

ABORDAGEM *BORROW STRENGTH* APLICADA A MODELO GEOESTATÍSTICO PARA ESTIMATIVA DE VOLUME

BORROW STRENGTH APPROACH APPLIED TO A GEOSTATISTICAL MODEL TO ESTIMATE VOLUME

Julio Cesar Wojciechowski¹ Julio Eduardo Arce² Saulo Henrique Weber³,
Paulo Justiniano Ribeiro Junior⁴ Carlos Alberto da Fonseca Pires⁵

RESUMO

O presente estudo teve como objetivo utilizar o compartilhamento de parâmetros de modelos geoestatísticos aplicado aos estimadores de máxima verossimilhança para predizer os volumes por hectare em três fragmentos de Floresta Estacional Subtropical localizados em Santa Teresa - RS empregando a abordagem *Borrow strength*. Os dados foram coletados em 56 unidades amostrais (U.A) de tamanho variável com aproximadamente 250 m² em um total de 9 ha, distribuídas em um *grid* sistemático de 40 x 40 m, sendo medidas as variáveis dendrométricas dos indivíduos com DAP \geq 10 cm próximas ao centro das unidades. Foram elaboradas duas abordagens para o conjunto de dados, sendo que a primeira considerou as áreas totalmente independentes entre si, subdivididas em dois tipos: ajuste ao modelo não espacial (NSM) e ajuste pelo método de máxima verossimilhança (MV) não compartilhado (ajuste individual). A segunda abordagem descreveu os ajustes dos modelos de máxima verossimilhança compartilhados em função do erro aleatório ou *nugget*, sendo: modelos sem *nugget* fixo (variabilidade entre as U.A) e com *nugget* fixo (variabilidade dentro das U.A), utilizando como correlação a função exponencial da família Matérn. Em seguida, os modelos foram comparados pelo critério de informação de Akaike (AIC) e grau de dependência espacial para posterior krigagem e elaboração das superfícies de predição dos modelos selecionados. Foi observado que os modelos combinados para estimativa do volume foram superiores para os valores de AIC e grau de dependência espacial em relação aos ajustes para as áreas individuais. Entre os modelos compartilhados, observou-se que houve um ganho nas estimativas dos parâmetros utilizando o *nugget* fixo, que resultaram em uma correlação das amostras e grau de dependência espacial maior (AP = 88 m), em relação aos modelos compartilhado sem *nugget* fixo (AP = 75 e 66 m). O AIC mostrou-se eficiente, uma vez que comparou os diferentes níveis de ajustes propostos na metodologia do trabalho, selecionando um modelo com parcimônia e compatível com os padrões de distribuição espacial encontrados nas áreas. Sugere-se o uso de modelos compartilhados para dados de amostragem em diferentes áreas, com introdução da estimativa do erro intraparcela (*nugget* fixo) nas equações de MV, para aumentar a correlação entre as U.A, com avaliação conjunta do AIC somado ao grau de dependência espacial na estimativa de variáveis dendrométricas.

Palavras-chave: máxima verossimilhança; silvicultura de precisão; critério de Akaike; inventário florestal.

1 Engenheiro Florestal, Dr., Professor do Departamento de Engenharia Florestal, Universidade do Estado de Mato Grosso, Campus II, Av. Perimetral Rogério Silva, s/n, Jardim Flamboyant, CEP 78580-000, Alta Floresta (MT), Brasil. juliocw@unemat.br

2 Engenheiro Florestal, Dr., Professor do Departamento de Ciências Florestais, Universidade Federal do Paraná, Av. Lothário Meissner, 900, Jardim Botânico, Campus III, CEP 80210-170, Curitiba (PR), Brasil. jarce@ufpr.br

3 Matemático, Dr., Professor da Pontifícia Universidade Católica, Rodovia BR 376, Km 14, Bairro Costeira, CEP 83010-500, São Jose dos Pinhais (PR), Brasil. sauloweber@gmail.com

4 Engenheiro Agrônomo, Ph.D., Professor do Departamento de Estatística, Centro Politécnico, Universidade Federal do Paraná, Ed. Administração, 3º andar, CEP 81531-990, Curitiba (PR), Brasil. paulojus@ufpr.br

5 Geólogo, Ph.D., Professor do Departamento de Geociências, Universidade Federal de Santa Maria, Av. Roraima, 1000, Cidade Universitária, CEP 97105-900, Santa Maria (RS), Brasil. calpis@terra.com.br

ABSTRACT

This study aimed to use the share parameters of the geo-statistical models applied to maximum likelihood estimators to predict the volumes per hectare in three fragments of a Deciduous Forest located in Santa Teresa, RS state, employing the 'Borrow Strength' approach. Data were collected in 56 sampling units (S.U) of variable sizes with approximately 250 m² for a total of nine ha, distributed in a systematic grid of 40 x 40 m. Dendrometric variables from individuals with DBH \geq 10 cm near the center of the S.U. were measured. Two approaches to the data set were prepared, the first of which considering both areas entirely independent themselves, subdivided into two types: a fit to non-spatial model (NSM) and a fit to the maximum likelihood (ML) not shared (individual adjustment) model. The second approach described the adjustment of the shared as a function of random error or nugget, comprising models: a shared model without fixed nugget (variability between S.U) and a shared model with fixed nugget (variability within S.U) models, using a logarithmic function of M.L applied to the *Matèrn* family of exponential correlation model. Then, the models were compared using Akaike information criterion (AIC) and by degree of spatial dependence for subsequent preparation of both kriging and prediction surfaces of the selected models. It was observed that the combined volume models to estimate values were higher for the AIC values and spatial dependence with respect to the adjustments for the individual areas. Among the shared models, it was observed that there was a gain in the parameter estimates using the fixed nugget, which resulted in a higher correlation of samples and spatial dependence (AP = 88 m), than the shared models without the fixed nugget (AP = 75 and 66 m). The AIC was efficient because it compared the different levels of proposed adjustments to the methodology of the study, selecting a model with parsimony and compatible with the spatial distribution patterns found in the areas. The use of combined models for data sampling in different areas with the introduction of the error estimate intra-plot (fixed nugget) in the equations of MV can be suggested to increase the correlation between the S.U and combined evaluation of the AIC plus the degree of spatial dependence in estimating dendrometric variables.

