



**ROMULO BARBOSA VELOSO**

**PROPOSTA DE UM MODELO NÃO LINEAR DE  
CRESCIMENTO E PRODUTIVIDADE  
FLORESTAL**

**LAVRAS - MG**

**2013**

**ROMULO BARBOSA VELOSO**

**PROPOSTA DE UM MODELO NÃO LINEAR DE CRESCIMENTO E  
PRODUTIVIDADE FLORESTAL**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Engenharia Florestal, área de Concentração em Ciências Florestais, para a obtenção do título de Doutor.

Orientador

PhD. Natalino Calegario

**LAVRAS - MG**

**2012**

**Ficha Catalográfica Preparada pela Divisão de Processos Técnicos da  
Biblioteca da UFLA**

Veloso, Romulo Barbosa.

Proposta de um modelo não linear de crescimento e  
produtividade florestal / Romulo Barbosa Veloso. – Lavras : UFLA,  
2012.

102 p.: il.

Tese (doutorado) – Universidade Federal de Lavras, 2012.

Orientador: Natalino Calegário.

Bibliografia.

1. Floresta. 2. Medidas de curvatura. 3. Modelos matemáticos. 4.  
Modelos não lineares. 5. Estimativa. I. Universidade Federal de  
Lavras. II. Título.

CDD – 634.9285

**ROMULO BARBOSA VELOSO**

**PROPOSTA DE UM MODELO NÃO LINEAR DE CRESCIMENTO E  
PRODUTIVIDADE FLORESTAL**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Engenharia Florestal, área de Concentração em Ciências Florestais, para a obtenção do título de Doutor.

APROVADA em 31 de julho de 2012.

Dr. Adriano Ribeiro de Mendonça	UFES
Dr. José Higino Dias Filho	UNIMONTES
Dr. Lucas Rezende Gomide	UFLA
Dr. Marcelo Ângelo Cirillo	UFLA

PhD. Natalino Calegario

Orientador

**LAVRAS - MG**

**2012**

A DEUS.

À Vida.

Aos felizes que transformam a vida em viver.

### **OFEREÇO**

Aos meus pais, Arthur e Edith, que, pela sabedoria de suas palavras e o calor de seus afagos, nunca estiveram onde eu não pudesse alcançá-los.

A minha esposa, Dora e meus filhos, Denny, Denise, Lucy, Jeane e Thiago, que, caminhando sempre ao meu lado, me apoiando e incentivando, me ajudaram a transformar as dificuldades que surgiram em obstáculos vencidos.

A meus irmãos, Fátima, Roberto, Artur e Rogério, que, sendo minha infância e minha juventude, se tornaram responsáveis por parte importante de tudo que hoje sou.

A minhas netas Julia, Leticia e Luisa que, com alegria, vivacidade e meninice, iluminaram a vida de todos nós.

Ao amigo Caju que, independente de como eu esteja, com um dom natural e mágico, tem sempre a capacidade de me fazer sentir a pessoa mais querida do mundo.

### **DEDICO**

## **AGRADECIMENTOS**

À Universidade Estadual de Montes Claros, em particular ao Departamento de Ciências Exatas (DCEX ), pela confiança.

À Universidade Federal de Lavras e ao Departamento de Ciências Florestal (DEX), pelos conhecimentos.

Ao amigo e professor Natalino Calegario, pela disponibilidade, paciência e orientação.

Aos professores e amigos Yule Roberta, Ronaldo e José Higino, pela confiança e o estímulo para iniciar o que foi esta jornada.

Aos colegas Samuel, Nory, Elliezer, Ernani, Sabrina, Michelle e Ayuni, pelo incentivo, troca de conhecimentos e colaboração inestimáveis.

A todos aqueles que, direta ou indiretamente, contribuíram para que eu pudesse atingir este objetivo.

## RESUMO

A estimativa de crescimento e de produtividade é ferramenta imprescindível para tomadas de decisão no meio florestal. Neste trabalho destaca-se a importância do uso de modelos não lineares para a estimativa dessas variáveis, com a proposta de utilização de uma variação do modelo exponencial que é comparado ao modelo logístico de três parâmetros. As medidas de curvatura são utilizadas para validação dos ajustes dos modelos pela avaliação da “aproximação linear”. Com o objetivo de obter maior precisão, estes modelos foram ajustados e comparados com e sem o acréscimo de covariáveis nos respectivos parâmetros assintóticos. Foi, ainda, utilizada a metodologia de simulação Monte Carlo, para gerar 10.000 amostras, o que permite produzir o estudo da distribuição dos parâmetros. Todas estas comparações dos modelos são feitas considerando-se três classes de sítio. Observou-se que o modelo proposto apresentou maior qualidade na aproximação linear, em todas as classes de sítio avaliadas.

Palavras-chave: Crescimento e produtividade. Modelos não lineares. Medidas de curvatura.

## **ABSTRACT**

The growth and productivity estimatives are indispensable tools for the decision making at the florestal studies. In this work is highlighted the importance of the usage of non-linear models to estimate those variables, with the proposal of utilization of a variation of the exponential model, that is compared with the three parameters logistic model. The curvature values are used to validate the adjustes of the models, through the evaluation of the "linear proximity". With the aim of acquire a higher acuracy, those models were adjusted and compared without adding co-variables at the respective assinthotic parameter. It is also used the methodology of Monte Carlo's simulation, to generate 10000 samples, what makes possible to produce a study of the parameters distribution. All those models comparations were done considering three classes of site was observed that the proposed model presents a higher quality of linear proximity at all classes of site evaluated.

**Keywords:** Growth and yield. Nonlinear models. Measures of curvature.



## SUMÁRIO

1	INTRODUÇÃO .....	10
2	OBJETIVOS .....	15
2.1	Objetivo geral.....	15
2.2	Objetivos específicos .....	15
3	REFERENCIAL TEÓRICO .....	16
3.1	O gênero <i>Eucalyptus</i> .....	16
3.2	Modelagem matemática.....	18
3.2.1	Modelos empíricos e modelos mecanicistas.....	19
3.2.2	Modelos estáticos e modelos dinâmicos .....	19
3.2.3	Modelos determinísticos e modelos estatísticos.....	20
3.3	Modelos lineares e modelos não lineares .....	22
3.4	Métodos de estimação do vetor de parâmetros .....	22
3.4.1	O método dos mínimos quadrados .....	23
3.4.1.1	Pressupostos da regressão.....	24
3.4.1.1.1	Ausência da normalidade dos erros .....	24
3.4.1.2	Ausência da homocedasticidade .....	26
3.4.1.3	Erros correlacionados.....	27
3.4.2	O método da máxima verossimilhança.....	32
3.5	Medidas de não linearidade .....	33
3.5.1	Representação geométrica da não linearidade.....	34
3.6	Estimativa da altura total e o uso de covariantes.....	40
3.7	Estimação do volume de povoamentos florestais .....	41
3.7.1	Método da tabela de volume .....	42
3.7.2	Método da tabela de produtividade.....	42
3.8	Modelos e funções de crescimento .....	43
3.8.1	O modelo proposto.....	45
3.9	Classificação da capacidade produtiva .....	45
3.10	Critério de informação de Akaike .....	48
3.11	Coefficiente de determinação ajustado - $R^2_{aj}$ .....	49
3.12	Erro de predição médio – EPM.....	50
4	MATERIAIS E MÉTODOS .....	51
4.1	Coleta de dados .....	51
4.2	Índices e curvas de sítio .....	54
4.2.1	Construção das curvas de índice local pelo método da curva guia .....	54
4.3	Modelos para estimativa da altura total.....	56
4.5	Avaliação dos modelos.....	58
4.4.1	Medidas de não linearidade ou de curvatura.....	58
4.4.2	Critério de informação de Akaike .....	60

4.4.3	Coefficiente de determinação ajustado - $R^2_{aj}$ .....	61
4.4.4	Erro de predição médio (EPM) .....	61
4.4.5	Erro relativo percentual .....	62
5	<b>RESULTADOS E DISCUSSÃO</b> .....	63
5.1	Ajustes dos modelos com decomposição de parâmetros .....	72
5.2	Estimativa do DAP para estabelecer a produtividade com o modelo completo .....	75
5.3	Modelo de produtividade compatível com o modelo de crescimento em altura.....	75
5.4	Modelos compatíveis de crescimento e produtividade de volume .....	78
6	<b>CONCLUSÃO GERAL</b> .....	83
	<b>REFERÊNCIAS</b> .....	84
	<b>ANEXOS</b> .....	88

## 1 INTRODUÇÃO

A crescente importância econômica e social das plantações florestais, gerando empregos de forma direta e indireta, com contribuição superior a 40 bilhões de reais para o Produto Interno Bruto de 2011, conforme anuário estatístico da Associação Brasileira de Produtores de Florestas Plantadas - ABRAF (2012), faz com que estratégias de trabalho a serem adotadas tenham relevada importância. Davis e Johnson (1987) expressam que, dentre outras medidas, a predição da produtividade futura é de imprescindível importância para o planejamento das atividades florestais, constituindo um dos elementos mais importantes do manejo florestal. Isto faz com que o estabelecimento preciso do crescimento, da produção ou da produtividade seja desejado e necessário.

Para Ferreira (2010), o crescimento é o ato de crescer e ele define crescer como aumentar em volume, grandeza ou extensão. Na teoria florestal, o termo crescimento é utilizado para indicar a taxa absoluta de aumento do organismo na variável em observação, e o termo produção é empregado para representar o crescimento acumulado em uma área específica, enquanto o termo produtividade determina a produção por unidade de área.

Diversas são as variáveis dendrométricas que sofrem modificações pelo crescimento de uma árvore. Dentre elas, estão diâmetro à altura do peito (DAP), altura total, área basal e os diversos tipos de volumes. Os parâmetros de crescimento destas variáveis em uma árvore têm comportamentos semelhantes (ENCINAS; SILVA; PINTO, 2005).

Graficamente, o crescimento pode ser bem apresentado com uso de uma curva sigmoide, a qual permite a verificação da existência de três períodos na vida da árvore: a idade juvenil, que descreve o crescimento inicial, seguido do crescimento chamado de idade madura ou retilíneo, quando observa-se que, a

intervalos de tempo iguais correspondem acréscimos aproximadamente iguais nas variáveis e, por último, o período ou a idade de senescência, quando o crescimento é cada vez menos significativo, conforme apresentado na Figura 1. Esta curva de crescimento tem formas distintas por espécies, pois as proporções relativas a cada fase têm diferentes tempos de ocorrência e de duração para estes períodos.

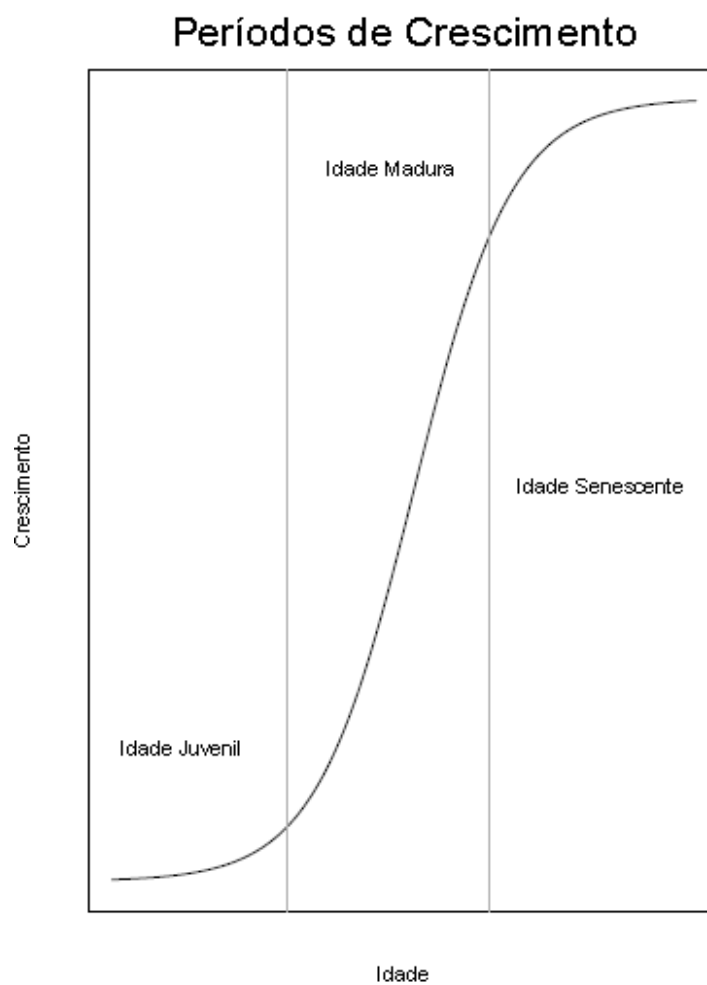


Figura 1 Representação dos períodos de crescimento em uma curva sigmoideal

A relação entre a variável que sofre o crescimento, dita variável resposta, e uma ou mais variáveis que o afetam, ditas variáveis explicativas, pode ser estabelecida por um modelo estatístico, pois esta atende a condições que permitem que o crescimento seja regido por uma lei que associa a cada valor de uma ou mais variáveis explicativas um valor da variável em estudo, ou seja, pode-se expressar a relação como um ou, em muitos casos, diversos modelos estatísticos, denominados de modelos de crescimento.

O incremento é uma variável que está matematicamente ligada ao crescimento, pois é a derivada primeira deste e, dentre os incrementos mais utilizados na bibliografia em trabalhos da área florestal, estão o Incremento Corrente Anual (ICA), que é a medida da variação do crescimento sofrida no período de um ano e ou o incremento médio anual (IMA), que expressa a variação relativa a diversos períodos anuais, apresentada como uma média dos períodos de interesse, ou seja, é a variação do crescimento, expressa como média dos anos em estudo.

A produtividade, segundo Campos e Ribeiro (1983), pode ser determinada pela análise da razão entre o volume, ou outra medida de crescimento, produzido em relação a outras variáveis, tais como a idade, a densidade, que pode ser expressa pela área basal por hectare e a qualidade de local, sendo de utilidade para a avaliação econômica de alternativas de manejo ou, mesmo, para o simples uso da madeira. Ainda segundo os mesmos autores, a análise desse inter-relacionamento entre as variáveis é uma das principais características que diferenciam os inventários florestais convencionais dos estudos de crescimento e produtividade. Os primeiros enfatizam, principalmente, a precisão das estimativas médias por uma unidade de área, independentemente do efeito das variáveis que são consideradas na decisão da época de corte.

Assim como o crescimento, a produtividade também pode ser modelada matematicamente para prognose, por meio de sua relação com as diversas

variáveis envolvidas no crescimento. Apesar dessa forte associação biológica e matemática entre o crescimento e a produtividade, apenas a partir de Buckman (1962) e Clutter (1963) é que surgiram modelos que relacionaram estas importantes variáveis da área florestal, chamadas de modelos de crescimento e produtividade compatíveis. Tais modelos baseiam-se no fato de que, sendo o crescimento uma taxa absoluta e a produtividade o acúmulo do crescimento, pode-se, então, a partir da integração do modelo de crescimento, estabelecer um modelo de produtividade que é plenamente compatível com os resultados apresentados. Devido ao forte apelo não linear, a transformação de variáveis para se obter a linearização de modelos que deveriam ser expressos na forma não linear ainda é a mais utilizada no meio florestal, pois, dessa forma, estende-se a aplicabilidade do uso da metodologia de regressão linear, obtida aplicando-se uma transformação antes de ajustar o modelo, ou seja, se a relação entre duas variáveis é não linear, então, frequentemente, são transformadas uma ou mais variáveis.

O uso de modelos não lineares torna-se, então, importantíssimo para o trabalho de manejo florestal, pois, segundo Ratkowsky (1983), transformações podem ser muito úteis em algumas circunstâncias, mas devem ser consideradas somente como um último recurso. Isso porque, quando uma, algumas ou todas as variáveis são transformadas, os coeficientes deixam de ter as mesmas interpretações diretas que tinham na equação original, o que equivale a dizer que o modelo resultante é capaz apenas de expressar uma relação entre as variáveis e este modelo se torna desprovido das informações biológicas importantes, que são, em geral, pertinentes à criação dos modelos não lineares.

Apesar da enorme importância dos modelos não lineares, estes ainda são pouco utilizados na área florestal, devido, sobretudo, conforme mencionado, à maior facilidade de ajuste dos modelos lineares, pois, nestes últimos, a solução do sistema de equações normais é obtida por meio algébrico, não sendo

necessária a utilização de métodos iterativos que necessitam de bons valores iniciais para alcançar a esperada convergência. Deve-se considerar também que modelos não lineares podem, ainda, ter seu uso restringido por seu grau de não linearidade, uma vez que, em caso da presença acentuada deste, diversas das inferências estatísticas obtidas podem ficar invalidadas.

Visando à maior precisão de modelos, Guimarães et al. (2009) apresentaram a importância da inclusão de covariantes na melhoria de modelos hipsométricos, ressaltando que a necessidade surge da existência de diversas variáveis que influenciam o crescimento em altura, o que se estende a outras variáveis de crescimento, tais como área basal e volume, que podem ser influenciadas por outras diversas variáveis dendrométricas, climáticas, etc.

Neste trabalho apresentam-se modelos estatísticos não lineares compatíveis em crescimento e em produtividade, acrescentando aos mesmos covariantes, para a melhoria da precisão das prognoses presente e futura de sortimentos de madeira em povoamentos florestais.

## **2 OBJETIVOS**

### **2.1 Objetivo geral**

Este trabalho foi realizado com o objetivo de avaliar o crescimento e a produtividade de volume em povoamentos florestais, com o uso de modelos estatísticos não lineares compatíveis para estas variáveis.

### **2.2 Objetivos específicos**

Especificamente, os objetivos foram:

- a) introduzir o uso de medidas de não linearidade em alguns dos principais modelos não lineares em aplicações florestais;
- b) utilizar e avaliar modelos não lineares com uso de covariante para estimativa de altura;
- c) apresentar um modelo de produtividade compatível com o modelo de crescimento.



### 3 REFERENCIAL TEÓRICO

#### 3.1 O gênero *Eucalyptus*

Observa-se crescente necessidade da sociedade por produtos de origem florestal, para atender à sobrevivência e ao conforto. Neste panorama de premências sociais, as florestas nativas, antes abundantes em todo o mundo, estão cada vez mais escassas e ameaçadas de desaparecer. As que restam tornaram-se indispensáveis para a manutenção da biodiversidade e de diversos serviços ambientais.

Neste contexto, as florestas plantadas surgem como opção, assumindo papel importante em todo o mundo. No Brasil, fatores silviculturais favoráveis, tais como condições climáticas, condições de solo, desenvolvimento de tecnologia, além de disponibilidade de área e de mão de obra, que são facilmente encontrados, favorecem a implantação destas florestas. Em virtude de seu rápido crescimento, grande capacidade de adaptação, alta produtividade e, ainda, por suas inúmeras aplicações em diversos setores, o eucalipto assume importância destacada nos plantios florestais.

O gênero *Eucalyptus* pertence à família Mirtaceae (subfamília Leptospermoidae). Geralmente, são árvores, podendo, em casos raros, apresentarem-se como arbustos. É nativo da Oceania, incluindo mais de 700 espécies, quase todas originárias da Austrália e é de fácil adaptação climática. A primeira descrição botânica do gênero foi realizada pelo francês Charles Louis L'Héritier de Brutelle, em 1788, e ele foi introduzido no Brasil no intuito de ajudar na produtividade de dormentes para as linhas férreas, implantadas a partir da segunda metade do século XIX. A distribuição do gênero, por área, no território brasileiro é apresentada na Tabela 1, com base no Anuário Estatístico da ABRAF 2011, ano base 2010 (ABRAF, 2012).

Tabela 1 Área e distribuição de plantios florestais com *Eucalyptus* no Brasil, 2010

UF	<i>Eucalyptus</i> (ha)
MG	1.400.000
SP	1.044.000
PR	161.422
BA	631.464
SC	102.399
RS	273.042
MS	378.195
ES	203.885
PA	148.656
MA	151.403
GO	58.519
AP	49.369
MT	61.950
TO	47.542
PI	37.025
Outros	4.650
Total	4.754.334

Fonte: Associadas individuais e coletivas da ABRAF (2012)

A alta importância econômica do eucalipto deve-se à multiplicidade de usos, assim como à dimensão da expressiva área plantada existente. Segundo a ABRAF (2012), esta planta está presente em todos os estados brasileiros, sendo utilizada para a produção de celulose, carvão vegetal para siderurgia, painéis de aglomerado de madeira, MDF, HDF, chapa de fibra, compensado, madeira serrada, lenha e biomassa como fonte de energia, óleos essenciais, postes e mourões, com fluxo de cadeia produtiva (Figura 2).

