



**AJUSTE E DIAGNÓSTICO DE MODELOS
ESTOCÁSTICOS LINEARES E NÃO-LINEARES
PARA A DESCRIÇÃO DO PERFIL
LONGITUDINAL DE ÁRVORES**

LEONARDO MACHADO PIRES

2004

LEONARDO MACHADO PIRES

**AJUSTE E DIAGNÓSTICO DE MODELOS
ESTOCÁSTICOS LINEARES E NÃO-LINEARES
PARA A DESCRIÇÃO DO PERFIL
LONGITUDINAL DE ÁRVORES**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Curso de Mestrado em Engenharia Florestal, área de concentração em Floresta de Produção, para obtenção do título de "Mestre".

Orientador

Prof. Dr. Natalino Calegario

LAVRAS
MINAS GERAIS – BRASIL
2004

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

Pires, Leonardo Machado

Ajuste e diagnóstico de modelos estocásticos lineares e não-lineares para
descrição do perfil longitudinal de árvores / Leonardo Machado Pires. -- Lavras :
UFLA, 2004.

127 p. : il.

Orientador: Natalino Calegario.
Dissertação (Mestrado) – UFLA.
Bibliografia.

1. Modelo linear. 2. Modelo não-linear. 3. Dendrometria. I. Universidade
Federal de Lavras. II. Título.

CDD-634.9285

LEONARDO MACHADO PIRES

**AJUSTE E DIAGNÓSTICO DE MODELOS
ESTOCÁSTICOS LINEARES E NÃO-LINEARES
PARA A DESCRIÇÃO DO PERFIL
LONGITUDINAL DE ÁRVORES**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Curso de Mestrado em Engenharia Florestal, área de concentração em Floresta de Produção, para obtenção do título de "Mestre".

APROVADA em 13 de agosto de 2004

Prof. Dr. José Roberto S. Scolforo

Prof. Dr. Luiz Marcelo T. de Carvalho



Prof. Dr. Natalino Calegario -UFLA
(Orientador)

LAVRAS
MINAS GERAIS - BRASIL

À minha querida mãe
À minha adorável esposa
Aos meus irmãos

DEDICO

SUMÁRIO

| | Página |
|--|---------------|
| RESUMO..... | i |
| ABSTRACT..... | ii |
| CAPÍTULO 1 | 1 |
| 1.1 Introdução | 1 |
| 1.2 Referencial Teórico... .. | 2 |
| 1.2.1 Forma da Árvore | 2 |
| 1.2.2 Equações de Perfil | 4 |
| 1.2.3 Quadrado médio linear versus não-linear | 14 |
| 1.3 Referências Bibliográficas..... | 15 |
| CAPÍTULO 2: Ajuste de Modelos Estocásticos Lineares e Não-Lineares para a Descrição do Perfil Longitudinal de Árvores | 19 |
| Resumo..... | 19 |
| Abstract..... | 20 |
| 2.1 Introdução..... | 21 |
| 2.2 Referencial Teórico... .. | 22 |
| 2.2.1 Convergibilidade.... .. | 22 |
| 2.2.2 Medidas de Acuracidade..... | 23 |
| 2.2.3 Gráfico de Resíduo. | 26 |
| 2.3 Material e métodos..... | 27 |
| 2.3.1 Dados | 27 |
| 2.3.2 Modelos | 28 |
| 2.3.2.1 Modelos Lineares | 30 |
| 2.3.2.2 Modelos Não-Lineares | 33 |
| 2.3.3 Medidas de acuracidade dos Modelos..... | 33 |
| 2.4 Resultados e discussão..... | 35 |

| | |
|--|----|
| 2.4.1 Modelos Lineares..... | 35 |
| 2.4.2 Comparação dos Modelos Lineares e Não Lineares..... | 41 |
| 2.4.2.1 Modificação no Modelo Logístico..... | 55 |
| 2.4.2.2 Comparação dos Perfis Longitudinais das Árvores..... | 59 |
| 2.5 Conclusões..... | 62 |
| 2.6 Referências Bibliográficas..... | 63 |
| CAPÍTULO 3: Diagnóstico de Modelos Estocásticos Lineares..... | 65 |
| Resumo..... | 65 |
| Abstract..... | 66 |
| 3.1 Introdução..... | 67 |
| 3.2 Referencial Teórico..... | 69 |
| 3.2.1 Diagnóstico da regressão e medidas de adequação dos modelos | 69 |
| 3.2.1.1 Pontos de influência segundo critério de Hoaglin e Welsch (1978).. | 69 |
| 3.2.1.2 “Outliers” | 70 |
| 3.2.1.3 PRESS e R^2_{predito} | 70 |
| 3.2.1.4 Influência nos coeficientes da regressão | 72 |
| 3.2.1.4.1 Distância de Cook | 72 |
| 3.2.1.4.2 $DFBETAS_{j,i}$ | 73 |
| 3.2.1.5 Influência nos valores ajustados | 74 |
| 3.2.1.6 Influência na precisão da estimativa | 77 |
| 3.2.2 Condicionantes do resíduo | 78 |
| 3.2.3 Multicolinearidade | 81 |
| 3.3 Material e métodos..... | 83 |
| 3.3.1 Dados | 83 |
| 3.3.2 Modelos Polinomiais | 83 |
| 3.3.3 Diagnóstico da regressão..... | 85 |
| 3.3.4 Teste de Hipóteses dos Parâmetros | 85 |
| 3.4 Resultados e discussão..... | 86 |

| | |
|---|-----|
| 3.4.1 Diagnósticos dos Modelos Lineares..... | 86 |
| 3.4.2 Condicionantes do Resíduo..... | 92 |
| 3.4.3 Multicolinearidade..... | 97 |
| 3.4.4 Teste de Hipótese dos Parâmetros..... | 99 |
| 3.4.5 Validação..... | 100 |
| 3.4.6 Retirada das Observações de Influência..... | 108 |
| 3.5 Conclusões..... | 112 |
| 3.6 Referências bibliográficas..... | 113 |
| ANEXOS | 115 |

ÍNDICE DE TABELAS

| | Página |
|---|--------|
| TABELA 2.1 – Coordenadas geográficas aproximadas e extensão da área da empresa Aracruz Celulose S.A..... | 27 |
| TABELA 2.2 – Medidas de Acuracidade do polinômio de 5ºGrau para todas as classes diamétricas e para cada classe, usando d/dap como variável dependente. | 35 |
| TABELA 2.3 – Medidas de Acuracidade do polinômio de Potências Fracionárias e Inteiras (Hradetzky, 1976) para todas as classes diamétricas e para cada classe, usando d/dap como variável dependente..... | 36 |
| TABELA 2.4 – Medidas de Acuracidade do polinômio proposto por Pires & Calegario (2004) para todas as classes diamétricas e para cada classe, usando d/dap como variável dependente..... | 36 |
| TABELA 2.5 – Avaliação do gráfico de resíduo em cada seção..... | 40 |
| TABELA 2.6 – Parâmetros estimados dos modelos não-lineares de Gompertz, Logístico, Weibull, e os modelos lineares do 5ºGrau, Hradetzky, Pires & Calegario, no agrupamento das classes 2, 3 e 4, 5 e 6, 7 e 8..... | 41 |
| TABELA 2.7 – Medidas de acuracidade dos modelos lineares e não-lineares, no agrupamento das classes 2, 3, 4; 5 e 6; 7 e 8..... | 43 |
| TABELA 2.8 – Correlações entre os parâmetros estimados pelos modelos lineares e não-lineares, na agrupamento das classes 2, 3 e 4..... | 45 |
| TABELA 2.9 – Correlações entre os parâmetros estimados pelos modelos lineares e não-lineares, na agrupamento das classes 5 e 6..... | 46 |
| TABELA 2.10 – Correlações entre os parâmetros estimados pelos modelos lineares e não-lineares, no agrupamento das classes 7 e 8..... | 47 |
| TABELA 2.11 – Parâmetros estimados no modelo não linear Logístico no agrupamento das classes 2, 3 e 4, 5 e 6, 7 e 8..... | 56 |
| TABELA 2.12 – Correlações entre os parâmetros estimados no modelo não linear Logístico no agrupamento das classes 2, 3 e 4, 5 e 6, 7 e 8..... | 56 |
| TABELA 3.1 - Contagem do número de observações que exercem influência no ajuste dos polinômios, utilizando o critério de Hoaglin e Welsch..... | 86 |
| TABELA 3.2 - Contagem do número de “outliers” dos três polinômios nas várias classes diamétricas..... | 87 |
| TABELA 3.3 – Poder preditivo dos modelos polinomiais testados, utilizando como parâmetros o resíduo PRESS e o $R^2_{predito}$ | 88 |

| | |
|--|-----|
| TABELA 3.4 – Contagem do número de observações que influenciam os valores ajustados e a precisão da estimativa, para os três polinômios testados..... | 91 |
| TABELA 3.5 - Avaliação da homocedasticidade do Polinômio de 5ºGrau nas diferentes classes diamétricas..... | 92 |
| TABELA 3.6- Avaliação da homocedasticidade do Polinômio de Potências Fracionárias e Inteiras nas diferentes classes diamétricas..... | 93 |
| TABELA 3.7- Avaliação da homocedasticidade do Polinômio proposto por Pires & Calegario nas diferentes classes diamétricas..... | 93 |
| TABELA 3.8- Teste de normalidade para os polinômios de 5ºGrau, Hradetzky e Pires & Calegario, nas diversas classes diamétricas, considerando o resíduo Padronizado e o Estudentizado com a probabilidade de 95%..... | 95 |
| TABELA 3.9 - VIF para os parâmetros do polinômio de 5ºGrau nas diferentes classes diamétricas..... | 97 |
| TABELA 3.10 - VIF para os parâmetros do polinômio de Potências Fracionárias e Inteiras nas diferentes classes diamétricas..... | 97 |
| TABELA 3.11 - VIF para os parâmetros do polinômio proposto por Pires & Calegario nas diferentes classes diamétricas..... | 98 |
| TABELA 3.12 – Coeficientes da regressão selecionados pelo teste t, considerando como variável dependente d, para o polinômio de 5ºGrau, testando os coeficientes das variáveis selecionadas, o seu valor, o erro padrão, o valor de t calculado, e o valor da probabilidade encontrado para o t calculado, para verificar a necessidade do controle por classe diamétrica..... | 99 |
| TABELA 3.13 – Medidas de Acuracidade do polinômio proposto por Pires & Calegario (2004) para todas as classes diamétricas e para cada classe, na validação dos dados..... | 101 |
| TABELA 3.14- Avaliação da homocedasticidade do Polinômio Pires & Calegario nas diferentes classes diamétricas, na validação dos dados..... | 103 |
| TABELA 3.15- Teste de normalidade para o polinômio Pires & Calegario, nas diversas classes diamétricas, considerando o resíduo Padronizado e o Estudentizado com a probabilidade de 95%, na validação dos dados..... | 104 |
| TABELA 3.16 - VIF para os parâmetros do polinômio proposto por Pires & Calegario nas diferentes classes diamétricas, na validação..... | 104 |
| TABELA 3.17 – Contagem das observações que necessitam de atenção e sua percentagem, utilizando alguns critérios estatísticos, para o polinômio Pires & Calegario, na validação dos dados..... | 105 |
| TABELA 3.18 – Poder preditivo do polinômio Pires & Calegario (2004), na validação dos dados. | 107 |

| | |
|--|-----|
| TABELA 3.19 – Observações de influência nos parâmetros ajustados, $DFBETAS_{j,i}$, para o polinômio Pires & Calegario, desconsiderando o controle por classe diamétrica. | 107 |
| TABELA 3.20 – Análises do polinômio proposto por Pires & Calegario, após a retirada das observações de influência, desconsiderando o controle por classes diamétrica. | 108 |
| TABELA 3.21 – Avaliação do resíduo, segundo as pressuposições básicas de homogeneidade e normalidade, para o polinômio proposto por Pires & Calegario (2004) após a retirada das observações de influência, desconsiderando as classes diamétricas. | 111 |

ÍNDICE DE FIGURAS

| | Página |
|---|--------|
| Figura 2.1- Perfis das árvores nas diferentes classes diamétricas, onde “d” corresponde ao diâmetro (cm) e “h” a altura (m)..... | 29 |
| Figura 2.2- Resíduos em função do valores estimados para o polinômio de 5º grau, considerando todas as classes diamétricas, usando d/dap como variável dependente..... | 38 |
| Figura 2.3- Resíduos versus os valores estimados, para o polinômio de Hradetzky (1976), com todos os dados ajustados, usando d/dap como variável dependente. | 39 |
| Figura 2.4- Resíduos versus os valores estimados, para o polinômio proposto por Pires & Calegario (2004), com todos os dados ajustados, usando d/dap como variável dependente. | 40 |
| Figura 2.5 – Resíduo, para o modelo de Gompertz, nos agrupamentos das classes (2, 3 e 4), (5 e 6), (7 e 8). | 49 |
| Figura 2.6 – Resíduo, para o modelo Logístico, nos agrupamentos das classes (2, 3 e 4), (5 e 6), (7 e 8). | 50 |
| Figura 2.7 –Resíduo, para o modelo de Weibull, nos agrupamentos das classes (2, 3 e 4), (5 e 6), (7 e 8). | 51 |
| Figura 2.8 –Resíduo, para o polinômio do 5ºGrau, nos agrupamentos das classes (2, 3 e 4), (5 e 6), (7 e 8). | 52 |
| Figura 2.9 –Resíduo, para o polinômio de Potências Fracionárias e Inteiras (Hradetzky), nos agrupamentos das classes (2, 3 e 4), (5 e 6), (7 e 8)..... | 53 |
| Figura 2.10 – Resíduo, para o polinômio proposto por Pires & Calegario, nos agrupamentos das classes (2, 3 e 4), (5 e 6), (7 e 8)..... | 54 |
| Figura 2.11 – Resíduo, para o modelo Logístico (2.10), nos agrupamentos das classes (2, 3 e 4), (5 e 6), (7 e 8)..... | 58 |
| Figura 2.12– Gráficos dos perfis médios reais, do modelo Logístico e do polinômio proposto por Pires & Calegario, nos agrupamentos das classes (2, 3 e 4), (5 e 6), (7 e 8)..... | 60 |
| Figura 3.1- Valor da distância de Cook no polinômio de 5ºGrau, para todas as classes diamétricas. | 89 |
| Figura 3.2- Valor da distância de Cook no polinômio de Potências Fracionárias e Inteiras, para todas as classes diamétricas..... | 90 |
| Figura 3.3- Valor da distância de Cook no polinômio proposto por Pires & Calegario, para todas as classes diamétricas..... | 90 |
| Figura 3.4- Comparação dos valores de t tabelado e o calculado, na verificação da homocedasticidade, para os diferentes polinômios nas diferentes classes diamétricas..... | 94 |

| | |
|--|-----|
| Figura 3.5 - Resíduo versus o valor estimado, para o polinômio proposto por Pires & Calegario (2004), com todos os dados ajustados, na validação..... | 102 |
| Figura 3.6- Comparação dos valores de t tabelado e calculado, na verificação da homocedasticidade, para o polinômios Pires & Calegario (2004), nas diferentes classes diamétricas. | 103 |
| Figura 3.7- Gráfico de Resíduo, para o polinômio proposto por Pires & Calegario, após a retirada das observações de influência, desconsiderando as classes diamétricas..... | 109 |
| Figura 3.8- Resíduo Padronizado para detecção de “outliers”, após a retirada das observações de influência, para o polinômio proposto por Pires & Calegario (2004),desconsiderando as classes diamétricas..... | 110 |

AGRADECIMENTOS

A Deus, que me acompanha e me guia pelo seu caminho luminoso.

À Universidade Federal de Lavras, à coordenadoria de Pós-Graduação e ao Departamento de Ciências Florestais pela oportunidade de realização de um sonho.

Ao professor, amigo e orientador Natalino Calegario, que esteve sempre me indicando o caminho, fazendo me trilhar por ele sem me desviar do objetivo.

Ao professor José Roberto Scolforo, por todos os ensinamentos e lições de vida. Que estas estarão guardadas comigo onde quer que esteja.

Aos professores Departamento de Ciências Florestais, pelos ensinamentos e pela colaboração.

As secretárias do Departamento de Ciências Florestais, Teresinha, Rose e Gláucia, pela auxílio sempre que necessário e pelo bom senso.

Aos companheiros de curso de graduação e pós-graduação pelo convívio durante estes períodos: Luis, Franciane, Fred, Guilherme, Samurai, Honório, Vivian, Awdréia, Yustane, Graciane, Pintado, Valéria, Vanete, Valdir, Cláudio, Dimas e Ana Carolina.

Em especial aos grandes amigos Lucas, Edmilson, Maria Zélia e Breno, pela ajuda, companheirismo e colaboração em todos os nossos trabalhos.

A todos que participaram direta ou indiretamente deste trabalho.

RESUMO

PIRES, Leonardo Machado. **Ajuste e Diagnóstico de Modelos Estocásticos Lineares e Não-Lineares para descrição do Perfil Longitudinal de Árvores.** 2004. 104 p. Dissertação (Mestrado em Engenharia Florestal) – Universidade Federal de Lavras, Lavras¹.

A madeira é um bem de consumo fundamental na vida atual, fornecendo diversos produtos advindos de uma floresta plantada, na maioria das vezes, ou de florestas naturais. A diversidade deste produto florestal numa empresa é uma forma de verticalização do produto, aumentando as receitas e reduzindo os custos. Uma forma de otimizar o uso múltiplo da árvore é através de modelos estocásticos lineares e não-lineares que descrevam o perfil longitudinal da mesma. Assim, o objetivo deste trabalho foi comparar os modelos lineares e os não-lineares e diagnosticar os modelos lineares através de avançadas medidas estatísticas. Os dados vieram da empresa Aracruz Celulose S/A, situada no litoral leste do Brasil, nos estados da Bahia e Espírito Santo originaram-se de cubagem rigorosa para a espécie *Eucalyptus* spp., com diversas idades, em várias classes diamétricas, num total de 140 árvores. A avaliação dos modelos foi através das medidas de acuracidade: R^2 , s_{yx} e, principalmente, do gráfico de resíduo. Os resultados mostraram o modelo não-linear Logístico como superior aos demais modelos, devido às suas características assintóticas, semelhantemente aos dados de cubagem. As medidas de diagnóstico utilizadas, resíduo, PRESS, R^2_{predito} , $DFBETAS_{j,i}$, distância de Cook, $DFITS_i$, $COVRATIO_i$, observações de influência, segundo critério de Hoaglin e Welsch, e os valores estranhos ("outliers"), foram eficientes para destacar as observações que influenciam os modelos lineares. O Fator de Inflação da Variância (VIF) também foi utilizado para observar a presença de multicolinearidade. Esta esteve presente em todos os modelos lineares testados. Pelo teste de hipótese dos parâmetros para o polinômio de 5º Grau foi construído um modelo único para todas as classes diamétricas com boa precisão. As condicionantes do resíduo apresentaram os resultados esperados, ou seja, todos os modelos lineares testados foram, de maneira geral, homocedásticos e com distribuição normal.

¹Comitê de Orientação: Natalino Calegario (Orientador) – UFLA.

1.2 REFERENCIAL TEÓRICO

1.2.1 Forma da árvore

De acordo com Husch et al. (1982), o tronco de uma árvore normalmente se assemelha a formas sólidas de figuras geométricas do tipo neilóide, conóide ou parabolóide, que são obtidos com a rotação da curva de forma geral $y = k x^r$, em torno do eixo x. Sempre que o seu perfil possa ser expresso por uma equação, a fórmula para o volume do tronco obtido pela rotação do gráfico da equação $y = f(x)$ sobre o eixo x (no intervalo de $x=a$ e $x=b$) pode ser escrita como: $v = \pi \int_a^b y^2 dx$ e, assim, obtêm-se o volume do sólido por integração. Como o valor do expoente “ r ” pode mudar na equação $y = k x^r$, diferentes sólidos podem ser produzidos. Assim, para $r=0$ na equação, a forma é cilíndrica; para $r=1/3$, a forma é parabolóide cúbico; para $r=1/2$, a forma é parabolóide quadrático; para $r=3/2$, a forma é neilóide; e para $r=1$, a forma é cônica. É sugerido considerar o tronco da árvore não somente com uma forma geométrica, mas uma composição de vários sólidos geométricos com 4 partes, ou seções, as quais sejam: a) Forma de cone ou parabolóide na seção apical da árvore; b) forma de cone ou cilíndrica podem ocorrer na seção arbórea central, predominando a forma parabolóide; c) forma neilóide na seção acima no tronco; e d) da base junto ao solo, embora pareça mais um neilóide, por razões práticas, é considerado como um cilindro.

Ocorrem grandes variações na forma dos fustes arbóreos que podem ser: de espécie para espécie e mesmo dentro de uma mesma espécie; de acordo com o tratamento e localização geográfica, e outros fatores que provoquem a interação genótipo ambiente. Vários autores, como Chapman & Meyer (1949),

Loetsch et al. (1973) e Husch et al. (1982) concluíram que a forma do tronco é basicamente descrita como um neilóide na base, um parabolóide no meio e um cone no topo, não sendo possível a determinação exata da transição de um sólido para o outro ao longo do tronco.