Keywords: maximum likelihood; precision silviculture; Akaike criterion; forest inventory.

INTRODUÇÃO

A complexidade dos levantamentos florestais em geral, em relação à organização, estrutura, coleta de dados e à logística nos locais de acesso, podem resultar em horas de trabalho e elevados custos operacionais, sendo proporcionais de uma população a ser inventariada.

Uma informação básica a ser respondida nos levantamentos e inventários florestais é sem dúvida o volume. Machado e Figueiredo Filho (2009) comentam que estimar o volume das árvores se configura como uma das principais finalidades dos levantamentos florestais. Assim sendo, a importância em estimar com precisão não apenas a variável em suas medidas usuais, mas também como ela se distribui no espaço, abre caminhos para inclusão de abordagens geoestatísticas criando novas concepções.

No contexto da amostragem para subpopulações em fragmentos florestais, diversas técnicas de modelagem avançadas são aplicadas para resolver estimativas e predições dos volumes e demais variáveis dendrométricas importantes. O uso de modelos mistos *mix-models*, (PINHEIRO; BATES, 2000), modelos de regressão condicional (BAS; PUIJENBROEK, 1989), modelos hierárquicos (FINLEY; BANERJEE; MACFARLANE, 2010), modelos heurísticos e meta-heurísticos (RONGXIA et al., 2010), modelos de multicritério (MALCZEWSKI, 1999; BUONGIORNO; GILLESS, 2003) e muitos outros usados nas diferentes áreas, são cada vez mais necessários no ramo da pesquisa florestal devido o crescente volume de dados e a complexidade de informações advindas de diferentes fontes de levantamentos.

Na modelagem quantitativa ajustam-se dados para uma determinada situação prática, e os modelos aceitos são muitas vezes testados e generalizados para outros locais ou regiões em condições semelhantes. Após a análise e comparação dos resultados, é escolhida a melhor estratégia a ser adotada frente a um problema proposto (MATTHIOPOULOS, 2011).

Nesse sentido, em um inventário baseado em várias subpopulações, ou classes consideradas *a priori* semelhantes, podem ser empregadas técnicas de *Borrow strenght* para "emprestar força" (KANGAS; MALTANO, 2006), entre uma área e outra, com a finalidade de compor a informação espacial de uma

determinada variável, através de um modelo único e uma distribuição conhecida, emprestando informação daquilo que lhes é comum.

Da mesma forma, é possível generalizar dados referentes à amostra para cobrir as árvores em que apenas as medidas básicas estão disponíveis, como o diâmetro, gerando assim, resultados com o potencial de unir informação de caráter dendrométrico e estimativas de caráter espacial. A proposta central do estudo foi utilizar os estimadores de máxima verossimilhança aplicados ao compartilhamento de parâmetros em uma abordagem *Borrow Strenght*, para avaliar e prever os volumes e sua estrutura de dependência espacial em três fragmentos de Floresta Estacional Subtropical.

MATERIAL E MÉTODO

Área de estudo

Os fragmentos localizam-se no Município de Santa Tereza, região nordeste do Estado do Rio Grande do Sul. As áreas somam 9 hectares sendo subdivididas por critérios de homogeneidade fisionômica e idade pós-intervenção em: floresta madura (FMA) com 75 anos; floresta em estágio secundário (FES) com 55 anos e floresta em estágio inicial (FEI) com 30 anos.

A distância entre as áreas FEI e FES é de 0,8 km, enquanto que a área mais afastada (FMA) dista 3,5 km em relação aos outros dois fragmentos. A FMA está situada nas coordenadas 29°11'00" S e 51°43'40" W, com altitude média de 280 m, exposição sudoeste e 4° de inclinação média. A FES está localizada nas coordenadas 29°09'29" S e 51°41'49" W, com altitude média de 180 m, exposição sudeste e 16° de inclinação média. A FEI, por sua vez, está localizada nas coordenadas 29°09'28" S e 51°42'05" W com altitude média de 215 m, exposição Norte e 5° de inclinação média.

O Clima é classificado como tipo Cfa, subtropical úmido segundo Köppen (MORENO, 1961), com chuvas bem distribuídas ao longo do ano, temperaturas do mês mais frio entre -3° e 18°C, e temperatura média do mês mais quente superior a 22°C. A temperatura média anual da região gira em torno de 18°C e a precipitação média anual é de aproximadamente 1800 mm.

A vegetação pertence à formação denominada Floresta Estacional Subtropical. O perfil do solo encontrado próximo ao centro de cada área experimental foi: Chernossolo Argilúvico Férrico típico na FMA, Cambissolo Háptico Ta Eutrófico típico na FES e um Neossolo Litólico Eutrófico típico na FEI (EMBRAPA, 2013).

Obtenção dos dados

As unidades amostrais (U.A) foram instaladas em um *grid* sistemático de 40 m, a partir de coordenadas métricas (UTM), com auxílio de imagens de satélite (Cbers 2B, Sensor CCD) e GPS (Garmin 12XL), distribuídas de forma a cobrir toda a extensão dos fragmentos sendo: 17 U.A na FEI, 20 na FES e 19 na FMA, totalizando 56 U.A com tamanho aproximado de 250 m² e com uma intensidade amostral mínima de 10 % (Figura 1)

Foram tomadas o diâmetro a altura do peito (DAP) e a altura das 8 árvores-amostra mais próximas do centro de cada parcela com DAP \geq 10 cm. O volume por hectare foi obtido a partir da metodologia descrita por Prodan (1965), utilizado um $f = 0,65$ como fator de forma do volume individual para florestas decíduais na região serrana do estado (SEMA-RS/UFSC, 2001).

