

# SELEÇÃO E AJUSTE DE MODELOS ESPACIAIS VISANDO A ESTIMAÇÃO DO VOLUME EM UM PLANTIO DE *EUCALYPTUS SP*

Júlio César PEREIRA<sup>1</sup>

Verônica SCALET<sup>2</sup>

Cláudio Roberto THIERSCH<sup>1</sup>

- RESUMO: O objetivo do trabalho foi utilizar modelos geoestatísticos para prever volume de um plantio de *Eucalyptus sp.* Para tanto, dispunha-se das covariáveis material genético (MG), densidade, altura dominante (Hd) e área basal (Ab). O primeiro passo foi selecionar um modelo para Hd, tomando como covariáveis MG e densidade. Em seguida ajustou-se modelo para Ab, em que a Hd também entrou como covariável. Finalmente, ajustou-se um modelo para a variável “volume”, sendo que todas as demais variáveis disponíveis foram utilizadas como covariáveis. Utilizou-se os métodos dos mínimos quadrados ordinários e ponderados, da máxima verossimilhança e da máxima verossimilhança restrita para ajuste dos modelos e para seleção utilizou-se o critério da informação de Akaike, o erro quadrático médio e o erro absoluto médio. O estudo realizado permitiu verificar que o uso de componente espacial e de covariáveis nos modelos para: altura dominante, área basal e volume, melhorou o ajuste e a capacidade de predição dessas variáveis, sendo esta melhoria mais evidente na variável “volume”.
- PALAVRAS-CHAVE: Geoestatística; inventário florestal; covariáveis.

## 1 Introdução

No setor florestal, em especial na silvicultura, conhecer a produtividade de uma área é de grande importância para o planejamento da colheita, logística e transporte de madeira. Tal produtividade depende de muitos fatores, tais como tipo de solo, clima, material genético, densidade de plantio, tratamentos silviculturais, entre outros.

Para se avaliar o crescimento e consequentemente a produtividade de uma floresta, o inventário florestal se torna uma ferramenta de grande importância (PÉLICO NETO e BRENA, 1997; MELLO, 2004; SOARES et al., 2007). Ao se realizar um inventário, uma das variáveis de maior interesse é o volume. Com o intuito de explicar a produção em volume e de se estimar de maneira precisa essa variável, tem-se tornado crescente o número de trabalhos na literatura que fazem o uso de covariáveis, tais como o material

---

<sup>1</sup> Universidade Federal de São Carlos - UFSCar, Centro de Ciências e Tecnologias para Sustentabilidade, Departamento de Ciências Ambientais, CEP: 13506-900, Sorocaba, SP, Brasil. E-mail: [julio.pereira.ufscar@gmail.com](mailto:julio.pereira.ufscar@gmail.com), [crthiersch@ufscar.br](mailto:crthiersch@ufscar.br)

<sup>2</sup> Universidade Federal de São Carlos - UFSCar, Centro de Ciências e Tecnologias para Sustentabilidade, Programa de Pós-graduação em Ciências dos Materiais, CEP: 13506-900, Sorocaba, SP, Brasil. E-mail: [ve.scalet@gmail.com](mailto:ve.scalet@gmail.com)

genético, densidade de plantio, área basal, altura dominante entre outras (ODA-SOUZA et al., 2008; BOGNOLA et al., 2008; MISIR, 2010; GUIMARÃES et al., 2009).

Geralmente os dados obtidos de um inventário florestal são submetidos a uma análise estatística, utilizando-se modelos de regressão linear tradicionais. Porém, esses modelos assumem que os erros são independentes, o que nem sempre ocorre com dados de inventário, pois medições das variáveis de interesse, realizadas em parcelas vizinhas, apresentam comportamento semelhante, isto é, dependência espacial. Dessa forma, o uso da geoestatística possibilitaria a modelagem dessa dependência (SCHMIDT e SANSÓ, 2006) e a obtenção de estimativas mais precisas do volume de madeira de um povoamento (MELLO et al., 2009).

Nesse sentido, o presente trabalho teve por objetivo comparar modelos geoestatísticos para a predição de volume de *Eucalyptus sp* em uma fazenda na região de Três Lagoas – MS a partir do uso de covariáveis como: densidade, material genético, altura dominante e área basal. Para tal objetivo foi, também, necessário selecionar modelos para a predição da altura dominante e da área basal.

## 2 Material e métodos

### 2.1 Material

O conjunto de dados utilizado no presente estudo provém de um plantio de *Eucalyptus SP*, com idade variando de 2,8 a 3,2 anos, pertencente a uma fazenda situada em Três Lagoas-MS. Foram amostradas 381 parcelas circulares de aproximadamente 400 m<sup>2</sup> cada, com intensidade amostral de uma parcela a cada dez hectares. Cada parcela avaliada foi georreferenciada no sistema SAD69, dada às coordenadas em metros, o que permitiu o cálculo das distâncias entre as parcelas. A Figura 1 apresenta mapa com a disposição das parcelas.

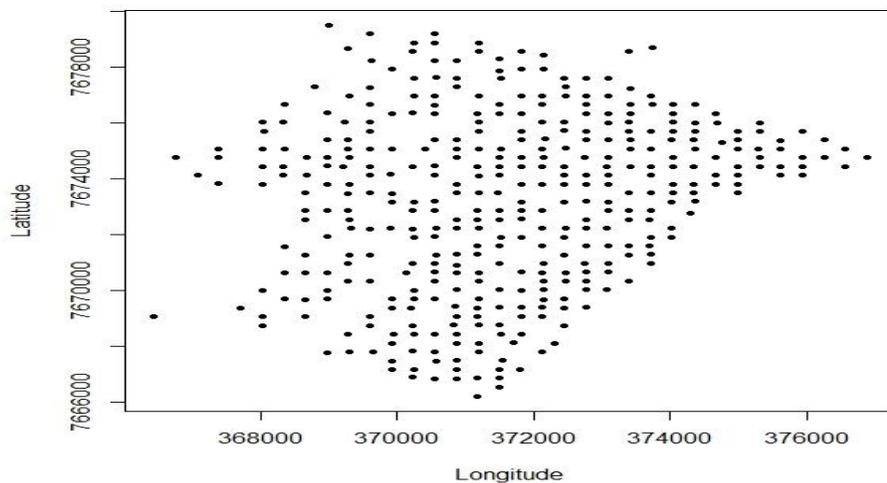


Figura 1 - Disposição das parcelas em campo.