Husch et al. (1982) ainda citam que os troncos das resinosas raramente se apresentam como cones, neilóides ou parabolóides típicos, e sim formas intermediárias entre o cone e o parabolóide. Entretanto, os troncos comerciais de folhosas são definidos como semelhantes a troncos de cones, neilóides, parabolóides e, eventualmente, cilíndricos.

O estudo sobre o comportamento da forma do fuste é extremamente importante, pois, além de afetar o volume total, também afeta a qualidade e a quantidade dos diversos produtos extraídos da árvore (Ferreira, 1999).

A forma do perfil longitudinal dos troncos das árvores em florestas naturais tropicais varia, principalmente, em função da espécie, da idade, do ecossistema, do sistema silvicultural aplicado, da competição entre plantas e da interação entre essas variáveis. Por isso, há troncos que possuem semelhança com algum tipo de figura geométrica sólida e há troncos que não se assemelham a uma geometria definida (Chichorro, 2000).

O desenvolvimento de modelos dendrométricos objetiva utilizar modelos matemáticos e, ou, estatísticos para analisar dados provenientes de cubagens de árvores, de modo que estas tenham suas formas naturais comparadas a sólidos geométricos de revolução, para que seus volumes sejam determinados. Estes sólidos são denominados “protótipos dendrométricos” ou sólidos padrões, sendo comparáveis com a forma do tronco ou parte deste (Finger, 1992). O mesmo autor acrescenta que o estudo matemático dos volumes das árvores considera suas seções circulares, assim, o fuste tridimensional é gerado pela rotação em torno de seu eixo principal, originando os sólidos de

revolução correspondentes. Estes, por sua vez, têm seu volume obtido através da integral da área seccional sobre o comprimento total do tronco.

1.2.2 Equações de Perfil

Os estudos sobre a modelagem da forma da árvore vêm sendo desenvolvidos intensamente desde o século XIX. A primeira tentativa de se expressar a forma média do tronco através de uma equação matemática e estatística ocorreu em 1903 com Höjer.

A partir de então, vários autores como Jonson (1911), Wright (1923), Heijbel (1928), Behre (1923 e 1927), Matte (1949), dentre outros, estudaram, de maneira pormenorizada, a possibilidade da forma da árvore ser representada por expressões matemáticas e, ou, estatísticas.

De acordo com Spurr (1952), no século XIX, o volume das árvores individuais eram obtidos através do fator de forma e, eventualmente, dos quocientes de formas. Entretanto, muitos profissionais da área florestal tem procurado equações com uma ou duas variáveis, envolvendo poucos parâmetros e variáveis, que possam ser usadas para representar o perfil inteiro da árvore.

Os modelos ou funções de taper foram inicialmente desenvolvidos com interesse de aplicação em florestas plantadas puras e florestas naturais mistas, em países de clima temperado, com a finalidade de estimar os volumes comercial e total, a altura comercial e o diâmetro superior (Kozak et al., 1969; Demaerschalk, 1972; Goulding & Murray, 1976; Burkhart, 1977; Biging, 1984).

No Brasil, estudos sobre a modelagem da forma de árvores, em sua grande maioria, estão relacionados a florestas equiâneas puras, com espécies como *Pinus spp.* e *Eucalyptus spp.* (Campos & Ribeiro, 1982; Guimarães & Leite, 1992; Schneider et al., 1996), sendo raro os estudos com espécies nativas.

Pode-se citar, como único exemplo de estudo com espécies nativas das florestas naturais tropicais, o caso do morototó (*Didymopanax morototonii*) (Garcia et al., 1993).

O pouco uso destes modelos em florestas naturais, segundo Chichorro (2000), está associado à baixa precisão, em razão da grande diversidade de espécies e idades que compõem estas florestas e, conseqüentemente, dos diferentes perfis das árvores.

A aplicação das funções de taper é de grande utilidade para quantificar o uso da madeira para os diferentes sortimentos, principalmente em termos de volume, pois descreve o número de toras obtido para as varias dimensões comercializáveis, conforme os diferentes produtos florestais (Gomes, 2001).

Segundo Husch et al. (1982), taper é o termo aplicado à taxa de decréscimo em diâmetro ao longo do fuste, sendo a razão fundamental da variação da forma e do volume deste. Equações de perfis de árvores significa o decréscimo gradual ou diminuição do diâmetro a partir da base em direção ao topo (Chichorro, 2000). A avaliação de multiprodutos da madeira das árvores exige uma descrição do perfil de seus fustes, ou seja, exige a determinação ou estimativa de vários diâmetros ao longo dos mesmos, através de equações de afilamento ou forma ou equações de taper (Husch et al., 1982).

Os modelos chamados de funções de afilamento, ou de forma, ou taper, são maneiras de se descrever matematicamente o perfil de um tronco. Para isto, a seção transversal em qualquer posição do tronco é considerada como circular e o volume da seção é calculado pela integração das funções de taper. Portanto, o fuste é tratado como um sólido de revolução (Ahrens & Holbert, 1981 citados por Assis, 1998).

Scolforo (1997) comenta que as funções de taper são opções obtidas de uma forma um pouco mais complexa que as equações de volume, no entanto muito mais flexíveis. Isto porque, através destas, pode-se quantificar o diâmetro

a qualquer altura ou vice-versa, possibilitando encontrar volume total ou comercial para uma mesma árvore e estes volumes com ou sem casca. Nestas estimativas, o nível de precisão varia, normalmente, de acordo com a equação ou modelo utilizado e, ou, com a forma do tronco.

Apesar das vantagens apresentadas pelas funções de taper em relação às tradicionais equações de volume, as funções de forma podem acarretar uma série de erros quando utilizadas para estimativas por unidade de área, pois os ajustes são feitos para um pequeno número de árvores comparado ao número de árvores envolvido num inventário florestal (Kichner et al., 1989). Estes autores enfatizam, como possíveis causas destes erros, vários fatores capazes de influenciar na forma da árvore, como a copa, idade, densidade, sítio, localização geográfica, etc. Entretanto, tais erros podem ser diminuídos quando se aumenta o espectro amostral, o que implica no aumento dos custos do processo.

Para Kozak et al. (1969), estimativas do volume total por unidade de área não são suficientes, porém o volume do material disponível em certos tamanhos e certas qualidades devem ser estimados com alto padrão de precisão e exatidão. Entretanto, os autores afirmam que a possibilidade de estimar vários diâmetros, alturas e volumes por meio de funções de taper, devidamente apropriadas às diferentes finalidades, só é garantido quando se obtêm no campo todas as informações necessárias, de maneira criteriosa e compreensiva.

Nos inventários tropicais, segundo Loetsch et al. (1973), para evitar irregularidades causadas por sapopemas, uma função de taper simples, começando acima do toco e terminando no ponto inicial da copa, pode oferecer resultados satisfatórios. No entanto, estes autores comentam que, se funções de taper forem usadas no inventário florestal, é necessário aceitar, previamente, as possíveis tendências nas estimativas.

Segundo Péllico Netto (1980), as funções que descrevem a forma da árvore podem ser construídas para uma série absoluta ou para uma série relativa

contínua de forma. Sendo que nas séries relativas contínuas de forma, a função que descreve o perfil da árvore é desenvolvida por uma série relativa de diâmetros correspondente a uma série relativa de alturas. Esta é a melhor maneira de se estimar volumes parciais ou totais das árvores, pois considera a hipótese de que as árvores com dimensões diferentes podem ter formas semelhantes. Assim, neste processo, todas as árvores são comparáveis independentemente de seus tamanhos e idade.

Atualmente há vários modelos ou funções de taper que são utilizados para descrever a forma do tronco, os quais apresentam-se com melhor ou pior qualidade de ajuste, dependendo do objetivo e, ou, da variável considerada (Chichorro, 2000).

Dentre os vários modelos existentes, Lima (1986) destacou os seguintes:

- **Modelos Polinomiais:** são empregadas técnicas de regressão no ajuste das equações, em que normalmente a variável dependente é dada pela razão entre os diâmetros superiores e o diâmetro medido a 1,30m do solo (d/D) ou, algumas vezes, a variável dependente é expressa pelo quadrado da razão destes diâmetros $(d/D)^2$. As variáveis independentes são expressas por razões de alturas comerciais e total (h/H), principalmente. Segundo Rios (1997), a dificuldade destes modelos em explicar as alterações na base da árvore levou à implementação de modelos polinomiais de grau elevado. Destaca-se, neste caso, o modelo de Kozak et al. (1969).
- **Modelos Sigmoidais:** São fundamentados no fato das funções permitirem pontos de inflexão, bem como apresentarem característica de grande flexibilidade no ajuste dos dados observados de diferentes relações biológicas. Essas equações

normalmente são derivadas da função de Chapman-Richards, e são apresentadas por Garay (1979) e Biging (1984). Rios (1997) acrescenta que os modelos sigmoidais utilizam transformações de variáveis de modo que o modelo se assemelha à forma natural dos troncos. Trata-se de expressões matemática estatística que geram um perfil semelhante a uma curva sigmóide.

- **Modelos Compatíveis:** São assim chamados em razão da compatibilidade nas estimativas entre equações de taper e equações de volume. A compatibilidade é verificada integrando-se as áreas seccionais ao longo do tronco, produzindo estimativas semelhantes àquelas obtidas pela equação de volume, da qual a equação de taper foi derivada. Um procedimento compatível com equações de volume foi desenvolvido por Demaerschalk (1971, 1972).

Além destes modelos, Guimarães & Leite (1992) mencionaram os modelos definidos por análise multivariada, que se baseiam na análise de componentes principais para investigar as variações existentes na estrutura dos dados e, então, definir o modelo de regressão a ser utilizado para descrever o perfil das árvores, citando Fries & Matern (1966) como precursores desta técnica. Estes autores mencionam também os modelos segmentados para estimar diâmetros e descrever o perfil arbóreo, que se constituem no uso de três submodelos justapostos, cada um representando uma porção do tronco, cuja união se dá por meio da imposição de restrições ao modelo geral, ajustados para seções inferiores, média e superior do tronco, mencionado, como exemplo, Max & Burkhart (1976) e Cao et al. (1980). Entretanto, admitem-se certas

dificuldades na predição de volumes parciais e na forma do tronco, quando aplicados os modelos de taper.

A utilização da técnica de regressão segmentada, representando cada parte do tronco por uma função polinomial, surgiu devido à dificuldade de descrição da forma da árvore por uma única função matemática ou estatística, pois cada parte do tronco possui uma forma geométrica particular. Então, utilizando-se a técnica da regressão segmentada, a curva de cada segmento é gerada pelo método dos mínimos quadrados e os pontos de união de cada segmento são necessariamente harmonizados (Kirchner et al., 1989).

Devido à pouca explicação de modelos polinomiais às deformações da base do tronco, Hradetzki (1976) estudou as potências mais apropriadas na descrição da forma da árvore e apontou como sugestão a construção de equações pelo método "stepwise", para expoentes variando de 0,005 a 25. Com isso, uma representação adequada do fuste através de polinômios exige uma combinação apropriada de potências e que estas sejam submetidas ao processo de seleção da regressão "passo a passo" (stepwise). Rosot (1989) concluiu que as equações construídas pelo processo de "stepwise" foram mais precisas que o polinômio de quinto grau. Também observou uma grande variação dos modelos nas diferentes classes diamétricas, o que o levou a sugerir o ajuste de modelos por classe diamétrica.

Segundo Demaerschalk (1972), é possível derivar uma razão de volume a partir de vários modelos, desde que estes forneçam estimativas precisas. Ressalta ainda que, para ser eficiente, um sistema não pode ser tendencioso nas diferentes classes de tamanho arbóreo, seja para estimativa de diâmetros a diferentes alturas ou vice-versa, ou ainda para estimativas de volume.

Lima (1986) analisou os modelos propostos por Biging (1984); Demaerschalk (1973); Kozak et al. (1969) e Ormerod (1973) em árvores de *Pinus elliottii*. Estes modelos foram avaliados quanto à estimativa de volume

total e comercial e de diâmetros e alturas comerciais. Conclusivamente, o modelo que apresentou maior precisão foi o de Kozak et al., e o menos preciso foi o de Ormerod. O modelo de Demaerschalk mostrou-se satisfatório para estimar todas as variáveis, exceto o volume total. Para esta variável, o modelo de Biging se adequou melhor.

Leite et al. (1995) propuseram um novo método para estimar volumes de partes do tronco, empregando um modelo volumétrico múltiplo que utiliza variáveis binárias além das variáveis usuais (diâmetro comercial, altura total e DAP). Partindo de transformações no modelo de Schumacher & Hall (1933), obtiveram um modelo múltiplo que permite estimar de maneira consistente e compatível os volumes com e sem casca de partes do tronco. Os resultados obtidos no teste do modelo foram satisfatórios quando comparados com as razões de volume.

Rios (1997) avaliou o desempenho de modelos polinomiais, de razões de volume e de função spline cúbica na estimativa de volumes comerciais. Neste trabalho, feito por classe diamétrica na espécie *Pinus elliottii*, utilizou-se três diâmetros mínimos pré-estabelecidos. O autor verificou que os modelos polinomiais proporcionaram estimativas mais acuradas na descrição dos perfis das árvores, onde o polinômio de quinto grau foi superior ao polinômio de potência fracionária. Em segundo lugar ficaram as razões de volume e as funções spline em terceiro. O autor ainda sugere que as equações de afilamento sejam ajustadas por classe diamétrica, que resultará em estimativas mais acuradas.

Assis (1998) trabalhou com o gênero *Eucalyptus* testando a acuracidade de modelos que melhor descrevem o perfil arbóreo observando, para isto, as diferentes classes diamétricas e as diferentes espécies. Concluiu-se que a equação de afilamento de potência fracionária é mais estável na estimativa dos diâmetros ao longo do fuste, tanto para um ajuste por classe diamétrica quanto

para um ajuste de um conjunto de árvores cubadas no estrato. O autor ainda denota que a equação de Amateis e Burkhart e o polinômio do quinto grau podem ser utilizados com segurança para estimar o diâmetro ao longo do fuste, embora com acuracidade inferior ao polinômio de potência fracionária, e descreve que a equação de afilamento de Clutter não é recomendada para estimar diâmetro ao longo do fuste de *Eucalyptus grandis* e *Eucalyptus urophylla* na região do município de Luiz Antônio, em São Paulo, pois resultou em estimativas pouco acuradas.

Thiersch (1999) avaliou a acuracidade de três modelos polinomiais segmentados e dois modelos não segmentados, para representar os perfis dos fustes de clones de *Eucalyptus grandis*. Para isto, o autor utilizou dados obtidos da cubagem rigorosa, com abate da árvore, e a cubagem através do penta-prisma de Wheeler. Os modelos mais acurados foram o de Clark et al. (1991) e o polinômio de potência fracionária para a cubagem rigorosa tradicional (abate das árvores) e para o penta-prisma, respectivamente. Em segundo lugar ficaram o modelo de Max e Burkhart e o polinômio de quinto grau. E, em último lugar, ficou o modelo segmentado de Parressol et al. (1987). Em outro trabalho, Chichorro (2000) considerou como melhor modelo para estimar diâmetros ao longo do tronco o proposto por Demaerschalk, pois, de acordo com as estatísticas de validação consideradas, foi superior aos demais.

Calegário & Daniels (2002) utilizaram informações do inventário florestal tradicional unido às informações de uma amostra arbórea, para verificar a modelagem do perfil de árvores individuais através de um modelo não linear de efeito misto. Novas variáveis foram incluídas no modelo para melhorar a representação dos dados, onde a altura dominante e os tipos de clones tiveram influências significativas no modelo. Neste mesmo estudo, usando parâmetros aleatórios e a técnica do sólido de revolução, pôde-se gerar volumes arbóreos individuais, confirmando a precisão destes modelos na estimativa do perfil da

árvore. Esta técnica tem a vantagem na captura de variações individuais na árvore. O modelo se mostrou robusto para estimar o perfil arbóreo com dados somente de DAP e altura. O método gerou estimativas precisas sendo uma alternativa para estimar o volume de árvores individuais e da parcela. As árvores com dados de DAP e altura total originaram uma curva estimada que cruzou os dois pontos da amostra. Isto ocorre devido a robustez do modelo de efeito misto. Em outra aproximação, com o usual método dos mínimos quadrados ordinários, esta situação seria um problema porque esta aproximação ajusta uma curva média para a população ou uma curva independente para cada árvore. Os autores ainda citam que, quando o perfil da árvore estimado pelo efeito misto é comparado com o valor real, pode-se notar que o modelo gera uma curva sigmóide, próximo da tendência observada. Esta situação tem uma forte aplicação prática, em estudos de predição florestal.

Os autores descrevem que o modelo de taper na forma não linear com quatro parâmetros advem do modelo logístico de três parâmetros. A inclusão do quarto parâmetro é motivada pelo fato de que o perfil da árvore possui evidências assintóticas superior e inferior. Este modelo logístico pode ser selecionado através de análise visual do perfil dos troncos e tem como qualidades a interpretação biológica dos parâmetros. Este fato simplifica a interpretação das características do perfil arbóreo facilitando, assim, as análises de estimativas.

Lindstrom & Bates (1990) apresentam um procedimento para modelos não lineares de taper, no intuito de predição da altura a diferentes diâmetros na árvore, de forma iterativa composta de dois passos para calcular os parâmetros estimados do modelo não linear geral. O primeiro passo usa o processo de iteração para obter a estimativa do efeito fixo (β), aleatório (b), variância e covariância dos componentes (θ), covariantes (X) e a matriz associada ao efeito aleatório (Z). Estas estimativas são baseadas nos fatores de Cholesky. O segundo

passo usa a estimativa obtida no primeiro passo e utiliza do processo iterativo para gerar as estimativas desejadas de (θ) , (σ) e (β) .

Scolforo (1997) afirma que, em termos de precisão, as funções de taper e as equações de volume são equivalentes, sendo as funções de afilamento mais interessantes, pois permitem estimar volume de qualquer porção da árvore. O autor observa que manualmente o procedimento é impraticável, já que para cada árvore que se deseja o sortimento, é necessário determinar novos parâmetros, de acordo com seu diâmetro tomado a 10% da altura total e de sua altura total. Entretanto, um software adequado torna o procedimento bastante operacional, já que ao operador bastará fornecer o diâmetro da árvore a 1,30m ou a 10% da altura, a altura total da árvore, além dos intervalos de altura que se deseja obter os valores de volume.

De acordo com Munro & Demershalk (1974), há várias maneiras de obter um sistema compatível de forma e volume, sendo mais importante um sistema baseado na forma e outro em volume. Assim, este sistema é constituído, basicamente, por uma função de forma. Para o sistema baseado no volume é necessário o ajuste da equação de volume e, posteriormente, substituição de seus coeficientes na função de forma compatível com a equação de volume. A premissa do procedimento é que a estimativa do volume total pela integração da função de afilamento seja igual à estimativa proporcionada pela equação de volume. Estes autores ressaltam que, dentro do primeiro sistema, a acuracidade das estimativas depende unicamente da precisão do modelo de forma. Muitas equações, principalmente aquelas sem pontos de inflexão, tendem a subestimar na base e superestimar no ápice da árvore, devendo ser consideradas no instante onde se estima o volume das seções na árvore. Para as funções de forma sigmóides, com um ponto de inflexão, não há solução para o problema de tendenciosidade. O padrão de tendência é inteiramente diferente, sendo apontados pequenos erros na parte superior.

1.2.3 Quadrado médio lineares versus não-linear

Dada uma validade da hipótese de independência e distribuição normal do resíduo, o mesmo pode ser uma afirmação geral sobre os estimadores do quadrado médio nos modelos de regressão linear e não-linear. Para um modelo de regressão linear, as estimativas dos parâmetros não são tendenciosas, são normalmente distribuídas, e tem uma possível variância mínima dentro de uma classe de estimadores conhecidos como estimadores regulares. Esta possível variância mínima é também chamada de limite mínimo da variância. Estas propriedades são, geralmente, aceitas para a maioria das características desejáveis dos estimadores. Outros estimadores, apesar de não serem tendenciosos, têm baixa “eficiência” devido às suas variâncias excederem àqueles dos estimadores dos mínimos quadrados (Ratkowsky, 1989).

O mesmo autor ressalta que os modelos de regressão não-linear diferem dos modelos lineares devido aos estimadores do quadrado médio de seus parâmetros não serem tendenciosos, não apresentam distribuição normal e a variância mínima. Os estimadores alcançam estas propriedades somente assintoticamente, com uma amostra de tamanho próximo do infinito. Alguns modelos de regressão não-linear tem estimadores que são “pessimamente” tendenciosos, com altíssima assimetria, distribuição não normal de “calda longa”, e com uma variância que excede grandemente o limite mínimo.

Ratkowsky (1989) declara que modelos não-lineares diferem muito entre si, no que diz respeito à extensão na qual o comportamento de seus estimadores de quadrado médio aproximam das propriedades assintóticas. Por exemplo, há alguns modelos não lineares cujos estimadores se igualam quando o tamanho amostral é relativamente pequeno, portanto são “não-tendenciosos”, normalmente distribuídos, e com variância mínima. O autor chama estes modelos de regressão não linear com comportamento próximo ao linear.

1.3 REFERÊNCIAS BIBLIOGRÁFICAS

ASSIS, A. L. de. **Acuracidade na estimativa de volumes comerciais de *Eucalyptus grandis* e *Eucalyptus urophylla***. Lavras: UFLA, 1998. 183 p. Monografia.