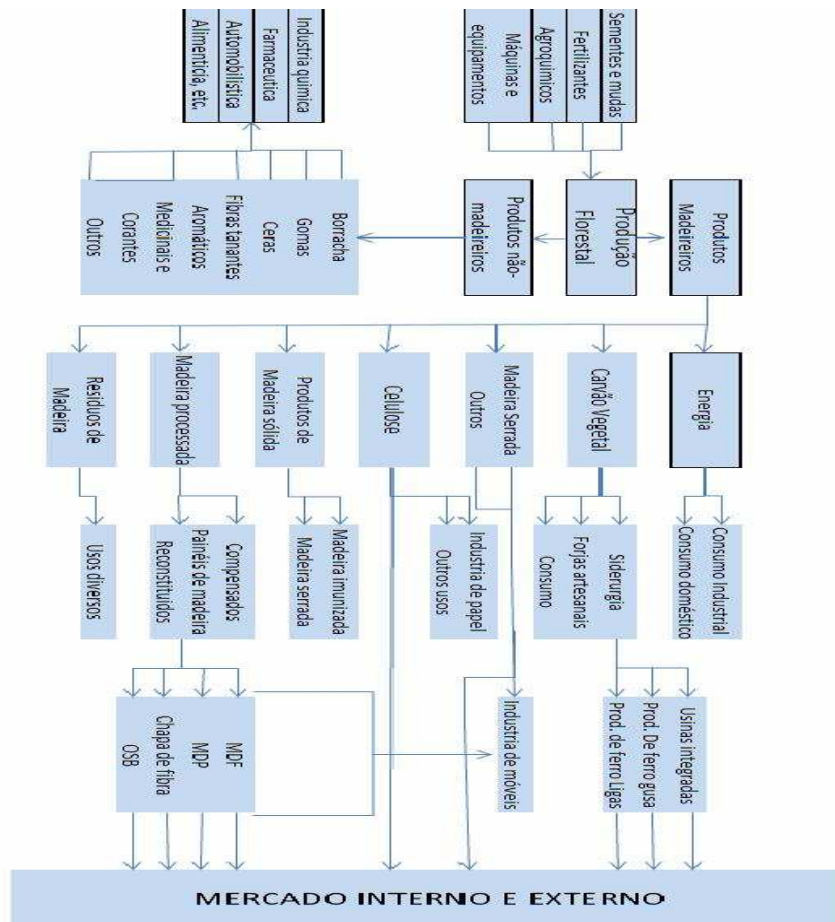


Figura 2 Fluxo da cadeia produtiva da madeira

Fonte: ABRAF (2012)

### 3.2 Modelagem matemática

Thornley e France (2004) dividem os modelos, quanto ao tipo, nas seguintes categorias: empíricos, mecânicos, estáticos, dinâmicos, determinísticos ou estocásticos.

### **3.2.1 Modelos empíricos e modelos mecanicistas**

Um modelo empírico busca, principalmente, descrever as respostas de um fenômeno, processo ou sistema em estudo, geralmente utilizando-se equações matemáticas ou estatísticas sem qualquer conteúdo científico ou restrições por princípios científicos que tenham dependência com os objetivos (THORNLEY; FRANCE, 2004). Já os modelos mecanicistas, além de descrever, buscam a compreensão do fenômeno ou objeto em estudo, provendo um grau de entendimento ou explicação intrínseco à modelagem do mesmo. Modelos empíricos estão sujeitos a menos ou a nenhuma restrição, o que torna sempre possível encontrar algum modelo deste tipo que ajuste melhor que um modelo mecanicista utilizado para o mesmo fim, sobretudo com um modelo empírico com maior número de parâmetros. Este trabalho foi desenvolvido sob a concepção de modelos mecanísticos, contudo, compreendendo que estes modelos dependem de elementos empíricos, nos quais a formulação e as inter-relações estão longe serem triviais.

### **3.2.2 Modelos estáticos e modelos dinâmicos**

Pode-se definir um modelo estático como um modelo que não considera a variável tempo, ou seja, a componente tempo do sistema é ignorada. Os modelos dinâmicos, por outro lado, têm a variável tempo explícita e predizem como as quantidades envolvidas variam com o tempo. Os modelos dinâmicos são, em geral, primeiramente, expressos como um conjunto de equações diferenciais ordinárias, com o tempo  $t$  como a variável independente. Muitas vezes, o modelo é apresentado como um conjunto de equações diferença, relacionando a variável em estudo num tempo  $t$ , a um tempo  $t+1$ , e assim sucessivamente. Se a taxa de mudança do sistema torna-se nula no tempo, então,

o modelo deixa de ser dinâmico e torna-se estático. Nestes modelos, variáveis são quantidades que mudam no tempo e elas podem ser consideradas sob quatro categorias: estado, taxa, auxiliares e autovariáveis.

Uma variável de estado é aquela que define ou ajuda a definir o estado de um sistema em dado ponto do tempo. A variável taxa é uma quantidade que define alguns processos no sistema a dado ponto no tempo. As variáveis ditas auxiliares podem ser vistas como variáveis extras no modelo, enquanto as autovariáveis são entradas do modelo, as quais variam autonomamente no tempo.

### 3.2.3 Modelos determinísticos e modelos estatísticos

Modelos determinísticos fazem previsões dos valores das variáveis respostas envolvidas sem nenhuma associação com probabilidades, o que é aceitável em muitos casos, mas, para muitas variáveis, tais como chuvas ou migração de animais, isso pode não ser aceitável.

A representação geral de um modelo determinístico pode ser feita da forma:

$$Y = f(X_1, X_2, \dots, X_n, \theta_1, \theta_2, \dots, \theta_k) = f(X_b, \theta). \quad (1)$$

No modelo,  $f(X_b, \theta)$  é a função, ou lei, que associa as N variáveis explicativas  $X_i$  ( $i=1, \dots, n$ ) à variável resposta Y. A utilização de  $f(X_b, \theta)$  simplifica a notação, sendo  $X_i$  o vetor de n variáveis explicativas e  $\theta$  o vetor de k parâmetros desconhecidos que se deseja determinar na função. As variáveis  $X_1, X_2, \dots, X_n$  representam os fatores que influenciam a resposta e os  $\theta$  parâmetros exprimem as quantidades em que estas serão expressas na lei de associação.

A introdução de elementos estocásticos ou aleatórios em um modelo matemático conduz a um modelo que é menos parcimonioso que o modelo

determinístico que o origina e conduz, ainda, à presença do elemento aleatório, o que faz com que as previsões tenham uma distribuição de probabilidades. Thornley e France (2004) apontam que um problema com os modelos estocásticos é que eles podem ser tecnicamente difíceis de construir e testar e, ainda, que a aparente estocasticidade de um sistema pode meramente refletir nosso desconhecimento de elementos envolvidos na resposta observada. Modelos estocásticos, ou modelos estatísticos, são aqueles que contêm constantes desconhecidas a serem estimadas a partir de dados empíricos. Os modelos estatísticos, de certa forma, podem ser considerados modelos determinísticos com algum componente extra, chamado de componente aleatório e pode ser representado, de forma geral, por

$$Y = f(X_1, X_2, \dots, X_n, \theta_1, \theta_2, \dots, \theta_k) + e_i \quad \Rightarrow Y = f(X_i, \theta) + e_i, \quad (2)$$

em que  $f(X, \theta)$  é o componente determinístico ou parte sistemática, conforme apresentado em (1), e é dado pela função média de um modelo e  $e_i$  é o componente aleatório, e este representa a diferença entre as observações e a função média, ou seja, o componente aleatório é responsável pela parte não controlada ou não conhecida das observações e está associado a uma função probabilística de densidade, por exemplo, distribuição Normal, Exponencial, Weibull, Gama ou Binomial, dentre outras. Um modelo estatístico de qualidade depende da melhor escolha do componente determinístico, assim como da melhor caracterização do componente aleatório. O pesquisador, para utilizá-lo, deverá, então, a partir dos dados observados, escolher um modelo apropriado.

### 3.3 Modelos lineares e modelos não lineares

Uma definição importante a ser atribuída a um modelo é quanto ao fato de este ser classificado como um modelo linear ou não linear. É importante ressaltar que não linearidade, neste caso, não se refere à curvatura da função média, mas, sim, ao espaço dos parâmetros envolvidos. Um modelo é dito linear se o espaço caracterizado por seus parâmetros for linear, ou seja, se a variável dependente aparece associada às variáveis independentes como uma combinação linear destas; caso contrário, é dito não linear. Uma forma prática muito usual de verificar se um conhecido modelo é linear ou não linear consiste em aplicar a este modelo as derivadas parciais relativas a cada parâmetro e observar se, nos resultados obtidos, o parâmetro utilizado como variável na derivação parcial continua presente. Caso isso aconteça com pelo menos um dos parâmetros do modelo, então, o modelo em estudo é não linear. Frequentemente, o efeito das variáveis independentes  $X_i$  na variável resposta  $Y$  é mais adequadamente aproximado por uma função que não é linear nos parâmetros. Como os parâmetros são valores desconhecidos, qualquer formulação de modelos tem seu uso dependente de que estes sejam determinados, pois dependem da amostra utilizada.

### 3.4 Métodos de estimação do vetor de parâmetros

A estimativa dos parâmetros envolvidos nos modelos é feita com a utilização da metodologia chamada de análise de regressão, por meio da qual se obtêm inferências sobre os modelos. Ela é responsável pelo entendimento do relacionamento entre as variáveis explicativas e dependentes, o qual é representado pelo modelo em uma equação e é chamado de modelo de regressão linear simples, se define uma relação linear simples, se a variável resposta é

descrita por uma única variável explicativa em um modelo linear. Mas, se em vez de uma, tivermos incorporadas ao modelo linear várias variáveis independentes, teremos, então, um modelo de regressão linear múltipla. Estas considerações são válidas também para os modelos regressão não lineares, que podem ser simples ou múltiplos, e são ditos não lineares quando a relação entre a variável resposta e as variáveis explicativas não ocorrer por meio de uma combinação linear. Os métodos dos mínimos quadrados e da máxima verossimilhança são os mais utilizados para a estimação do vetor de parâmetros nos estudos de regressão.

#### **3.4.1 O método dos mínimos quadrados**

O método dos mínimos quadrados tem variações conforme as pressuposições assumidas quanto aos erros, podendo ser método dos mínimos quadrados ordinários, mínimos quadrados ponderados ou mínimos quadrados generalizados. Em um modelo de regressão linear simples, o ajuste utilizando o método dos mínimos quadrados ordinários exige as seguintes pressuposições:

- a)  $E[e_i]=0$  a esperança matemática de cada erro é nula, em que  $e_i$  é o erro associado a cada variável dependente, também chamada de variável resposta;
- b) que os erros sejam não correlacionados;
- c) os erros devem ser variáveis aleatórias com distribuição normal.

Com estas pressuposições, segundo Hoffmann e Vieira (1977), obtêm-se estimadores não tendenciosos e de variância mínima. Draper e Smith (1998) afirmam que pode ocorrer que algumas das observações utilizadas sejam menos fidedignas que outras, o que pode acontecer pela existência de situações



indesejáveis nas medições ou em qualquer forma de coleta utilizada na obtenção das mesmas. A ausência de qualquer uma das pressuposições deve ser corrigida para que se possam aplicar técnicas de regressão paramétricas e obter outras estatísticas, que permitirão extrair informações importantes para inferências sobre a população em estudo.

#### **3.4.1.1 Pressupostos da regressão**

A metodologia de regressão é construída sobre os seguintes pressupostos:

- a) normalidade dos erros;
- b) variância residual constante;
- c) independência dos erros.

A ausência de algum destes pressupostos pode invalidar as inferências estatísticas.

##### **3.4.1.1.1 Ausência da normalidade dos erros**

Segundo Neter, Wasserman e Kutner (1990), nos resíduos, a heterocedasticidade, que é a presença de erros com variâncias não constantes, e a não normalidade, frequentemente, aparecem juntas. Pode-se fazer uma transformação na variável resposta, desde que as formas e o espalhamento da distribuição desta variável necessitem ser modificados, considerando o caso particular do modelo de regressão linear simples com a formulação a seguir:

$$Y = \alpha + \beta X + \varepsilon, \quad (3)$$

em que

$Y$  – variável resposta;

$X$  – variável explicativa;

$\alpha$  e  $\beta$  - parâmetros do modelo;

$\varepsilon$  – erro.

É comum a utilização de alguma transformação em  $Y$ , que provoque uma linearização na relação curvilínea entre as variáveis, caso exista. Neter, Wasserman e Kutner (1990) afirmam, ainda, que é frequente o fato de a não normalidade e a heterocedasticidade serem crescentes em seu espalhamento e em sua variabilidade e, nesta situação, as transformações mais pertinentes a serem utilizadas são, geralmente,

$$Y' = \sqrt{Y}, Y' = \log_{10} Y \text{ ou ainda } Y' = \frac{1}{Y}.$$

em que  $Y'$  - representa a variável resposta transformada.

Box e Cox (1964) apresentaram o desenvolvimento de um procedimento para escolher uma transformação em uma família de potência de  $Y$ . Este é um procedimento útil para corrigir o espalhamento da distribuição dos erros e a não linearidade da função em uso. A família de potências é apresentada na forma  $Y^\lambda$ , em que  $\lambda$  é o parâmetro a ser determinado para o conjunto de dados. O critério para a determinação de  $\lambda$  é encontrar aquele que minimiza a soma dos quadrados dos erros para uma regressão linear com estes dados transformados. Observe que, para

$$\lambda = 2,0 \Rightarrow Y' = Y^2$$

$$\lambda = 0,5 \Rightarrow Y' = \sqrt{Y}$$

$$\lambda = 0 \Rightarrow Y' = \log_e Y$$

$$\lambda = -0,5 \Rightarrow Y' = \frac{1}{\sqrt{Y}}$$

$$\lambda = -1 \Rightarrow Y' = \frac{1}{Y}.$$

Na prática, após a transformação, um gráfico dos resíduos deve ser efetuado para verificar se a mesma foi, de fato, apropriada. Desde que a transformação altera a magnitude da soma dos quadrados, é necessário que se ajuste essa soma ou que seja utilizada uma variável padronizada, o que faz com que a soma dos quadrados dos erros se mantenha inalterada pelo valor de  $\lambda$ . Para a segunda abordagem, a variável padronizada empregada é

$$W = \begin{cases} K_1(Y^\lambda - 1) & \lambda \neq 0 \\ K_2(\log_e Y) & \lambda = 0 \end{cases} \quad (4)$$

em que

$$K_2 = \left( \prod_{i=1}^n Y_i \right)^{\frac{1}{n}} \quad \text{média geométrica} \quad (5)$$

$$K_1 = \frac{1}{\lambda K_2^{(\lambda-1)}} \quad (6)$$

#### 3.4.1.2 Ausência da homocedasticidade

A presença da heterocedasticidade, ou seja, a quebra do pressuposto da homocedasticidade, significa, na representação matricial, a existência de valores diferentes na diagonal principal da matriz de variâncias e covariâncias, o que equivale à afirmação da presença de diferentes variâncias. Neste caso, a

estimação pelo método dos mínimos quadrados ordinários não se aplica. Busca-se, então, trocar a variável de observações  $Y$  por outra variável, de forma a satisfazer às suposições que garantam o uso do método dos mínimos quadrados ordinários ou, ainda, trabalhando diretamente na matriz de variâncias e covariâncias com a inserção de alguma ponderação nesta, levando-a a assumir a forma  $\sigma^2(X'X)^{-1}$ . Esta prática equivale a utilizar a metodologia que é chamada de mínimos quadrados ponderados ou, ainda, de forma mais geral, podem-se utilizar os mínimos quadrados generalizados para corrigir a heterocedasticidade.

Pinheiro e Bates (2000) apresentam diversas funções de variância para modelagem da heterocedasticidade. Dentre elas, destacam-se duas classes que são utilizadas neste trabalho. A primeira é representada pelo modelo  $var(\varepsilon_{ij}) = \sigma^2 |v_{ij}|^{\delta}$ , que corresponde à função de variância  $g(v_{ij}, \delta) = |v_{ij}|^{\delta}$ , que é uma potência do valor absoluto da covariante; a segunda classe é representada pela função  $var(\varepsilon_{ij}) = \theta^2 (\delta_1 + |v_{ij}|^{\delta_2})$ , correspondendo à função de variância  $g(v_{ij}, \delta) = \delta_1 + |v_{ij}|^{\delta_2}$ . Em ambos os modelos,  $\delta$  é o vetor dos parâmetros da variância,  $v_{ij}$  é o vetor de covariáveis da variância e  $g(\cdot)$  é a função variância é assumida contínua em  $\delta$ , sendo essa função uma constante mais uma potência do valor absoluto da covariante com  $\delta_1$  restrito (valor constante) e  $\delta_2$  não restrito (pode assumir valores reais distintos). Trabalhos envolvendo modelagem desta estrutura podem ser encontrados em Calegario et al. (2005) e Fang e Bailey (2001).

### 3.4.1.3 Erros correlacionados

Pode, ainda, ocorrer que, em alguns fenômenos observados, os elementos fora da diagonal principal da matriz  $V(\varepsilon)$  de variâncias e covariâncias dos erros não sejam todos nulos, isto é, as observações são correlacionadas.

Amostragens ocorridas repetidamente em tempos distintos adjacentes num mesmo objeto ou organismo estão fortemente propensas a introduzir uma correlação que pode causar restrições à aplicabilidade de alguns métodos estatísticos convencionais que dependem da independência dos erros, ou seja, da não correlação.

Os estudos pertinentes a dados correlacionados no tempo são examinados em análise de séries temporais, podendo acontecer em duas diferentes abordagens: uma que ocorre no domínio do tempo e a outra, no domínio da frequência. A primeira baseia-se no fato de que a correlação entre as observações adjacentes é mais bem explicada em termos de uma regressão de valores presentes sobre valores passados, enquanto a segunda abordagem supõe que a série temporal é mais bem explicada quando considerada como uma soma ou superposição linear de ondas periódicas de senos e cossenos de diferentes períodos ou frequências. Neste trabalho, adotou-se a primeira abordagem.

Como dito anteriormente, a estimativa de parâmetros em modelos de regressão é comumente obtida pelo método dos mínimos quadrados, o qual considera as pressuposições de independência, normalidade e variância constante dos erros. Em caso da ausência de algum destes pressupostos, tornam-se necessárias intervenções. Seja o modelo linear expresso por

$$Y = X\beta + \varepsilon \quad (7)$$

em que

$Y$  é um vetor  $n \times 1$  contendo os valores da variável resposta em estudo;

$X$  é uma matriz  $n \times (k+1)$ , contendo os  $n$  valores observados das variáveis explicativas e  $k$  é o número de parâmetros envolvidos;

$\beta$  é um  $(k+1) \times 1$  vetor de parâmetros;

$\varepsilon$  é um  $n \times 1$  vetor de erros.

Assumindo que o  $\varepsilon \sim N_n(0, \sigma^2 I_n)$ , isto conduz aos mínimos quadrados ordinários (MQO), com o estimador de  $\beta$

$$b_{MQO} = (X'X)^{-1}X'y \quad (8)$$

e com a matriz de variância e covariância dada por

$$V_{(b_{MQO})} = \sigma^2(X'X)^{-1} \quad (9)$$

No entanto, de forma mais geral, deve-se assumir que  $\varepsilon \sim N_n\left(0, \sum_{i=1}^n \sigma_i^2 \varepsilon_i^2\right)$ , em que  $\sum_{i=1}^n \sigma_i^2 \varepsilon_i^2$ , a matriz de covariâncias dos erros, é simétrica e positivamente

definida. Diferentes entradas na diagonal principal de  $\sum_{i=1}^n \sigma_i^2 \varepsilon_i^2$  correspondem a variâncias dos erros não constantes, enquanto valores diferentes de zero fora desta diagonal correspondem a erros correlacionados. Se supusermos  $\sum_{i=1}^n \sigma_i^2 \varepsilon_i^2$  conhecida, a log-verossimilhança para o modelo será dada por

$$\log_{\varepsilon} L(\beta) = -\frac{n}{2} \log_{\varepsilon} 2\pi - \frac{1}{2} \log_{\varepsilon} \left( \det \sum_{i=1}^n \sigma_i^2 \varepsilon_i^2 \right) - \frac{1}{2} (y - X\beta)' \sum_{i=1}^n \sigma_i^{-2} \varepsilon_i^{-2} (y - X\beta),$$

que tem o estimador de  $\beta$  maximizado pelos Mínimos Quadrados Generalizados (MQG) dado por

$$b_{MQG} = \left( X' \sum_{i=1}^n \Omega X \right)^{-1} X' \sum_{i=1}^n \Omega y \quad (10)$$

com matriz de covariância

$$\text{Var}(b_{MQG}) = \left( X' \sum_{i=1}^n \Omega X \right)^{-1} \quad (11)$$

Conforme Fox e Weisberg (2011), o caso mais real é que, para entradas distintas, ocorram diferentes valores em sua diagonal principal, o que corresponde à variância dos erros não constante, além de apresentar valores diferentes de zero fora desta diagonal, ou seja, erros correlacionados. Acrescenta-se que essa matriz é desconhecida com um número elevado de elementos,  $n(n+1)/2$  elementos distintos e, por este motivo, a estimativa do modelo é realizada impondo-se adequadas restrições em sua parametrização, o que é feito com o emprego da metodologia dos mínimos quadrados generalizados (MQG).

