

# O USO DE MÉTODOS BAYESIANOS EMPÍRICOS PARA AJUSTAR RELAÇÕES HIPSOMÉTRICAS COM O MODELO DE CURTIS COM RESTRIÇÕES NOS PARÂMETROS

Monica Fabiana Bento MOREIRA<sup>1</sup>  
Selene LOIBEL<sup>1</sup>  
Cláudio Roberto THIERSCH<sup>2</sup>  
Marinho Gomes de ANDRADE<sup>1</sup>

- RESUMO: O objetivo deste trabalho é propor o uso de um método Bayesiano empírico, para ajustar relações hipsométricas (diâmetro e altura) em clones de *Eucalyptus sp*, com o modelo de Curtis com restrições biológicas impostas aos parâmetros. As densidades de probabilidades *a posteriori* são obtidas por métodos de simulação de Monte Carlo em Cadeia de Markov (MCMC). O método proposto foi aplicado para cinco conjuntos de dados reais e os resultados foram comparados aos obtidos pelo método de mínimos quadrados, a superioridade do método Bayesiano é destacada.
- PALAVRAS-CHAVE: Modelo hipsométrico; modelo de Curtis; método bayesiano empírico; simulação MCMC; modelo de regressão não-linear.

## 1 Introdução

No inventário florestal a quantificação das alturas das árvores é fundamental para classificar os povoamentos florestais de acordo com o potencial de produção de volume de madeira. Por outro lado, medir a altura de uma árvore grande exige muito tempo para sua quantificação e geralmente é feita de forma indireta por isso é sujeita a erros, estes fatores tornam a quantificação das alturas das árvores uma operação onerosa.

---

<sup>1</sup>Departamento de Matemática Aplicada e Estatística, Instituto de Ciências Matemáticas e de Computação – ICMC, Universidade de São Paulo – USP, Caixa Postal 668, CEP: 13560-970, São Carlos, SP, Brasil. E-mail: [monica@icmc.usp.br](mailto:monica@icmc.usp.br) / [sloibel@icmc.usp.br](mailto:sloibel@icmc.usp.br) / [marinho@icmc.usp.br](mailto:marinho@icmc.usp.br)

<sup>2</sup>Biometria e Inventário Florestal, Votorantim Celulose Papel, Rod. Gal Euryale Jesus Zerbine, s/n - São Silvestre - Km: 84, CEP:12300-000, Jacareí, SP, Brasil. E-mail: [claudio.thiersch@vcp.com.br](mailto:claudio.thiersch@vcp.com.br)

Para contornar estes problemas, adota-se como prática medir em uma certa área (chamada parcela) a altura e o diâmetros de uma amostra das árvores e estas medidas de altura-diâmetro, são usadas para identificar modelos matemáticos que possibilitem estimar as alturas das demais árvores contidas nas parcelas. Estes modelos matemáticos que relacionam a altura da árvore ( $H$ ) com o diâmetro do tronco, medidos a 1,3m do solo, ( $Dap$ ) são conhecidos como relações hipsométricas.

A grande dificuldade da escolha do melhor modelo para representar estas relações hipsométricas deve-se a não linearidade da relação entre as variáveis envolvidas e as restrições impostas aos parâmetros dos modelos, por razões biológicas (ver Plácido et al., 2004 e Soares et al., 2004). Estes dois fatores aliados tornam o ajuste de modelos matemáticos para representar as relações hipsométricas um problema de regressão não linear com restrições nos parâmetros. Para que as estimativas obtidas com estas equações sejam precisas os métodos e modelos adotados tornam-se cada vez mais sofisticados (ver Soares et al., 2004 e Machado; Basso; Bevilacqua, 1993). A abordagem desse problema com os métodos convencionais de inferência estatística, tais como método de máxima verossimilhança e método de mínimos quadrados torna-se ineficiente para considerar as restrições nos parâmetros de vários modelos propostos.

Neste trabalho vamos considerar um destes modelos proposto na literatura conhecido como o modelo de Curtis (Curtis, 1967). O modelo de Curtis é um modelo relativamente simples devido a possibilidade de linearização do modelo. Mas a presença de restrições nos parâmetros torna o problema de inferência complexo.

Para resolver o problema de inferência com restrições nos parâmetros para o modelo de Curtis, propomos neste trabalho uma abordagem Bayesiana (ver Box and Tiao, 1973). As restrições dos parâmetros são modeladas considerando-se uma densidade *a priori* empírica. As estimativas Bayesianas *a posteriori* são calculadas por simulação de Monte Carlo considerando o algoritmo amostrador de Gibbs (*Gibbs sampling*), Gilks; Richardson; Spiegelhalter (1998).

Os dados utilizados para testar a metodologia proposta foram coletados em florestas plantadas de clones de *Eucalyptus sp.* da Votorantim Celulose e Papel (VCP), na região de Itapetininga na região sul do estado de São Paulo.

## 2 Modelo de Curtis

O modelo de Curtis para representa a relação hipsométrica de árvores é dado por:

$$H_i = \exp \{ \beta_0 - \beta_1 Dap_i^{-1} \} \quad (1)$$

Em que  $H_i$  representa a altura total de uma árvore e  $Dap_i$  o diâmetro do tronco desta árvore a uma altura de 1,30m do solo. Por razões biológicas, os parâmetros do modelo (1) estão sujeitos as restrições do tipo  $\beta_0 > 0$  e  $\beta_1 > 0$ .

### 2.1 Estimador de mínimos quadrados para o modelo de Curtis

Considerando o conjunto de observações  $\mathbf{D} = \{(H_i, Dap_i), i = 1, \dots, n\}$ . Sob a hipótese de que a relação entre a altura  $H_i$  e o diâmetro  $Dap_i$  para cada árvore desta

amostra é dada pelo modelo de Curtis, podemos escrever para cada par  $(H_i, Dap_i)$  a relação linear:

$$y_i = \beta_0 - \beta_1 x_i + e_i \quad (2)$$

em que estamos considerando  $y_i = \log H_i$ ,  $x_i = 1/Dap_i$  e  $e_i$  é um erro de ajuste. Considerando as  $n$  observações podemos escrever a relação (2) na forma matricial como:

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{e} \quad (3)$$

com  $\mathbf{Y} = (y_1, \dots, y_n)'$ ,  $\mathbf{X} = (\mathbf{1}', -\mathbf{x}')$  sendo  $\mathbf{1} = (1, \dots, 1)'$  um vetor  $(n \times 1)$  somente de uns e  $\mathbf{x} = (x_1, \dots, x_n)'$ . Além disso temos  $\mathbf{e} = (e_1, \dots, e_n)'$  e  $\beta = (\beta_0, \beta_1)'$ . Portanto o estimador de mínimos quadrados para o vetor de parâmetros  $\beta$  irrestrito, pode ser calculado de forma direta por:

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \quad (4)$$

A variância do erro  $e$  pode ser expressa como  $\sigma^2 = E(\mathbf{e}'\mathbf{e})/n$ . Portanto um estimador não viciado para a variância  $\sigma^2$  pode ser calculado como:

$$\sigma^2 = \frac{1}{n-1}(\mathbf{Y} - \mathbf{X}\hat{\beta})'(\mathbf{Y} - \mathbf{X}\hat{\beta}) \quad (5)$$

Devido ao pequeno tamanho ( $n$ ) das amostras que geralmente são utilizadas em situações práticas, as estimativas de mínimos quadrados irrestritas para os parâmetros dadas em (4) resultam, em muitos casos, em valores inconsistentes biologicamente, ou seja estimadores negativos que não permitem prever corretamente a relação entre alturas e diâmetros das árvores (além do que modelos com parâmetros negativos são inconsistentes para predição de valores extremos tais como altura mínima e máxima). Para resolver este problema a solução é a inclusão das restrições no problema de estimação dos parâmetros, isso nos leva a considerar juntamente com a equação (3) as restrições do tipo  $\beta \in R$ , onde  $R = \{a_i < \beta_i < b_i, i = 0, \dots, p$ . Na definição destas restrições estamos considerando inclusive os casos em que alguns ou todos os  $a_i$  podem ser  $-\infty$  e os  $b_i$  podem ser  $+\infty$ . A solução numérica deste problema pode ser encontrada mas exige condições iniciais adequadas caso contrário implica em muitas dificuldades de convergência.

Para contornar essa dificuldade propomos uma abordagem Bayesiana para o cálculo dessas estimativas. Na abordagem Bayesiana as informações *a priori* sobre os parâmetros podem compensar parcialmente a pouca informação extraída somente dos dados quando estes constituem uma pequena amostra.

### 3 Abordagem bayesiana do modelo de Curtis

A abordagem Bayesiana do problema de inferência dos parâmetros de um modelo assume que estes parâmetros são variáveis aleatórias e qualquer informação inicial sobre eles pode ser modelada por uma função densidade de probabilidade *a priori*. Combinando-se estas densidades *a priori* com a função de verossimilhança dos dados, através do Teorema de Bayes, chega-se a função densidade de

probabilidade *a posteriori*. Denotando-se o vetor de parâmetros de um modelo por  $\theta$ , a densidade *a priori* por  $\pi_0(\theta)$  e a função de verossimilhança associada a um conjunto de observações  $\mathbf{D}$  por  $L(\theta|\mathbf{D})$ , então a função densidade de propabilidade *a posteriori* é dada por:

$$\pi(\theta|\mathbf{D}) = \frac{L(\theta|\mathbf{D})\pi_0(\theta)}{\int_{\Theta} L(\theta|\mathbf{D})\pi_0(\theta)d\theta} \quad (6)$$

A integral em (6) é uma integral múltipla sobre o domínio de definição dos parâmetros  $\theta \in \Theta$  que representa a constante normalizadora da densidade *a posteriori*, portanto é uma função somente dos dados  $\mathbf{D}$  observados. É comum se adotar a notação de proporcionalidade para representar a densidade *a posteriori*, dada por:

$$\pi(\theta|\mathbf{D}) \propto L(\theta|\mathbf{D})\pi_0(\theta) \quad (7)$$

Para maiores detalhes sobre inferência bayesiana ver [□andTiao\(1973\)](#).

### 3.1 A função de verossimilhança

Para escrever a função de verossimilhança do modelo (1) vamos considerar que na equação (2) o erro de ajuste  $e_i$  é uma variável aleatória independente e identicamente distribuída com distribuição normal  $N(0, \sigma^2)$ . Desta forma a função de verossimilhança para os dados  $\mathbf{D} = \{(H_i, Dap_i), i = 1, \dots, n$  pode ser escrita como:

$$L(\beta, \tau|\mathbf{D}) = \left(\frac{\tau}{2\pi}\right)^{n/2} \exp\left\{-\frac{\tau}{2}(\mathbf{Y} - \mathbf{X}\beta)'(\mathbf{Y} - \mathbf{X}\beta)\right\} \quad (8)$$

na qual estamos denotando por  $\tau = 1/\sigma^2 > 0$ . Podemos reescrever (8) considerando o estimador de mínimos quadrados  $\hat{\beta}$  dado em (4), assim temos:

$$L(\beta, \tau|\mathbf{D}) = \left(\frac{\tau}{2\pi}\right)^{n/2} \exp\left\{-\frac{\tau}{2}\left[(\beta - \hat{\beta})'\mathbf{X}'\mathbf{X}(\beta - \hat{\beta}) + S^2(\hat{\beta})\right]\right\} \quad (9)$$

onde estamos considerando  $S^2(\hat{\beta}) = (\mathbf{Y} - \mathbf{X}\hat{\beta})'(\mathbf{Y} - \mathbf{X}\hat{\beta})$ .

### 3.2 Densidade de probabilidade *a priori* empírica

Para considerar as restrições do tipo  $\beta \in R$ , onde  $R = \{a_i < \beta_i < b_i, i = 1, \dots, p$ , vamos considerar densidades *a priori* conjugadas Normal-Gama Truncada para os parâmetros assim, vamos considerar  $\beta|\tau \sim \text{NT}(\mu, (\tau\mathbf{P})^{-1})$  e  $\tau \sim \text{G}(\nu_0, \lambda_0)$ , ou seja, temos:

$$\pi_1(\beta|\tau) \propto \frac{\tau^{1/2}}{K(\mu, \mathbf{P})} \exp\left\{-\frac{\tau}{2}(\beta - \mu)'\mathbf{P}(\beta - \mu)\right\} \mathbf{I}_R(\beta) \quad (10)$$