As variáveis observadas foram: material genético, espaçamento e densidade do plantio, idade, altura dominante e área basal. Foram, também, obtidas amostras de volume de árvores próximas de cada parcela, a fim de se obter uma medida do volume dentro da parcela, porém as equações utilizadas para gerar o volume por parcela não foram divulgadas pela empresa detentora dos dados. A variável “espaçamento” é composta por dois níveis,  $3,6 \times 2,2\text{m}$  e  $3,6 \times 2,5\text{m}$ , enquanto que a variável “material genético” é composta por doze tipos distintos de clones. A densidade de plantio foi obtida pela razão entre o número de fustes e área de cada parcela, a qual variou de 1,75 a 3,375 fustes/m<sup>2</sup>, devido aos diferentes espaçamentos e por se tratar de uma floresta em segunda rotação com vários fustes por árvore.

## 2.2 Métodos

Aos dados descritos na seção anterior foram ajustados modelos com e sem componente espacial. Porém, foram removidas aleatoriamente 60 parcelas (aproximadamente 15% do total), isto é, estas parcelas foram deixadas de fora do processo de inferência a fim de se fazer nelas, predições do volume, usando os modelos ajustados, para que posteriormente fosse possível a comparação entre os modelos quanto à sua capacidade preditiva.

Convém observar que as covariáveis “material genético” e “densidade” eram conhecidas em todos os talhões, inclusive para as parcelas retiradas do processo de inferência. Sendo assim, foram utilizadas na predição da altura dominante, da área basal e do volume.

Foram, então, ajustados e selecionados os modelos para as variáveis: altura dominante, área basal e volume, nessa ordem. De posse de um modelo selecionado para a altura dominante, foram utilizadas as covariáveis consideradas conhecidas em todo o talhão para a predição dessa variável nas parcelas deixadas de fora do processo de inferência. Após a predição da altura dominante, utilizou-se um modelo selecionado para a área basal e realizaram-se predições dessa variável nas parcelas deixadas de fora do processo de inferência, sendo que as alturas dominantes preditas foram, neste momento, utilizadas como covariáveis. E finalmente, foi selecionado um modelo para o volume, em que a altura dominante e a área basal preditas entraram como covariáveis para a predição do volume naquelas parcelas deixadas de fora do processo de inferência.

No que se segue são descritos os modelos empregados nas análises e comparados entre si.

### Modelos sem componente espacial

Neste trabalho um modelo para qualquer uma das características, altura dominante, área basal ou volume, considerando suas observações espacialmente independentes, foi escrito genericamente como:

$$Y = X\beta + \varepsilon \quad (1)$$

em que  $X$  é a matriz delineamento,  $\beta$  é o vetor de coeficientes do modelo e  $\varepsilon$  é o vetor de erros aleatórios, independentes e identicamente distribuídos, tais que  $\varepsilon_i \sim N(0, \sigma_y^2)$ ,  $i = 1, 2, \dots, n$ .

Considerando o modelo para a altura dominante, na Equação 1  $Y_{n \times 1}$  é um vetor composto de valores de altura dominante observados nas parcelas. Considerando ainda material genético e densidade como covariáveis, a matriz  $X$  é composta de 13 colunas, cuja primeira coluna é constituída de uns, as colunas 2 a 12 referem-se às variáveis *dummy* ( $d_1, d_2, \dots, d_{11}$ ) que caracterizam o tipo de material genético e a coluna 13 é composta de observações da densidade. Para uma linha qualquer da matriz  $X$ , se esta linha se refere a uma parcela de material genético 1, por exemplo, então  $d_1 = d_2 = \dots = d_{11} = 0$ ; se se refere ao material genético 2, tem-se  $d_1 = 1$  e todos os demais  $d$ 's iguais a zero.

O modelo para a área basal também pode ser escrito de forma geral, como apresentado na Equação 1, porém, neste caso  $Y_{n \times 1}$  é o vetor composto pelos valores observados da área basal; a matriz  $X$  é composta de 14 colunas, em que as 13 primeiras colunas são como descritas no caso anterior e a 14ª coluna é constituída dos valores de altura dominante observadas nas parcelas amostradas.

Para o volume o modelo pode, ainda, ser representado de maneira geral como na equação 1, porém neste caso  $Y_{n \times 1}$  é o vetor das observações dos volumes, a matriz  $X$  tem 15 colunas, sendo a 15ª composta pelos valores da área basal.

### Modelos com componente espacial

Além dos modelos de regressão linear tradicionais, representados de forma geral pela Equação 1, foram ajustados modelos com componentes espaciais, para cada uma das variáveis (altura dominante, área basal e volume). Seja  $y(s_1), y(s_2), \dots, y(s_n)$ , uma amostra da variável de interesse observada ao longo da região em estudo, considerando a amostra como uma realização parcial de um processo gaussiano, um modelo com componente espacial pode ser escrito da seguinte forma:

$$Y(s) = X(s)\beta + v(s) + \varepsilon(s) \quad (2)$$

em que  $X(s)\beta$  é uma componente de tendência,  $X(s)$  é a matriz contendo os níveis das possíveis covariáveis, em uma parcela de localização denotada por  $s$ ,  $\beta$  é o vetor de coeficientes do modelo;  $v(s)$  é um efeito aleatório com estrutura de correlação espacial tal que  $(v(s) | \sigma, \varphi)$  segue um processo gaussiano,  $(v(s) | \sigma, \varphi) \sim PG(0, \sigma^2 \rho(\|s - s'\|; \varphi))$ , sendo  $\sigma^2$  a variância do processo e  $\rho(\cdot; \varphi)$  representa a função de correlação que depende do parâmetro  $\varphi$ ;  $\varepsilon(s)$  é uma componente de ruído branco, normalmente distribuída, com média 0 e variância  $\tau^2$  (efeito pepita), que descreve a variação de microescala.

