



Revista Árvore

ISSN: 0100-6762

r.arvore@ufv.br

Universidade Federal de Viçosa  
Brasil

Cunha Silva da, Ulisses; Machado Amaral, Sebastião do; Filho Figueiredo, Afonso  
Uso de análise exploratória de dados e de regressão robusta na avaliação do crescimento de  
espécies comerciais de terra firme da Amazônia  
Revista Árvore, vol. 26, núm. 4, julho-agosto, 2002, pp. 391-402  
Universidade Federal de Viçosa  
Viçosa, Brasil

Disponível em: <http://www.redalyc.org/articulo.oa?id=48826401>

- Como citar este artigo
- Número completo
- Mais artigos
- Home da revista no Redalyc

redalyc.org

Sistema de Informação Científica  
Rede de Revistas Científicas da América Latina, Caribe, Espanha e Portugal  
Projeto acadêmico sem fins lucrativos desenvolvido no âmbito da iniciativa Acesso Aberto

# USO DE ANÁLISE EXPLORATÓRIA DE DADOS E DE REGRESSÃO ROBUSTA NA AVALIAÇÃO DO CRESCIMENTO DE ESPÉCIES COMERCIAIS DE TERRA FIRME DA AMAZÔNIA<sup>1</sup>

Ulisses Silva da Cunha<sup>2</sup>, Sebastião do Amaral Machado<sup>3</sup> e Afonso Figueiredo Filho<sup>4</sup>

**RESUMO** - Esta pesquisa originou-se de uma base de dados de dois inventários a 100%, realizados em 1984 e 2000 sobre uma mesma área de 576 ha de floresta tropical primária, localizada na Floresta Nacional do Tapajós, Belterra, Pará. O objetivo da pesquisa foi utilizar a análise exploratória de dados e a regressão robusta para modelar o crescimento em diâmetro e área basal. Nos dois inventários as circunferências à altura do peito (CAP) foram medidas com fitamétrica, enquanto as alturas comerciais em 1984 foram medidas com o hipsômetro de Weiss e em 2000, estimadas com o uso de varas; o DAP mínimo em 1984 foi de 55 cm para todas as espécies e em 2000 foi adotado o DAP mínimo de 35 cm. As análises estatísticas foram precedidas de análise exploratória de dados (AED), em que foram utilizados o *box plot* (caixa-de-bigodes) na detecção de *outliers* (observações discrepantes) e o gráfico *stem-and-leaf* (tronco-e-folhas) para filtrar as observações extremas. Utilizou-se a regressão robusta para ajustar os modelos na presença de *outliers*. A pesquisa mostrou que, apesar das variações intra e interespecíficas, as seis espécies, em conjunto, apresentaram taxas de crescimento medianas observadas e esperadas em diâmetro e área basal de 4,56 e 4,28 mm/ano e 13,00 e 13,09 cm<sup>2</sup>/ha/ano, respectivamente. A pesquisa também mostrou que o uso da análise exploratória de dados e da regressão robusta viabilizou a análise e a determinação dos incrementos periódicos em diâmetro e área basal em bases consistentes. A metodologia empregada no caso específico dos dados disponíveis e tipos de variáveis mostrou ser mais eficiente com o uso de regressão robusta pelo método *Least Trimmed Square* do que pelo método dos mínimos quadrados ordinários.

**Palavras-chave:** Incremento diamétrico e em área basal, análise exploratória de dados, regressão robusta e floresta tropical.

## USE OF EXPLORATORY DATA ANALYSIS AND ROBUST REGRESSION TO EVALUATE THE GROWTH OF COMMERCIAL SPECIES IN UPLAND TROPICAL AMAZONIAN FOREST

**ABSTRACT** - This research originated from a data base of two forest inventories at 100%, carried out in 1984 and 2000 on an area of 576 hectares of an upland primary forest, located in the National Forest of Tapajós, in Belterra, Pará, Brazil. The objective of the research was to use exploratory data analysis and robust regression for modeling diameter and basal area growths. In the two inventories, circumference breast height (CBH)-cm, was measured with a common tape; the commercial heights were measured with a Weise hypsometer in 1984 and, estimated with the use of sticks in 2000. The minimum DBH was 55 cm in 1984 and 35 cm in 2000, for all species. The statistical analyses were preceded by exploratory data analysis (EDA), where the box plot was used for outliers detection, and the stem-and-leaf graph to filter the extreme observations. Robust regression was used to adjust the models in the outliers presence. This research showed that, in spite of the intra and inter specific

<sup>1</sup> Recebido para publicação em 1.10.2001.

Aceito para publicação em 20.6.2002.

<sup>2</sup> Professor do Curso de Engenharia Florestal da Universidade Federal do Amazonas, Doutorando em Manejo Florestal pela Universidade Federal do Paraná, <uscfaa@ig.com.br/ucunha@fua.br>, Fone/Fax: (0xx92) 647-4042/647-4043-, Av. Gal. Rodrigo Otávio J. Ramos, 3000, Aleixo, Mini-Campus, 69077-000 Manaus-AM. <sup>3</sup> Professor Senior, Dr. em Manejo Florestal, do Curso de Engenharia Florestal da Universidade Federal do Paraná, <sammac@floresta.ufpr.br>, PQ 1A - CNPq. <sup>4</sup> Professor Senior, Dr. em Manejo Florestal, do Curso de Engenharia Florestal da Universidade Federal do Paraná. <afig@floresta.ufpr.br>, PQ 2A - CNPq.

*variations the six species together presented an observed and expected growth rate in diameter and basal area of 4.56 and 4.28 mm/year and 13.00 and 13.09 cm<sup>2</sup>/ha/year, respectively. It was also shown that the use of exploratory data analysis and robust regression made possible the analysis and determination of periodic increments in diameter and basal area in a consistent way. The methodology used for the available data and variable types was found to be more efficient when using robust regression for the method Least Trimmed Square than for the ordinary minimum square method.*

**Key words:** Diameter and basal area increment, exploratory data analysis, robust regression and tropical forest.

## 1. INTRODUÇÃO

A Floresta Nacional do Tapajós reúne um dos maiores acervos em pesquisas florestais da Amazônia, seja de natureza científica ou técnica. Mesmo considerando os avanços da pesquisa florestal na Amazônia durante a última década, vários componentes do ecossistema ainda necessitam ser investigados. Como exemplo pode-se citar a falta de melhor conhecimento sobre a dinâmica florestal, que depende de um período mais longo de observação, quase sempre exigindo o estabelecimento de parcelas permanentes.

Apesar de serem mais eficientes para estudos de dinâmica e mudanças estruturais em uma floresta, nem sempre existem dados disponíveis de parcelas permanentes com a qualidade exigida para o desenvolvimento de modelos de crescimento e produção, pois, em muitos casos, existem problemas de procedimento de amostragem na condução e no acompanhamento das medições, perda de dados, abandono total ou parcial ou até eventuais distúrbios naturais sobre determinadas parcelas, o que dificulta o controle e a recuperação de dados.