Dados coletados como medidas repetidas em tempos distintos são susceptíveis de não serem independentes. Considerando o erro com esperança (zero), variância constante ( $\sigma^2$ ) e com a covariância entre dois erros dependendo apenas de sua separação (d) no tempo, que consideramos igualmente espaçados, tem-se a covariância dada por

$$C(\varepsilon_{t'}, \varepsilon_{t+d}) = C(\varepsilon_{t'}, \varepsilon_{t-d}) = \sigma^2 \rho_d \quad (12)$$

em que  $\rho_d$  é a autocorrelação dos erros no intervalo (lag) d. Nesta situação, a matriz de covariância tem a seguinte estrutura:

$$\Sigma = \sigma^2 \cdot \begin{bmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{n-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{n-2} \\ \rho_2 & \rho_1 & 1 & \dots & \rho_{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{n-1} & \rho_{n-2} & \rho_{n-3} & \dots & 1 \end{bmatrix} = \sigma^2 \cdot P \quad (13)$$

Se os valores de  $\sigma^2$  e de  $\rho_d$  forem conhecidos, podem-se utilizar os mínimos quadrados ponderados para calcular o estimador  $\beta$ , mas isso não é o que realmente acontece. O grande número de  $\rho_d$  torna a estimação impossível sem uma especificação adicional para a estrutura dos erros correlacionados. Existem diversos modelos para séries temporais estacionárias e o mais comum para regressão de erros autocorrelacionados é o processo autorregressivo de primeira ordem, AR(1) :

$$\varepsilon_t = \phi \varepsilon_{t-1} + u_t \quad (14)$$

em que  $u_t$  é assumido ser o ruído branco gaussiano,  $N(0, \sigma_u^2)$ . Sob este modelo,

$$\rho_1 = \theta \text{ e } \rho_d = \theta^d \text{ e } \sigma^2 = \frac{\sigma_u^2}{1 - \phi^2}. \quad (15)$$

Dessa maneira, como  $|\phi| < 1$ , então, os erros autocorrelacionados decaem exponencialmente em direção a 0 (zero), com o aumento de  $d$ . Modelos regressivos de ordens superiores, AR(p), são generalizações dos modelos de primeira ordem; para o de segunda AR(2) ordem, teremos

$$\varepsilon_t = \phi_1 \varepsilon_{t-1} + \phi_2 \varepsilon_{t-2} + u_t. \quad (16)$$



Por outro lado, em processos de médias móveis MA(1), o erro corrente depende diretamente de  $u_t$  corrente e do período anterior.

$$\varepsilon_t = u_t + \psi u_{t-1} \quad (17)$$

Os processos de ordens superiores MA(q) são definidos de maneira similar. Os termos são combinados em processos ARMA(p,q), em que ARMA(1,1) é dado por

$$\varepsilon_t = \phi \varepsilon_{t-1} + u_t + \psi u_{t-1}. \quad (18)$$

### 3.4.2 O método da máxima verossimilhança

Um método também muito utilizado na estimação de parâmetros é a máxima verossimilhança. Mood, Graybill e Boes (1974) definem da seguinte forma uma função de verossimilhança: a função de  $n$  variáveis aleatórias  $X_1, X_2, \dots, X_n$  é uma densidade conjunta de  $n$  variáveis, dita

$$f_{X_1, X_2, \dots, X_n}(X_1, X_2, \dots, X_n; \theta),$$

a qual é considerada uma função de  $\theta$ . Em particular, se  $X_1, X_2, \dots, X_n$  é uma amostra aleatória de uma densidade  $f(x; \theta)$ , então, a função de verossimilhança é

$$f(x_1; \theta) \cdot f(x_2; \theta) \cdot \dots \cdot f(x_n; \theta).$$

Os mesmos autores apresentam a função de verossimilhança como

$$L(\theta) = f(x_1; \theta) \cdot f(x_2; \theta) \cdot \dots \cdot f(x_n; \theta).$$

Muitas vezes, a função de verossimilhança satisfaz às condições de regularidade e o estimador de máxima verossimilhança é a solução da equação

$$L(\theta) = \frac{d(L(\theta))}{d(\theta)} = 0$$

Assim, são determinados os pontos críticos da função de verossimilhança. O método da máxima verossimilhança consiste em, dentre os infinitos valores que  $\hat{\theta}$  pode assumir, tomar aquele que mais provavelmente gerou os dados. A máxima verossimilhança é determinada pelo ponto de máximo da função de verossimilhança. As funções  $L(\theta)$  e  $\log(L(\theta))$ , esta última chamada log-verossimilhança, têm seus máximos no mesmo ponto e, portanto, pode-se utilizar tanto uma quanto a outra para se determinar os esses valores.

### 3.5 Medidas de não linearidade

Conforme Hoffmann e Vieira (1977), em modelos de regressão não linear, apesar da ampla aplicação, apenas assintoticamente o estimador de mínimos quadrados tem as propriedades do mesmo estimador em regressão linear, ou seja, quando o tamanho da amostra tende para infinito, pode-se garantir que as propriedades destes estimadores se aproximam. Para amostras finitas, deve-se assumir que as variáveis respostas  $Y_i$  são normalmente distribuídas em torno de sua média, com variância finita  $\sigma^2$  para todo  $i = 1, 2, \dots, n$ .

Ainda segundo Hoffmann e Vieira (1977), o estimador de parâmetros não é uma combinação linear de  $Y_i$  e, em geral, não é normalmente distribuído, como também é um estimador tendencioso e não se pode garantir que seja de variância mínima. Por estes motivos, as propriedades do estimador de mínimos quadrados nos modelos não lineares são desconhecidas para amostras de tamanho finito.

A partir desses aspectos, a teoria assintótica mostra que o estimador de mínimos quadrados torna-se cada vez menos viesado, cada vez mais

normalmente distribuído e aproxima-se da variância mínima possível, à medida que o tamanho da amostra tende para infinito. No entanto, não existem regras para saber se o tamanho da amostra é grande o suficiente para que sejam válidas as propriedades assintóticas. Bates e Watts (1980) apresentaram as medidas de não linearidade intrínseca e não linearidade devido ao efeito dos parâmetros que permitem avaliar se o modelo e amostra em uso atendem às condições para aplicação dos mínimos quadrados. Estas são as chamadas medidas de curvatura.

### 3.5.1 Representação geométrica da não linearidade

Ratkowsky (1983) apresenta o seguinte exemplo, para a representação geométrica dos modelos de regressão não lineares no espaço amostral. Seja o modelo representado da forma

$$Y = f(\theta) + e \quad (19)$$

em que

$$Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \quad f(\theta) = \begin{bmatrix} f(x_1, \theta) \\ \vdots \\ f(x_n, \theta) \end{bmatrix}, \quad e = \begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix}, \quad \theta = \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_n \end{bmatrix}$$

$$X_t = [X_{t1}, \dots, X_{tk}], \quad t = 1 \dots n.$$

$$\text{com } E[e] = 0 \text{ e } \text{var}(e) = \sigma^2 I.$$

Inicialmente, é considerado o caso  $n=2$  e  $p=1$ , com  $n$  representando o número de observações e  $p$ , o número de parâmetros. Como exposto em

Ratkowsky (1983), iniciaremos pelo estudo do caso linear, que propiciará maior clareza para o caso não linear.

**a) Caso linear**

As componentes do modelo são representadas como segue:

$$Y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, \quad f(\theta) = x\theta \quad \text{com } x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \text{ e } \theta \in \mathbb{R}.$$

Temos, neste primeiro exemplo,  $n=2$  representando a dimensão do espaço amostral e  $p=1$ , indicando que a dimensão do espaço de estimação, ou dimensão do locus de solução, aqui igual a 1, constituída pelos pontos  $\begin{bmatrix} x_1\theta \\ x_2\theta \end{bmatrix}$  que representam uma reta cuja direção é a do vetor  $X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ . Deve-se observar que, no caso linear, tomando valores igualmente espaçados de  $\theta$  no espaço paramétrico, então, os pontos correspondentes  $x_i\theta$  no espaço solução serão também igualmente espaçados, conforme apresentado na Figura 3.

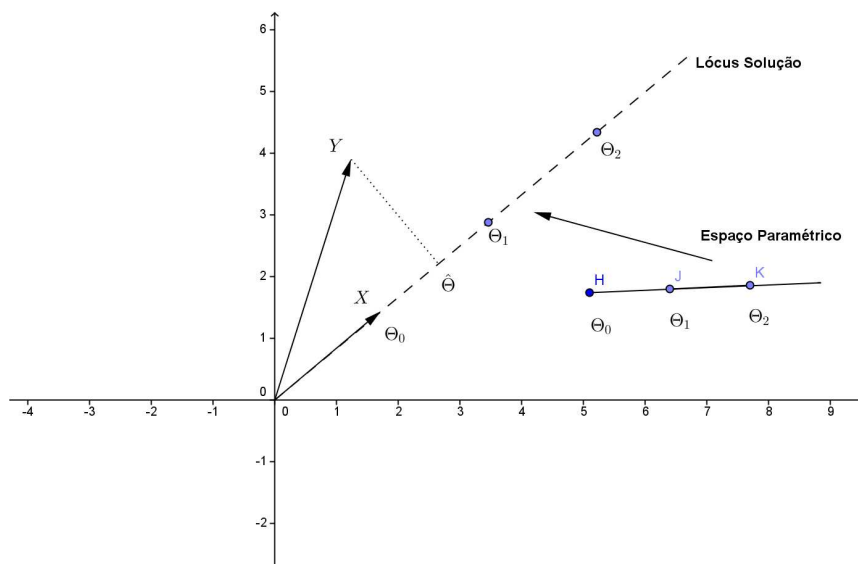


Figura 3 Curvatura do espaço de estimação no espaço amostral no caso linear  
Fonte: Ratkoswsk (1983)

Os eixos de abscissa e ordenada equivalem, respectivamente, à coordenada 1 e à coordenada 2 e são tomados como componentes da base do espaço amostral. Temos, ainda,  $Y = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ ,  $X = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ ,  $\theta_0=1$ ,  $\theta_1=2$ ,  $\theta_2=3, \dots$

A função

$$h(\theta) = \sum_{i=1}^2 (y_i - \theta x_i)^2 \quad (20)$$

representa a distância de um ponto genérico  $\theta x$  ao valor observado  $Y$ . Assim, minimizar esta função em relação a  $\theta$  representa determinar o ponto no locus solução, estimador de mínimos quadrados  $\theta$  aqui chamado de  $\hat{\theta}$ , que está mais próximo do valor  $y$  observado, ou seja, determinar  $\theta$  tal que a distância de  $\theta x_i$  ao ponto observado  $y_i$  seja a mínima possível.

### b) Caso não linear

Considerando o mesmo exemplo de  $n=2$  e  $p=1$ , tem-se que o espaço solução agora não é dado por uma reta, mas uma curva gerada pelos pontos da forma

$$f(\theta) = [f(x_1, \theta) \quad f(x_2, \theta)]$$

Para estas funções, não mais teremos espaçamentos iguais no lócus solução correspondentes a  $\theta$  igualmente espaçados conforme a Figura 4, na qual está representado no espaço amostral o modelo

$$Y = f(\theta) + e \quad (21)$$

em que

$$Y = [2.5 \quad 10] , \quad x = [2 \quad 3] \quad \text{com } f(\theta) = [f(x_1, \theta) \quad f(x_2, \theta)] = [x_1 \cdot 1^{\theta} \quad x_2 \cdot 2^{\theta}] = [2^{\theta} \quad 3^{\theta}]$$

com  $\theta \in \mathbf{R}$ . Observa-se, na Figura 4, que, para valores de  $\theta$  variando de 0,5 em 0,5, igualmente espaçados, os espaçamentos no lócus solução diferenciam-se entre si.

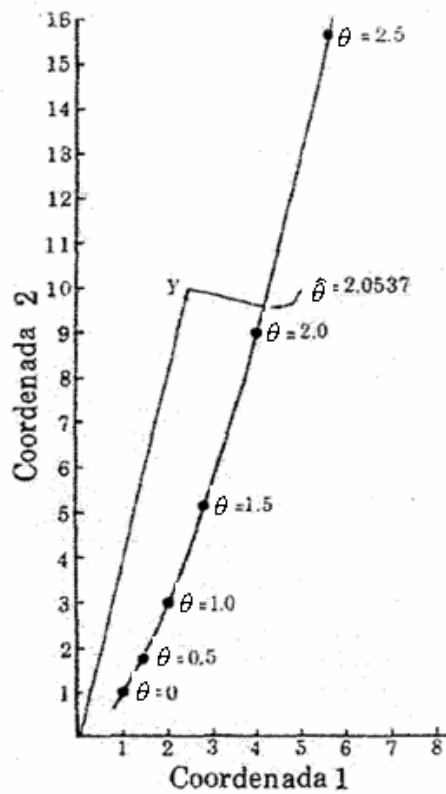


Figura 4 Curvatura do espaço de estimação no espaço amostral para o caso não linear

Fonte: Ratkowsky (1983)

Bates e Watts (1980) observaram que a curvatura existente do modelo em relação a um conjunto de dados pode ser separada em duas componentes chamadas de curvatura intrínseca e curvatura devido à parametrização. Prudente (2009) afirma que, para um modelo normal não linear expresso na forma clássica por

$$Y = g(\theta; X) + \varepsilon = \mu(\theta) + \varepsilon \quad (21),$$

em que os  $\varepsilon_{it}$ s têm distribuição normal com média zero e variância constante  $\sigma^2$ ,  $g(\theta; X) = \mu(\theta)$  é uma função diferenciável em  $\theta$ ,  $\theta = (\theta_1, \dots, \theta_p)$  contém os  $p$  parâmetros desconhecidos a serem estimados e  $X = (x^{(1)}, \dots, x^{(n)})^T$  representa a matriz de dimensões  $n \times s$  dos valores das  $s$  variáveis explicativas. Neste caso, a equação de uma reta no espaço paramétrico passando por  $\hat{\theta}$  pode ser expressa por

$$\theta(b) = \hat{\theta} + bh \quad (22)$$

com  $b$  escalar e  $h = (h_1, \dots, h_p)^T$  é um vetor de valores não nulos. A aplicação dos pontos desta reta no espaço paramétrico gera a curva

$$\mu(b) = \mu(\hat{\theta} + bh) \quad (23)$$

Com a tangente no ponto  $b=0$  dada pelo produto da matriz jacobiana  $\tilde{X}$  da transformação de  $\mu(\theta)$  em  $\theta = \hat{\theta}$ , dada por

$$\dot{\mu}_h = \tilde{X}h$$

tal que o plano tangente em  $\mu(\hat{\theta})$  é dado pelo conjunto de todas as combinações lineares deste tipo e a aceleração da curva  $\ddot{\mu}_h$  é estimada por

$$\ddot{\mu}_h = h^T \hat{W} h ,$$



em que  $w$  é uma matriz de dimensão  $n \times (p \times p)$  com a  $i$ -ésima face expressa como

$$w_i = \left( \frac{\partial^2 \mu_i}{\partial \theta_r \partial \theta_s} \right)$$

$$i=1 \dots n \quad e \quad r, s=1, \dots, p. \quad (24)$$

O vetor aceleração pode ser decomposto em três componentes: O componente  $\ddot{\mu}^{IN}$  que, representando uma derivada segunda, determina a variação na direção do vetor da velocidade instantânea  $\dot{\mu}_h$  e o segundo e o terceiro componentes, cuja normal será dada por  $\ddot{\mu}^{PE}$ , que determinam a variação paralela ao plano tangente e a variação no ponto móvel.

### 3.6 Estimativa da altura total e o uso de covariantes

As medições da altura de árvores em pé podem ocorrer de forma direta ou indireta. Quando ocorrem da forma direta, essas medidas são efetuadas com uso de hastes extensíveis e, quando medidas indiretamente, utilizam-se, mais comumente, instrumentos chamados de hipsômetros. Contudo, o transporte de hastes extensíveis na floresta é quase sempre muito difícil e a visualização das diversas alturas depende extremamente da habilidade de quem realiza as medições, tornando-se, assim, motivo para aumento no erro para a obtenção dessa variável. Outra opção é a derrubada de um número de árvores que represente suficiência amostral. Em seguida, essas árvores são medidas e essas medidas utilizadas para estimar as demais.

Diversos são os estudos existentes para se estimar a altura total, os quais, em geral, são realizados com o uso de modelos lineares ou linearizados. Calegário et al. (2005) descrevem que a evolução dos computadores e dos softwares minimizou as dificuldades do uso de modelos não lineares e apresenta

uma alternativa utilizando modelos de efeitos mistos não lineares para ajuste de altura. Estes modelos diferenciam-se da ANOVA, que consideram apenas os efeitos fixos, pois levam em consideração também os efeitos aleatórios.

Calegário et al. (2004) apresentam a evolução da abordagem de modelos lineares de efeitos mistos aplicados ao crescimento e à produtividade clonal de *Eucalyptus*. Nos modelos de efeito misto estão separados os efeitos fixos e os efeitos aleatórios; enquanto os efeitos fixos estão relacionados ao tratamento e ao efeito de interesse, os efeitos aleatórios relacionam-se ao efeito na variância da resposta. A incorporação de covariáveis ao modelo elimina a necessidade de alguns ou de todos os efeitos aleatórios.

Guimarães et al. (2009) afirmam que a dificuldade da modelagem da relação hipsométrica é o grande número de variáveis que a influenciam e afirmam, ainda, que técnicas de decomposição dos parâmetros do modelo, com a inclusão de covariantes relacionadas com árvores individuais ou com o povoamento, podem melhorar a precisão do mesmo.

### **3.7 Estimação do volume de povoamentos florestais**

O volume em povoamentos florestais é apresentado por Gomes (1957) e sua estimativa é dividida em dois grupos, que são:

- a) o que estima a partir de árvores individuais, conhecido como o método da tabela de volume;
- b) o que estima o volume tomando a unidade de superfície como base por meio das chamadas tabelas de produtividade.

Enquanto o primeiro grupo trabalha com o conhecimento de situações do presente, o segundo está mais diretamente associado a previsões das

disponibilidades futuras ou incremento proporcionado, possibilitando também considerações de caráter econômico. Para tal, se faz necessário conhecer a capacidade produtiva da espécie por meio do índice de local.

### **3.7.1 Método da tabela de volume**

Cotta (1804) enunciou o seguinte postulado: o volume de uma árvore depende do seu diâmetro, altura e forma. Quando o volume de uma árvore for determinado corretamente, o valor achado é válido para outra árvore de igual diâmetro, altura e forma.

Segundo Mackay (1964), este postulado foi o ponto de partida para os métodos existentes de determinação do volume de árvores em pé. O mesmo autor afirma que, para indivíduos de uma mesma espécie, vegetando em iguais condições de clima e solo, sujeitos aos mesmos regimes, pertencentes às mesmas classes de diâmetro e altura, pode-se admitir que estas árvores tenham a mesma forma e, conseqüentemente, o mesmo volume, relacionando o volume de uma árvore diretamente com sua altura e diâmetro, possibilitando seu conhecimento por meio de técnicas como métodos gráficos e analiticamente por técnicas de regressão

Com o desenvolvimento de softwares e computadores, o método dos mínimos quadrados superou totalmente os demais métodos de construção de tabelas volumétricas, segundo Campos e Kronka (1973).

### **3.7.2 Método da tabela de produtividade**

Tabelas de produtividade podem ser definidas como uma representação tabular do volume por unidade de área para povoamentos equiâneos,

classificadas de acordo com a idade, o índice local e a densidade. Os europeus, particularmente os alemães, foram precursores no uso de tabelas de produtividade, no final do século XVIII; no final do século XIX, foram realizados trabalhos nos Estados Unidos da América. Inicialmente, utilizavam-se as tabelas de produtividade normal baseadas somente na idade e na qualidade de local. Eram feitas por meio de métodos gráficos e, posteriormente, com o uso de regressão segundo Moser (1980). Mackinney, Schumacher e Chaiken (1937) foram os primeiros a utilizar a técnica dos mínimos quadrados aplicada a modelos de regressão; eles construíram a primeira tabela de produtividade para uma população não normal de *Pinus taeda*. Schumacher (1939) também demonstrou a superioridade do método dos mínimos quadrados sobre os métodos gráficos.

Até 1962, os modelos de crescimento e produtividade não consideravam a compatibilidade, ou seja, as funções de predição do volume ou área basal em idades estabelecidas e as funções de produtividade em intervalos de tempo eram consideradas independentes e, dessa forma, em geral, não representavam valores iguais de volume ou área basal, quando eram acrescentados sucessivos incrementos de produtividade a uma idade inicial até uma segunda idade estimada pelo modelo de crescimento.

Clutter (1963) construiu um modelo de crescimento compatível com o modelo de produtividade que satisfaz às pressuposições do uso de regressão. As equações do crescimento foram obtidas por diferenciação das taxas de variação do volume no tempo.

### **3.8 Modelos e funções de crescimento**

Existe uma relação funcional entre duas variáveis quando a variável dita dependente pode ser determinada por algum tipo de associação com a outra, a

variável independente (que podem ser diversas), por meio de uma lei chamada de função. Seja, por exemplo,  $W$  uma variável qualquer, cujo crescimento pode ser expresso em função do tempo  $t$ . Então,

$$W=f(t) \quad (25)$$

em que  $f$  é a relação funcional.

Segundo Thornley e France (2004), é conveniente, e também tradicional, a aproximação das funções de crescimento acontecerem primeiramente pela discussão das taxas de crescimento  $dW/dt$ , ou seja, a razão da variação de  $W$  pela variação do tempo, de onde tem-se:

$$\frac{dW}{dt} = g(t), \quad \text{em que } g(t) = \frac{df(t)}{dt} \quad (26)$$

Eliminando-se a variável tempo entre as equações 25 e 26, tem-se que

$$\frac{dW}{dt} = h(W) \quad (27)$$

em que  $h$  é uma função de  $W$ . Dessa maneira, considerando que se está trabalhando com taxas de variações de organismos, a equação 27 pode ser interpretada biologicamente.

Funções de crescimento têm tido enorme desenvolvimento e novos modelos são originados de modelos propostos, tal como a generalização de Chapman-Richards (CHAPMAN, 1961; RICHARDS, 1959) para o modelo de Bertalanffy (BERTALANFFY, 1957) discutida por Pienaar e Turnbull (1973) e, ainda, em trabalhos como de Prodan (1968).



potencial para a produtividade de madeira de determinado lugar, para determinada espécie ou clone”. Os mesmos autores afirmam, ainda, que essa qualidade pode ser expressa, de modo empírico, em alta, média e baixa, ou por meio de algum índice quantitativo, como o índice de local, o qual é a altura dominante média do povoamento em uma idade específica.

Para a condução de povoamentos e no planejamento da produtividade madeireira de uma empresa florestal, o conhecimento da capacidade produtiva é essencial. A construção de curvas de índice de sítio torna-se uma muito útil para orientar o inventário florestal, a construção de tabelas de volume, as previsões de produtividade, os estudos de variedades e espécies e os estudos de práticas adequadas de manejo florestal. Fatores como solo, topografia, competição com outros vegetais e animais, fatores genéticos e climáticos influenciam uma árvore. A interação destes sobre a planta expressa a qualidade do sítio, segundo Husch, Miller e Beers (1972).

Nas ciências florestais, é comum aplicar o termo "sítio" para designar a influência do ambiente sobre a produtividade florestal. Qualidade do sítio é definida para cada espécie, pois o sítio que pode ser bom para uma pode não ser para outra. O índice de sítio é uma variável que procura quantificar a qualidade do sítio por meio de parâmetros dendrométricos (volume de madeira, altura das árvores, área basal, etc.). Segundo Spurr (1951), a qualidade absoluta do sítio é teoricamente medida pela quantidade de madeira produzida. Contudo, é condicionada por fatores genéticos, idade de rotação, densidade de povoamento, etc. e seu uso como medida do sítio só deve ocorrer quando estes fatores forem constantes, o que é impraticável, tornando questionável o uso de volume de madeira como índice de sítio.