Considerando o modelo com componente espacial (Equação 2) para a altura dominante, com material genético e densidade como covariáveis,  $Y(s)$  representa a variável altura dominante em uma localização  $s$  qualquer, por:

$$\mathbf{X}(s) = [1 \ d_1(s) \ d_2(s) \ \dots \ d_{11}(s) \ dens(s)]$$

em que  $d_1(s) \ d_2(s) \ \dots \ d_{11}(s)$  são variáveis *dummy*, como descritas para o modelo sem componente espacial, utilizadas para descrever o material genético na parcela de localização  $s$  e  $dens(s)$  representa a densidade observada na parcela de localização  $s$  e  $\beta$  é um vetor coluna de dimensão igual ao número de colunas de  $\mathbf{X}(s)$ .

O modelo com componente espacial para a área basal pode ser representado também pela equação 2, porém, neste caso,  $Y(s)$  representa a área basal em uma localização  $s$  e a matriz  $\mathbf{X}(s)$  contém uma coluna a mais representando a altura dominante ( $hdom(s)$ ), a qual foi considerada como uma covariável para a área basal.

Para a variável volume o modelo com componente espacial, também é representado pela Equação 2, entretanto,  $Y(s)$  representa o volume para a parcela de localização  $s$ ,  $\mathbf{X}(s)$  tem 15 colunas, pois altura dominante e área basal entram como covariáveis no modelo para o volume.

Observa-se que a componente aleatória  $v(s)$  na Equação 2, depende de uma função de correlação  $\rho(\cdot; \varphi)$ . Em Diggle e Ribeiro Júnior (2007), pode ser encontrado uma descrição de diversas funções que podem ser utilizadas para descrever a correlação espacial. Assim, para cada uma das variáveis (altura dominante, área basal e volume), foram testadas as funções de correlação exponencial, gaussiana, esférica e Matérn ( $K = 0,3$ ), a fim de selecionar aquelas que melhor descrevem a estrutura de correlação espacial para cada uma das variáveis analisadas.

A fim de se verificar a importância das covariáveis dada a presença da componente espacial no modelo, foram ajustados também os modelos com componente espacial sem a presença das covariáveis. Dessa forma, foi possível a comparação entre o modelo sem componente espacial, os modelos com componente espacial e os modelos com componente espacial sem a presença das covariáveis.

### Método de ajustes e critérios de comparação

A estimação dos parâmetros dos modelos ajustados foi feita pelos métodos da Máxima Verossimilhança, da Máxima Verossimilhança Restrita, dos Mínimos Quadrados Ordinários e dos Mínimos Quadrados Ponderados. Tanto para o ajuste dos modelos, quanto para a realização das previsões, utilizou-se o pacote geoR (RIBEIRO JÚNIOR e DIGGLE, 2001) e o *software* R (R CORE TEAM, 2014).

A comparação dos modelos foi realizada utilizando-se o critério de informação de Akaike (1974) e técnicas de validação cruzada. Valores baixos do critério de informação de Akaike (AIC) são considerados como representativos de um melhor ajuste, dessa forma, os modelos são selecionados visando a obter um mínimo AIC. Na comparação de modelos pelo critério AIC, segundo Hardin e Hilbe (2007) dois modelos que apresentam uma diferença entre AIC maior que 2 podem ser considerados distintos.

Foram realizadas predições nas 60 parcelas removidas e a comparação entre tal predição e o valor real observado foi feita através do Erro Quadrático Médio (EQM) e do Erro Absoluto Médio (EAM).

### 3 Resultados

Das variáveis apresentadas na seção 2.1, a idade e o espaçamento não foram considerados como covariáveis nos modelos descritos nas Equações 1 e 2, pois não foram significativas para explicar a variação da altura dominante, da área basal e do volume, conforme os resultados do algoritmo de seleção de variáveis, usando a função *stepAIC* do software R (R CORE TEAM, 2014).

A fim de se ter uma noção do comportamento das variáveis: altura dominante, área basal e volume, em termos de tendência central, posição e dispersão, uma análise descritiva é apresentada na Tabela 1. Foram realizados, também, os testes de correlação linear de Pearson, cujos resultados são apresentados na Tabela 2.

Tabela 1 - Medidas resumo, sendo *min* o valor mínimo, *Q1*, *Q2* e *Q3* os três quartis,  $\bar{x}$  a média, *máx* o valor máximo e  $s_x$  o desvio padrão amostral das variáveis altura dominante, área basal e volume

	min	$Q_1$	$Q_2$	$\bar{x}$	$Q_3$	máx	$s_x$
Altura dominante	14,30	17,23	18,05	18,09	18,83	24,13	1,41
Área Basal	7,20	11,89	12,94	13,07	13,98	19,24	1,77
Volume	49,28	94,56	104,10	106,10	114,10	169,30	18,85

Tabela 2 - Resultado do teste de correlação linear de Pearson, sendo apresentados: a estimativa do coeficiente de correlação linear (*r*), a estatística do teste (*t*) e o *p*-valor entre as variáveis: altura dominante, área basal e volume

	r	t	p-valor
Altura dominante × Área Basal	0,6416	16,29	$2,2 \times 10^{-16}$
Altura dominante × Volume	0,7506	22,11	$2,2 \times 10^{-16}$
Área Basal × Volume	0,9788	93,01	$2,2 \times 10^{-16}$

Em todos os casos, as variáveis apresentaram correlação positiva e significativa, sendo mais forte a correlação entre a área basal e o volume. Isso evidencia a importância

de se ajustarem modelos para área basal em função da altura dominante e modelos para o volume em função da área basal e da altura dominante.