Em (10)  $\mu = (\mu_0, \mu_1)'$  é o vetor de hiperparâmetros de locação e  $\mathbf{P}$  é a matriz de precisão, ambos conhecidos e  $K(\mu, \mathbf{P})$  é uma constante normalizadora que não precisa ser calculada quando se usa o algoritmo *Gibbs sampling*.  $\mathbf{I}_R(\beta)$  é uma função

indicadora tal que  $\mathbf{I}_R(\beta) = 1$  se  $\beta \in R$  e  $\mathbf{I}_R(\beta) = 0$  caso contrário. A densidade *a priori* para  $\tau$  é dada por:

$$\pi_2(\tau) \propto \tau^{\nu_0-1} \exp\{-\lambda_0\tau\} \quad (11)$$

Em (11)  $\nu_0$  e  $\lambda_0$  são hiperparâmetros também conhecidos. A *a priori* conjunta é dada por:

$$\pi_0(\beta, \tau) = \pi_1(\beta|\tau)\pi_2(\tau) \quad (12)$$

Neste trabalho vamos adotar uma abordagem empírica para a determinação dos hiperparâmetros  $\mu_0$ ,  $\mu_1$  e  $\mathbf{P}$ . Para isso consideramos os percentis ( $H_p, Dap_p$ ) e ( $H_q, Dap_q$ ), com  $p < q$ , encontrados em dados históricos coletados anteriormente às observações da amostra  $\mathbf{D}$ , os quais não ocorreram obrigatoriamente de forma pareada (neste trabalho consideramos nas aplicações  $p = 25\%$  e  $q = 75\%$ , mas outros percentis podem ser considerados sem perda de generalidade). Assumindo que o modelo de Curtis (Curtis, 1967) é válido para esses percentis, podemos escrever:

$$\mu_0^* = y_p + \left( \frac{y_q - y_p}{x_q - x_p} \right) x_p \quad (13)$$

$$\mu_1^* = - \left( \frac{y_q - y_p}{x_q - x_p} \right) \quad (14)$$

em que estamos considerando  $y_k = \ln(H_k)$ ,  $x_k = \ln(Dap_k)$  para  $k = p, q$  e  $\mu = (\mu_0^*, \mu_1^*)'$ . A matriz de precisão  $\mathbf{P}$  também pode ser encontrada de forma empírica considerando um coeficiente de variação constante  $CV_j = \gamma_j/\mu_j^*$ ,  $j = 0, 1$  que expresse a confiança nos hiperparâmetros  $\mu$ . Desta forma temos:

$$P_{j,j} = \frac{1}{\gamma_j^2} = \frac{1}{(CV_j \mu_j^*)^2}, j = 0, 1 \quad (15)$$

Vamos assumir também que  $\beta_0$  e  $\beta_1$  são independentes a priori, de forma que  $P_{i,j} = 0$  para  $i \neq j$ , ( $i, j = 0, 1$ ).

### 3.3 A densidade de probabilidade *a posteriori*

A densidade de probabilidade *a posteriori* para  $\beta$  e  $\tau$  é dada por:

$$\pi(\beta, \tau | D) \propto L(\beta, \tau) \pi_0(\beta, \tau) \quad (16)$$

substituindo-se (9), (10) e (11) em (16), temos:

$$\pi(\beta, \tau | \mathbf{D}) \propto \tau^{n/2+\nu_0} \exp \left\{ -\frac{\tau}{2} \left[ (\beta - \tilde{\beta})' \mathbf{V} (\beta - \tilde{\beta}) + \mathbf{W} + 2\lambda_0 \right] \right\} \mathbf{I}_R(\beta) \quad (17)$$

onde  $\mathbf{W} = \mathbf{Y}'\mathbf{Y} + \tilde{\beta}'\mathbf{V}\tilde{\beta}$ ,  $\mathbf{V} = (\mathbf{X}'\mathbf{X} + \mathbf{P})$  e  $\tilde{\beta}$  é o estimador Bayesiano irrestrito, dado por

$$\tilde{\beta} = (\mathbf{X}'\mathbf{X} + \mathbf{P})^{-1}(\mathbf{X}'\mathbf{X}\hat{\beta} + \mathbf{P}\mu) \quad (18)$$

sendo que  $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$  é o estimador de mínimos quadrado.

Para encontrar os estimadores Bayesianos a partir da densidade *a posteriori* conjunta (17) vamos adotar o algoritmo amostrador de Gibbs. Para isso vamos gerar amostras das densidades condicionais *posteriori*,  $\pi(\beta|\tau, \mathbf{D})$  e  $\pi(\tau|\beta, \mathbf{D})$ .

Da equação (17) encontramos que  $\beta$  tem uma densidade condicional *a posteriori* Normal-Truncada,  $\beta|\tau \sim \text{NT}(\tilde{\beta}, (\tau\mathbf{V})^{-1})\mathbf{I}_R(\beta)$ . Portanto para considerar restrições nos parâmetros *a posteriori*, podemos escrever:

$$\pi(\beta|\tau, D) \propto \frac{\tau^{n/2+\nu_0}}{K(\tilde{\beta}, \mathbf{V})} \exp\left\{-\frac{\tau}{2} [(\beta-\tilde{\beta})'\mathbf{V}(\beta-\tilde{\beta})]\right\} \mathbf{I}_R(\beta) \quad (19)$$

onde  $K(\tilde{\beta}, \mathbf{V})$  é a constante normalizadora da função densidade de probabilidade truncada que não depende dos parâmetros do modelo, portanto não precisa ser avaliada quando se usa o algoritmo *Gibbs sampling*

Para gerar amostras de  $\beta$ , vamos considerar uma variável aleatória  $\mathbf{z} = (z_1, z_2)'$  com distribuição normal,  $\mathbf{z}|\tau \sim \text{N}(\tilde{\beta}, (\tau\mathbf{V})^{-1})$ . Podemos gerar amostras de  $\mathbf{z}|\tau$ , considerando uma variável aleatória uniforme  $u \sim U(0, 1)$  e usando a relação dada por:

$$z_0 = \tilde{\mu}_0 + \varpi_0 \Phi^{-1}(u_0) \quad (20)$$

$$z_1 = \tilde{\mu}_1 + \varpi_1 \Phi^{-1}(u_1) \quad (21)$$

onde  $\varpi_0 = \sqrt{\sigma_{11}}$  é a raiz quadrada do primeiro elemento da diagonal de  $(\tau\mathbf{V})^{-1}$  e  $\tilde{\mu}_0 = \tilde{\beta}_0$  e  $z_1$  é gerada considerando-se a distribuição de  $z_1$  condicionada a  $z_0$  cuja a média e desvio-padrão são dados por:

$$\tilde{\mu}_1 = \tilde{\beta}_1 + \sigma_{12}\sigma_{11}^{-1}(z_0 - \tilde{\beta}_0) \quad (22)$$

$$\varpi_1 = (\sigma_{22} - \sigma_{12}\sigma_{11}^{-1}\sigma_{21})^{1/2} \quad (23)$$