BIGING, G.S. Taper equations for second mixed-conifers of Northern California. **Forest Science**, Washington, v.30, n.4, p.1103-1117, 1984.

BURKHART, H.E. Cubic-foot volume of loblolly pine to any merchantable top limit. **Southern Journal of Applied Forestry**, Bethesda v.1, n.2 p.7-9, 1977.

CALEGARIO, N.; DANIELS, R.F. Modeling individual tree profile based on a nonlinear mixed effect model: an application in *Eucalyptus* stands. 2002. 30p.

CAMPOS, J.C.C., RIBEIRO, J.C. Avaliação de dois modelos de taper em árvores de *Pinus patula*. **Revista Árvore**, v.6, n.2, p.140-149, Viçosa jul/dez 1982.

CHAPMAN, H.H.; MEYER, W.H. **Forest Mensuration**. New York: McGraw-Hill, 1949. 522p.

CHICHORRO, J.F. **Análise estrutural e econômica de multiprodutos da madeira em florestas naturais**. Viçosa: UFV, 2000.241p. Tese (Doutorado em Ciência Florestal) – Universidade Federal de Viçosa, 2000.

DEMAERSCHALK, J.P.. Converting volume equations to compatible taper equations. **Forest Science**, Washington, v.18,n.3,p.241-245, sept.1972.

DEMAERSCHALK, J.P. Taper equations can be converted to volume equations and point sampling factors. **The Forestry Chronicle**, Ottawa Dec, v. 47, n.6, p.352-354, 1971.

FERREIRA, S. **O Estudo da forma do fuste de *Eucalyptus grandis* e *Eucalyptus cloeziana***. Lavras: UFLA, 1999. 132p. (Dissertação de Mestrado em Engenharia Florestal).

FINGER, C. A. G. **Fundamentos de biometria florestal**. Santa Maria, UFSM/CEPEF/FATEC, 1992, 269p.

GARAY, L. **Tropical forest utilization system**. VIII. A taper model for entire stemprofile including buttressing. Coll. Forest. Resour., Inst. Forest Prod. Univ. Wash. 1979. 64p. (contrib. 36).

GARCIA, S.L.R.; LEITE, H.G.; YARED, J.A.G. Análise do perfil do tronco de morototó (*Didymopax morototonii*) em função do espaçamento. In: CONGRESSO FLORESTAL PANAMERICANO e CONGRESSO FLORESTAL BRASILEIRO, 1993, Curitiba, PR, **Anais...** Curitiba, PR, Sociedade Brasileira de Silvicultura e Sociedade Brasileira de Engenheiros Florestais, V.2, São Paulo, Unipress, 1993, p.485-491.

GOMES, E. R. **Avaliação de modelos polinomiais segmentados e não segmentados na estimativa de diâmetros e volumes comerciais de *Eucalyptus cloeziana***. Lavras: UFLA, 2001. 46p. Monografia.

GOULDING, C.J., MURRAY, J.C. Polynomial taper equations that are compatible with tree volume equations. **New Zealand Journal of Forestry Science**, v.5, n.3, p.313-322, 1976.

GUIMARÃES, D.P.; LEITE, H.G. Um novo modelo para descrever o perfil do tronco. **Revista Árvore**, v.16, n.2, p.170-180, Viçosa abr/jun 1992.

HRADETZKY, J. **Analyse und interpretation statistischer abränger keiten (Biometrische Beiträge zu aktuellen forschung projekten)**. Baden: Württemberg Mitteilungen der FVA, 1976. 146p. (Abt. Biometric und Informatik, 21).

HUSCH, B; MILLER, C.L.; BEERS, T.E. **Forest Mensuration**. 3. ed. New York: J. Wiley & Sons, 1982.397p.

KIRCHNER, F.F.; FIGUEIREDO FILHO, A.; SCOLFORO, J.R.S.; MACHADO, S.A.; MITISHITA, E.A.. "O uso de funções Spline no calculo de volumes de árvores." **Floresta**, Curitiba v.19, n.1-2, p.116-122. Curitiba, 1989.

KOZAK, A.; MUNRO, D.D.; SMITH, J.G.H. Taper function and their application in forest inventory. **The Forestry Chronicle**, Ottawa Aug, v.45, n.4, p.278-283, 1969.

LEITE, H.G.; GUIMARÃES, D.P.; CAMPOS, J.C.C.. "Descrição e emprego de um modelo para estimar múltiplos volumes de árvores." **Revista árvore**, v.19,n.1,p.65-79. Viçosa, jan/mar 1995.

LIMA, F.S. "Análise de funções de Taper destinadas à avaliação de multiprodutos de árvores de *Pinus elliottii*" Viçosa, 1986. 79p. (Dissertação de Mestrado em Engenharia Florestal). UFV.

LINDSTROM, J.M.; BATES, D.M. Nonlinear mixed effects models for repeated measures data. *Biometrics*, Washington, v.46, n.3, p.673-687, sept. 1990.

LOETSCH, F.; ZOHRER, F.; HALLER, K.E. "Forest Inventory" Volume II. BLV Verlagsgesellschaft München Bern Wien. 469p. 1973.

PÉLLICO NETTO, S. "Estimativas volumétricas de árvores individuais – síntese teórica." *Floresta*, Curitiba, v.11, n.2, p.63-73. Curitiba, dez. 1980.

RATKOWSKY, D.A.; REEDY, T.J. Choosing near-linear parameters in the four-parameter logistic model for radioligand and related assays. *Biometrics*. Washington, n.3, p.575-583, sept. 1986.

RATKOWSKY, D.A. *Handbook of nonlinear regression models*. New York and Basel: Marcel Dekker, 1989. 241p.

RIOS, M. S. "A eficiência das funções polinomiais, da função Spline cúbica, e razões de volume para representar o perfil da árvore e estimar os sortimentos de *Pinus elliottii*". Lavras, 1997. 116p. (Dissertação de mestrado em Engenharia Florestal). UFLA.

ROSOT, M.A.D. Estudo comparativo de métodos para avaliação volumétrica por unidade de área em povoamento de *Pinus taeda* L. Curitiba: UFPR, 1989. 163p. (Dissertação de mestrado em Engenharia Florestal).

SCOLFORO, J.R.S. "Biometria florestal 2: Técnica de regressão aplicada para estimar: volume, biomassa, relação hipsométrica e múltiplos produtos de madeira." Lavras: UFLA/FAEPE, 1997. 292p.

SCHNEIDER, P.R.; FINGER, C.A.G.; KLEIN, J.E.M.; TOTTI, J.A.; BAZZO, J.L. Forma de tronco e sortimentos de madeira de *Eucalyptus grandis* Maiden para o estado do Rio Grande do Sul. *Ciência Florestal*, v.6, n.1, p.79-88, Santa Maria nov. 1996.

SPURR, S.H. *Forest Inventory*. Ronald Press Co., New York, N.Y. 1952. 476p.

THIERSCH, C. R..”Acuracidade dos modelos polinomiais para representar os perfis dos fustes de clones de *Eucalyptus grandis*” Lavras: UFLA, 1999. 35 p. Monografia

CAPÍTULO 2

AJUSTE DE MODELOS ESTOCÁSTICOS LINEARES E NÃO-LINEARES PARA A DESCRIÇÃO DO PERFIL LONGITUDINAL DE ÁRVORES

RESUMO

PIRES, Leonardo Machado. **Ajuste de Modelos Estocásticos Lineares e Não-Lineares para a Descrição do Perfil Longitudinal de Árvores**. In: _____: Modelos Estocásticos Lineares e Não-Lineares para a Descrição do Perfil Longitudinal de Árvores. 2004. Cap. 2, p. 20-59. Dissertação (Mestrado em Engenharia Florestal) – Universidade Federal de Lavras, Lavras¹.

Os modelos polinomiais são mais difundidos e usuais no meio florestal brasileiro por diversos motivos, tais como a facilidade de ajuste e o desconhecimento de outros modelos. A utilização de modelos de regressão não-linear clássicos na descrição do perfil, como Gompertz, Logístico e Weibull, é pouco difundida no Brasil. Portanto, este estudo visa verificar qual é o melhor modelo não-linear e o melhor modelo linear, e qual é o melhor entre estes. Os modelos lineares foram apresentados de duas formas, com d/dap como variável dependente, para posterior validação e diagnóstico, e com h_i como variável dependente para comparar com os modelos não-lineares. As medidas de comparação foram: R^2 , syx , $R^2_{ajustado}$, gráfico de resíduo, correlação entre parâmetros, e facilidade de ajuste. Os resultados ressaltam que, dentre os modelos não-lineares, o que obteve melhor desempenho, de forma geral, foi o modelo Logístico, apesar do modelo de Gompertz ser superior a este e ao modelo de Weibull, em termos de erro padrão residual (syx). Nos modelos lineares, o polinômio proposto por Pires & Calegario (2004) foi superior aos demais. Ao comparar os modelos não-lineares com os modelos lineares, o modelo Logístico foi melhor devido, principalmente, ao comportamento dos dados serem não-lineares, à baixa correlação entre os parâmetros e a interpretabilidade dos mesmos, facilitando a convergência e o ajuste.

¹Comitê de Orientação: Natalino Calegario (Orientador).

FITTING LINEAR AND NONLINEAR STOCHASTIC MODELS FOR THE DESCRIPTION OF THE LONGITUDINAL TREE PROFILE

ABSTRACT

PIRES, Leonardo Machado. Fitting Linear and Nonlinear Stochastic model for the Description of the Longitudinal Tree Profile. In: _____: Linear and Nonlinear Stochastic models for the Description of the Longitudinal Tree Profile. 2004. Chap. 2, p. 20-59. Dissertation (Master in Forest Engineering) – Universidade Federal de Lavras, Lavras¹.

The use of polynomial models has spreading and becoming usual in the Brazilian forest modeling for several reasons. The use of models of regression nonlinear classic, like Gompertz, Logistic and Weibull, is not very diffused in Brazil. Therefore, this study seeks to verify which it is the best nonlinear model and the best linear model, and which one is the best among them. The linear models were presented in two ways, with di/dap as dependent variable, for posterior validation and diagnosis, and with hi as dependent variable to compare with the nonlinear models. The comparison measures were: R^2 , syx , $R^2_{ajustado}$, residual graph, correlation among parameters, and adjustment quality. The results showed that, among the nonlinear models, the logistic was the best, in spite of the model of Gompertz to be superior in terms of residual standard error (syx). In the linear models, the polynomial by Pires and Calegario (2004) was superior to the others. When comparing the nonlinear models with the linear models, the logistic was better due, mainly, to the behavior of the data in being nonlinear, the low correlation between the parameters and the biological interpretation of them, facilitating the convergence and the adjustment.

¹Guidance Committee: Natalino Calegario (Major Professor) – UFLA.

2.1 INTRODUÇÃO

O Brasil é uma potência mundial em produção madeireira, graças ao seu clima favorável e a sua grande extensão territorial. As plantas no Brasil crescem rapidamente, quando comparado com países de clima temperado, pode-se citar como exemplo o *Eucalyptus* spp., que é originário da Austrália mas é mais produtivo, em uma escala menor de tempo, no Brasil.

Há uma demanda muito grande de madeira para diferentes finalidades nas fábricas das grandes empresas. Celulose, madeira serrada, compensado e carvão, são apenas alguns usos encontrados para a mesma. Acontece que o valor pago a estes produtos varia imensamente (Assis, 2000).

A utilização de modelos lineares, principalmente os polinomiais, é muito difundida no meio florestal brasileiro, pode-se citar os trabalhos de Rios (1997), Assis (1998) e Thiersch (1999). Todavia, modelos não-lineares como Gompertz, Logístico e Weibull, raramente são utilizados para o ajuste de funções de afilamento, pode-se citar o trabalho de Calegario & Daniels (2002). Uma das justificativas da não utilização dos mesmos é a dificuldade de ajuste e de conversão. Porém, com a evolução tecnológica atual, estes ajustes são facilmente obtidos.

Portanto, o presente estudo consiste em: comparar os modelos lineares (polinomiais) com os modelos não-lineares, com o intuito de verificar qual o melhor modelo; e determinar qual o melhor modelo linear e qual o melhor modelo não-linear.

2.2 REFERENCIAL TEÓRICO

2.2.1 Convergibilidade

O quadrado médio estimado dos parâmetros no modelo de regressão não-linear é diferente daquele apresentado para os modelos de regressão linear, pois não pode ser determinado por expressão matemática explícita. Os modelos não lineares tem o dever de obter uma soma de quadrados mínima por algumas médias como num método iterativo, iniciando com um subconjunto de parâmetros iniciais (Ratkowsky, 1989). Há uma extensa literatura que discute e avalia estes métodos como Chambers (1973) e Schmidt (1982). Entretanto, o conceito de iniciar a convergibilidade, que é, a facilidade de convergência dos parâmetros iniciais estimados, e sua relação com a amplitude na qual um modelo apresenta o comportamento “close-to-linear”, é utilizado considerando somente o método de Marquardt (mais exemplos, veja Bard, 1974), que tem a vantagem sobre os demais métodos, como o método do Gradiente (Steepest-Descent) e de Gauss-Newton, pois agrega as melhores propriedades dos dois, apresentando facilidade e rapidez na convergência. A convergibilidade de um modelo para uma combinação de dados depende intensamente na forma do contorno da idêntica soma de quadrados do resíduo na superfície de resposta esboçada no espaço paramétrico (veja mais informações em Ratkowsky, 1983).

Uma das limitações dos modelos não lineares é a escolha correta dos valores iniciais dos parâmetros para o processo de iteração, os quais irão gerar as estimativas pela convergência do algoritmo baseado, na maioria das vezes, no método de Gauss-Newton (Calegario, 2004). O mesmo autor comenta que quando os valores iniciais dos parâmetros estão distantes das estimativas para a base de dados em questão, o processo de convergência do algoritmo falha e não

ocorre a estimativa, ou o processo gera estimativas não confiáveis. Uma das formas de contornar o problema é usar expressões geradas com base na interpretação dos parâmetros do modelo e, ou, em expressões geradas pelo isolamento do parâmetro em função de outros parâmetros conhecidos e pontos específicos na base de dados.

2.2.2 Medidas de Acuracidade

O teste F é uma importante medida de acuracidade. Este é obtido pela tabela de análise de variância consistindo na divisão do quadrado médio da regressão pelo quadrado médio do resíduo. Quanto maior o F calculado melhor o ajuste do modelo. Este teste F é um teste de significância da regressão que determina se há relação linear entre a resposta y e qualquer uma das variáveis regressoras x_1, x_2, \dots, x_k . O teste de hipótese apropriado é

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$$

$$H_1 : \beta_j \neq 0 \text{ para pelo menos um } j$$

A rejeição de H_0 implica que pelo menos um dos regressores x_1, x_2, \dots, x_k contribui significativamente para o modelo. Rejeita-se H_0 quando o F calculado for maior que o F tabelado. O teste é importante para medir a adequação do modelo mas, para isso, deve-se manter a pressuposição de normalidade dos resíduos.

Outra medida de acuracidade adotada é a Soma de Quadrado do Resíduo (ou do Erro) simbolizada por SQ_{Res} . Quanto menor a SQ_{Res} menor o erro e, assim, melhor o modelo testado.

O coeficiente de determinação (R^2) que também é utilizado na seleção de modelos é dado por:

$$R^2 = \frac{SQ_{Reg}}{SQT} = 1 - \frac{SQ_{Res}}{SQT}$$

onde: SQ_{Reg} = Soma de Quadrados da Regressão; SQT = Soma de Quadrados Total; SQ_{Res} = já fora definida.

É costume pensar que o R^2 é uma medida de redução da variabilidade de y obtida pelo uso das variáveis regressoras x_1, x_2, \dots, x_k . O coeficiente de determinação varia de $0 \leq R^2 \leq 1$, ou de 0 a 100%. Entretanto, maior valor de R^2 não necessariamente implica que o modelo de regressão é melhor. Com a adição de um regressor no modelo sempre aumenta o R^2 , alguns cuidados devem ser tomados no uso de tal índice. É comum modelos com elevados valores de R^2 mas com uma baixa performance na estimativa e predição.

Pode-se também utilizar a raiz quadrada do R^2 a qual se processa pela correlação múltipla entre a variável y e o conjunto de variáveis regressoras x_1, x_2, \dots, x_k . Isto é, R é uma medida de associação linear entre y e x_1, x_2, \dots, x_k . Também pode ser mostrado que o R^2 é o quadrado da correlação entre o vetor de observações y e o vetor de valores ajustados \hat{y} .

Muitas análises oferecem o uso estatístico do R^2 ajustado desde que o R^2 ordinário sempre aumenta (não decresça em média) quando um novo regressor é adicionado ao modelo de regressão. O R^2 ajustado penaliza as análises que incluem, desnecessariamente, variáveis no modelo.

Pode-se definir R^2 ajustado pela substituição de SQ_{Res} e SQT pelos seus correspondentes quadrados médios, isto é

$$R^2_{\text{ajustado}} = 1 - \frac{SQ_{\text{Res}} / (n - p)}{SQT / (n - 1)} = 1 - \frac{(n - 1)}{(n - p)} (1 - R^2)$$

Quando o R^2_{ajustado} e R^2 são muito diferentes para um mesmo modelo, há chance do modelo ter sido super específico, isto é, variáveis que não contribuem significativamente para o ajuste tenham sido incluídos (Montgomery & Peck,

1992). Portanto, ambos devem ser analisados. Esta medida de acuracidade expressa o quanto a variável y é explicada pela(s) variável(is) independente(s).

O Erro Padrão Residual (syx) expressa o quanto, em termos médios, os valores observados variam em relação aos valores estimados. É a partir do syx que se consegue o intervalo de confiança (Scolforo, 1997). O mesmo é dado por

$$syx = \sqrt{QMRes}$$

onde: $QMRes$ = Quadrado Médio do Resíduo ou variável residual.

A unidade do syx é a mesma da variável dependente y . Quanto mais próximo de zero este valor, mais eficiente tende a ser a regressão. Toda vez que ao se ajustar um modelo em que o y esteja ora numa escala ora noutra, é necessário retransformar o syx para que todos fiquem na mesma escala e possam ser comparáveis. Esse novo syx é dado por

$$syx = \sqrt{\frac{\sum (y - \hat{y})^2}{n - p - 1}} = \sqrt{QMRes}$$

O R^2 só pode ser comparado dentro dos modelos não-lineares e dentro dos modelos lineares. Os valores de syx podem ser utilizados para comparação entre os modelos lineares e não-lineares, pois os mesmos estão na mesma dimensão. No nosso caso, h_i em metros. Outras formas de comparação entre os modelos lineares e não-lineares são o gráfico de resíduo (item 2.2.4) e a correlação entre os parâmetros estimados, pois quanto maior a correlação, menos eficiente é o modelo, isto porque dois ou mais parâmetros podem estar explicando a mesma coisa no modelo.

padrão residual (syx). Também foram comparados pela distribuição do gráfico de resíduo.

Ao comparar os modelos não-lineares com os lineares, as medidas de acuracidade utilizadas foram o coeficiente de determinação (R^2) e o erro padrão residual (syx). Foram examinadas as correlações entre os parâmetros dos modelos e o gráfico de resíduo. O melhor modelo linear foi comparado com o melhor modelo não-linear pelo gráfico de perfil longitudinal médio da árvore.

2.4 RESULTADOS E DISCUSSÃO

2.4.1 Modelos Lineares

Nas tabelas 2.2, 2.3 e 2.4 são apresentadas as medidas de acuracidade dos modelos lineares testados para todas as classes diamétricas e para cada classe, usando d/dap como variável dependente.

TABELA 2.2 – Medidas de Acuracidade do polinômio de 5ºGrau para todas as classes diamétricas e para cada classe, usando d/dap como variável dependente.

| Classe Diamétrica | F | SQRes | R ² (%) | R ² _{ajustado} (%) | Syx |
|----------------------|---------------|----------------|--------------------|---|----------------|
| 2 | 1032,49 | 0,1631 | 98,62 | 98,53 | 0,0476 |
| 3 | 830,48 | 0,26613 | 98,06 | 97,95 | 0,05697 |
| 4 | 2206,72 | 0,11169 | 99,12 | 99,07 | 0,03376 |
| 5 | 3200,98 | 0,10014 | 99,23 | 99,2 | 0,02842 |
| 6 | 5077,23 | 0,08459 | 99,41 | 99,39 | 0,02375 |
| 7 | 3080,27 | 0,22222 | 98,62 | 98,58 | 0,03207 |
| 8 | 2543 | 0,27625 | 98,39 | 98,35 | 0,03644 |
| Todas | 8564,6 | 2,26009 | 97,75 | 97,74 | 0,04788 |

Os valores destacados em negrito mostram onde foram menos eficientes as medidas de acuracidade. Pode se notar que o polinômio de 5ºGrau foi menos eficiente na classe 3 para as medidas de acuracidade do teste F e do Erro Padrão Residual (syx). Como era de se esperar, para todas as classes diamétricas, a SQRes, R² e R²_{ajustado} obtiveram os piores resultados devido à heterogeneidade dos dados agrupados nesta única classe. Os valores de R² e o R²_{ajustado} foram semelhantes, demonstrando que este polinômio não é super específico, ou seja, todos os termos contribuem significativamente para o ajuste do modelo.

