Modelagem do crescimento de clones de *Eucalyptus* via modelo de Chapman-Richards com diferentes distribuições simétricas

Luiz Medeiros de Araujo Lima Filho[1], José Antônio Aleixo da Silva[2], Gauss Moutinho Cordeiro[3] e Rinaldo Luiz Caraciolo Ferreira[2]

[1] Programa de Pós-Graduação em Biometria e Estatística Aplicada, Universidade Federal Rural de Pernambuco, CEP 52171-900 - Recife (PE) - Brazil
[2] Departamento de Ciência Florestal, Universidade Federal Rural de Pernambuco, CEP: 52171-900 - Recife (PE)- Brazil
[3] Departamento de Estatística e Informática, Universidade Federal Rural de Pernambuco, CEP: 52171-900 - Recife (PE)- Brazil
`luiz@de.ufpb.br`

Resumo O Pólo Gesseiro do Araripe em Pernambuco é um grande consumidor de madeira para produção de gesso. Devido à grande necessidade de se buscar uma alternativa econômica e ambiental para a região é de interesse obter uma produção sustentável para o *Eucalyptus* ssp, uma vez que esta é um gênero de rápido crescimento e grande versatilidade. No planejamento do manejo florestal sustentado uma variável é de extrema importância: o crescimento. Sua modelagem é fundamental na prognose da produtividade, qualidade do local e dinâmica de populações. Geralmente, as curvas de crescimento são estudadas por meio de modelos não-lineares desenvolvidos empiricamente para relacionar, por exemplo, circunferência e idade. Um modelo não-linear bastante utilizado na prática para modelar curvas de crescimento é o modelo de Chapman-Richards. Em estudos deste tipo, em geral, assume-se que os erros seguem distribuição normal. Contudo, a modelagem sob a suposição de erros com distribuição normal é bastante sensível a valores atípicos que por ventura possam ocorrer, podendo distorcer as estimativas dos parâmetros. Uma alternativa para corrigir esse problema é adotar distribuições mais robustas que a distribuição normal. Desta forma, a classe de modelos simétricos se torna uma alternativa viável para corrigir tal problema. Com a expectativa de obter melhores estimativas de crescimento de *Eucalyptus* ssp, aplicaram-se ao modelo de Chapman-Richards as seguintes distribuições: normal, t de Student, Cauchy, exponencial potência e logística II que apresentou a distribuição Cauchy com melhores estimativas de crescimento em circunferência de *Eucalyptus* ssp no Pólo Gesseiro de Pernambuco.

Palavras-chave: Log-Verossimilhança, Modelos Não-Lineares, Modelos Simétricos.

1 Introdução

O Polo Gesseiro do Araripe, localizado na microregião de Araripina, semiárido Pernambucano, é um grande consumidor de biomassa vegetal que é usada na calcinação da gipsita. Essa microregião abrange 10 municípios e é responsável por 95% do gesso industrializado no Brasil (ALBUQUERQUE, 2002).

O bioma caatinga, no qual está localizado o Pólo Gesseiro do Araripe-PE, vem sofrendo pressão visto que é explorado de forma desordenada. Esse fato se deve, principalmente, a crescente demanda por recursos naturais renováveis, aumentando gradativamente a sua degradação. Uma alternativa econômica e ambiental viável é a implementação e o manejo sustentado de povoamentos florestais nativos ou o reflorestamento com florestas de rápido crescimento, com destaque para o *Eucalyptus* ssp por sua elevada taxa de crescimento, a facilidade de reprodução, a rusticidade e o altíssimo nível de melhoramento genético em produtividade e qualidade da madeira.

Em face desse fato, torna-se de interesse quantificar o crescimento e a produção de florestas, promovendo um planejamento criterioso da produção através da prescrição de regimes de manejos adequados visando à qualidade do produto final. Sendo assim, pode-se dizer que a predição do crescimento e da produção é parte fundamental do processo de planejamento dos povoamentos florestais. Geralmente, as curvas de crescimento são estudadas por meio de modelos não-lineares. Um modelo não-linear bastante utilizado para descrever tais fenômenos em ciências florestais é o modelo de Chapman-Richards.