Primeiramente foi realizada uma análise descritiva e exploratória com a finalidade de verificar a natureza dos dados, sua distribuição, quantidades e necessidades de transformações. A partir da tabulação do conjunto dos valores observados de volume e representação gráfica por meio do *Box-plot*, foram observadas as medidas estatísticas usuais como média; mediana e variância.

Fez-se a transformação dos dados pelo método de *Box-Cox* (BOX; COX, 1964), encontrando como parâmetro de transformação o valor de lambda igual a zero ($\lambda = 0$). A função Matérn (MATÈRN, 1986) foi utilizada para explicar a estrutura da covariância dos dados de volume nas áreas, com parâmetro de suavização $kappa = 0,5$; que resulta na aproximação da função ao modelo exponencial (1):

$$\rho(u) = \exp \left\{ - \left(\frac{u}{\Phi} \right)^K \right\} \quad (1)$$

Em que: $\rho(u)$ é a correlação entre os pares de pontos separados pela distância “ u ”, Φ é o alcance do modelo, e K é parâmetro de suavização *kappa*.

Na realização desse estudo foram elaborados dois métodos, o primeiro considerando as áreas independentes entre si, subdivididos em dois tipos: ajuste aos modelos de regressão simples e ajuste aos modelos não compartilhados (amostras analisadas e modeladas para cada fragmento separadamente). O segundo método descreve os ajustes dos modelos com compartilhamento dos parâmetros utilizando abordagem *Borrow Strength*, subdividida em relação ao *nugget* sendo: modelos compartilhados sem *nugget* fixo (variabilidade entre as U.A) e modelos compartilhados com *nugget* fixo (variabilidade dentro das U.A), nos quais serão descritos em detalhes nos itens que seguem.

Ajuste de parâmetros do modelo sem componente espacial

Primeiramente, testou-se a hipótese de que a amostragem das áreas em separado era suficiente para explicar a variabilidade dos volumes apenas com seus valores médios comparados aos modelos sem compartilhamento de parâmetros, ou seja, ajustados para cada área individualmente, sendo essa alternativa um caminho mais curto que envolveria menor esforço computacional.

Para tanto, admitiu-se que observações “ y_i ” do volume na posição “ x_i ” dado (x_i, y_i) , sendo $i = 1, 2, 3, \dots, n$, e $x_i = (x_{1i}, x_{2i})$ o vetor da localização espacial dos pares de pontos. Foi ajustado um modelo sem componente espacial “ $S(x_i)$ ” para as observações do volume $y_i = \mu(x_i)$, resultando na média associada a uma relação linear com as coordenadas x_{1i} e x_{2i} da área (2):

$$\mu(x_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} \quad (2)$$

Ajuste de parâmetros do modelo não compartilhado

O ajuste de uma função para cada área com os parâmetros, média (β), *nugget* (τ^2), variância espacial (σ^2) e alcance (Φ) independentes e não compartilhados foi obtido com a incorporação do modelo gaussiano à função da verossimilhança adicionando uma superfície de tendência linear “ $S(x_i)$ ” que segue a distribuição normal multivariada (*NMV*) por aproximação, conforme expressão (3) (DIGGLE; RIBEIRO JUNIOR, 2007).

$$S(x_i) \sim NMV [D\beta, \sigma^2 R(\Phi)] \quad (3)$$

Foi utilizado o modelo geoestatístico como base para o ajuste por máxima verossimilhança dos volumes em $m^3 ha^{-1}$. Sendo as observações $y_i = y_1, y_2, \dots, y_n$, normalmente distribuídas, ficando reescrito como (4):

$$y_i = D\beta + S(x_i) + \varepsilon_i \quad (4)$$

Em que: y_i = vetor da variável resposta; D = matriz das variáveis explicativas ou covariáveis; β = vetor dos parâmetros da regressão; $S(x_i)$ = matriz de efeitos aleatórios dada por $S(x_i) \sim NMV [0, \sigma^2 R(\Phi)]$, com média zero e variância σ^2 ; e $R(\Phi)$ é a matriz de covariâncias do elemento $\rho(u; \Phi, K)$. “ u ” na função de correlação “ ρ ” é a distância euclidiana entre os pontos observados ($u = \|x_i - x_j\|$), em função do grau de associação espacial Φ , ε_i é um termo independente associado ao erro aleatório com média “0” e variância “ τ^2 ”, chamado de *nugget effect* ou efeito pepita, “ I ” é a matriz identidade (RIBEIRO JUNIOR, 1995).

Ajuste de parâmetros do modelo compartilhado

Uma segunda análise refere-se à ideia de que, em uma estimativa conjunta, parâmetros do modelo

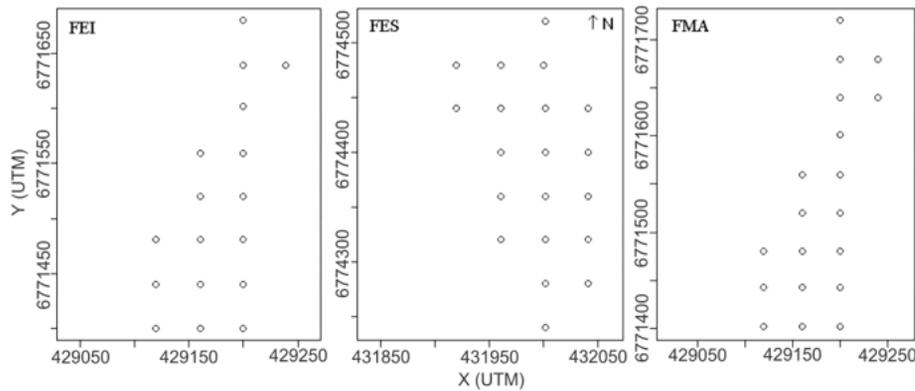


FIGURA 1: Distribuição das unidades amostrais nas áreas de FEI = floresta em estágio inicial, FES = floresta em estágio secundário, FMA = floresta madura.

