

# Ajuste de Relações Hipsométricas com heterocedasticidade usando GAMLSS

Márcia Ap. Centanin Macera <sup>1</sup>, ICMC-USP, São Carlos-SP, marciam@icmc.usp.br

Marinho Gomes de Andrade Filho, ICMC-USP, São Carlos-SP, marinho@icmc.usp.br

**Resumo:** Este trabalho propõe o ajuste dos modelos de Stofel, Curtis, Petterson e Prodan para representar a relação hipsométrica em um conjunto de dados (altura, diâmetro) de *Eucalyptus sp.*, considerando heterocedasticidade e utilizando os modelos generalizados aditivos para localização, escala e forma (GAMLSS) para o ajuste das relações propostas. Supõe-se para a variável resposta (altura) distribuições da família simétrica tais como normal, logística do tipo *II*, normal generalizada, t de student e família normal. Os parâmetros de localização e escala são modelados como funções da variável explicativa (diâmetro), em que para o parâmetro de localização são utilizados os modelos de hipsometria citados acima. Utiliza-se para as estimativas dos parâmetros o pacote NLGAMLSS, com o qual é possível trabalhar com modelos não-lineares. Por fim, os critérios de seleção utilizados para comparar os modelos são os critérios de informação de Akaike (AIC) e de Schwarz (BIC).

**Palavras-chave:** GAMLSS, relações hipsométricas, heterocedasticidade, distribuições simétricas.

## 1 Introdução

A quantificação das alturas das árvores é fundamental principalmente para auxiliar no planejamento sustentável do estoque de inventários florestais. Por outro lado, medir a altura de uma árvore grande é inviável exigindo muito tempo e geralmente sendo feita de forma indireta, o que leva a erros.

Desse modo, na prática, procura-se medir em uma certa área (chamada parcela) a altura de parte das árvores e o diâmetro de todas e, através das medidas altura-diâmetro estabelecer uma relação matemática que possibilite estimar as alturas das demais árvores contidas nas parcelas. Estes modelos matemáticos que relacionam a altura da árvore ( $H$ ) com o diâmetro do tronco ( $D$ ), medidos a  $1,3m$  do solo, são chamados modelos de hipsometria.

Neste trabalho, pretende-se utilizar as relações hipsométricas dadas pelos modelos de Stofel, Curtis, Petterson e Prodan para modelar o parâmetro de localização e, pretende-se também modelar a variância, em que um parâmetro de escala é atribuído para cada observação, sendo

---

<sup>1</sup>Agradecimento ao CNPq pelo apoio financeiro.

relacionado através de combinações das variáveis explicativas por meio de uma função de ligação conhecida.

Utiliza-se o GAMLSS pois, segundo Stasinopoulos & Rigby (2007) pode-se trabalhar com uma família mais geral de distribuições para a variável resposta fora do contexto da família exponencial, em que a parte sistemática do modelo é expandida como funções lineares e/ou não-lineares, paramétricas e/ou não-paramétricas aditivas das variáveis explicativas e/ou efeitos aleatórios, permitindo a modelagem não só do parâmetro de locação mas também dos outros parâmetros da distribuição da variável resposta.

## 2 Metodologia

De forma geral é possível representar a relação entre o diâmetro ( $D$ ) do tronco da árvore e a altura total da árvore ( $H$ ) usando a relação

$$\mathbf{H} = \mathbf{h}(\mathbf{D}; \boldsymbol{\beta}). \quad (1)$$

Os modelos de regressão geralmente adotam a seguinte forma

$$\ln(H_i) = \boldsymbol{\eta}_1(D_i, \boldsymbol{\beta}) + \epsilon_i, \quad (2)$$

sendo  $\epsilon_i$  um erro aleatório independente e identicamente distribuído, com  $\epsilon_i \sim N(0, \sigma^2)$  e  $\boldsymbol{\eta}_1(D_i, \boldsymbol{\beta}) = \ln(\mathbf{h}(D_i; \boldsymbol{\beta}))$  para  $i = 1, \dots, n$ .