Foram obtidos também, os semivariogramas de cada uma das variáveis juntamente com os gráficos de envelope (Figura 2). Esses gráficos permitem constatar a dependência espacial em cada uma das variáveis, pois conforme há um aumento na distância, a semivariância aumenta até se aproximar de um valor constante. Além disso, a região do envelope representa a região sem dependência espacial e como alguns pontos dos semivariogramas estão fora do envelope, fica mais evidente a existência de correlação espacial.

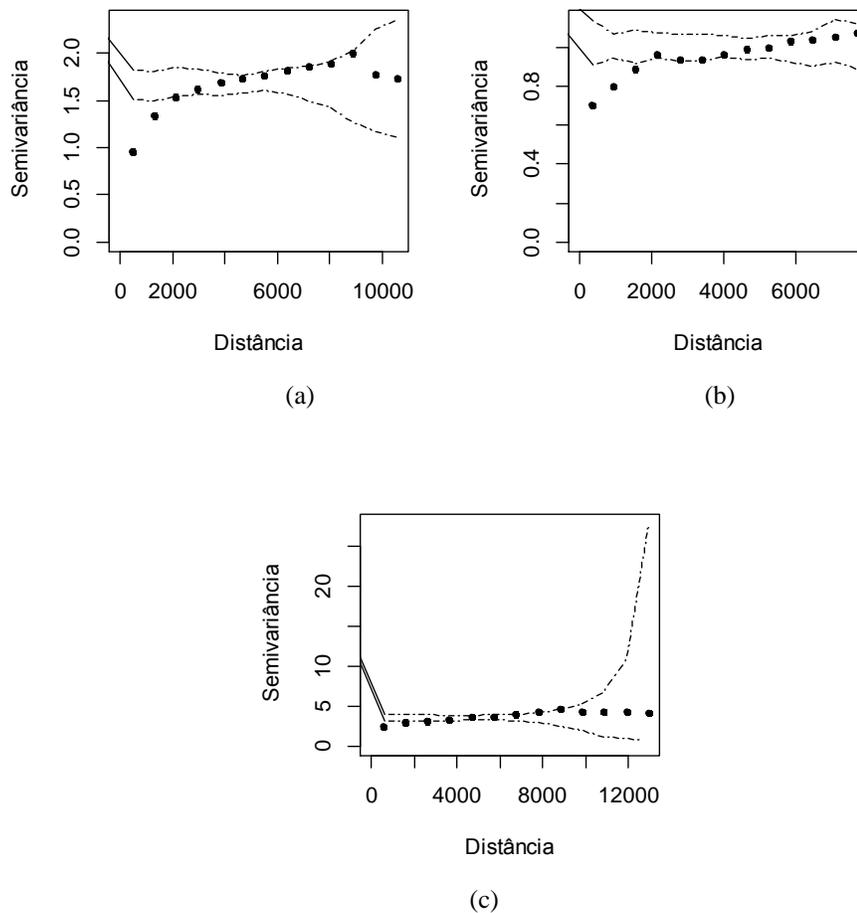


Figura 2 - Semivariogramas juntamente com os respectivos gráficos de envelope, para as variáveis: (a) altura dominante, (b) área basal e (c) volume.

### Comparação dos modelos ajustados por MQO e MQP quanto à capacidade de predição

A fim de ilustrar o ajuste do modelo com componente espacial pelos métodos de MQO e MQP, tomou-se como exemplo o modelo ajustado com função de correlação exponencial, para a variável altura. Na Figura 3 são apresentados os semivariogramas dos resíduos após eliminar o efeito das covariáveis, juntamente com a função de correlação exponencial ajustada, pelos métodos MQO e MQP. Observa-se comportamento similar no ajuste por ambos os métodos, assim como se observa que as estimativas dos parâmetros ficaram próximas. As estimativas dos parâmetros da função de correlação exponencial ajustada por MQO foram  $\widehat{\tau^2}=0,7420$ ,  $\widehat{\sigma^2}=1,1043$  e  $\widehat{\phi} = 1769,1577$  e o ajuste por MQP resultou em  $\widehat{\tau^2}=0,8486$ ,  $\widehat{\sigma^2}=1,0208$  e  $\widehat{\phi} = 2083,5144$ .

O parâmetro  $\sigma^2$  é chamado de variância do processo ou variância do sinal enquanto que o parâmetro  $\tau^2$  é a variância de micro escala ou devido a erro de medida. Quanto maior o valor de  $\sigma^2$  maior a variância devida à dependência espacial. Dessa forma, as estimativas desses parâmetros mostram que a variância devida à dependência espacial é maior que a variância do erro ou de microescala para os dados em questão.

O parâmetro  $\phi$ , está relacionado com o alcance da dependência espacial capturada pela função ajustada, em termos de distância. O alcance efetivo é aproximadamente 3 vezes o valor de  $\phi$ , sendo assim o alcance ficou estimado em 5299,9 m e 6241,6 m pelos métodos MQO e MQP respectivamente.