Então, podemos estabelecer a relação entre  $\mathbf{z}|\tau \sim \text{N}(\tilde{\beta}, (\tau\mathbf{V})^{-1})$  e  $\beta|\tau \sim \text{NT}(\tilde{\beta}, (\tau\mathbf{V})^{-1})\mathbf{I}_R(\beta)$  considerando a relação encontrada em Albert and Chib (1996), dada por:

$$u_i = \frac{\Phi\left(\frac{\beta_i - \tilde{\mu}_i}{\varpi_i}\right) - \Phi\left(\frac{a_i - \tilde{\mu}_i}{\varpi_i}\right)}{\Phi\left(\frac{b_i - \tilde{\mu}_i}{\varpi_i}\right) - \Phi\left(\frac{a_i - \tilde{\mu}_i}{\varpi_i}\right)}, \quad i = 0, 1 \quad (24)$$

assim temos:

$$\beta_0 = \tilde{\mu}_0 + \varpi_0 \Phi^{-1} \left[ \Phi\left(\frac{a_0 - \tilde{\mu}_0}{\varpi_0}\right) + u_0 \left( \Phi\left(\frac{b_0 - \tilde{\mu}_0}{\varpi_0}\right) - \Phi\left(\frac{a_0 - \tilde{\mu}_0}{\varpi_0}\right) \right) \right] \quad (25)$$

$$\beta_1 = \tilde{\mu}_1 + \varpi_1 \Phi^{-1} \left[ \Phi\left(\frac{a_1 - \tilde{\mu}_1}{\varpi_1}\right) + u_1 \left( \Phi\left(\frac{b_1 - \tilde{\mu}_1}{\varpi_1}\right) - \Phi\left(\frac{a_1 - \tilde{\mu}_1}{\varpi_1}\right) \right) \right] \quad (26)$$

As equações (25)-(26) indicam como podemos gerar valores dos parâmetros restritos  $\beta = (\beta_0, \beta_1)' \in R$  da distribuição Normal truncada, a partir da geração de variáveis aleatórias uniforme  $U(0, 1)$ .

Portanto, a equação (24) juntamente com as equações (20)-(21) podem ser usadas para gerar valores dos parâmetros  $\beta = (\beta_0, \beta_1)'$  a partir dos valores gerados da distribuição normal irrestrita  $\mathbf{z} = (z_0, z_1)'$ . Existe uma relação determinística entre as variáveis aleatórias  $\beta$  e  $\mathbf{z}$  de forma que  $\pi(\beta|\mathbf{z}, \tau, \mathbf{D}) = 1$

Estamos agora em posição de elaborar um algoritmo *Gibbs Sampling* para encontrar o estimador Bayesiano de  $\beta$  e  $\tau$ . Escrevendo a densidade *a posteriori* conjunta  $\pi(\beta, \mathbf{z}, \tau|\mathbf{D})$ , como:

$$\pi(\beta, \mathbf{z}, \tau|\mathbf{D}) = \pi(\beta|\mathbf{z}, \tau, \mathbf{D})\pi(\mathbf{z}|\tau, \mathbf{D})\pi(\tau|\mathbf{D})$$

O algoritmo *Gibbs Sampling* para gerar amostras da densidade  $\pi(\beta, \mathbf{z}, \tau|\mathbf{D})$ , é obtido com as densidades condicionais  $\pi(\mathbf{z}|\tau, \mathbf{D})$ ,  $\pi(\beta|\mathbf{z}, \tau, \mathbf{D})$  e a densidades  $\pi(\tau|\beta, \mathbf{D})$  que pode ser obtida de (17).

Considerando-se a relação determinística entre  $\beta$  e  $\mathbf{z}$  dada pela equação (24) combinada com as equações (20) e (21) então a densidade condicional  $\pi(\beta|\mathbf{z}, \tau, \mathbf{D})$ , é dada por:

$$\pi(\beta|\mathbf{z}, \tau, \mathbf{D}) = \begin{cases} 1 & \text{se (24) e (20)-(21) vale.} \\ 0 & \text{caso contrário} \end{cases} \quad (27)$$

Uma vez que dado  $\mathbf{z}$  temos  $\beta$ , então podemos considerar simplesmente a densidade condicional  $\pi(\tau|\beta, \mathbf{D})$  para gerar amostras de  $\tau$  no algoritmo *Gibbs Sampling* e a densidade  $\pi(\tau|\beta, \mathbf{D})$ , é dada por:

$$\pi(\tau|\beta, \mathbf{D}) \propto \tau^{n/2+\nu_0} \exp \left\{ -\frac{\tau}{2} \left[ (\beta - \tilde{\beta})' \mathbf{V} (\beta - \tilde{\beta}) + \mathbf{W} + 2\lambda_0 \right] \right\} \quad (28)$$

No algoritmo *Gibbs Sampling* vamos adotar a seguinte notação:

$$\beta|\tau, \mathbf{D} \sim \text{NT} \left( \tilde{\beta}, (\tau \mathbf{V})^{-1} \right) \quad (29)$$

$$\tau|\beta, \mathbf{D} \sim \text{G}(n/2 + \nu_0 + 1, \mathbf{\Lambda}(\beta) + 2\lambda_0) \quad (30)$$

onde  $\mathbf{\Lambda}(\beta) = (\beta - \tilde{\beta})' \mathbf{V} (\beta - \tilde{\beta}) + \mathbf{W}$ .

O algoritmo amostrador de Gibbs utilizado para gerar amostras da densidade *a posteriori* conjunta  $\pi(\beta, \tau|\mathbf{D})$  é dado por:

- Algoritmo *Gibbs Sampling*:

1. Faça  $j = 0$  e considere a condição inicial  $\tau^{(0)}$ .
2. Gerar  $\mathbf{z}^{(j+1)}|\tau^{(j)}, \mathbf{D} \sim \text{N} \left( \tilde{\beta}, (\tau^{(j+1)} \mathbf{V})^{-1} \right)$  e calcular  $\mathbf{u} = (u_0, u_1)$  de (24)
3. Gerar os valores de  $\beta^{(j+1)}$  usando as equações (25) e (26).
4. Gerar  $\tau^{(j+1)}$  de  $\tau^{(j+1)}|\beta^{(j)}, \mathbf{D} \sim \text{G}(n/2 + \nu_0 + 1, \mathbf{\Lambda}(\beta^{(j)}) + 2\lambda_0)$
5. Faça  $j \leftarrow j + 1$  e repita os passos (2), (3) e (4).