A busca de fontes alternativas de obtenção e geração de dados para análise deve ser um desafio constante dos pesquisadores. Neste sentido, considerando que normalmente os inventários a 100%, em escala comercial, podem abranger áreas em módulos de 50, 96, 100 ou mais hectares, a disponibilidade de dados em áreas desse porte é um aspecto positivo em termos de potencial de pesquisas.

Em florestas tropicais, o crescimento ao longo de um período de anos tem três componentes básicos: i) incremento individual da árvore; ii) mortalidade; e iii) ingresso ou aparecimento de novas árvores em tamanho mensurável das classes de DAP em função da regeneração (Alder & Synnott, 1992).

De acordo com Nist (2001), a análise exploratória de dados (AED) emprega uma variedade de técnicas gráficas para: 1) maximizar o conhecimento sobre um conjunto de dados; 2) investigar hipóteses sobre a estrutura dos dados; 3) detectar *outliers*; 4) examinar relações entre variáveis; e 5) analisar aspectos distributivos. Segundo Ellison (1993), a AED é mais apropriada quando existe grande variabilidade nos dados e quando os efeitos dos tratamentos não são imediatamente aparentes.

Segundo S-Plus (1997), as técnicas de regressão robusta são um importante complemento às técnicas clássicas de quadrados mínimos, uma vez que fornecem respostas similares aos quadrados mínimos quando existe relação linear entre variáveis com os erros normalmente distribuídos, porém diferem significativamente dos ajustes dos quadrados mínimos quando os erros não satisfazem as condições de normalidade ou quando os dados contêm *outliers* significantes.

O objetivo da pesquisa foi utilizar a análise exploratória de dados e a regressão robusta para modelar o crescimento em diâmetro e área basal.

## 2. MATERIAL E MÉTODOS

A área utilizada na presente pesquisa, com destaque na Figura 1, é quadrada, com 2,4 km de lado, tem 576 ha, é composta de seis blocos contíguos de 96 ha de floresta primária, cada um medindo 800 m no sentido leste-oeste e 1.200 m no sentido norte-sul, e faz parte da Floresta Nacional do Tapajós (FLONA TAPAJÓS), que tem 600.000 ha e está situada entre os paralelos de 2°40' a 4°10' de latitude sul e os meridianos de 54°45' a 55°30' de longitude oeste de Greenwich, município de Belterra, Estado do Pará. O seu manejo está previsto para 2001, como parte do programa a ser executado pela TREVISIO AGROPECUÁRIA LTDA., através de convênio de cooperação entre o Instituto Brasileiro do Meio Ambiente

e dos Recursos Naturais Renováveis (IBAMA) e a *International Tropical Timber Organization* (ITTO), em andamento desde 1999.

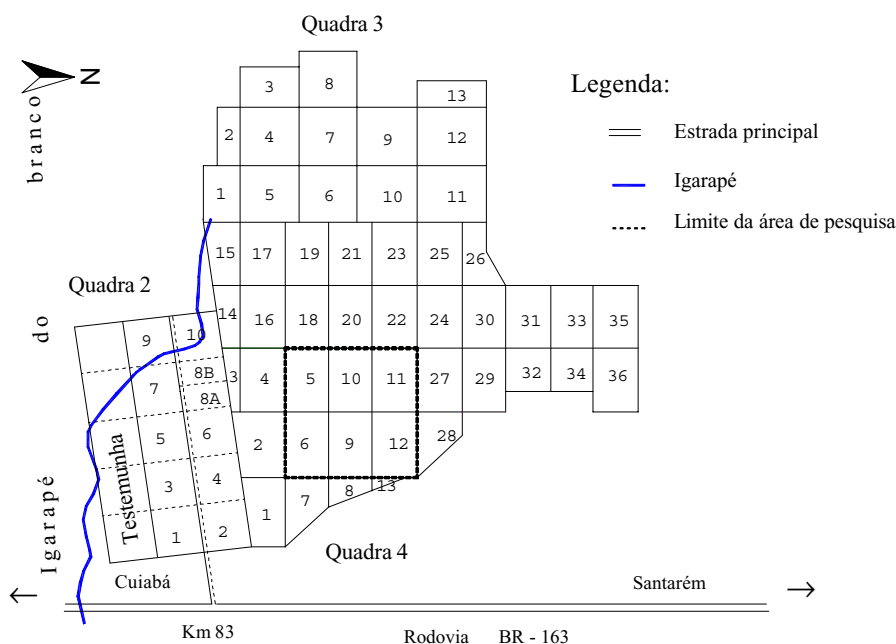
A Floresta Ombrófila Densa, subdividida pelos quadrantes ecológicos (litologia, morfologia e clima), em termos de aspectos fisionômicos e florísticos, caracteriza a tipologia florestal dominante na região. Apresenta-se também na área uma fisionomia florestal densa, com indivíduos emergentes ocupando as áreas aplainadas dos platôs do Xingu-Tapajós, caracterizada não só por apresentar alto volume, como também pela presença de espécies de alta cotação de preços nos mercados externo e interno (IBAMA/PNUD, 1997).

## 2.1. Base de Dados

A partir de informações de dois inventários, realizados em períodos diferentes, sobre uma mesma área de 576 ha, desenvolveu-se uma metodologia de recuperação, de análise e de tratamento estatístico desses dados, visando avaliar a produção presente e prever a produção futura com margem aceitável de erro.

Os dados são provenientes de duas fontes distintas, mas referentes à mesma área: 1) inventário a 100% (FUPEF, 1984); e 2) inventário a 100% (TREVISÓ, 2000). Em 1984, as equipes da FUPEF percorreram as linhas de inventário distribuídas a cada 100 m, cobrindo sempre o lado de uma das faixas de 50 m, dispostas no sentido norte-sul. Em 2000, as equipes da TREVISÓ percorreram as linhas de inventário distribuídas a cada 50 m, com o caminhar simultâneo de duas equipes por linha, recobrindo duas faixas laterais internas de 25 m, ora se deslocando para dentro ora para fora da linha-base, situada em posição frontal às linhas de inventário.

Os dois inventários foram previamente processados, e a partir das coordenadas de locação das árvores ( $x_i, y_i$ ) foram elaborados dois tipos de mapas por espécie, objetivando cruzar informações que permitissem recuperar por diferença diamétrica, numeração seqüencial e propriedade de vizinhança a mesma árvore abordada pelos dois inventários. O primeiro, contendo informações sobre o diâmetro à altura do peito (DAP) de cada indivíduo, em centímetro, e o segundo, contendo a numeração atribuída pelas equipes em cada inventário.



Fonte: FUPEF (1984); IBAMA & PNUD (1997).