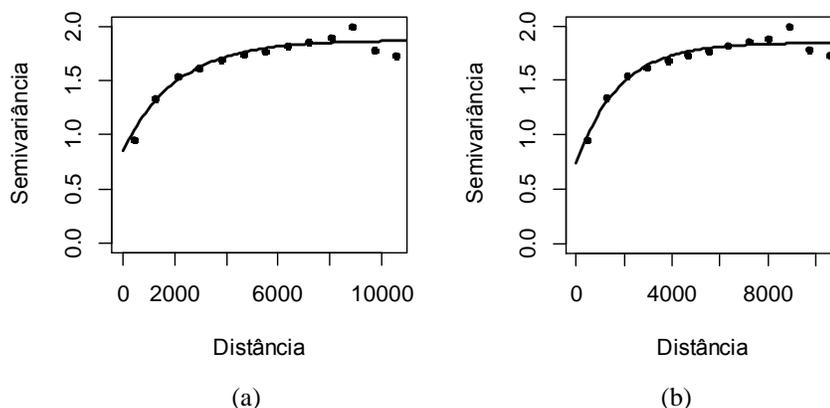


Figura 3 – Semivariograma teórico (exponencial) ajustado ao semivariograma experimental do resíduo da variável altura: (a) ajuste por MQO e (b) ajuste por MQP.

Na Tabela 3 são apresentados os valores de EQM e EAM das predições feitas para a altura dominante, segundo cada um dos modelos ajustados. São apresentados, também, os EQM's e EAM's para as predições da área basal. Observa-se que as predições da área basal foram realizadas nas mesmas parcelas em que foram feitas as predições da altura dominante e que todos os modelos ajustados para a área basal são em função da altura

dominante. Desse modo, os valores preditos da altura dominante foram utilizados para se realizar as predições da área basal. Da mesma forma são apresentados os EQM's e EAM's das predições feitas para o volume, onde foram utilizados os valores preditos da altura dominante e da área basal.

Tabela 3 - Valores de EQM e EAM das predições da altura dominante, área basal e volume, sendo os modelos ajustados por MQO e MQP

Modelo Ajustado	Função de Correlação		Altura Dominante		Área Basal		Volume		
			MQO	MQP	MQO	MQP	MQO	MQP	
Sem componente espacial	EQM		1,934	1,843	8,113	6,702	588,760	539,413	
		EAM	1,551	1,006	5,982	4,833	17,015	16,925	
Com componente espacial	Esférica	EQM	1,517	1,521	3,244	3,256	387,806	419,953	
		EAM	0,966	0,962	1,330	1,224	13,260	14,378	
	Exponencial	EQM	1,276	1,224	3,234	3,368	387,644	389,466	
		EAM	0,883	0,851	1,223	1,249	13,256	13,280	
	Gaussiana	EQM	1,238	1,244	3,295	3,604	378,944	388,535	
		EAM	0,847	0,844	1,216	1,319	13,268	13,269	
	Matern (k = 0,3)	EQM	1,252	1,224	3,226	3,256	387,644	389,466	
		EAM	0,866	0,851	1,224	1,224	13,256	13,280	
	Com componente espacial sem covariáveis	Esférica	EQM	1,523	1,524	3,412	3,412	419,792	435,632
			EAM	0,951	0,956	1,400	1,400	14,350	15,113
Exponencial		EQM	1,283	1,275	3,276	3,284	387,824	398,116	
		EAM	0,884	0,880	1,253	1,261	13,260	13,554	
Gaussiana		EQM	1,302	1,223	3,296	3,286	434,745	424,614	
		EAM	0,905	0,849	1,260	1,256	14,504	14,408	
Matern (k = 0,3)		EQM	1,276	1,275	3,273	3,282	387,691	398,116	
		EAM	0,883	0,880	1,256	1,265	13,259	13,554	

Como pode ser observado na Tabela 3, qualquer modelo que utilize componente espacial, independente da função de correlação, apresentou melhor predição, quando comparado ao modelo sem a componente espacial. Avaliando o modelo ajustado com a componente espacial, porém sem a presença das covariáveis, pode-se observar que estes modelos apresentaram maiores valores de EQM e EAM do que os mesmos modelos com a presença das covariáveis. Ou seja, ao se adicionar as covariáveis a um modelo com componente espacial, há um ganho na predição. E tal ganho é mais evidente na predição da variável volume.

Para o volume a função de correlação que forneceu as melhores predições foi a gaussiana pelo método dos mínimos quadrados ordinários. Para as variáveis “altura dominante” e “área basal” não houve uma única função de correlação que apresentou

sempre os melhores resultados. Porém, com qualquer função de correlação, o modelo com componente espacial e a presença das covariáveis, foi sempre o que forneceu as melhores predições.

### Avaliação dos modelos ajustados pelos métodos MV e MVR

Para a avaliação dos modelos quanto à capacidade de predição, quando ajustados por MV e MVR, foram calculados os EQM e EAM. E para se avaliar o ajuste, obteve-se os valores de AIC. A Tabela 4 apresenta os valores de EQM, EAM e AIC para cada um dos modelos ajustados.

Tabela 4 - Valores de AIC, EQM e EAM dos modelos para altura dominante, área basal e volume, ajustados por MV e MVR

Modelo Ajustado	Função de Correlação		Altura Dominante			Área Basal			Volume			
			EQM	EAM	AIC	EQM	EAM	AIC	EQM	EAM	AIC	
Sem component e espacial		MV	1,934	1,551	1048	4,982	2,811	1273	589,72	17,01	2786	
		MVR	1,843	1,006	1067	4,833	2,670	1270	539,41	16,92	2779	
Com component e espacial	Esférica	MV	1,356	0,943	982	3,251	1,237	860	405,294	14,428	1282	
		MVR	1,336	0,934	952	3,256	1,238	834	406,855	14,470	1235	
	Exponencia l	MV	1,330	0,913	985	3,259	1,244	865	388,668	13,693	1277	
		MVR	1,339	0,917	955	3,261	1,249	839	386,724	13,494	1229	
	Gaussiana	MV	1,341	0,934	983	3,283	1,251	860	378,053	13,311	1279	
		MVR	1,342	0,935	953	3,289	1,253	834	379,057	13,336	1231	
	Matern (k = 0,3)	MV	1,296	0,894	987	3,236	1,254	868	386,286	13,493	1278	
		MVR	1,328	0,914	955	3,261	1,249	839	386,948	13,541	1228	
	Com component e espacial sem covariáveis	Esférica	MV	1,328	0,923	1044	3,312	1,253	1171	393,564	13,544	2672
			MVR	1,328	0,924	1040	3,347	1,285	1173	403,654	13,705	2664
Exponencia l		MV	1,351	0,914	1040	3,346	1,293	1171	403,088	13,927	2670	
		MVR	1,327	0,909	1051	3,351	1,290	1168	404,428	14,005	2659	
Gaussiana		MV	1,341	0,922	1047	3,333	1,279	1173	394,898	13,484	2673	
		MVR	1,338	0,921	1043	3,344	1,281	1173	395,193	13,486	2664	
Matern (k = 0,3)		MV	1,327	0,909	1036	3,323	1,285	1172	394,328	13,629	2673	
		MVR	1,327	0,909	1038	3,351	1,290	1168	395,098	13,639	2662	

