Tabela 2. Estimativa dos parâmetros efeito pepita (t^2), variação estruturada (S^2), patamar ($t^2 + S^2$), alcance (q) e relação variação estruturada/patamar (DE%) do modelo exponencial ajustados aos semivariogramas, para cada característica dendrométrica (CAR.DEND.) volume total (VT); diâmetro médio quadrático (DG); altura média das dominantes (HD) e área basal (G) avaliada dentro dos respectivos procedimentos de amostragem

| CAR. DEND. | PROCEDIMENTO 1 | | | | | PROCEDIMENTO 2 | | | | |
|---------------|----------------|-------------|-------------|-------|-------|----------------|-------------|-------------|-------|--------|
| | $t^2_{(1)}$ | $S^2_{(1)}$ | $t^2 + S^2$ | q (m) | DE(%) | $t^2_{(1)}$ | $S^2_{(1)}$ | $t^2 + S^2$ | q (m) | DE(%) |
| VT | 261,34 | 948,90 | 1210,24 | 250 | 78,8 | 260,0 | 920,27 | 1180,27 | 183 | 78,00 |
| DG | 0,50 | 0,83 | 1,33 | 188 | 62,4 | 0,44 | 0,43 | 0,87 | 573 | 49,4 |
| HD | 0,83 | 2,61 | 3,44 | 279 | 75,9 | 0,00 | 3,26 | 3,26 | 176 | 100,00 |
| G | 1,80 | 4,03 | 5,83 | 238 | 69,1 | 2,00 | 3,79 | 5,79 | 185 | 65,46 |

⁽¹⁾ unidade da característica ao quadrado

Para verificar o grau de dependência espacial, entre as características, utilizou-se da relação entre a variação estruturada e o patamar (DE%). Essa relação foi relatada em trabalho desenvolvido por Biondi et al. (1994), que estudaram o comportamento espacial do diâmetro a altura do peito e da área basal em floresta natural de *Pinus ponderosa*. Pela Tabela 2 verificou-se que a altura dominante e o volume total foram as características que apresentaram melhor estruturação espacial. Pela classificação de Cambardella et al. (1994), que trabalharam com atributos do solo, quando a relação for maior ou igual a 75%, há uma forte dependência espacial. Tanto o volume quanto a altura dominante apresentaram grau de dependência espacial (DE%) maior do que 75% para os dois procedimentos, podendo ser classificadas como características com dependência espacial forte. Isto mostra que as estimativas de volume e altura dominante com o modelo exponencial, considerando a componente espacial, independente do procedimento de amostragem, trará otimização nas estimativas. O fato da altura dominante ter apresentado forte correlação espacial é um ótimo indicativo do potencial

da técnica de krigagem na classificação de sítio, concordando com os resultados obtidos por Höck et al. (1998). O diâmetro médio quadrático e a área basal foram as características que apresentaram os menores níveis de dependência espacial. Pela classificação de Cambardella et al. (1994) podem ser classificadas de média dependência espacial.

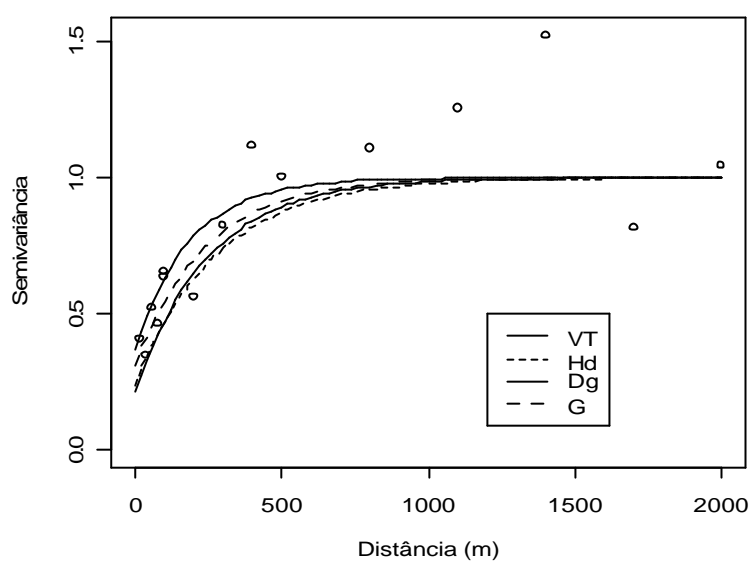


Figura 4 - Semivariogramas ajustados para cada uma das variáveis: volume total médio por hectare (VT); altura dominante média (Hd); diâmetro médio quadrático (Dg); área basal média por hectare (G), para o procedimento de amostragem 1

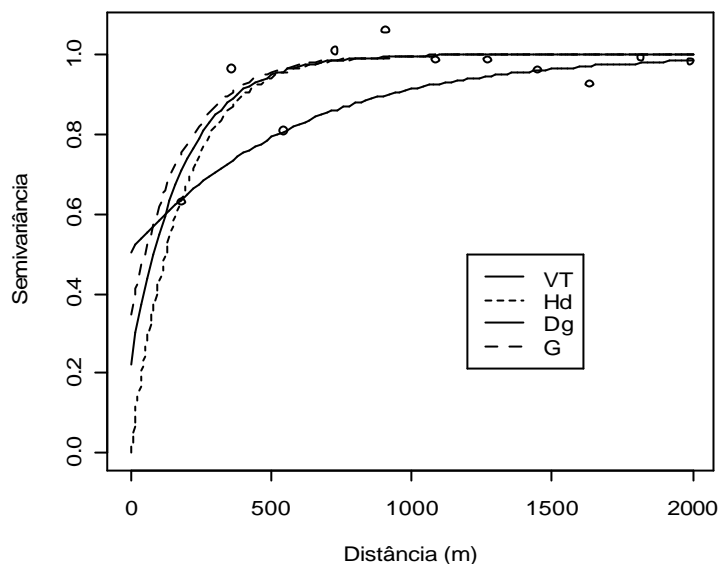


Figura 5 - Semivariogramas ajustados para cada uma das variáveis: volume médio por hectare (VT); altura dominante média (Hd); diâmetro médio quadrático (Dg); área basal média por hectare (G), para o procedimento de amostragem 2

Pelos semivariogramas escalonados verificou-se que, em cada procedimento de amostragem, as curvas de semivariâncias ajustadas para o volume e a altura dominante foram semelhantes, o que está de acordo com os resultados do Tabela 2 (DE%). O diâmetro médio quadrático foi a característica com menor grau de dependência espacial (DE%), e também, a característica que apresentou a curva de semivariância com maior diferença em relação às demais.

O alcance apresentado na Tabela 2 é chamado de alcance teórico, uma vez que este parâmetro no modelo exponencial é considerado infinito (Journel & Huijbregts, 1978). Para os alcances gerados por este modelo, utiliza-se o alcance prático, que é três vezes o alcance teórico (Journel & Huijbregts, 1978 e McBratney & Webster, 1986). Assim sendo, o alcance prático médio, independente do procedimento de amostragem, foi de 645 metros para volume, 690 metros para altura dominante e 630 metros para área basal. Para o diâmetro médio quadrático, o valor do alcance variou significativamente de um

procedimento para o outro. No procedimento 1, o alcance prático foi de 564 metros enquanto que, no procedimento 2 este valor foi acima de 1500 metros. Este fato pode ser verificado na curva do semivariograma experimental, onde a assíntota da curva ocorreu acima de 1000 metros. Estes valores de alcance correspondem ao raio das áreas consideradas homogêneas para cada característica.

O alcance encontrado para a variável volume mostrou que, em levantamentos efetuados nesta área de estudo, as parcelas para serem consideradas independentes, devem ser lançadas com um espaçamento mínimo de 700 metros. Na atual situação, qualquer método estatístico empregado para análise, deve considerar a dependência espacial. Ou seja, as parcelas não devem ser tratadas de forma independente, perdendo assim, qualidade nas estimativas e desperdício de recursos.

Estes resultados são importantes para o planejamento amostral e devem ser levados em consideração, na determinação do número de parcelas para obtenção de representatividade ótima do ponto de vista estatístico e econômico. Em face dos pontos levantados até o momento neste trabalho, fica difícil encontrar uma resposta para o número ideal de parcelas para amostrar adequadamente a área. A utilização da equação que determina o número de amostra (parcelas) amplamente divulgada em livros estatísticos e de inventário florestal, baseia-se na variação da característica e exige que as mesmas tenham distribuição normal e sejam independentes. Porém, os resultados contrariam essa exigência, isto é, não há como supor que volume, diâmetro médio quadrático, altura média dominante e área basal, sejam características essencialmente aleatórias. Como as análises geoestatísticas revelaram dependência espacial, propõe-se, a princípio, considerar o alcance na realização de futuras amostragens para qualquer uma das características avaliadas, para os dois procedimentos de amostragem avaliados.

Plotaram-se os semivariogramas para as quatro direções 0° , 45° , 90° , e 135° , para os dois procedimentos de amostragem, como pode ser visto nas Figuras 6 e 7. Nestas figuras plotou-se o semivariograma isotrópico (linha tracejada cheia). Como houve considerável semelhança entre os semivariogramas em todas as direções, é apropriado considerar o semivariograma isotrópico para a krigagem. Estes resultados revelaram que

as características avaliadas têm a propriedade de possuir a mesma estrutura de dependência espacial em todas as direções, caracterizando a isotropia.

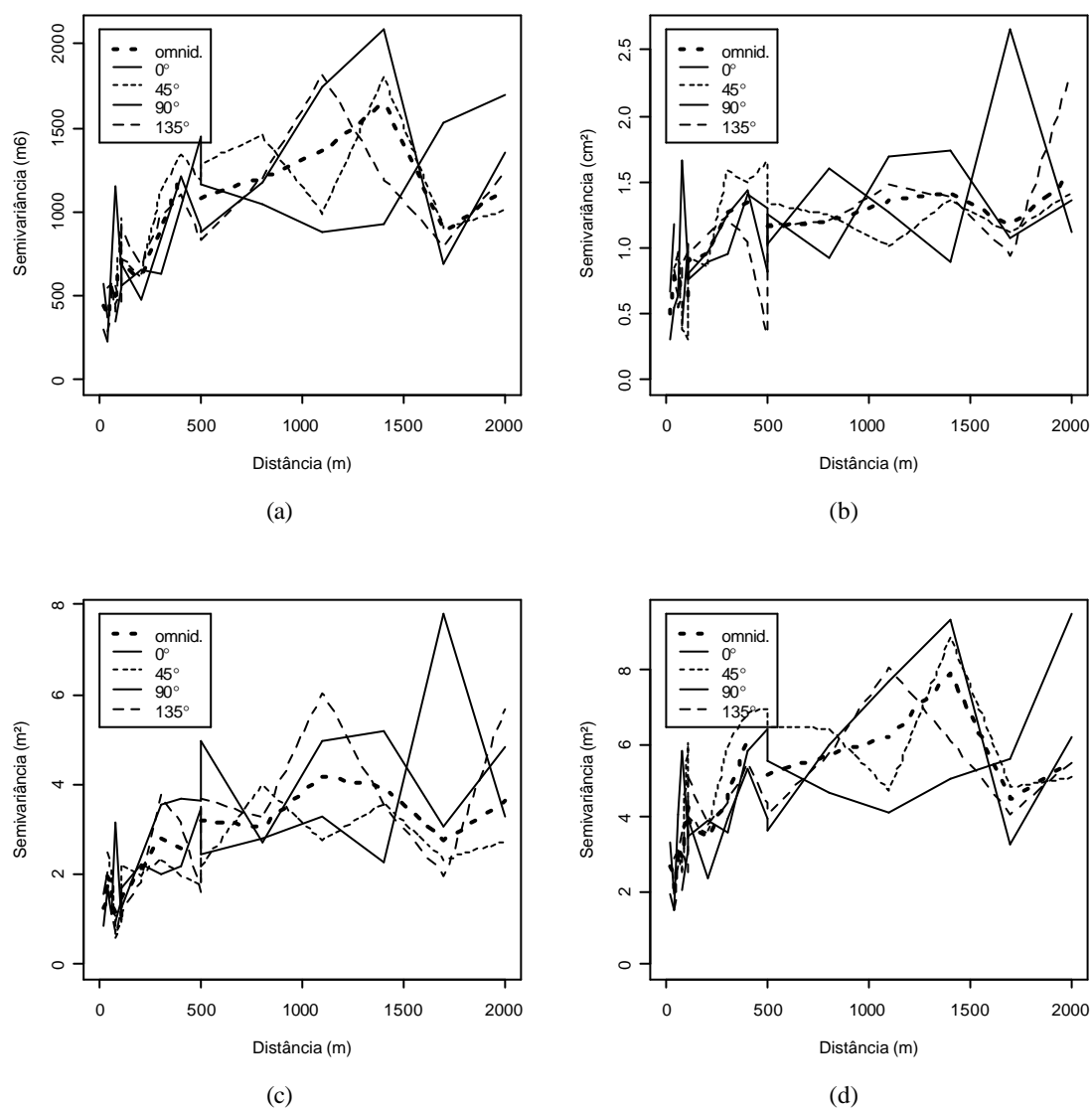


Figura 6 - Avaliação da estrutura de dependência espacial nas diferentes direções para as características dendrométricas: volume total (a); diâmetro médio quadrático (b); altura média das dominantes (c) e área basal (d) para o procedimento de amostragem 1

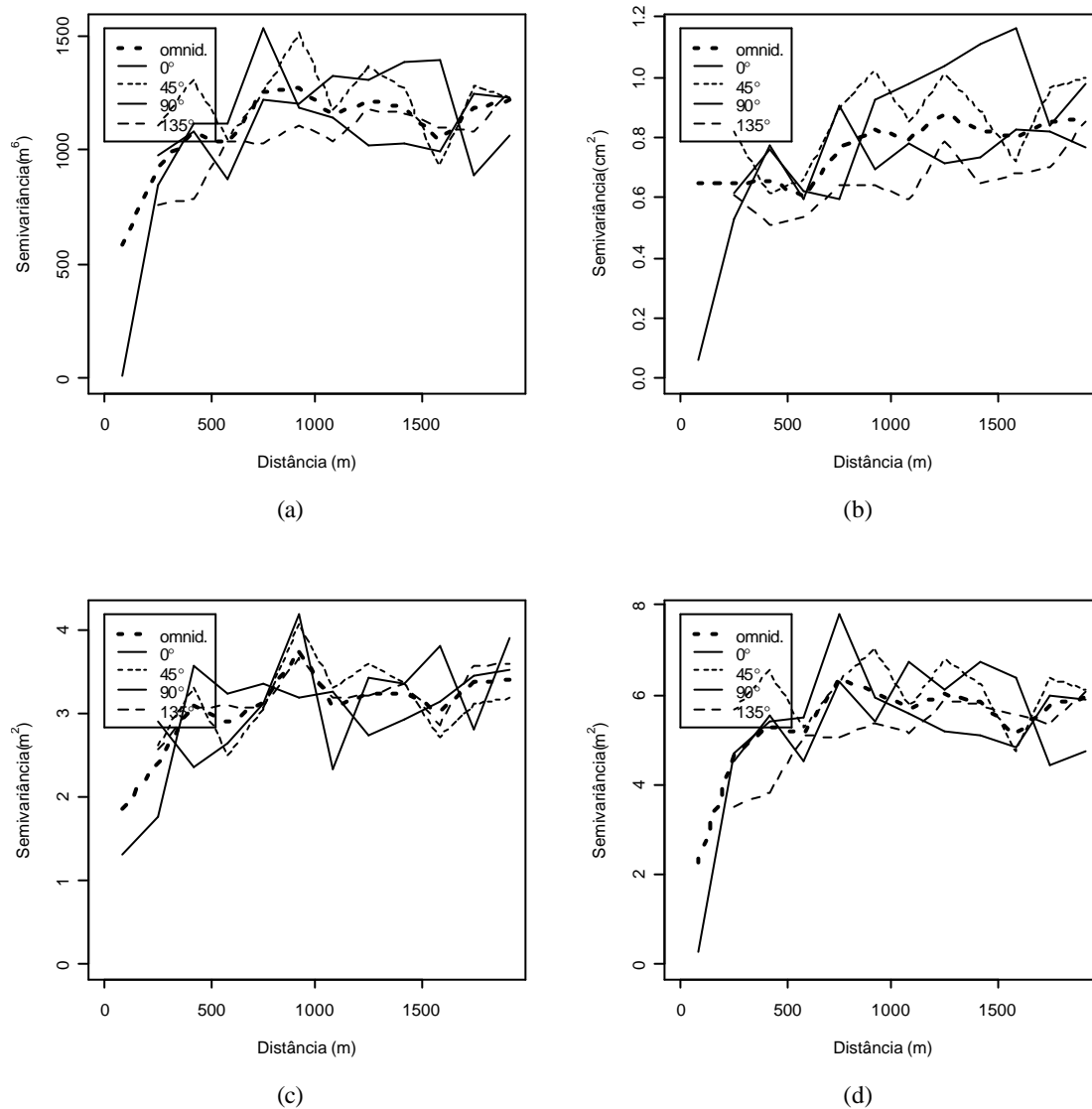


Figura 7 - Avaliação da estrutura de dependência espacial nas diferentes direções para as características dendrométricas: volume total (a); diâmetro médio quadrático (b); altura média das dominantes (c) e área basal (d) para o procedimento de amostragem 2

A Figura 8 representa a estrutura de dependência espacial para cada característica dendrométrica, através do semivariograma experimental, nos respectivos procedimentos de amostragem. Foi possível detectar que volume total, altura dominante e área basal apresentaram estrutura de dependência espacial semelhante nos dois procedimentos de amostragem. Já para diâmetro médio quadrático, a estrutura de dependência espacial teve comportamento diferente entre os dois procedimentos de amostragem. Este comportamento também pôde ser verificado nos semivariogramas escalonados (Figuras 4 e 5). Observou-se pela Figura 4 e 5 que o efeito pepita para as características avaliadas foi elevado. Estes resultados concordam com Aubry & Debouzie (2001), ao concluírem que o efeito pepita é alto em estudos que envolvem variáveis ecológicas.

Os resultados do presente estudo revelaram que a estrutura de dependência espacial para as características avaliadas foi semelhante nos dois procedimentos de amostragem, exceto para o diâmetro médio quadrático. Pelos parâmetros do modelo exponencial ajustado ao semivariograma experimental, para as características avaliadas, verificou-se ligeira superioridade do procedimento sistemático quadrado em relação ao sistemático desconstruído. Estes resultados contrariam os estudos teóricos de Quenouille (1949) e estão de acordo com Burgess et al. (1981) e Oliveira (1991), os quais demonstram que a variância mínima de krigagem é obtida a partir de uma grade regular de pontos.