Para encontrar os estimadores bayesianos obtidos via simulação de Monte Carlo adotamos o procedimento sugerido em Gilks; Richardson e Spiegelhalter (1998). Consideramos um período de aquecimento (burning) de 50% das iterações iniciais. Os passos (2), (3) e (4) do algoritmo *Gibbs Sampling* são repetidos até que a convergência monitorada graficamente e com o critério de Geweke (Geweke, 1992) seja verificada. Então uma amostra é selecionada a cada 10 iterações. Denotando por  $\theta^{(j)} = \{\beta_0^{(j)}, \beta_1^{(j)}, \beta_{0,R}^{(j)}, \beta_{1,R}^{(j)}, \tau^{(j)}\}$ ,  $j = 1, \dots, M$  a amostra resultante do algoritmo *Gibbs Sampling*, os estimadores bayesianos obtidos via simulação de Monte Carlo são dados por:

$$\hat{g}(\theta_k^{MC}) = \frac{1}{M} \sum_{j=1}^M g(\theta_k^{(j)})$$

E os intervalos de credibilidade  $[\hat{\theta}_k^{MC}(\alpha), \hat{\theta}_k^{MC}(1-\alpha)]$  são estimados com base nos percentis da amostra, dados por:

$$P[\hat{\theta}_k^{MC}(\alpha) \leq \theta_k^{MC} \leq \hat{\theta}_k^{MC}(1-\alpha)] = 1 - \alpha$$

## 4 Aplicações

### 4.1 Localização e caracterização da área

Os dados utilizados para testar a metodologia proposta foram coletados em florestas plantadas de clones de *Eucalyptus sp.* da Votorantim Celulose e Papel (VCP), na região de Itapetininga na região sul do estado de São Paulo, com coordenadas geográficas dadas por latitude  $23^{\circ}56'S$ , longitude  $42^{\circ}21'W$  e altitude média de  $655m$ . Além disso, de acordo com Golfari; Caser e Moura (1978), nesta região a precipitação média anual é de  $1600mm$  e temperatura média anual de  $20^{\circ}C$ . Os solos são predominantemente Latossolo Vermelho distrófico, descrevendo uma situação favorável a cultura de *Eucalyptus sp.*

### 4.2 Amostragem

Foi adotada a técnica de amostragem sistemática com intensidade amostral de uma parcela a cada  $10ha$ . As parcelas são mensuradas anualmente a partir do segundo ano dos povoamentos. As parcelas têm forma circular com área de  $500m^2$ . Dentro de cada parcela são medidos os diâmetros (*Dap*) de todas as árvores e a altura de 10 árvores centrais e das 5 árvores mais altas (chamadas árvores com alturas dominantes). Este procedimento resulta em amostras com, no máximo, 15 medidas (*H* e *Dap*).

Os dados usados neste trabalho são referentes a cinco destas parcelas (que denominaremos pelas letras A, B, C, D e E) que estão localizadas em povoamentos com idades variando de 2 a 8 anos e diferentes espaçamento entre as árvores. A parcela A foi mensurada em povoamento de 2,33 anos de idade e espaçamento de  $3,0 \times 3,0m$ , a parcela B em povoamento de 3,46 anos e espaçamento de  $4,0 \times 2,2m$ ,



Tabela 1 - Parcelas para árvores estratificadas por idade  $Dap$  (cm) - diâmetro a 1,30m do solo e  $H$  (m) - altura total

Parc. A		Parc. B		Parc. C		Parc. D		Parc. E	
$Dap$	$H$	$Dap$	$H$	$Dap$	$H$	$Dap$	$H$	$Dap$	$H$
10,92	12,8	12,96	17,5	16,87	25,5	22,73	30,8	16,46	25,5
10,76	12,0	13,27	17,3	16,07	27,5	20,31	31,0	16,49	25,9
11,59	13,5	12,16	17,0	17,25	27,0	24,35	39,5	16,62	26,1
10,79	13,5	13,46	17,0	16,58	27,3	23,40	31,0	17,09	26,5
10,06	13,3	13,11	17,0	17,95	26,8	23,78	32,0	17,22	25,5
11,52	13,0	12,38	17,5	17,41	26,5	23,40	31,0	16,81	25,8
10,73	13,0	14,10	18,0	16,11	26,3	20,53	30,5	17,19	26,0
10,57	13,5	12,99	18,0	15,37	26,5	23,68	30,3	16,87	25,8
10,73	14,0	12,99	17,3	16,77	27,0	22,63	30,5	16,84	25,5
12,16	13,3	13,11	17,0	15,50	27,3	21,14	31,5	17,63	26,3
11,59	13,0	14,39	17,8	18,78	26,5	25,34	30,5	17,79	25,6
11,55	13,0	14,26	18,0	17,67	26,8	24,00	30,0	18,18	25,6
11,68	12,8	14,25	17,0	17,60	27,5	25,43	31,5	18,11	25,7
—	—	14,61	17,3	18,18	26,0	24,41	32,0	17,83	25,2
—	—	—	—	—	—	26,32	29,5	—	—

a parcela  $C$  em povoamento de 4,94 anos e espaçamento de  $5,0 \times 1,5m$ , a parcela  $D$  em povoamento de 6,12 anos e espaçamento de  $3,0 \times 3,0m$  e a parcela  $E$  em povoamento de 7,64 anos com espaçamento de  $5,0 \times 1,8m$ . Os dados das amostras utilizados nos ajustes, por parcelas são apresentados na Tabela 1.

#### 4.3 Densidades *a priori* empíricas

Os parâmetros das densidades *a priori* empíricas consideradas para cada parcela são dados na Tabela 2. Estes valores foram calculados usando as equações (13)-(15) com dados de parcelas já analisadas. Consideramos nas equações (13) e (14) os percentis de  $p = 25\%$  e  $q = 75\%$ , e na equação (15) fizemos os coeficientes de variação  $CV_0 = 1$  e  $CV_1 = 10$ . Estes valores são arbitrários, porém a escolha de coeficientes de variações grandes é sempre recomendável para garantir que as densidades *a priori* sejam curvas aplainadas (flat curves).