Veiga (1976, 1980) define que a área basal não é um bom índice de sítio, pois é dependente do espaçamento, além de sensível às condições climáticas, fisiográficas e edáficas. Por outro lado, a área basal é uma boa representante do

potencial produtivo, pois tende à estagnação, uma vez que a área basal máxima é fixa na chamada “idade de estagnação”. Pela facilidade de medição torna-se vantajosa, no entanto, pelo fato de, para mesma área basal, poder acontecer diferentes produções volumétricas, torna-se pouco recomendável.

Em geral, o índice de sítio é representado pela altura média ou máxima de um povoamento numa dada idade. Segundo Spurr (1951), a altura das árvores é o melhor índice para representar o sítio. O índice de sítio é quase invariavelmente apresentado como a altura das árvores dominantes e codominantes de um povoamento numa determinada idade, a qual é chamada de idade de referência ou idade base.

Como solução para problemas apresentados na medição das alturas, Spurr (1951) propõe o conceito atualmente mais utilizado e aceito, o qual define o índice de sítio como a média aritmética da altura das 100 árvores de maior diâmetro por hectare. As curvas de índice de sítio descrevem o crescimento em altura das árvores hipotéticas de índices de sítios específicos. O valor do índice de sítio de um povoamento é, então, estimado por interpolação entre duas curvas.

As curvas de índice de sítio podem ser classificadas em dois tipos: anamórficas e polimórficas. Conforme Spurr (1951), as curvas anamórficas são construídas a partir de dados de parcelas temporárias, ajustando-se uma curva guia da altura em função da idade. Adiciona-se, então, uma série de curvas similares acima e abaixo da curva guia, as quais têm a mesma forma desta, diferindo em magnitude apenas por uma porcentagem fixa. Machado (1980) afirma que Osborne e Schumacher, analisando o coeficiente de variação de tais curvas, concluíram que assumiam valores diferentes em idades distintas. Esses dois autores desenvolveram um método gráfico para construir curvas polimórficas aplicando diferentes pesos a diferentes classes de idade, conforme os respectivos coeficientes de variação.



Os diversos modelos encontrados na literatura para a construção de curvas de índice de sítio podem ser divididos em dois grupos: modelos lineares ou linearizáveis, que podem ser ajustados pelos métodos de regressão linear simples ou múltipla e os modelos não lineares, que são ajustados por algum dos métodos de regressão não linear.

### 3.10 Critério de informação de Akaike

Mazerolle (2004) define a seleção de modelos como a tarefa um modelo estatístico de um conjunto de modelos. Segundo Emiliano et al. (2009), utilizar modelos implica em inevitável perda de informação. Dentre muitas estatísticas utilizadas para quantificar a perda de informação está a conhecida como informação de Kullback-Leibler, que representa esta perda devido ao modelo ajustado a partir da realidade.

A informação de Kullback-Leibler é dada por

$$I(g;f) = E_g \left[ \log \left( \frac{g(X)}{f(X)} \right) \right] = \int_{-\infty}^{+\infty} g(\infty) \log \left( \frac{g(\infty)}{f(\infty|B)} \right) d\infty \quad (28)$$

em que  $f$  e  $g$  são funções de distribuição e a equação 12 pode ser escrita como

$$I(g;f) = \int_{-\infty}^{+\infty} g(\infty) \log[g(\infty)] d\infty - \int_{-\infty}^{+\infty} g(\infty) \log[f(\infty|B)] d\infty \quad (29)$$

Dessa forma, para os dois modelos,  $f_1(\infty|B)$  e  $f_2(\infty|B)$ , é possível obter, da equação 12, que

$$I(g; f_2) = \int_{-\infty}^{+\infty} g(x) \log[g(x)] dx - \int_{-\infty}^{+\infty} g(x) \log[f_2(x|\theta)] dx \quad (30)$$

e

$$I(g; f_2) = \int_{-\infty}^{+\infty} g(x) \log[g(x)] dx - \int_{-\infty}^{+\infty} g(x) \log[f_2(x|\theta)] dx \quad (31)$$

Somente o segundo termo da equação 28 é importante na avaliação do modelo estatístico  $f_1(\square)$  e  $f_2(\square)$ , desde que o primeiro dependa apenas do modelo verdadeiro  $g$ . Contudo, o segundo termo depende também de uma distribuição  $g$  desconhecida, sendo representado por

$$E_g[\log f(X)] = \int g(x) \log[f(x|\theta)] dx \quad (32)$$

Akaike (1974) propôs usar a informação de Kullback-Leibler para selecionar modelos, estabelecendo uma relação entre a máxima verossimilhança e esta informação, e desenvolveu um critério para estimá-la, o qual foi chamado, mais tarde, de critério de informação de Akaike (1974).

### 3.11 Coeficiente de determinação ajustado - $R^2_{aj}$

O coeficiente  $R^2_{aj}$  ajustado para o número de parâmetros da regressão é dado por

$$R^2 = 1 - \frac{SQR}{SQT} \quad (33)$$

$$R_{aj}^2 = 1 - \left[ \frac{(n-i)(1-R^2)}{n-p} \right] \quad (34)$$

em que

SQR- soma de quadrados dos resíduos;

SQT - soma de quadrado total;

n – número de observações utilizadas para ajustar a curva;

p – número de parâmetros da função;

i – ajuste do intercepto: 1 se houver intercepto e 0 se não houver intercepto na função.

### 3.12 Erro de predição médio – EPM

A estimativa do EPM é obtida calculando-se a média de todos os erros de predição (EP), a partir de cada observação, como um desvio entre altura observada ( $h_o$ ) e altura estimada ( $h_e$ ), dividido pela altura observada e multiplicado por 100, também chamado de erro de predição percentual.

$$EP = 100 \left[ \frac{h_o - h_e}{h_o} \right]$$

Este valor leva um sinal, designando se a função subestimou (+) ou superestimou (-) o peso observado.

## 4 MATERIAIS E MÉTODOS

### 4.1 Coleta de dados

Os dados são provenientes do inventário de *Eucalyptus* sp. com idades entre 2 e 12 anos, na cidade de Luminárias, estado de Minas Gerais, latitude de 21,274962° S e longitude de 44,584422° W, entre 880 e 1.001 m de elevação. As temperaturas médias anuais variam entre 14 °C e 26 °C, precipitação média anual de 1.385 mm e clima tipo Cwa (VEIGA, 1975).

Os dados foram obtidos de parcelas permanentes, medidas anualmente no período de 1999 a 2009, para a obtenção da variável diâmetro à altura do peito (DAP), que é o diâmetro do tronco da árvore medido a 1,30 m do solo, em todas as árvores das parcelas selecionadas. Foi realizada a cubagem rigorosa de 1.781 árvores que tiveram sua altura total medida e, para as demais, a altura foi estimada com emprego do modelo não linear logístico de três parâmetros, comparado ao modelo proposto, que foram ajustados com e sem ponderação da matriz de variâncias e covariâncias.

As autocorrelações existentes foram detectadas com o uso da função de autocorrelação parcial e, em seguida, os três modelos que apresentaram melhor ajuste, um em cada classe, foram modificados e reajustados com uso de covariáveis. Todos os modelos não lineares utilizados foram testados quanto à propriedade da aproximação linear por meio das medidas de não linearidade, propostas por Bates e Watts (1980), não linearidade intrínseca e não linearidade quanto ao efeito dos parâmetros. Foi realizada, ainda, a medição do viés de Box por meio de simulação Monte Carlo, conforme apresentado na Figura 5.

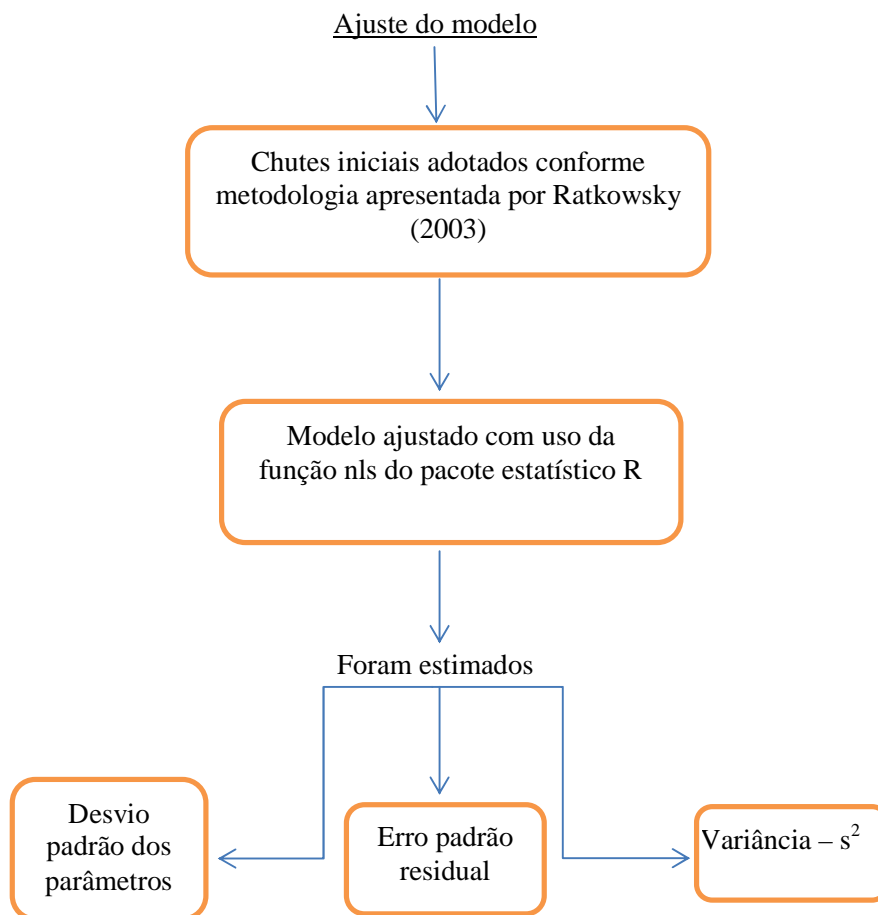
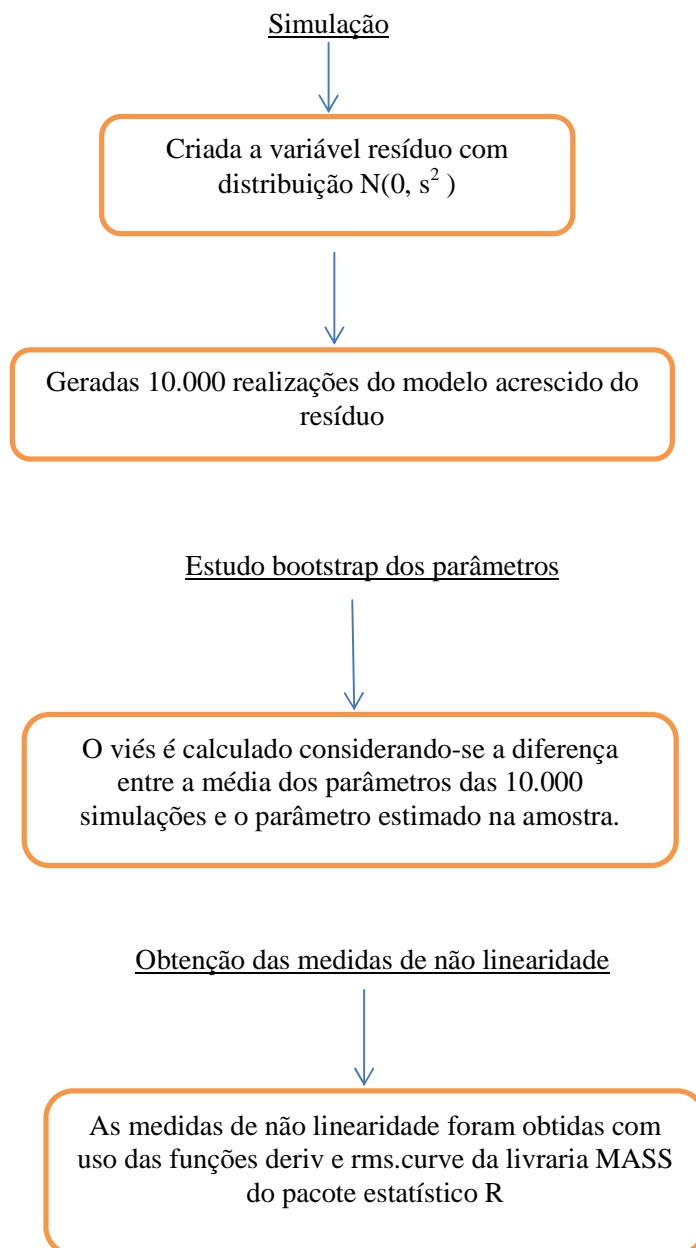


Figura 5 Fluxograma para estudo bootstrap do viés e das medidas de não linearidade

(...continua...)



## 4.2 Índices e curvas de sítio

Os índices de sítio foram estabelecidos pelo método da curva guia seguindo o modelo sigmoidal não linear logístico de três parâmetros, conforme modelo 1.

$$HT_i = \frac{a}{1 + b \exp(-c * idade)} + \varepsilon_i \quad \text{modelo 1}$$

em que

$HT_i$ : altura total da  $i$ -ésima árvore;

Idade <sub>$i$</sub> : idade da  $i$ -ésima árvore no dia da medida;  $a$ ,  $b$  e  $c$ : parâmetros a serem estimados por regressão para o modelo logístico;

$\varepsilon_i$ : erro estocástico com distribuição normal, média zero e variância constante.

### 4.2.1 Construção das curvas de índice local pelo método da curva guia

Na idade-índice, ou idade de referência, a altura dominante é igual ao índice local

$$HT = S \Leftrightarrow I = I_{ref}$$

em que

$HT$  - altura dominante;

$S$  - índice de sítio;

$I_i$  - idade do indivíduo  $i$  de altura  $HT_i$  no momento da medição;

$I_{ref}$  - idade tomada como referência para a construção da curva guia (neste trabalho, adotou-se a idade de 6 anos).

Seja, então,

$$HT_i = \frac{a}{1 + b \exp(-c * I_{ref})} \quad (35)$$

$$S = \frac{a}{1 + b \exp(-c * I_{ref})} \quad (36)$$

Isolando-se o parâmetro  $a$  na equação 34,

$$a = HT_i * (1 + b \exp(-c * I_i)) \quad (37)$$

Substituindo  $a$  da equação 36 na equação 35, tem-se:

$$S = \frac{HT_i * (1 + b \exp(-c * I_i))}{1 + b \exp(-c * I_{ref})} \quad (38)$$

Isolando  $HT_i$  da equação 37 obtém-se a expressão

$$HT_i = \frac{S * (1 + b \exp(-c * I_{ref}))}{1 + b \exp(-c * I_i)} \quad (39)$$

que permite construir as curvas de sítio, do índice de sítio e dos parâmetros do modelo 1, o qual foi ajustado conforme Tabela 2.



Tabela 2 Parâmetros de ajuste da curva média para a construção das curvas de limitantes das classes de sítio

Parâmetros	Estimativas	Erro padrão	Estatística t	Pr(> t )
a	27,543	0,154116	197,97	< 2e-16
b	23,9054	0,269386	26,15	< 2e-16
c	18,0401	0,003237	39,35	< 2e-16

Foram adotadas três classes de sítio, representadas por curvas anamórficas, tomando como idade de referência os seis anos e os índices de sítio para as distintas classes foram estabelecidos tomando a maior e a menor alturas na idade de referência e dividindo-se em três intervalos, conforme Tabela 3.

Tabela 3 Intervalos das classes de sítio tomadas à idade de referência de 6 anos

Classe	Intervalo (m)
I	23,4 --27,6
II	19,2  --23,4
III	15,0  --19,2

### 4.3 Modelos para estimativa da altura total

Na literatura, são encontrados diversos modelos que expressam a altura total das árvores. Para este trabalho, os dados foram ajustados por um modelo proposto e o modelo logístico de três parâmetros apresentado por Pinheiro e Bates (2000), aplicado de acordo representado pela equação 39, que foi ajustado com e sem o acréscimo de covariantes.

Os modelos com o uso de covariantes aqui foram chamados de modelos completos, enquanto os modelos aos quais não se acrescentaram estas variáveis auxiliares foram chamados de modelo reduzidos. Em seguida, o modelo que

apresentou melhor ajuste foi selecionado conforme os critérios já mencionados. Inicialmente, foram determinadas as medidas de não linearidade de Bates e Watts (1980). O modelo proposto neste trabalho é uma variação do modelo exponencial e é representado na equação 40. Este modelo, juntamente com o modelo da equação 39, foi utilizado para ajustar a altura.

Modelos reduzidos:

$$HT_i = \frac{a}{1 + b \exp(-c * Idade)} + \varepsilon_i \quad (40)$$

$$HT_i = \theta_1 + \theta_3^{(-idade)} e^{(\theta_2 + \theta_3^{(-idade)})} + \varepsilon_i \quad (41)$$

em que

$HT_i$ : altura total da i-ésima árvore;

Idade: idade da i-ésima árvore no dia da medida;

$\theta_1, \theta_2, \theta_3$ : parâmetros a serem estimados por regressão para o modelo proposto;

a, b, c: parâmetros a serem estimados por regressão para o modelo logístico;

$\varepsilon_i$ : erro estocástico com distribuição normal, média zero e variância constante.

Modelos completos:

$$HT_i = (\theta_1 + \theta_4 * dap_i) - \theta_2^{(-idade)} * e^{(\theta_2 + \theta_3^{idade})} + \varepsilon_i$$

$$HT_i = \frac{\theta_1 + \theta_4 * dap_i}{1 + \theta_2 * \exp(-\theta_2 * idade_i)} + \varepsilon_i$$

As variáveis e os parâmetros dos modelos têm os mesmos significados que nos modelos reduzidos ajustados sem o uso da decomposição de parâmetros, com  $\theta_1 + \theta_4 = \text{dap}_i$  representando a decomposição do parâmetro  $\theta_1$  sob a influência do DAP.

#### 4.5 Avaliação dos modelos

Os modelos foram testados na estimativa do crescimento em altura e em volume de povoamentos de *Eucalyptus* sp., de acordo com os critérios descritos a seguir.

##### 4.4.1 Medidas de não linearidade ou de curvatura

Conforme desenvolvimento e notação apresentados por Souza (2008), as medidas de curvatura propostas por Bates e Watts (1980) estão diretamente associadas à aproximação pela série de Taylor e, com o uso desta, pode-se escrever, numa vizinhança de  $\tilde{\theta}$ , que

$$\begin{aligned} f(x_t, \theta) &= f(x_t, \tilde{\theta}) + \frac{\partial f(x_t, \tilde{\theta})}{\partial \theta} (\theta - \tilde{\theta}) + \frac{1}{2} (\theta - \tilde{\theta})' (\partial^2 f(x_t, \tilde{\theta})) (\theta - \tilde{\theta}) \\ &= f(x_t, \tilde{\theta}) + \frac{\partial f(x_t, \tilde{\theta})}{\partial \theta} \delta + \frac{1}{2} \delta' (\partial^2 f(x_t, \tilde{\theta})) \delta \end{aligned}$$

Com  $\delta = (\theta - \tilde{\theta})$

Seja, ainda,

$$H = \begin{pmatrix} H_1 \\ \vdots \\ H_n \end{pmatrix}$$

De dimensão  $n \times p$ , em que o bloco

$$H_i = \frac{\partial^2 f(x_i, \theta)}{\partial \theta' \partial \theta}$$

tem dimensão  $p \times p$ . Daí segue que

$$f(\theta) = f(\theta^*) + F(\theta^*)\delta + 1/2 \delta' H \delta$$

em que, por definição,

$$\delta' H \delta = \begin{pmatrix} \delta' H_1 \delta \\ \vdots \\ \delta' H_n \delta \end{pmatrix}$$

é um vetor  $n \times 1$ .

Bates e Watts (1980) utilizam duas componentes ortogonais para comparar o tamanho do vetor  $\delta' H \delta$  relativamente ao termo linear  $F(\theta^*)\delta$ , a projeção sobre o plano tangente e a normal ao plano tangente, no intuito de medir o grau de linearização de  $f(\theta)$ . Seja  $P_F(\theta^*)$  a matriz da projeção sobre  $C(F(\theta^*))$  ao longo de seu ortogonal. Pode-se escrever

$$\delta' H \delta = C^T + C^N$$

com

$$C^T = P_{F(\hat{\theta})} \delta' H \delta \quad \text{e} \quad C^N = (I - P_{F(\hat{\theta})}) \delta' H \delta$$

O vetor  $C^T$  é chamado componente tangencial e  $C^N$  componente normal de  $\delta' H \delta$ . Daí são definidas as medidas de curvatura:

$$K^T(\hat{\theta}) = \frac{\|C^T\|}{\|F(\hat{\theta})\delta\|^2} \quad \text{- curvatura tangencial ou paramétrica}$$

$$K^N(\hat{\theta}) = \frac{\|C^N\|}{\|F(\hat{\theta})\delta\|^2} \quad \text{- curvatura normal ou intrínseca.}$$

Souza (2008) afirma também que as quantidades  $K^T(\hat{\theta})$  e  $K^N(\hat{\theta})$  devem ser uniformemente pequenas para que  $f(\theta)$  seja linearizável.

#### 4.4.2 Critério de informação de Akaike

$$AIC = -2 \ln(mv) + 2p \quad (42)$$

em que

$p$  = número de parâmetros do modelo;

$mv$  = valor da máxima verossimilhança;

$n$  = número de observações.

Quanto menor o valor de AIC melhor o modelo avaliado.