TABELA 2.3 – Medidas de Acuracidade do polinômio de Potências Fracionárias e Inteiras (Hradetzky, 1976) para todas as classes diamétricas e para cada classe, usando d/dap como variável dependente.

| Classe Diamétrica | F | SQRes | R ² (%) | R ² _{ajustado} (%) | syx |
|----------------------|----------------|----------------|--------------------|---|----------------|
| 2 | 1778,38 | 0,16222 | 98,63 | 98,58 | 0,04682 |
| 3 | 1083,42 | 0,25826 | 98,12 | 98,03 | 0,05578 |
| 4 | 3930,62 | 0,10668 | 99,16 | 99,13 | 0,03266 |
| 5 | 5431,09 | 0,09996 | 99,23 | 99,21 | 0,02817 |
| 6 | 8751,15 | 0,0829 | 99,42 | 99,41 | 0,02335 |
| 7 | 3283,95 | 0,17342 | 98,92 | 98,89 | 0,0284 |
| 8 | 2629,18 | 0,26733 | 98,44 | 98,4 | 0,03585 |
| Todas | 9409,47 | 2,06132 | 97,95 | 97,94 | 0,04572 |

Os valores destacados em negrito mostram onde foram menos eficientes as medidas de acuracidade. Pode se notar que o polinômio de Potências Fracionárias e Inteiras têm um comportamento semelhante àquele observado no polinômio de 5º Grau. Na classe 3, as medidas de acuracidade teste F e o Erro Padrão Residual (syx), demonstraram menor eficiência que as demais. Para todas as classes diamétricas, a SQRes, R² e R²_{ajustado} tiveram os piores resultados devido à heterogeneidade dos dados agrupados nesta única classe. Assim, como o polinômio de quinto grau, o R² e o R²_{ajustado} apresentaram valores semelhantes, demonstrando que este polinômio não é super específico, ou seja, todos os termos contribuem significativamente para o ajuste do modelo.

TABELA 2.4 – Medidas de Acuracidade do polinômio Pires & Calegario (2004) para todas as classes diamétricas e para cada classe, usando d/dap como variável dependente.

| Classe diamétrica | F | SQRes | R ² (%) | R ² _{ajustado} (%) | syx |
|----------------------|---------------|---------|--------------------|---|----------------|
| 2 | 1417,04 | 0,15077 | 98,73 | 98,66 | 0,04545 |
| 3 | 1039,3 | 0,21349 | 98,45 | 98,35 | 0,05102 |
| 4 | 5602,35 | 0,11334 | 99,11 | 99,09 | 0,0335 |
| 5 | 4081,38 | 0,09898 | 99,24 | 99,22 | 0,02814 |

Continua...

| Continuação... | | | | | |
|----------------|---------|---------|-------|-------|---------|
| 6 | 4569,5 | 0,06632 | 99,54 | 99,52 | 0,02117 |
| 7 | 4785,66 | 0,24031 | 98,5 | 98,48 | 0,0332 |
| 8 | 2600,94 | 0,27019 | 98,43 | 98,39 | 0,03604 |
| Todas | 6188,33 | 1,95507 | 98,05 | 98,04 | 0,0446 |

Pode se notar que o polinômio proposto por Pires & Calegario (2004) possui um comportamento semelhante àquele observado no polinômio de 5ºGrau e no Polinômio de Potências Fracionárias e Inteiras . Na classe 3, as medidas de acuracidade relativas ao teste F e ao Erro Padrão Residual (syx) mostraram uma menor eficiência que as demais classes. Para todas as classes diamétricas, a SQRes, R^2 e $R^2_{ajustado}$ tiveram os piores resultados devido à heterogeneidade dos dados agrupados nesta única classe. Uma observação importante é quanto à comparação do R^2 e o $R^2_{ajustado}$, que apresentaram valores semelhantes, demonstrando que este polinômio não é super específico, ou seja, todos os termos contribuem significativamente para o ajuste do modelo.

A análise gráfica de resíduos foi feita para todas as classes diamétricas. Nos gráficos 2.2, 2.3, 2.4 são apresentadas as análises de resíduo para os polinômios de 5ºGrau, Hradetzky e Pires & Calegario (2004), respectivamente, representando os resíduos versus os valores estimados.

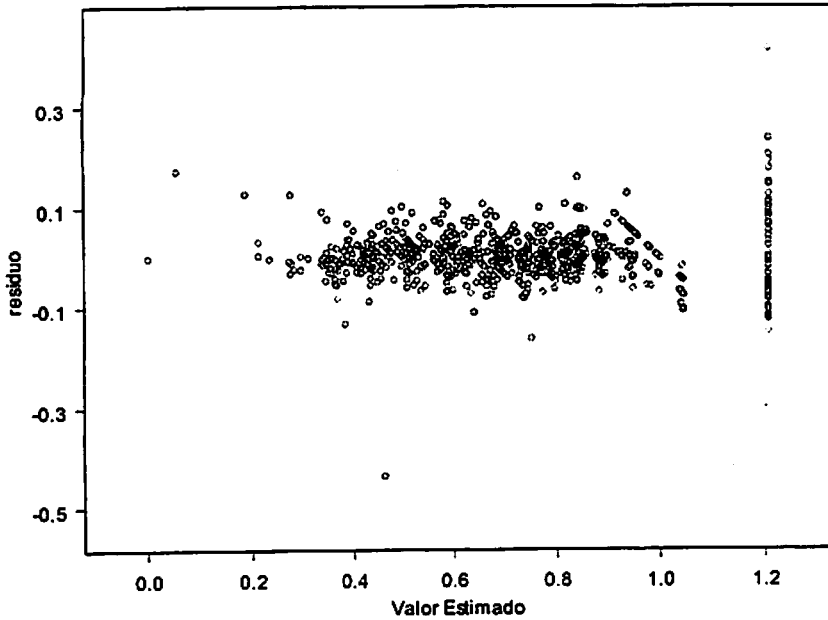


Figura 2.2- Resíduos em função do valores estimados, para o polinômio de 5º grau, considerando todas as classes diamétricas, usando d/d_{ap} como variável dependente.

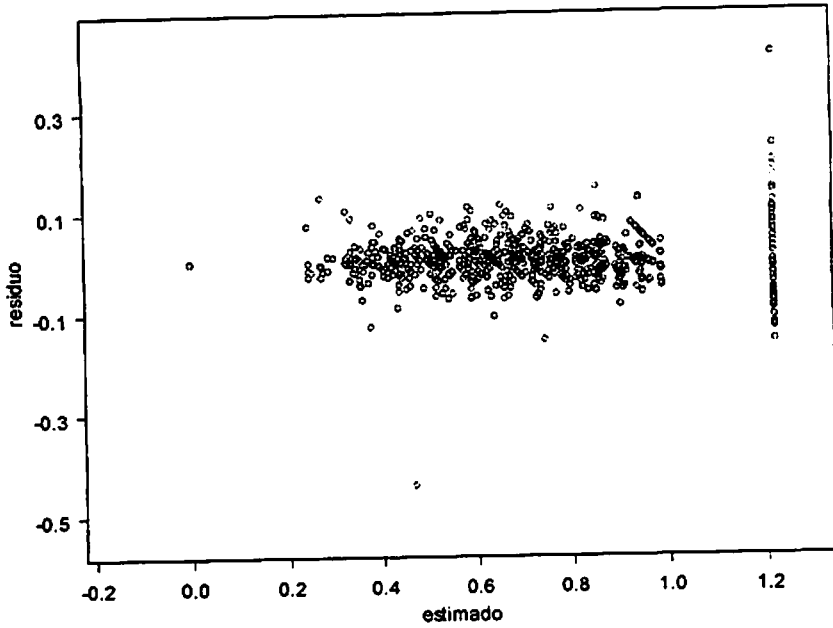


Figura 2.3- Resíduos versus os valores estimados, para o polinômio de Hradetzky (1976), com todos os dados ajustados, usando d/dap como variável dependente.

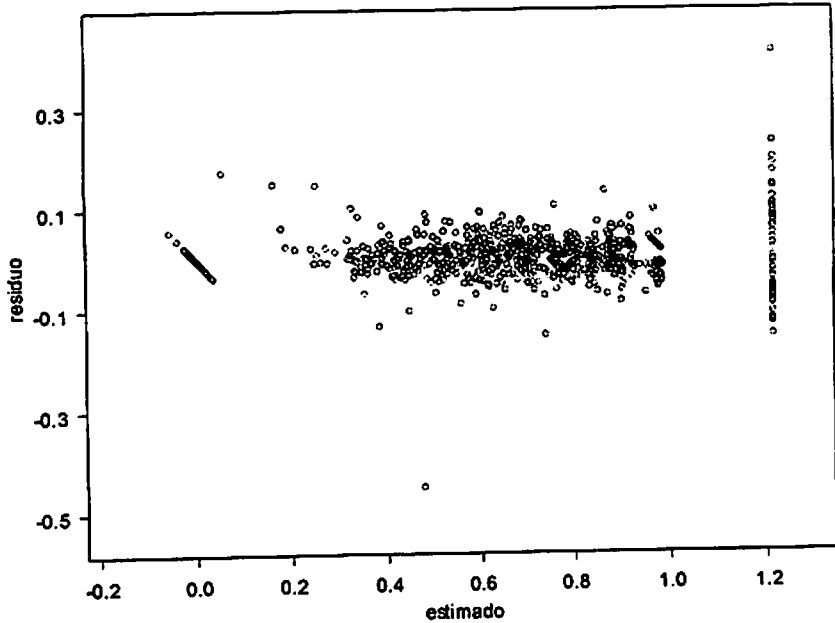


Figura 2.4- Resíduos versus os valores estimados, para o polinômio proposto por Pires & Calegario (2004), com todos os dados ajustados, usando d/dap como variável dependente.

Para se fazer uma avaliação mais apurada do resíduo, o mesmo foi dividido em 3 partes, da seguinte forma: nos resíduos de -0,1 a 0,1 e nos valores estimados de 0,0 a 0,4, de 0,4 a 0,8 e de 0,8 a 1,2. Em seguida foram avaliados cada uma destas seções, em caráter qualitativo, onde obteve-se o resultados apresentados na tabela 2.5.

TABELA 2.5 – Avaliação do gráfico de resíduo em cada seção.

| Polinômio | Seções | | |
|-------------------|---------------|---------------|---------------|
| | 0,0 a 0,4 | 0,4 a 0,8 | 0,8 a 1,2 |
| 5ºGrau | Intermediário | Intermediário | Pior |
| Hradetzky | Melhor | Intermediário | Intermediário |
| Pires & Calegario | Pior | Melhor | Melhor |

Ao analisar detalhadamente o gráfico de resíduo, pode-se notar que o polinômio de Potências Fracionárias e Inteiras (Hradetzky, 1976) apresentou melhor distribuição de resíduo, mais estável, que os demais polinômios testados, apesar do Polinômio Pires & Calegario (2004) ser melhor nos demais itens. Contudo, este último será utilizado para validação e diagnóstico dos dados no próximo capítulo.

2.4.2 Comparação dos Modelos Lineares e Não Lineares

Ao comparar os modelos não-lineares com os modelos lineares, apresenta-se, primeiramente, os parâmetros estimados dos mesmos, na tabela 2.6.

TABELA 2.6 – Parâmetros estimados dos modelos não-lineares de Gompertz, Logístico, Weibull, e o modelos lineares do 5º Grau, Hradetzky, Pires & Calegario (2004), no agrupamento das classes 2, 3 e 4, 5 e 6, 7 e 8.

| Modelo | parâmetro | Classes | | |
|-----------|-----------|----------|---------|---------|
| | | 2, 3 e 4 | 5 e 6 | 7 e 8 |
| Gompertz | ϕ_1 | 23,9399 | 19,6503 | 27,4812 |
| | ϕ_2 | 0,7011 | 0,1530 | 0,0985 |
| | ϕ_3 | -0,1348 | -0,2108 | -0,1849 |
| Logístico | ϕ_1 | 15,4157 | 17,4684 | 24,9916 |
| | ϕ_2 | 3,6851 | 7,8437 | 11,1039 |
| | ϕ_3 | -3,043 | -2,6147 | -3,1719 |
| Weibull | ϕ_1 | 11,9351 | 16,694 | 24,5205 |
| | ϕ_2 | 71,0524 | 29,9411 | 67,1983 |
| | ϕ_3 | 4,2417 | 13,381 | 14,8135 |
| | ϕ_4 | -0,339 | -1,133 | -0,8993 |

Continua...

Continuação...

| | | | | |
|--------------------------|-----------|----------|----------|-----------|
| 5º Grau | β_0 | 11,93811 | 16,75807 | 25,03139 |
| | β_1 | -2,59804 | -0,27774 | 0,01588 |
| | β_2 | 0,70141 | -0,02217 | -0,1334 |
| | β_3 | -0,09485 | -0,00883 | 0,01132 |
| | β_4 | 0,00527 | 0,000705 | -0,00051 |
| | β_5 | -0,0001 | -1,4E-05 | 8,48E-06 |
| Hradetzky | β_0 | 11,94103 | 16,75692 | 25,02115 |
| | β_1 | -99577 | 4,22455 | 156,2809 |
| | β_2 | 101163 | -1,26262 | 43,57733 |
| | β_3 | 15,83486 | 1,35E-06 | -164,782 |
| | β_4 | -1674,83 | - | -17,6298 |
| | β_5 | - | - | 2,66E-21 |
| | β_6 | - | - | -7,87E-36 |
| | β_7 | - | - | 1,91E-43 |
| Pires e Calegario | β_0 | 11,93265 | 16,76795 | 25,02479 |
| | β_1 | -0,28236 | -2,77968 | 1,62755 |
| | β_2 | -1,90884 | 0,17838 | -0,06305 |
| | β_3 | 0,16635 | -0,03005 | 6,34E-05 |
| | β_4 | -0,00048 | -0,00022 | -0,00014 |
| | β_5 | 0,000107 | 8,62E-05 | 3,89E-06 |
| | β_6 | 4,27E-12 | - | -1,91E-14 |

Onde: ϕ_{is} e β_{is} = parâmetros estimados.

As variáveis selecionadas no polinômio Pires & Calegario (2004), e as potências selecionadas nos polinômios de Hradetzky, encontram-se em anexo. A obtenção da estimativa dos parâmetros dos modelos lineares é feita baseada no método dos mínimos quadrados ordinários, ou seja, minimizando a soma do quadrado do erro de forma matemática explícita.

O modelo logístico apresentou maior facilidade de interpretação de parâmetros e, conseqüentemente, facilidade no ajuste dos mesmos. O primeiro parâmetro representa a assintótica horizontal superior, sendo o mesmo para todos os outros modelos não-lineares, a qual pode ser visualizada na Figura 2.1. Isto facilita a convergência do modelo Logístico neste primeiro parâmetro. O segundo parâmetro é o ponto de inflexão, que também pode ser observado nas curvas da figura 2.1. O terceiro parâmetro é a escala, a qual é a diferença entre os diâmetros à aproximadamente 73% da reposta média máxima e o diâmetro no ponto de inflexão, resultando em um valor negativo.

No modelo de Gompertz, os dois parâmetros que não representam a assintótica horizontal superior são obtidos por um modelo de regressão linear simples, onde a variável dependente é função do primeiro parâmetro e dos valores observados de altura (h_i) e a variável independente é o diâmetro (d_i). Desta forma são obtidas as estimativas dos valores iniciais dos outros dois parâmetros.

Para o modelo de Weibull, baseado na distribuição de mesmo nome, a seqüência de obtenção dos valores iniciais dos parâmetros é, basicamente, a mesma. O valor inicial do quarto parâmetro é estimado como uma função da altura (h_i) no ponto de inflexão e do primeiro e segundo parâmetro. Maiores detalhes veja Ratkowsky (1983) e Calegario (2004).

Na tabela 2.7 é apresentada as medidas de acuracidade dos modelos lineares e não-lineares.

TABELA 2.7 – Medidas de acuracidade dos modelos lineares e não-lineares, no agrupamento das classes 2, 3, 4; 5 e 6; 7 e 8.

| modelo | Medida de Acuracidade | Classes | | |
|-----------|-----------------------|----------|--------|--------|
| | | 2, 3 e 4 | 5 e 6 | 7 e 8 |
| Gompertz | R^2 (%) | 79,8 | 79,8 | 96,17 |
| | syx | 1,7176 | 1,1502 | 1,4066 |
| Logístico | R^2 (%) | 79,49 | 94,58 | 96,17 |
| | syx | 1,7309 | 1,1605 | 1,4869 |

Continua...

Continuação...

| | | | | |
|--------------------------|----------------|---------|---------|---------|
| Weibull | R^2 (%) | 79,86 | 94,54 | 95,42 |
| | syx | 1,7185 | 1,1671 | 1,5395 |
| 5º Grau | R^2_{aj} (%) | 89,13 | 95,55 | 98,19 |
| | syx | 1,2553 | 1,04846 | 0,96346 |
| Hradetzky | R^2_{aj} (%) | 89,28 | 95,54 | 98,22 |
| | syx | 1,24644 | 1,04981 | 0,95699 |
| Pires e Calegario | R^2_{aj} (%) | 93,96 | 97,76 | 98,55 |
| | syx | 0,93615 | 0,74343 | 0,86268 |

Onde: R^2 = coeficiente de determinação; R^2_{aj} = coeficiente de determinação ajustado; e syx = erro padrão residual.

A Tabela 2.7 apresenta as medidas de acuracidade dos modelos lineares e não-lineares. Os modelos não-lineares foram comparados pelo R^2 e syx. Os modelos lineares forma comparados pelo R^2_{aj} . A comparação entre lineares e não lineares foi feita pelo syx.

O modelo Logístico foi melhor que os demais modelos não-lineares na classe 5 e 6, juntamente com o modelo de Gompertz, na classe 7 e 8. Na classe 2, 3 e 4, o melhor foi o modelo de Weibull, apresentando maior valor de (R^2). Ainda dentre os modelos não-lineares, o melhor, em termos de syx (erro padrão residual), foi o modelo de Gompertz, para todas as classes avaliadas, porém todos obtiveram valores muito próximos entre si, com pequenas diferenças.

Para os modelos lineares, o melhor em todas as classes avaliadas, para R^2_{aj} e syx, foi o modelo de Pires & Calegario (2004), que obteve maior valor de (R^2_{aj}) em todas as classes e menor valor de syx.

Ao comparar os modelos não-lineares com os lineares, utilizando o erro padrão residual (syx), o melhor de todos os modelos foi o modelo linear Pires & Calegario (2004), para todas as classes avaliadas.

Vale comentar que apesar da melhor performance dos modelos lineares em termos de syx, os modelos não lineares possuem as seguintes características desejáveis: parcimoniosidade, interpretabilidade e extrapolabilidade.

A seguir, são apresentadas as tabelas 2.8, 2.9 e 2.10, que avaliaram a correlação entre os parâmetros estimados dos modelos de regressão linear e não-linear.