Ao longo dos anos, modelos supondo erros normais vêm sendo utilizados para descrever a maioria dos fenômenos aleatórios, dado que a suposição de normalidade sempre é muito atrativa para os erros dos modelos de regressão com resposta contínua. Contudo, observa-se que as estimativas obtidas para os coeficientes dos modelos normais se mostram sensíveis à presença de observações extremas. Desta forma, alternativas à suposição de erros normais têm sido propostas na literatura. Lange et al. (1989) propuseram o modelo baseado na suposição de erros t de Student. Taylor (1992) propôs o ajuste de um modelo de regressão linear supondo erro com distribuição exponencial potência com um parâmetro extra de forma. Galea et al. (2005) apresentaram alguns resultados sobre modelagem, em particular sobre o desenvolvimento da análise inferencial e de diagnóstico na classe não-lineares com erros simétricos independentes.

O objetivo deste trabalho consiste em estimar a circunferência dos *Eucalyptus* ssp, através de modelos simétricos não-lineares baseados em erros com distribuições mais robustas que a distribuição normal.

2 Distribuições Simétricos

A família de distribuições simétricas gera uma classe geral de distribuições com a mesma simetria que a distribuição normal padrão. Entre essas distribuições podemos citar: t de Student, Cauchy, exponencial potência e logística II. Para maiores detalhes sobre a família de distribuições simétricas em modelos de regressão, podem ser encontradas em Cysneiros e Paula (2005a).

Diz-se que a variável aleatória Y tem distribuição simétrica, com suporte em $\Re$, com parâmetros de locação $\mu \in \Re$ e de escala $\phi > 0$, se sua função de densidade de probabilidade é dada por

$$f(y; \mu, \phi) = \frac{1}{\sqrt{\phi}} g\left\{u\right\}, \qquad y \in \Re, \tag{1}$$

para alguma função $g(\cdot)$ denominada função geradora de densidade, em que $u = \frac{(y-\mu)^2}{\phi}$, com $g(u) > 0$, para $u > 0$ e $\int_0^\infty u^{-1/2} g(u) du = 1$. Essa condição é necessária para que $f(y; \mu, \phi)$ seja uma função densidade de probabilidade. Assim, denota-se por $Y \sim S(\mu, \phi)$ e denomina-se de variável aleatória simétrica.

A seguir, são apresentadas algumas distribuições simétricas com suporte na reta real para $Y \sim S(\mu, \phi)$ em que $u = (y - \mu)^2/\phi$.

2.1 Distribuição Normal

Diz-se que $Y \sim S(\mu, \phi)$ tem distribuição normal se sua função geradora de densidade $g(\cdot)$ é da forma

$$g(u) = \frac{1}{\sqrt{2\phi}} exp\left(-u/2\right), \qquad u > 0, \tag{2}$$

então, Y tem distribuição normal denotada por $Y \sim N(\mu, \phi)$. O coeficiente de curtose desta distribuição é $\gamma_2 = 3$.

2.2 Distribuição t de Student

A variável aleatória $Y \sim S(\mu, \phi)$ tem distribuição t de Student se sua função geradora de densidade $g(\cdot)$ é da forma

$$g(u) = \frac{\nu^{\nu/2}}{B(1/2, \nu/2)} \left(\nu + u\right)^{-\frac{\nu+1}{2}}, \qquad \nu > 0, u > 0, \tag{3}$$

em que $B(\cdot, \cdot)$ é a função Beta. Assim, Y é denotada por $Y \sim t(\mu, \phi, \nu)$. O coeficiente de curtose é $\gamma_2 = 3 + \frac{6}{\nu-4}$, para $\nu > 4$. Este coeficiente é maior que o coeficiente da distribuição normal.