**Figura 1** – Localização e delimitação da área de pesquisa.  
**Figure 1** – Location and delimitation of the research area.

Inicialmente, procurou-se identificar e marcar os pares de árvores mais próximas, observando e comparando os DAP medidos nas duas ocasiões. Em seguida, os pares marcados foram transferidos para o segundo mapa, contendo a numeração dos dois inventários. O passo seguinte consistiu na ordenação dos dados árvore-a-árvore, de acordo com a respectiva numeração em cada ocasião, visando visualizar o comportamento das mesmas variáveis durante o período 1984-2000.

Nos dois inventários as circunferências à altura do peito (CAP) foram medidas com fitamétrica, enquanto as alturas comerciais em 1984 foram medidas com o hipsômetro de Weiss e em 2000 estimadas com o uso de varas; o DAP mínimo em 1984 foi de 55 cm e em 2000 adotou-se o DAP mínimo de 35 cm. Em 1984 foram inventariados 100% dos indivíduos com  $DAP \geq 55$  cm, o que permitiu que 210 espécies fossem identificadas. Em 2000, foram inventariados 100% dos indivíduos com  $DAP \geq 35$  cm, à exceção de 30 espécies estabelecidas em um acordo técnico entre IBAMA/TREVISÓ, dentre as quais se destacam a envira-preta (*Guatteria poeppigiana* Mart.) e o tachi-preto (*Tachigalia paniculata* Aubl.).

Os dados foram analisados a partir da base diamétrica de 55 cm, por ter sido o valor mínimo de referência utilizado em 1984. Para os estudos realizados foram escolhidas as seguintes espécies: 1) *Minquartia guianensis* Aubl. (acariquara); 2) *Carapa guianensis* Aubl. (andiroba); 3) *Bertholletia excelsa* Humb. & Bonpl. (castanheira); 4) *Goupia glabra* Aubl. (cupiúba); 5) *Manilkara huberi* (Ducke) Standl. (maçaranduba); e 6) *Bagassa guianensis* Aubl. (tatajuba). A presente lista é composta por espécies de importância comercial, aparentemente sem problemas de identificação, conforme relatado por EMBRAPA (2000). O principal critério de escolha baseou-se na facilidade de identificação por mateiros da região, tendo sido considerado que nas duas ocasiões diferentes mateiros não tiveram problemas na identificação.

Preferiu-se utilizar a mediana do DAP ( $d_1, \dots, d_{med}, \dots, d_n$ ) e da área transversal  $g$  ( $g_1, \dots, g_{med}, \dots, g_n$ ), em vez da média aritmética nas análises de crescimento, com base nas seguintes justificativas demonstradas por Mosteller & Tukey (1977): i) o uso da média aritmética deve ser precedido de análise exploratória, quando o tipo de dados apresentar *outliers* ou quando, por conveniência, for requerida a linearidade de um modelo de regressão; ii) a média aritmética pode apresentar estimativas com tendências para dados que não seguem uma distribuição normal; e iii) por ser comprovadamente o protótipo de uma

medida resistente, a mediana é usada com alta eficiência na maioria das situações de análise exploratória e quando os dados não seguem uma distribuição normal, o que reflete o caso das variáveis de crescimento em florestas tropicais. A opção pelo uso da mediana em sentido técnico deveu-se à impossibilidade de sua determinação em termos de observações individuais, o que exigiu que a área de estudo fosse dividida em módulos.

## 2.2. Outliers

Barnett & Lewis (1995) definiram *outlier* em um conjunto de dados como sendo uma observação que parece ser inconsistente com o conjunto de dados remanescentes. Os *outliers* podem indicar algumas características importantes sobre um modelo, como modelo incompatível com os dados e omissão de variáveis importantes. Para Draper & Smith (1981), a rejeição automática de *outliers* não é um procedimento correto e as regras propostas para rejeição de *outliers* devem incluir a reanálise sem essas observações, que, dependendo das circunstâncias, podem ser portadoras de informações vitais dos indivíduos de uma população.

Na rejeição de *outliers* teve-se a preocupação em diagnosticar os verdadeiros *outliers*, diferenciando-os das observações influentes como os *breakdown points* e *leverage points*. Gnanadesikan (1997) definiu *breakdown point* como a maior fração das observações em uma amostra, que podem ser valores extremos sem distorcer o valor do estimador. Statgraphics (1999) definiu *leverage point* como a estatística que mede o quanto cada coeficiente estimado poderá mudar se cada observação for removida dos dados e constatou que um ponto tem influência significativa quando o valor do seu *leverage* for maior do que três vezes o número de parâmetros dividido pelo número de observações.

## 2.3. Análise Exploratória de Dados (AED)

A AED é necessária antes de se explorar adequadamente a natureza dos dados e suas propriedades, sendo um complemento de suma importância que deve anteceder o uso de certas ferramentas estatísticas, como a regressão, pois evita erros grosseiros, e a análise parcial com a geração de resultados duvidosos. Fez-se a análise de dados por processo de monitoramento, visando revelar *outliers* através de técnicas gráficas como *stem-and-leaf* (tronco-e-folhas) e *box plot* (caixa-de-bigodes).

O *box plot* é uma excelente ferramenta, pois permitir visualizar a locação, a dispersão, a simetria, as barreiras de *outliers* e os *outliers*, independentemente da forma da distribuição de um conjunto de dados. Além disto, o *box plot* é construído com base na mediana e nos quartos associados à coleção de dados, o que o torna resistente a valores perturbadores dentro das barreiras de *outliers* e, conseqüentemente, atraente em análise exploratória de dados (Hoaglin et al., 1992).

O *stem-and-leaf* é usado para organizar e sumarizar valores numéricos, produzindo uma figura compacta das propriedades de um conjunto de dados sem teoria elaborada. Por regra, qualquer ponto que exceder dois intervalos interquartis (ex. de 0,25 a 0,75), a partir do quartil mais próximo, é considerado um valor extremo.

## 2.4. Regressão Robusta

Optou-se pelo uso da regressão robusta por ser esta considerada uma técnica robusta não somente com respeito aos *outliers*, mas também com relação aos pontos extremos, que são pontos no modelo matricial com excessiva influência sobre o resultado, e porque quanto maior o número de variáveis de um modelo, mais difícil se torna a identificação de *outliers* com o uso das técnicas de regressão clássicas. Utilizou-se a técnica robusta *Least Trimmed Squares* disponível em S-Plus 1997.

O método LTS, introduzido por ROUSSEEUW (1984), é altamente robusto para ajuste de modelo linear, uma vez que, se aproximadamente a metade dos dados com referência à mediana apresentar *outliers*, o estimador LTS ajustará bem uma massa de dados com cerca de 50% das observações remanescentes.