FIGURE 1: Distribution of sample units in FEI, FES, FMA areas.

como “ β , τ^2 , σ^2 e Φ ” de um determinado local podem “emprestar força” ou informações daquilo em que uma área estiver em comum com a outra. De maneira que as estimativas do modelo possam explicar a variável resposta em outras áreas através de uma distribuição assumida, abordagem conhecida na literatura por *Borrow strenght* (BRILLINGER, 1997; KANGAS; MALTANO, 2006), nessa metodologia, os modelos foram subdivididos em: modelos estimados por máxima verossimilhança, sem *nugget* fixo e com *nugget* fixo, separando a variância dos volumes entre as U.A (parcelas) e dentro das U.A (árvores do mesmo ponto ou *kernel*) respectivamente.

A variância dentro das U.A é similar à variabilidade não explicada do efeito pepita, sendo esse parâmetro determinado *a priori* por meio de uma análise de variância (ANOVA), usando como fator o logaritmo dos volumes e o quadrado médio dos resíduos como estimativa do erro (τ^2) fora do variograma, reduzindo assim o número de parâmetros a serem estimados, na tentativa de que as relações de dependência espacial entre as U.A ocorressem de forma mais nítida, podendo assim, serem estimadas com maior precisão.

A expressão final do modelo geoestatístico compartilhado montada a partir da Equação (4), com 1.000 iterações e especificação nos parâmetros da função do logaritmo da verossimilhança L [β , θ (τ^2 , σ^2 , Φ)] em função da composição das áreas ficou escrita como (5):

$$\begin{aligned}
 &L\{(\beta^{(i)(j)(k^*)}), \theta [(\tau^{2(i)(j)(k^*)}), (\sigma^{2(i)(j)(k^*)}), (\Phi^{(i)(j)(k^*)})]\} \\
 &= - 0,5 \{n^{(i)(j)(k^*)} \log (2\mu) + \log \{(\sigma^{2(i)(j)(k^*)} R(\Phi^{(i)(j)(k^*)}) + \tau^{2(i)(j)(k^*)} I)\} \\
 &+ (y^{(i)(j)(k^*)} - D\beta^{(i)(j)(k^*)})^T (\sigma^{2(i)(j)(k^*)} R(\Phi^{(i)(j)(k^*)}) + \tau^{2(i)(j)(k^*)} I)^{-1} (y^{(i)(j)(k^*)} - D\beta^{(i)(j)(k^*)}) \} \quad (5)
 \end{aligned}$$

Em que: L = função log-verossimilhança do modelo geoestatístico; $\beta^{(i)(j)(k^*)}$; $\tau^{2(i)(j)(k^*)}$; $\sigma^{2(i)(j)(k^*)}$; $\Phi^{(i)(j)(k^*)}$ - parâmetros da média, efeito pepita, variância espacial e alcance para as áreas de FEI⁽ⁱ⁾, FES^(j), FMA^(k^{*}), e ^{(i)(j)(k^{*})} os índices que indicam as áreas onde os parâmetros podem ou não ser comuns.

Critério adotado para construção dos modelos de compartilhamento de força

Foram selecionados alguns modelos dos 2.880 possíveis com maior chance de aderência do ponto de vista biológico, assim sendo, muitos outros modelos foram testados e não entraram no *roll* da análise por fatores como: explicação biológica pouco provável ou problemas de não convergência devido à ocorrência de matrizes com covariâncias próximas a zero. Dessa forma, foi proposto o ajuste a um gradiente de modelos baseados na complexidade de parâmetros do modelo mais parcimonioso ou modelo mínimo em que todas as áreas são semelhantes para β , τ^2 , σ^2 e Φ (1,1,1 - 1,1,1 - 1,1,1 - 1,1,1), até o modelo máximo em que todas as áreas são distintas para cada parâmetro (1,2,3 - 1,2,3 - 1,2,3 - 1,2,3), de modo que as interpretações

TABELA 1: Sequência de compartilhamentos possíveis para os parâmetros β , τ^2 , σ^2 e Φ nos modelos.
 TABLE 1: Sequence of possible shares for the parameters β , τ^2 , σ^2 and Φ in models.

Compartilhamento	Interpretação
1,1,1	FEI = FES = FMA
1,1,2	(FEI = FES) \neq FMA
1,2,1	(FEI=FMA) \neq FES
1,2,2	FEI \neq (FES=FMA)
1,2,3	FEI \neq FES \neq FMA

Em que FEI = floresta em estágio inicial; FES = floresta em estágio secundário; FMA = floresta madura.

biológicas advindas dos ajustes em conjunto permitam um melhor uso da informação proveniente das diferentes áreas, ampliando o poder das análises, além de inferir em que aspectos as áreas são semelhantes ou diferentes (Tabela 1).

Critério de Seleção do modelo para estimativa de volume

Adotou-se o critério de informação de Akaike (AKAIKE, 1974) conhecido pela sigla “AIC”, para seleção do modelo, sendo considerado melhor o modelo com menor valor de AIC, calculado pela expressão (6):

$$AIC = - 2*(L-V) + k * Npar \quad (6)$$

Em que: L-V é o logaritmo da verossimilhança maximizada; k é o termo da fórmula que penaliza o modelo quanto maior for seu número de parâmetros ($k = 2$, para a fórmula clássica do AIC) e $Npar$ é o número de parâmetros estimados pelo modelo, sendo o mínimo 3 e o máximo é 12 parâmetros. O termo (L-V) tem uma interpretação oposta ao da função de verossimilhança e indica que a função com maior verossimilhança terá o menor valor da log-verossimilhança negativa, resultado da mudança de sinal decorrente da transformação dos dados. Foi utilizado para avaliar o comportamento da função frente à transformação logarítmica do conjunto de dados da variável resposta, no caso o volume por hectare.