#### 4.4 Análise *a posteriori*

As estimativas bayesianas foram obtidas via simulação de Monte Carlo com o algoritmo *Gibbs Sampling* gerou uma amostra de tamanho 20.000 da qual descartamos as primeiras 10.000 (*burning*). Selecionamos da segunda parte um valor a cada 10 valores gerados (ou seja, 10.001, 10.010, 10020, ..., 20.000) perfazendo

Tabela 2 - Parâmetros das densidades *a priori* empírica das parcelas

Parâmetros da <i>Priori</i>	Parcelas				
	A	B	C	D	E
$\mu_0$	3.1217	2.9786	3.6114	3.8143	3.5899
$\mu_1$	6.0147	1.6334	5.4272	8.9840	5.9081
$P_{1,1}$	0.1026	0.1127	0.0766	0.0687	0.0776
$P_{2,2}$	0.0276	0.3748	0.0339	0.0124	0.0287

assim uma amostra de tamanho 1000 que foi utilizada para o cálculo das estimativas bayesianas via MCMC.

As amostras geradas de cada parâmetro são apresentadas na Figura 1 e as estimativas do critério de Geweke (1992), utilizado para verificar a convergência da cadeia gerada são  $Z(\beta_0) = 0.0147$ ,  $Z(\beta_1) = 0.1022$  e  $Z(\sigma^2) = 0.030843$ , como critério de Geweke a convergência ocorre quando  $|Z| < Z_c$  para  $Z_c = 1.96$ .

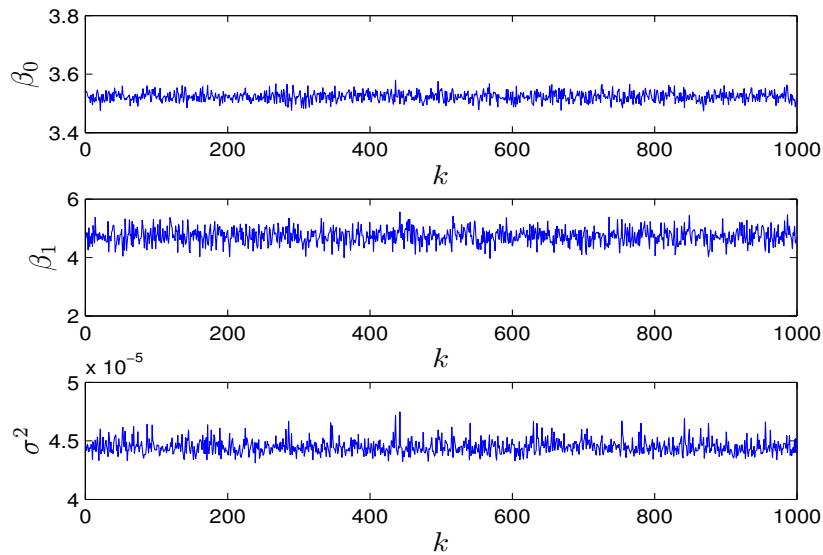


Figura 1 - Trajetória das amostras geradas como o algoritmo de simulação MCMC após ser verificada a convergência.

Um sumário das estimativas bayesianas e de mínimos quadrados são apresentados na Tabela 3. Comparando-se as estimativas bayesianas às de mínimos quadrados sem restrições apresentadas na Tabela 3, notamos que várias estimativas

Tabela 3 - Estimativas Bayesianas e de Mínimos Quadrados

Parcela	Estimativa Bayesiana			
<b>A</b>	<b>Média</b>	<b>D.P.</b>	<b>IC (95%)</b>	<b>M. Q.</b>
$\beta_0$	3.1111	0.0106	[3.0905; 3.1298]	2.5079
$\beta_1$	5.9474	0.1099	[5.7350; 6.1596]	-0.7369
$\sigma^2$	3.52e-4	8.91e - 6	[3.34e-4; 3.71e-4]	1.37e-3
<b>B</b>	<b>Média</b>	<b>D.P.</b>	<b>IC (95%)</b>	<b>M. Q.</b>
$\beta_0$	2.9789	0.0030	[2.9728; 2.9842]	2.9786
$\beta_1$	1.6328	0.0177	[1.5984; 1.6673]	1.6334
$\sigma^2$	1.20e-4	1.79e - 6	[1.17e-4; 1.24e-4]	4.77e-4
<b>C</b>	<b>Média</b>	<b>D.P.</b>	<b>IC (95%)</b>	<b>M. Q.</b>
$\beta_0$	3.6045	0.0047	[3.5949; 3.6137]	3.2091
$\beta_1$	5.3923	0.0619	[5.2671; 5.5098]	-1.3087
$\sigma^2$	1.15e-4	1.69e - 6	[1.11e-4; 1.18e-4]	4.54e-4
<b>D</b>	<b>Média</b>	<b>D.P.</b>	<b>IC (95%)</b>	<b>M. Q.</b>
$\beta_0$	3.8090	0.0049	[3.7999; 3.8177]	3.3764
$\beta_1$	8.8602	0.0953	[8.6812; 9.0559]	-1.2162
$\sigma^2$	1.28e-4	1.95e - 6	[1.25e-4; 1.32e-4]	5.00e-4
<b>E</b>	<b>Média</b>	<b>D.P.</b>	<b>IC (95%)</b>	<b>M. Q.</b>
$\beta_0$	3.5245	0.0114	[3.4933; 3.5541]	3.1679
$\beta_1$	4.7258	0.1522	[4.2018; 5.2697]	-1.4071
$\sigma^2$	4.45e-5	6.15e - 7	[4.35e-5; 4.59e-5]	1.76e-4

D.P. é o desvio padrão; IC é o intervalo de credibilidade e M.Q. é a estimativa de mínimos quadrados.

de mínimos quadrados apresentam valores negativos, violando portanto as restrições biológicas impostas aos parâmetros. Por outro lado as estimativas Bayesianas, dadas pela a média das amostrais geradas da densidade *a posteriori* via MCMC, apresentam sempre valores positivos. Os intervalos de credibilidade bayesianos *IC*(95%) dados na Tabela 3 revelam a precisão das estimativas bayesianas. Notamos na Tabela 3 que, quando os estimadores de Mínimos Quadrados (M.Q.) resultam em valores positivos, estas estimativas coincidem com as estimativas Bayesianas, como pode ser visto na Tabela 3, com os resultados da parcela B.

As densidades marginais *a posteriori* obtidas via MCMC juntamente com as densidades *a priori* empíricas são apresentadas na Figura 2, para cada um dos parâmetros do modelo de Curtis ajustado aos dados da amostra da parcela E. Os resultados para as outras parcelas mostram resultados semelhantes.