Apesar da ligeira superioridade do procedimento sistemático em relação ao sistemático desconstruído, este resultado não condena o procedimento usado pela VCP para lançamento de parcelas permanentes. O importante, independente dos dois procedimentos, é amostrar sempre na menor escala para se ter uma boa estimativa do efeito pepita (variação aleatória). A sugestão seria a implantação de um transecto com parcelas separadas por distâncias de 50 a 300 metros para compor o conjunto de parcelas permanentes. Seguramente este conjunto de parcelas proporcionaria uma ótima estimativa do semivariograma.

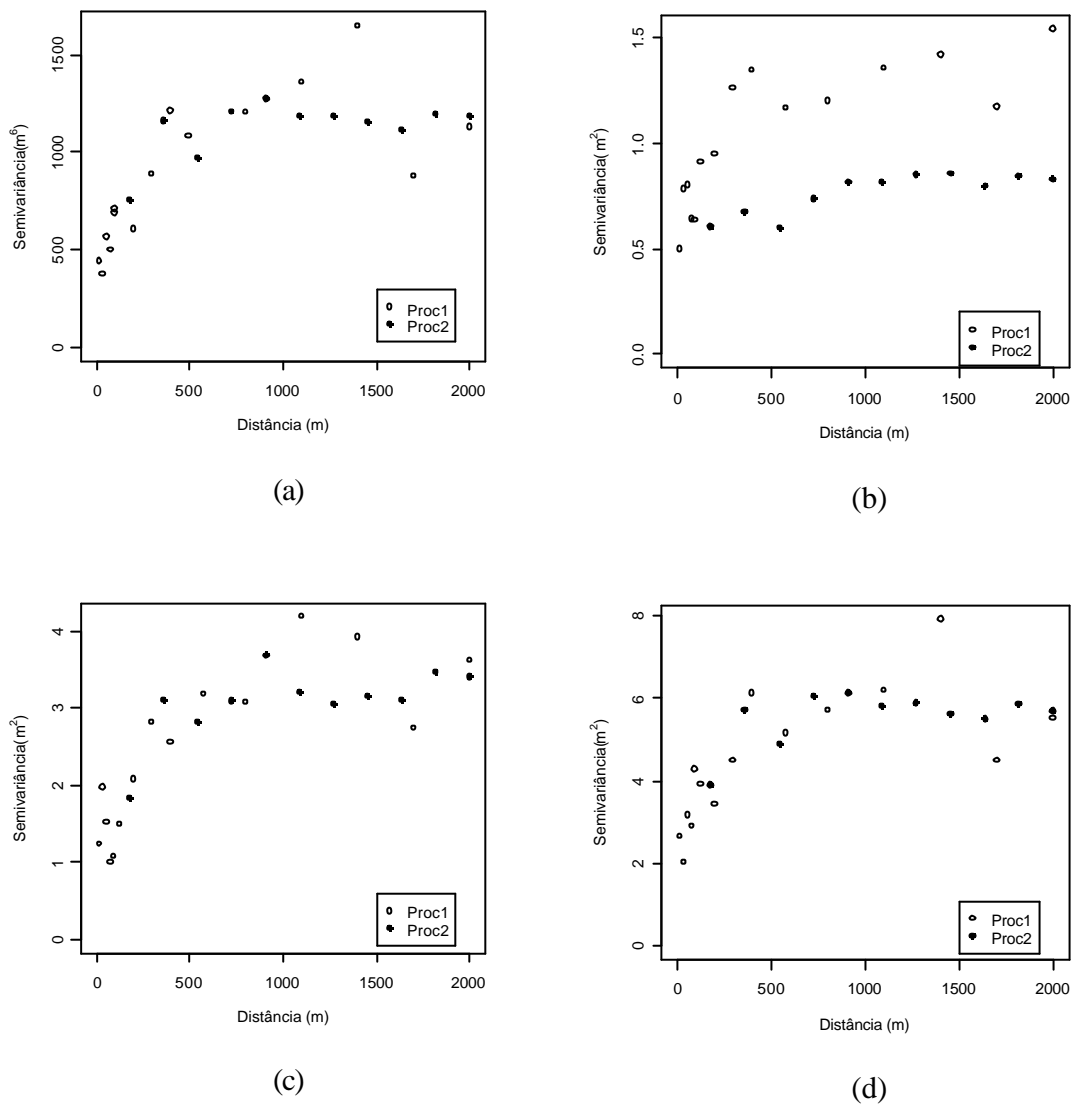


Figura 8 - Semivariogramas experimentais para as características dendrométricas volume total (a); diâmetro médio quadrático (b); altura dominante média (c) e área basal (d), nos procedimentos de amostragem 1e 2

3.4 Conclusões

- a) As características dendrométricas mensuradas no inventário florestal, por meio dos dois procedimentos de amostragem, apresentaram-se estruturadas espacialmente. Este fato revelou que, independente do procedimento de amostragem, a dependência espacial se manifesta, quando a mesma existe, desde que se tome cuidado para também amostrar a pequenas distâncias. Estes resultados sugerem que, no processamento do inventário florestal, deve-se considerar a componente espacial, ou seja, as unidades amostrais (parcelas) não devem ser tratadas de forma independente.
- b) Para as características consideradas neste trabalho, a estrutura de dependência espacial foi semelhante em todas as direções. Este resultado sugere o uso do semivariograma onidirecional no processo de krigagem, pressupondo, portanto, essa continuidade espacial como isotrópica.
- c) Independente do procedimento de amostragem, as características volume e altura dominante apresentaram forte dependência espacial. O volume é o produto final mais importante no processo de mensuração florestal. Para esta característica, o alcance prático foi de 700 metros, mostrando que as parcelas do inventário florestal, lançadas a distância inferior a 700 metros, não devem ser tratadas como independentes.
- d) A estrutura de dependência espacial não foi afetada pelos planos de amostragem avaliados. Para estudos geoestatísticos com as parcelas permanentes lançadas pela empresa, sugere-se agregar ao método atual de lançamento de parcelas, um sistema que seja capaz de amostrar também em pequenas distâncias entre parcelas.

4 AJUSTE E SELEÇÃO DE MODELOS ESPACIAIS DE SEMIVARIOGRAMA VISANDO À ESTIMATIVA VOLUMÉTRICA DE *EUCALYPTUS GRANDIS*

Resumo

O semivariograma é a parte central dos estudos geoestatísticos, sendo capaz de descrever tanto qualitativamente 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 preditiva indicaram o mesmo modelo e o mesmo método de ajuste com melhor desempenho. Este resultado é interessante, uma vez que nem sempre tem-se 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

Summary

The semi variogramme is the central part of the geoestatistic studies, describing qualitatively and quantitatively the spatial variatio n, and can be considered the key point in the determination geostatistic predictor – krigage, 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 of 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 quite interesting, because sometimes a new data base is not available for the predictive validation. It seems, therefore, that conjugating AIC and cross validation criterion for selecting semi variance function models is recommended.

Key-words: Spatial models, semivariogrammes, validation, AIC criterion

4.1 Introdução

Estimativas 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). Como não deveria ser diferente, estes estudos tiveram como ponto de partida à avaliação da dependência espacial por meio do semivariograma. O semivariograma é a parte central da Geoestatística, segundo Mc Bratney & 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 (Journel & Huijbregts, 1978). Em seguida, sua preocupação se volta para o ajuste da função matemática ao semivariograma experimental ou aos dados. Segundo Mc Bratney & 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, em 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 & 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 Mc Bratney & 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 & Wendroth (2003), a validação é um meio efetivo para avaliar a qualidade das estimativas de Krigagem, bem como 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.

4.2 Material e Métodos

4.2.1 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 a 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, para o estudo, 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 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 & Bera, 1980). A verificação da normalidade das diferenças se justificou pelo fato de que a Geoestatística 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.

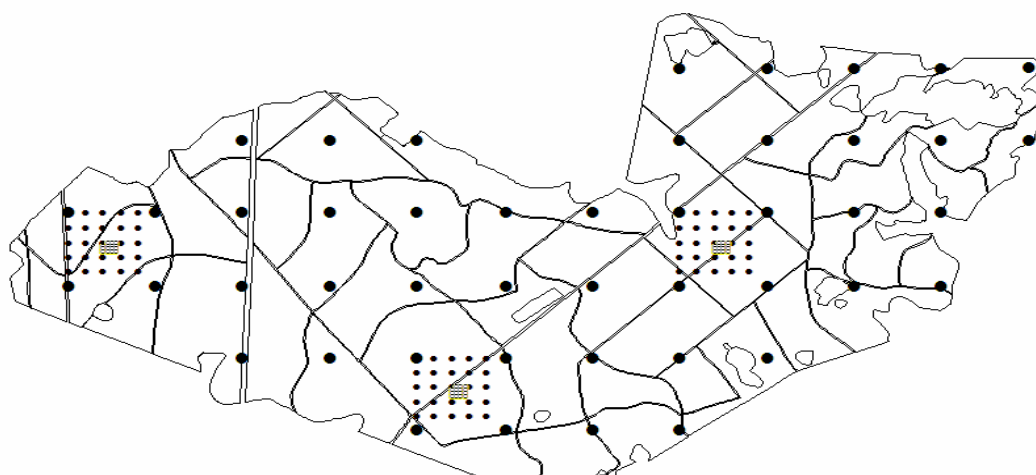


Figura 1 - Croqui da área experimental com as parcelas lançadas sistematicamente

4.2.2 Semivariograma

Considere 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 correlação destas variáveis aleatórias no espaço, quando x varia, é dado pelo semivariograma. Na Figura 2 está apresentado um semivariograma típico para dados sem tendência e com ruído. A linha

tracejada refere-se a um semivariograma sem ruído. Neste semivariograma estão apresentados os parâmetros que caracterizam a estrutura de correlação: efeito pepita ou “nugget” (τ^2), contribuição ou “sill partial” (σ^2) e alcance ou “range” (ϕ).

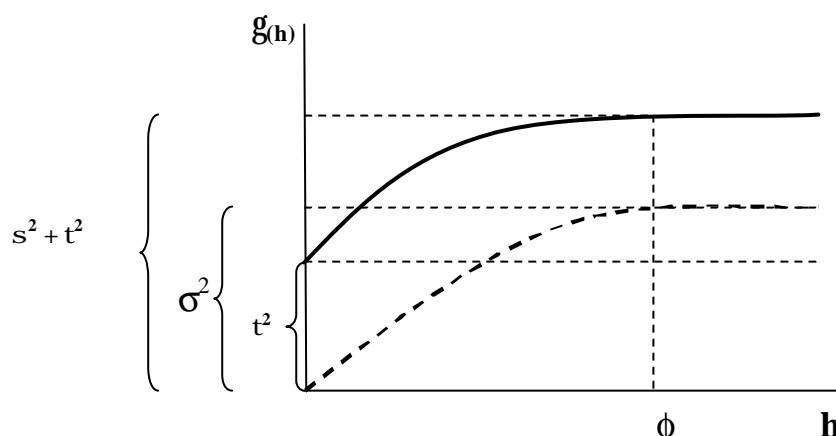


Figura 2 - Semivariograma típico para dados sem tendência e com ruído

Considere dois pontos separados por uma distância h . A variação entre eles é caracterizada por uma função variograma $2g(x, h)$ (Mc Bratney & Webster, 1986). Esta função é definida como

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

Note que a função variograma como representado 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, há uma relação direta entre a correlação e a semivariância. Esta relação é dada pela seguinte expressão:

$$\mathbf{g}(h) = c(0) - \mathbf{r}(h) \quad (2)$$

em que $c(0)$ é a variância dos dados “sill total”, igual à $\tau^2 + \sigma^2$, $\mathbf{r}(h)$ é a correlação entre dois pontos separados por uma distância h .

Na Teoria das Variáveis Regionalizadas, outro nome dado a Geoestatística (Journel & Huijbregts, 1978), a quantidade $\mathbf{g}(h)$ 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 $\mathbf{g}(\cdot)$ é o semivariograma. Obedecidas as condições da hipótese intrínseca, a semivariância pode ser estimada simplesmente através de uma única realização do processo estocástico por:

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

onde $\hat{\mathbf{g}}(h)$ 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 expressa 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.

$$\mathbf{r}(h) = \exp(-h/\mathbf{f}) \quad (4)$$

$$\mathbf{r}(h) = \left\{ 2^{k-1} \Gamma(k) \right\}^{-1} (h/\mathbf{f})^k K_k(h/\mathbf{f}) \quad (5)$$

onde $\mathbf{r}(h)$ é a correlação entre pares de pontos separados pela distância h , \mathbf{f} 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.

4.2.3 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)

4.2.3.1 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”.

4.2.3.2 Métodos dos Quadrados Mínimos Ordinários e Quadrados Mínimos Ponderados

Segundo Bussab & 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:

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

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 $\mathbf{g}(h_i; \mathbf{q})$ é a semivariância calculada pelo modelo e depende dos parâmetros estimados θ e após minimização, a função $g(h_i)$ se torna a estimativa $\hat{g}(h_i)$ de quadrados mínimos. David (1977) sugere dividir o erro quadrático médio pelo número de pares de pontos em cada distância no semivariograma $m(h_j)$, conforme a expressão a seguir:

$$Q(\mathbf{q}) = \frac{\sum_{i=1}^k [g(h_j) - \mathbf{g}(h_j; \mathbf{q})]^2}{m(h_j)} \quad (7)$$

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.

4.2.3.3 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 & Wonnacott, 1972). A idéia do método é obter, a partir de uma amostra, o estimador “mais verossímil” dos parâmetros de um certo modelo probabilístico. Considere 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:

$$Z_i = \mathbf{m}(x_i) + S(x_i) + \mathbf{e}_i \quad i = 1, \dots, n \quad (8)$$

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

$$\mathbf{m}(x_i) = \sum_{k=1}^p f_k(x_i) \mathbf{b}_k \quad (9)$$

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 qual no presente trabalho, foram avaliados os modelos exponencial e Matérn. A componente \mathbf{e}_i é a parte totalmente aleatória do modelo, sendo $\mathbf{e}_1, \dots, \mathbf{e}_n$ independentes, identicamente distribuídos, com média zero e variância τ^2 , denominada de nugget ou efeito pepita (Diggle & Ribeiro Júnior, 2000).

É conveniente escrever o modelo em forma de matriz, a fim de derivar a função de verossimilhança, para obter sua maximização. Considere que $\mu_{(x)} = F\mathbf{b}$, em que F é a matriz com os valores das funções f_k e \mathbf{b} 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 $\mathbf{q} = (\mathbf{t}^2, \mathbf{s}^2, \mathbf{f})$, descrita matricialmente como $VAR(Z) = K = \mathbf{t}^2 I + \mathbf{s}^2 \Sigma$, 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: $Z \sim MVN(F\mathbf{b}, \mathbf{s}^2 \Sigma + \mathbf{t}^2 I)$. Logo, a função de verossimilhança a ser maximizada tem a seguinte forma:

$$L(\mathbf{b}, \mathbf{q}) = -\frac{1}{2} \left\{ \log |\mathbf{K}| + (\mathbf{Z} - \mathbf{F}\mathbf{b})' [\mathbf{K}]^{-1} (\mathbf{Z} - \mathbf{F}\mathbf{b}) \right\} \quad (10)$$

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 (Diggle & Ribeiro Júnior, 2000).

4.2.3.4 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 & 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 & 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 a expressão do estimador é:

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

em que $\hat{z}(x_o)$ é o valor estimado para o ponto x_o ; I_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 Mc Bratney & Webster (1986) e Cressie (1993), o erro reduzido é definido como:

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

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

$$S_{ER} = \sqrt{\frac{1}{n} \left\{ \sum_{i=1}^n \frac{z(x_{i0}) - \hat{z}(x_{i0})}{S(x_{i0})} \right\}^2} \quad (15)$$

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, utilizando-se pacote geoR (Ribeiro Júnior & Diggle, 2001).

4.3 Resultados e Discussão

4.3.1 Análise Exploratória de Dados

Na Figura 3 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 & Ribeiro Júnior, 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 3, 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 distribuição 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.

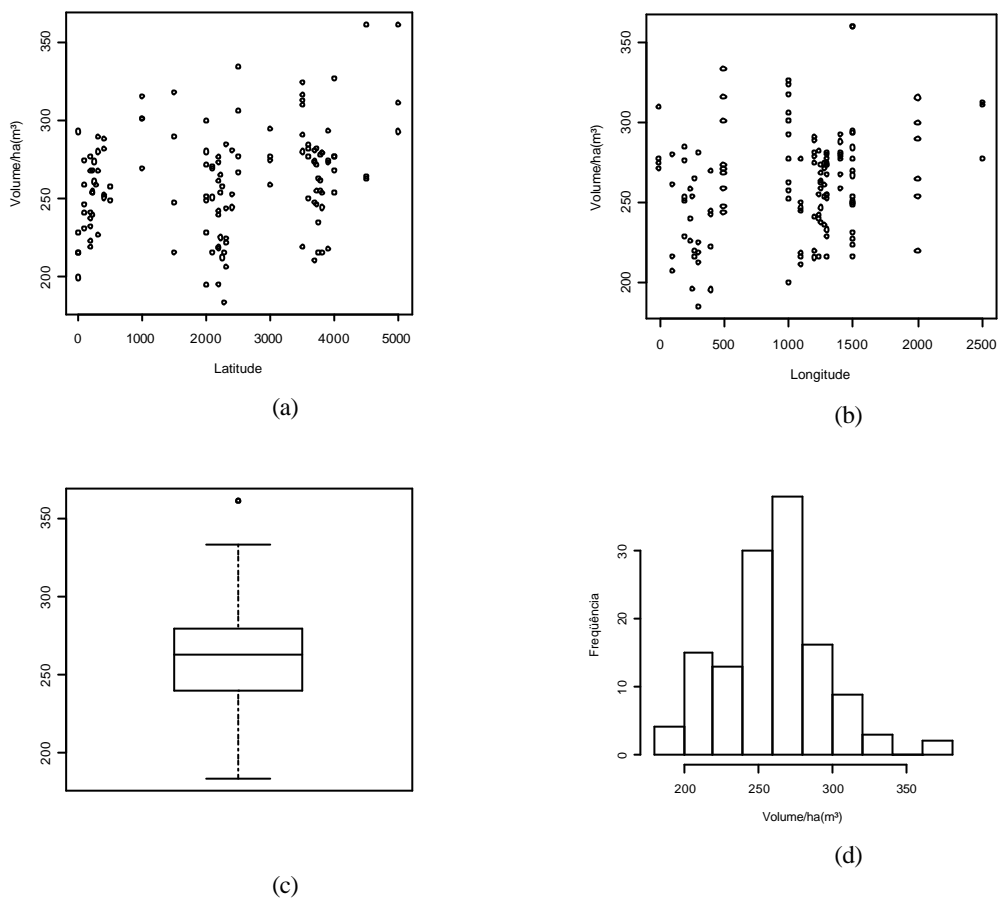


Figura 3 - Gráficos exploratórios do volume por hectare; (a) relação do volume com a latitude; (b) relação do volume com a longitude; (c) histograma de frequência e (d) gráfico de Box Plot

Tabela 1. Teste de Jarque.Bera para as diferenças nas respectivas distâncias do semivariograma empírico

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

D – estatística de Jarque.Bera; NS e S – teste não significativo e significativo respectivamente; n – número de pares de pontos em cada distância do semivariograma.