TABELA 2.8 – Correlações entre os parâmetros estimados pelos modelos lineares e não-lineares, na agrupamento das classes 2, 3 e 4.

| Modelos / Parâmetros | β_1 | β_2 | β_3 | β_4 | β_5 | β_6 | |
|----------------------|-----------|-----------|-----------|-----------|-----------|-----------|---------|
| 5º Grau | b1 | 1 | -0,9938 | 0,9799 | -0,9607 | 0,938 | - |
| | b2 | -0,9938 | 1 | -0,9957 | 0,9844 | -0,9681 | - |
| | b3 | 0,9799 | -0,9957 | 1 | -0,9964 | 0,9867 | - |
| | b4 | -0,9607 | 0,9844 | -0,9964 | 1 | -0,9969 | - |
| | b5 | 0,938 | -0,9681 | 0,9867 | -0,9969 | 1 | - |
| Hradetzky | b1 | 1 | -1 | -0,9939 | 0,9997 | - | - |
| | b2 | -1 | 1 | 0,994 | -0,9997 | - | - |
| | b3 | -0,9939 | 0,994 | 1 | -0,9964 | - | - |
| | b4 | 0,9997 | -0,9997 | -0,9964 | 1 | - | - |
| Pires | b1 | 1 | -0,3235 | -0,8851 | 0,8225 | 0,1868 | -0,7602 |
| | b2 | -0,3235 | 1 | -0,124 | 0,1543 | -0,8464 | 0,1977 |
| | b3 | -0,8851 | -0,124 | 1 | -0,9764 | 0,2041 | 0,772 |
| | b4 | 0,8225 | 0,1543 | -0,9764 | 1 | -0,2696 | -0,8236 |
| | b5 | 0,1868 | -0,8464 | 0,2041 | -0,2696 | 1 | -0,1847 |
| | b6 | -0,7602 | 0,1977 | 0,772 | -0,8236 | -0,1847 | 1 |
| Gompertz | b1 | 1 | 0,99549 | 0,98253 | - | - | - |
| | b2 | 0,99549 | 1 | 0,99331 | - | - | - |
| | b3 | 0,98253 | 0,99331 | 1 | - | - | - |
| Logístico | b1 | 1 | -0,9820 | -0,9217 | - | - | - |
| | b2 | -0,9820 | 1 | 0,94605 | - | - | - |
| | b3 | -0,9217 | 0,94605 | 1 | - | - | - |
| Weibull | b1 | 1 | 0,05182 | 0,06829 | -0,0373 | - | - |
| | b2 | 0,05182 | 1 | 0,99804 | 0,98574 | - | - |
| | b3 | 0,06829 | 0,99804 | 1 | 0,97381 | - | - |
| | b4 | -0,0373 | 0,98574 | 0,97381 | 1 | - | - |

TABELA 2.9 – Correlações entre os parâmetros estimados pelos modelos lineares e não-lineares, no agrupamento das classes 5 e 6.

| Modelos/Parâmetros | β_1 | β_2 | β_3 | β_4 | β_5 | |
|--------------------|-----------|-----------|-----------|-----------|-----------|---------|
| 5ºGrau | β_1 | 1 | -0,9908 | 0,9739 | -0,9534 | 0,931 |
| | β_2 | -0,9908 | 1 | -0,9952 | 0,9835 | -0,9678 |
| | β_3 | 0,9739 | -0,9952 | 1 | -0,9964 | 0,9874 |
| | β_4 | -0,9534 | 0,9835 | -0,9964 | 1 | -0,9972 |
| | β_5 | 0,931 | -0,9678 | 0,9874 | -0,9972 | 1 |
| Hradetzky | β_1 | 1 | -0,821 | -0,8623 | - | - |
| | β_2 | -0,821 | 1 | 0,651 | - | - |
| | β_3 | -0,8623 | 0,651 | 1 | - | - |
| Pires | β_1 | 1 | -0,2522 | -0,483 | 0,6175 | -0,5967 |
| | β_2 | -0,2522 | 1 | -0,7082 | -0,6415 | 0,597 |
| | β_3 | -0,483 | -0,7082 | 1 | 0,0399 | -0,0439 |
| | β_4 | 0,6175 | -0,6415 | 0,0399 | 1 | -0,9448 |
| | β_5 | -0,5967 | 0,597 | -0,0439 | -0,9448 | 1 |
| Gompertz | β_1 | 1 | 0,9321 | 0,8785 | - | - |
| | β_2 | 0,9321 | 1 | 0,9852 | - | - |
| | β_3 | 0,8785 | 0,9852 | 1 | - | - |
| Logístico | β_1 | 1 | -0,8948 | -0,6753 | - | - |
| | β_2 | -0,8948 | 1 | 0,6788 | - | - |
| | β_3 | -0,6753 | 0,6788 | 1 | - | - |
| Weibull | β_1 | 1 | 0,2631 | -0,4065 | 0,2861 | - |
| | β_2 | 0,2631 | 1 | -0,9401 | 0,9814 | - |
| | β_3 | -0,4065 | -0,9401 | 1 | -0,9835 | - |
| | β_4 | 0,2861 | 0,9814 | -0,9835 | 1 | - |

TABELA 2.10 – Correlações entre os parâmetros estimados pelos modelos lineares e não-lineares, no agrupamento das classes 7 e 8.

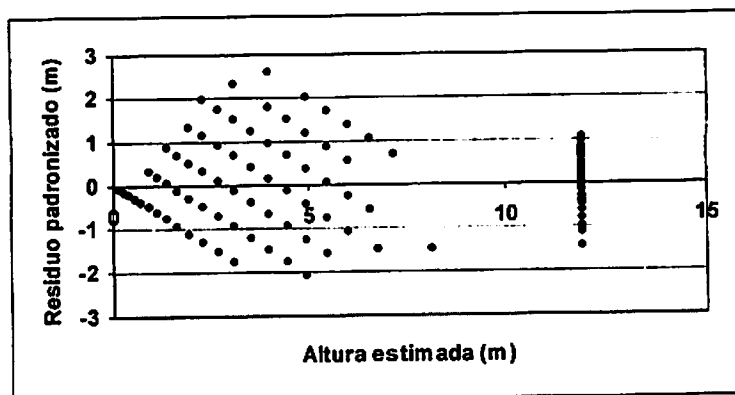
| Modelos / Parâmetros | b1 | b2 | b3 | b4 | b5 | b6 | b7 | |
|------------------------------|----|--------|--------|-------|--------|--------|-------|--------|
| 5ºGrau | b1 | 1 | -0,983 | 0,956 | -0,927 | 0,8979 | - | - |
| | b2 | -0,982 | 1 | -0,99 | 0,977 | -0,957 | - | - |
| | b3 | 0,9559 | -0,993 | 1 | -0,995 | 0,9843 | - | - |
| | b4 | -0,926 | 0,977 | -1 | 1 | -0,996 | - | - |
| | b5 | 0,8979 | -0,957 | 0,984 | -0,997 | 1 | - | - |
| Hradetzky | b1 | 1 | 0,9944 | -1 | -0,992 | 0,6455 | -0,54 | 0,5039 |
| | b2 | 0,9944 | 1 | -1 | -1 | 0,6923 | -0,58 | 0,5465 |
| | b3 | -0,999 | -0,997 | 1 | 0,9948 | -0,657 | 0,548 | -0,515 |
| | b4 | -0,991 | -1 | 0,995 | 1 | -0,703 | 0,592 | -0,557 |
| | b5 | 0,6455 | 0,6923 | -0,66 | -0,703 | 1 | -0,98 | 0,958 |
| | b6 | -0,537 | -0,581 | 0,548 | 0,5917 | -0,976 | 1 | -0,997 |
| | b7 | 0,5039 | 0,5465 | -0,51 | -0,557 | 0,958 | -1 | 1 |
| Pires e Calegario | b1 | 1 | -0,982 | 0,916 | -0,481 | 0,3752 | -0,62 | - |
| | b2 | -0,981 | 1 | -0,96 | 0,52 | -0,413 | 0,699 | - |
| | b3 | 0,9157 | -0,962 | 1 | -0,698 | 0,6017 | -0,84 | - |
| | b4 | -0,481 | 0,52 | -0,7 | 1 | -0,985 | 0,822 | - |
| | b5 | 0,3752 | -0,414 | 0,602 | -0,986 | 1 | -0,8 | - |
| | b6 | -0,623 | 0,6991 | -0,84 | 0,8217 | -0,801 | 1 | - |
| Gompertz | b1 | 1 | 0,9241 | 0,875 | - | - | - | - |
| | b2 | 0,9241 | 1 | 0,988 | - | - | - | - |
| | b3 | 0,8752 | 0,9878 | 1 | - | - | - | - |
| Logístico | b1 | 1 | -0,901 | -0,72 | - | - | - | - |
| | b2 | -0,901 | 1 | 0,7 | - | - | - | - |
| | b3 | -0,719 | 0,72 | 1 | - | - | - | - |
| Weibull | b1 | 1 | 0,4001 | 0,479 | -0,595 | - | - | - |
| | b2 | 0,4001 | 1 | 0,985 | -0,94 | - | - | - |
| | b3 | 0,4785 | 0,9852 | 1 | -0,983 | - | - | - |
| | b4 | -0,595 | -0,94 | -0,98 | 1 | - | - | - |

Pode-se notar que o polinômio de 5ºGrau apresentou altíssima correlação entre seus parâmetros em todas as classes diamétricas agrupadas, indicando que há um número excessivo de parâmetros. Considerando somente o agrupamento 2, 3 e 4, observa-se que o polinômio de Hradetzky, Pires & Calegario (2004), modelo de Gompertz, Logístico e o modelo de Weibull,

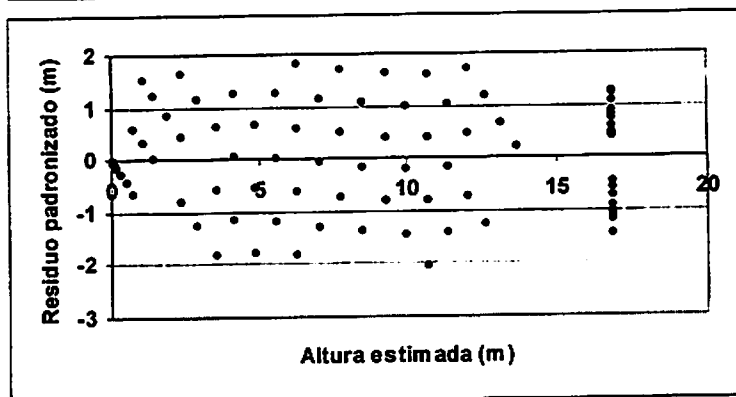
apresentaram valores altíssimos de correlação, indicando um certo excesso de parâmetros. A alta correlação entre os parâmetros é devido a junção das classes na mesma equação, podendo-se concluir que as classes diferem entre si significativamente, devendo ser testadas separadamente. Outra explicação é o comportamento das mesmas, conforme figura 2.1. Pode haver valores estranhos em um pequeno grupo de dados, dificultando assim a modelagem dos mesmos. No primeiro grupo (classes 2, 3 e 4), em termos de correlação entre os parâmetros, o melhor modelo foi o polinômio proposto por Pires & Calegario (2004). No agrupamento das classes 5 e 6, somente os modelos Logístico e o polinômio de Hradetzky devem ser usados (correlação abaixo de 0,90). Mesmo assim, o Hradetzky foi superior ao Logístico. Finalmente, no agrupamento das classes 7 e 8, pode-se observar que somente o modelo Logístico apresentou, para todos os seus parâmetros, uma correlação abaixo de 0,902, sendo a melhor escolha dentre os demais. De maneira geral, a correlação entre os parâmetros foi menor nos modelos não-lineares.

Nas Figuras 2.5, 2.6, 2.7, 2.8, 2.9 e 2.10, estão representados os gráficos dos resíduos padronizados versus a altura estimada, para todos os modelos lineares e não-lineares, nos agrupamentos das classes (2, 3 e 4), (5 e 6), (7 e 8).

Agrupamento
das classes
2, 3 e 4



Agrupamento
das classes
5 e 6



Agrupamento
das classes
7 e 8

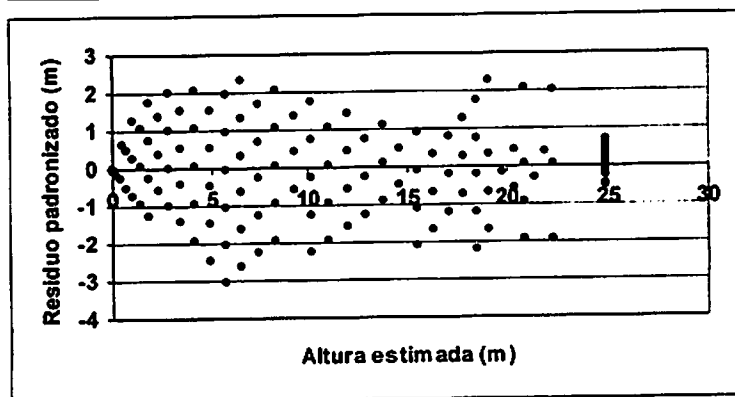
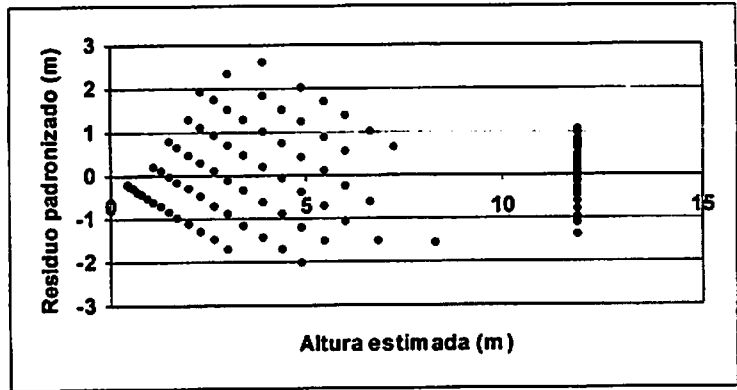
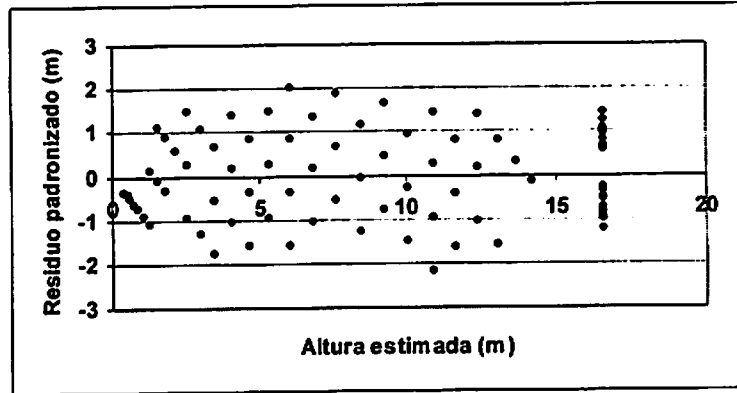


Figura 2.5 – Residuo, para o modelo de Gompertz, nos agrupamentos das classes (2, 3 e 4), (5 e 6) e (7 e 8).

Agrupamento das classes 2, 3 e 4



Agrupamento das classes 5 e 6



Agrupamento das classes 7 e 8

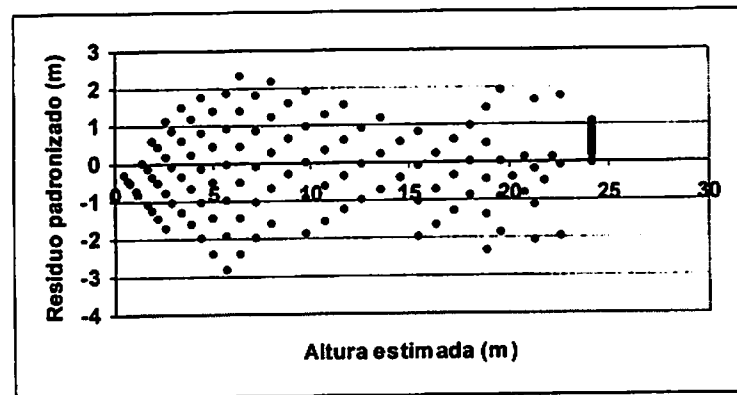
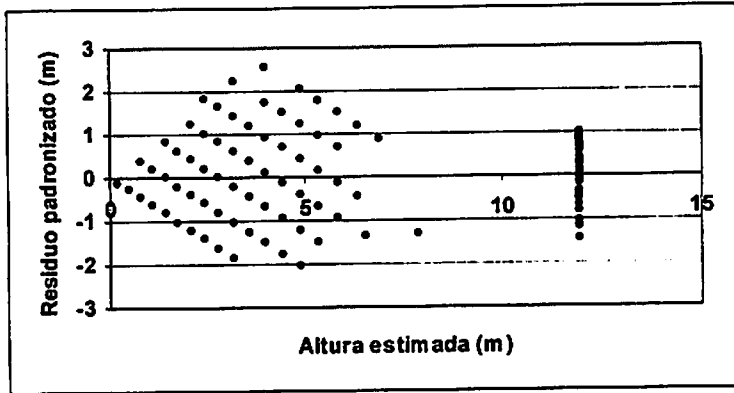
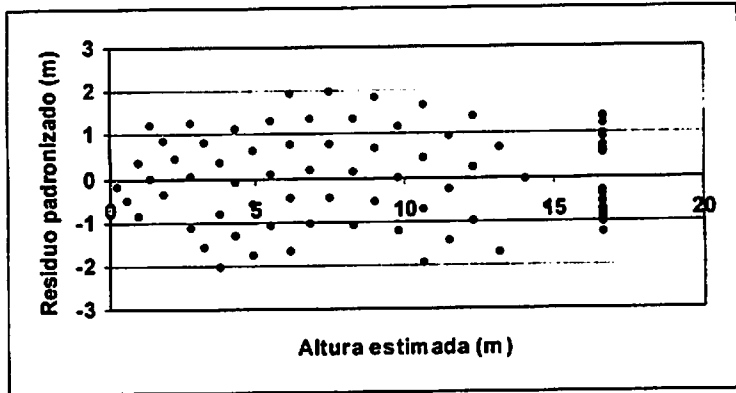


Figura 2.6 – Resíduo, para o modelo Logístico, nos agrupamentos das classes (2, 3 e 4), (5 e 6) e (7 e 8).

Agrupamento
das classes
2, 3 e 4



Agrupamento
das classes
5 e 6



Agrupamento
das classes
7 e 8

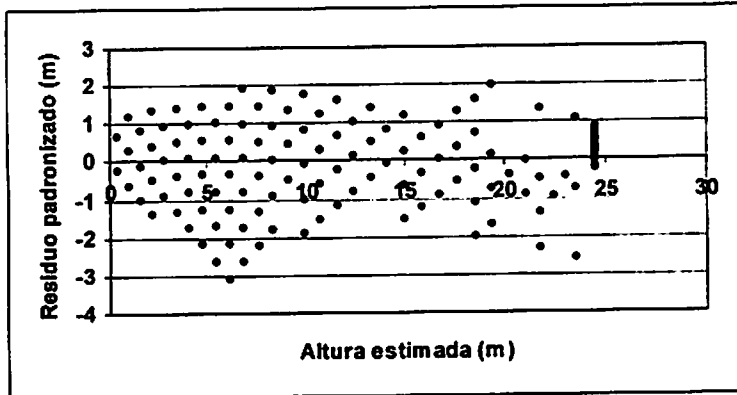
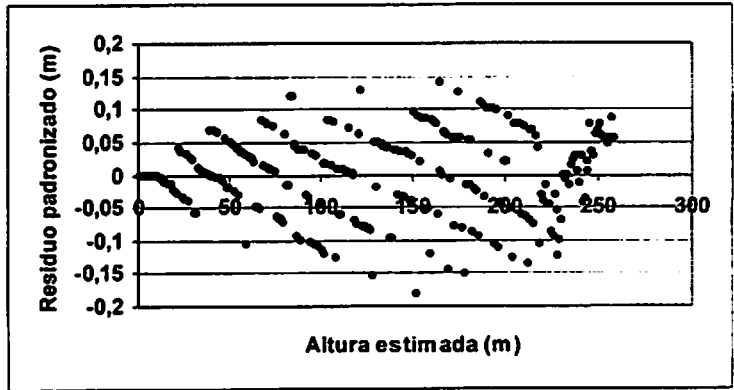
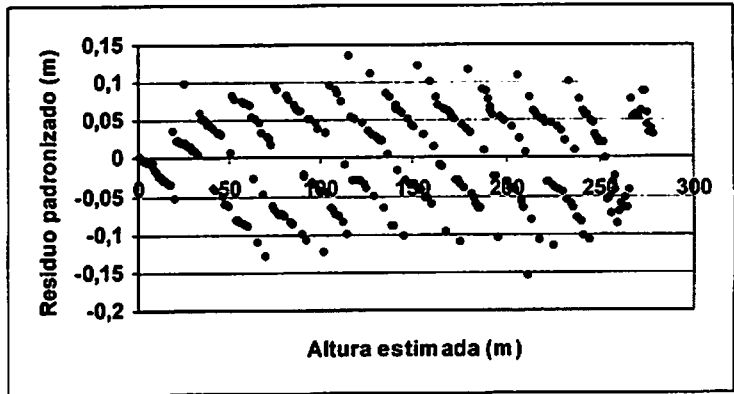


Figura 2.7 –Resíduo, para o modelo de Weibull, nos agrupamentos das classes (2, 3 e 4), (5 e 6) e (7 e 8).

Agrupamento
das classes
2, 3 e 4



Agrupamento
das classes
5 e 6



Agrupamento
das classes
7 e 8

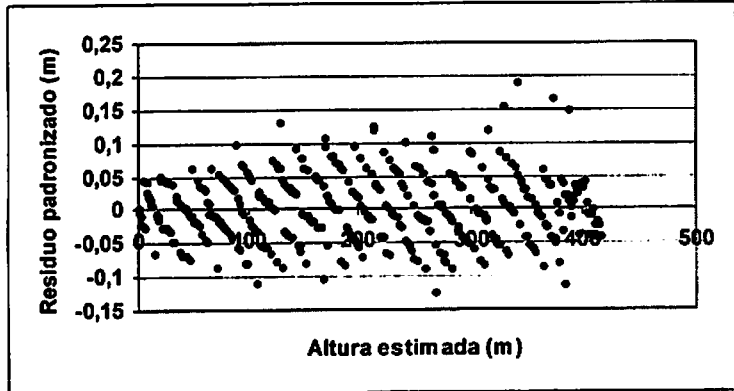
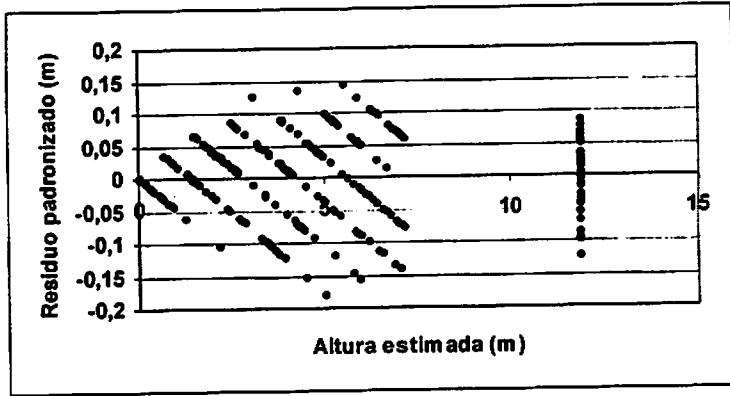
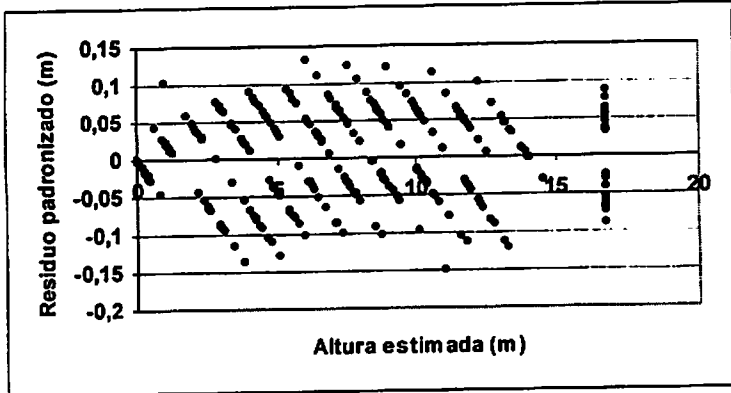


Figura 2.8 –Resíduo, para o polinômio do 5ºGrau, nos agrupamentos das classes (2, 3 e 4), (5 e 6) e (7 e 8).