Analisando as densidades apresentadas na Figura 2, observamos que a combinação das informações fornecidas *a priori* com as informações extraídas dos dados através da função de verossimilhança, localiza o domínio de definição das densidades *a posteriori* dentro do domínio das densidades *a priori* empíricas,

definido pelas restrições impostas de forma biologicamente aceitável, além disso, a análise *a posteriori* aumenta a precisão das estimativas dos parâmetros, isso pode ser constatado nos gráficos da Figura 2, observando-se densidades *a posteriori* mais fechadas em relação as densidades *a priori*. Estes dois aspectos (localização e forma das densidades *a posteriori*) destacam a validade da proposta empírica de construção das densidades *a priori* e destacam a vantagem da abordagem bayesiana para tratar o problema de inferência dos parâmetros, das relações hipsométricas com restrições nos parâmetros.

As curvas estimadas usando os estimadores bayesianos e de mínimos quadrados para cada uma das amostras dadas na Tabela 3, são apresentadas na Figuras 3.

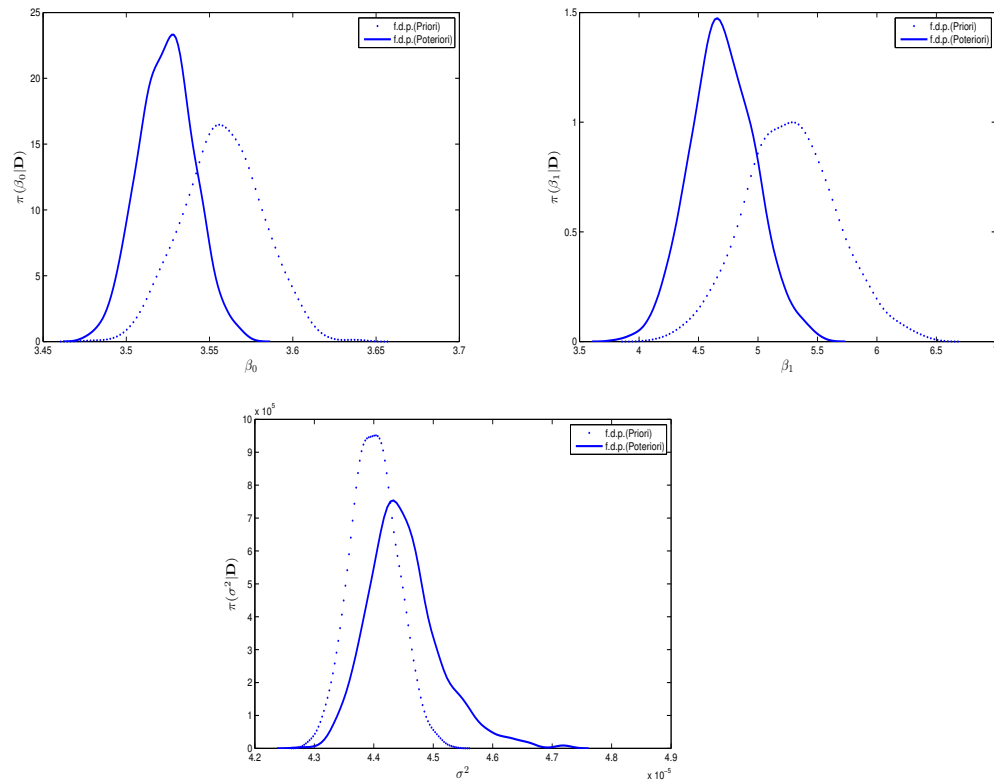


Figura 2 - Densidades *a priori* e *posteriori* para os parâmetros  $\beta_0$ ,  $\beta_1$  e  $\sigma^2$  do modelo de Curtis ajustado para a Parcela E.

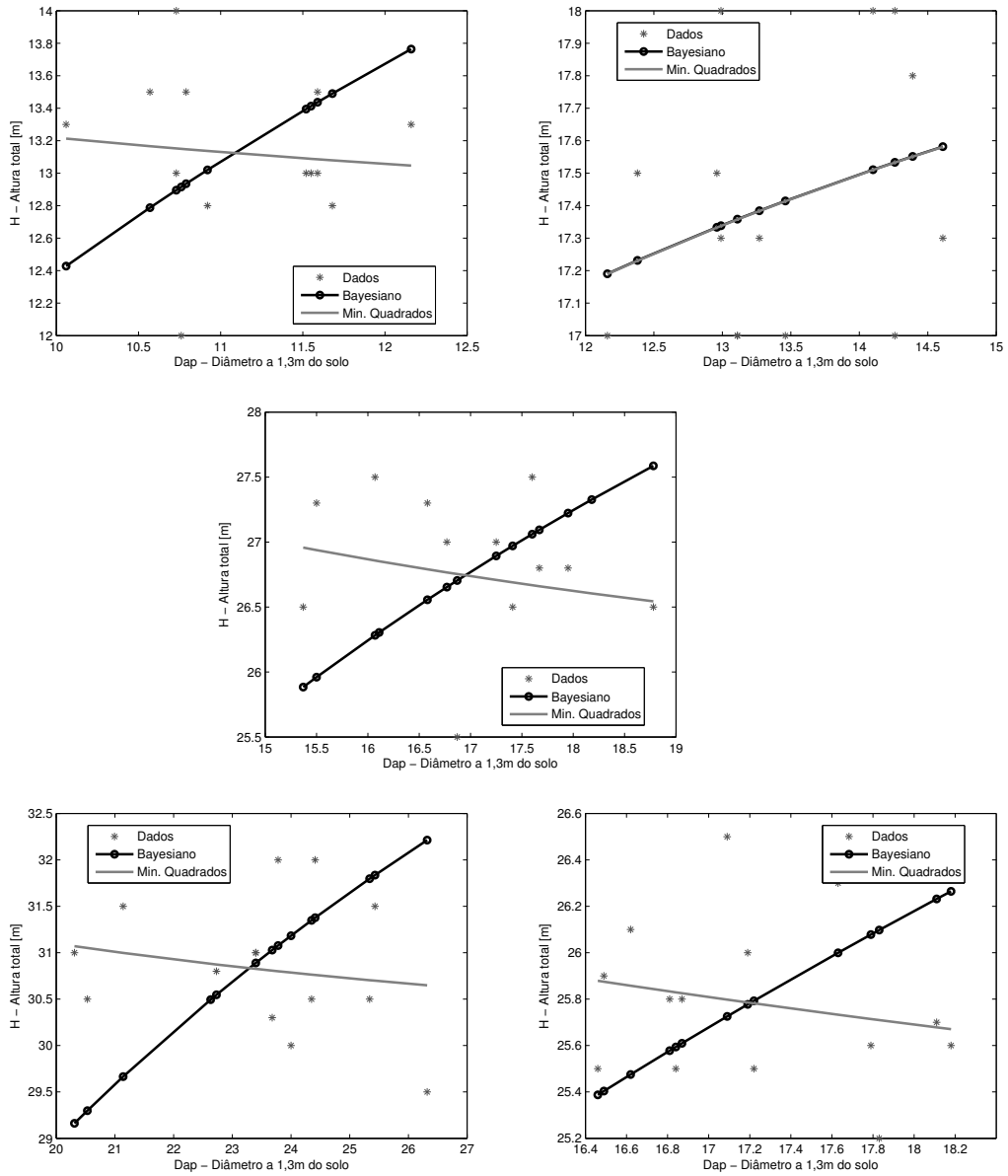


Figura 3 - Ajuste do modelo de Curtis para as parcelas A, B, C, D e E.