4.3.2 Avaliação do Critério de Informação de Akaike (AIC)

Nas Figuras 4 e 5 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 se confundiram 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 & 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 métodos 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, é mais aplicável para a seleção do melhor modelo.

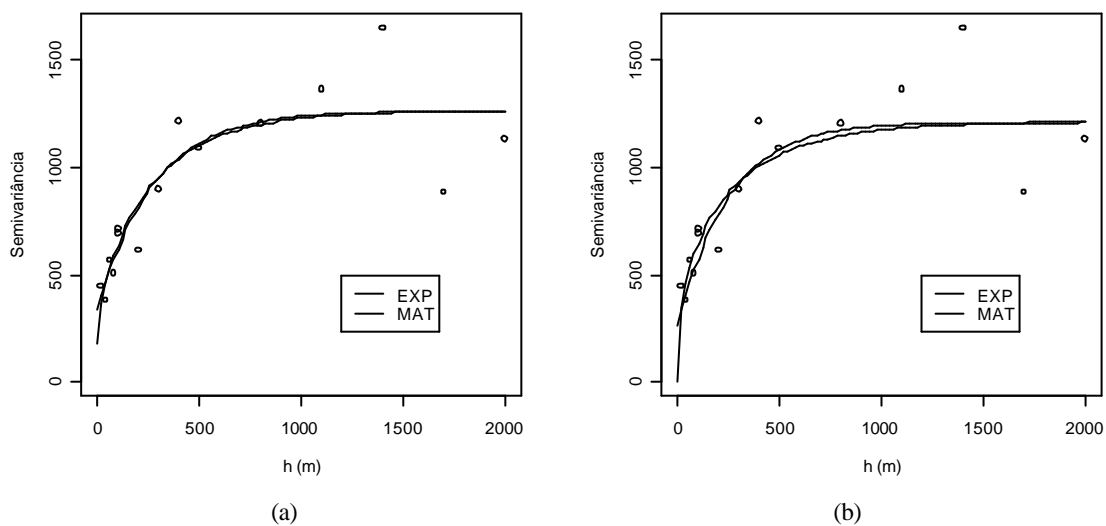


Figura 4 - Modelos ajustados ao semivariograma conforme os Métodos dos Quadrados Mínimos Ordinários (a) e Quadrados Mínimos Ponderados (b)

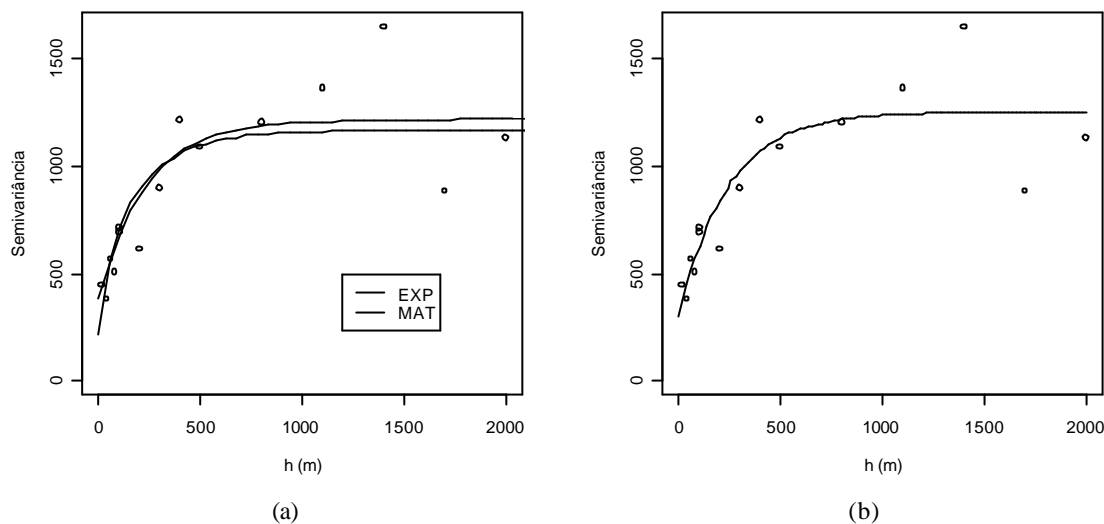


Figura 5 - Modelos ajustados ao semivariograma conforme o Método da Máxima Verossimilhança (a) e “a sentimento” para o modelo exponencial (b)

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 & Anderson (2002). A diferença de AIC (Δ_i) entre os modelos exponenciais 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 estimados para cada um dos modelos (Tabela 3). Estes resultados contrariam a análise da Figura 5(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.

Tabela 2. Critério de Akaike (AIC), para cada ajuste conforme o método e o modelo avaliado

| 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 | - |

OLS – quadrados mínimos ordinário; WLS – quadrados mínimos ponderados; ML – Método da Máxima Verossimilhança.

Tabela 3. Parâmetros estimados para cada modelo através dos respectivos métodos de ajustes

| MÉTODO DE AJUSTE | Exponencial | | | Matérn | | |
|------------------|-------------|--------|-------|--------|---------|-------|
| | t^2 | s^2 | f | t^2 | s^2 | f |
| 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 & 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 & 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 forem 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 5b), 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 curva 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.

4.3.3 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.

Tabela 4. Estatísticas da validação cruzada para cada modelo nos respectivos métodos de ajuste

| MODELO | MÉTODO DE AJUSTE | (\overline{ER}) | (S_{ER}) | EA (m ³ /ha) |
|-------------|------------------|-------------------|------------|----------------------------|
| Exponencial | “à 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 |

Conforme McBratney & Webster (1986), Cressie (1991) 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.

Tabela 5. Estatísticas da validação preditiva para cada modelo nos respectivos métodos de ajuste

| MODELO | MÉTODO DE AJUSTE | $\overline{(ER)}$ | (S_{ER}) | EA (m ³ /ha) |
|-------------|------------------|-------------------|------------|----------------------------|
| Exponencial | “à 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 |

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.

4.4 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.

5 MÉDIA ESPACIAL PARA VOLUME DE *EUCALYPTUS GRANDIS* ESTIMADA POR GEOESTATÍSTICA BASEADA EM MODELO

Resumo

O volume de madeira de uma população florestal é determinado através de uma amostra, selecionada com base em algum procedimento de amostragem. Os procedimentos de amostragem comumente utilizados nos inventários florestais foram desenvolvidos com base nos conceitos da teoria clássica de amostragem que pressupõem independência. O desenvolvimento dos estimadores desta teoria foram efetuados considerando o conceito de população fixa. A utilização destes estimadores numa população estocástica (não fixa) e com dependência espacial pode aumentar os problemas estatísticos. Este trabalho teve como objetivo principal desenvolver uma metodologia de análise de inventário florestal, que considerasse a estrutura de dependência espacial, da característica volume. Especificamente, pretendeu-se avaliar as coberturas observadas do intervalo de confiança dos estimadores geoestatístico e clássico para uma população estocástica e uma população fixa. A forma de seleção das unidades amostrais, na população florestal, determinou as diferenças entre essas populações. Os volumes utilizados para realização do presente estudo foram coletados em 130 parcelas de 525 m², numa área de 986,78 hectares plantada com *Eucalyptus grandis*. A partir da análise da normalidade das diferenças em cada distância do semivariograma experimental permitiu-se assumir um modelo multivariado normal para o processo estocástico. Para a realização das simulações estatísticas, consideraram-se duas situações: (i) a floresta como população gerada pela realização de um processo estocástico, e (ii) como uma população fixa. No caso de população estocástica, as simulações foram efetuadas em diferentes grades e em parcelas aleatorizadas. Para cada

uma destas situações, as simulações geraram os intervalos de confiança clássico e espacial. Posteriormente, foram comparadas as coberturas observadas com a cobertura nominal. Esta comparação permitiu estabelecer uma correção para o valor de t de Student, utilizado na definição do intervalo de confiança da média espacial e de sua variância. No caso de população estocástica, o comportamento entre os dois estimadores foi semelhante nas duas configurações amostrais avaliadas. Nesta situação, a aleatorização não atenuou o efeito da dependência espacial. Na configuração amostral sistemática, na grade de 700 metros, a cobertura observada igualou a cobertura nominal. A distância entre parcelas nesta grade foi superior ao alcance do modelo espacial, mostrando, portanto, que para distâncias superiores ao alcance, os dois estimadores foram semelhantes. Estes resultados evidenciaram que não é indicado fazer uso do Teorema Central do Limite para geração do intervalo de confiança, sem que nenhuma correção possa ser efetuada. Os resultados mostraram que, em população fixa, independentemente da estrutura de dependência espacial do fenômeno, a aleatorização proporciona condições de se usar o Teorema Central do Limite para geração do intervalo de confiança.

Palavras-chave: *Eucalyptus grandis*, simulação, geoestatística baseada em modelo, inventário florestal

Summary

Wood volume of a forest stand (population) is determined departing from a sample, which is selected through sampling procedure. The sampling procedures commonly used in forestry inventories were developed by the classical sampling theory. The development of this estimator theory was based upon fix population concepts. The use of these estimators in a stochastic population may increase statistical problems. This research developed a methodology of forestry inventory that considers the volume dependence spatial structure. Specifically, it evaluated the observed coverings of the confidence interval of the geostatistic and classic estimators for stochastic and for fixed population. For selecting samples units, in the forest population, the differences among

them were determined. Data were collected in 130 plots of 525 m² each, in 986.78 hectares of *Eucalyptus grandis* plantation. The normality of the differences in each distance of the experimental semi variogramme analysis allowed assuming a multivariate normal model for the stochastic process. For the statistical simulations, the whole forest was considered as a population generated by the stochastic process and as a fixed population. For stochastic population, the simulations were done in different grating and in randomized portions. For each situation, the simulations generated confidence intervals. After, the observed covering and the nominal covering were compared, allowing the correction of the value of t of Student, used in the definition of the confidence interval using its spatial average and variance. In the case of stochastic population, the behavior of the two estimators was similar for the two appraised sampling schedule. In this case, randomization didn't lessen the spatial dependence effect. In the systematic sampling configuration, departing from the grating of 700 meters, the observed covering equaled the nominal covering. The distance among plots, in this grating, was superior to the distance that the spatial model could reach, showing, therefore, that for long distances the two estimators were similar. Thus, it is not indicated to use the Central Limit Theorem for generating confidence interval, if no correction can be done. The results showed that, in fix population, independently of the dependence of the spatial structure of the characteristic evaluated, randomization provides the necessary conditions for using the Central Limit Theorem, for generating the confidence interval; however, it does not lessen the spatial dependence effect.

Key-words: *Eucalyptus grandis*, simulations, geostatistic based model, forestry inventory

5.1 Introdução

Um dos objetivos básicos do inventário florestal é estimar o volume total de madeira existente em uma população florestal, sendo a estimativa efetuada a partir de uma amostra retirada desta população. As unidades amostrais são selecionadas conforme o procedimento de um delineamento amostral. Tais procedimentos, comumente empregados nos inventários florestais, foram desenvolvidos com base na teoria clássica

de amostragem (Péllico Netto & Brena, 1997). O desenvolvimento teórico da amostragem clássica foi efetuado com base no conceito de populações fixas, ou seja, aquelas cujo conjunto de unidades amostrais é finito e que, para cada unidade amostral, o valor da variável avaliada é fixada (Thompson & Seber, 1996; Hoef, 2002).

Os procedimentos de amostragem mais utilizados nos inventários florestais são: casual simples, casual estratificado e sistemático. Segundo Matérn (1960) e Cressie (1993), estes procedimentos também são considerados na amostragem espacial. Os estimadores dos procedimentos de amostragem casualizados foram construídos com base nas probabilidades geradas pela aleatorização dos mesmos sobre a população fixa (Brus & Gruijter, 1997). A descrição detalhada de cada um destes procedimentos pode ser encontrada em Cochran (1977), Thompson (1992), Péllico Netto & Brena (1997) e Scolforo & Mello (1997), dentre outros.

Na estimativa de parâmetros populacionais em estudos ecológicos, especificamente em inventários florestais, o uso de procedimentos aleatórios somente aparece descrito na metodologia empregada nos levantamentos. Segundo Aubry & Debouzie (2001), em uma análise feita em cinco periódicos da área de ecologia, em 60% dos trabalhos que envolviam levantamento, foram utilizados procedimentos de amostragem sistemático. Nos inventários florestais este percentual deve ser maior, e a sistematização não ocorre necessariamente na forma de uma grade regular. Na grande maioria das amostragens, efetuadas em florestas plantadas, há uma restrição da aleatoriedade visando melhor cobertura espacial das parcelas. Cochran (1977) denomina este tipo de distribuição das parcelas como “Amostragem Sistemática Desencontrada”. O procedimento de amostragem mais utilizado nos inventários de florestas plantadas é o estratificado. Porém, em cada estrato as parcelas são lançadas de forma a cobrir o máximo da variação espacial. Percebe-se, portanto, o caráter sistemático na distribuição das parcelas nos levantamentos florestais. Os estimadores utilizados, tanto para a amostragem sistemática quanto para a amostragem estratificada, foram desenvolvidos para amostragem casual simples sobre uma população fixa.

Segundo Aubry & Debouzie (2000), o uso do estimador de variância da média da amostragem casual simples para a amostragem sistemática aumenta os problemas

estatísticos, uma vez que na amostragem sistemática percebe-se melhor a dependência entre unidades amostrais (correlação espacial). A estimativa de \bar{z} da média populacional, tanto na ausência quanto na presença de correlação, não tem tendência (viés). Porém, o estimador clássico para o cálculo da variância da média em dados com correlação conduz a um viés substancial (Koch & Gillings, 1983). Segundo estes autores, na ausência de correlação, os dois estimadores proporcionam valores iguais de variância da média. A fórmula apresentada por Cochran (1977) e Thompson (1992) para estimativa da variância da média na amostragem sistemática também deixa implícita a presença de viés no cálculo da variância da média, ao considerar a correlação entre unidades amostrais:

$$s_{\bar{x}}^2 = \left[\frac{s^2}{n} \left(1 - \frac{n}{N} \right) \right] \cdot [1 + (n-1) \cdot r] \quad (1)$$

em que σ^2 é a variância populacional; n tamanho da amostra; N é o número total de parcelas cabíveis na área (população fixa) e r é o coeficiente de correlação entre as parcelas. Conforme descrito por Cochran (1977), a presença de correlação positiva entre unidades da amostra inflaciona a variância da média. Mesmo uma pequena correlação positiva pode ter um grande efeito por causa do multiplicador $(n-1)$.

Segundo Thompson & Seber (1996), as situações amostrais no espaço, nas áreas ambiental e ecológica, estão além da estrutura conceitual da amostragem clássica apresentada até o momento. A variável de interesse é definida sobre uma região contínua no espaço (“população estocástica”), ao invés de ser um conjunto finito de unidades amostrais (“população fixa”). A primeira situação refere-se a um processo estocástico, onde a abordagem apropriada para a amostragem deve ser baseada em modelo. A segunda trata-se de uma população com unidades amostrais fixas. Neste segundo caso, a abordagem mais apropriada da amostragem seria com base no delineamento (teoria clássica). Segundo Koch & Gillings (1983), as duas abordagens têm o mesmo objetivo: estimar corretamente a variância da média. A escolha entre uma abordagem e outra deve ser realizada observando-se as fontes de aleatorização envolvidas na amostragem.

Na abordagem baseada no delineamento, a estocasticidade é introduzida no estágio da amostragem e os valores são fixados. Os locais amostrados são selecionados por um procedimento de seleção aleatório pré-determinado, definido pelo procedimento de amostragem, o qual determina a probabilidade de seleção para cada amostra. Esta fonte de aleatorização permite construir estimadores sem viés da variância da média baseado somente na probabilidade do esquema de amostragem (Bruss & Guijter, 1993). Uma vez selecionada a amostra, determina-se a média espacial. O que aconteceria se outra amostra de tamanho igual fosse retirada na mesma área, usando o mesmo esquema de amostragem e o mesmo estimador? Esta é a questão colocada na abordagem por delineamento.

Na abordagem baseada em modelo, a população é definida por um modelo de processo estocástico. A amostra é formada como uma observação de uma realização do processo, dentre um número infinito de possíveis realizações. As unidades amostrais podem ser aleatorizadas de duas formas: os locais amostrados podem ser aleatorizados e os valores fixados (realização), ou os locais são fixados e os valores são variáveis (processo estocástico). Nesta abordagem as variáveis são aleatórias por causa do modelo probabilístico. Nota-se que estas variáveis aleatórias diferem daquelas geradas pela replicação do procedimento (baseada no delineamento) e elas não são necessariamente independentes (independência estatística). Entretanto, na abordagem baseada em modelo, a amostra não necessita ser selecionada por um esquema de amostragem probabilístico. Qualquer esquema de amostragem pode ser adotado, tal como a amostragem sistemática, freqüentemente utilizada nos inventários florestais. Nesta abordagem, a estimativa da variância da média é inteiramente baseada no modelo estocástico ajustado para a floresta.

Referindo-se a levantamentos florestais, a escolha de qual tipo de população deve ser utilizada na amostragem pode ser definida na seguinte questão: “como é efetuado o procedimento de amostragem no campo? São selecionadas por sorteio as unidades amostrais na população fixa ou selecionadas posições onde serão lançadas as parcelas?” As respostas para estas questões são fundamentais na escolha do “tipo de população” no contexto de amostragem. Nos inventários florestais são selecionadas as posições onde

são lançadas as unidades amostrais. Portanto, intuitivamente, tratou-se a população como uma população espacialmente contínua ou estocástica no processo de amostragem. Porém, as análises são efetuadas com estimadores construídos sobre uma população fixa. Mediante o exposto, adotou-se a abordagem baseada em modelo para efetuar inferências a respeito dos parâmetros populacionais. Segundo Thompson & Seber (1996), no caso de populações estocásticas, somente a abordagem baseada em modelo é apropriada.

O método para análise dos inventários florestais, na atualidade, fundamenta-se na teoria clássica de amostragem, ou seja, àquela desenvolvida para populações fixas. Assim, o intervalo de confiança é gerado a partir do Teorema Central do Limite, o qual trata as amostras de forma independente. Este capítulo teve como objetivo geral desenvolver um método de análise do inventário florestal que considere a estrutura de continuidade espacial, da característica volume na definição do intervalo de confiança. Especificamente pretendeu-se:

- comparar, através de simulações, as coberturas observadas nos intervalos de confiança dos estimadores geoestatístico e clássico com a cobertura nominal, para os procedimentos de amostragem casual simples e sistemático, numa população contínua;
- avaliar, através de simulações, as coberturas observadas dos intervalos de confiança gerados pelos estimadores geoestatístico e clássico, a partir de amostras aleatórias retiradas de uma população fixa.