Agrupamento
das classes
2, 3 e 4



Agrupamento
das classes
5 e 6



Agrupamento
das classes
7 e 8

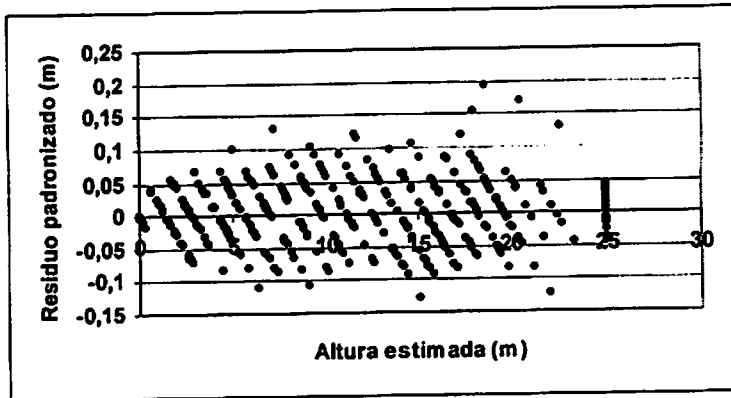
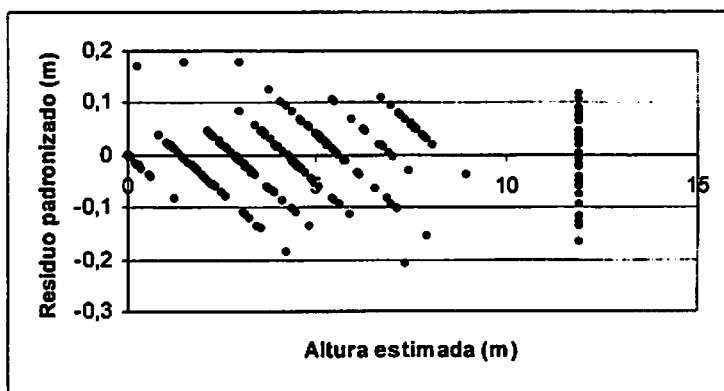
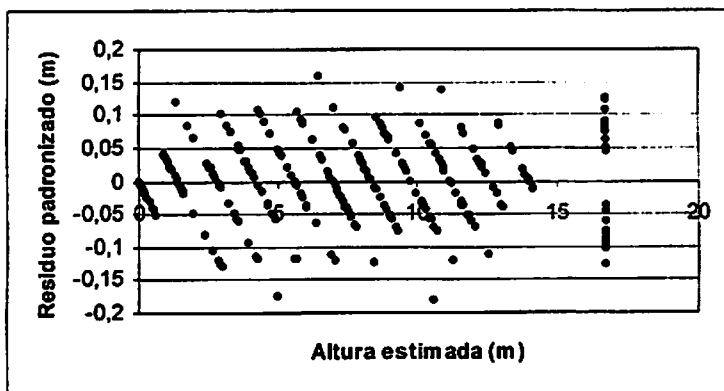


Figura 2.9 –Resíduo, para o polinômio de Potências Fracionárias e Inteiras (Hradetzky), nos agrupamentos das classes (2, 3 e 4), (5 e 6) e (7 e 8).

Agrupamento
das classes
2, 3 e 4



Agrupamento
das classes
5 e 6



Agrupamento
das classes
7 e 8

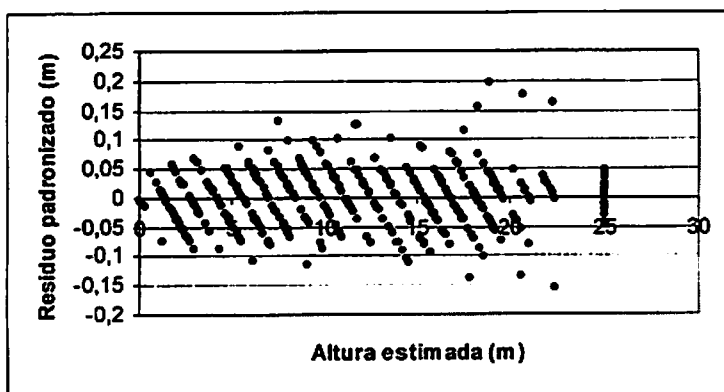


Figura 2.10 – Resíduo, para o polinômio proposto por Pires & Calegario, nos agrupamentos das classes (2, 3 e 4), (5 e 6) e (7 e 8).

Pode-se notar nos gráficos de resíduo padronizado acima, Figuras 2.5, 2.6, 2.7, 2.8, 2.9 e 2.10, que o comportamento do resíduo para os modelos não-

lineares e os modelos lineares foi semelhante, porém em escalas diferentes, para todos os agrupamentos de classe. O polinômio de 5º Grau teve problemas na distribuição do resíduo pois, o mesmo apresentou uma dispersão maior dos resíduos no sentido horizontal que os demais modelos. Os modelos lineares obtiveram uma junção maior dos dados, enquanto os modelos não-lineares estes resíduos ficaram mais espaçados entre si. Nos agrupamentos de classes 5 e 6, e 7 e 8, não houve maiores problemas, em nenhum modelo, quanto ao comportamento do resíduo. Já no agrupamento das classes 2, 3 e 4, o resíduo, para todos os modelos, apresentou um formato de um quadrado a 45°. Nos polinômios este formato foi menos característico. Isto demonstra que para este agrupamento de classe, houve algum problema nesta união, ou em seus dados amostrais. O polinômio proposto por Pires & Calegario (2004) apresentou melhores distribuições gráficas do resíduo que os demais modelos, para todas as classes diamétricas testadas na forma de agrupamento.

2.4.2.1 Modificação do Modelo Logístico

Devido às diferenças na matriz de covariância entre os modelos lineares e os não-lineares, o modelo logístico foi modificado para que o mesmo pudesse ser comparável com o melhor modelo linear. Portanto o mesmo é apresentado da seguinte forma:

$$h_{ij} = \frac{\phi_1}{1 + e^{-\left(ht_j \left(\frac{d_{ij}}{dap_j}\right) - \phi_2\right) / \phi_3}} + \varepsilon_{ij} \quad (2.10)$$

onde: ht_j , dap_j , d_{ij} , h_{ij} , e , ϕ_1 , ϕ_2 , ϕ_3 e ε_{ij} já foram definidos anteriormente.

Com as características descritas acima, o modelo logístico apresenta os seguintes parâmetros ajustados e as respectivas medidas de precisão descrita na tabela 2.11.

TABELA 2.11 – Parâmetros estimados do modelo não-linear Logístico e medidas de acuracidade no agrupamento das classes 2, 3 e 4, 5 e 6, 7 e 8.

| Característica | Classes | | | |
|------------------------|-----------|----------|---------|---------|
| | Grupos | 2, 3 e 4 | 5 e 6 | 7 e 8 |
| Parâmetros Estimados | ϕ_1 | 12,6121 | 17,4035 | 24,5516 |
| | ϕ_2 | 7,4049 | 9,8888 | 15,9939 |
| | ϕ_3 | -2,6965 | -3,2364 | -4,3350 |
| Medidas de Acuracidade | R^2 (%) | 88,65 | 95,42 | 97,03 |
| | syx | 1,2874 | 1,0671 | 1,2377 |

Com relação ao ajuste anterior do modelo Logístico, este logicamente obteve novos parâmetros ajustados e suas medidas de precisão foram melhores em todos os grupos. Isto se deve ao incremento que se dá no ajuste ao adicionar novas variáveis de alto grau de relacionamento, como é o caso da altura total e o diâmetro a altura do peito (dap). Ao comparar este ajuste com o melhor modelo linear ajustado, este é o caso do polinômio proposto por Pires & Calegario (2004), observa-se que ainda assim o modelo linear é superior ao modelo Logístico, em todas as medidas de acuracidade e em todos os grupos.

Em seguida, tabela 2,12, há a avaliação da correlação entre os parâmetros estimados no modelo Logístico após alteração da variável independente.

TABELA 2.12 – Correlações entre os parâmetros estimados no modelo não linear Logístico no agrupamento das classes 2, 3 e 4, 5 e 6, 7 e 8.

| Parâmetros | | ϕ_1 | ϕ_2 | ϕ_3 |
|---------------------|----------|-----------|-----------|-----------|
| Classes 2, 3 e 4 | ϕ_1 | 1 | -0,866052 | -0,689326 |
| | ϕ_2 | -0,866052 | 1 | 0,790708 |
| | ϕ_3 | -0,689326 | 0,790708 | 1 |

Continua...

Continuação...

| | | | | |
|--------------------------|----------|-----------|-----------|-----------|
| Classes 5 e 6 | ϕ_1 | 1 | -0,891820 | -0,671941 |
| | ϕ_2 | -0,891820 | 1 | 0,673383 |
| | ϕ_3 | -0,671941 | 0,673383 | 1 |
| Classes 7 e 8 | ϕ_1 | 1 | -0,893479 | -0,716932 |
| | ϕ_2 | -0,893479 | 1 | 0,692015 |
| | ϕ_3 | -0,716932 | 0,692015 | 1 |

Ao observar a tabela 2.12 pode-se notar que o modelo Logístico (2.10) obteve uma menor correlação entre os seus parâmetros, quando comparado com o ajuste (2.8). Outra observação que se faz, ao comparar o modelo Logístico com o polinômio proposto por Pires & Calegario (2004), é que seus maiores valores de correlação entre os parâmetro, são os maiores valores de correlação do polinômio de Pires & Calegario (2004) no agrupamento das classes 2, 3 e 4 (-0,9764), e no agrupamento das classes 7 e 8 (-0,986), e o menor valor de correlação no agrupamento das classes 5 e 6 (-0,7082). Por outro lado, o modelo logístico em todos os agrupamentos, apesar de obter um valor maior nas classes 5 e 6 (-0,8918), obteve valores de correção entre os parâmetros inferior a 0,9, ou seja, é o melhor modelo no que tange a suposição de que as variáveis independentes não são fortemente correlacionadas e que a presença de multicolinearidade é menor que nos demais modelos. Quando a correlação entre os parâmetros é alta, a eficiência destes é significativamente afetada, tornando-os instáveis, conseqüentemente, aumentando a variância da estimativa e o erro-padrão. Isto faz com que o valor de t se reduza, implicando na hipótese de efeito nulo ser aceita, quando deveria ser rejeitada. Outro fato é que os parâmetros se tornam imprecisos porque apresentam elevada sensibilidade a pequenas alterações na base de dados. Portanto, a interpretação dos resultados fica

prejudicada pois torna-se difícil isolar a influência relativa das variáveis explanatórias.

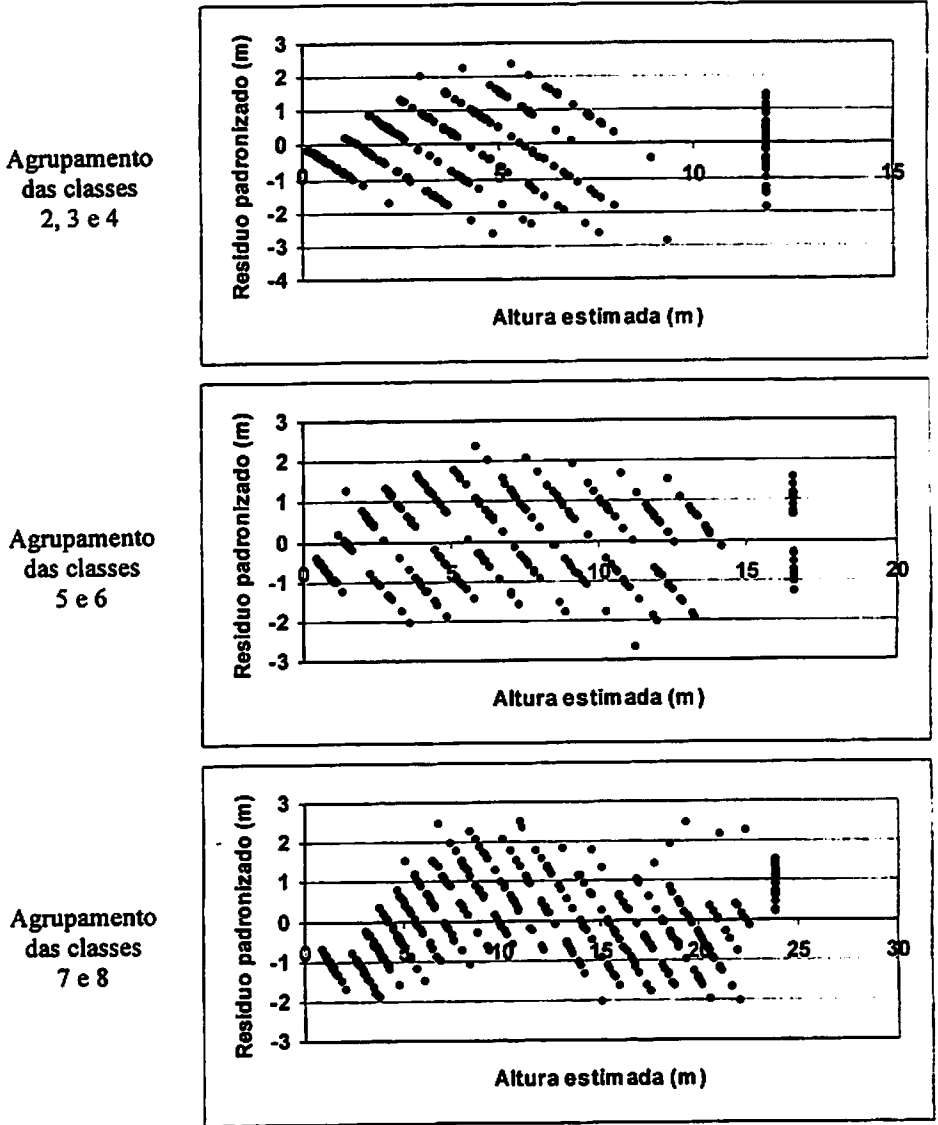


Figura 2.11 – Resíduo, para o modelo Logístico (2.10), nos agrupamentos das classes (2, 3 e 4), (5 e 6) e (7 e 8).

Na figura 2.11 são apresentados os gráficos de resíduo padronizado versus a altura estimada para o modelo logístico (2.10) nos agrupamentos das classes (2, 3 e 4), (5 e 6), (7 e 8).

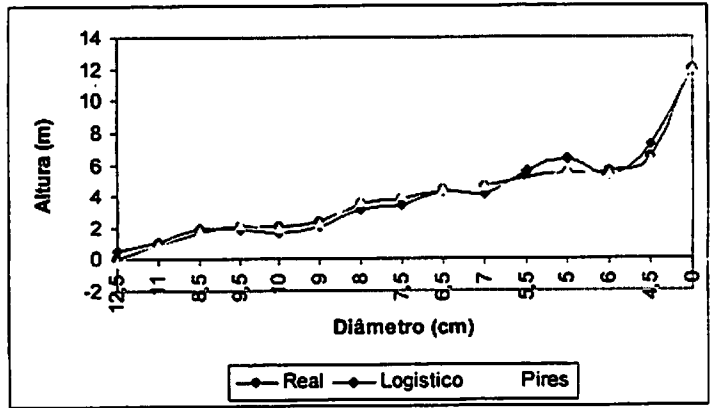
Comparando os gráficos de resíduo padronizado do primeiro e do segundo ajuste do modelo logístico, pode se verificar que o novo ajuste melhorou a distribuição do resíduo no agrupamento das classes 2, 3 e 4. O mesmo obteve uma maior amplitude de erro, em torno de 3 a 6 metros de altura estimada, no agrupamento das classes 5 e 6, que no ajuste anterior e uma menor amplitude de erro no agrupamento das classes 7 e 8. Contudo, como pode-se observar na figura 2.10 em comparação com a figura 2.11, o gráfico de resíduo padronizado do polinômio proposto por Pires & Calegario (2004) continua sendo superior pois apresenta menor amplitude de erro e melhor distribuição do mesmo.

Comparação dos Perfis Longitudinais das Árvores

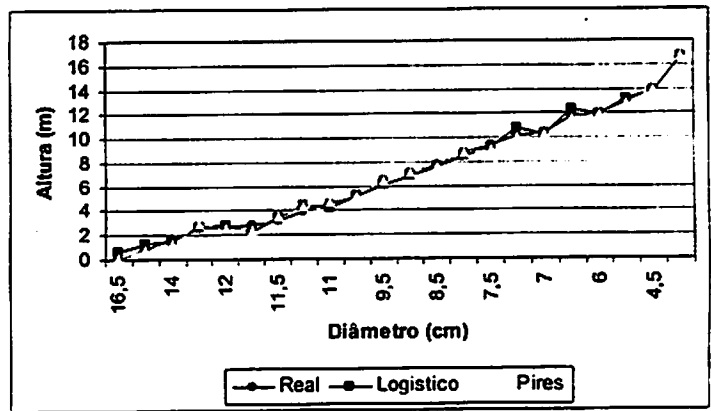
Comparar os perfis longitudinais das árvores é uma forma de verificar a precisão do modelo utilizado e seus respectivos valores estimados com o valor real médio, em cada posição seccional da árvore. Desta forma, pode se antever em qual a posição o modelo é mais preciso e em qual posição é menos preciso.

A figura 2.12 apresenta os gráficos de perfil médio real, e este comparado com o modelo Logístico e o polinômio proposto por Pires & Calegario (2004) nos agrupamentos das classes (2, 3 e 4), (5 e 6) e (7 e 8).

Agrupamento
das classes
2, 3 e 4



Agrupamento
das classes
5 e 6



Agrupamento
das classes
7 e 8

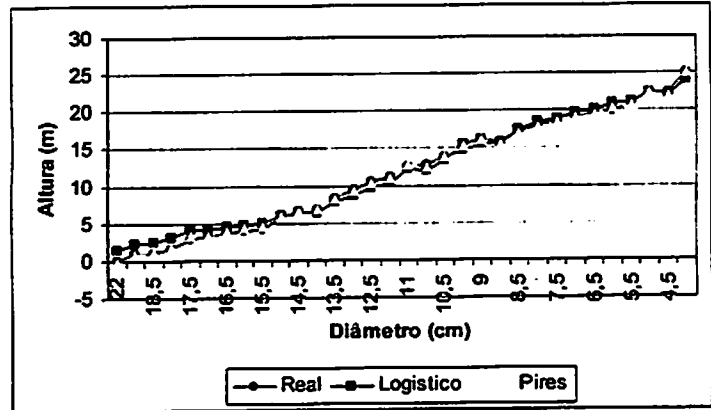


Figura 2.12- Gráficos dos perfis médios reais, do modelo Logístico e do polinômio proposto por Pires & Calegario, nos agrupamentos das classes (2, 3 e 4), (5 e 6) e (7 e 8).

Pode-se verificar na figura 2.12 que o polinômio proposto por Pires & Calegario (2004) se adequou melhor que o modelo logístico, principalmente no agrupamento das classes 2, 3 e 4. Nos demais agrupamentos este continuou superior porém com menor diferença entre este e o modelo Logístico. Também pode se observar que a altura estimada pelo modelo Logístico foi satisfatório nos agrupamentos das classes 2, 3 e 4, com uma pequena tendência no início do tronco, e obteve um resultado bom nos agrupamentos das classes 5 e 6, e classes 7 e 8.

Algumas considerações devem ser feitas ao comparar o modelo Logístico com o polinômio proposto por Pires & Calegario (2004). Primeiramente, o polinômio proposto por Pires & Calegario (2004) possui maior correlação entre os parâmetros podendo ocasionar tendências na estimativa, devido a alta multicolinearidade. Outra observação a ser feita é quanto às “novas” variáveis adicionadas no mesmo (“coni” e “base”), estas proporcionam melhor ajuste do modelo aos dados, pois oferecem uma medida de forma da árvore, porém as mesmas são de difícil obtenção ou pouco operacional a nível de campo. Isto implica diretamente no custo da operação, promovendo um custo adicional.

2.5 CONCLUSÕES

- O melhor modelo linear, utilizando como variável dependente d/dap , foi o polinômio proposto por Pires & Calegario devido às suas melhores medidas de acuracidade.

- O polinômio de Hradetzky, com a variável dependente d/dap , apresentou melhor distribuição dos resíduos de uma maneira geral.

- O polinômio de 5º Grau, com a variável dependente d/dap , apresentou os piores resultados nas medidas de acuracidade e a pior distribuição dos resíduos.