Como já observado para o ajuste com MQO e MQP, a componente espacial melhorou a predição, reduzindo os valores de EQM e EAM também no ajuste por MV e MVR (Tabela 2). Foi possível observar, também, que o uso das covariáveis melhorou a capacidade de predição em relação ao ajuste sem covariáveis, pois as predições dos modelos com componente espacial adicionados das covariáveis apresentaram menores EQM e EAM do que os respectivos modelos sem as covariáveis.

Considerando os valores de AIC, também foi possível observar enorme ganho no ajuste (redução nos valores de AIC) quando se utilizou a componente espacial com qualquer função de correlação; e que mesmo na presença da componente espacial o uso das covariáveis ainda é importante para se obter um melhor modelo ajustado para as três variáveis analisadas, pois o uso das covariáveis ajuda a reduzir ainda mais os valores de AIC.

Quanto a função de correlação, levando-se em consideração os valores de AIC, para a variáveis altura dominante e área basal, as funções de correlação gaussiana e esférica, foram as que apresentaram os melhores ajustes, não apresentando diferença significativa ( $AIC \leq 2$ ). E em termos de predição, ambas as funções apresentaram bom desempenho, não apresentando diferença relevante. Quanto a variável volume, as funções de correlação exponencial e Matérn ( $K=0,3$ ) foram as que apresentaram os melhores ajustes, com diferença de AIC inferior a 2. A capacidade preditiva não apresentou diferença relevante quando comparadas as duas funções de correlação.

#### 4 Discussão

Avaliando-se todos os métodos de ajuste, para a altura dominante o método dos MQP se mostrou mais eficaz na predição, apresentando os menores valores de EQM e EAM. No entanto, para as variáveis “área basal” e “volume”, as estimativas dos parâmetros obtidas pelo método da MVR proporcionaram os modelos mais eficazes. Em todos os casos, a componente espacial proporcionou ganho na predição, sendo superior ao modelo sem componente espacial. Além disso, o uso das covariáveis mostrou-se de grande valia na realização de predições mais precisas, gerando valores de EQM e EAM inferiores aos obtidos nos ajustes com componente espacial sem o uso das covariáveis.

A superioridade de modelos com inclusão de covariáveis e o uso da geoestatística na modelagem de variáveis dendrométricas já foi, também, observada em outros contextos, como verificado, por exemplo, Guimarães et al. (2009), que em um estudo de relação hipsométrica constataram melhoria na precisão adicionando-se covariáveis disponíveis ao modelo. Oda-Souza et al. (2008) é outro exemplo, nesse caso os autores verificaram a superioridade dos modelos com componente espacial e as covariáveis plantas mortas e tempo de ocorrência da morte na análise de experimento em delineamento sistemático com *Eucalyptus dunnii*.

O volume por hectare, como observado nos resultados, não foi influenciado pelo espaçamento e pela idade, ao contrário do encontrado por Leles et al. (2001), Muller et al. (2005) e Machado (2014). Porém, a densidade de plantio foi significativa para explicar o volume. Visto que espaçamento e densidade são variáveis diretamente relacionadas, é de se esperar que apenas uma delas seja suficiente para explicar a relação com o volume. A idade por sua vez, talvez não tenha sido significativa pela pequena variação (2,8 a 3,2 anos) apresentada nos dados disponíveis.

O uso de qualquer informação disponível que ajude a explicar e estimar o volume ou a variável de interesse é sempre interessante desde que não torne o modelo muito complexo. Wang, Upadhyay e Zang (2010), por exemplo, também utilizaram variáveis auxiliares para avaliar o volume em um plantio de *Tectonagrandis* na Índia, porém esses autores adotaram uma abordagem em que propuseram um modelo trivariado para as variáveis: diâmetro, altura e volume. Apesar dos autores levarem em conta a dependência

das três variáveis entre si, ao contrário do presente trabalho, não consideraram a possível correlação espacial para cada uma delas. Seguindo essa linha de pensamento Bognol et al. (2008) avaliou o incremento médio anual (IMA) volumétrico em um reflorestamento com *Pinus taeda* no estado de Santa Catarina. Os autores utilizaram um modelo geoestatístico univariado para IMA e também um modelo geoestatístico bivariado onde modelou conjuntamente IMA e teor de argila no solo. O uso do modelo bivariado permitiu uma predição mais detalhada do IMA, uma vez que esse modelo leva em conta não só a dependência espacial no IMA, mas também no teor de argila. À semelhança do trabalho de Bognola et al. (2008), no presente trabalho, não somente foi considerada a dependência espacial na variável volume, mas também foram modeladas a altura e a área basal com dependência espacial antes de se incluí-las na equação do volume.

Num contexto semelhante ao desenvolvido no presente trabalho, Zonete (2009) também constatou melhor desempenho ao usar modelos geoestatístico para a predição da variável volume, mesmo tendo encontrado menor alcance em seus dados do que o encontrado neste trabalho.