Cada modelo foi relacionado com seu respectivo parâmetro (Φ) que representa o maior alcance resultante do compartilhamento de suas matrizes de covariância, visto que a melhor equação selecionada deveria necessariamente atender a duas condições: um menor valor do AIC e um alcance prático ($AP = 3\Phi$) superior à menor distância entre os pontos (40 m), em função do uso da equação selecionada nos mapas de predição dos volumes por hectare.

Todos os dados utilizados para os diferentes métodos foram tabulados e analisados utilizando o programa computacional R Core Team (2013) e os pacotes geoR (RIBEIRO JÚNIOR; DIGGLE, 2001), MASS (VENABLES; RIPLEY, 2002) e sp (PEBESMA; BIVAND, 2005). As etapas que envolveram a análise da dependência espacial, os tratamentos, processos e ajustes dos modelos bem como, as estimativas para o volume, encontram-se ordenados de acordo com o fluxograma, sendo: AED – análise exploratória de dados; MV – máxima verossimilhança; Nsm – modelo sem componente espacial; MV compartilhado sem tau (nugget) fixo; MV compartilhado com tau (nugget) fixo, AIC + AP – critério de informação de Akaike + alcance prático (Figura 2).

RESULTADO E DISCUSSÃO

Estatística descritiva e análise exploratória

O comportamento da distribuição e simetria do conjunto de dados em relação aos valores médios do volume ($m^3 ha^{-1}$) podem ser observados nos gráficos de *Box-plot* (Figura 3). Os valores localmente

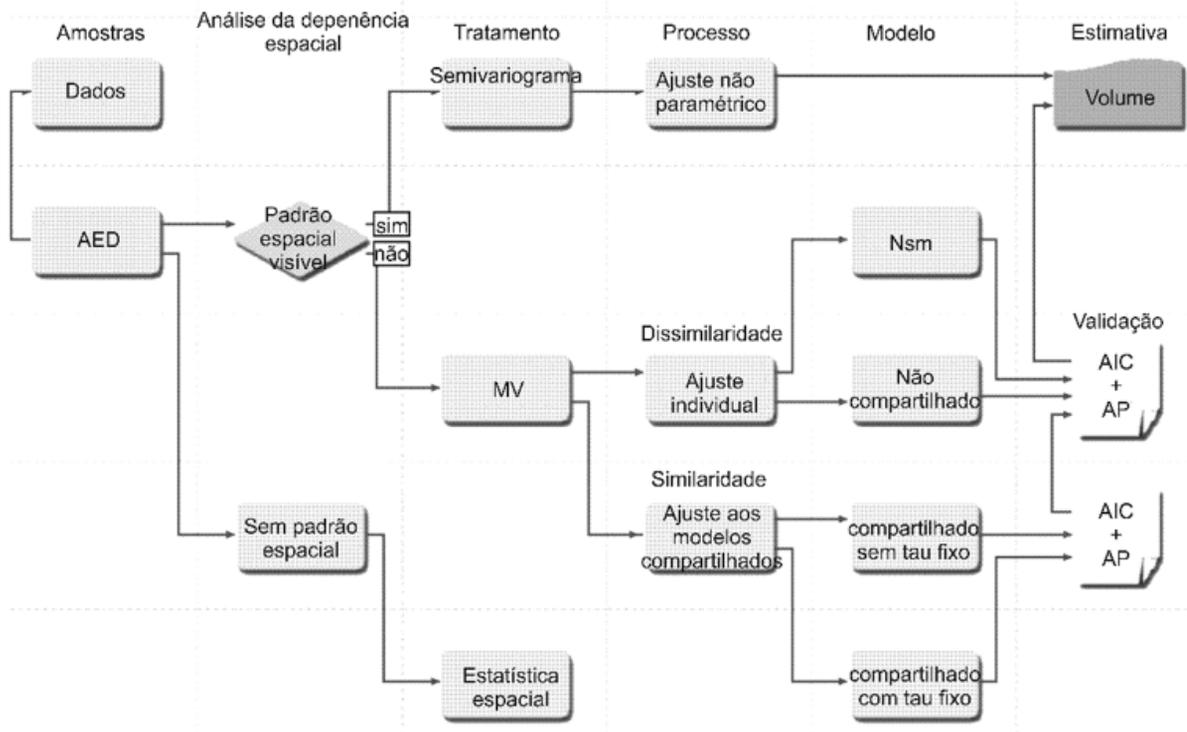


FIGURA 2: Fluxograma da metodologia de análise dos dados, ajuste e seleção dos modelos.

FIGURE 2: Flowchart of the methodology of data analysis, setting and selection of models.

atípicos foram removidos antes dos ajustes dos modelos, por influenciarem negativamente pressupostos da normalidade assumidos (MELLO, 2004).

Na FEI, os valores mínimos e máximos da distribuição foram, respectivamente, 26,3 m³ e 125,0 m³ com média de 56,8 m³ por hectare, divididos em intervalos interquartílicos de 40,2; 50,2 e 66,6 m³ respectivamente para 25%, 50% e 75% do total do conjunto de dados observados.

Na FES, os valores mínimos e máximos da distribuição foram 25,1 m³ e 249,0 m³ respectivamente, com média de 90,3 m³ por hectare, com intervalos interquartílicos de 42,4; 73,1 e 111 m³ por hectare, respectivamente para 25%, 50% e 75% dos dados.