A análise destas curvas revela que, em conformidade com as estimativas apresentadas na Tabela 3, as estimativas de mínimos quadrados (MQ) resultam em uma relação biologicamente inconsistente para representar a relação hipsométrica para as amostras das parcelas A, C, D e E. Para a amostra da parcela B ambas abordagens coincidem, mostrando que há consistência da abordagem bayesiana proposta com a de mínimos quadrados, quando esta última é factível.

## Conclusões

Neste trabalho tratamos o problema de inferência para o modelo de Curtis com restrições nos parâmetros, considerando uma modelagem Bayesiana. As restrições dos parâmetros foram modeladas considerando-se uma densidade *a priori* empírica Normal-Truncada. As estimativas Bayesianas *a posteriori* foram calculadas com um algoritmo de simulação de Monte Carlo em Cadeia de Markov (MCMC).

Os dados utilizados para testar a metodologia proposta foram coletados em florestas plantadas de clones de *Eucalyptus sp.* da Votorantim Celulose e Papel (VCP), na região de Itapetininga na região sul do estado de São Paulo, com coordenadas geográficas dadas por latitude  $23^{\circ}56'S$ , longitude  $42^{\circ}21'W$ , e altitude média de  $655m$ .

A abordagem Bayesiana empírica proposta em conjunto com a técnica de solução via MCMC mostram-se vantajosa na consideração das restrições dos parâmetros do modelo. As estimativas obtidas pelo método de mínimos quadrados (MQ) resultam em uma relação biologicamente inconsistente para representar a relação hipsométrica em quatro dos cinco conjuntos de dados analisados, enquanto que as estimativas Bayesianas respeitaram os limites impostos aos parâmetros para todos os conjuntos de dados analisados.

A modelagem Bayesiana proposta é portanto recomendável como uma técnica alternativa e de fácil implementação para a estimação dos parâmetros do modelo de Curtis com restrições. O esforço computacional dos algoritmos MCMC depende principalmente do número de iterações necessárias até que a convergência seja verificada. Neste sentido o uso de funções densidades de probabilidade *a priori* empírica contribui para que essa convergência seja atingida mais rapidamente. Este é um fator importante para viabilizar a implementação da abordagem Bayesiana proposta para tratar um grande número de parcelas, como geralmente ocorre na prática.

MOREIRA, M. F. B; LOIBEL, S.; THIERSCH, C. R.; ANDRADE, M. G. The use of empirical Bayesian methods to fit hypsometric relationship with the Curtis model with constraint in the parameter. *Rev. Bras. Biom.*, São Paulo, v.25, n.3, p.23-38, 2007.

- **ABSTRACT:** *The objective of this paper is to propose the use of an empirical Bayesian methods for fitting hypsometric relationships (diameter and height) for clones of Eucalyptus sp., with Curtis model with biological constraint imposed to the parameters. The posterior probability density is obtained via simulation technique with Monte Carlo Markov Chain (MCMC) methods. The proposed approach was applied for five real data sets and the results were compared to the obtained by the minimum square method, the superiority of the Bayesian method is highlighted.*
- **KEYWORDS:** *Hypsometric model, Curtis model; empirical Bayesian methods; simulation MCMC; non linear regression model.*

## Referências

- ALBERT, J. H.; CHIB, S. Computation in bayesian econometrics: an introduction to Markov chain Monte Carlo. In: Hill, R.C. (Ed.). *Advances in econometrics*. Greenwich: JAI Press, 1996. p.3-24. (Computational Methods and Applications, Volume 11A ).
- BOX, G. E.; TIAO, G. C. *Bayesian inference in statistical Analysis*. New York: Addison-Wesley, 1973. 588p.
- COUTO, H.T. Z.; BASTOS, N. L. M. Modelos de equações de volume e relações hipsométricas para plantações de eucalyptus no estado de São Paulo. IPEF, n.37, p.33-44, 1987.
- CURTIS, R. O. Height diameter and height diameter age equations for second growth Douglas-fir. *For. Sci.*, v.13, n.4, p.365-375, 1967.
- GEWEKE, J. Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In: BERNARDO, J. M., BERGER, J. O., DAWID, A. P. and SMITH, F. M. (Ed.). *Bayesian statistics*. Oxford: Clarendon Press, 1992. p.1-31.
- GILKS, W. R.; RICHARDSON, S.; SPIEGELHALTER, D. J. *Markov chain Monte Carlo in practice*. Chapman & Hall, 1998. p.486.
- GOLFARI, L.; CASER, R. L.; MOURA, V. P. G. *Zoneamento ecológico esquemático para reflorestamento no Brasil: 2ª aproximação*. Centro de pesquisas florestais da região do cerrado, 1978. 66p.
- MACHADO, S. A.; BASSO, S. F.; BEVILACQUA Jr, V. G. Teste de modelos matemáticos para o ajuste da relação hipsométrica em diferentes sítios e idades para plantações de *Pinus elliottii* no Estado do Paraná. In: CONGRESSO FLORESTAL PANAMERICANO, 1.; CONGRESSO FLORESTAL BRASILEIRO, 7., 1993, Curitiba. *Anais...* Curitiba: SBS, 1993. v.2, p. 553-556.
- PLÁCIDO, A. C. et al. Dinâmica da relação hipsométrica em função da idade, do sítio e da densidade inicial de povoamentos de Bracatinga da região metropolitana de Curitiba; PR. *Rev. Árvore*, Viçosa, v.28, n.4, p.517-533, 2004.

SOARES T. S. et al. Uso de diferentes alternativas para viabilizar a relação hipsométrica no povoamento florestal. *Rev. Árvore*, Viçosa-MG, v.28, n.6, p.845-854, 2004.

Recebido em 05.09.2007.

Aprovado após revisão em 12.11.2007.