Neste trabalho, considera-se a variável aleatória  $\mathbf{Y} = \ln(\mathbf{H})$  e assume-se que  $\mathbf{Y}$  tem distribuição na família simétrica com  $E(\mathbf{Y}) = \boldsymbol{\mu}$ .

No contexto dos modelos generalizados (lineares ou não-lineares) considera-se conhecida uma função de ligação  $g_1(\cdot)$  biunívoca, contínua e duas vezes diferenciável tal que

$$g_1(\boldsymbol{\mu}) = \boldsymbol{\eta}_1(\mathbf{D}, \boldsymbol{\beta}). \quad (3)$$

Adota-se, para o problema considerado neste trabalho,  $g_1(\cdot)$  como sendo a função identidade tal que

$$\boldsymbol{\mu} = \boldsymbol{\eta}_1(\mathbf{D}, \boldsymbol{\beta}), \quad (4)$$

em que  $\boldsymbol{\eta}_1(\mathbf{D}, \boldsymbol{\beta})$  são os modelos hipsométricos descritos na seção 2.2.

Deseja-se também modelar a variância dos dados. Dessa forma, considera-se uma função conhecida biunívoca, contínua e diferenciável  $g_2(\cdot)$ , que deve ser positiva, tal que

$$g_2(\boldsymbol{\phi}) = \boldsymbol{\eta}_2(\mathbf{D}, \boldsymbol{\alpha}) \quad (5)$$

Admite-se  $g_2(\cdot)$  como sendo a função log e supõe-se que a variabilidade é descrita da seguinte forma

$$\log(\phi) = \alpha_0 + \alpha_1[1/\ln(\mathbf{D})] \quad (6)$$

A seguir são apresentadas as distribuições simétricas para a variável resposta bem como os modelos de hipsometria considerados neste trabalho.

## 2.1 Distribuições

### Distribuição Normal

Uma variável aleatória tem distribuição normal, denotada por  $N(\mu, \phi)$ , se sua função densidade de probabilidade é dada por

$$f_y(y|\mu, \phi) = \frac{1}{\sqrt{2\pi\phi}} \exp\left\{-\frac{(y-\mu)^2}{2\phi}\right\}$$

para  $-\infty < y < \infty$  em que  $-\infty < \mu < \infty$  e  $\phi > 0$ , sendo  $E(Y) = \mu$  e  $Var(Y) = \phi$ .

### Distribuição Logística-II

Uma variável aleatória tem distribuição logística-II, denotada por  $LII(\mu, \phi)$ , se sua função densidade de probabilidade é da forma

$$f_y(y|\mu, \phi) = \frac{1}{\sqrt{\phi}} \left\{ \exp\left[-\left(\frac{y-\mu}{\sqrt{\phi}}\right)\right] \right\} \left\{ 1 + \exp\left[-\left(\frac{y-\mu}{\sqrt{\phi}}\right)\right] \right\}^{-2}$$

para  $-\infty < y < \infty$  em que  $-\infty < \mu < \infty$  e  $\phi > 0$ , sendo  $E(Y) = \mu$  e  $Var(Y) = \pi^2\phi/3$ .

### Distribuição Normal Generalizada

Uma variável aleatória tem distribuição normal generalizada, denotada por  $NG(\mu, \phi, \nu)$ , se sua função densidade de probabilidade é da forma

$$f_y(y|\mu, \phi, \nu) = \frac{\nu}{2\sqrt{\phi}\Gamma\left(\frac{1}{\nu}\right)} \exp\left\{-\left|\frac{y-\mu}{\sqrt{\phi}}\right|^\nu\right\}$$

para  $-\infty < y < \infty$  em que  $-\infty < \mu < \infty$ ,  $\phi > 0$  e  $\nu > 0$  sendo  $E(Y) = \mu$  e  $Var(Y) = \phi\Gamma\left(\frac{3}{\nu}\right)/\Gamma\left(\frac{1}{\nu}\right)$ .