5.2 Material e Métodos

5.2.1 Descrição da área e obtenção dos dados

Os dados de volume por parcela para realização do trabalho foram coletados num povoamento de 986,78 hectares de *Eucalyptus grandis*, pertencente a Votorantim Celulose e Papel. O povoamento localizou-se no município de Itapetininga, na região Sul do estado de São Paulo, apresentando localização geográfica 23°56' de latitude sul e 42°21' de longitude oeste, 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 para o estudo, dados coletados em parcelas de 525 m² (25 X 21m), sistematizadas em toda a área, distantes entre si de 500 metros. Com o objetivo de melhor estimar o efeito pepita, foram selecionados três pontos na área para a realização de uma amostragem mais intensa. Ao todo foi estimado o volume por hectare de 130 parcelas. No contexto de análise exploratória, foram avaliados a normalidade dos dados originais, bem como a normalidade das diferenças em cada distância do semivariograma experimental. O teste utilizado para testar a normalidade foi o Jarque.Bera conforme descrito por Jarque & Bera (1980). Esta avaliação serviu como indicativo para assumir um Modelo Multivariado Normal para o processo.

5.2.2 Modelo Geoestatístico

Considerando que as observações medidas para o volume nos diferentes pontos amostrais apresentaram-se estruturadas espacialmente, obteve-se um vetor multidimensional de observações correlacionadas em duas dimensões espaciais, as quais foram avaliadas por um modelo linear misto

$$z_i = \mathbf{m} + s(x_i) + \mathbf{e}_i \quad (2)$$

onde:

z_i : valor observado no ponto i ;

\mathbf{m} : constante determinística na condição da hipótese intrínseca;

$s(x_i)$: componente estocástica que explica o comportamento espacial da variável Y em toda a área, sendo x_i a localização do ponto i ;

\mathbf{e}_i : erro aleatório com distribuição normal e variância τ^2 .

$S(\cdot)$ é um processo Gaussiano estacionário com $E[S(x)] = \mathbf{m}$ $VAR[s(x)] = \sigma^2$ e $r(h) = CORR[s(x), s(x+h)]$, sendo h o deslocamento em relação à posição x . A variável

resposta Z tem uma distribuição conjunta multivariada normal (Diggle & Ribeiro Júnior, 2000),

$$Z \sim MVN(\mathbf{m}, \mathbf{s}^2 \Sigma + \mathbf{t}^2 I) \quad (3)$$

onde:

$\mathbf{1}$: vetor de valores 1

σ^2 : variância populacional

Σ : matriz de correlação ou de semivariância

I : matriz identidade.

τ^2 : variância do erro aleatório.

A função de densidade da variável resposta Z é definida pela seguinte expressão:

$$f(\underline{z}) = \frac{1}{(2\mathbf{p})^{n/2} |\mathbf{K}|^{1/2}} e^{\left[-\frac{1}{2} (\underline{z} - \mathbf{m})' \mathbf{K}^{-1} (\underline{z} - \mathbf{m}) \right]} \quad (4)$$

onde:

\underline{z} : vetor da variável resposta

$\underline{\mathbf{m}}$: vetor de média (constante na condição da hipótese intrínseca)

\mathbf{K} : igual a, $\mathbf{s}^2 \Sigma + \mathbf{t}^2 I$, matriz de covariância ou de semivariância em função de h .

Na presença de correlação a matriz de semivariância é definida por uma função de dependência espacial, a partir da distância (h) entre as observações. No presente estudo, o modelo selecionado para expressar a dependência espacial foi o exponencial, tendo sido esse modelo ajustado previamente aos dados coletados pelo método de Máxima Verossimilhança.

$$\mathbf{g}(h) = 382,9987 + 831,6125 \left[1 - \exp\left(-\frac{h}{233} \right) \right] \quad (5)$$

onde:

$g(h)$: é a semivariância entre dois pontos;

$t^2 = 382,9987$ (efeito pepita “nugget”);

$s^2 = 831,6125$ (contribuição “sill partial”);

$q = 233$ (alcance “range”)

5.2.3 Simulação Estocástica

A utilização da simulação como ferramenta de análise permite realizar diversos cenários num mesmo ponto ou em toda a área (Braga, 1990). O objetivo da simulação geoestatística é reproduzir a Função Aleatória, cujas características foram identificadas na análise estrutural. As simulações reproduzem, portanto, a mesma variabilidade dos dados sob a área de estudo.

A simulação geoestatística, no presente estudo, gerou 20000 amostras a partir do modelo espacial de volume, definido para a floresta através dos pontos amostrados. Para a realização das simulações (20000), considerou-se a floresta ora como uma população estocástica ora como fixa. No caso da população estocástica, trabalhou-se a simulação em dois níveis: grades – pontos sistematizados e pontos aleatórios (Tabela 1). Nesta situação, os locais foram fixados e os valores variaram, conforme a estocasticidade do modelo geoestatístico. Utilizou-se o programa R, através do pacote geoR (Ribeiro Júnior & Diggle, 2001), para efetuar as simulações estocásticas. A função usada no geoR foi a `grf`, que simula campos aleatórios gaussianos dado um modelo espacial. Os campos aleatórios foram gerados a partir do produto da raiz quadrada da matriz K por um vetor de erros aleatórios, que são independentes e distribuídos identicamente. Desta forma, têm-se campos aleatórios multivariados normais e com a mesma estrutura de correlação (Cressie, 2003; Ribeiro Júnior & Diggle, 2001).

Tabela 1. Tamanho de grades e número de parcelas aleatorizadas, avaliadas no processo de simulação

| GRADE (metros) | Número de parcelas aleatorizadas (n) |
|-------------------|---|
| 200 x 200 | 268 |
| 300 x 300 | 106 |
| 400 x 400 | 66 |
| 500 x 500 | 46 |
| 600 x 600 | 33 |
| 700 x 700 | 22 |
| 800 x 800 | 12 |
| 900 x 900 | 11 |

Para efetuar a abordagem por delineamento para variáveis contínuas no espaço, denominada de população fixa, foi simulado o mesmo número de parcelas aleatórias apresentadas na Tabela 1. Thompson & Seber (1996) mostraram que é necessário estabelecer 3 conexões para utilizar a teoria clássica de amostragem no contexto de amostragem espacial. Desta forma:

- a partir da área total e da área da parcela, definiu-se o número total de parcelas cabíveis na área. Ao todo foram 19750 parcelas (N). Cada parcela teve a sua coordenada de referência;
- através do modelo espacial definido para a floresta, utilizando-se o pacote RandomFields, que simula campos aleatórios gaussianos, através de um modelo com seus parâmetros conhecidos, gerou-se o valor de volume para cada uma das 19750 parcelas. Cada ponto teve a sua respectiva coordenada e o valor do volume correspondente;
- a aplicação da teoria clássica de amostragem para população fixa foi estendida ao padrão espacial fixo, no qual a análise baseada em delineamento foi efetuada condicional à realização específica de um modelo. A estocasticidade surgiu nas

diferentes realizações da amostragem aleatória sobre os valores fixos de volume em cada parcela.

A estes números aleatórios gerados soma-se a média espacial (μ), que é a parte fixa do modelo. Realizaram-se 20000 simulações para cada um dos diferentes cenários apresentados na Tabela 1, bem como para as parcelas aleatorizadas na população fixa.

5.2.4 Processamento dos dados simulados

O processamento dos dados simulados teve como objetivo avaliar o desempenho dos intervalos de confiança clássico e geoestatístico, para a amostragem baseada em modelo ou em delineamento. O desempenho foi analisado a partir da comparação entre a cobertura nominal e a cobertura observada para cada uma das 20000 simulações. Para cada simulação, por modelo ou por delineamento, obteve-se a média espacial estimada ($\hat{\mathbf{m}}$), a variância estimada da média espacial [$V\hat{A}R(\hat{\mathbf{m}})$], a média aritmética (\bar{z}) e a estimativa da variância da média [$V\hat{A}R(\bar{z})$]. Estas estimativas permitiram gerar os intervalos de confiança clássico e geoestatístico. O intervalo de confiança clássico foi gerado com a seguinte expressão:

$$\bar{z} \pm t_{\left(\frac{\alpha}{2}; n-1\right)} \sqrt{V\hat{A}R(\bar{z})} \quad (6)$$

onde $1 - \alpha$ é o coeficiente de confiança, sendo testados os coeficiente de 90, 91, 92, 93, 94, 95, 96, 97, 98 e 99%.

Para o intervalo geoestatístico, a fórmula clássica sofreu algumas mudanças:

$$\hat{\mathbf{m}} \pm t_{\left(\frac{\alpha}{2}; n-1\right)} \sqrt{V\hat{A}R(\hat{\mathbf{m}})} \quad (7)$$

O valor t de Student, tanto para o intervalo clássico quanto para o geoestatístico, foi considerado como t tabelado para amostras independentes, conforme o nível de probabilidade (cobertura nominal) e a intensidade amostral (n). Avaliou-se, para cada um dos 20000 intervalos clássicos e geoestatísticos, quantos deles abrangiam a média espacial do modelo ($\hat{\mathbf{m}}$), para o caso de população contínua, obtendo assim o percentual de cobertura observado do intervalo clássico e do intervalo geoestatístico. Para o caso de população fixa, após a geração dos 19750 valores de volume, determinou-se a média populacional (\bar{Z}). Posteriormente, verificaram-se quantos dos 20000 intervalos clássicos e geoestatísticos abrangiam a média populacional. A seguir, comparou-se a cobertura observada com a nominal para cada nível de probabilidade. Esta comparação permitiu inferir sobre a eficiência da cobertura observada contra a nominal.

Como descrito no item 5.1 deste capítulo, o objetivo foi utilizar a fórmula da teoria clássica com as estimativas espaciais para gerar o intervalo de confiança. Porém, não se conhece qual a distribuição que $\hat{\mathbf{m}}$ assume e, portanto, não se sabe qual a estatística a ser utilizada no lugar de t. Para tal, após a comparação entre cobertura observada e a nominal, nas diferentes grades e nas amostras casualizadas, estabeleceu-se uma relação entre o valor t e o percentual de cobertura observada obtida nas simulações. Com isto, foi possível gerar um modelo linear, onde a variável dependente foi o valor t e a variável independente a cobertura nominal. Com este modelo efetuou-se a estimativa do valor t, a fim de se obter a melhor aproximação da cobertura observada com a nominal.

Com os volumes por hectare obtidos no experimento, processou-se o inventário florestal por meio dos estimadores clássico e geoestatístico. Para o estimador clássico, obteve-se a média aritmética dos volumes, o desvio padrão da média e o intervalo de confiança. As estatísticas geradas pelo estimador geoestatístico, para a determinação do intervalo de confiança, foram a média espacial ($\hat{\mathbf{m}}$) e a variância da média espacial [$\widehat{VAR}(\hat{\mathbf{m}})$], respectivamente por:

$$\hat{\mathbf{m}} = (\mathbf{f}'\Sigma^{-1}\mathbf{f})^{-1}(\mathbf{f}'\Sigma^{-1}\mathbf{z}) \quad (8)$$

$$V\hat{AR}(\hat{\mathbf{m}}) = (\mathbf{f}'\Sigma^{-1}\mathbf{f})^{-1} \quad (9)$$

onde:

\mathbf{f} : vetor de 1's (média constante)

\mathbf{K} : matriz de covariância ou semivariância

\mathbf{z} : vetor de dados observados.

5.3 Resultados e Discussão

5.3.1 Simulação com base em população estocástica

Nas Tabelas 2 e 3 apresenta-se o percentual de cobertura observada para os estimadores geoestatístico e clássico, considerando-se grades de 200 a 900 metros para a configuração amostral sistemática. Para a configuração aleatória, aleatorizou-se o mesmo número de parcelas conforme o número de parcelas das grades.

Verificou-se que os valores de cobertura observada, gerados pelos estimadores geoestatístico (EG) e clássico (EC), no processo de simulação, na grade 700 (Tabela 2) e de 22 parcelas aleatorizadas (Tabela 3), tiveram comportamento semelhante em relação à cobertura nominal. Para grades menores que 700 metros, observou-se maior diferença entre os valores gerados pelos estimadores geoestatístico e clássico. Quanto menor a grade (maior número de parcelas), maiores são as diferenças. Nas duas configurações amostrais avaliadas, o estimador geoestatístico sempre proporcionou valores de cobertura observada mais próximos da nominal, independente da configuração amostral (Figura 1). A semelhança da cobertura observada do estimador geoestatístico entre as duas configurações amostrais revelou que a aleatorização não atenuou o efeito da dependência espacial sobre a estimativa da variância da média. Na Figura 1 pode-se observar o comportamento dos dois estimadores, nas duas configurações amostrais, dentro do nível de significância de 95%. Este comportamento também foi verificado nos demais níveis de significância avaliados. Verificou-se um crescimento suave da cobertura observada, gerada pelo estimador geoestatístico, para a amostragem

sistemática (Figura 1a), e um decréscimo suave para a amostragem aleatória (Figura 1b). Na cobertura observada do estimador clássico, o crescimento foi acentuado até a grade de 700 metros, apresentando forte decréscimo a partir de 22 parcelas aleatorizadas.

Tabela 2. Percentual da cobertura observada para os estimadores geoestatístico (EG) e clássico (EC) na configuração amostral sistemática

| GRADE | ESTIMADOR | COBERTURA NOMINAL | | | | | | | | | |
|-------|-----------|-------------------|----|----|----|----|----|----|----|----|----|
| | | 90 | 91 | 92 | 93 | 94 | 95 | 96 | 97 | 98 | 99 |
| 200 | EG | 86 | 87 | 89 | 90 | 91 | 92 | 94 | 95 | 96 | 97 |
| | EC | 54 | 56 | 57 | 59 | 60 | 62 | 64 | 67 | 70 | 74 |
| 300 | EG | 87 | 88 | 90 | 91 | 92 | 93 | 94 | 95 | 96 | 97 |
| | EC | 71 | 73 | 74 | 76 | 78 | 80 | 82 | 84 | 87 | 91 |
| 400 | EG | 86 | 87 | 88 | 89 | 90 | 92 | 93 | 94 | 95 | 97 |
| | EC | 79 | 80 | 81 | 83 | 84 | 86 | 88 | 90 | 92 | 95 |
| 500 | EG | 88 | 89 | 90 | 91 | 92 | 93 | 94 | 96 | 97 | 98 |
| | EC | 84 | 85 | 86 | 87 | 89 | 90 | 92 | 94 | 95 | 97 |
| 600 | EG | 89 | 90 | 91 | 92 | 93 | 94 | 95 | 96 | 97 | 98 |
| | EC | 86 | 87 | 88 | 89 | 91 | 92 | 93 | 95 | 96 | 98 |
| 700 | EG | 89 | 90 | 91 | 92 | 93 | 94 | 95 | 96 | 98 | 99 |
| | EC | 87 | 88 | 89 | 90 | 91 | 93 | 94 | 95 | 97 | 98 |
| 800 | EG | 90 | 90 | 92 | 93 | 93 | 95 | 96 | 97 | 98 | 99 |
| | EC | 89 | 90 | 91 | 92 | 93 | 94 | 95 | 96 | 98 | 99 |
| 900 | EG | 90 | 91 | 92 | 93 | 94 | 95 | 96 | 97 | 98 | 99 |
| | EC | 89 | 90 | 91 | 93 | 94 | 95 | 96 | 97 | 98 | 99 |

Tabela 3. Percentual da cobertura observada para os estimadores geoestatístico (EG) e clássico (EC) na configuração amostral aleatória

| TAMANHO DE AMOSTRA | ESTIMADOR | COBERTURA NOMINAL | | | | | | | | | |
|--------------------------|-----------|-------------------|----|----|----|----|----|----|----|----|----|
| | | 90 | 91 | 92 | 93 | 94 | 95 | 96 | 97 | 98 | 99 |
| 268 | EG | 86 | 87 | 88 | 90 | 90 | 92 | 93 | 94 | 95 | 97 |
| | EC | 31 | 31 | 32 | 33 | 34 | 35 | 36 | 38 | 40 | 43 |
| 106 | EG | 87 | 88 | 90 | 91 | 92 | 93 | 94 | 95 | 97 | 98 |
| | EC | 71 | 72 | 74 | 76 | 77 | 79 | 81 | 84 | 87 | 91 |
| 66 | EG | 87 | 88 | 90 | 91 | 92 | 93 | 94 | 95 | 97 | 98 |
| | EC | 77 | 79 | 80 | 82 | 83 | 85 | 87 | 89 | 92 | 95 |
| 44 | EG | 88 | 89 | 90 | 91 | 92 | 93 | 94 | 95 | 96 | 98 |
| | EC | 81 | 82 | 83 | 85 | 86 | 88 | 90 | 92 | 94 | 96 |
| 33 | EG | 88 | 89 | 90 | 91 | 92 | 93 | 94 | 95 | 97 | 98 |
| | EC | 83 | 84 | 86 | 87 | 89 | 90 | 92 | 93 | 94 | 97 |
| 22 | EG | 88 | 89 | 90 | 91 | 92 | 94 | 95 | 96 | 97 | 98 |
| | EC | 85 | 86 | 88 | 89 | 90 | 92 | 93 | 95 | 96 | 98 |
| 12 | EG | 89 | 90 | 91 | 92 | 93 | 94 | 95 | 96 | 97 | 99 |
| | EC | 87 | 88 | 89 | 91 | 92 | 93 | 94 | 96 | 97 | 98 |
| 11 | EG | 90 | 90 | 91 | 92 | 93 | 95 | 96 | 96 | 97 | 99 |
| | EC | 87 | 89 | 91 | 91 | 92 | 94 | 95 | 96 | 97 | 99 |

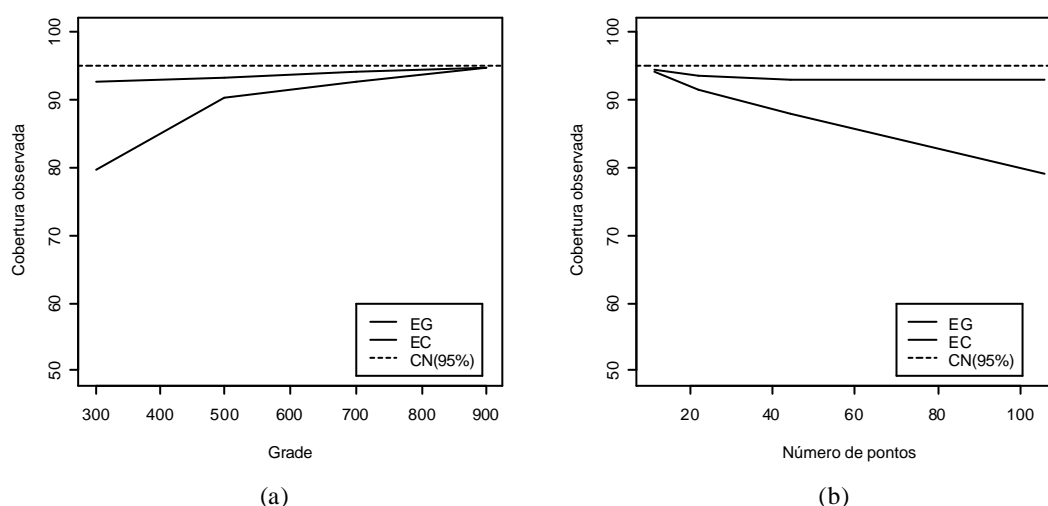


Figura 1 - Comportamento dos intervalos de confiança dos estimadores clássico (EC) e geoestatístico (EG) para coeficiente de confiança de 95% (CN – cobertura nominal), para as configurações amostrais sistemática (a) e aleatória (b)

Analisando o percentual de cobertura dos dois estimadores, nas diferentes grades, verificou-se que ambos geraram o mesmo resultado à medida que se aumentou a distância entre os pontos da grade. Além do mais, a cobertura observada praticamente se igualou à cobertura nominal nas maiores grades e na condição de poucas parcelas por amostra (Figura 1). Este comportamento foi detectado em todos os níveis de significância avaliados. Essa tendência ocorreu porque, nas grades com distância entre pontos maior do que o alcance prático (699 metros), o efeito da correlação entre parcelas foi quase nulo, ou seja, na condição de não existência de dependência espacial, os dois estimadores gerariam o mesmo resultado da estimativa do intervalo de confiança, o que concorda com os resultados obtidos por Koch & Gillings (1983). Na configuração aleatória, observou-se o mesmo comportamento, quanto menor o número de parcelas na área, menor o efeito da correlação entre parcelas sobre a estimativa da variância da média.