#### 4.4.3 Coeficiente de determinação ajustado - $R^2_{aj}$ .

O coeficiente  $R^2_{aj}$  ajustado para o número de parâmetros da regressão é dado por

$$R^2 = 1 - \frac{SQR}{SQT}$$

$$R^2_{aj} = 1 - \left[ \frac{(n - i)(1 - R^2)}{(n - p)} \right] \quad (43)$$

em que

SQR - soma de quadrados dos resíduos;

SQT – soma de quadrado total;

n – número de observações utilizadas para ajustar a curva;

p – número de parâmetros da função;

i – ajuste do intercepto: 1 se houver intercepto e 0 se não houver intercepto na função.

#### 4.4.4 Erro de predição médio (EPM)

A estimativa do EPM é obtida calculando-se a média de todos os erros de predição (EP), a partir de cada observação, como um desvio entre altura observada ( $h_o$ ) e altura estimada ( $h_e$ ), dividido pelo peso observado e multiplicado por 100, também chamado de erro de predição percentual.

$$EP = 100 \left[ \frac{h_o - h_e}{h_o} \right]$$

Este valor leva um sinal, designando se a função subestimou (+) ou superestimou (-) a altura observada.

#### 4.4.5 Erro relativo percentual

O erro relativo percentual (ERP) é dado por

$$ERP = \frac{EPR}{\frac{\sum_{i=1}^n Hd_i}{n}} * 100$$

em que ERP é o erro padrão residual

$$ERP = \sqrt{\frac{\sum_{i=1}^n (Hd_i - \widehat{H\bar{a}}_i)^2}{n - p}}$$

## 5 RESULTADOS E DISCUSSÃO

Nas Figuras 6 e 7 estão plotadas as alturas observadas, distribuídas em três classes de sítio. Na Figura 5 está representada toda a base de dados e, na Figura 6, estão separadas por classe. É possível observar visualmente, em todas as classes, um aumento da variabilidade destas alturas ao longo do tempo, o que

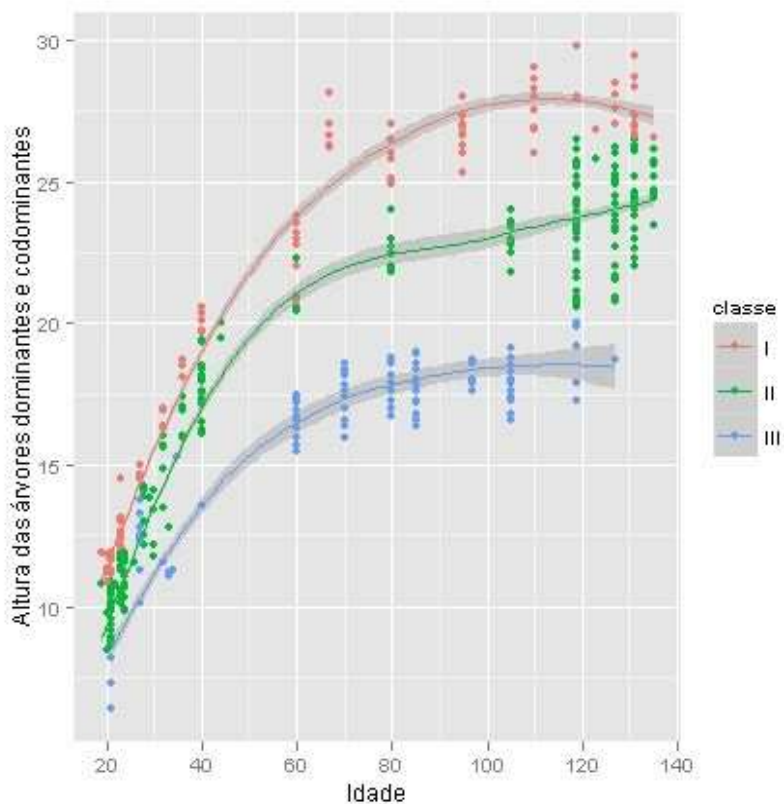


Figura 6 Dados das alturas observadas ao longo do tempo, representadas por classes de sítio



indica a presença de heterocedasticidade e, segundo Mazzini (2001), deve-se considerar, ainda, que, no caso de dados longitudinais, as suposições de erros não correlacionados e homogeneidade de variância não são realísticas.

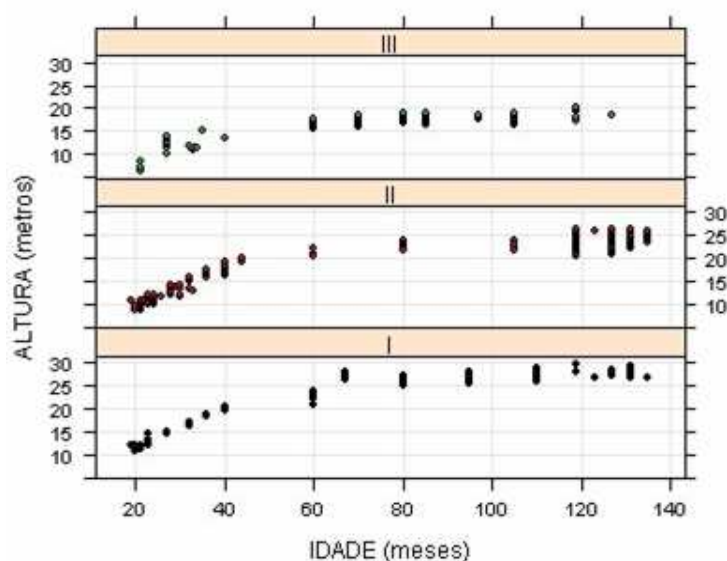


Figura 7 Gráficos dos dados apresentados em classes de sítio

#### **Ajuste e seleção dos modelos de estimativa de altura**

As alturas totais das árvores foram ajustadas tomando-se o modelo logístico de três parâmetros, apresentado por Pinheiro e Bates (2000), confrontado com o modelo exponencial modificado aqui chamado de modelo proposto ou simplesmente proposto, como indicado na Tabela 4. Os resultados obtidos para o primeiro critério de seleção dos modelos são apresentados na Tabela 4, em que se apresentam os parâmetros estimados e as medidas de não linearidade intrínseca IN e devido ao efeito dos parâmetros PE. Foi adotado

como valor de corte para estas medidas aquele proposto por Robinson e Hamann (2010), que sugerem 0,3 como limite aceitável para estas duas medidas.

Tabela 4 Parâmetros IN e PE dos modelos de crescimento em altura não lineares logístico e proposto, ajustado por classe de sítio

Classe	Modelo	Parâmetros			Med. não linearidade	
		$\Theta_1$	$\Theta_2$	$\Theta_3$	IN	PE
I	Logístico	27,5430	4,8492	0,0615	0,0332	0,1661
	Proposto	28,7497	2,6996	1,0232	0,0156	0,2273
II	Logístico	23,9054	6,9324	0,0744	0,0209	0,2498
	Proposto	24,51907	2,6905	1,0271	0,0099	0,2219
III	Logístico	18,0401	6,8650	0,0838	0,1013	<b>0,9130</b>
	Proposto	18,7873	2,7126	1,0382	0,0225	<b>0,3203</b>

Todos os parâmetros estimados para os dois modelos nas três classes apresentaram resultados estatisticamente significativos, a 95% de confiança. Observa-se, pelos dados da Tabela 4, que a medida de não linearidade devido ao efeito dos parâmetros, para os dados relativos à classe III, foram superiores a 0,3, em ambos os modelos, indicando que, mesmo ocorrendo ajuste, a não linearidade existente na aplicação do modelo à base de dados pode ser suficientemente grande para comprometer o uso de diversas metodologias estatísticas que estão sustentadas pela aproximação linear. Deve-se, ainda, observar que o excesso apresentado pelo modelo proposto é mínimo. Apesar de identificar a existência do excesso de não linearidade provocado pelos parâmetros, a medida proposta por Bates e Watts (1980) é incapaz de fornecer informações que permitam identificar o parâmetro responsável. A avaliação de 10.000 amostras geradas aleatoriamente por meio de simulação Monte Carlo

possibilitou conhecer o viés, conforme Tabela 5. Foi adotado o valor de 1%, sugerido por Ratkowsky (1983), como um valor de corte para identificação do parâmetro que afeta a aproximação linear.

Tabela 5 Medidas do viés para os parâmetros dos modelos em uso (em termos percentuais)

		Parâmetros		
Classe	Modelo	$\Theta_1$	$\Theta_2$	$\Theta_3$
I	Logístico	0,010173	0,040922	-0,012418
	Proposto	0,012226	0,008088	0,000466
II	Logístico	0,002020	0,254326	0,044880
	Proposto	0,032023	0,016272	0,001162
III	Logístico	0,061502	<b>2,486479</b>	0,241213
	Proposto	0,221612	1,004477	0,056044

Observa-se que o parâmetro  $\Theta_2$  excedeu minimamente o valor limite estabelecido nos dois modelos. Pode-se afirmar que, conforme a medida de viés proposta por Box, este é o parâmetro responsável pelo valor apresentado na medida de não linearidade, devido ao efeito dos parâmetros. O modelo proposto apresenta valores para a medida do viés aceitáveis para todos os parâmetros.

Os gráficos de perfil foram desenvolvidos por Bates e Watts (1988) e são apresentados em painéis, que proveem uma informação da normalidade dos parâmetros sob a condição da distribuição dos parâmetros estimados. Este gráfico mostra, para cada par de parâmetros x e y de uma distribuição, duas curvas que fazem intersecção na estimativa de máxima verosimilhança, que dão os loci de pontos onde as tangentes aos contornos do perfil da probabilidade bivariada tornaram vertical e horizontal, respectivamente. Para o perfil de uma distribuição exatamente normal bivariada, o traço dessas duas curvas é uma

linha reta. Os gráficos obtidos corroboram as medidas de viés quanto aos parâmetros que estão provocando a incompatibilidade de modelo com o conjunto de dados. Nas Figuras 8 a 13 são apresentados os painéis dos gráficos de perfil para o conjunto modelo/base de dados, de acordo com cada uma das três classes de índices de sítio. Conforme mencionado, na interpretação dos gráficos espera-se que a linha tracejada vertical ocorra aproximadamente ao meio do intervalo de confiança do parâmetro em avaliação. A não simetria é um indicativo de que a aproximação linear está comprometida localmente. Os gráficos foram construídos adotando-se que os parâmetros são apresentados com a seguinte correspondência para sua representação:  $\Theta_1=a$ ,  $\Theta_2=b$  e  $\Theta_3=c$ .

#### CLASSE I

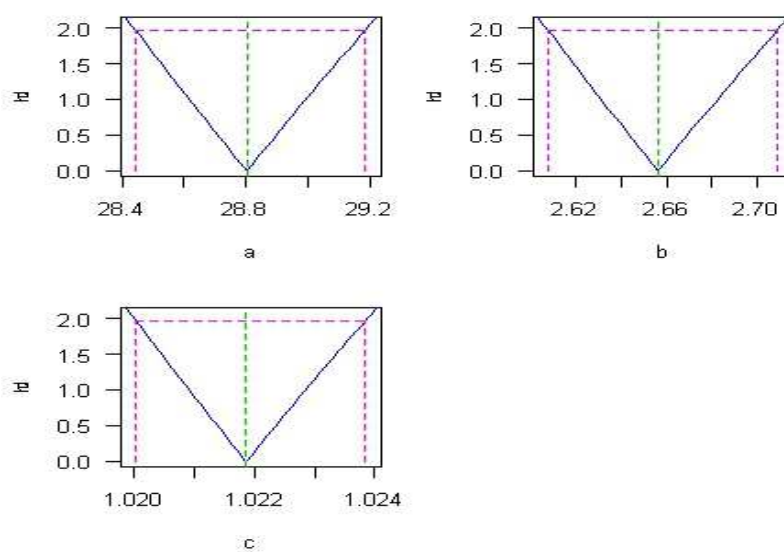


Figura 8 Gráficos de perfil para avaliação visual do comprometimento da aproximação linear localmente, devido aos parâmetros (modelo proposto/dados da classe de sítio I)

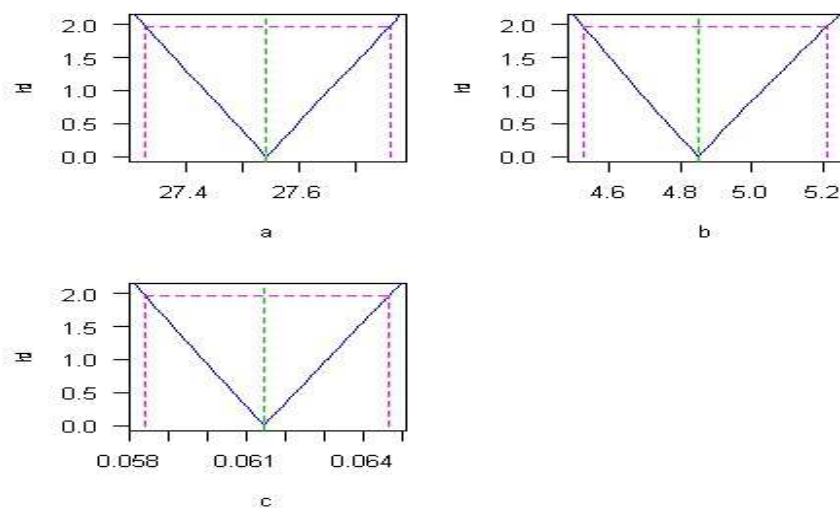


Figura 9 Gráficos de perfil para avaliação visual do comprometimento da aproximação linear localmente, devido aos parâmetros (modelo logístico/dados da classe de sítio I)

#### CLASSE II

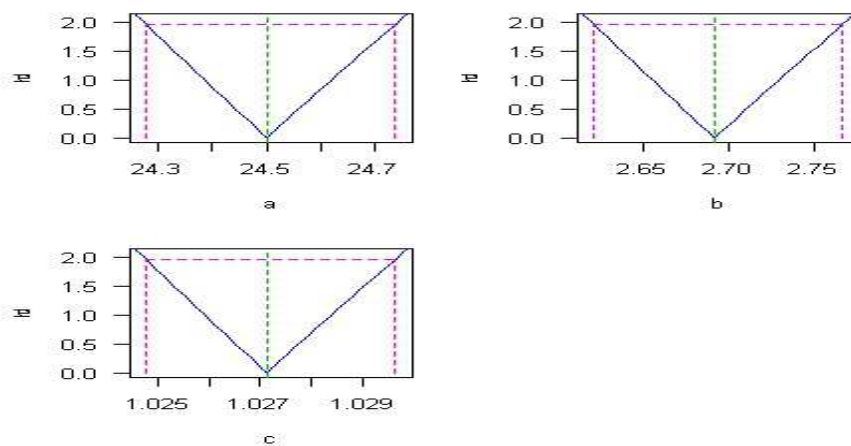


Figura 10 Gráficos de perfil para avaliação visual do comprometimento da aproximação linear localmente, devido aos parâmetros (modelo proposto/ dados da classe de sítio II)

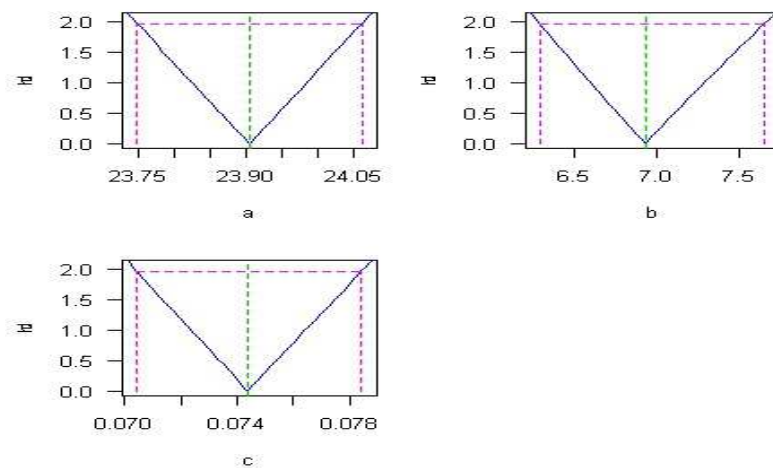


Figura 11 Gráficos de perfil para avaliação visual do comprometimento da aproximação linear localmente, devido aos parâmetros (modelo logístico/ dados da classe de sítio II)

### CLASSE III

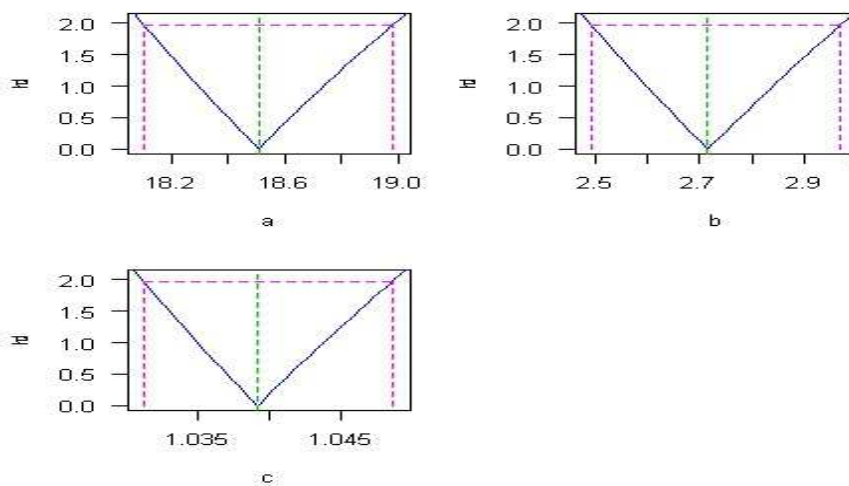


Figura 12 Gráficos de perfil para avaliação visual do comprometimento da aproximação linear localmente, devido aos parâmetros (modelo proposto/dados da classe de sítio III)

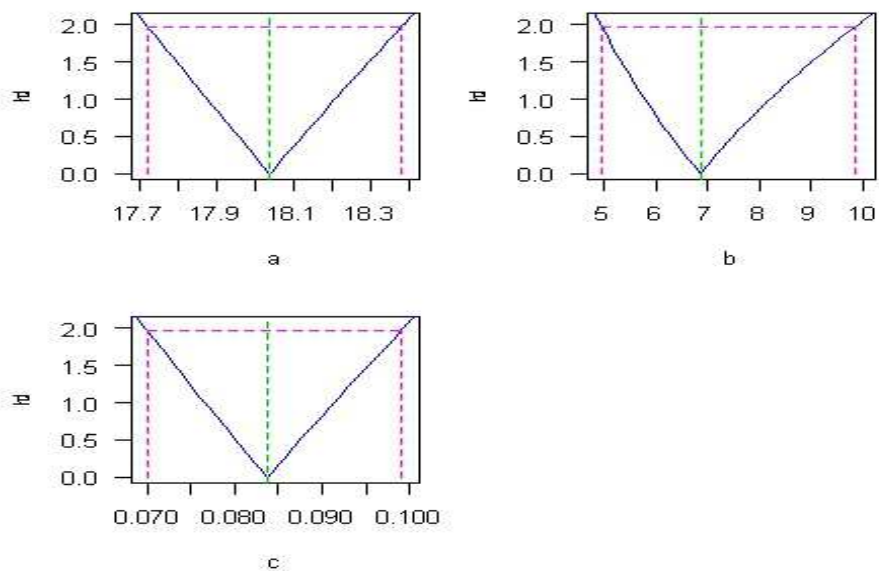


Figura 13 Gráficos de perfil para avaliação visual do comprometimento da aproximação linear localmente, devido aos parâmetros (modelo logístico/ dados da classe de sítio III)

Observa-se que os parâmetros  $\theta_2$  que, nos gráficos de perfil, estão representados por b, quando aplicados ao ajuste dos dados para a classe de sítio III, apresentam falta de simetria, indicando serem os responsáveis pela medida de não linearidade, devido ao efeito dos parâmetros em cada um dos modelos. O teste F do viés indica que apenas o valor referente ao modelo logístico é significativo e o modelo proposto é, portanto, satisfatório para o uso quanto às medidas de curvatura. No anexo A são apresentadas as Figuras 7A a 12A, que são os histogramas das distribuições dos parâmetros obtidos por meio de 10.000 simulações realizadas com uso de simulação Monte Carlo.

Ratkowsky (1983) afirma que o problema, geralmente, pode ser solucionado pela parametrização do modelo e sugere, ainda, que uma opção de transformação do parâmetro é a sua logaritmização. Esta transformação permite

que avanços iguais nos valores dos parâmetros no espaço paramétrico resultem em intervalos aproximadamente iguais no lócus de solução.

### Seleção e diagnóstico do modelo

Na Tabela 6 apresentam-se as estatísticas critério de informação de Akaike (AIC), erro padrão relativo (ERP), erro de predição médio (EPM) e coeficiente de determinação ajustado  $R^2_{aj}$ , que foram utilizadas para a seleção dos modelos.

Tabela 6 Estatísticas para a seleção do modelo de crescimento em altura por classe de sítio, para os modelos reduzidos

CLASSE	MODELO	EPM	ERP	AIC	$R^2_{aj}$
I	Logístico	-0,224336	4,723731	662,2671	0,9857
	Proposto	-0,183744	5,128140	584,2384	0,9838
II	Logístico	-0,652373	7,383233	2000,321	0,9637
	Proposto	-0,589527	7,538343	1914,445	0,9638
III	Logístico	-1,458352	9,214695	606,4061	0,9168
	Proposto	-1.019528	12,66928	373,7108	0,8656

em que

EPM – erro de predição percentual;

ERP – erro residual padrão;

AIC - critério de informação de Akaike

$R^2_{aj}$  - coeficiente de determinação ajustado.

Observa-se, pelos valores de EPM, que todos os modelos, em média, superestimaram as alturas. O modelo proposto apresentou melhores valores nas estatísticas de AIC, com resultados distanciando mais que duas unidades dos valores determinados para o modelo logístico, o que torna as diferenças significativas. Superou, ainda, o modelo logístico em EPM, ou seja, o erro médio provocado pelo modelo proposto foi inferior ao do modelo logístico.