A FMA obteve valores de média igual a 82,0 m³ por hectare, com distâncias interquartílicas de

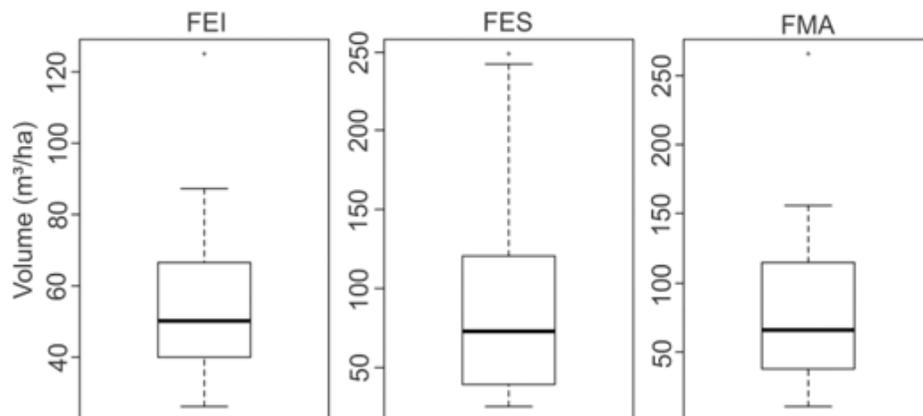


FIGURA 3: Gráfico *Box-plot* da dispersão do volume (m³ ha⁻¹) nas áreas de FEI = floresta em estágio inicial; FES = floresta em estágio secundário; FMA = floresta madura.

FIGURE 3: Box-plot graph of the volume dispersion (m³ ha⁻¹) in FEI, FES, FMA areas.

37,8, 66,1 e 115,1 m³ respectivamente para o 1º, 2º e 3º quartil da distribuição de dados. As áreas obtiveram valores de coeficiente de variação máximo de 30%.

Seleção dos modelos não compartilhados pelo critério de AIC

Observam-se os resultados na Tabela 2 do teste de AIC para os modelos sem o componente espacial (Nsm) e os modelos de máxima verossimilhança (MV) não combinados ajustados individualmente para cada área. Todos os modelos lineares Nsm obtiveram um valor de AIC menor que os modelos ajustados individualmente. Mesmo com a inclusão de dois parâmetros nos modelos, os valores do AIC foram próximos, ficando a diferença entre os métodos, nesse caso, apenas relacionada ao número de parâmetros.

Tanto os modelos lineares Nsm, quanto os modelos ajustados por MV obtiveram uma relação de continuidade espacial para os volumes abaixo da distância mínima entre os pontos das U.A (40 m), com exceção da área FES na qual teve um alcance prático de 44,03 m.

Seleção dos modelos compartilhados pelo critério de AIC

Pode-se observar na Tabela 3, o ranking dos modelos ajustados por MV logarítmica sem *nugget* fixo (fit01 ao fit08) e com *nugget* fixo (fixnug09 ao fixnug13) em ordem crescente de AIC. O erro estimado fora do semivariograma para os modelos com *nugget* fixo foi: 1,597; 2,001 e 3,076 respectivamente para FEI, FES e FMA.

O modelo “fixnug11” foi o melhor, com um AIC de 120,10 e um AP de 82 m nas áreas FEI e FES, ficando a FMA com um AP de 10 m resultante dos efeitos fixos aplicados à matriz de covariância do modelo. Calegario et al. (2005) em estudos com modelos não lineares de efeito misto em dois níveis com efeitos fixos e aleatórios (erros nas U.A e nas árvores dentro de cada unidade) para estimativas de alturas com dados desbalanceados, concluíram que a inclusão de tais níveis reduziu significativamente o erro associado aos estimadores no modelo.

No modelo “fixnug12” a correlação entre as amostras foi maior na área FES com um valor do AP de 47 m, enquanto que na FEI e FMA atingiu 6,14 m, nota-se que a dependência espacial nessa condição, explicou a variabilidade dos volumes em apenas uma das áreas, no caso a área “FES”, enquanto que o modelo de menor AIC “fixnug11” cobriu a variabilidade em duas das três áreas “FEI-FES”. O valor do AIC para o modelo “fixnug12” foi próximo do modelo “fixnug11” menos parcimonioso.

Para os modelos combinados sem *nugget* fixo, o melhor valor de AIC encontrado foi no “fit01”, porém, o alcance nas áreas foi pequeno (AP = 9 m), indicativo de que a correlação entre os pontos na

TABELA 2: Critério de seleção de Akaike (AIC) aplicado aos modelos linear sem o componente espacial (Nsm) e Máxima verossimilhança (MV) nas áreas FEI = floresta em estágio inicial, FES = floresta em estágio secundário, FMA = floresta madura.

TABLE 2: Selection Akaike criteria (AIC) applied to linear models without the spatial component (Nsm) and maximum likelihood (ML) in FEI, FES, FMA areas.

Modelo	Áreas	L-V	Npar	AP	AIC
Nsm	FEI	-75,66	2	8,30	155,3
	FES	-111,9	2	44,03	227,8
	FMA	-105,1	2	0,01	214,3
MV	FEI	-75,66	4	8,30	159,3
	FES	-111,9	4	44,03	231,3
	FMA	-105,1	4	0,01	218,3

Em que: L-V = log-verossimilhança; Npar = número de parâmetros; (AP) = alcance prático (3Φ para $\kappa = 0,5$), (AIC) = Critério de Informação de Akaike; FEI = floresta em estágio inicial; FES = floresta em estágio secundário; FMA = floresta madura.

TABELA3: Critério de seleção de Akaike (AIC) aplicado aos modelos com compartilhamento de força nas áreas FEI = floresta em estágio inicial; FES = floresta em estágio secundário; FMA = floresta madura.

TABLE 3: Selection Akaike criteria (AIC) applied to models with power sharing in FEI, FES, FMA areas.