Esta diferença desfavorável ao estimador clássico foi devido à falta de um termo que considerasse a correlação entre as unidades amostrais no cálculo da variância da média. O estimador geoestatístico, ao considerar a correlação entre as unidades amostrais, proporcionou estimativas mais realistas da variância da média, nas duas configurações amostrais avaliadas, o que certamente contribuiu para o aumento da cobertura observada em relação à nominal. Resultados semelhantes foram obtidos por Koch & Gillings (1983), quando compararam a variância da média, obtida pela formulação clássica, com a variância geoestatística. Os autores detectaram um viés substancial da variância da média, gerada pela formulação clássica, em dados ecológicos que apresentaram correlação espacial. No presente estudo, o viés provocou uma subestimativa da variância da média.

O comportamento das coberturas observadas para os estimadores geoestatístico e clássico foi semelhante nas duas configurações amostrais avaliadas em cada intensidade amostral. Na grade 200 X 200, que se refere a 268 parcelas aleatorizadas, o efeito da correlação foi mais acentuado, provocando maior viés na estimativa da variância da média, e destacando a diferença entre as coberturas observadas dos estimadores geoestatístico e clássico. Estes resultados estão em consonância com a teoria da

abordagem baseada em modelo, onde a amostra não necessita ser selecionada por um esquema de amostragem probabilístico, uma vez que a variância da média é inteiramente baseada no modelo estocástico (Aubry & Debouzie, 2001). Pelos resultados do estudo de uma população contínua com dependência espacial, foi possível verificar que a cobertura observada foi inferior à cobertura nominal até a distância correspondente ao alcance do modelo. Assim, pode-se concluir que não foi possível utilizar o teorema central do limite para a geração do intervalo de confiança.

5.3.2 Simulação com base em população fixa

É importante ressaltar que a população espacialmente fixa foi gerada do modelo espacial do volume para a floresta estudada, que foi o mesmo que gerou as simulações da população estocástica. Portanto, todas as análises foram condicionadas a uma única realização do modelo. Na Tabela 4 é possível avaliar algumas informações sobre as simulações das diferentes amostras retiradas da população fixa, para cada uma das intensidades amostrais. O percentual de amostras sem dependência espacial, retiradas da população, variou de zero a 32,7%. O percentual de dependência espacial foi determinado através da razão entre a contribuição (σ^2) e o sill total ($\sigma^2 + \tau^2$), para cada amostra retirada da população fixa. À medida que o número de parcelas foi diminuindo nas amostras simuladas, verificou-se aumento do percentual de amostra sem dependência espacial. Com um número maior de parcelas em cada amostra, há uma maior probabilidade da distância média entre as parcelas da amostra ser menor do que o alcance do modelo (699m). Desta forma, as amostras apresentaram-se estruturadas espacialmente. Porém, para as situações com menor número de parcelas em cada amostra simulada, aumentou-se a probabilidade da distância média entre parcelas ser maior do que o alcance e, conseqüentemente, houve um maior número de amostras sem dependência espacial. Se a população fixada foi gerada por um modelo com continuidade espacial, a aleatorização não significa, necessariamente, amostras sem dependência espacial (Tabela 4). Estes resultados estão de acordo com Särndal (1978), Brus & Gruijter (1997).

Tabela 4. Percentual de amostra sem dependência espacial e média do grau de dependência espacial para as amostras retiradas da população fixa, nas diferentes intensidades de amostragem

| NÚMERO DE PARCELAS ALEATORIZADAS | DE (%) | \bar{X} |
|---|---------------|-----------------------------|
| 268 | 0 | 68,5 |
| 106 | 0,3 | 70,9 |
| 66 | 2,1 | 73,0 |
| 44 | 6,0 | 74,7 |
| 33 | 9,9 | 74,8 |
| 22 | 17,7 | 72,2 |
| 12 | 30,1 | 65,1 |
| 11 | 32,7 | 63,4 |

Pela Tabela 5, analisa-se que a cobertura observada do estimador clássico foi igual à cobertura nominal para todas as intensidades amostrais simuladas. Isto mostra mesmo que, independente da estrutura de continuidade espacial do fenômeno, a aleatorização em população fixa, proporciona condições para se utilizar o Teorema Central do Limite na geração do intervalo de confiança.

Este resultado permite concluir que a teoria clássica de amostragem, amplamente utilizada nos inventários florestais, foi desenvolvida com base em população fixa. Portanto, a maneira como se aborda a questão da amostragem na floresta determinará qual o tipo de população. Se a floresta for tratada como população estocástica, a formulação da amostragem clássica provoca um viés na estimativa da variância da média. O estimador geoestatístico apresentou cobertura observada inflacionada em relação à cobertura nominal. Isto ocorreu, porque este estimador considera a questão da redundância de informação no cálculo da variância da média, quando há dependência espacial. A partir de um número pequeno de parcelas na amostra, os dois estimadores apresentaram cobertura observada igual. Neste caso, houve um percentual alto de amostras que não apresentaram estrutura de dependência espacial do volume (Tabela 4). Assim, na ausência de dependência espacial, os estimadores clássico e geoestatístico proporcionaram a mesma cobertura. Resultados semelhantes foram obtidos para população contínua, ou seja, na ausência de dependência espacial, pode-se utilizar o

Teorema Central do Limite na definição do intervalo de confiança, tanto para população fixa quanto estocástica. Para população fixa, pode-se gerar o intervalo de confiança, utilizando a formulação clássica, com ou sem dependência espacial.

Tabela 5. Percentual da cobertura observada para os estimadores geoestatístico (EG) e clássico (EC) na configuração amostral aleatória com base em população fixa

| N ^o DE AMOSTRA | ESTIMADOR | COBERTURA NOMINAL | | | | | | | | | |
|---------------------------|-----------|-------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| | | 90 | 91 | 92 | 93 | 94 | 95 | 96 | 97 | 98 | 99 |
| 268 | EG | 99 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | EC | 90 | 91 | 92 | 93 | 94 | 95 | 96 | 97 | 98 | 99 |
| 106 | EG | 99 | 99 | 99 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | EC | 90 | 91 | 92 | 93 | 94 | 95 | 96 | 97 | 98 | 99 |
| 66 | EG | 98 | 99 | 99 | 99 | 99 | 99 | 100 | 100 | 100 | 100 |
| | EC | 90 | 91 | 92 | 93 | 94 | 95 | 96 | 97 | 98 | 99 |
| 44 | EG | 97 | 97 | 98 | 98 | 98 | 99 | 99 | 99 | 100 | 100 |
| | EC | 90 | 91 | 92 | 93 | 94 | 95 | 96 | 97 | 98 | 99 |
| 33 | EG | 96 | 96 | 97 | 97 | 98 | 98 | 99 | 99 | 99 | 100 |
| | EC | 90 | 91 | 92 | 93 | 94 | 95 | 96 | 97 | 98 | 99 |
| 22 | EG | 94 | 95 | 95 | 96 | 97 | 97 | 98 | 98 | 99 | 100 |
| | EC | 90 | 91 | 92 | 93 | 94 | 95 | 96 | 97 | 98 | 99 |
| 12 | EG | 91 | 92 | 93 | 94 | 94 | 95 | 96 | 97 | 98 | 99 |
| | EC | 90 | 91 | 92 | 93 | 94 | 95 | 96 | 97 | 98 | 99 |
| 11 | EG | 91 | 92 | 93 | 94 | 95 | 95 | 96 | 97 | 98 | 99 |
| | EC | 90 | 91 | 92 | 93 | 94 | 95 | 96 | 97 | 98 | 99 |

5.3.3 Simulação estocástica no contexto do inventário florestal

Analisando a população florestal no contexto de população contínua, observou-se que, ao se considerar inventários com configuração amostral sistemática, através do estimador da amostragem casual simples, o nível de probabilidade de acerto assumido não foi verdadeiro na presença de autocorrelação espacial. Assim, ao assumir um certo nível de probabilidade, superestima-se a precisão final do levantamento. No presente estudo, adotou-se como referência a grade 300 X 300, uma vez que as parcelas permanentes estabelecidas nos plantios da área estudada, foram distribuídas de forma

“sistemática” com distância irregular de aproximadamente 300 metros uma da outra. Nesta situação, definindo-se o coeficiente de confiança de 95% para os intervalos, esse coeficiente seria na verdade de 80%. Esta precisão inflacionada pode comprometer a informação quantitativa que se pretende extrair de um levantamento. Este fato ocorreu porque não foi considerada a correlação entre unidades amostrais no cálculo da variância da média (Cochran, 1977; Thompson, 1992).

A cobertura real do estimador clássico foi sensivelmente inferior à cobertura nominal, principalmente nas menores grades da configuração sistemática e na configuração aleatória com alta intensidade amostral. Em ambas configurações amostrais avaliadas na presença de autocorrelação, quanto maior o número de parcelas maior a correlação entre elas e, portanto, maior viés na estimativa da variância. Esta situação revela a gravidade do problema quando se pensa no inventário pré-corte. Neste tipo de inventário, normalmente, lança-se um número maior de parcelas a fim de se obter uma melhor estimativa em cada talhão. Os resultados obtidos até o momento, revelaram que na presença de autocorrelação, amostrar mais não implica necessariamente uma melhor precisão. A fórmula para o cálculo da variância da média tem o número de parcelas no denominador ($\text{variância}/n$), isto é, quanto maior o número de parcelas, menor será a variância e, conseqüentemente, menor a amplitude do intervalo de confiança e maior a precisão. Porém, esta variância não é a verdadeira, uma vez que há redundância de informação sobre a variabilidade contida nas parcelas. A formulação clássica considera o número total de parcelas, pois as assume independentes e, portanto, não redundantes, mas na presença de continuidade espacial o número de parcelas que efetivamente carregam informação não redundantes é menor. Esta é a principal diferença entre os estimadores clássicos e geoestatístico.

A questão da redundância de informações pode ser analisada a partir da Figura 2. Esta teve como propósito relacionar o número de parcelas lançadas (n) com o número de parcelas úteis (n^*), na presença de continuidade espacial. O número de parcelas úteis foi gerado para a amostra sistemática e para a amostra casual simples, através da razão entre o patamar ($\tau^2 + \sigma^2$) e a variância da média espacial.

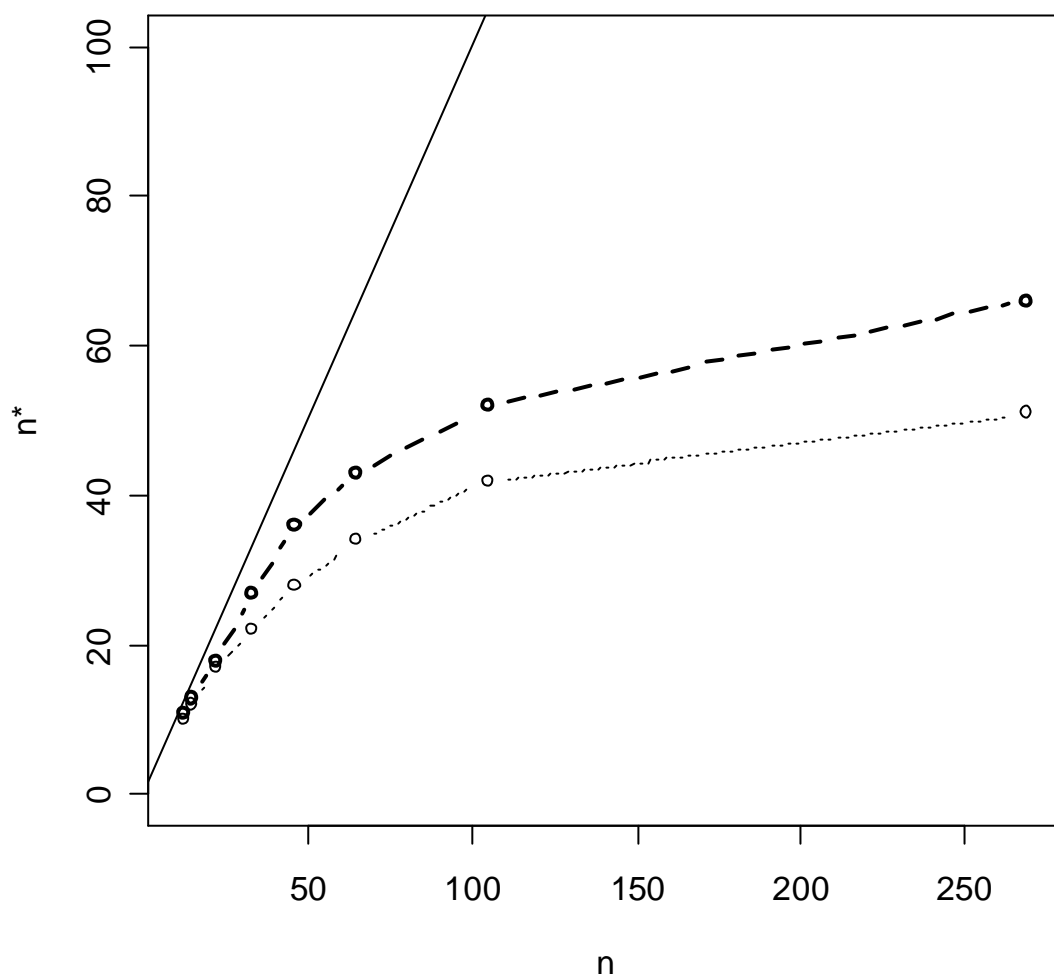


Figura 2 - Relação entre o número de parcelas lançadas e o número de parcelas úteis para o cálculo da variância da média na presença de continuidade espacial, para o procedimento de amostragem casual simples (linha pontilhada) e amostragem sistemática (linha tracejada)

Na ausência de continuidade espacial, os valores de n e n^* são iguais e, portanto, a relação entre estes valores coincide com a bissetriz traçada na Figura 2. A presença de

continuidade espacial revelou que, do número total de parcelas lançadas na área, parte delas não carregava informação sobre a variabilidade espacial da variável regionalizada. Esta parte provocou a redundância de informação, sendo responsável pela inflação da variância da média quando há continuidade espacial.

Pela Figura 2 foi possível verificar que o número de parcelas úteis (n^*) diferiu entre os dois procedimentos de amostragem, para um mesmo número de parcelas lançadas (n). Esta diferença mostrou que a amostragem sistemática utilizou mais parcelas no cálculo da variância da média, conseguindo, assim, expressar melhor a estrutura de continuidade espacial da variável regionalizada - volume.

Num padrão completamente aleatório, a distância parcela-parcela-mais-próxima segue a distribuição exponencial, com valor esperado inversamente proporcional à distância (Cressie, 1993). Desta forma, na amostragem casual simples houve maior concentração de parcelas separadas por pequenas distâncias, refletindo maior grau de continuidade espacial e conseqüentemente menor número de parcelas úteis (n^*).

5.3.4 Correção empírica para o intervalo de confiança do estimador geoestatístico

Para a utilização da formulação da estatística clássica na geração de intervalos de confiança a partir da média espacial e de sua variância, efetuou-se uma correção no valor t de tabela. Esta correção foi necessária, uma vez que a distribuição probabilística da média espacial não é conhecida. Assim, não foi possível definir uma estatística que pudesse quantificar a incerteza associada com as informações obtidas com a amostra. O modelo matemático apresentado a seguir permitiu corrigir o valor t para o coeficiente de cobertura (CC) desejado nos intervalos de confiança para a média espacial.

$$\hat{t} = -6,70141 + 0,09464 * CC$$

$$R^2 = 90\%$$

No presente estudo, independentemente da configuração utilizada, da intensidade amostral e do nível de significância, esta expressão pode ser usada para estimar o valor

\hat{t} na fórmula 7. Assim, atinge-se o valor nominal de cobertura para o intervalo geoestatístico.

Segundo Aubry & Debouzie (2001), a confiabilidade que pode ser colocada no intervalo de confiança, obtido na abordagem baseada em modelo, depende da estrutura espacial da população. Observe que este trabalho foi executado com base em uma população que tem sua própria estrutura espacial. Assim, a metodologia proposta, para ser aplicada em povoamentos de *Eucaliptus grandis* com estrutura espacial diferenciada, demanda novos estudos de simulações, os quais devem ser realizados a fim de se obter a função apropriada para estimativa do t . Sugere-se avaliar várias populações de *Eucaliptus grandis* na idade de corte com diferentes estruturas espaciais, com o objetivo de gerar uma tabela com valores de t em função dos parâmetros do modelo espacial.

5.4 Conclusões

- a) No caso de população estocástica espacialmente contínua, os estimadores geoestatístico e clássico apresentaram cobertura observada inferior à cobertura nominal, sendo os valores do estimador clássico menores do que os geoestatísticos. Esta observação foi constatada para os procedimentos de amostragem baseados em modelo.
- b) No caso de população estocástica, para as grades cuja distância entre parcelas foi maior do que o alcance, os estimadores clássico e geoestatístico apresentaram cobertura observada semelhante e próxima da cobertura nominal. Este fato mostrou que, na ausência de continuidade espacial, os estimadores proporcionaram cobertura similares. Esta observação também foi verificada para a amostragem casual simples na população fixa.
- c) Avaliando a cobertura observada em relação à nominal, em cada situação estudada, verificou-se que ambas tiveram comportamento semelhante nos dois procedimentos de amostragem baseados em modelo. Este resultado evidenciou que, para populações contínuas, o tipo de procedimento de amostragem não influenciou nos valores dos intervalos de confiança.