Segundo S-PLUS (1997), a estimativa de  $\hat{\beta}_{LTS}$  por LTS minimiza a soma do menor quadrado de resíduo de  $q$  dado por  $\sum_{i=1}^q r_{(i)}^2(\beta)$ , em que  $r_{(i)}^2(\beta)$  é o  $i$ -ésimo resíduo ordenado. O valor de  $q$  é freqüentemente maior do que a metade de  $n$ , ou seja, o número de observações. Por contraste, a estimativa por mínimos quadrados ordinários de  $\hat{\beta}_{MQO}$  minimiza a soma de todos os quadrados de resíduos dados por  $\sum_{i=1}^n r_{(i)}^2(\beta)$ .

Comprovada a existência de diferentes fontes de variação nos dados: erros de medidas, erros de execução

e variação inerente, com potencial de produzirem *outliers* de natureza determinística ou aleatória, optou-se pela identificação e rejeição controlada de variações severas, seguidas pela acomodação (LTS) e incorporação (adaptação do modelo aos dados), conforme recomenda Barnett & Lewis (1995). A rejeição incidiu somente sobre as observações como sendo potenciais *outliers*, onde apenas uma fração dos dados caracterizados como observações de alta influência foi removida, usando o critério *Trimmed* do gráfico *stem-and-leaf*.

Optou-se por eliminar apenas os pontos com valores extremos, embora tenha sido possível identificar na estrutura interna dos dados os pontos de mínimos como sendo potenciais *outliers*, uma vez que, dependendo das circunstâncias em uma floresta primária, alguns indivíduos podem apresentar incrementos mínimos, nulos ou até mesmo negativos.

## 2.5. Modelos de Incremento em Diâmetro e Área Basal

As análises foram aplicadas às variáveis DAP= $d$ , área basal ( $g$ ) e seus incrementos. Além disto, foram utilizadas outras variáveis auxiliares como locação ( $x_i, y_i$ ), número da árvore e altura comercial média.

O incremento periódico do diâmetro e da área basal foi calculado da seguinte forma:

$$ipdmed = dmed00 - dmed84 \quad [1]$$

$$ipgmed = gmed00 - gmed84 \quad [2]$$

Foram utilizados os seguintes modelos para incremento em diâmetro e área basal, com base em Alder (1995):

### Diâmetro

$$\ln(ipadmed) = \beta_1 \ln(dmed84) + \beta_2 dmed84 \quad \text{Vanclay} \quad [3]$$

### Área Basal

$$\ln(ipagmed) = \beta_1 \ln(gmed84) + \beta_2 gmed84 \quad \text{Power-Exponential} \quad [4]$$

em que  $ipadmed$  = incremento periódico anual do diâmetro, em mm;  $dmed00$  = diâmetro mediano em 2000, em cm;  $dmed84$  = diâmetro mediano em 1984, em cm;  $ipagmed$  = incremento periódico anual da área basal, em cm<sup>2</sup>/ha;  $gmed00$  = área basal mediana em 2000, em m<sup>2</sup>/ha;  $gmed84$  = área basal mediana em 1984, em m<sup>2</sup>/ha; e  $\beta_1, \beta_2$  = parâmetros a serem estimados.

## 2.6. Análises Estatísticas

As análises estatísticas foram precedidas pela AED, por duas razões principais: 1) suspeita de possíveis erros de medição do CAP; e 2) presença de *outliers* decorrentes dos processos de medição utilizados ou devidos à variação biológica dos fatores de sítio, como solo, espécie etc. De todos esses erros, a posição de medida do CAP é a mais importante, por variar em cada ocasião nos inventários comerciais e por ser fixa em remediações de parcelas permanentes.

O erro-padrão da estimativa recalculado  $EPE_{rec}$  em % foi obtido por:

$$EPE_{rec} \% = 100 \times \frac{\sqrt{\frac{\sum (Y_i - \hat{Y}_i)^2}{n - p}}}{\bar{Y}_i} \quad [5]$$

em que  $EPE_{rec} \%$  = erro-padrão da estimativa, recalculado em porcentagem;  $Y_i$ ,  $\hat{Y}_i$  e  $\bar{Y}_i$  são, respectivamente, valor observado, valor estimado e média aritmética da variável resposta;  $n$  = número de observações; e  $p$  = número de parâmetros do modelo.

Para cálculo das medianas, os dados foram organizados em classes de área, dividindo-se a área de 576 ha em módulos de 4 ha, cujo tamanho foi considerado adequado por permitir um total de até 144 observações potenciais para uso nos modelos de regressão. Em cada módulo de 4 ha foi calculada a mediana do diâmetro e da área basal, assim como de seus respectivos incrementos periódicos. A coleção de dados resultantes foi utilizada no ajuste das equações [3] e [4]. O método LTS foi programado para monitorar cerca de 10% das observações com possíveis *outliers*, valor este considerado satisfatório em função da quantidade de potenciais *outliers* por espécie, revelados após a utilização do gráfico *stem-and-leaf*. A qualidade final do ajuste foi avaliada pelo coeficiente de determinação ( $R^2$ ), pelo erro-padrão da estimativa recalculado ( $EPE_{rec} \%$ ) e pela análise gráfica de resíduos.

## 3. RESULTADOS E DISCUSSÃO

Utilizaram-se os *box plot* por espécie para melhor visualização das principais diferenças entre os incrementos periódicos por classes de DAP (cm) e possíveis *outliers*.

Observou-se que os potenciais *outliers* podem ser devido à diversidade biológica dos fatores de crescimento e competição, característicos de uma floresta em estágio clímax, ou podem ser erros de medição. Tanto a variabilidade biológica quanto o erro de medição são comuns em levantamentos deste tipo, o que torna difícil discriminar os erros de medição, pois alguns tipos de *outliers* são inerentes da população. Neste caso, dispõe-se da possibilidade de identificar os *outliers* mais severos através da detecção das observações extremas reveladas pelo gráfico *stem-and-leaf*. Tal fato permitiu remover uma parte significativa dos prováveis erros de medição, mantendo-se os demais *outliers* na análise.

Na etapa seguinte da análise utilizou-se a técnica de regressão robusta LTS, controlando cerca de 10% das observações detectadas e consideradas pelo algoritmo do LTS como sendo possíveis *outliers*, o que significa dizer que após a identificação dos *outliers* o LTS se utilizou de 90% dos dados remanescentes para fazer o ajuste dos parâmetros.

Para ilustrar os resultados obtidos em termos de *box plot* e gráfico de resíduos para os incrementos em diâmetro e área basal, foram escolhidas as espécies constantes nas Figuras 3, 4, 5 e 6.