- O polinômio proposto por Pires & Calegario (2004), com a variável dependente h , foi o melhor dentre os polinômios nas medidas de acuracidade, $R^2_{ajustado}$, $syx_{..}$, na distribuição gráfica dos resíduos padronizados, e na correlação entre os parâmetros.

- O modelo Logístico é o melhor modelo não-linear devido à sua baixa correlação entre os parâmetros e melhor interpretação dos mesmos, facilitando no processo de convergência, e uma melhor distribuição dos resíduos padronizados.

- O modelo de Gompertz pode ser utilizado com segurança sendo superior ao modelo de Weibull.

- O modelo de Weibull, com quatro parâmetros, não se apresentou adequado para os dados de cubagem testados, sendo o pior dentre os modelos não-lineares.

- Ao comparar os modelos lineares com os modelos não-lineares, o melhor modelo é o Logístico.

2.6 REFERÊNCIAS BIBLIOGRÁFICAS

- ASSIS, A. L. de. **Acuracidade na estimativa de volumes comerciais de *Eucalyptus grandis* e *Eucalyptus urophylla***. Lavras: UFLA, 1998. 183 p. Monografia.
- ASSIS, A.L. de. **Avaliação de modelos polinomiais segmentados e não-segmentados na estimativa de diâmetros e volumes comerciais de *Pinus taeda***. Lavras: UFLA, 2000.189p. Tese (Mestrado em Ciência Florestal) – Universidade Federal de Lavras, 2000.
- BARD, Y. **Nonlinear Parameter Estimation**, New York: Academic Press. 1974. 320p.
- CALEGARIO, N. **Melhoria da Qualidade de Ajuste de Modelos Biométricos Florestais pelo Emprego da Teoria dos Modelos Não Lineares Generalizados**. IPEF. (Submetido). 2004. 17p.
- CALEGARIO, N.; DANIELS, R.F. **Modeling individual tree profile based on a nonlinear mixed effect model: an application in *Eucalyptus* stands**. 2002. 30p.
- CHAMBERS, J. **Fitting Nonlinear Models: Numerical Techniques**, *Biometrika* London, n.1.,p1-13.,apr.1973.
- CHATTERJEE, S.; PRICE, B. **Regression Analysis by Example**. Nova York: John Wiley . 1977. 408p.
- EMBRAPA – EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA. **Levantamento generalizado e semidetalhado dos solos da Aracruz celulose S. A. no estudo do Espírito Santo e no extremo sul da Bahia e sua aplicação aos plantios de eucalipto**. Rio de Janeiro, 2000. p.13-17
- HRADETZKY, J. **Analyse und interpretation statistischer abhänger keiten (Biometrische Beiträge zu aktuellen forschungs projekten)**. Baden: Württemberg Mitteilungen der FVA, 1976. 146p. (Abt. Biometric und Informatik, 21).
- MONTGOMERY, D.C.; PECK, E.A. **Introduction to linear regression analysis**. Nova York: John Wiley 2 Ed. 1992. 526p.

PIRES, L.M.; CALEGARIO N. **Modelos Estocásticos Lineares para a Descrição do Perfil Longitudinal de Árvores**. 49ºRBRAS. Uberlândia, 2004. P.1-5.

RATKOWSKY, D.A. **Handbook of nonlinear regression models**. New York and Basel: Marcel Dekker, 1989. 241p.

RATKOWSKY, D.A. **Nonlinear Regression Modeling: a Unified Practical Approach**, New York: Marcel Dekker. 1983. ???p.

RIOS, M. S. “A eficiência das funções polinomiais, da função Spline cúbica, e razões de volume para representar o perfil da árvore e estimar os sortimentos de *Pinus elliottii*”.Lavras, 1997.116p. (Dissertação de mestrado em Engenharia Florestal). UFLA.

SCHMIDT, R. **Advances in Nonlinear Parameter Optimization**, Berlim: Springer-Verlag. 1982.420p.

SCOLFORO, J.R.S. “Biometria florestal 2: Técnica de regressão aplicada para estimar: volume, biomassa, relação hipsométrica e múltiplos produtos de madeira.” Lavras: UFLA/FAEPE, 1997. 292p.

THIERSCH, C. R..”Acuracidade dos modelos polinomiais para representar os perfis dos fustes de clones de *Eucalyptus grandis*” Lavras: UFLA, 1999. 35 p. Monografia

CAPÍTULO 3

DIAGNÓSTICO DE MODELOS ESTOCÁSTICOS LINEARES

PIRES, Leonardo Machado. **Diagnóstico de Modelos Estocásticos Lineares**. In: _____: Modelos Estocásticos Lineares e Não-Lineares para a Descrição do Perfil Longitudinal de Árvores. 2004. Cap. 3 p. 59-109. Dissertação (Mestrado em Engenharia Florestal) – Universidade Federal de Lavras, Lavras¹.

RESUMO

Ao analisar um modelo estocástico que descreva o perfil longitudinal da árvore, normalmente, são feitas as medidas de acuracidade e o gráfico de resíduo, para se concluir qual é o melhor modelo a ser adotado. Esta seria a finalização do trabalho, portanto o objetivo deste trabalho é analisar, diagnosticar e validar o modelo de regressão linear e retirar as observações de influência coincidentes para verificação de possíveis melhorias no modelo. Um objetivo secundário é testar a necessidade do controle por classe para o polinômio de 5ºGrau, ou se não é necessário este controle, e observar as condicionantes do resíduo. Na análise, entre os polinômios de 5ºGrau, Hradetzky e o polinômio proposto por Pires & Calegario (2004), utilizou-se as medidas de acuracidade conhecidas, $R^2_{ajustado}$, s_{yx} , o gráfico de resíduo, observou-se os valores estranhos e as observações de influência. O polinômio proposto por Pires & Calegario (2004) foi melhor que os demais, sendo validado e obtendo-se resultados adversos e esperados e, em seguida, foram retiradas as observações de influência coincidentes e “outliers”. Esta retirada proporciono melhoria na qualidade do polinômio proposto por Pires & Calegario (2004). O polinômio de 5ºGrau utilizado para o teste de hipóteses dos parâmetros mostrou ser desnecessária a utilização de um modelo para cada classe, pois um só modelo pode ser usado para todas com excelente precisão. As condicionantes do resíduo foram boas, dentro do esperado, ou seja, em sua grande maioria, a homocedasticidade e a normalidade foram aceitas para todos os polinômios e em todas as classes diamétricas.

¹Comitê de Orientação: Natalino Calegario (Orientador) – UFLA.

as classes (isto só será feito para o polinômio de 5ºGrau), e verificar a presença de multicolinearidade.

3.1 INTRODUÇÃO

A eficiência das funções de afilamento e suas vantagens são claras, pois permitem a obtenção do diâmetro a qualquer altura e vice-versa. Normalmente estas funções são avaliadas baseando-se apenas nas medidas de acuracidade e no gráfico de resíduo. Técnicas estatísticas modernas de diagnóstico a serem aplicadas são pouco conhecidas e utilizadas pelos pesquisadores florestais. Mesmo quando estas são conhecidas, há uma resistência grande destas serem aplicadas e utilizadas.

Formas estatísticas modernas foram desenvolvidas para localização de possíveis distorções no ajuste do modelo, podendo-se citar o trabalho de Montgomery & Peck (1992). Dentre as citadas técnicas estão a validação do modelo, retirada de valores estranhos, diagnósticos de influência, poder preditivo do modelo e verificarão se o modelo necessita de um controle por classe diamétrica ou se somente um modelo com todas as classes diamétricas pode ser usado. Além destas análises, verificar as violações das pressuposições básicas de independência, variância constante e normalidade dos erros é importante para as avaliações das funções de afilamento. Outra análise primordial é a verificação da multicolinearidade, que afeta diretamente na utilização do modelo.

Portanto, o presente capítulo tem como objetivos principais: verificar as condicionantes do resíduo (homogeneidade e normalidade) nos polinômios testados; observar a presença de valores estranhos (“outliers”) nos polinômios e retirá-los se necessário; diagnosticar as influências nos polinômios; validar o melhor polinômio; retirar das observações de influência e verificar as possíveis melhorias na acuracidade do mesmo; testar se há a necessidade de um ajuste para cada classe diamétrica ou se somente um modelo pode ser usado para todas

as classes (isto só será feito para o polinômio de 5ºGrau), e verificar a presença de multicolinearidade.

3.2 REFERENCIAL TEÓRICO

3.2.1 Diagnóstico da regressão e medidas de adequação dos modelos

Avaliar a adequação dos modelos é importante para solucionar possíveis problemas na regressão. Uma situação comum é a presença de um pequeno grupo de observações gerando influência desproporcional no modelo de regressão, isto é, a estimativa dos parâmetros ou previsões podem depender mais da influência do grupo que da maioria dos dados. Então é necessário localizar estes pontos de influência e avaliar seu impacto no modelo. Se estes pontos de influência são valores altos, eles devem ser eliminados. Por outro lado, esta atitude pode ser errônea, pois estes pontos podem controlar as propriedades chave do modelo. Portanto, deve-se estar ciente disto e como isto pode afetar o uso do modelo (Montgomery & Peck, 1992).

3.2.1.1 Pontos de influência segundo critério de Hoaglin e Welsch (1978)

Hoaglin & Welsch (1978) discutem o papel da matriz “hat” $H = X(X'X)^{-1}X'$ na identificação dos pontos de influência (observações influentes). H representa as variâncias e as covariâncias entre \hat{y} e e , desde $V(\hat{y}) = \sigma^2 H$ e $V(e) = \sigma^2 (I - H)$. Os elementos h_{ij} da matriz H podem ser interpretados como a soma de influências mostrada pelo y_i no \hat{y}_j . Portanto, a inspeção dos elementos de H podem mostrar pontos que são, potencialmente, influentes em virtude de suas localizações no espaço x . A atenção é, usualmente, focada nos elementos da diagonal principal de h_{ii} . Desde que $\sum_{i=1}^n h_{ii} = \text{ordem}(H) = \text{ordem}(x) = p$, o tamanho médio de um elemento da

diagonal da matriz H é p/n . Se um elemento da diagonal $h_{ii} > 2p/n$, a observação i é um alto ponto de influência.

Algumas propriedades e usos dos elementos da matriz “hat” no diagnóstico da regressão são discutidos por Belsley, Kuh, e Welsch (1980).

3.2.1.2 “Outliers”

“Outliers” são valores estranhos observados na análise de resíduo. Para detecção destes valores estranhos, deve-se construir um gráfico de resíduo padronizado (maiores detalhes no item 3.2.2) em função do número de observações. Considerando sempre as pressuposições que os resíduos possuem média zero e variância igual a 1. Espera-se que 95% dos valores desta razão estejam entre os limites de $-1,96$ a $+1,96$, quando há um número de observações superior a 120. Os valores acima desta região são considerados “outliers” e só devem ser removidos, se e somente se, a anormalidade constatada não puder ser contornada (Scolforo, 1997).

3.2.1.3 PRESS e R^2_{predito}

A Predição da Soma de Quadrados do Resíduo (PRESS) proposta por Allen[1971, 1974] propõe uma útil escala de resíduos. Para calcular o resíduo PRESS, seleciona-se uma observação, por exemplo i , ajusta o modelo de regressão para os dados remanescentes, $n-1$ observados, e usa-se equação gerada para prever a observação y_i retida. Chamando o valor predito $\hat{y}_{(i)}$, pode-se encontrar o erro de predição para o ponto i como $e_{(i)} = y_i - \hat{y}_{(i)}$. O erro de predição é muitas vezes chamado de iésimo resíduo PRESS. Este procedimento é repetido para cada observação $i = 1, 2, \dots, n$, produzindo um conjunto de n

resíduos PRESS $e_{(1)}, e_{(2)}, \dots, e_{(n)}$. Então a estatística PRESS é definida como a soma de quadrados dos n resíduos PRESS, como em

$$\text{PRESS} = \sum_{i=1}^n e_{(i)}^2 = \sum_{i=1}^n [y_i - \hat{y}_{(i)}]^2$$

Conseqüentemente, o PRESS usado para cada possível subconjunto de $n-1$ observações está como o conjunto de dados estimados, e toda a observação em volta é usada para formar a predição do conjunto de dados.

Inicialmente, parece que o calculo do PRESS requer ajuste de n diferentes regressões. Entretanto, é possível calcular o PRESS para os resultados de um simples ajuste do quadrado médio para todas as n observações. Pode se produzir o i ésimo resíduo PRESS como

$$e_{(i)} = \frac{e_i}{1 - h_{ii}}$$

Conseqüentemente, já que o PRESS é exatamente a soma de quadrados dos resíduos PRESS, uma simples fórmula de computação é

$$\text{PRESS} = \sum_{i=1}^n \left(\frac{e_i}{1 - h_{ii}} \right)^2$$

É fácil perceber que o resíduo PRESS é exatamente o resíduo ponderado de acordo com os elementos da diagonal principal da matriz "hat" h_{ii} . Os resíduos associados com estes pontos para o qual h_{ii} é grande, terá um resíduo PRESS grande. Estes pontos são, geralmente, pontos de alta influência. Em geral, esta grande diferença entre os resíduos e o resíduo PRESS indicará um ponto onde o modelo ajusta os dados satisfatoriamente, porém um modelo construído de fora dos dados fará uma predição ruim.

Finalmente, nota-se que o PRESS pode ser usado para gerar uma aproximação do R^2 predito, da seguinte forma:

$$R^2_{\text{Predito}} = 1 - \frac{\text{PRESS}}{\text{SQT}} \quad (3.4)$$

Esta estatística dá uma grande indicação da capacidade preditiva do modelo de regressão.

3.2.1.4 Influência nos coeficientes da regressão

3.2.1.4.1 Distância de Cook

O critério de Hoaglin e Welsch identifica os pontos que são, potencialmente, influentes em sua localização no espaço x . É desejável considerar as localizações dos pontos e a variável resposta para mensurar a influência. Cook (1977, 1979) sugeriu usar a medida do quadrado das distâncias entre o quadrado médio estimado baseado em todos “ n ” pontos $\hat{\beta}$ e a estimativa obtida pela retirada do i ésimo ponto, convencionado como $\hat{\beta}_{(i)}$. Esta medida de distância pode ser expressa, de uma forma geral, como

$$D_i(M, c) = \frac{(\hat{\beta}_{(i)} - \hat{\beta})' M (\hat{\beta}_{(i)} - \hat{\beta})}{c}, \quad i = 1, 2, \dots, n \quad (3.5)$$

Pontos com alto valor de D_i tem considerável influencia na estimativa do quadrado médio do $\hat{\beta}$. A significância de D_i pode ser avaliada pelo teste $F_{\alpha, p, n-p}$. Se $D_i > F_{0.5, p, n-p}$, então retirando o ponto i vamos mover $\hat{\beta}$ para o limite de uma região de confiança de 50% de β , baseado no conjunto de dados completos. Este é um alto deslocamento e indica que a estimativa do quadrado médio é sensível ao i ésimo ponto dos dados. Desde que $F_{0.5, p, n-p} > 1$, pode se considerar que os pontos no qual $D_i > 1$ como ponto de influência. É ideal que

cada estimativa de $\hat{\beta}_{(i)}$ permaneça dentro do limite de confiança de 10 ou 20% (Montgomery & Peck, 1992).

A estatística D_i pode ser reescrita como

$$D_i = \frac{r_i^2}{p} \frac{V(\hat{y}_i)}{V(e_i)} = \frac{r_i^2}{p} \frac{h_{ii}}{(1-h_{ii})}, \quad i = 1, 2, \dots, n \quad (3.6)$$

Entretanto, pode se ver que, com “p” constante, D_i é o produto do quadrado do iésimo resíduo estudentizado e $h_{ii}/(1-h_{ii})$. Esta razão pode ser mostrada como sendo a distância do vetor x_i para a centróide dos dados remanescentes. Portanto, D_i é um componente que reflete o quanto a precisão do ajuste do modelo na estimativa da iésima observação e medindo o quanto o ponto está distante do resto dos dados. Qualquer componente da expressão (3.6) pode contribuir para um alto valor de D_i .

3.2.1.4.2 DFBETAS_{j,i}

Belsley et al. (1980) sugerem uma estatística que indique o quanto varia os coeficiente da regressão $\hat{\beta}_j$, em unidades de desvio padrão, se a iésima observação estiver em um conjunto a parte. Esta estatística é dada por:

$$DFBETAS_{j,i} = \frac{\hat{\beta}_j - \hat{\beta}_{j(i)}}{\sqrt{S_{(i)}^2 C_{jj}}}$$

Onde C_{jj} é o j-ésimo elemento da diagonal de $(X'X)^{-1}$ e $\hat{\beta}_{j(i)}$ é o j-ésimo coeficiente da regressão, computado sem usar a iésima observação. Valores elevados de $DFBETAS_{j,i}$ indicam que a observação i tem considerável influência no j-ésimo coeficiente da regressão. Note que $DFBETAS_{j,i}$ é uma matriz $n \times p$

que possui informações similares às da composição da informação de influência na medida de distância de Cook.

Na computação de $DFBETAS_{j,i}$ é gerada a matriz $p \times n$

$$R = (X'X)^{-1} X'$$

onde os “n” elementos na j-ésima linha de R produzem a influência das “n” observações na amostra para o $\hat{\beta}_j$. Considerando r'_j como a j-ésima linha de R, então pode ser mostrado que:

$$DFBETAS_{j,i} = \frac{r_{ji}}{\sqrt{r'_j r_j}} \frac{e_i}{S_{(i)}(1-h_{ii})}$$

$$DFBETAS_{j,i} = \frac{r_{j,i}}{\sqrt{r'_j r_j}} \frac{t_i}{\sqrt{1-h_{ii}}}$$

onde t_i é o Resíduo Estudantizado. Note que $DFBETAS_{j,i}$ mede ambas as influências $(r_{j,i}/\sqrt{r'_j r_j})$, o que é uma medida de impacto na i-ésima observação em $\hat{\beta}_j$) e $\frac{t_i}{\sqrt{1-h_{ii}}}$, o qual é o efeito de um resíduo alto. Belsley et al. (1980)

sugerem uma remoção de $2/\sqrt{n}$ para $DFBETAS_{j,i}$; ou seja, se $DFBETAS_{j,i} > 2/\sqrt{n}$, então a i-ésima observação é passível de um exame.

3.2.1.5 Influência nos valores ajustados

Apesar de não ter sido utilizado diretamente neste trabalho, o Resíduo Estudantizado será abordado a seguir.

O resíduo estudantizado r_i é muitas vezes considerado um diagnóstico de outliers. É costume usar o QMRes como uma estimativa da σ^2 no r_i . Isto se refere a uma escala interna do resíduo porque o QMRes é uma estimativa

internamente gerada da σ^2 , a qual é obtida pelo ajuste do modelo para todas as “n” observações. Outro enfoque será usar uma estimativa da σ^2 baseada em um conjunto de dados com a *i*ésima observação removida. Denota-se a estimativa da σ^2 como a obtida por $S_{(i)}^2$. Esta estimativa pode ser obtida pela seguinte expressão:

$$S_{(i)}^2 = \frac{(n-p) \text{QMRes} - \frac{e_i^2}{(1-h_{ii})}}{n-p-1}$$

O QMRes pode ser substituído pela seguinte expressão:

$$t_i = \frac{e_i}{\sqrt{S_{(i)}^2 (1-h_{ii})}}, \quad i=1,2,\dots,n$$

Em muitas situações t_i vai diferir um pouco do resíduo estudentizado r_i . Entretanto, se a *i*ésima observação está influenciando, então $S_{(i)}^2$ pode ser significativamente diferente do QMRes e, com isto, a estatística Resíduo Estudantizado será mais sensível para este ponto. Além disso, t_i segue a distribuição t_{n-p-1} . Deste modo, o Resíduo Estudantizado oferece um procedimento mais formal para detecção de outliers via teste de hipóteses. Pode ser usada a aproximação do tipo Bonferroni e comparar todos os n valores de $|t_i|$ para $t_{(\alpha/2n), n-p-1}$, suprimindo a orientação que considera os outliers. Entretanto, alguns autores acham que esta aproximação formal não é, usualmente, necessária e que somente os valores brutos removidos necessitam ser considerados. Em geral, um diagnóstico visual é mais apropriado, quando comparado a uma estatística mais rigorosa, como teste de hipóteses. Além disso, a detecção de outliers precisa ser considerada simultaneamente com a detecção de observações de influência.

Pode-se investigar a influência da i ésima observação na predição ou no valor ajustado. O diagnóstico utilizado é o

$$DFFITS_i = \frac{\hat{y}_i - \hat{y}_{(i)}}{\sqrt{S_{(i)}^2 h_{ii}}}, \quad i = 1, 2, \dots, n$$

onde $\hat{y}_{(i)}$ é o valor ajustado de y_i obtido sem usar a i ésima observação. O denominador é exatamente uma padronização, desde que $V(\hat{y}_i) = \sigma^2 h_{ii}$. Portanto, $DFFITS_i$ é o número do desvio padrão que o valor ajustado \hat{y}_i representa a observação i é removida.