- d) Para populações fixas o estimador clássico proporcionou cobertura observada igual a cobertura nominal, em todas as situações avaliadas. Este resultado comprovou a validade da aleatorização para uso do Teorema Central do Limite, na definição do intervalo de confiança, independente da estrutura de continuidade espacial. Todavia, esta aleatorização não atenuou o efeito da dependência espacial.

6 USO DO ESTIMADOR GEOESTATÍSTICO PARA PREDIÇÃO VOLUMÉTRICA POR TALHÃO

Resumo

A estimativa de volume de madeira por talhão em um povoamento florestal é efetuada através do inventário pré-corte, sendo importante para o planejamento, colheita florestal e produção de celulose com qualidade. Para realização do inventário pré-corte, lançam-se parcelas numa intensidade amostral maior do que a do inventário florestal contínuo de cada talhão. Assim, além do custo de medição das parcelas do Inventário Florestal Contínuo, agrega-se um novo custo referente à medição das parcelas do Inventário Pré-Corte. A análise das informações do pré-corte é efetuada com base nos conceitos da teoria clássica de amostragem, sem consideração da estrutura de continuidade espacial. Uma alternativa para obtenção da informação volumétrica por talhão seria utilizar um método que fosse capaz de estimar o volume por talhão, através das parcelas permanentes, considerando a estrutura de continuidade espacial. A krigagem de bloco é o método de predição da Geoestatística capaz de predizer o volume por talhão. O presente estudo teve como objetivo comparar as estimativas efetuadas por talhão, quanto aos estimadores clássico e geoestatístico (krigagem de bloco). Os dados para realização do trabalho foram provenientes de cinco projetos florestais de *Eucalyptus grandis*, pertencentes a Votorantin Celulose e Papel, situados no município de Luís Antônio – SP. De cada projeto, obteve-se o volume total com casca das parcelas permanentes e as informações do inventário pré-corte. As informações do pré-corte foram processadas conforme os conceitos da teoria clássica de amostragem. Ajustou-se,

para cada um dos projetos um modelo espacial, o qual foi utilizado para efetuar a predição volumétrica por talhão, através da krigagem de bloco. Foram comparadas as informações das estatísticas do inventário pré-corte com as do estimador geoestatístico. A diferença média entre o volume total, considerando todos os projetos, gerado pelo estimador clássico e o estimador geoestatístico, foi de 1%. O erro amostral por talhão do estimador geoestatístico variou menos do que o erro amostral do estimador clássico. A amplitude do intervalo de confiança do estimador geoestatístico oscilou pouco entre os talhões avaliados, enquanto que esta oscilação no intervalo clássico foi alta. As estimativas do estimador geoestatístico foram muito semelhantes entre si. Considerando que a informação do pré-corte é atualmente a mais confiável, sugere-se o uso do estimador geoestatístico para estimativa por talhão, produzindo mais economia e precisão do que o inventário pré-corte, realizado segundo a teoria clássica de amostragem.

Palavras-chave: *Eucalyptus grandis*, inventário florestal, simulação, krigagem de bloco

Summary

The estimate of wood volume per stand, in a forest plantation, is made through the before felling inventory. This activity is important for planning, harvesting and pulp production with quality. For accomplishing the inventory a larger number of plots is established in the stand. Therefore, besides the cost of measuring the plots of the Continuous Forest Inventory, it is added the cost of the before felling forest inventory. The analysis of the information of the before felling inventory is made using classic statistics concepts, not considering the spatial continuity structure. One way of getting the volumetric information per stand would be to use a method capable of estimating the volume per stand using the information of the Continuous Forestry Inventory, considering the spatial continuity structure. Block Krigage is a geostatistic prediction method capable of predicting volume per stand. This method predicts the volume, considering the hole forest population spatial continuity to which the stands belong. This research compared the estimates made by stand using classic estimators and geostatistic

(block krigage) estimators. Data were collected in five *Eucalyptus grandis* stands, belonging Votorantim Pulp and Paper, located in Luís Antônio (SP) County. In each project, total volume with bark of the fixed plots and the information of the before felling inventory were obtained. The information of the before felling was processed according to the classic statistics concepts. For each project, a spatial model was adjusted for predicting volume per each stand, using block krigage. The statistics of the before felling inventory and those of the geostatistic estimator were compared. The average difference among the total volume, considering the five stands, generated by the classic estimator and the geostatistic estimator was 1%. The sampling error, per stand, of the geostatistic estimator varied less than that of the classic estimator. The width of the geostatistic estimator confidence interval varied very little among the appraised stands, while in the classic interval the variation was high. The estimates of the geostatistic estimator were very similar amongst themselves. Considering, therefore, that the information of the before felling inventory is the most reliable one, the use of the geostatistic estimator for volume estimates, per stand, is recommended, because it cost less and is more precise than the classical before felling forestry inventory.

Key-words: *Eucalyptus grandis*, forestry inventory, simulations, block kriging

6.1 Introdução

A importância da estimativa volumétrica por talhão numa empresa se deve a diversas razões. Dentre elas, destacam-se o setor de planejamento e colheita florestal e a qualidade da celulose. A formação do “mix” de madeira com diferentes densidades básicas proporciona celulose de boa qualidade, sendo que para isto são necessárias informações precisas sobre a predição volumétrica por talhão.

O acompanhamento do crescimento de uma floresta ocorre através das avaliações efetuadas sobre uma rede de parcelas permanentes, denominado de Inventário Florestal Contínuo (IFC). A intensidade amostral das parcelas permanentes varia de empresa para empresa (Péllico Netto & Brena, 1997; Scolforo & Mello, 1997). A manutenção e o processamento destas parcelas, durante o ciclo da floresta, tem um custo significativo.

Porém, as informações obtidas com o processamento destas parcelas não são suficientes para gerar um número seguro sobre o volume e sobre o erro do inventário por talhão na idade de corte. Assim, há necessidade de se efetuar o inventário pré-corte para obter as informações por talhão. Este inventário faz uso de um maior número de parcelas por talhão em relação ao Inventário Florestal Contínuo, agregando um novo componente ao custo da madeira posto fábrica. Deve-se considerar ainda que, se houver estrutura de continuidade espacial do volume, a precisão do inventário fica inflacionada (Aubry & Debouzie, 2001).

Uma outra alternativa, sem agregação de custo, seria o uso de um estimador que fosse capaz de efetuar a predição volumétrica com eficácia, em relação às parcelas permanentes remedidas no ano em que a floresta fosse cortada. A krigagem pode ser este estimador devido a vários motivos, destacando, principalmente, sua capacidade de considerar a autocorrelação entre as parcelas, a fim de aumentar a eficiência da estimativa. Krigagem é o interpolador geoestatístico para locais não amostrados, em combinação linear dos pontos amostrados, considerando pesos para cada um deles. São vários os tipos de krigagem (Isaaks & Serivastava, 1989). Segundo Yamamoto (2001), conforme o domínio que se deseja estimar, tem-se a krigagem pontual e a krigagem de bloco. Krigagem pontual é a estimativa ponto a ponto na área avaliada. Krigagem de bloco envolve estimativas de valores da variável regionalizada para uma sub-área da área total. A krigagem de bloco é apropriada (em relação a krigagem pontual), para situações em que a informação média é mais útil do que o valor exato no ponto (Burgess & Webster, 1980). A krigagem de bloco foi utilizada por Biondi et al. (1994) para mapear o diâmetro e a área basal de *Pinus ponderosa*. Conforme Trangmar et al. (1985), a krigagem é realizada a partir da média ponderada dos valores observados na vizinhança. Procedimento detalhado para a solução analítica, através das funções auxiliares, para a krigagem de bloco, pode ser visto em Clark (1979). Com o desenvolvimento da informática, a solução numérica pela simulação, é mais viável do que a solução analítica (Soares, 2000). Conforme Bussab & Morettin (2002), muitas vezes, mesmo construindo um modelo probabilístico, certas questões não podem ser resolvidas analiticamente ou, na melhor das hipóteses, bastante trabalhosa. Desta forma,

deve-se recorrer a estudos de simulação para obter aproximações de quantidades de interesse. A partir do exposto, o presente trabalho se propôs a desenvolver um método de análise para predição volumétrica por talhão de *Eucalyptus grandis*, considerando a estrutura de continuidade espacial da variável regionalizada. Especificamente, objetiva-se:

- avaliar a estrutura de dependência espacial da característica volume por hectare, para cinco projetos avaliados na idade de corte;
- ajustar um modelo geoestatístico que seja capaz de reunir toda a informação da estrutura de continuidade espacial do volume num único conjunto de parâmetros;
- a partir do modelo geoestatístico, efetuar a krigagem de bloco com simulação em cada talhão dos projetos avaliados, utilizando-se as parcelas permanentes, a fim de obter o volume médio e o erro de estimativa por talhão;
- comparar as medidas de precisão geradas pelo estimador geoestatístico e pelo inventário pré-corte (estimador clássico).

6.2 Material e Métodos

6.2.1 Descrição geral

Os dados, para realização do presente estudo, foram obtidos em cinco projetos florestais de *Eucalyptus grandis* (clone e semente), na idade de corte, pertencentes a Votorantim Celulose e Papel - VCP, situados no município de Luís Antônio – SP. A região localiza-se nas coordenadas 21°34'48'' de latitude Sul e 47°35'44'' de longitude Oeste, a uma altitude de 700 metros. A temperatura média da região é de 22°C e a precipitação total anual é de 1450mm. O solo predominante é Latossolo Vermelho Amarelo com relevo plano. As parcelas mensuradas foram as do Inventário Florestal Contínuo (IFC), estabelecidas entre dois e três anos de idade, e monitoradas anualmente até a idade de corte. Através das informações de volume por hectare de cada projeto na idade de corte, efetuou-se o estudo variográfico e a predição espacial por talhão. Na Tabela 1 está apresentado um resumo das principais características de cada projeto.

Tabela 1. Informações dos projetos utilizados para estudo da predição espacial

| PROJETO Nº | IDADE (Anos) | Nº DE PARCELAS (IFC) | Nº DE PARCELAS (PRÉ-CORTE) | Nº DE TALHÕES |
|---------------|-----------------|-------------------------|-------------------------------|------------------|
| C131 | 6,8 | 48 | 60 | 17 |
| C141 | 7,0 | 80 | 190 | 28 |
| C153/154 | 6,0 | 43 | 53 | 13 |
| C564 | 5,8 | 23 | 49 | 16 |
| C565 | 5,9 | 16 | 36 | 08 |

IFC. Inventário Florestal Contínuo

6.2.2 Inventário pré-corte

Para cada projeto florestal, a Votorantim Celulose e Papel realizou o inventário pré-corte em todos os talhões, com a intensidade amostral conforme a Tabela 1. Este inventário foi efetuado por ocasião da última remedição das parcelas permanentes usadas para a predição espacial. Estes inventários foram processados conforme os princípios da teoria clássica de amostragem, ou seja, sem considerar a estrutura de continuidade espacial. Obteve-se a média aritmética e sua variância e posteriormente o intervalo de confiança e o erro do inventário para cada talhão. Estas estatísticas foram posteriormente comparadas com àquelas geradas pelo estimador geoestatístico.

6.2.3 Estudo variográfico

Para cada projeto efetuou-se o estudo variográfico a fim de verificar a estrutura de continuidade espacial. Considere 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 , $i=1,2,\dots,n$, onde 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. O estudo da correlação entre $Z(x_i), \dots, Z(x_n)$, é dado pelo semivariograma.

Considere dois pontos separados por uma distância h . A variação da variável regionalizada (volume) entre eles, é caracterizada por uma função variograma $2g(x, h)$. Esta função é definida como

$$2\mathbf{g}(x,h) = E[Z(x) - Z(x+h)]^2 \quad (1)$$

Note que a função variograma, como representado 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 é a mesma para toda a área avaliada (estacionaridade) e a variância entre pares de pontos depende apenas do vetor de separação h (Journel & Huijbregts, 1978).

Na Teoria das Variáveis Regionalizadas, outro nome dado à Geoestatística, a quantidade $\mathbf{g}(h)$ é conhecida como semivariância e é, como definido acima, a metade da esperança da variância entre pares de pontos separados por uma distância (h). Ao se obedecer as condições da hipótese intrínseca, a semivariância pode ser estimada simplesmente por uma única realização do processo estocástico pela equação:

$$\hat{\mathbf{g}}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i) - Z(x_i + h)]^2 \quad (2)$$

em que $\hat{\mathbf{g}}(h)$ 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 por um vetor h . Esta função permite gerar o semivariograma experimental. Neste semivariograma, é possível ajustar uma função matemática que expressa a estrutura de dependência espacial da característica avaliada. Para cada projeto, foi ajustado e selecionado o melhor modelo de função de semivariância aos dados, através do Método da Máxima Verossimilhança. O semivariograma experimental e o ajuste dos modelos foram efetuados através do programa R, utilizando-se o pacote geoR (Ribeiro Jr. & Diggle, 2001). Conforme Biondi et al. (1994), a avaliação do grau de dependência espacial (DE) do volume por hectare foi realizada através do percentual de variação estruturada (σ^2) em relação ao patamar ($\tau^2 + \sigma^2$).

$$DE = \frac{\mathbf{s}^2}{\mathbf{t}^2 + \mathbf{s}^2} * 100 \quad (3)$$

6.2.4 Modelo geoestatístico

Considerando que as observações medidas para o volume nos diferentes pontos amostrais (parcelas) apresentam-se estruturadas espacialmente, tem-se um vetor multidimensional de observações correlacionadas em duas dimensões espaciais, as quais foram avaliadas por um modelo linear misto:

$$z_i = \mathbf{m} + S(x_i) + \mathbf{e}_i \quad (4)$$

onde:

z_i : valor observado no ponto i ;

\mathbf{m} : constante determinística na condição da hipótese intrínseca

$S(x_i)$: componente estocástica que explica o comportamento espacial autocorrelacionada da variável Z em toda a área, sendo x_i a localização do ponto i ;

\mathbf{e}_i : erro aleatório não correlacionado espacialmente com distribuição normal e variância τ^2 .

$S(\cdot)$ é um processo Gaussiano estacionário com $E[S(x)] = \mathbf{m}$, $VAR[S(x)] = \sigma^2$ e $\mathbf{r}(h) = CORR[S(x), S(x+h)]$, sendo h o deslocamento em relação à posição x . A variável resposta Z tem uma distribuição conjunta multivariada normal (Diggle & Ribeiro Júnior, 2000),

$$Z \sim MVN(\mathbf{m}, \mathbf{s}^2 \Sigma + \mathbf{t}^2 I) \quad (5)$$

onde:

$\mathbf{1}$: vetor de valores 1

σ^2 : variância populacional

Σ : matriz de correlação, que pode ser transformada em matriz de semivariância e vice versa, definida conforme os parâmetros de cada modelo selecionado

I : matriz identidade.

τ^2 : variância do erro aleatório.

A função de densidade da variável resposta Z é definida pela seguinte expressão:

$$f(\underline{z}) = \frac{1}{(2\mathbf{p})^{n/2} |\mathbf{K}|^{1/2}} e^{(\underline{z}-\underline{m})' \mathbf{K}^{-1} (\underline{z}-\underline{m})} \quad (6)$$

onde:

\underline{z} : vetor da variável resposta

\underline{m} : vetor de média (constante na condição da hipótese intrínseca)

\mathbf{K} : igual a $\sigma^2 \Sigma + \tau^2 \mathbf{I}$.

Na presença de autocorrelação, a matriz de semivariância (Σ) é definida por uma função de dependência espacial, a partir da distância (h) entre as observações de cada projeto. Assim, para cada projeto obteve-se o modelo espacial, do qual todas as inferências foram efetuadas.

6.2.5 Krigagem com simulação

Foi definido o volume por hectare de cada uma das parcelas permanentes nos respectivos projetos. Assim, deseja-se agora, predizer valores em locais não avaliados $[\hat{z}(x_0)]$. Esta estimativa pode ser obtida através da seguinte fórmula:

$$\hat{z}(x_0) = \sum_{i=1}^n I_i z(x_i) \quad (7)$$

em que $\hat{z}(x_0)$ é a estimativa no ponto não amostrado (x_0); I_i é o peso que cada valor avaliado recebe, conforme a estrutura de continuidade espacial do volume e $z(x_i)$ é o valor do ponto amostrado. Este processo é efetuado numa grade quadrada ou retangular com as coordenadas pré-definidas, sendo esta uma krigagem do tipo pontual.

A krigagem com simulação permite realizar diversos cenários sobre a área total a ser krigada. O objetivo da simulação é reproduzir a função aleatória, cujas características

foram identificadas na análise estrutural (modelo geoestatístico). Pode ser entendida como uma particular realização do modelo Multivariado Normal definido para cada um dos projetos. Portanto, os valores simulados podem ser vistos como diferentes amostras retiradas da população florestal. As simulações reproduzem, portanto, a mesma variabilidade dos dados sob toda a área. Ela gera um grande número de valores aleatórios em cada ponto amostrado (x_o), a partir do método de simulação de Monte Carlo (Diggle e Ribeiro Júnior., 2000).

6.2.6 Predição espacial ou krigagem de bloco

Para efetuar a predição espacial do volume em cada talhão dos respectivos projetos, utilizou-se da krigagem com simulação em cada talhão (bloco). Desta forma, a krigagem de bloco foi realizada numericamente. Com o modelo geoestatístico de cada projeto, realizaram-se 3000 simulações em cada ponto não amostrado dentro do talhão. Para cada simulação, obteve-se a média aritmética dos pontos estimados nos respectivos talhões. Portanto, ao final, obtiveram-se 3000 valores médios estimados de volume para cada talhão. Com estes valores gerou-se a média aritmética dos mesmos e a variância entre estas médias. Estas duas medidas estatísticas possibilitaram estabelecer o intervalo de confiança e o erro do inventário em percentagem para cada talhão nos respectivos projetos florestais. Este método foi denominado por Journel & Huijbregts (1978) como simulação de bloco condicionada aos pontos amostrados. Este método de predição também foi descrito por Soares (2000). Segundo este autor, quando se pretende obter o valor médio da variável regionalizada Z numa subárea R , este pode ser obtido pela média dos valores krigados pontuais que compõem a subárea R .