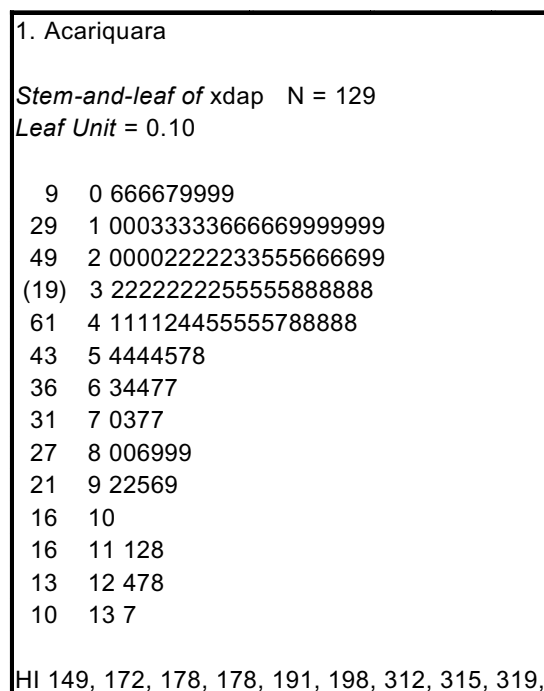
De acordo com os *high leverage points* indicados na Figura 2, foram eliminados da base de dados inicial da acariquara todos os valores acima de 14,8 cm de incremento periódico em diâmetro, com redução de 131 para 122 observações individuais e de 81 para 74 observações computadas pela mediana.

De acordo com a Figura 3(a), observa-se que mesmo após a remoção dos valores extremos os potenciais *outliers* de natureza biológica continuam presentes, exceto na classe 90, em que se verifica a maior variância do incremento periódico.

Por meio do gráfico de resíduos ilustrado pela Figura 3(b), observa-se a tendência inicial mais forte para subestimativas, acompanhada de variações positivas mais acentuadas, o que retrata que nestas condições as estimativas não são estáveis ao longo das classes diamétricas.

As observações discrepantes, constantes na Figura 4(a), mostram que mesmo após a remoção dos possíveis erros de medição resultantes dos incrementos anormais os potenciais *outliers* estão presentes em classes alternadas, inclusive em classes menores, que são mais fáceis de ser medidas, o que se deve ao erro sistemático ou

aleatório. As maiores variâncias em termos de incremento periódico ocorreram na classe 70-80, o que evidencia que nesta classe a castanheira apresenta maiores níveis de incremento.



**Figura 2** – Ilustração do modo de exibição do gráfico *stem-and-leaf*.

**Figure 2** – Illustration of the *stem-and-leaf* graphic configuration.

A Figura 4(b) apresenta um gráfico de resíduos mais bem estruturado, principalmente na faixa central das classes diamétricas, com a maioria das estimativas centradas na faixa de -1,0 a +1,0 desvio-padrão, o que demonstra a boa qualidade do ajuste. O comportamento da Figura 5(a) é idêntico ao da Figura 3(a), diferindo em termos de números de observações discrepantes, que neste caso é menor.

A Figura 5(b) mostra que os resíduos em termos de área basal estão mais bem distribuídos e aparentam melhor estabilidade, com poucas variações fluando acima de +2,0 desvios.

Na Figura 6(a) poucos *outliers* remanescentes podem ser observados, mesmo considerando a maior variação dos dados na classe 95-100.

A distribuição dos resíduos mostrados pela Figura 6(b) permite visualizar um ajuste de excelente qualidade e praticamente sem tendências em função da uniformidade dos pontos em torno da média.

De acordo com o Quadro 1, o ajuste dos modelos [3] e [4] pelo método dos mínimos quadrados ordinários (MQO) produziu erros maiores do que os resultados obtidos pelo método LTS e, em alguns casos, resultou em valores quase 100% maiores. Apenas com base na avaliação prévia do erro de estimativa pode-se afirmar que a metodologia adotada, além de preservar a propriedade dos dados, permitiu obter estimativas mais precisas.

Os resultados do Quadro 1, obtidos pelo método LTS, mostram que para os ajustes do modelo [3] do incremento em diâmetro a cupiúba foi a que apresentou o pior resultado entre as espécies, tendo produzido um erro estimado de 8,72%, enquanto para a maçaranduba obteve-se o melhor resultado, com erro de apenas 3,70% contra um erro de 6,80%, caso o ajuste tivesse sido feito por mínimos quadrados ordinários. Enquanto isso, os ajustes do modelo [4] do incremento em área basal revelaram que a maçaranduba, com 4,56%, superou as demais espécies, com pequena diferença em relação a andiroba, cujo erro estimado foi de 4,67%. De todas as espécies, apenas a tatajuba apresentou erro superior a 10%, relativo ao modelo [4].

No Quadro 2 estão mostrados os valores observados e os resultados das estimativas das taxas de crescimento por espécie, com base nos parâmetros dos modelos [3] e [4] indicados no Quadro 1, ajustados pelo método de regressão robusta LTS. O valor mínimo esperado para taxa de crescimento em diâmetro e área basal variou de 1,70-2,72 mm/ano e 6,37-6,83 cm<sup>2</sup>/ha/ano, respectivamente, para acariquara, espécie considerada de crescimento lento. Sem as estimativas da última classe, a espécie que mais se destacou em termos de crescimento foi a castanheira, com valores mínimos e máximos esperados em diâmetro e área basal de 5,27-7,98 mm/ano e 22,51-36,70 cm<sup>2</sup>/ha/ano, respectivamente.

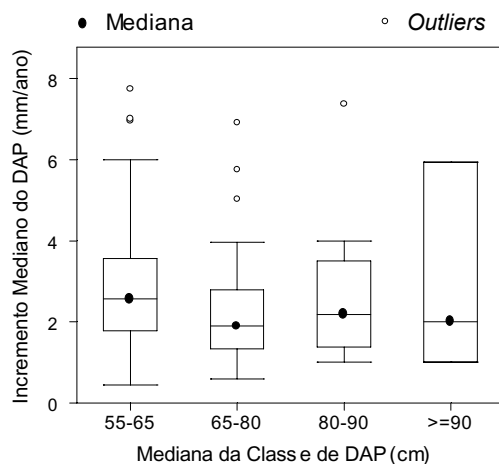
De acordo com o Quadro 2, as espécies cupiúba, maçaranduba e tatajuba foram as que apresentaram superestimativas nas taxas de crescimento em área basal (cm<sup>2</sup>/ha/ano), na última classe de DAP, resultando nos seguintes pares de valores (*Obs.*, *Est.*): (15,88; 38,04), (11,38; 45,92) e (15,74; 58,25), respectivamente. Tais resultados estão compatíveis com a análise gráfica de resíduos, conforme pode ser notado pela dispersão da Figura 6(b), referente à maçaranduba, não indicando falta



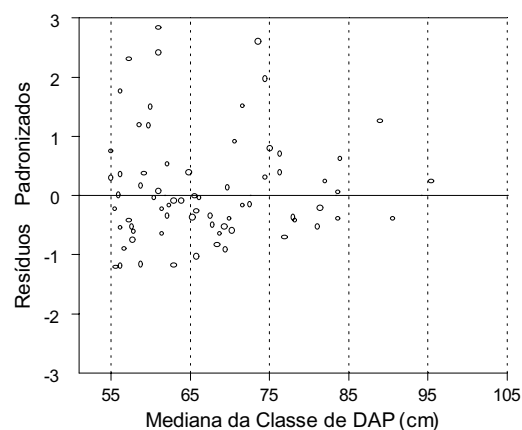
de consistência no ajuste, mas perda de precisão na determinação das estimativas na última classe diamétrica, o que na prática é um resultado esperado.