Modelo	Parâmetros				L-V	Npar	(AP)	AIC
	β	τ^2	σ^2	Φ				
fixnug11*	1,1,2	-	1,2,2	1,1,2	-54,0	6	(82) (82) (10)	120,10
fixnug12	1,2,3	-	1,2,2	1,2,1	-53,9	7	(6) (47) (6)	120,11
fixnug09	1,1,1	-	1,1,1	1,1,1	-57,1	3	(1) (1) (1)	120,24
fixnug13	1,2,3	-	1,2,3	1,2,3	-52,9	9	(1) (46) (5)	123,83
fixnug10	1,1,1	-	1,1,1	1,1,2	-57,0	5	(44) (44) (8)	124,15
fit01	1,1,1	1,1,1	1,1,1	1,1,1	-58,4	4	(9) (9) (9)	124,99
fit04	1,1,2	1,1,2	1,2,3	1,2,2	-54,0	9	(2) (66) (66)	125,02
fit06	1,1,2	1,1,2	1,2,3	1,1,2	-54,0	9	(75) (75) (1)	125,11
fit02	1,2,3	1,1,2	1,1,1	1,1,1	-55,5	7	(16) (16) (16)	125,18
fit07	1,2,3	1,2,3	1,1,1	1,2,3	-52,9	10	(3) (47) (9)	125,86
fit05	1,1,2	1,2,3	1,2,2	1,2,2	-54,2	9	(8) (35) (35)	126,59
fit03	1,1,2	1,1,2	1,1,2	1,1,2	-56,7	8	(39) (39) (10)	129,55
fit08	1,2,3	1,2,3	1,2,3	1,2,3	-52,8	12	(1) (44) (1)	129,77

Em que: * = melhor modelo pelo critério de Akaike; β = vetor de médias; τ^2 = *nugget*; σ^2 e Φ = correlação espacial; L-V = log-verossimilhança; Npar = número de parâmetros; (AP) = alcance prático (3Φ para $\kappa = 0,5$), na ordem FEI = floresta em estágio inicial; FES = floresta em estágio secundário; FMA = floresta madura.

condição de igualdade para as 3 áreas é fraca.

Os modelos “fit04” e “fit06” aparecem como boas alternativas para estimativas de superfície, com valores de AP de 66 m (FES-FMA) e 75m (FEI-FES) respectivamente, bem como os valores de AIC próximos ao modelo “fit01”.

A análise em conjunto do grau de dependência espacial somado ao AIC é importante, pois facilita a escolha de um modelo mais adequado para a interpolação da superfície por meio da krigagem. Não houve entre os modelos testados um que obtivesse grau de dependência espacial para os volumes em todas as áreas acima do *grid* mínimo da amostragem (40 m), indicando que a variabilidade espacial possa estar ocorrendo em distâncias menores, pelos menos em uma das três áreas analisadas. Kent e Dress (1979) comentaram que em florestas naturais, a mortalidade e ingresso de florestas jovens e maduras são fatores que influenciam na formação de padrões espaciais de uma população durante o seu desenvolvimento, interferindo pela ação competitiva de seus vizinhos.

Comparando os valores de AIC entre os métodos de ajuste não compartilhados e compartilhados, nota-se que a combinação das áreas obteve sucesso em melhorar as estimativas dos parâmetros em todos os modelos, pois, o maior valor de AIC dos modelos compartilhados (fit08, AIC = 129,77), é inferior ao menor valor encontrado nos modelos não compartilhados (Nsm, AIC = 160,2).

Na comparação entre os modelos compartilhados, os modelos com *nugget* fixo foram superiores aos sem *nugget* fixo, evidenciando que houve um ganho em relação às estimativas expressas pelos resultados do AIC quando modelado o efeito aleatório dentro das U.A. Gregoire e Schabenberger (1994) utilizaram técnica similar ao *nugget* fixo, para efeitos mistos de modelos não lineares estimados por MV restrita e também por meio de equações de estimativas generalizadas (GEE) para prever o volume de árvores individuais, encontraram melhorias nas estimativas dos modelos quando separadas a variância residual aleatória da variância total.

O modelo “fit08” com 12 parâmetros que relaciona a média dos volumes, o *nugget*, a variabilidade espacial e o alcance como características individuais de cada área resultaram em um valor alto pelo critério do AIC (129,77). De acordo com os critérios estabelecidos de parcimônia, maximização da função log-

verossimilhança, e grau de dependência espacial, esse modelo obteve uma relação fraca do ponto de vista da correlação espacial dos volumes nos locais amostrados, explicado novamente pelos valores pequenos do alcance “AP = 1 m” nas áreas FEI e FMA e “AP = 44 m” na área FES.

Em relação aos valores de AIC obtidos por cada técnica de ajuste, observa-se que foi visível o escalonamento dos modelos em relação ao refinamento da modelagem empregada nos ajustes não espaciais, MV ajustado as áreas individualmente e os modelos compartilhados, relacionando corretamente os ganhos obtidos na comparação entre um ajuste e outro. Para Burnham e Anderson (2002), o teste de AIC não somente seleciona o modelo mais correto, mas também o quanto, sendo utilizado para qualquer tipo de modelo, tanto lineares, como não lineares, aninhados ou não aninhados.

Os resultados dos volumes totais obtidos pelo método de Prodan foram 295,35 m³/ha na FEI e 435,08 m³/ha na FES, enquanto que os totais estimados pelo modelo compartilhado “fixnug11” foram 300,95 e 435,03 m³/há, respectivamente para FEI e FES. Não houve estimativa do volume para a área FMA pois o alcance ficou abaixo (AP = 10 m) do espaçamento mínimo entre as unidades amostrais (40 m).

CONCLUSÕES

O modelo “fixnug11” foi o melhor dentre todos testados apresentando um AIC de 120,10 e um alcance prático para as áreas de FEI e FES de 82 m.