Contudo, para as medidas de ERP e  $R^2_{aj.}$ , o modelo logístico apresentou diferenças desprezíveis para os modelos das classes de sítio I e II em relação ao modelo proposto, enquanto, para a classe de sítio III, superou apenas em  $R^2_{aj.}$  Nas Figuras 1A a 6A, do anexo A, são apresentadas as elipses de confiança paramétricas dos modelos ajustados.

Os modelos que apresentaram medidas satisfatórias foram ajustados com decomposição dos parâmetros  $\Theta_1$ .

Todos os ajustes, gráficos e simulações foram executados com auxílio do software estatístico R. As simulações e o estudo bootstrap foram executados adequando-se a modelos de três parâmetros o script para modelos de dois parâmetros apresentado por Souza (1998).

### **5.1 Ajustes dos modelos com decomposição de parâmetros**

Os modelos tiveram os parâmetros  $\Theta_1$  decompostos em função da variável DAP, por ser esta uma variável fortemente correlacionada com a altura. Em seguida, estes novos modelos foram ajustados apresentando os valores de parâmetros conforme a Tabela 7 e estatísticas para a seleção dos modelos conforme a Tabela 8.

Tabela 7 Parâmetros estimados para modelos de altura por classe de sítio para os modelos completos

CLASSE	MODELO	Parâmetros			
		$\Theta_1$	$\Theta_2$	$\Theta_3$	$\Theta_4$
I	Logístico	25,8332	4,6326	0,0622	2,4786
	p-valor	0,0000	0,0000	0,0000	0,0007
	Proposto	26,3276	26,3276	1,0220	11,947
	p-valor	0,0000	0,0000	0,0000	0,000
II	Logístico	22,8635	6,8212	0,0753	3,5223
	p-valor	0,0000	0,0000	0,0000	<b>0,0969</b>
	Proposto	23,5743	2,6810	1,0278	4,9191
	p-valor	0,0000	0,0000	0,0000	<b>0,1190</b>
III	Proposto	17,4890	2,6885	1,0396	7,7236
	p-valor	0,0000	0,0000	0,0000	0,0157

Os parâmetros ajustados foram significativos para todos os modelos, exceto nos modelos para o índice de classe de sítio II. Por este motivo, foi selecionado entre os modelos reduzidos para esta classe de sítio. Na Figura 14 apresenta-se a distribuição, em três dimensões, dos dados para altura, expressa em função da idade e do DAP.

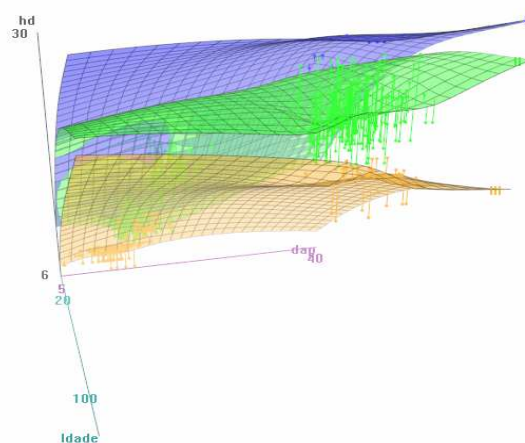


Figura 14 Distribuição tridimensional da altura de árvores, expressa em relação à idade e ao DAP

Tabela 8 Estatísticas para seleção do modelo de altura por classe de sítio, para os modelos completos utilizados

CLASSE	MODELO	EPM	ERP	AIC	R <sup>2</sup> aj.
I	Logístico	-0,0022	4,6330	652,8913	0,98630
	Proposto	-0,1813	4,7566	666,9541	0,98560
III	Proposto	-0,9785	8,3875	573,33	0,93150

em que

EPM – erro de predição percentual;

ERP – erro residual padrão;

AIC - critério de informação de Akaike;

R<sup>2</sup>aj.- coeficiente de determinação ajustado.

Na Tabela 8 observam-se os resultados das estatísticas para seleção entre os modelos de ajuste de altura com decomposição do parâmetro  $\Theta_4$  dos modelos para ajuste das alturas na classe de sítio I, apresentando o modelo completo logístico com melhores estatísticas. Com estes resultados foram obtidos os modelos selecionados de acordo com a tabela 9.

Tabela 9 Modelos selecionados

CLASSE MODELO	TIPO do MODELO
I	Logístico completo
II	Proposto reduzido
III	Proposto completo

## 5.2 Estimativa do DAP para estabelecer a produtividade com o modelo completo

Os modelos de crescimento completos selecionados têm como covariável o DAP e seu conhecimento nas diversas idades é necessário para se estabelecer a produtividade. Estas estimativas são apresentadas na Tabela 10.

Tabela 10 Parâmetros, estatísticas IN e PE dos modelos de crescimento em DAP logístico e proposto, ajustado nas classes de índice de sítio em que serão utilizados como covariável no modelo completo

Classe	Modelo	Parâmetros			Med. não linearidade	
		$\Theta_1$	$\Theta_2$	$\Theta_3$	IN	PE
I	Logístico	0,185471	1,916289	0,0324	0,1188	0,2536
III	Proposto	0,2244	-2.531616	1.0101	0.0497	0.2952

Todos os parâmetros estimados foram estatisticamente significativos, a 95% de confiança. Não consta ajuste do DAP para a classe II, pois o modelo selecionado para ajuste de altura foi o modelo reduzido. No anexo A, nas Figuras 13A a 16A, estão representadas as elipses de confiança e os histogramas gerados a partir de 10.000 simulações Monte Carlo, para os dois modelos utilizados no ajuste do crescimento em altura.

## 5.3 Modelo de produtividade compatível com o modelo de crescimento em altura

Um modelo de produtividade compatível pode ser obtido pela derivação modelo de crescimento e o crescimento ocorrido em um intervalo de tempo será

determinado pela integral definida neste período de tempo. Dessa forma, os modelos de produtividade foram dados para as classes por:

**Modelo completo logístico selecionado para o ajuste dos dados na classe de sítio I**

$$\begin{aligned}
 \text{Produção em HT} &= \int_{idade_1}^{idade_2} \frac{\theta_1 + \theta_4 * dap}{1 + \theta_2 * e^{(-\theta_3 * idade)}} = -\frac{\theta_1 * \ln(e^{(-\theta_3 * idade_2)})}{\theta_3} \\
 &+ \frac{\theta_1 \ln(1 + \theta_2 * e^{(-\theta_3 * idade_2)})}{\theta_3} - \frac{\theta_4 * DAP_2 * \ln(e^{(-\theta_3 * idade_2)})}{\theta_3} \\
 &+ \frac{\theta_2 * DAP_2 * \ln(1 + \theta_2 * e^{(-\theta_3 * idade_2)})}{\theta_3} + \frac{\theta_1 * \ln(e^{(-\theta_3 * idade_1)})}{\theta_3} \\
 &- \frac{\theta_1 \ln(1 + \theta_2 * e^{(-\theta_3 * idade_1)})}{\theta_3} + \frac{\theta_4 * DAP_2 * \ln(e^{(-\theta_3 * idade_1)})}{\theta_3} \\
 &- \frac{\theta_2 * DAP_2 * \ln(1 + \theta_2 * e^{(-\theta_3 * idade_1)})}{\theta_3} \quad (44)
 \end{aligned}$$

em que

$\theta_1$ ,  $\theta_2$  e  $\theta_3$  - parâmetros ajustados no modelo logístico de crescimento;

DAP<sub>1</sub> – diâmetro à altura do peito, tomado a 1,30 m, no momento inicial do intervalo de produtividade;

DAP<sub>2</sub> - diâmetro à altura do peito, tomado a 1,30 m, no momento final do intervalo de produtividade;

idade<sub>1</sub> – idade da árvore no momento inicial do intervalo de produtividade;

idade<sub>2</sub> - idade da árvore no momento final do intervalo de produtividade.

**Modelo reduzido proposto selecionado para o ajuste dos dados na classe de sítio II**

$$\begin{aligned}
 \text{Produção em Hd} &= \int_{idade_1}^{idade_2} (\theta_1 + \theta_2^{-idade}) * e^{(\theta_2 + \theta_3^{-idade})} = \\
 &\theta_1 * idade_2 + \frac{e^{(\theta_2 + \theta_3^{-idade_2})}}{\ln(\theta_3)} - \theta_1 * idade_1 - \frac{e^{(\theta_2 + \theta_3^{-idade_1})}}{\ln(\theta_3)} \quad (45)
 \end{aligned}$$

As expressões da equação 15 têm o mesmo significado apresentado para equação 14.

**Modelo completo proposto selecionado para o ajuste dos dados na classe de sítio III**

$$\begin{aligned}
 \text{Produção em HT} &= \int_{idade_1}^{idade_2} (\theta_1 + \theta_2 * dap_1) + \theta_3^{-idade} * e^{(\theta_2 + \theta_3^{-idade})} = \\
 &\theta_1 * idade_2 + \theta_4 * dap_2 * idade_2 + \frac{e^{(\theta_2 + \theta_3^{-idade_2})}}{\ln(\theta_3)} - \theta_1 * dap_1 * idade_1 \\
 &- \frac{e^{(\theta_2 + \theta_3^{-idade_1})}}{\ln(\theta_3)} \quad (46)
 \end{aligned}$$

As expressões da equação 46 têm o mesmo significado apresentado para equação 44.

#### 5.4 Modelos compatíveis de crescimento e produtividade de volume

##### Modelagem da produtividade

O modelo exponencial da expressão 47 foi utilizado para ajustar os dados em cada uma das classes de sítio.

$$v_{tot_i} = e^{\left(\theta_1 + \frac{\theta_2}{10000t_i}\right)} + \varepsilon_i \quad (47)$$

em que

$v_{tot_i}$  – volume total da i-ésima árvore

$\theta_1$  e  $\theta_2$  – parâmetros

$t_i$  - idade da i-ésima árvore

$\varepsilon_i$  - erro

Na Tabela 11 apresentam-se os valores ajustados dos parâmetros e as medidas de curvatura intrínseca e devido ao efeito dos parâmetros.

Tabela 11 Parâmetros, IN e PE dos modelos de produtividade em altura não lineares exponenciais ajustados por classe de sítio

Classe	Parâmetros		Medidas de não linearidade	
	$\theta_1$	$\theta_2$	IN	PE
I	-0,026102	-40,586596	0,0026	0,0101
II	0,072816	-54,577724	0,0027	0,0109
III	0,28317	-61,70824	0,0076	0,0239

Conforme Ratkowsky (1983), o modelo apresenta valores nas medidas de curvatura que garantem o não comprometimento da teoria de aproximação linear, o que torna confiáveis as estatísticas de seleção e diagnóstico a serem

calculadas. Todos os parâmetros dos três modelos avaliados apresentaram valores estatisticamente significativos, a 95% de confiança. Foram realizadas 10.000 simulações Monte Carlo para estudo dos parâmetros. Nas Figuras 15 a 17 apresentam-se os histogramas das distribuições dos parâmetros para cada modelo ajustado.

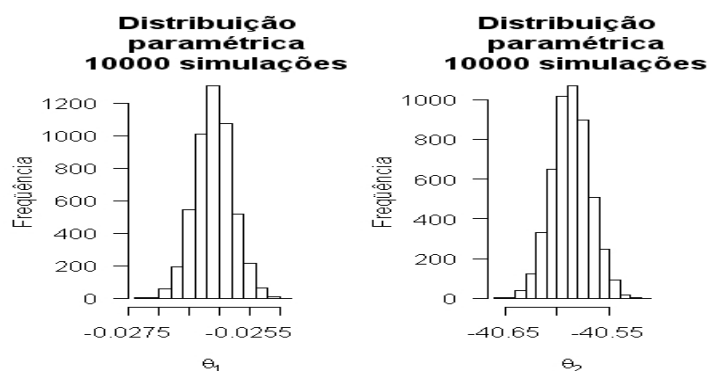


Figura 15 Distribuição paramétrica do modelo exponencial de ajuste dos dados da classe de sítio I

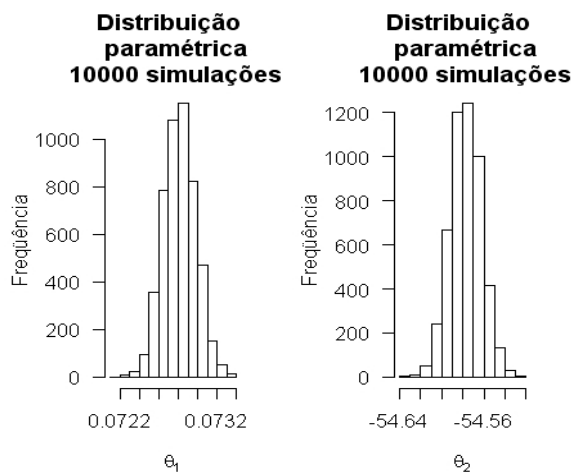


Figura 16 Distribuição paramétrica do modelo exponencial de ajuste dos dados da classe de sítio II



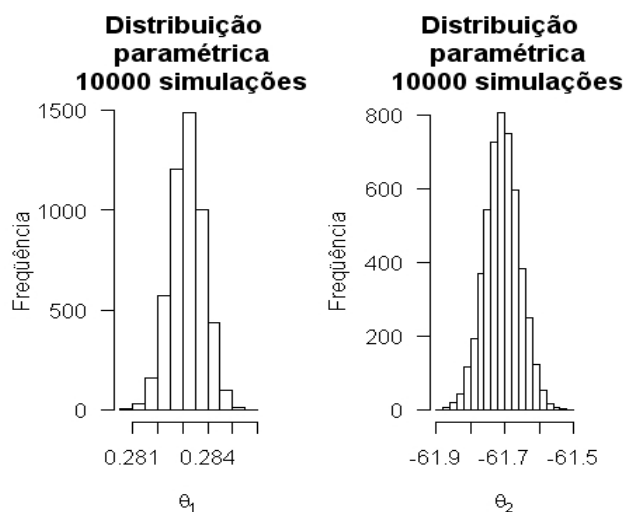


Figura 17 Distribuição paramétrica do modelo exponencial de ajuste dos dados da classe de sítio III

Os histogramas confirmam o diagnóstico das medidas de não linearidade, verificando-se simetria na distribuição dos parâmetros. As elipses de confiança resultantes do estudo dos parâmetros por simulação são apresentadas no anexo A, nas Figuras 17A a 19A.

Na Tabela 12 apresentam-se as estatísticas de diagnóstico dos modelos ajustados.

Tabela 12 Estatísticas para diagnóstico do modelo de produtividade em volume, por classe de sítio

CLASSE	EPM	ERP	AIC	$R^2_{aj.}$
I	0,03577835	4,85591	-1016,735	0,9773
II	-9,876636	2,808812	-2461,63	0,9876
III	-40,90037	3,99843	-637,7517	0,9797

em que

EPM – erro de predição percentual;

ERP – erro residual padrão;

AIC - critério de informação de Akaike;

$R^2_{aj}$  - coeficiente de determinação ajustado.

O modelo de ajuste dos dados classificados na classe de sítio I subestimou em sua estimativa, enquanto os demais superestimaram. O erro de predição, assim como o coeficiente de determinação, qualifica os modelos ao uso; o primeiro, por seus valores pequenos, inferiores a 5% e o segundo, por apresentarem resultados superiores a 0,97, ou seja, 97%. Nas Figuras de 20A a 21A, do anexo, observam-se as representações das curvas de ajustes dos dados.

#### **Modelo de crescimento compatível com o modelo de produtividade**

A expressão 47 representa o modelo de produtividade compatível com o modelo de produtividade estudado.

Obtenção do modelo de crescimento:

$$\text{Modelo de produtividade para voltot} = \int_{idade_1}^{idade_2} e^{\left(\theta_1 - \frac{\theta_2}{x}\right)} dx = -\frac{\theta_2 e^{\left(\theta_1 - \frac{\theta_2}{x}\right)}}{x^2}$$

em que

x – idade de interesse;

V – volume na idade de interesse.

Na Tabela 13 são apresentados os valores estimados para somatório de crescimento, comparado a períodos em meses da produtividade.

Tabela 13 Comparação dos resultados de produtividade e somatório de crescimento em diversos períodos

Período (meses)	Produtividade (m <sup>3</sup> )			Somatório do crescimento ( m <sup>3</sup> )		
	Classe			Classe		
	I	II	III	I	II	III
0 a 10	0,023	0,005	0,004	0,024	0,006	0,004
0 a 30	0,343	0,174	0,174	0,348	0,176	0,175
0 a 50	0,592	0,361	0,361	0,594	0,364	0,364
0 a 90	0,845	0,586	0,586	0,845	0,588	0,571
0 a 120	0,946	0,684	0,676	0,946	0,684	0,672

Observa-se que as diferenças ocorridas são desprezíveis, portanto, o uso de um resultado é tão confiável quanto o uso do outro.

## **6 CONCLUSÃO GERAL**

Os dados foram ajustados por classe de sítio com modelos não lineares. A verificação das medidas de não linearidade assumindo valores inferiores a 0,3 asseguraram as propriedades de “aproximação linear” em relação ao conjunto de dados utilizado, o que garantiu a validade das estatísticas apresentadas. Verificou-se, ainda, que, nas classes I e III de sítio em altura, os modelos sem covariáveis apresentaram melhores resultados. Para a estimação do volume em crescimento compatível com produtividade, o modelo exponencial sobressaiu-se, com melhores resultados.

## REFERÊNCIAS

- AKAIKE, H. A new look at the statistical model identification. **IEEE-Transactions on Automatic Control**, Boston, v. 19, n. 6, p. 716-723, 1974.
- ASSOCIAÇÃO BRASILEIRA DE PRODUTORES DE FLORESTAS PLANTADAS. **Anuário estatístico da ABRAF 2011, ano base 2010**. Disponível em: <<http://www.abraflor.org.br/estatisticas.asp>>. Acesso em: 18 jan. 2012.
- BATES, D. M.; WATTS, D. G. Relative curvature measures of nonlinearity. **Journal of Royal Statistical Society B**, London, v. 42, p. 1-25, 1980.
- BERTALANFFY, L. von. Quantitative laws for metabolism and growth. **Quarterly Review of Biology**, Stony Brook, v. 32, p. 217-231, 1957.
- BOX, G. E. P.; COX, D. R. An analysis of transformations. **Journal of Royal Statistical Society B**, London, v. 26, p. 211-243, 1964.
- BUCKMAN, R. E. **Growth and yield of red pine in Minnesota**. Washington: Buckman, 1962. 50 p. (Technical Bulletin, 1272).
- CALEGARIO, N. et al. Desenvolvimento de um modelo linear de efeito misto na estimativa do crescimento e produtividade de povoamentos clonais de *Eucalyptus*. **Cerne**, Lavras, v. 10, n. 1, p. 67-86, 2004.
- \_\_\_\_\_. Estimativa de crescimento de povoamentos de *Eucalyptus* baseada em modelos lineares em multiníveis de efeito misto. **Revista Árvore**, Viçosa, MG, v. 29, n. 2, p. 251-261, mar./abr. 2005.
- CAMPOS, J. C. C.; KRONKA, J. J. N. Tabela de volume comercial para *Pinus elliotii* Engelm. **Silvicultura em São Paulo**, São Paulo, v. 8, p. 75-80, 1973.
- CAMPOS, J. C. C.; LEITE, H. G. **Mensuração florestal: perguntas e respostas**. Viçosa, MG: UFV, 2006. 470 p.
- CAMPOS, J. C. C.; RIBEIRO, J. C. **Produtividade dos povoamentos de *Eucalyptus* spp na região do Triângulo Mineiro**. Viçosa, MG: SIF, 1983. 43 p. (Boletim Técnico, 1).

- CHAPMAN, D. E. Statistical problems in dynamics of exploited fishing populations. In: BERKELEY SYMPOSIUM ON MATHEMATICAL STATISTIC PROBABILITY, 1., 1961, Berkeley. **Proceedings...** Berkeley: SMSP, 1961. p. 153-158.
- CLUTTER, J. L. Compatible growth and yield models for loblolly pine. **Forest Science**, Bethesda, v. 9, n. 3, p. 355-371, 1963.
- COTTA, H. **systematische anleitung zur taxation der waldungen**. Berlin: J. D. Sander, 1804.
- DAVIS, L. S.; JOHNSON, K. N. **Forest management**. 3<sup>rd</sup> ed. New York: McGraw-Hill, 1987. 720 p.
- DRAPER, N. R.; SMITH, H. **Applied regression analysis**. 3<sup>rd</sup> ed. New York: J. Wiley, 1998. 706 p.
- EMILIANO, P. C. et al. Foundations and comparisons of information criteria Akaike e Bayesian. **Revista Brasileira de Biometria**, São Paulo, v. 27, n. 3, p. 394 -411, 2009.
- ENCINAS, J. I.; SILVA, G. F.; PINTO, J. R. R. Idade e crescimento das árvores. **Comunicações Técnicas Florestais**, Brasília, v. 7, n. 1, p. 1-43, 2005.
- FANG, Z. X.; BAILEY, R. L. Nonlinear mixed effects modeling for slash pine dominant height following growth intensive silvicultural treatments. **Forest Science**, Bethesda, v. 47, n. 3, p. 287-300, Aug. 2001.
- FERREIRA, A. B. H. **Novo dicionário da língua portuguesa**. Rio de Janeiro: Positivo, 2010. 1838 p.
- FOX, J.; WEISBERG, S. **An R companion to applied regression**. London: Sage, 2011. 449 p.
- GOMES, A. M. A. **Medição dos arvoredos**. Lisboa: S. da Costa, 1957. 413 p.
- GUIMARÃES, M. A. M. et al. Height-diameter models in forestry with inclusion of covariates. **Cerne**, Lavras, v. 15, n. 3, p. 313-321, 2009.
- HOFFMANN, R.; VIEIRA, S. **Análise de regressão: uma introdução à econometria**. 2. ed. São Paulo: Hucitec, 1977. 379 p.

HUSCH, B.; MILLER, C. I.; BEERS, T. W. **Forest mensuration**. New York: The Ronald, 1972. 410 p.