O fato do modelo com componente espacial ter se mostrado mais preciso, na predição da altura dominante e da área basal, pode ser atribuído à forte dependência espacial encontrada nessas variáveis. Observa-se, por exemplo, que as estimativas da variância devida à correlação espacial ( $\sigma^2$ ) são maiores do que as estimativas da variância de microescala ( $\tau^2$ ) e, por esse motivo, era esperado obter predições mais precisas com o modelo com componente espacial (PEREIRA et al., 2011; MELLO et al., 2006). Além disso, o fato do volume ser diretamente dependente da altura dominante e da área basal reforçou a superioridade do mesmo modelo, também nas predições do volume.

Outro fato interessante a se observar é a estimativa do alcance obtido para a variável “volume”, que pelo método dos mínimos quadrados ordinários foi de 8622 metros. Tal valor quando comparado a trabalhos similares é bastante elevado. Zonete (2009), por exemplo, encontrou um alcance de 1100 metros para a variável volume em um povoamento de clones de Eucaliptus em Mogi-Guaçu, SP. O alcance encontrado para a variável IMA por Bognola et al. (2008) em um reflorestamento com *Pinus taeda* no estado de Santa Catarina ficou em torno de 1428 m. Mello et al. (2005) também analisando modelos espaciais para um povoamento de *Eucaliptus grandis* localizado na região de Itapetininga – SP, constatou forte correlação espacial na variável volume. No modelo geoestatístico ajustado obteve estimativas do parâmetro  $\phi$  entre 240 e 362,1. Uma vez que o alcance efetivo é dado por aproximadamente  $3\phi$ , o alcance efetivo da dependência espacial na variável volume ficou entre 720 e 1086 m.

Por outro lado, se tratando de floresta nativa, resultados muito distintos podem ser encontrados, uma vez que se está sujeito a maior variabilidade. Sales et al., (2007), por exemplo, encontraram forte correlação espacial e um alcance espacial de cerca de 300 km para floresta nativa em Rondônia.

Provavelmente a elevada estimativa do parâmetro de alcance contribuiu para o ganho em precisão nas predições obtidas com os modelos adicionados da componente espacial no presente trabalho. Além de se um ter ganho em precisão nas predições ao se incluir uma componente aleatória para modelar a dependência espacial, a inclusão de tal componente se faz necessária como evidenciado por Zhang et al. (2004) e Zhang, Ma e Gu (2009). Esses últimos autores avaliaram efeitos de 6 modelos de regressão espacial para a predição da altura em três diferentes florestas no Canadá incluindo-se várias espécies em cada uma. Os autores mostram que a não modelagem da correlação espacial

quando presente no termo de erro dos modelos conduz à estimativas viesadas da variância e dessa forma, testes de significância para parâmetros dos modelos podem ser falhos.

Na prática o uso de modelos geoestatísticos, acurados e precisos, para predição do volume ao longo de uma área, também têm consequências econômicas, pois permite adotar manejos adaptados à variabilidade local, que potencialmente reduz custo e otimiza a produção (PELLISSARI, 2012; KANEGAE Jr. et al., 2007).

## Conclusões

O estudo realizado permitiu verificar que o uso de componente espacial e de covariáveis nos modelos para altura dominante, área basal e volume, melhorou o ajuste e a capacidade de predição dessas variáveis, sendo esta melhoria mais evidente na variável volume. Permitiu verificar, também, os melhores métodos de ajuste e selecionar funções de correlação mais adequadas para cada variável.

Dessa forma, o uso de modelos com componente espacial e covariáveis pode proporcionar a obtenção de mapas de volume mais precisos, permitindo um melhor planejamento no sistema de colheita, logística, inventário florestal e reduzindo custos.

PEREIRA, J. C.; SCALET, V.; THIRSCH, C. R. Selection and fit of spatial models to estimate the volume in a *Eucalyptus sp* plantation. *Rev. Bras. Biom.*, Lavras, v.34, n.3, p.507-521, 2016.

- **ABSTRACT:** *This paper is aimed at geostatistical models of prediction of the Eucalyptus' volume. The available characteristics were genetic material (MG), density, dominant height (Hd), and basal area (Ab). Firstly, a model was chosen at the Hd, in which the MG and density were used as covariates. Next, we fitted a model to the Ab, where the Hd was a covariate, as well as the MG and density. Finally, we selected a model to volume, where all the others variables were used as covariates. We fitted the models with OLS, WLS, ML and REML methods. The MSE, the MSA and the AIC criteria were used to evaluate the predictions and to select the models. The WLS method gave the best adjustment to the dominant height, while the REML gave the best adjustment of the models to basal area and volume. The study allows us to verify that the use of a spatial component in the model and the inclusion of covariables at Hd, Ab and volume models improved their fit. As well as the prediction capacity of those variables were improved by using the covariates and spatial component and the improvement was more evident for the volume variable.*
- **KEYWORDS:** *Geoestatistics; forest inventory; covariates.*

## Referências

- BOGNOLA, I. A.; RIBEIRO JÚNIOR, P. J.; SILVA, E. A. A.; LINGNAU, C.; HIGA, A. R. Modelagem uni e bivariada da variabilidade espacial de rendimento de *Pinus taeda* L. *Revista Floresta*, Curitiba. v.38, n.2, p.373-385, 2008.
- DIGGLE, P. J.; RIBEIRO JÚNIO, P. J. *Model-based geostatistics*. Londres: Springer, 2007. 228p.
- EMBRAPA. *Cultivo de pinus: Sistemas de produção*. 2005. Disponível em: <<http://sistemasdeproducao.cnptia.embrapa.br/FontesHTML/Pinus/CultivodoPinus/index.htm>>. Acesso em: 02/10/2010.