6.3 Resultados e Discussão

6.3.1 Análise variográfica

Na Tabela 2 estão apresentados os modelos de função de semivariância com os parâmetros ajustados pelo Método da Máxima Verossimilhança, para cada um dos projetos de *Eucalyptus grandis* avaliados. Por esta tabela observou-se que a

característica volume por hectare de todos os projetos apresentou-se estruturada espacialmente. Exceto o projeto C153, os demais tiveram grau de dependência espacial (DE) superior a 75%. Em trabalhos de geoestatística na área de Ciência do Solo, segundo Cambardella et al. (1994), quando o grau de dependência espacial for superior a 75%, significa que a característica tem forte dependência espacial. Conforme esta mesma classificação, o projeto C153 possuiu moderado grau de dependência espacial. Assim, de acordo com esta classificação, pode-se dizer que o grau de dependência espacial do volume por hectare foi de moderado a forte. Estes resultados mostraram que as estimativas de volume, considerando a componente espacial, proporcionará otimização no processo de krigagem.

Tabela 2. Parâmetros do modelo espacial efeito pepita (t^2), contribuição (s^2), patamar ($t^2 + s^2$), alcance (q) e o grau de dependência espacial (DE%) de cada projeto

| PROJETOS/ MODELO | PARÂMETROS | | | | |
|---------------------|------------|--------|-------------|-------|-------|
| | t^2 | s^2 | $t^2 + s^2$ | q (m) | DE(%) |
| C131/EXP | 0 | 1105,9 | 1105,9 | 251 | 100,0 |
| C141/EXP | 380 | 1543,2 | 1923,2 | 283 | 80,2 |
| C153/ESF. | 350 | 766,4 | 1116,4 | 1634 | 68,6 |
| C564/EXP | 100 | 620,5 | 720,5 | 66 | 86,1 |
| C565/EXP | 300 | 1302,8 | 1602,8 | 145 | 81,3 |

Nos projetos C564 e C565 houve poucas parcelas permanentes para ajuste dos modelos (Tabela 1). Apesar dos mesmos apresentarem graus de dependência espacial elevados, existe a incerteza com relação à estimativa dos parâmetros, sobretudo do efeito pepita, devido à insuficiência amostral na pequena escala (parcelas próximas). Os valores de alcance destes projetos foram de 66 e 145 metros. O parâmetro alcance é indicativo da magnitude da continuidade espacial. Comparando os valores de alcance com os demais projetos, pode-se dizer que estes dois possuem fraca dependência. Portanto, o valor do grau de dependência espacial destes projetos (DE), está fortemente influenciado pelo baixo valor do efeito pepita. Os valores do efeito pepita destes projetos estão fortemente influenciados pela insuficiência amostral na pequena escala, e não

propriamente pelo grau de continuidade espacial do volume. Esta situação deve ser motivo para avaliação sobre a forma de amostragem para variáveis com continuidade espacial. É de extrema importância amostrar adequadamente na pequena escala, a fim de obter uma boa estimativa do efeito pepita, uma vez que ele atua diretamente na qualidade da krigagem (Soares, 2002). Com os valores dos parâmetros da Tabela 2, fica difícil supor que a característica volume por hectare seja um fenômeno essencialmente aleatório. Logo, o uso da teoria clássica de amostragem no processo de inferência, estará provocando um viés substancial na precisão.

6.3.2 Predição espacial

A Tabela 3 contém informações médias de volume por projeto. Verificou-se que os dois estimadores geraram valores médios de volume total por projeto praticamente iguais. A razão média em percentagem do volume total estimada pela teoria clássica de amostragem (pré-corte) e do estimador geoestatístico foi inferior a 1%. Pela Tabela 4 observa-se que a teoria clássica de amostragem teve erro inferior a 10% em 40 talhões e o estimador geoestatístico teve este mesmo desempenho em 63 talhões.

Tabela 3. Informações médias volume total em m³ (VT), volume médio por hectare em m³ (VM), erro médio em percentagem (EM) para os dois estimadores

| PROJETO | PRÉ-CORTE (VPC) | | | GEOESTATÍSTICO (VG) | | | |
|----------|-----------------|----------|-----|---------------------|----------|-----|------|
| | Nº | VT | VM | EM | VT | VM | EM |
| C131 | | 105893,3 | 247 | 14,6 | 104327,9 | 243 | 8,0 |
| C141 | | 229099,0 | 275 | 9,0 | 228542,0 | 275 | 10,0 |
| C151/153 | | 104279,9 | 225 | 12,0 | 104032,0 | 226 | 8,0 |
| C564 | | 51892,0 | 193 | 32,0 | 52620,2 | 205 | 9,0 |
| C565 | | 39629,0 | 252 | 14,0 | 39308,2 | 253 | 12,0 |

Na Figura 1 apresenta-se o volume, em m^3/ha , determinado pela teoria clássica de amostragem (pré-corte), com o respectivo valor estimado pela krigagem geoestatística. Foram plotados os valores de volume, considerando-se uma diferença entre ambos métodos de $\pm 10\%$. Foi possível verificar que dos 81 talhões estimados, apenas 12 estão fora do intervalo de erro de 10%, produzindo-se erros superiores a este valor. Mesmo nestes pontos fora deste intervalo de erro, observam-se diferenças próximas a 10%, com apenas dois talhões com erros consideravelmente maiores. Há grande quantidade de pontos (talhões) em torno da bissetriz central, mostrando pequena dispersão dos valores. Estas observações permitiram avaliar que o estimador geoestatístico produziu estimativa de volume muito próxima daquela gerada pelo estimador clássico (pré-corte), considerada a estimativa mais confiável de volume por talhão utilizada pela empresa florestal.

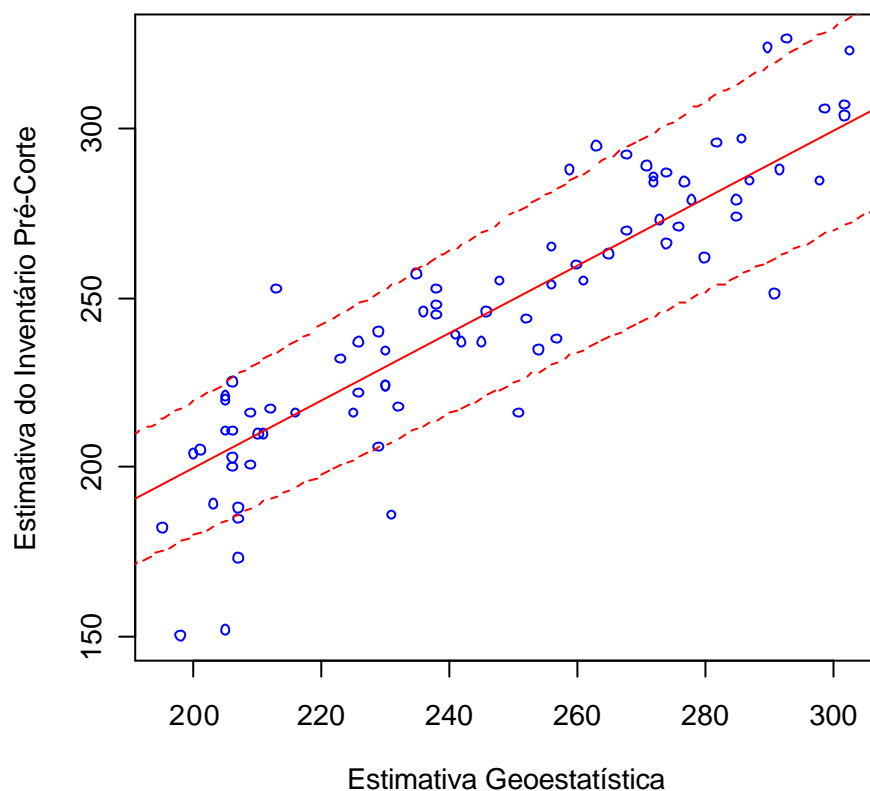


Figura 1 - Relação entre os volumes (m^3) estimados pela teoria clássica de amostragem e pela geoestatística, com erro de 10% delimitado pela linha tracejada

Pela Figura 2, pode-se observar o comportamento do erro amostral (%) em função do volume estimado pela geoestatística e pela teoria clássica de amostragem. Observou-se que o erro amostral gerado pelo estimador geoestatístico possuiu comportamento consideravelmente mais estável do que o estimador clássico, havendo erro máximo em torno de 15% para o primeiro e 40% para o segundo. Analisou-se também que o erro amostral médio do estimador geoestatístico está em torno de 10%, com baixa variabilidade. Segundo Trangmar et al. (1985), a krigagem de bloco produz mapas suaves através da interpolação dos valores médios de cada bloco. Este efeito influenciou na dispersão do erro amostral do estimador geoestatístico, fazendo com que

o mesmo apresentasse a medida de precisão mais estável. Já o estimador da teoria clássica de amostragem apresentou oscilações que variaram aproximadamente de 5% a 40%. Estas oscilações do estimador clássico estão de acordo com os resultados que o inventário pré-corte gera. Houve situações em que este estimador, num mesmo projeto, gerou erros abaixo de 10% em alguns talhões e erros que atingiram 40% em outros talhões. Parte desta oscilação pode ser explicada pela variação existente em um mesmo talhão. Pela Figura 3 é possível verificar a magnitude da variação quanto a um mesmo talhão. Esta figura se refere à krigagem com simulação do projeto C141, sendo destacado o talhão três com erro de 2,8% e o talhão 14 com erro de 17,6%. O estimador clássico entende que o talhão é uma subpopulação do projeto, assim, ele usa apenas as informações contidas nas parcelas lançadas no talhão. No estimador geoestatístico esta oscilação foi menor. Neste caso, ao gerar as estimativas por talhão (krigagem de bloco), o estimador utiliza as informações espaciais do volume de todas as parcelas contidas no projeto. Esta é uma diferença marcante entre os dois estimadores.

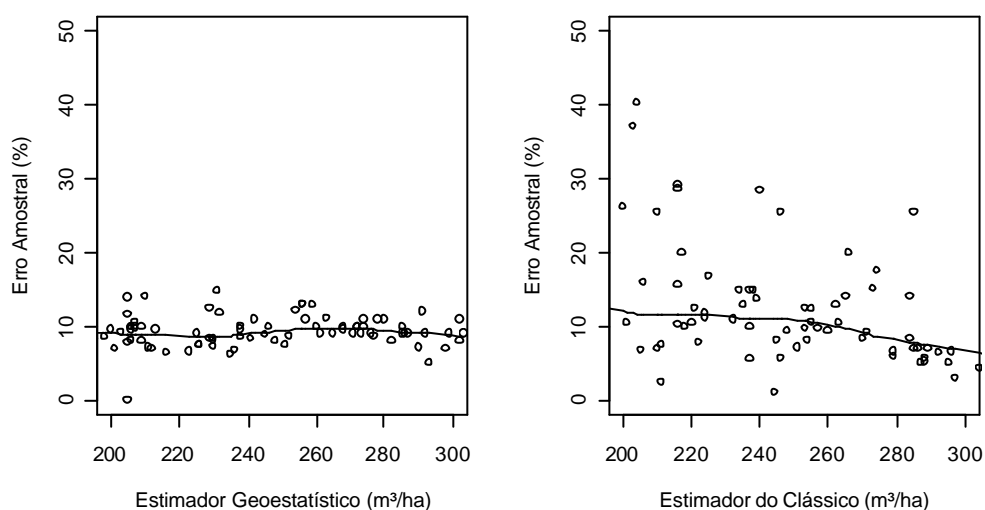


Figura 2 - Relação do erro percentual para os dois estimadores considerando todos os talhões

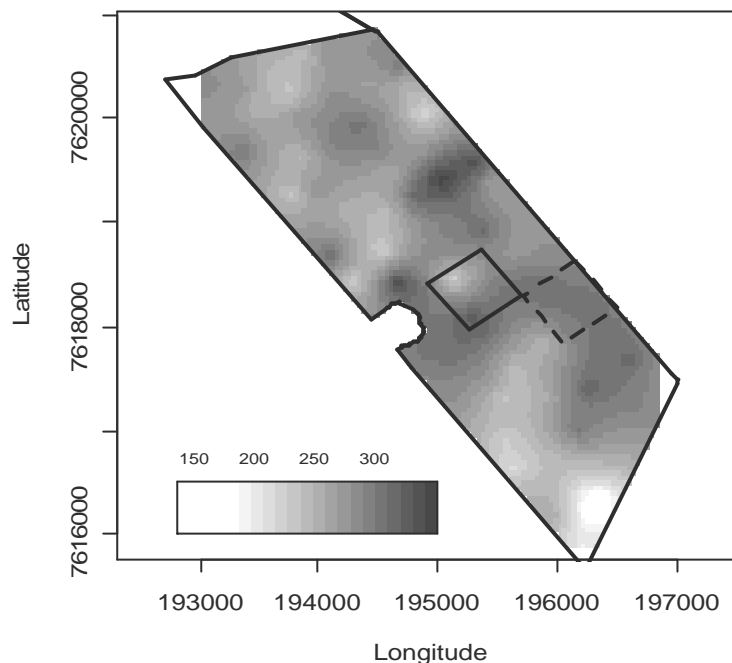


Figura 3 - Mapa de krigagem do projeto C141, destacando os talhões 3 (tracejado) e 14 (linha cheia)

Na Figura 4 apresentam-se as diferenças percentuais entre os estimadores por talhão nos respectivos projetos avaliados. Considerou-se uma diferença entre os métodos de $\pm 10\%$. Observa-se que as diferenças estão centradas próximas do eixo zero, mostrando uma semelhança evidente entre as estimativas efetuadas pelos dois estimadores. Esta tendência não foi confirmada para o projeto C565 (Figura 4e), o qual apresentou alta dispersão das diferenças, sendo explicada pela baixa intensidade amostral usada tanto no estimador clássico, quanto no estimador geoestatístico para a definição do modelo. Por outro lado, o projeto C141 (Figura 4b) apresentou as diferenças mais concentradas próximas do eixo zero. Este projeto foi o que teve a maior intensidade amostral (80) para gerar o modelo geoestatístico e, conseqüentemente, o maior número de vizinhos usados na krigagem de bloco. Certamente, a intensidade amostral bem dispersa em todo projeto, proporcionou maior estabilidade do estimador no processo de inferência espacial (krigagem de bloco).

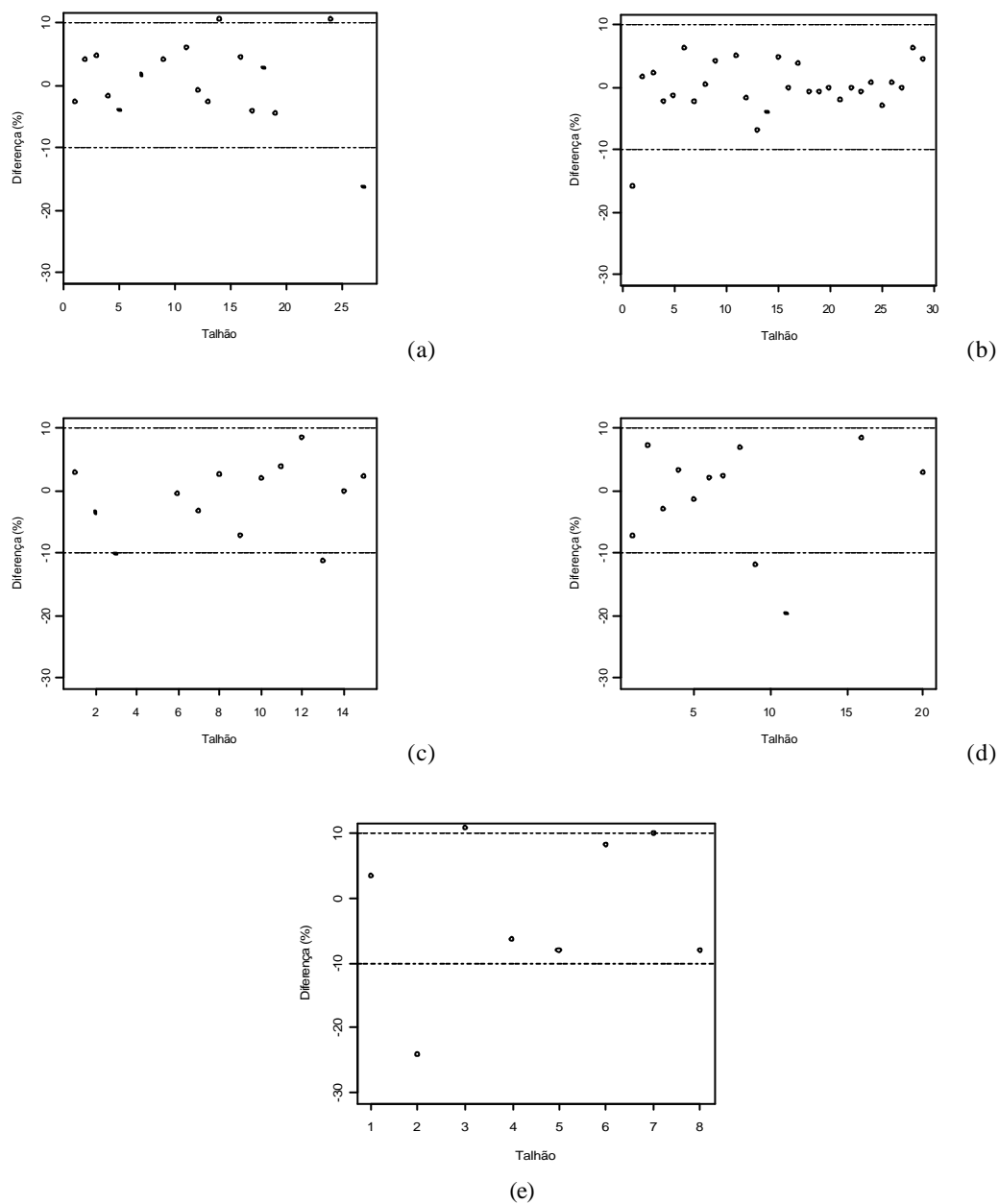


Figura 4 - Diferença percentual entre os estimadores por talhão nos respectivos projetos avaliados, C131(a), C141(b), C153(c), C564(d) e C565(e), com uma diferença de $\pm 10\%$ marcada pela linha tracejada

Outro aspecto importante a ser observado na Figura 4, foi a proximidade dos pontos médios dos intervalos de confiança gerados pelos dois estimadores, na maioria dos talhões avaliados nos projetos. Em 72% dos talhões avaliados, a amplitude do intervalo de confiança do estimador geoestatístico foi maior do que a amplitude do intervalo do estimador clássico. Conforme demonstrado no item 3 deste trabalho, na presença de continuidade espacial, há uma inflação da precisão do estimador clássico. Assim, este estimador tende a gerar intervalos com amplitude menor, mais preciso, porém ilusório.

Na Figura 5 apresenta-se uma comparação dos intervalos de confiança gerados pelos dois estimadores. Avalia-se que a amplitude do intervalo de confiança do estimador geoestatístico oscilou pouco entre os talhões de cada projeto. Isto mostra que houve regularidade na precisão em todos os talhões no mesmo projeto. Parte desta regularidade foi devido à distribuição sistemática desencontrada (Cochran, 1977) das parcelas do inventário permanente em cada projeto. O uso de todas as parcelas do projeto, para efetuar krigagem de bloco (estimativa por talhão), foi outro fator que exerceu influência na regularidade de estimador geoestatístico. O intervalo de confiança do estimador clássico apresentou grandes variações de amplitude nos diferentes talhões dentro de um mesmo projeto. Esta variação se reflete na incerteza com relação à precisão gerada pelo estimador clássico. Ela está fortemente associada às variações existente em um mesmo talhão (Figura 3).