MACHADO, S. A. Curvas de índice de sítio para plantações de *Pinus taeda* L. na região central do Estado do Paraná. **Floresta**, Curitiba, v. 11, n. 2, p. 4-18, 1980.

MACKAY, E. **Dasometria**. Madri: Escuela Técnica de Ingenieros de Montes, 1964. 759 p.

MACKINEY, A. L.; SCHUMACHER, F. X.; CHAIKEN, L. E. Construction of yield tables for nonnormal loblolly pine stands. **Journal of Agricultural Research**, Washington, v. 54, p. 531-545, 1937.

MAZEROLLE, M. J. **Movements et reproduction des amphibiens en tourbières perturbées**. 2004. 78 f. Thesis (Ph.D. in Sciences Forestières) - Université Laval, Quebec, 2004.

MAZZINI, A. R. A. **Análise da curva de crescimento de machos Hereford considerando heterogeneidade de variâncias e autocorrelação dos erros**. 2001. 96 p. Dissertação (Mestrado em Estatística e Experimentação Agronômica) - Universidade Federal de Lavras, Lavras, 2001.

MOOD, A. M.; GRAYBILL, F. A.; BOES, D. C. **Introduction to theory of statistics**. 3<sup>rd</sup> ed. New York: Wiley, 1974. 564 p.

MOSER, J. W. Historical chapters in the development of modern forest growth and yield theory. In: BROWN, K. M.; CLARKE, F. R. (Ed.). **Forecasting forest stand dynamics**. Ontario: Thunder Bay, 1980. p. 42-61.

NETER, J.; WASSERMAN, W.; KUTNER, M. H. **Applied linear statistical models**. 3<sup>rd</sup> ed. Baldwinville: Irwin, 1990. 1181 p.

PIENAAR, L. V.; TURNBULL, K. J. The Chapman-Richards generalization of Von Bertalanffy's growth model for basal area growth and yield in even-aged stands. **Forest Science**, Bethesda, v. 19, n. 1, p. 2-22, 1973.

PINHEIRO, J. C.; BATES, D. **Mixed-effects models in S and S-PLUS**. New York: Springer Verlag, 2000. 528 p.

PRODAN, M. **Forest biometrics**. New York: Pergamon, 1968. 447 p.

PRUDENTE, A. A. **Modelos não-lineares de regressão: alguns aspectos de teoria assintótica**. 2009. 109 p. Dissertação (Mestrado em Modelagem Estatístico e Computacional) - Universidade Federal Rural de Pernambuco, Recife, 2009.

RATKOSWISKY, D. A. **Nonlinear regression modeling, a unified practical approach**. New York: M. Decker, 1983. 276 p.

RICHARDS, F. J. A flexible growth function ofr empirical use. **Journal of Experimental Botany**, Oxford, v. 10, p. 290-300, 1959.

ROBINSON, A. P.; HAMANN, J. D. **Forest analytics with R: an introduction**. New York: Springer, 2010. 339 p.

SCHUMACHER, F. X. A new growth curve and its application to timber-yield. **Journal of Forestry**, Washington, v. 37, p. 817-820, 1939.

SOUZA, E. M. **Modelagem não linear da extração de zinco em solo tratado com lodo de esgoto**. 2008. 95 p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras, 2008.

SOUZA, G. S. **Introdução aos modelos de regressão linear e não linear**. Brasília: EMBRAPA-SPI/EMBRAPA-SEA, 1998. 489 p.

SPURR, S. H. **Forest inventory**. New York: The Ronald, 1951. 476 p.

THORNLEY, J. H. M.; FRANCE, J. **Mathematical models in agriculture: quantitative methods for the plant, animal and ecological sciences**. 2<sup>nd</sup> ed. Cambridge: CABI, 2004. 906 p.

VEIGA, A. A. Balanços hídricos das dependências da divisão de florestas e estações experimentais. **Boletim Técnico do Instituto Florestal**, São Paulo, v. 18, p. 1-34, 1975.

\_\_\_\_\_. Determinação do “site”. **A semente**, São Paulo, v. 6, p. 5-6, jan. 1976.

\_\_\_\_\_. Proposição para trabalhos de campo na determinação de “site” em povoamentos naturais dissetâneos. **Silvicultura em São Paulo**, São Paulo, v. 16A, p. 1142-1120, 1980.



## ANEXOS

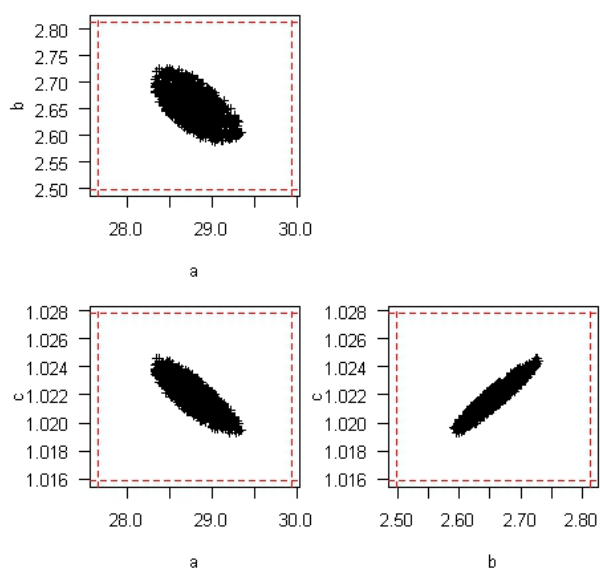


Figura 1A Elipses de confiança dos parâmetros do modelo de crescimento em altura (proposto/dados da classe de sítio I)

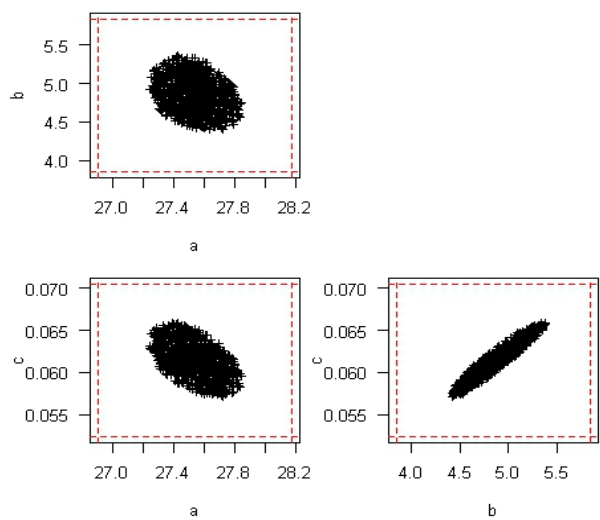


Figura 2A Elipses de confiança dos parâmetros do modelo de crescimento em altura (logístico/dados da classe de sítio I)

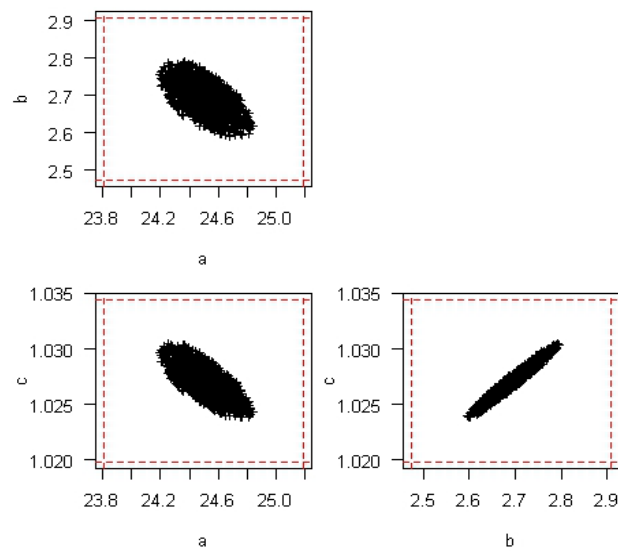


Figura 3A Elipses de confiança dos parâmetros do modelo de crescimento em altura (proposto/dados da classe de sítio II)

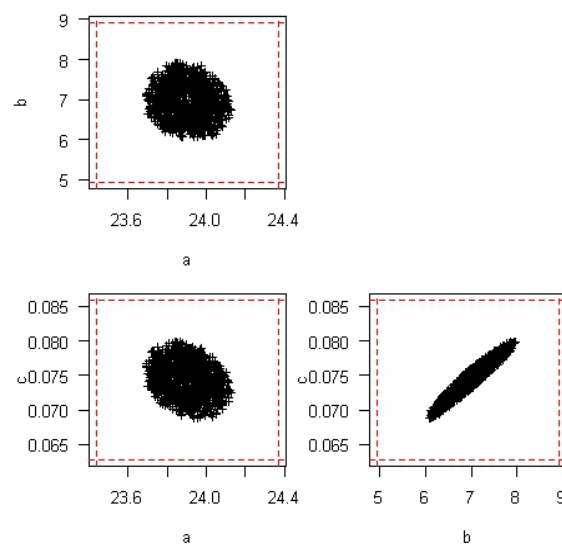


Figura 4A Elipses de confiança dos parâmetros do modelo de crescimento em altura (logístico/dados da classe de sítio II)

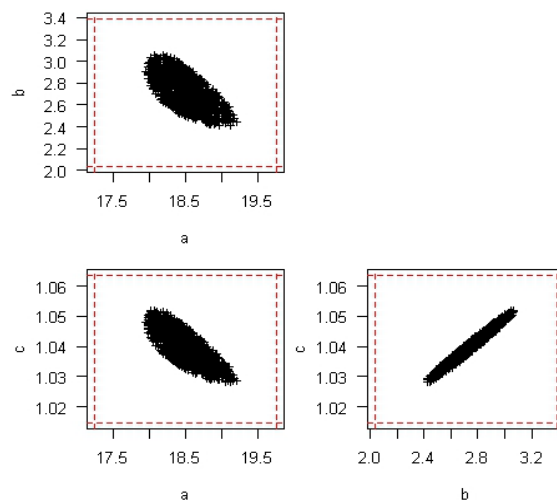


Figura 5A Elipses de confiança dos parâmetros do modelo de crescimento em altura (proposto/dados da classe de sítio III)

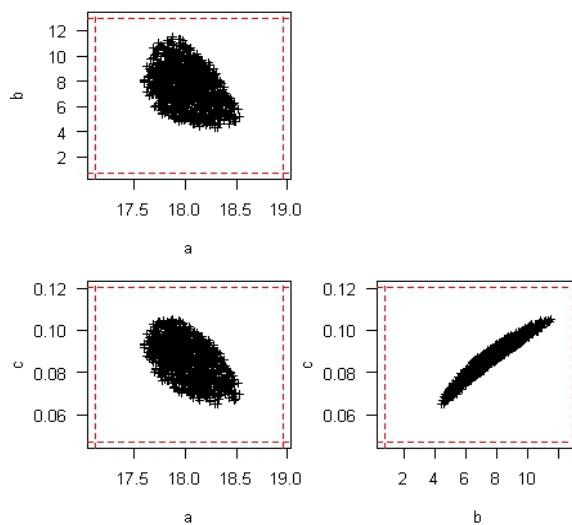


Figura 6A Elipses de confiança dos parâmetros do modelo de crescimento em altura (logístico/dados da classe de sítio III)

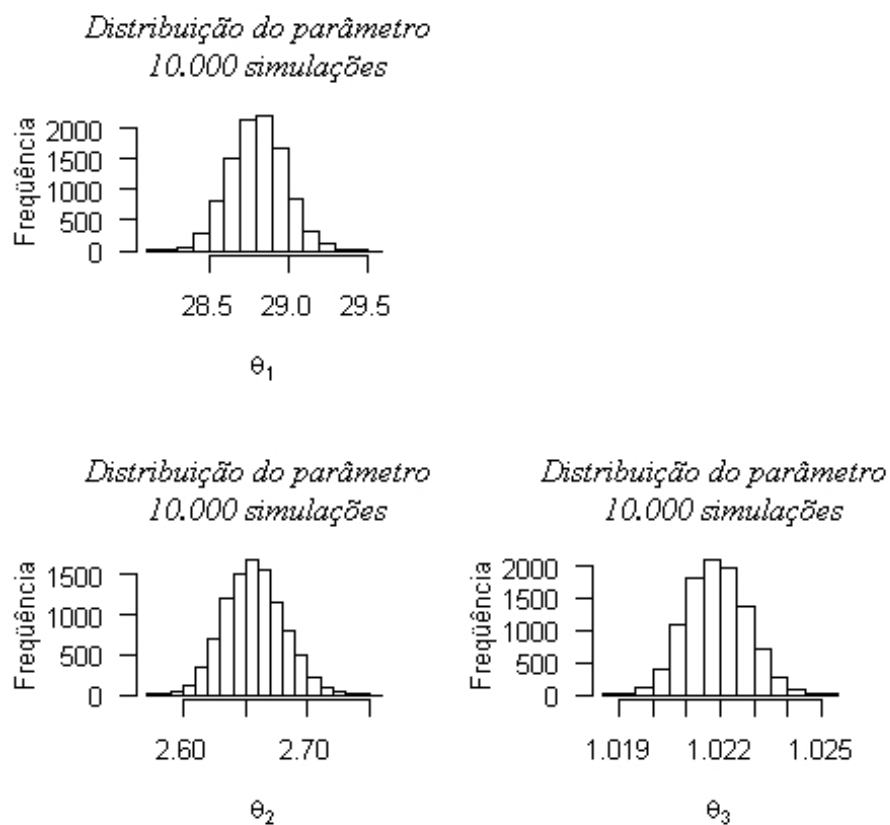


Figura 7A Histograma da distribuição dos parâmetros para avaliação visual do comprometimento local da aproximação linear, devido aos parâmetros modelo de crescimento em altura (proposto/dados da classe de sítio I)

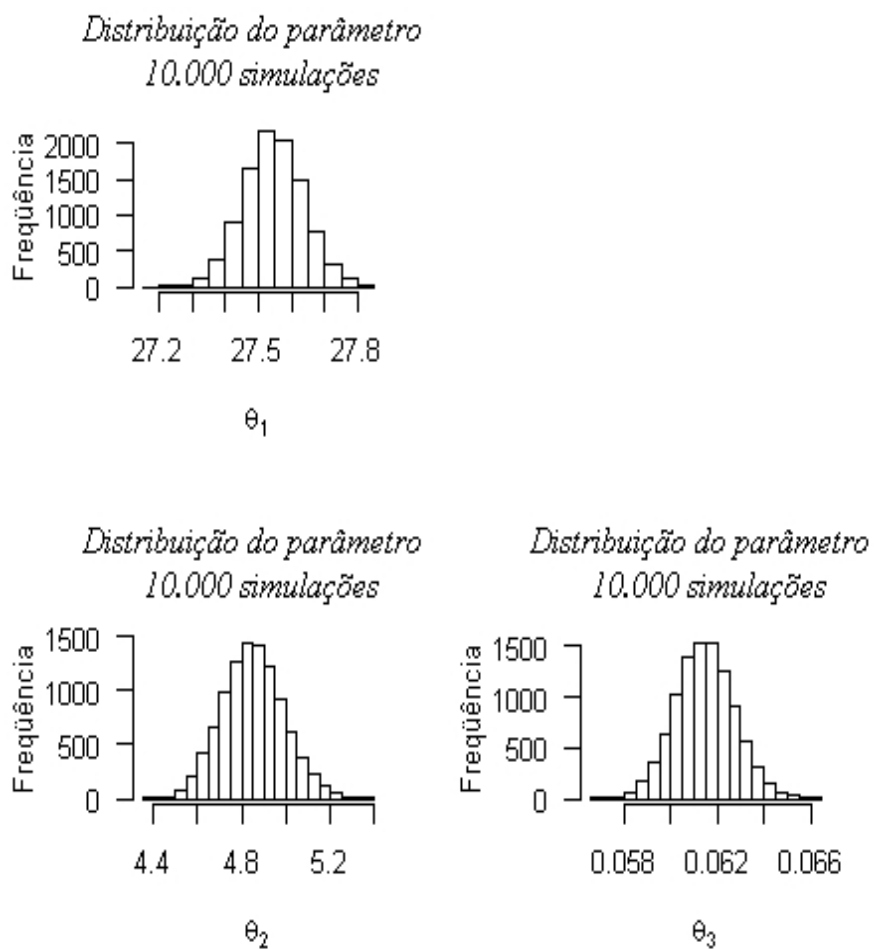


Figura 8A Histograma da distribuição dos parâmetros para avaliação visual do comprometimento local da aproximação linear, devido aos parâmetros modelo de crescimento em altura (logístico/dados da classe de sítio I)

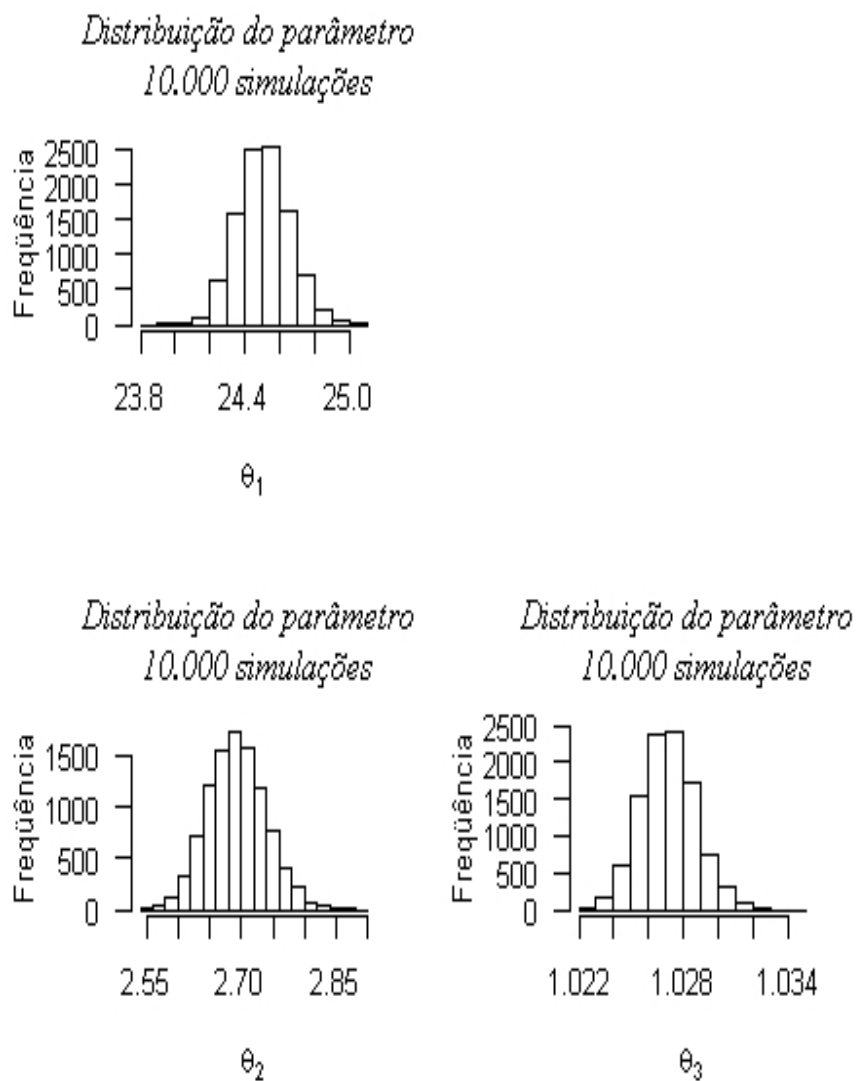


Figura 9A Histograma da distribuição dos parâmetros para avaliação visual do comprometimento local da aproximação linear, devido aos parâmetros modelo de crescimento em altura (proposto/dados da classe de sítio II)

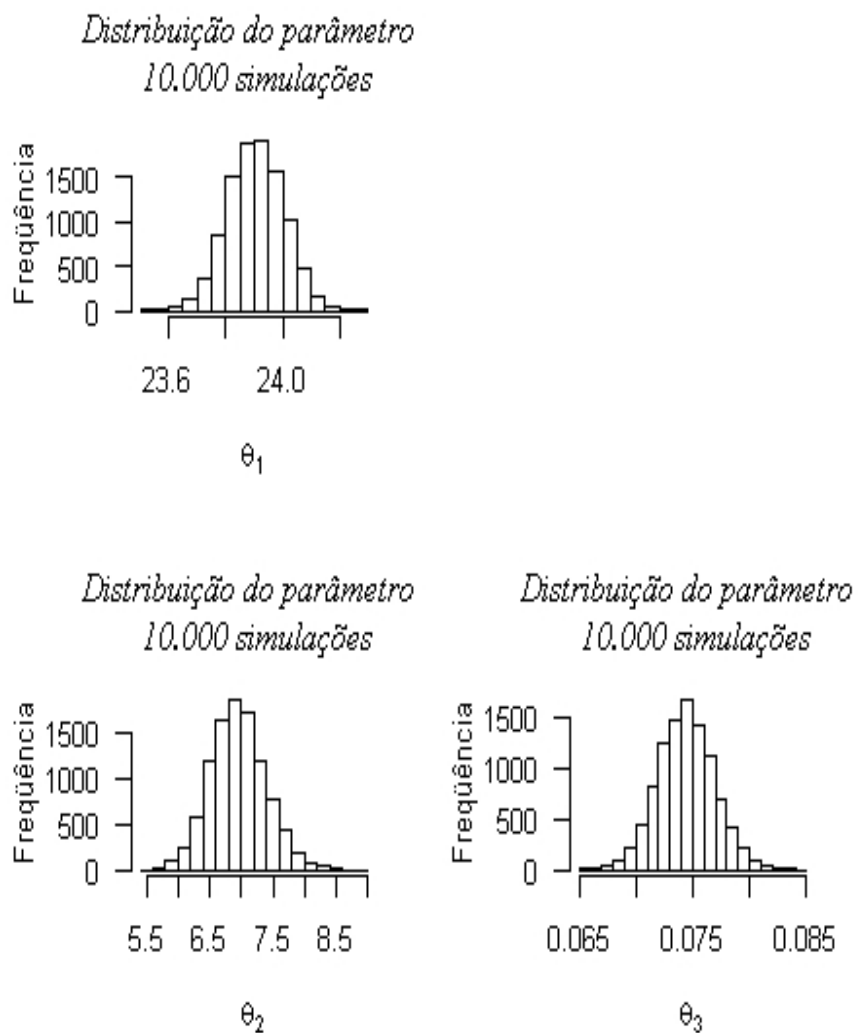


Figura 10A Histograma da distribuição dos parâmetros para avaliação visual do comprometimento local da aproximação linear, devido aos parâmetros modelo de crescimento em altura (logístico/dados da classe de sítio II)