- GUIMARÃES, M. A. M.; CALEGÁRIO, N.; CARVALHO, L. M. T.; TRUGILHO, P. F. Height-diameter models in forestry with inclusion of covariates. *Revista Cerne*. v.15, n.3, 313-321, 2009.
- HARDIN, J. W.; HILBE, J. M. *Generalized linear models and extensions*. 2. ed., College Station: Stata Press, 2007. 552p.
- LELES, P. S. S.; REIS, G. G. REIS, M. G. F. MORAIS, E. J. Crescimento, produção e alocação de matéria seca de *Eucalyptus camaldulensis* e *E. pellita* sob diferentes espaçamentos na região de cerrado, MG. *Scientia Forestalis*, v.59, n.1, p.77-87, 2001.
- KANEGAE Jr., H.; MELLO, J. M. de; SCOLFORO, J. R. S.; OLIVEIRA, A. D. de. Avaliação da continuidade espacial de características dendrométricas em diferentes idades de povoamentos clonais de *Eucalyptus* sp. *Revista Árvore*, v.31, n.5, p.859-866, 2007.
- MACHADO, F. C. *Crescimento inicial de um clone de Eucalyptus grandis em diferentes arranjos de plantio no sistema de curta rotação*. Dissertação (Mestrado em Agronomia) - Faculdade de Ciências Agrônômicas, Universidade Estadual Paulista, Botucatu, 2014.
- MELLO, J. M. *Geoestatística Aplicada ao Inventário Florestal*. Tese (Doutorado em Ciências Florestais) - Escola Superior de Agricultura Luiz de Queiroz, Piracicaba. 2004.
- MELLO, J. M.; OLIVEIRA, M. S.; BATISTA, J. L. F.; JÚNIOR, P. R. J.; JÚNIOR, H. K. Uso do estimador geoestatístico para predição volumétrica por talhão. *Revista Floresta*, v.36, n.2, p.251-260, 2006.
- MELLO, J. M.; DINIZ, F. S.; OLIVEIRA, A. D.; SCOLFRO, J. R. S.; ACERBI JÚNIOR, F. W.; THIERSCH, C. R. Métodos de amostragem e geoestatística para estimativa do número de fustes e volume em plantios de *Eucalyptus grandis*. *Revista Floresta*, v.39, n.1, p.158-166, 2009.
- MISIR, N. Generalized height-diameter models for *Populustremula* L. stands. *African Journal of Biotechnology*, v.9, n.28, p.4348-4355, 2010.
- MONTENEGRO, S. M. G. L.; MONTENEGRO, A. A. A.; MACKAY, R. Caracterização da variabilidade espacial de parâmetros hidráulicos em solos aluviais no estado de Pernambuco. *Revista Brasileira de Recursos Hídricos*. v.4, n.2, p.3-20, 1999.
- MULLER, M. D.; COUTO, L.; LEITE, H. G.; BRITO, J. O. Avaliação de um clone de eucalipto estabelecido em diferentes densidades de plantio para produção de biomassa e energia. *Biomassa & Energia*, v.2, n.3, p.177-186, 2005.
- ODA-SOUZA, M.; BARBIN, D.; RIBEIRO JÚNIOR, P. J.; STAPE, J. L. Aplicação de métodos geoestatísticos para identificação de dependência espacial na análise de dados de um ensaio de espaçamento florestal em delineamento sistemático tipo leque. *Revista Árvore*. v.32, n.3, p.499-509, 2008.
- R CORE TEAM. *R: A language and environment for statistical computing*, 2014. Disponível em: <www.R-project.org>. Acesso em: Outubro de 2014.
- RIBEIRO JUNIOR., P. J.; DIGGLE, P. J. geoR: A package for geostatistical analysis. *R-News*, v.1, n.2, p.15-18, 2001. Disponível para download em: <www.cran.r-project.org/doc/Rnews>. Acesso em: Outubro de 2014.
- PELICO NETO, S.; BRENA, D. A. *Inventário Florestal*. Curitiba: UFPR, 1997. 316p.

SALES, M. H.; SOUZA, C. M.; PHAEDON, C.; KYRIAKIDISB, D. A.; ROBERTSB, D. A.; VIDAL, E. V. Improving spatial distribution estimation of forest biomass with geostatistics: A case study for Rondônia, Brazil. *Ecological Modelling*, v.205, p.221–230, 2007.

PELISSARI, A.L. *Silvicultura de precisão aplicada ao desenvolvimento de Tectonagrandis L.f. na Região Sul do Estado de Mato Grosso*. Dissertação (Mestrado em Ciências Florestais e Ambientais) - Universidade Federal de Mato Grosso, Cuiabá, 2012.

SCHMIDT, A. M.; SANSÓ. B. *Modelagem bayesiana da estrutura de covariância de processos espaciais e espaço-temporais*, Caxambu: ABE, 2006. 151 p.

SOARES, C. P. B.; PAULA NETO, F.; SOUZA, A. L. *Dendrometria e inventário florestal*. Viçosa: UFV, 2007. 276 p.

WANG, M.; UPADHYAY, A.; ZHANG, L. Trivariate distribution modeling of tree diameter, height, and volume. *Forest Science*, v.56, n.3, p.290-300, 2010.

ZHANG, L.; BI, H.; CHENG, P.; DAVIS, C. J. Modeling spatial variation in tree diameter–height relationships. *Forest Ecology and Management*, v.189, n.1, p.317-329, 2004.

ZHANG, L.; MA, Z.; GUO, L. An evaluation of spatial autocorrelation and heterogeneity in the residuals of six regression models. *Forest Science*, v.55, n.6, p.533-548, 2009.

ZONETE, M. C. *Avaliação do uso de técnicas de interpolação para estimativa de volume em florestas clonais de Eucalyptus sp.* Tese (Doutorado em Ciências Florestais), Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba. 2009.

Recebido em 12.03.2015

Aprovado após revisão em 27.04.2016