2.3 Distribuição Cauchy

A variável aleatória $Y \sim S(\mu, \phi)$ tem distribuição de Cauchy se sua função geradora de densidade $g(\cdot)$ é da forma

$$g(u) = \frac{1}{\pi(1 + u)}, \qquad u > 0. \tag{4}$$

Essa distribuição, denotada por $Y \sim C(\mu, \phi)$, é também conhecida como distribuição de Pearson Tipo VII.

Por questão de brevidade, outras distribuições simétricas podem ser encontradas em Cysneiros et al.(2005).

3 Modelos Simétricos

A família simétrica de densidades de locação-dispersão guarda a estrutura da distribuição normal, mas elimina a forma específica da densidade normal para incluir densidades simétricas com caudas mais leves ou mais pesadas do que as caudas da normal.

Para introduzir uma estrutura regressora na classe de distribuições (1), toma-se a componente sistemática do modelo linear generalizado para o vetor da média $\mu = E(Y)$ dado por

$$g(\mu) = \eta_i(\beta) = h(x_i, \beta), \tag{5}$$

em que $g(\cdot)$ é conhecida e duas vezes diferenciável, $\eta_i(\beta)$ é o preditor não-linear, X é uma matriz $n \times p$ de posto completo e $\beta = (\beta_1, ..., \beta_p)^T$ é um conjunto de parâmetros não-lineares desconhecidos a serem estimados.

Os modelos simétricos assumem que as variáveis aleatórias $Y_1, ..., Y_n$ podem ser tratadas como distribuidas independentemente seguindo a componente aleatória (1) e a componente sistemática (5). Desta forma, o modelo definido em (1) e (5) é dito modelo simétrico não-linear.

O principal objetivo na análise de modelos simétricos é fazer inferências no vetor de parâmetros β e no parâmetro de dispersão ϕ. A log-verossimilhança para os parâmetros do modelo pode ser expressa como:

$$l(\beta, \phi) = -\frac{n}{2} \log \phi + \sum_{i=1}^{n} \log g \left\{ \phi^{-1}(y_i - \mu_i)^2 \right\}. \tag{6}$$

A log-verossimilhança apresentada pode ser maximizada incondicionalmente usando alguns softwares como o SAS, Matlab, R ou a linguagem de programação Ox.

A função escore para o vetor de parâmetro β e parâmetro de dispersão ϕ é

$$U_\beta = \phi^{-1} X^T H D(y - \mu) \tag{7}$$

e

$$U_\phi = (2\phi)^{-1}\{\phi^{-1} Q(\mu, y) - n\}. \tag{8}$$

Os parâmetros β e ϕ são globalmente ortogonais. Assim, as estimativas de máxima verossimilhança de $\hat{\beta}$ e $\hat{\phi}$ são assintoticamente independentes devido a sua normalidade assintótica e a estrutura bloco diagonal da matriz de informação conjunta, isto é, $K = diag(K_\beta, k_\phi)$, em que $K_\beta = 4a\phi^{-1} X^T H^2 X$ e $k_\phi = n(4b - 1)/(4\phi^2)$ são, respectivamente, as matrizes de informação para β e ϕ.

Os estimadores de máxima verossimilhança de β e ϕ não podem ser resolvidos explicitamente. No entanto, podem ser resolvidas por meio de um método iterativo, como por exemplo, o método escore de Fisher. Assim, o processo iterativo se reduz a

$$\beta^{(m+1)} = (X^T H^{(m)2} X)^{-1} X^T H^{(m)2} \delta^{(m)} \tag{9}$$

e

$$\phi^{(m+1)} = \frac{1}{n}Q(\mu^{(m)}, y), \tag{10}$$

em que

$$\delta^{(m)} = \eta^m + (4a)^{-1}H^{(m)(-1)}D^{(m)}(y - \mu^{(m)}). \tag{11}$$

Aproximações iniciais $\beta^{(1)}$ e $\phi^{(1)}$ escolhidas são usadas para avaliar $H^{(1)}$, $D^{(1)}$ e $\delta^{(1)}$ e assim produzirem a próxima estimativa para $\beta^{(2)}$. Então, atualiza-se $\mu^{(2)}$ e $Q(\mu^{(2)}, y)$ para encontrar $\phi^{(2)}$ e assim continuando as iterações até que as estimativas β e ϕ sejam obtidas.