Com base nos resultados do Quadro 2, obteve-se a seguinte disposição em ordem decrescente por espécie,

das maiores taxas de crescimento diamétrico em condições de clímax, sem considerar a última classe de DAP: castanheira (7,98 mm/ano); tatajuba (5,98 mm/ano); cupiúba (5,75 mm/ano); maçaranduba (5,26 mm/ano); andiroba (3,65 mm/ano); e acariquara (2,72 mm/ano).



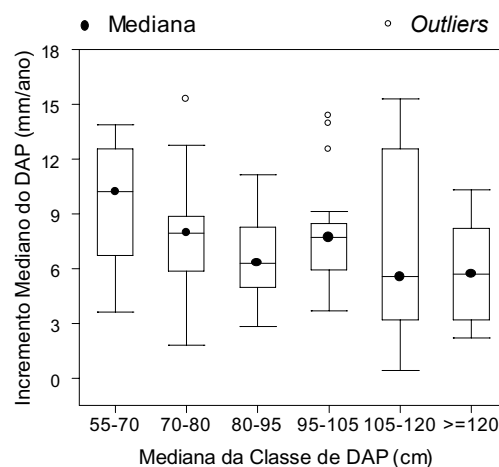
(a)



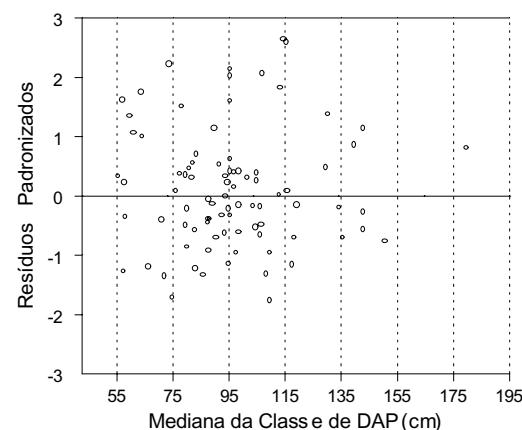
(b)

**Figura 3** – Box plot do incremento periódico com potenciais outliers (a) e distribuição de resíduos por classe de DAP em cm (b) para *Minquartia guianensis* Aubl. (acariquara).

**Figure 3** – Box plot of the periodic increment with potential outliers (a) and residual distribution per diameter class-DBH in cm (b) for *Minquartia guianensis* Aubl. (acariquara).



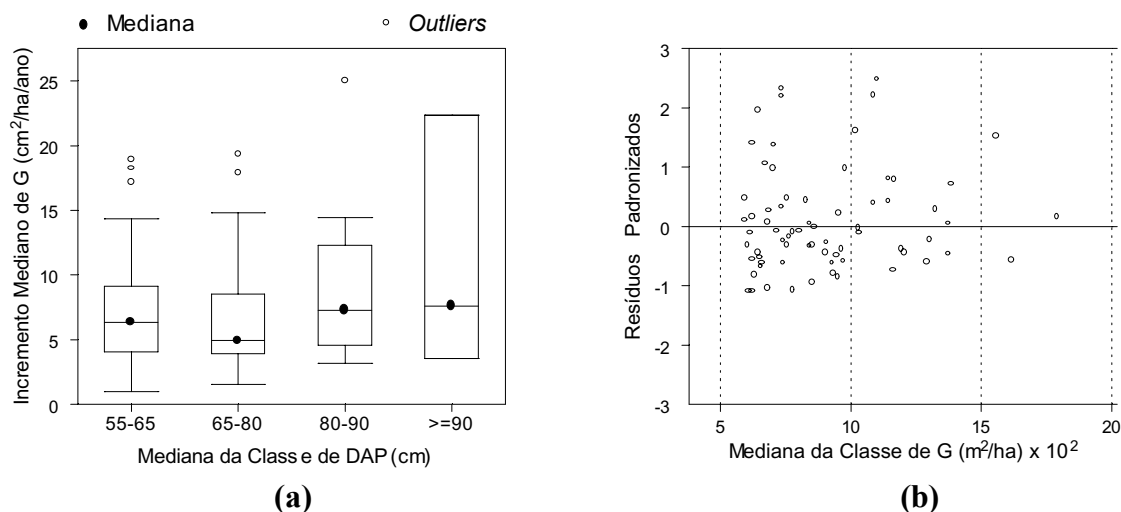
(a)



(b)

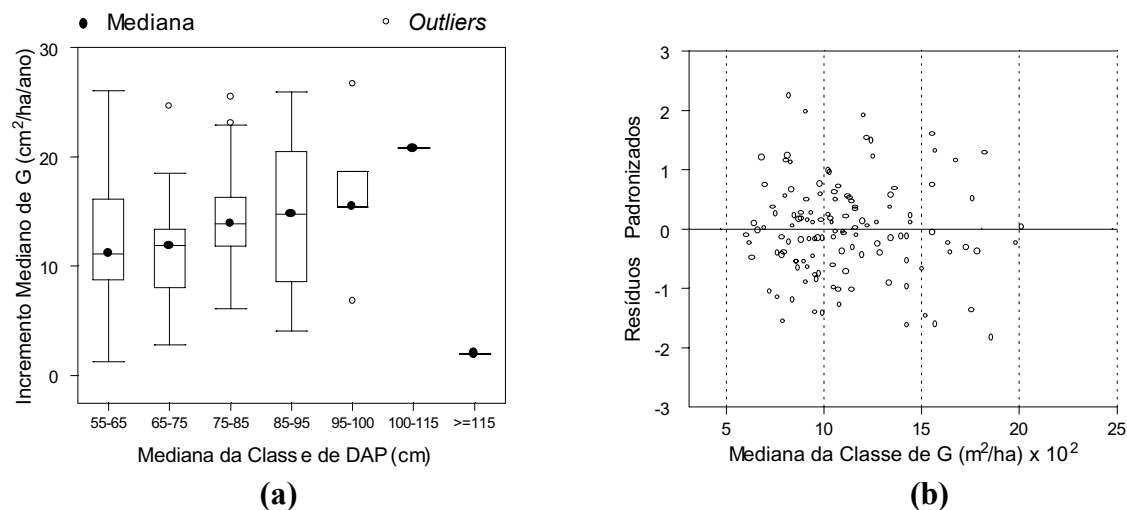
**Figura 4** – Box plot do incremento periódico com potenciais outliers (a) e distribuição de resíduos por classe de DAP em cm (b) para *Bertholletia excelsa* Humb. & Bonpl. (castanheira).