Os modelos compartilhados com *nugget* fixo foram superiores pelo critério de Akaike em relação aos modelos sem *nugget* fixo, sugerindo que o efeito do erro aleatório (τ^2) mensurado fora do semivariograma (intraparcela) pode aumentar a correlação da dependência espacial dos volumes nas áreas.

Em comparação com os modelos com parâmetros não compartilhados, os modelos compartilhados foram superiores para os valores de AIC, grau de dependência espacial e valor máximo de função do logaritmo da verossimilhança (L-V).

O critério de informação de Akaike (AIC) foi eficiente, uma vez que comparou os diferentes níveis de ajustes propostos na metodologia do trabalho, selecionando um modelo compatível com os padrões de distribuição espacial dos volumes nas áreas.

Foi possível recuperar informações de locais com baixa amostragem através dos modelos compartilhados, gerando estimativas de volume por hectare distribuídas espacialmente.

A pequena diferença encontrada pela estimativa utilizando o método tradicional em relação ao volume estimado pelo modelo compartilhado “fixnug11” torna esse método viável com a grande vantagem de possibilitar a informação especializada, de difícil obtenção principalmente em florestas naturais.

Sugere-se o uso de modelos geoestatísticos com abordagem compartilhada *Borrow strenght* para estudos de variabilidade espacial em áreas fragmentadas com tipologias semelhantes e presença de correlação espacial.

AGRADECIMENTOS

O autor agradece ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela concessão da bolsa de doutorado.

REFERÊNCIAS

- AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on Automatic Control**, Boston, v. 19, n. 6, p. 716–723, dec. 1974.
- BAS, D.; VAN PUIJENBROEK, R. A. G. Conditional Regression Analysis. **Quality & Quantity**, Dordrecht, v. 23, n. 1, p. 83-108, feb. 1989.
- BOX, G. E. P.; COX, D. R. An Analysis of Transformations. **Journal of the Royal Statistical Society**, London, v. 26, n. 2, p. 211-252, 1964.
- BUONGIORNO, J.; GILLESS, J. K. **Decision methods for forest resource management**. San Diego: Academic Press, 2003. 429 p.
- BRILLINGER, D. **Selected works in probability and statistics**. Oslo: Springer, 1997. 609 p.

- BURNHAM, P. K.; ANDERSON, D. R. **Model selection and multimodel inference – a practical information - theoretic approach**. 2. ed. Colorado: Springer, 2002. 347 p.
- DIGGLE, P. J.; RIBEIRO JUNIOR, P. J. **Model-Based geostatistics**. New York: Springer, 2007. 228 p.
- EMBRAPA. **Sistema brasileiro de classificação de solos**. 3. ed. Brasília: EMBRAPA, 2013. 353 p.
- FINLEY, A. O.; BANERJEE, S.; MACFARLANE, D. W. **A hierarchical model for quantifying forest variables over large heterogeneous landscapes with uncertain forest areas**. Michigan: Department of Forestry, 2010. 88 p.
- GREGOIRE, T. G.; SCHABENBERGER, O. Fitting bole-volume equations to spatially correlated within-tree data. In: ANNUAL CONFERENCE OF APPLIED STATISTICS IN AGRICULTURE, 6., 1994, Manhattan. **Proceedings...** Kansas, 1994. p. 120-133.
- KANGAS, A.; MALTANO, M. **Forest inventory – methodology and applications**. New York: Springer, 2006. 350 p.
- MACHADO, S. A.; FIGUEIREDO FILHO, A. **Dendrometria**. Curitiba: Os autores, 2003. 309 p.
- MALCZEWSKI, J. **GIS and multicriteria decision analysis**. New York: Wiley, 1999. 298 p.
- MATÈRN, B. **Spatial Variation**. Berlin: Springer, 1986. 594 p.
- MATTHIOPOUOS, J. **How to be a quantitative ecologist – the ‘A to R’ of green mathematics and statistics**. Scotland: Wiley, 2011. 1091 p.
- MELLO, J. M. **Geoestatística aplicada ao inventário florestal**. 2004. 110 f. Tese (Doutorado em Recursos Florestais) - Escola Superior de Agricultura Luiz de Queiroz, Piracicaba, 2004.
- MORENO, J. A. **Clima do Rio Grande do Sul**. Porto Alegre: Secretaria da Agricultura, 1961. 73 p.
- PEBESMA, E. J.; BIVAND, R. S. Classes and methods for spatial data in R. **R News**, Ontario, v. 5, n. 2, 2005.
- PINHEIRO, J.; BATES, D. **Mixed-Effects Models in S and S-PLUS**. New York: Springer, 2000. 530 p.
- PRODAN, M. **Holzmesslehre**. Frankfurt: Sauerlände's Verlag, 1965. 644 p.
- R CORE TEAM. **R: A language and environment for statistical computing**. Vienna: R Foundation for Statistical Computing, 2013. Disponível em: <<http://www.R-project.org/>>. Acesso em: 12 abr. 2014.
- RIBEIRO JÚNIOR, P. J. **Métodos geoestatísticos no estudo da variabilidade espacial dos parâmetros do solo**. 1995. 99 f. Dissertação (Mestrado) - Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 1995.
- RIBEIRO JUNIOR, P. J.; DIGGLE, P. J. The geoR package functions for geostatistical data analysis. **R News**, Ontario, v. 1, n. 2, p. 15-18, 2001.
- RONGXIA, L.; PETE, B.; KEVIN, B. Informed development of meta heuristics for spatial forest planning problems. **The Open Operational Research Journal**, Sharjah, n. 4, p. 1-11, 2010. RIO GRANDE DO SUL. Secretaria Estadual do Meio Ambiente. **Relatório Final do Inventário Florestal Contínuo do Rio Grande do Sul**. Porto Alegre: SEMA/RS, 2001. 2 v. 706 p.
- VENABLES, W. N.; RIPLEY, B. D. **Modern applied statistics with S**. 4th ed. New York: Springer, 2002.