4 Material e método

Em março de 2002, foi implantado na Estação Experimental do Instituto Agronômico de Pernambuco na Chapada do Araripe - PE o Módulo de Experimentação Florestal para o Pólo Gesseiro do Araripe. Foram utilizadas 83 árvores, sobreviventes das 100 árvores plantadas no início do experimento, sendo 25 árvores em cada uma das quatro repetições. A variável circunferência foi medida à altura do peito (CAP) em todas as árvores ao longo do tempo, durante cinco anos e meio. O tempo inicial considerado foi de doze meses.

Para estimar a altura dos *Eucalyptus* ssp foi utilizado o modelo de Chapman-Richards, na estrutura de modelos lineares generalizados para a média $\mu = E(Y)$, definido por

$$\mu = U(1 - \exp(-kt))^{\theta}, \tag{12}$$

com t correspondendo à idade da árvore em meses. Os modelos foram ajustados supondo diferentes distribuições simétricas para os erros. Este procedimento foi realizado por meio da Proc NLP do SAS.

Para comparar os modelos ajustados aos dados se utilizou o critério de informação de Akaike (AIC), critério de informação Bayesiana (BIC) e o erro percentual absoluto médio (MAPE).

5 Aplicação a dados reais

Para a seleção dos modelos ajustados, apresentamos na Tabela 1 o critério AIC, BIC e o MAPE. Assim, para os dados de circunferência o modelo supondo erro com distribuição Cauchy obteve menores valores para os três critérios adotados, ou seja, AIC (130,65), BIC (141,10) e MAPE (9,01).

Tabela 1. Estatísticas para seleção dos modelos.

Distribuições	AIC	BIC	MAPE
Normal	150,70	161,14	10,63
Student t_2	178,69	189,12	9,37
Exp. Potência (l=0,1)	224,27	234,70	10,19
Cauchy	130,65	141,10	9,01
Logística II	224,71	235,16	9,72

6 Conclusão

Desta forma, é possível concluir que os modelos simétricos são bastantes relevantes para os estudos de modelos de crescimento de forma prática e bastante útil para análise de dados reais, contribuindo assim de forma efetiva, no sentido de ampliar as possibilidades de análise para os modelos de crescimento adotados em Ciências Florestais.

Referências

1. ALBUQUERQUE, J.L. **Diagnóstico ambiental e questões estratégicas: Uma análise considerando o Pólo Gesseiro do Sertão do Araripe - Estado de Pernambuco**. 2002. Tese (Doutorado em Ciências Florestais) - Universidade Federal do Paraná, Brazil.
2. CYSNEIROS, F.J.A.; PAULA, G.A. Restricted Methods in Symmetrical Linear Regression Models. **Computational Statistics and Data Analysis**. V. 49, n. 3, p. 689-708, 2005a.
3. CYSNEIROS, F.J.A.; PAULA, G.A.; GALEA, M. **Modelos Simétricos Aplicados**. São Pedro: 9ª Escola de Modelos de Regressão, 2005b.
4. GALEA, M.; PAULA, G.A.; CYSNEIROS, F.J.A. On Diagnostic in Symmetrical Nonlinear Models. **Statistics and Probability Letters**. V. 73, n. 4, p. 459-467, 2005.
5. LANGE, K.L.; LITTLE, R.J.A.; TAYLOR, J.M.G. Robust statistical modeling using the t distribution. **Journal of the American Statistical Association**, v. 84, p. 881-896, 1989.
6. TAYLOR, J.M.G. Properties of modelling the error distribution with an extra shape parameter. **Computational statistics and data analysis**, v. 13, p. 33-46, 1992.