**Figure 4** – Box plot of the periodic increment with potential outliers (a) and residual distribution per diameter class-DBH in cm (b) for *Bertholletia excelsa* Humb. & Bonpl. (castanheira).



**Figura 5** – Box plot do incremento periódico com potenciais outliers por classe de DAP em cm (a) e distribuição de resíduos por classe de G em  $\text{m}^2/\text{ha}$  (b) para *Minquartia guianensis* Aubl. (acariquara).

**Figure 5** – Box plot of the periodic increment per diameter class-DBH in cm with potential outliers (a) and residual distribution per basal area class-G in  $\text{m}^2/\text{ha}$  (b) for *Minquartia guianensis* Aubl.(acariquara).



**Figura 6** – Box plot do incremento periódico por classe de DAP em cm (a) e distribuição de resíduos por classe de G em  $\text{m}^2/\text{ha}$  (b) para a *Manilkara huberi* (Ducke) Standl. (maçaranduba).

**Figure 6** – Box plot of the periodic increment with potential outliers per diameter class-DBH em cm (a) and residual distribution per basal area class-G in  $\text{m}^2/\text{ha}$  (b) for *Manilkara huberi* (Ducke) Standl.(maçaranduba).

Dispondo em ordem decrescente os valores do Quadro 2, referentes às maiores taxas de crescimento em área basal por espécie, sem considerar a última classe de DAP, foram obtidas para a castanheira (36,70  $\text{cm}^2/\text{ha}/\text{ano}$ ); cupiúba (20,91  $\text{cm}^2/\text{ha}/\text{ano}$ ); tatajuba (28,16  $\text{cm}^2/\text{ha}/\text{ano}$ );

maçaranduba (20,59  $\text{cm}^2/\text{ha}/\text{ano}$ ); andiroba (10,01  $\text{cm}^2/\text{ha}/\text{ano}$ ); e acariquara (6,83  $\text{cm}^2/\text{ha}/\text{ano}$ ). Os valores medianos observados e esperados em crescimento diamétrico e área basal, para as seis espécies, foram 4,56 e 4,28 mm/ano e 13,00 e 13,09  $\text{cm}^2/\text{ha}/\text{ano}$ , respectivamente.

**Quadro 1** – Parâmetros estimados e estatísticas por modelo e por espécie**Table 1** – Estimated parameters and statistics by model and by species

Espécie	MQO				LTS			
	$b_1$	$b_2$	$R^2$	$EPE_{rec} \%$	$b_1$	$b_2$	$R^2$	$EPE_{rec} \%$
$\ln(ipadmed) = b_1 \ln(dmed84) + b_2 dmed84$								
Acariquara	0,3077	-0,0054	0,584	9,63	0,5689	-0,0224	0,647	6,79
Andiroba	0,4090	-0,0076	0,781	6,78	0,4549	-0,0095	0,881	4,47
Castanheira	0,6994	-0,0133	0,914	5,56	0,6901	-0,0125	0,954	5,14
Cupiúba	0,3515	0,0025	0,810	8,98	0,5579	-0,0089	0,858	8,72
Maçaranduba	0,5462	-0,0120	0,835	6,80	0,3171	0,001	0,939	3,70
Tatajuba	0,5772	-0,0104	0,847	12,27	0,5976	-0,011	0,932	6,48
$\ln(ipagmed) = b_1 \ln(gmed84) + b_2 gmed84$								
Acariquara	-0,4553	9,0808	0,848	11,29	-0,5887	4,9791	0,909	7,11
Andiroba	-0,5601	9,4357	0,919	7,35	-0,5854	8,7754	0,960	4,67
Castanheira	-1,1590	6,1644	0,962	6,82	-1,1912	5,9628	0,979	6,47
Cupiúba	-0,7576	10,3853	0,914	11,18	-0,8488	8,0396	0,941	8,68
Maçaranduba	-0,7850	7,3301	0,935	7,60	-0,6537	9,8261	0,982	4,56
Tatajuba	-0,9601	6,4215	0,932	16,26	-0,8384	8,2522	0,967	14,44

em que  $b_1$ ,  $b_2$  = estimativas dos parâmetros dos modelos,  $R^2$  = coeficiente de determinação múltipla,  $EPE_{rec} \%$  = erro-padrão da estimativa, recalculado em %, MQO= mínimos quadrados ordinários e LTS= *least Trimmed Square*.

**Quadro 2** – Incrementos periódicos anuais do DAP (mm/ano) e da área basal G (cm<sup>2</sup>/ha/ano) no período 1984-2000, por espécie e classe de DAP (cm)**Table 2** – Periodic annual increments in DBH (mm/year) and basal area G (cm<sup>2</sup>/ha/year) during (1984-2000) by species and class of DBH (cm)

Espécie	$CD_{med}$	$ipadmed$		$ipagmed$		Espécie	$CD_{med}$	$ipadmed$		$ipagmed$	
		Obs.	Est.	Obs.	Est.			Obs.	Est.	Obs.	Est.
Acariquara	55-65	2,56	2,72	6,34	6,83	Andiroba	55-60	3,17	3,65	7,33	8,76
	65-80	1,89	2,33	4,94	6,40		60-65	2,94	3,62	7,41	8,81
	80-90	2,19	1,93	7,27	6,37		65-70	3,19	3,58	9,01	9,01
	$\geq 90$	2,00	1,70	7,58	6,54		70-75	3,19	3,52	9,23	9,35
Médias	→	2,16	2,17	6,53	6,53		75-80	4,78	3,45	15,63	10,01
							$\geq 80$	3,18	3,38	10,21	10,78
Castanheira	55-70	10,22	7,98	27,27	36,70	Médias	→	3,41	3,53	9,80	9,46
	70-80	8,36	7,65	29,21	25,94	Cupiúba	55-60	7,48	5,75	11,20	16,88
	80-95	6,31	7,32	24,98	23,37		60-70	9,20	5,75	15,68	15,94
	95-105	7,70	6,98	32,76	22,51		70-85	7,65	5,67	14,70	15,88
	105-120	5,56	6,36	21,87	23,17		85-95	9,20	5,51	22,53	17,28
	$\geq 120$	6,56	5,27	39,16	30,82		95-100	11,25	5,44	29,57	18,17
Médias	→	7,45	6,93	29,21	27,08		100-110	8,60	5,27	23,06	20,91
Maçaranduba	55-65	4,38	3,95	11,13	11,38		$\geq 110$	6,60	4,81	21,50	38,04
	65-75	3,88	4,12	11,88	11,91	Médias	→	8,90	5,57	19,46	17,51
	75-85	4,63	4,31	13,89	13,04	Tatajuba	55-60	8,95	5,98	13,65	16,84
	85-95	4,00	4,54	14,77	15,60		60-75	9,90	5,85	16,39	15,74
	95-100	4,00	4,69	15,42	18,29		75-90	7,90	5,70	18,39	16,02
	100-115	7,94	4,78	20,78	20,59		90-100	11,32	5,33	28,05	18,54
	$\geq 115$	0,38	5,26	1,97	45,92		100-140	6,05	4,82	17,37	28,16
Médias	→	4,17	4,52	12,83	19,53		$\geq 140$	4,10	4,04	14,50	58,25
						Médias	→	8,04	5,29	18,06	25,59

em que  $CD_{med}$  = classe de DAP mediana,  $ipadmed$  = incremento periódico anual do diâmetro mediano,  $ipagmed$  = incremento periódico anual da área basal mediana, Obs. = valores observados e Est. = valores estimados.