### Distribuição da família t de Student

Uma variável aleatória tem distribuição da família t de student, denotada por  $TF(\mu, \phi, \nu)$ , se sua função densidade de probabilidade é dada por

$$f_y(y|\mu, \phi, \nu) = \frac{1}{\sqrt{\phi}B\left(\frac{1}{2}, \frac{\nu}{2}\right)\nu^{1/2}} \left[1 + \frac{(y-\mu)^2}{\phi\nu}\right]^{-\frac{\nu+1}{2}}$$

para  $-\infty < y < \infty$  em que  $-\infty < \mu < \infty$ ,  $\phi > 0$  e  $\nu > 0$  sendo  $E(Y) = \mu$  e  $Var(Y) = \phi\nu/(\nu - 2)$ .

### Distribuição da família Normal

Uma variável aleatória tem distribuição da família normal, denotada por  $NF(\mu, \phi, \nu)$ , se sua função densidade de probabilidade é dada por

$$f_y(y|\mu, \phi, \nu) = \frac{1}{\sqrt{2\pi\phi}|\mu|^{\nu/2}} \exp\left\{\frac{-(y - \mu)^2}{2\phi|\mu|^\nu}\right\}$$

para  $-\infty < y < \infty$  em que  $-\infty < \mu < \infty$ ,  $\phi > 0$  e  $-\infty < \nu < \infty$  sendo  $E(Y) = \mu$  e  $Var(Y) = \phi|\mu|^\nu$ .

## 2.2 Relações Hipsométricas

Existe uma grande quantidade de modelos que descrevem a relação entre o diâmetro ( $D$ ) do tronco da árvore e a sua altura total ( $H$ ) (Do Couto & Bastos, 1987; Curtis, 1967; Soares *et al.*, 2004). Neste trabalho serão considerados apenas quatro deles que podem ser vistos na Tabela 1.

**Tabela 1:** Relação dos modelos de hipsometria.

Autor	$y$	$\eta_1(D, \beta)$
Stofel	$\ln(H)$	$\beta_0 + \beta_1 \ln(D)$
Curtis	$\ln(H)$	$\beta_0 - \beta_1(1/D)$
Petterson	$H$	$[\beta_0 + \beta_1(1/D)]^{-3}$
Prodan	$H$	$[\beta_0 + \beta_1(1/D) + \beta_2(1/D)^2]^{-1}$

## 3 Resultados e Discussão

A metodologia proposta foi aplicada ao conjunto de dados contendo 1309 pares mensuráveis (altura, diâmetro) de *Eucalyptus sp.*. As distribuições assumidas para a variável resposta  $Y$  se encontram na seção 2.1 e o ajuste dos modelos com essas distribuições foi feito a partir da função `nlglmss` do software R. Por fim, os modelos foram comparados utilizando os critérios de informação AIC e BIC, em que o modelo mais versátil aos dados foi escolhido pelo menor valor desses critérios.

Os valores dos critérios de seleção calculados para cada modelo são apresentados na Tabela 2.

**Tabela 2:** Critérios de seleção AIC e BIC para os modelos considerados.

Distribuições	Modelos							
	Stofel		Curtis		Petterson		Prodan	
	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
Normal	-2182.3	-2161.6	-2147.8	-2127.1	-2164.9	-2144.2	-2161.0	-2135.1
Logística-II	-2126.8	-2106.1	-2089.7	-2069.0	-2107.9	-2087.2	-2120.0	-2094.1
Normal Generalizada	-2197.1	-2171.2	-2168.0	-2142.1	-2182.9	-2157.0	-2197.8	-2166.7
$t_2$	-2180.3	-2154.4	-2145.8	-2120.0	9100.5	9126.3	-2181.5	-2150.4
Família Normal	-2225.7	-2199.8	-2150.2	-2124.3	-2108.6	-2134.3	-2125.3	-2110.0

Ambos os critérios, AIC e BIC, determinam que o modelo que melhor representa o conjunto de dados analisado é dado pela família normal com a relação hipsométrica de Stofel.