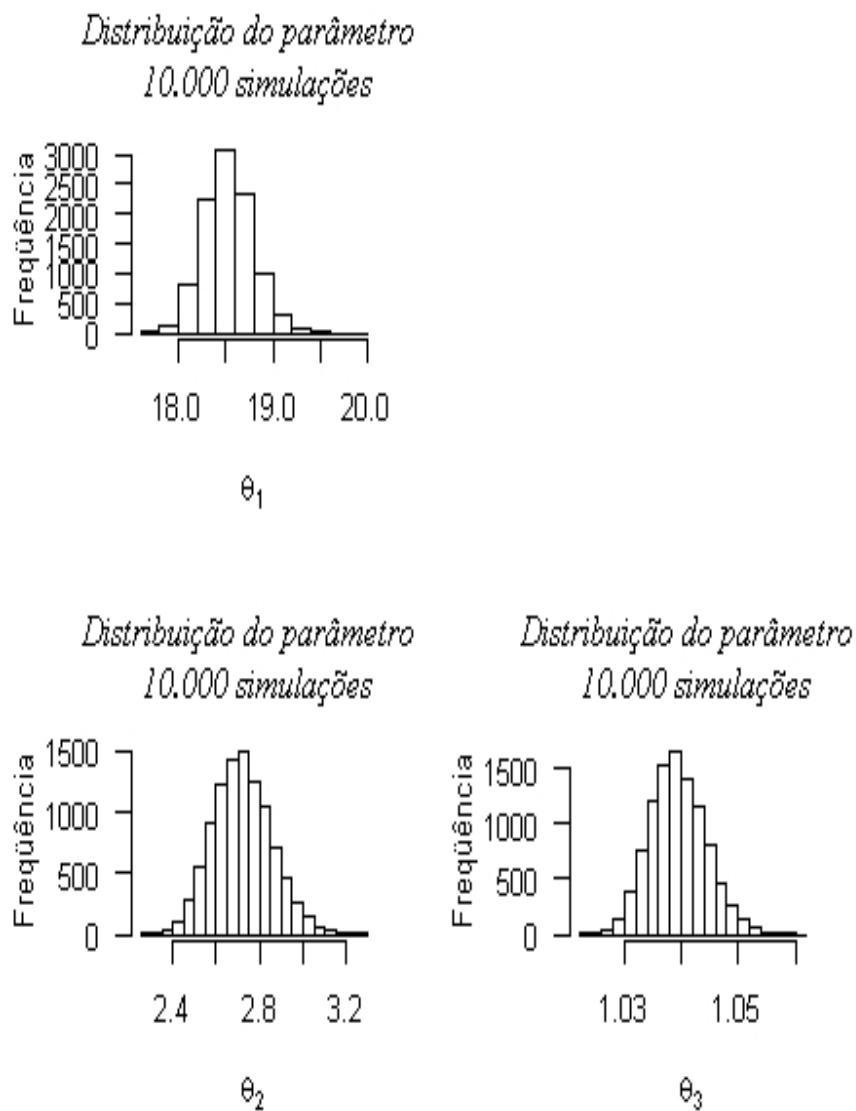


Figura 11A Histograma da distribuição dos parâmetros para avaliação visual do comprometimento local da aproximação linear, devido aos parâmetros modelo de crescimento em altura (proposto/dados da classe de sítio III)



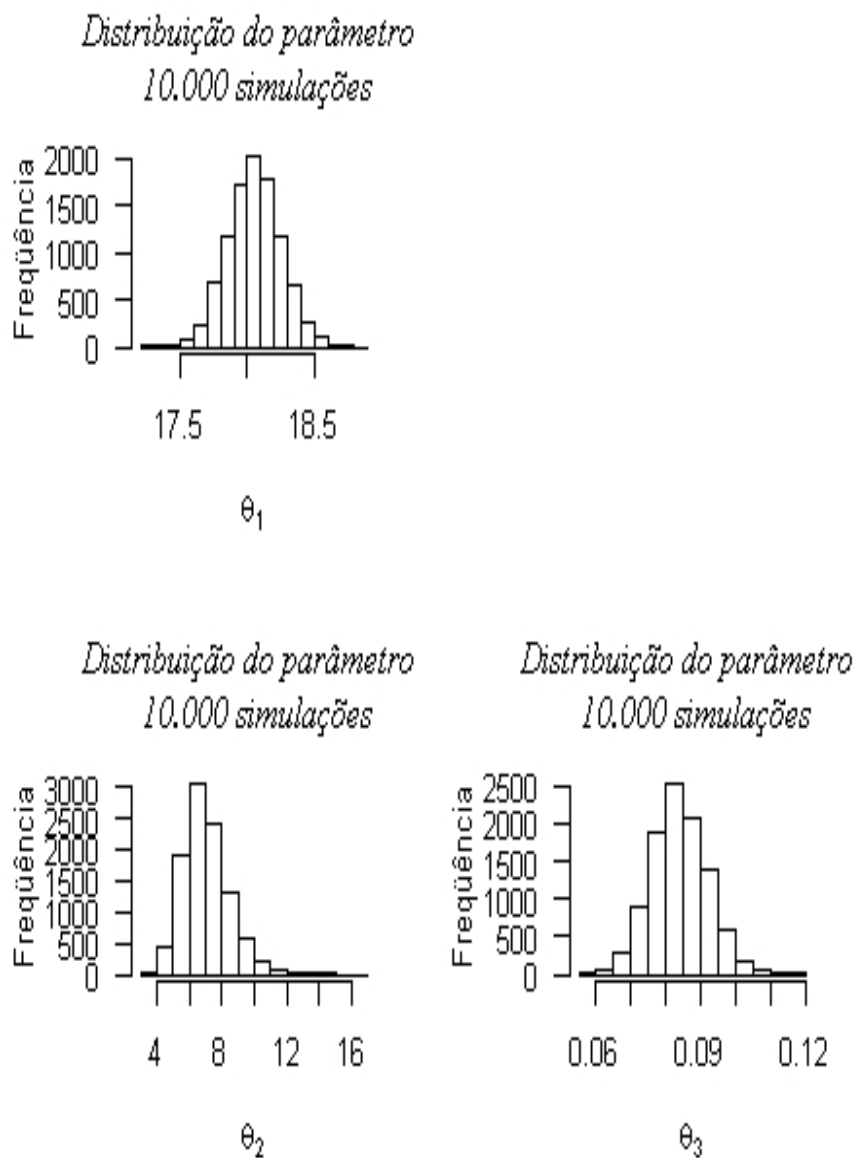


Figura 12A Histograma da distribuição dos parâmetros para avaliação visual do comprometimento local da aproximação linear, devido aos parâmetros modelo de crescimento em altura (logístico/dados da classe de sítio III)

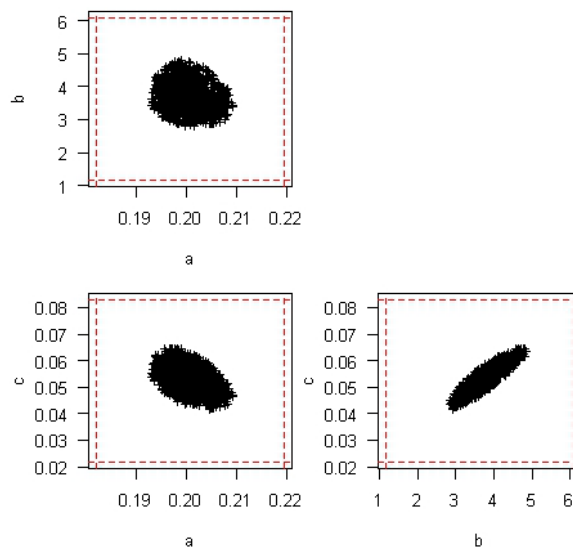


Figura 13A Elipses de confiança dos parâmetros modelo de crescimento em DAP (logístico/dados da classe de sítio I)

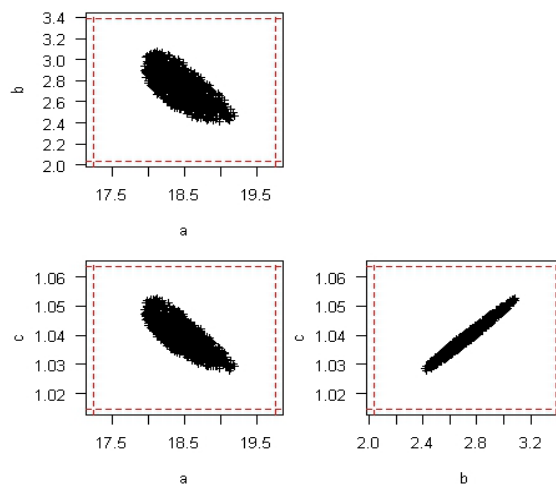


Figura 14A Elipses de confiança dos parâmetros modelo de crescimento em DAP (proposto/dados da classe de sítio III)

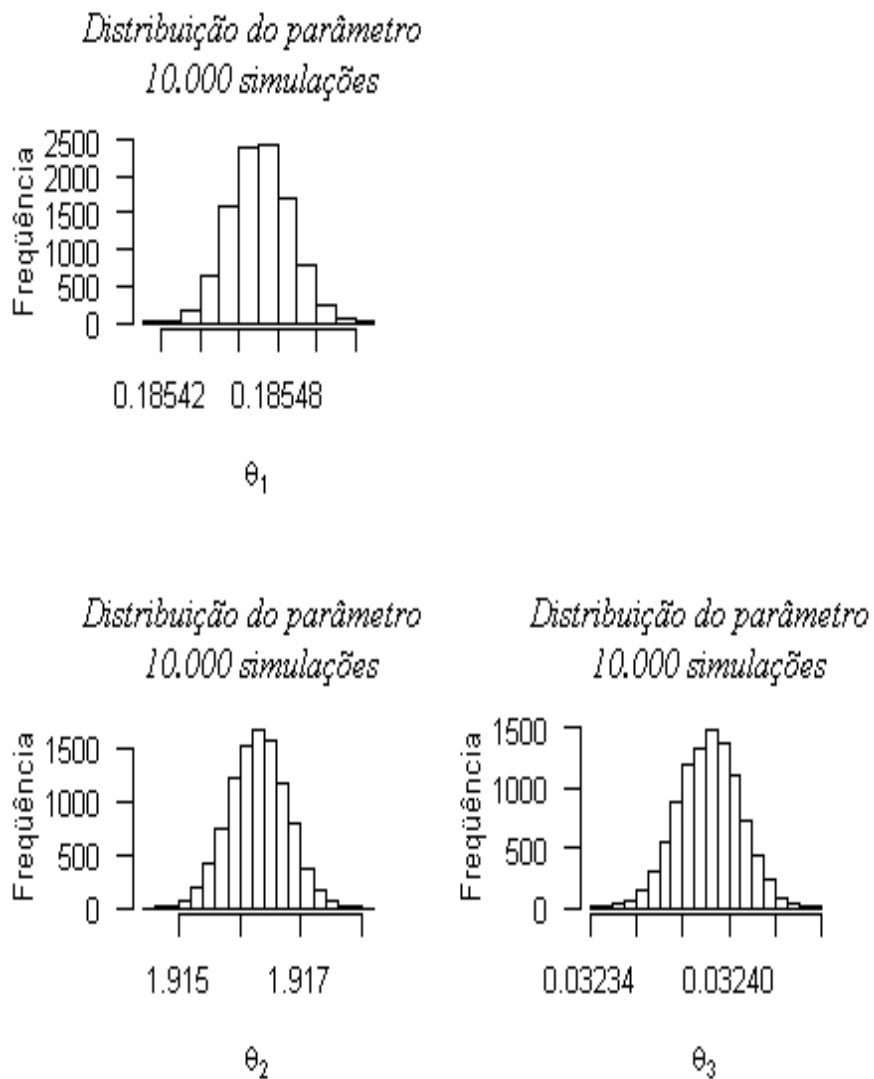


Figura 15A Histograma da distribuição dos parâmetros para avaliação visual do comprometimento local da aproximação linear devido aos parâmetros modelo de crescimento em DAP (logístico/dados da classe de sítio I)

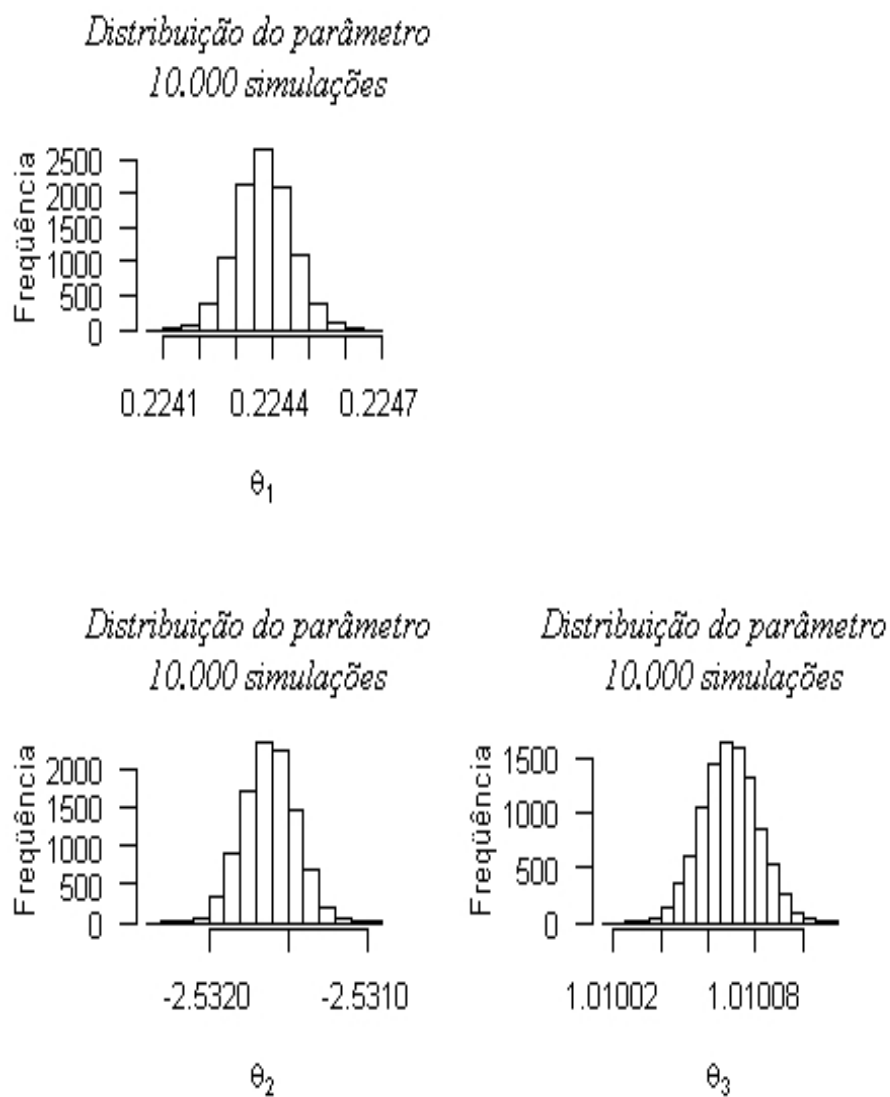


Figura 16A Histograma da distribuição dos parâmetros para avaliação visual do comprometimento local da aproximação linear devido aos parâmetros modelo de crescimento em DAP (proposto/dados da classe de sítio III)

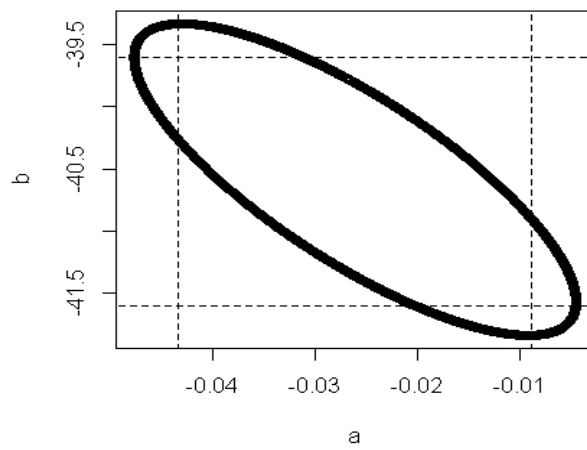


Figura 17A Elipses de confiança dos parâmetros modelo de crescimento em volume (proposto/dados da classe de sítio I)

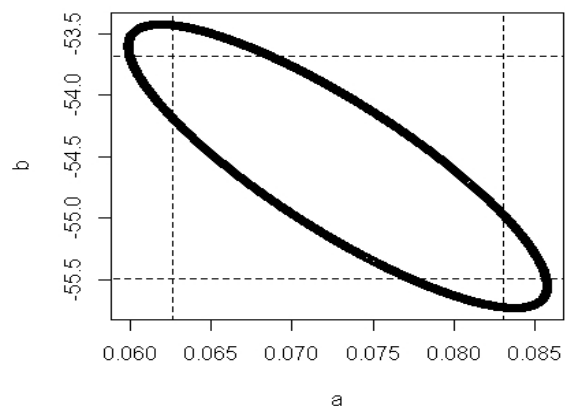


Figura 18A Elipses de confiança dos parâmetros modelo de crescimento em volume (proposto/dados da classe de sítio II)

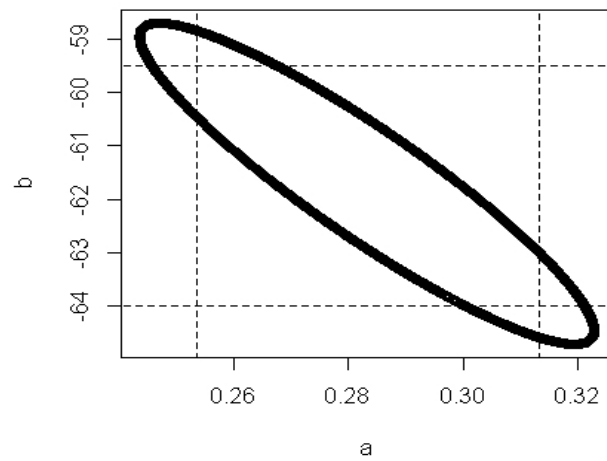


Figura 19A Elipses de confiança dos parâmetros modelo de crescimento em volume (dados da classe de sítio I)

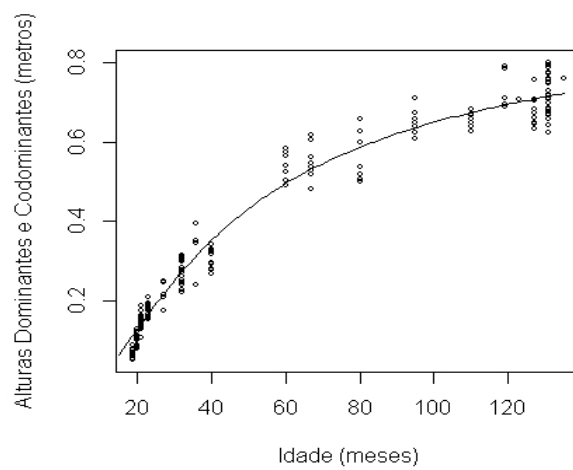


Figura 20A Curva de ajuste dos dados da classe de sítio I

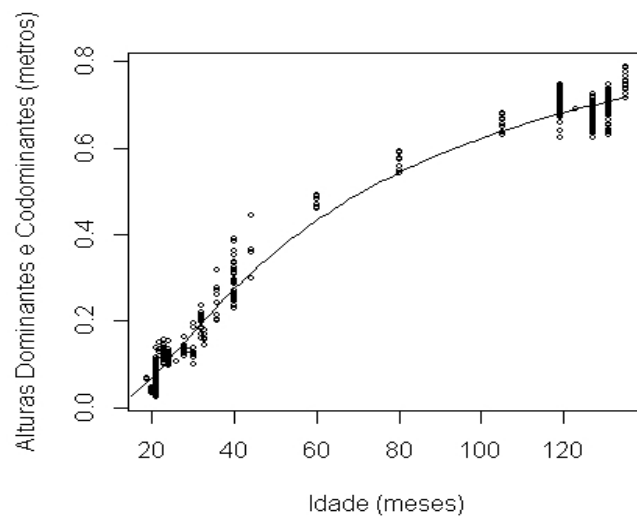


Figura 21A Curva de ajuste dos dados da classe de sítio II

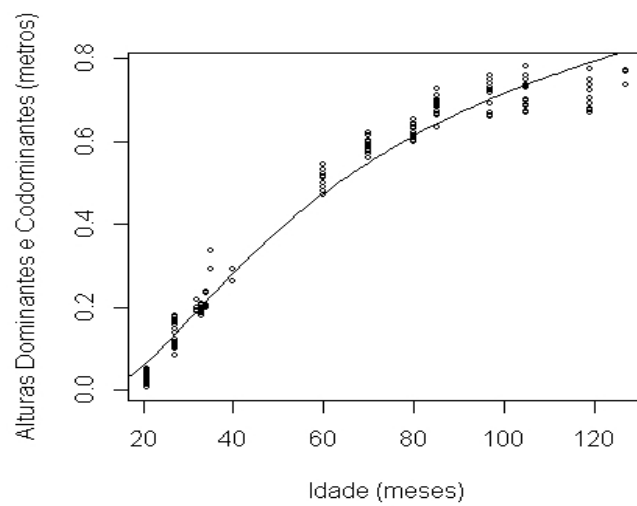


Figura 22A Curva de ajuste dos dados da classe de sítio III