Esses resultados apresentam variações intra e interespecíficas, porém estão muito próximos de resultados encontrados por Chambers et al. (1998), que determinaram, com o uso do carbono 14, uma taxa de crescimento médio de 1,0-6,4 mm/ano para 13 espécies em condições clímax da Amazônia Central, dentre as quais inclui-se a *Bertholletia excelsa* Humb. & Bonpl. (castanheira).

Em termos de fatores que explicam certas interferências nos espaços de crescimento, é preciso considerar a forte presença de cipós em grande parte da área, algumas vezes dominando o sub-bosque, sendo certamente um fator de competição limitante para o crescimento. Mas esse é apenas um dos fatores que podem trazer implicações e causar interferências no processo de crescimento individual e da floresta. Dentre outros aspectos deve-se considerar também que: i) as condições de solo não são homogêneas; ii) a competição entre as árvores de menores diâmetros é mais intensa; e iii) dependendo do hábito da espécie, da iluminação de copa e da posição sociológica é possível observar indivíduos com incrementos maiores para diâmetros maiores, o que acontece porque as árvores de menores diâmetros não são favorecidas pela intensidade de luz na copa (posição sociológica), pois geralmente estão no estrato inferior ou médio onde os efeitos da competição por luz e nutrientes são mais sensíveis.

Os resultados mostraram que a eliminação de dados discrepantes em estudos de incrementos, independentemente da origem dos dados, seja de parcelas permanentes ou de inventários convencionais, deve ser precedida de criteriosa análise, devendo-se evitar remover dados dentro dos intervalos críticos que apresentem principalmente variâncias altas, com base apenas na hipótese de que podem se tratar de dados que contenham erros sistemáticos de medição ou remedição.

Em virtude de a variação dos incrementos ser um fato absolutamente normal, deve-se recorrer às técnicas disponíveis de tratamento de dados como análise exploratória de dados (AED) e regressão robusta, visando obter resultados consistentes com a validação das estimativas dos incrementos periódicos.

A combinação dos recursos gráficos e analíticos, associada à metodologia desenvolvida, permitiu analisar de maneira satisfatória os dados de incrementos com base nos dois inventários, o que é uma contribuição importante, levando-se em consideração que por falta de metodologias adequadas deixa-se de analisar certos fenômenos

relevantes da floresta, ou que em virtude de a maioria dos dados de florestas tropicais ainda ser proveniente de distribuições truncadas corre-se o risco de formular conclusões equivocadas.

#### 4. CONCLUSÃO

Os resultados obtidos na presente pesquisa permitiram mostrar que o uso da análise exploratória de dados e da regressão robusta viabilizou a análise e determinação dos incrementos periódicos em diâmetro e área basal, em bases consistentes.

A metodologia empregada no caso específico dos dados disponíveis e tipos de variáveis mostrou ser mais eficiente com o uso de regressão robusta pelo método *Least Trimmed Square* do que pelo método dos mínimos quadrados ordinários.

#### 5. REFERÊNCIAS BIBLIOGRÁFICAS

- ALDER, D. **Growth modeling for mixed tropical forests.** Oxford: Oxford Forestry Institute, 1995. 231 p. (Tropical Forestry Paper, 30)
- ALDER, D.; SYNNOTT, T. J. **Permanent sample plot techniques for mixed tropical forest.** Oxford: Forestry Institute, 1992. 24 p. (Tropical Forestry Paper, 25)
- ASSMANN, E. **The principles of forest yield study.** Oxford: Pergamon Press, 1970. 506 p.
- BARNETT, V.; LEWIS, T. **Outliers in statistical data.** Chichester: John Wiley, 1995. 584 p.
- CHAMBERS, J. Q.; HIGUCHI, N.; SCHIMEL, J.P. Ancient trees in Amazonian. **Nature**, n. 391, p. 135-136. 1998.
- DRAPER, N. R.; SMITH, H. **Applied regression analysis.** 2.ed. New York: John Wiley, 1981. 709 p.
- ELLISON, A. M. **Exploratory data analysis and graphic display: design and analysis of ecological experiments.** New York: Chapman & Hall, 1993. p. 14-41.
- FUNDAÇÃO DE PESQUISAS FLORESTAIS DO PARANÁ - FUPF. **Inventário comercial da quadra de exploração nº 4 da floresta nacional do Tapajós.** Curitiba: 1984. 429 p.

GNANADESIKAN, R. **Methods for statistical data analysis of multivariate observations**. New Jersey: John Wiley, 1997. 353 p.

HOAGLIN, D. C.; MOSTELLER, F.; TUKEY, J.W. **Análise exploratória de dados: técnicas robustas - um guia**. Salamandra: Almada, 1992. 446 p.

INSTITUTO BRASILEIRO DO MEIO AMBIENTE E DOS RECURSOS NATURAIS - IBAMA/PNUD. **Plano de manejo florestal da floresta nacional do Tapajós, Pará, Brasil**. Belém: 1997. 109 p.

MOSTELLER, F.; TUKEY, J. W. **Data analysis and regression: a second course in statistics**. Reading: Addison-Wesley Publishing Company, 1977. 588 p.

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY-NIST. **Statistics for Scientists & Engineers: exploratory data analysis**. Disponível em: <http://www.itl.nist.gov>.

ROUSSEEUW, P. J. Least median of squares regression. **Journal of the American Statistical Association**, n. 79, p. 871-880. 1984.

SILVA, J. N. M. et al. Equações de volume para a floresta nacional dos Tapajós. Curitiba, Embrapa, **Boletim de Pesquisa Florestal**, n. 8/9, jun/dez, 1984. p. 50-63.

S-PLUS. **S-Plus 1997: guide to statistics**, volume 1. 1997. 395 p. Disponível em: <http://www.mathsoft.com>.

STATGRAPHICS. **User guide**. 1999. Disponível em: <http://www.statgraphics.com>.