Na tabela a seguir é apresentado um sumário do modelo escolhido.

**Tabela 3:** Sumário do modelo.

Parâmetros	$\beta_0$	$\beta_1$	$\alpha_0$	$\alpha_1$	$\nu$
Estimavas	1.3095	0.6457	35.29	-29.08	-47.88
Desvio Padrão	0.05594	0.01972	7.813	5.827	9.895
p-valor	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001

Na figura a seguir são apresentados os gráficos do comportamento do parâmetro de locação  $\mu$  comparando com os dados e do desvio padrão de  $\phi$  do modelo escolhido (família normal com a relação de Stofel).

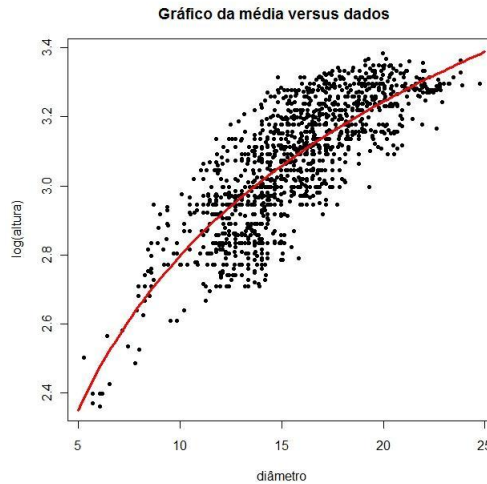


Figura 1: Gráfico do parâmetro de locação ( $\mu$ ) com os dados.

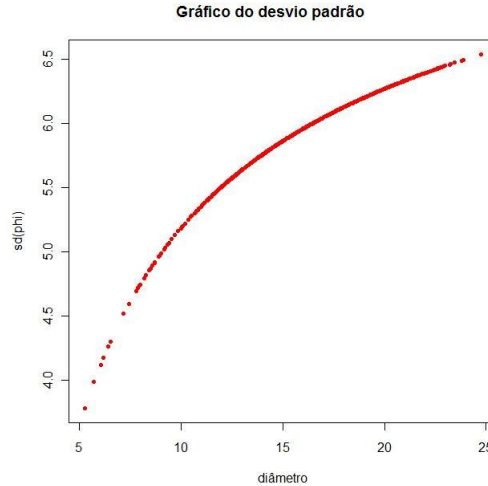


Figura 2: Gráfico do desvio padrão de  $\phi$ .

## 4 Conclusões

Dentre os modelos propostos e de acordo com os resultados obtidos se conclui que de fato, o modelo que melhor acomoda os dados é aquele representado pela relação de Stofel com a suposição da distribuição da família normal para a variável resposta. Em relação a heterocedasticidade, pode-se notar que a mesma está relacionada ao diâmetro das árvores, não descartando a possibilidade da influência de outros fatores que ainda devem ser estudados.

## Referências

- Curtis, R. O. (1967). Height-diameter and height-diameter-age equations for second-growth Douglas-fir. *Forest Science*, **13**(4), 365–375.
- Do Couto, H. T. Z. & Bastos, N. L. M. (1987). Modelos de equações de volume e relações hipsométricas para plantações de Eucalyptus no Estado de São Paulo. *IPEF*, (37), 33–44.
- Soares, T. S., Scolforo, J. R. S., Ferreira, S. O. & Mello, J. M. D. (2004). Uso de diferentes alternativas para viabilizar a relação hipsométrica no povoamento florestal. *Rev. Árvore*, **28**(6), 845–854.
- Stasinopoulos, D. M. & Rigby, R. A. (2007). Generalized Additive Models for Location Scale and Shape (GAMLSS) in R. *Journal of Statistical Software*, **23**(7), 1–46.