Todas as observações constatadas permitiram avaliar que o estimador geoestatístico gerou estimativa de volume semelhante ao do inventário pré-corte, conjecturando-se sobre uma possível substituição deste último pela estimativa geoestatística, o que proporcionará redução considerável no custo de medição da floresta. No presente estudo, houve um acréscimo no custo de medição referente a 388 parcelas mensuradas no pré-corte dos cinco projetos avaliados. Ao que tudo indica, o estimador geoestatístico proporcionou resultados, se não melhores, pelo menos iguais ao inventário pré-corte, sem a medição de mais 388 parcelas.

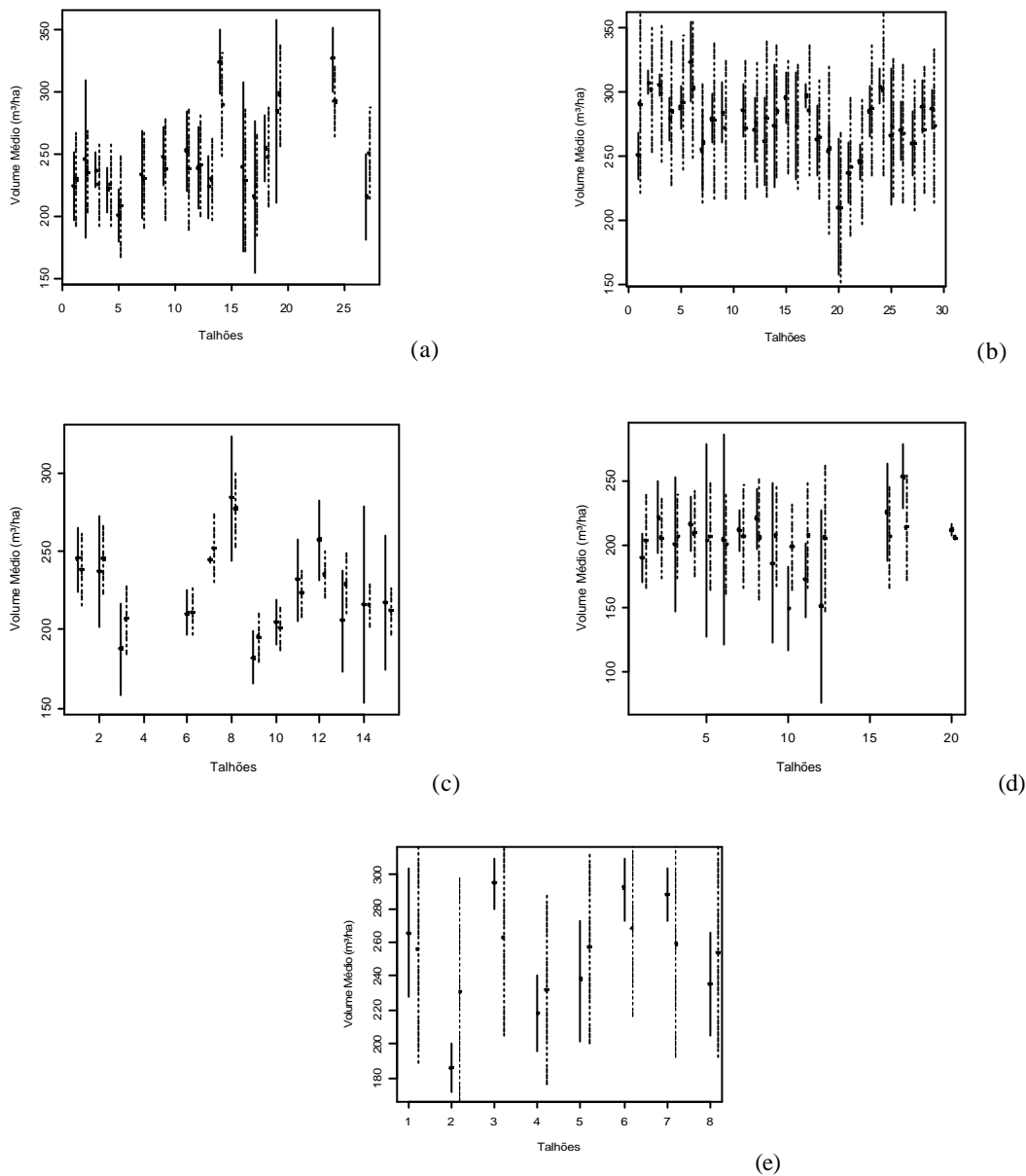


Figura 5 - Comparação entre as coberturas dos intervalos de confiança gerada pelo estimador geoestatístico (linha tracejada) e pelo estimador clássico (linha cheia) para os respectivos talhões dos projetos C131(a), C141(b), C153(c), C564(d) e C565(e)

6.4 Conclusões

Conforme as condições em que o trabalho foi desenvolvido, é possível concluir que:

- os cinco projetos avaliados apresentaram grau de continuidade espacial de moderada a forte para a característica dendrométrica volume por hectare, devendo-se considerar a componente espacial no processamento dos inventários;
- a diferença média entre o volume total gerado pelo estimador clássico e o geoestatístico foi de 1%, com a variação do erro de 9 a 32% para o primeiro e 8 a 12% para o segundo;
- o erro amostral por talhão em percentagem do estimador geoestatístico variou menos do que o estimador clássico;
- a amplitude do intervalo de confiança do estimador geoestatístico oscilou pouco entre os talhões avaliados;
- de maneira geral as estimativas do estimador geoestatístico foram semelhantes quando comparadas com as do estimador clássico. Na presença de continuidade espacial o uso do estimador geoestatístico para estimativa de volume por talhão, é preferido por duas razões: proporciona estimativa da precisão sem viés e é de baixo custo quando comparada com o estimador clássico.

6 CONCLUSÕES GERAIS

O desenvolvimento deste trabalho permitiu verificar que as principais características dendrométricas avaliadas no inventário florestal, apresentaram-se estruturadas espacialmente. Estes resultados sugerem que, no processamento do inventário florestal, na presença de estrutura de continuidade espacial, as parcelas (unidades amostrais) não devem ser tratadas de forma independente.

Outro ponto relevante observado neste trabalho, refere-se à forma de como amostrar a floresta. Se a amostragem for efetuada sobre uma população fixa a teoria clássica de amostragem para o estabelecimento do intervalo de confiança é válida. Caso a amostragem seja efetuada sobre um conjunto infinito de pontos sendo esta situação mais genérica na amostragem florestal, o intervalo de confiança estabelecido com base nos conceitos da teoria clássica de amostragem inflaciona a precisão do levantamento. Ao incorporar a componente espacial no modelo, as coberturas observadas versus as coberturas nominais dos intervalos de confiança tornam-se iguais.

A relação entre o número de parcelas lançadas e o número de parcelas úteis para definição da variabilidade da variável de interesse é outro ponto interessante observado neste trabalho. Esta relação é unitária se não houver estrutura de continuidade espacial. A teoria clássica de amostragem considera todas as parcelas lançadas na área, pois assume que as mesmas são independentes e, portanto, não redundantes. Porém, na presença de continuidade espacial o número de parcelas que efetivamente carregam informação não redundante é menor (parcelas úteis). A partir desta constatação, sugere-se para estudos futuros, incorporar ou desenvolver uma metodologia para definição do número de parcelas a serem avaliadas numa floresta com estrutura de continuidade espacial da característica avaliada.

As empresas florestais, de maneira geral, conduzem um grupo de parcelas permanentes nos projetos florestais implantados, permitindo avaliar o crescimento da floresta. Na idade de corte da floresta, é preciso estimar o volume de cada talhão que compõe os projetos, uma vez que somente as informações das parcelas permanentes não proporcionam estimativas precisas por talhão. É necessário portanto, efetuar o inventário pré-corte, o qual lança uma intensidade amostral maior em cada talhão. Assim, há um novo custo adicionado no processo de medição da floresta. Fazendo uso do estimador geoestatístico, o qual considera a estrutura de continuidade espacial, é possível efetuar com precisão a estimativa volumétrica por talhão fazendo uso somente das parcelas permanentes.

Em síntese, a aplicação da teoria clássica de amostragem no inventário florestal, pode aumentar os problemas estatísticos, caso houver estrutura de continuidade espacial. Assim, é necessário, primeiramente, na análise do inventário florestal, verificar se há ou não estrutura de continuidade espacial da característica avaliada. Havendo continuidade espacial, deve-se fazer uso dos benefícios da mesma através da introdução da componente espacial no modelo.

REFERÊNCIAS BIBLIOGRÁFICAS

- AKAIKE, H. Information measures and model selection. **International Statistical Institute**, v.44, p.277-291, 1983.
- AUBRY, P.; DEBOUZIE, D. Geostatistical estimation variance for the spatial mean in two-dimensional systematic sampling. **Ecology**, v.81, n.2, p.543-553, 2000.
- AUBRY, P.; DEBOUZIE, D. Estimation of the mean from a two-dimensional sample: The geostatistical model-based approach. **Ecology**, v.82, n.5, p.1484-1494, 2001.
- BIONDI, F.; MYERS, D.E.; AVERY, C.C. Geostatistically modeling stem size and increment in an old-growth forest. **Canadian Journal of Forest Research-Revue Canadienne de Recherche Forestiere**, v.24, n.7, p.1354-1368, 1994.
- BRAGA, L.P.V. Geoestatística e aplicações. In: SIMPÓSIO BRASILEIRO DE PROBABILIDADE E ESTATÍSTICA, 9., São Paulo, 1990. **Resumos**, São Paulo: IME-USP, 1990, p.36.
- BRUS, D.J.D.; GRUIJTER, J. J. Design-based versus model-based estimates of spatial means: theory and application in environmental soil science. **Environmetrics**, v.4, n.2, p.123-152, 1993.

- BRUS, D. J.D.; GRUIJTER, J.J. Random sampling or geostatistical modelling? Choosing between design-based and model-based sampling strategies for soil (with discussion). **Geoderma**, v.80, n.1/2, p.1-59, 1997.
- BURGESS, T.M.; WEBSTER, R. Optimal interpolation and isarithmic mapping of soil properties. 1.The semivariogram and punctual kriging. **Journal of Soil Science**, v.31, n.2, p.315-331, 1980.
- BURGESS, T.M.; WEBSTER, R.; McBRATNEY, A.B. Optimal interpolation and isarithmic mapping of soil properties IV. Sampling Strategy. **Journal of Soil Science**, v.32, n.4, p.643-659, 1981.
- 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.
- CAMBARDELLA, C. A.; MOORMAN, T. B.; NOVAK, J. M.; PARKIN, T. B.; KARLEN, D. L.; TURCO, R. F.; KONOPKA, A. E. Field scale variability of soil properties in central Iowa soils. **Soil Science Society of America Journal**, v.58, n.5, p.1501-1511, 1994.
- CLARK, I. **Practical geostatistics**. London: Applied Science Publishers, 1979. 128p.
- COCHRAN, W. G. **Sampling techniques**. 3.ed. New York: Wiley, 1977. 555p.
- COHEN, W.B.; SPIES, T.A.; BRADSHAW, G.A. Semivariograms of digital imagery for analysis of conifer canopy structure. **Remote Sensing of Environment**, v.34, n.3, p.167-178, 1990.

CRESSIE, N. Fitting variogram models using weighted least-squares, **Journal of the International Association for Mathematical Geology**, v.17, n.5, p.563-586, 1985.

CRESSIE, N. **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., Caxambú, 2000. **Resumos**, Caxambú: Associação Brasileira de Estatística, 2000, p.192.

DUARTE, J.B. Sobre o emprego e a análise estatística do delineamento em blocos aumentados no melhoramento genético vegetal. Piracicaba, 2000. 292p. Tese (Doutorado) – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo.

GUNNARSON, F.; HOLM, P.H.; THURESONN, T. On the potential of krigagem for forest mannegement planning. **Scandinavian Journal of Forest Research**, v.13, n.2, p.237-245,1998.

HIGUCHI, N. Amostragem sistemática versus amostragem aleatória em floresta tropical úmida de terra firme na região de Manaus. **Acta Amazônica**, v.16/17, p.393-400, 1986/1987.

HOAGLIN, D. C. Letras-resumo: um conjunto de estatísticas ordinais selecionadas. In: HOAGLIN, D. C.; MOSTELLER, F.; TUKEY, J.W. (Ed.). **Análise exploratória de dados**. Técnicas robustas: um guia. Lisboa: Edições Salamandra, 1992. cap.2, p.32-58.

- HÖCK, B.K.; PAYN, T.W.; SHIRLEY, J. Using a geographic information system and geostatistics to estimate site index of *Pinus radiata* for Kaingaroa forest. **New Zealand Journal of Forestry Science**, v.23, p.264-277, 1998.
- HOEF, J.V. Sampling and geostatistics for spatial. **Ecoscience**, v.9, n.2, p.152-161, 2002.
- ISAAKS, E.H.; SRIVASTAVA, R.M. **An introduction to applied geostatistics**. New York: Oxford University Press, 1989. 560p.
- JARQUE, C.M.; BERA, A.K. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. **Economics Letters**, v.6, n.3, p.255-259, 1980.
- JOURNEL, A.G.; HUIJBREGTS, C.J. **Mining geostatistics**. London: Academic Press, 1978. 600p.
- KOCH, G. G.; GILLINGS, D.B. Inference, design-based vs. Model-based. In: KOTZ, S.; JOHNSON, N.L (Ed.). **Encyclopedia of statistical sciences**. New York: Wiley, 1983. v.4, p.84-88.
- KORNING, J.; THOMSEN, K.; OLLGAARD, B. Composition and structure of a species rich Amazonian rain forest obtained by two different sample methods. **Nordic Journal of Botany**, v.11, n.1, p.103-110, 1991.
- KRIGE, D.G. A statistical approach to some basic mine valuation problems on the Witwatersrand. **Journal of the Chemical, Metallurgical and Mining Society of South Africa**, v.52, p.119-139, 1951.

- KULBACK, S.; LEIBLER, R.A. On information and sufficiency. **Annals of Mathematical Statistics**, v.22, p.79-86, 1951.
- MACHADO, S. A. Complete enumeration forest inventory versus Cluster sampling methods applied in the Amazonian Rain forest. **Revista Floresta**, v.18, n.1/2, p.122-130, 1988.
- MATÉRN, B. **Spatial variation:** stochastic models and their application to some problems in forest surveys and other sampling investigations. Stockholm: Skogsforsknings Institut, 1960. 144p. (Skogsforsknings Institut. Meddelanden fran Statens Skogsforsknings Institut, v.49, n.5)
- MATHERON, G. Principles of geostatistics. **Economic Geology**, v.58, p.1246-1266, 1963.
- Mc BRATNEY, A.G.; WEBSTER, A.G. Choosing functions for semi-variograms and fitting them to sampling estimates. **Journal of Soil Science**, v.37, p.617-639, 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**, v.27, n.5, p.925-933, 2003.
- MELLO, J.M.; SCOLFORO, J.R.S. Análise comparativa de procedimentos de amostragem em um remanescente de floresta estacional semidecídua montana. **Revista Árvore**, v.24, n.1, p.55-62, 2000.
- NETER, J.; WASSERMAN, W.; KUTNER, M. H. **Applied linear statistical models:** regression analysis of variance, and experimental designs. 3.ed. New York: Irwin, 1990. 471p.

- OLIVEIRA, M. S. Planos amostrais para variáveis espaciais utilizando geoestatística. Campinas, 1991. 100p. Dissertação (Mestrado) – Universidade Estadual de Campinas.
- OSBORNE, J.G. Sampling errors of systematic and random surveys of cover-type areas. **Journal Statistics Association**, v.37, p.256-264, 1942.
- PÉLLICO NETTO, S.; BRENA, D. A. **Inventário Florestal**. Curitiba: Os autores, 1997. 316p.
- PRODAN, M., PETERS, R., COX, F., REAL, P. **Mensura Florestal**, Costa Rica: San José, 1997. 586p.
- QUENOUILLE, M. H. Problems in plane sampling. **Annals of Mathematical Statistics**, v.20, n.3, p.355-375, 1949.
- R Development Core Team. R: **A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-00-3, URL <http://www.R-project.org>. (2003)
- RIBEIRO JÚNIOR, P.J.; DIGGLE, P.J. GeoR: a package for geostatistical analysis. **R-NEWS**, v.1, n.2, p.15-18, 2001.
- RIBEIRO JÚNIOR, P.J. Métodos geoestatísticos no estudo da variabilidade espacial de parâmetros do solo. Piracicaba, 1995. 99p. Dissertação (Mestrado) – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo.
- SAMRA, J.S. GILL, H.S.; BHATIA, V.K. Spatial stochastic modelling of growth and forest resource evaluation. **Forest Science**, v.35, n.3, p.663-676, 1989.

- SÄRNDAL, C. E. Design-based and model-based inference in survey sampling. **Scandinavian Journal of Statistics**, v.5, n.1, p.27-52, 1978.
- SCOLFORO, J.R.S.; MELLO, J.M. **Inventário Florestal**. Lavras: UFLA-FAEPE, 1997. 344p.
- SOARES, A.G. **Geoestatística para as ciências da terra e do ambiente**. Lisboa: IST Press, 2000. 186p.
- THOMPSON, S.K. **Sampling**. New York: Wiley, 1992. 343p.
- THOMPSON, S.K.; SEBER, G.A.F. **Adaptative Sampling**. New York: Wiley, 1996. 350p.
- TRANGMAR, B.B.; YOST, R.S.; UEHARA, G. Application of geostatistics to spacial studies of soil properties. **Advances in Agronomy**, v.38, p.45-94, 1985.
- VIEIRA, S.R. Geoestatística em estudos de variabilidade espacial do solo. In: NOVAES, R.S.; ALVAREZ V, V.H.; SCHAESER, C.E.G.R. (Ed.). **Tópicos em Ciências do Solo**. Viçosa: Sociedade Brasileira de Ciência do Solo, 2000. cap.1, p.1-54.
- WONNACOTT, T.H.;WONNACOTT, R.J. **Introductory Statistics**. 2.ed. New York: John Wiley, 1972. 510p.
- WULDER, M.; LAVIGNE, M.; FRANKLIN, S. High spatial resolution optical image texture for improved estimation of forest stand leaf area index. **Canadian Journal of Remote Sensing**, v.22, n.4, p.441-449, 1996.
- YAMAMOTO, J. K. **Avaliação e classificação de reservas minerais**. São Paulo: Editora da Universidade de São Paulo, 2001. 226p.