Computacionalmente, pode-se ter

$$DFFITS_i = \left(\frac{h_{ii}}{1 - h_{ii}} \right)^{1/2} \frac{e_i}{S_{(i)} (1 - h_{ii})^{1/2}}$$

$$DFFITS_i = \left(\frac{h_{ii}}{1 - h_{ii}} \right)^{1/2} t_i$$

onde t_i é o Resíduo Estudantizado. Portanto, $DFFITS_i$ é o valor do Resíduo Estudantizado multiplicado pela influência da i ésima observação $[h_{ii}/(1-h_{ii})]^{1/2}$. Se um ponto nos dados for um "outlier", então o Resíduo Estudantizado terá alta magnitude. Também se um ponto nos dados tiver alta influência, h_{ii} estará concordando com a unidade. Em qualquer um destes casos, o $DFFITS_i$ pode ser alto. Entretanto, se $h_{ii} > 0$, o efeito do Resíduo Estudantizado será moderado. Similarmente, um Resíduo Estudantizado próximo de zero, combinado com um alto valor de influência, poderá produzir um pequeno valor de $DFFITS_i$. Conseqüentemente, $DFFITS_i$ é afetado pela influência e pelo erro de predição. Geralmente, cada observação para a qual $|DFFITS_i| > 2\sqrt{p/n}$ requer atenção.

3.2.1.6 Influência na precisão da estimativa

Os diagnósticos D_i , $DFBETAS_{j,i}$ e $DFFITS_i$ dão uma introspecção sobre o efeito das observações na estimativa dos coeficientes $\hat{\beta}_j$ e os valores ajustados \hat{y}_i . Eles não fornecem qualquer informação sobre a completa precisão da estimativa. É uma prática comum usar o determinante da matriz de covariância como uma conveniente medida escalar de precisão, chamada de “variância generalizada”. A variância generalizada de $\hat{\beta}$ pode ser definida como

$$VG(\hat{\beta}) = |V(\hat{\beta})| = |\sigma^2 (X'X)^{-1}|$$

Para expressar o papel da i ésima observação na precisão da estimativa, pode-se definir

$$COVRATIO_i = \frac{|(X'_{(i)}X_{(i)})^{-1} S^2_{(i)}|}{|(X'X)^{-1} QMRes|}, \quad i = 1, 2, \dots, n$$

Claramente, se $COVRATIO_i > 1$, a i ésima observação melhora a precisão da estimativa, embora se $COVRATIO_i < 1$, a inclusão do i ésimo ponto diminui a precisão. Computacionalmente tem-se que,

$$COVRATIO_i = \frac{(S^2_{(i)})^p}{QMRes^p} \left(\frac{1}{1 - h_{ii}} \right)$$

Note que $[1/(1-h_{ii})]$ é a razão de $|(X'_{(i)}X_{(i)})^{-1}|$ por $|(X'X)^{-1}|$, deste modo, um alto ponto de influência possuirá um alto valor de $COVRATIO_i$. Isto é de se esperar, desde que um alto ponto de influência melhore a precisão, exceto

se o ponto é um “outlier” no espaço y . Se a i ésima observação é um “outlier”, $S_{(i)}^2/QMRes$ será menor do que uma unidade.

Obter valores de $COVRATIO_i$ que possam ser removidos não é uma tarefa simples. Belsley et al. (1980) sugerem que se $o\left(1 + \frac{3p}{n}\right) < COVRATIO_i < \left(1 - \frac{3p}{n}\right)$ o i ésimo ponto deverá ser considerado influente. O menor limite é somente apropriado quando $n > 3p$. Os valores serão removidos de forma mais apropriada em grandes amostras.

3.2.2 Condicionantes do resíduo

Afifi & Azen (1974) comentam que a grande importância da análise dos resíduos é no estudo das violações das pressuposições básicas de independência, variância constante e normalidade dos erros. Dentre estas condicionantes, somente a independência dos resíduos não será tratada neste trabalho, assim considerando os mesmos como independentes.

Para Seber (1977), Mosteller & Tukey (1977) e Weisberg (1980), uma maior ênfase deve ser dada ao estudo dos resíduos padronizados relacionados com \hat{y}_i .

Para Draper e Smith (1966), o resíduo padronizado é dado por

$$d_i = \frac{e_i}{\sqrt{QMRes}}, \quad i=1,2,\dots,n$$

onde $e_i = y_i - \hat{y}_i$. Se a especificação do modelo é correta, os resíduos terão distribuição normal. Neste caso, 95% dos d_i 's deverão situar-se entre -2 e +2. Um desvio substancial destes limites indica uma potencial violação da hipótese de normalidade.

Alguns analistas preferem utilizar o resíduo Estudentizado para verificar a normalidade.

Para o desenvolvimento geral do resíduo Estudentizado, usa-se o vetor de resíduos como

$$e = (I - H)y$$

onde: H é a matriz “hat” (veja extrapolação “oculta” no anexo). A matriz “hat” tem algumas propriedades úteis. Ela é simétrica ($H' = H$) e idempotente ($HH = H$). Similarmente, a matriz I-H que também é simétrica e idempotente. Substituindo $y = X\beta + \varepsilon$ na equação anterior, tem-se

$$\begin{aligned} e &= (I - H)(X\beta + \varepsilon) \\ e &= X\beta - HX\beta + (I - H)\varepsilon \\ e &= X\beta - X(X'X)^{-1}X'X\beta + (I - H)\varepsilon \\ e &= (I - H)\varepsilon \end{aligned}$$

Deste modo, os resíduos são a mesma transformação linear das observações y e dos erros ε .

A matriz de covariância dos resíduos é dada por:

$$\begin{aligned} V(e) &= V[(I - H)\varepsilon] \\ V(e) &= (I - H)V(\varepsilon)(I - H)' \\ V(e) &= \sigma^2(I - H) \end{aligned} \tag{3.7}$$

Desde que $V(\varepsilon) = \sigma^2 I$ e I-H seja simétrica e idempotente. A matriz I-H é, geralmente, não diagonal. Assim, os resíduos têm diferentes variâncias e estão correlacionados.

A variância do iésimo resíduo é dada por:

$$V(\varepsilon_i) = \sigma^2(1 - h_{ii})$$

onde h_{ii} foi definido anteriormente. Desde que, $0 \leq h_{ii} \leq 1$, usando o quadrado médio do resíduo (QMRes) para estimar a variância dos resíduos ao invés das superestimadas de $V(e_i)$. Além disso, já que h_{ii} é uma medida de locação do i ésimo ponto no espaço x , a variância de e_i depende de onde o ponto x_i é falso. Geralmente, pontos próximos ao centro do espaço x tem maior variância (ajuste do quadrado médio ruim) do que os resíduos localizados mais distantes. Violações das pressuposições são muito prováveis nos pontos distantes. Estas violações podem ser difíceis de detecção pela inspeção do e_i (ou d_i) devido aos seus resíduos serem, usualmente, pequenos.

Alguns autores [(Behnken & Draper, 1972), Davies & Hutton (1975), e Huber(1975)] sugerem considerar a heterogeneidade da variância dentro da avaliação no instante de escalonamento do resíduo. Eles recomendam a representação gráfica dos resíduos estudentizados

$$r_i = \frac{e_i}{\sqrt{\text{QMRes}(1-h_{ii})}}, \quad i = 1, 2, 3, \dots, n$$

no lugar de e_i (ou d_i). Os resíduos estudentizados têm variância constante $V(r_i)=1$, indiferente da localização de x_i , quando a forma do modelo é correta. Em muitas situações, a variância dos resíduos estabiliza, particularmente para um grande conjunto de dados. Nestes casos, pode-se ter uma pequena diferença entre os resíduos padronizados e estudentizados. Conseqüentemente, os resíduos padronizados e estudentizados, muitas vezes, conduzem à informações equivalentes. Entretanto, visto que cada ponto com um alto resíduo e um alto h_{ii} é potencialmente um ponto de alta influência, o exame dos resíduos estudentizados é recomendado no ajuste do quadrado médio,.

Para Anscombe (1973), citado por Seber (1977) e Draper & Smith (1966), uma forma de verificar a homocedasticidade de um modelo é utilizar um teste que consista basicamente no ajuste de uma equação de regressão linear

simples, onde \hat{y}_i seja considerada a variável independente e e_i^2 a dependente. A homocedasticidade é aceita se, pelo teste t, o coeficiente β_1 for estatisticamente igual a zero.

3.2.3 Multicolinearidade

A interpretação de uma equação de regressão múltipla depende implicitamente da suposição de que as variáveis independentes não sejam fortemente correlacionadas.

Segundo Mattioli (1983), é usual interpretar o coeficiente da regressão como a quantidade de variação sofrida pela variável dependente, considerando que a variável correspondente a este coeficiente é incrementada em uma unidade e as demais variáveis mantidas constantes. Este tipo de interpretação pode não ter validade se é constatada a existência de acentuadas correlações entre as variáveis independentes, também denominadas explanatórias (Weisberg, 1980). Na maioria dos processos estudados é praticamente impossível modificar o valor de apenas uma variável e manter as demais inalteradas, tornando assim não válida a interpretação usual do coeficiente da regressão como um efeito marginal.

Segundo Chatterjee & Price (1977), pode-se observar alguns indicativos de multicolinearidade no modelo. Estes autores destacam os seguintes:

- 1) Mudanças bruscas nos coeficientes da regressão quando uma variável é incluída ou excluída da equação.
- 2) Mudanças bruscas nos coeficientes da regressão quando uma observação é descartada, substituída ou adicionada. Mesmo que a análise de resíduos tenha sido satisfatória, a multicolinearidade pode estar presente:
- 3) Os sinais algébricos dos coeficientes não estão em conformidade com os esperados, a priori.

4) São grandes os erros padrões dos coeficientes de variáveis consideradas importantes.

Uma forma visual de detecção de multicolinearidade é a verificação da correlação entre os elementos da diagonal principal da matriz $(X'X)^{-1}$. Dentre as formas algébricas os “fatores de inflação da variância” (VIF) são importantes diagnósticos de multicolinearidade. De uma forma geral, VIF para o j-ésimo coeficiente da regressão pode ser escrito como

$$VIF_j = \frac{1}{1 - R_j^2}$$

onde: R_j^2 é o coeficiente de determinação obtido para o regressor x_j sobre as outras variáveis regressoras. Claramente, se x_j é linearmente dependente dos outros regressores, então R_j^2 estará próximo da homogeneidade e o VIF_j será alto. Autores sugerem que um $VIF > 10$ implica em sérios problemas com multicolinearidade.

3.3 MATERIAL E MÉTODOS

3.3.1 Dados

Os dados deste estudo são provenientes de 140 árvores de *Eucalyptus* spp., distribuídos em sete classes diamétricas, pertencentes à empresa Aracruz Celulose S.A., localizada nos estados do Espírito Santo e Bahia, na região leste do Brasil. A caracterização da região da empresa se encontra no capítulo 2.

Foi obtido um total de 1522 observações, deste total 35% foi retirado nas diferentes classes para a validação. Esta retirada foi sistematizada nas 7 classes, de onde foram retiradas às ultimas 7 árvores, restando 13 árvores.

A cubagem absoluta foi feita segundo a fórmula de Smalian nas seguintes posições da árvores: 0; 1,4; 2,8; 4,2; 5,6 metros, e assim por diante a cada 1,4 metros, até um diâmetro mínimo de 5 cm.

3.3.2 Modelos Polinomiais

O ajuste dos modelos polinomiais foi feito por classe e também desconsiderando-se o controle das classes diamétricas.

Os três polinômios testados são:

Polinômio de quinto grau (Prodan, 1971), dado por:

$$\frac{d_{ij}}{dap_j} = \beta_0 + \beta_1 \left(\frac{h_{ij}}{ht_j} \right) + \beta_2 \left(\frac{h_{ij}}{ht_j} \right)^2 + \beta_3 \left(\frac{h_{ij}}{ht_j} \right)^3 + \beta_4 \left(\frac{h_{ij}}{ht_j} \right)^4 + \beta_5 \left(\frac{h_{ij}}{ht_j} \right)^5 + \varepsilon_{ij} \quad (3.1)$$

onde: β_{is} = parâmetros a serem estimados; d_{ij} = diâmetro correspondente a qualquer altura h_{ij} , especificada ou comercial (cm); dap_j = diâmetro a 1,3 metros

de altura (cm); ht_j = altura total da árvore (m); h_{ij} = altura comercial (m); e ε_{ij} = erro de estimativa.

- O polinômio de potências fracionárias e inteiras (Hradetzky, 1976), é dado por:

$$\frac{d_{ij}}{dap_j} = \beta_0 + \beta_1 \left(\frac{h_{ij}}{ht_j} \right)^{p_1} + \beta_2 \left(\frac{h_{ij}}{ht_j} \right)^{p_2} + \dots + \beta_n \left(\frac{h_{ij}}{ht_j} \right)^{p_n} + \varepsilon_{ij} \quad (3.2)$$

onde d_{ij} , dap_j , h_{ij} , ht_j , β_i , e ε_{ij} já foram definidos.

Os expoentes são selecionados pelo processo Stepwise do software SAS. Os expoentes testados para construir este modelo foram: 0,00001; 0,00005; 0,0009; 0,0007; 0,0006; 0,0004; 0,0002; 0,0001; 0,009; 0,008; 0,007; 0,006; 0,005; 0,004; 0,09; 0,08; 0,07; 0,06; 0,05; 0,04; 0,03; 0,02; 0,01; 0,9; 0,8; 0,7; 0,6; 0,5; 0,4; 0,3; 0,2; 0,1; 1; 2; 3; 4; 5; 10; 15; 20; 25; 30; 35; 40; 45; 50; 55; 60; 65; 70; 75; 80; 85; 90 e 95.

- O polinômio proposto por Pires & Calegario (2004) é apresentado a seguir:

$$\begin{aligned} \frac{d_{ij}}{dap_j} = & \beta_0 + \beta_1 \left[\left(\frac{h_{ij}}{ht_j} \right)^{p_1} \right] + \beta_2 \left[\left(\frac{h_{ij}}{con_i_j} \right)^{p_1} \right] + \beta_3 \left[\left(\frac{h_{ij}}{base_j} \right)^{p_1} \right] + \dots \\ & \dots + \beta_{n-2} \left[\left(\frac{h_{ij}}{ht_j} \right)^{p_{n-2}} \right] + \beta_{n-1} \left[\left(\frac{h_{ij}}{con_i_j} \right)^{p_{n-1}} \right] + \beta_n \left[\left(\frac{h_{ij}}{base_j} \right)^{p_n} \right] + \varepsilon_{ij} \end{aligned} \quad (3.3)$$

onde d_{ij} , d_{apj} , h_{ij} , ht_j , β_i , e ε_{ij} já foram definidos; sendo $base_j = \frac{d_{1,4mj}}{d_{0,0mj}}$ e

$$con_i_j = \left[\frac{\left(\frac{d_{2,8mj}}{d_{0,0mj}} + \frac{d_{5,6mj}}{d_{2,8mj}} \right)}{2} \right] \quad (3.3.1)$$

Os expoentes foram selecionados pelo processo Stepwise. Os expoentes testados foram 0,5; 1; 2; 3; 4; e 5.

3.3.3 Diagnósticos da regressão

As medidas de diagnósticos utilizadas para encontrar observações que geram influências nos polinômios foram: Pontos de influência segundo o critério de Hoaglin & Welsch (1978), "outliers", Predição da soma de quadrado do resíduo (PRESS), Coeficiente de determinação predito ($R^2_{predito}$), Distância de Cook, (D_i), $DFBETAS_{j,i}$, $DFFITS_i$, e $COVRATIO_i$.

Além das medidas de diagnóstico os condicionantes do resíduo e a multicolinearidade também foram observados.

3.3.4 Teste de Hipóteses dos Parâmetros

O objetivo principal deste teste de hipóteses é verificar se há necessidade de utilização do polinômio para cada classe diamétrica ou se podem ser agrupadas algumas classes ou todas. Este teste é feito no software SPLUS que interage cada variável, testando sua significância, segundo o teste t de student. Os valores não significativos são rejeitados pelo teste t e podem ser descartados do modelo e o contrário também é válido.

3.4 RESULTADOS E DISCUSSÃO

3.4.1- Diagnóstico dos Modelos Lineares

Na tabela 3.1 tem-se a contagem de valores de influência, de acordo com o critério de Hoaglin & Welsch (1978). Detalhes das observações de influência encontram-se anexos.

TABELA 3.1 - Contagem do número de observações que exercem influência no ajuste dos polinômios, utilizando o critério de Hoaglin e Welsch.

| Polinômio | Classes | | | | | | | |
|------------------------------|---------|---|----|----|----|----|----|----|
| | Todas | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 5ºGrau | 31 | 2 | 5 | 2 | 1 | 0 | 33 | 31 |
| Hradetzky | 20 | 0 | 5 | 13 | 26 | 26 | 32 | 31 |
| Pires & Calegario | 73 | 5 | 11 | 13 | 0 | 12 | 27 | 30 |

Na tabela 3.1 pode-se notar que os valores destacados em negrito são os menores em cada classe. Percebe-se que em três classes cada modelo foi melhor. Porém, ao se fazer uma análise mais aprofundada, verifica-se que o polinômio proposto por Pires & Calegario (2004) teve um número muito grande de observações de influência para todas as classes. O polinômio de Hradetzky apresentou uma quantidade alta de observações de influência nas classes 5 e 6. Após esta análise, pode-se dizer que o polinômio de 5ºGrau é mais estável, em termos de observações de influência (diagonal principal da matriz “hat”), que os demais polinômios.

Na tabela 3.2 é apresentada a contagem dos valores estranhos (“outliers”), para os três polinômios, em todas as classes e em cada classe individualmente.

TABELA 3.2 - Contagem do número de “outliers” dos três polinômios nas várias classes diamétricas.

| Polinômios/Classes | Polinômio 5ºGrau | Polinômio de Potências Fracionárias | Polinômio Pires & Calegario |
|---------------------|---------------------|---|--------------------------------|
| Classe diamétrica 2 | 4 | 3 | 4 |
| Classe diamétrica 3 | 3 | 3 | 3 |
| Classe diamétrica 4 | 5 | 4 | 5 |
| Classe diamétrica 5 | 10 | 9 | 9 |
| Classe diamétrica 6 | 7 | 8 | 9 |
| Classe diamétrica 7 | 8 | 9 | 8 |
| Classe diamétrica 8 | 2 | 1 | 1 |
| Todas as classes | 55 | 49 | 51 |

Verifica-se que os valores destacados em negrito são os polinômios que obtiveram o menor número de valores estranhos na classe em questão. Por exemplo, para todas as classes, o polinômio de Potências Fracionárias foi o que obteve o menor número de “outliers”. Em caso de empate, ambos foram destacados em negrito. Nota-se que o polinômio de Potências Fracionárias e Inteiras obteve, de maneira geral, um menor número de “outliers”, ficando em segundo lugar o polinômio proposto por Pires & Calegario (2004) e, na seqüência, o polinômio de 5ºGrau. A diferença entre estes polinômios é muito pequena e pode ser, dependendo do estudo, considerado insignificante.

Na tabela 3.3 é apresentado o poder preditivo dos polinômios de 5ºGrau, Hradetzky e de Pires & Calegario (2004), utilizando-se o resíduo PRESS e o R^2_{predito} como parâmetros.

TABELA 3.3 – Poder preditivo dos modelos polinomiais testados, utilizando-se como parâmetros o resíduo PRESS e o R^2_{predito} .

| Polinômio | Classe | SQRes | PRESS | $R^2_{\text{predito}}(\%)$ |
|-------------------|-----------|----------|----------|----------------------------|
| 5ºGrau | Todas | 2,260089 | 2,29346 | 97,716 |
| | 2 | 0,163102 | 0,211067 | 98,220 |
| | 3 | 0,266128 | 0,305921 | 97,774 |
| | 4 | 0,111689 | 0,127095 | 98,998 |
| | 5 | 0,100145 | 0,112011 | 99,140 |
| | 6 | 0,084589 | 0,091598 | 99,364 |
| | 7 | 0,222216 | 0,238218 | 98,517 |
| | 8 | 0,276246 | 0,288911 | 98,317 |
| | Hradetzky | Todas | 2,061323 | 2,098969 |
| 2 | | 0,162219 | 0,181209 | 98,472 |
| 3 | | 0,258257 | 0,290267 | 97,888 |
| 4 | | 0,106683 | 0,11796 | 99,070 |
| 5 | | 0,09996 | 0,109697 | 99,158 |
| 6 | | 0,082895 | 0,08731 | 99,394 |
| 7 | | 0,173423 | 6,073304 | 62,200 |
| 8 | | 0,267332 | 0,278721 | 98,376 |
| Pires & Calegario | | Todas | 1,955074 | 1,994166 |
| | 2 | 0,150772 | 0,168004 | 98,583 |
| | 3 | 0,213489 | 0,240261 | 98,252 |
| | 4 | 0,113336 | 0,121278 | 99,044 |
| | 5 | 0,098979 | 0,109735 | 99,158 |
| | 6 | 0,066324 | 0,072517 | 99,496 |
| | 7 | 0,240313 | 0,252389 | 98,429 |
| | 8 | 0,270189 | 0,282738 | 98,353 |

Quando os valores de SQRes e do PRESS estão próximos, há poucos pontos de influência. Pode se notar que todos os polinômios, em todas as classes, apresentam alto poder preditivo, exceto o polinômio de Potências Fracionárias e Inteiras (Hradetzky, 1976) que apresentou baixa predição na classe 7. Os valores de R^2_{predito} podem ser interpretados da seguinte forma: espera-se que o polinômio, por exemplo, de Pires & Calegario (2004) explique 98% da variabilidade na predição de novas observações, quando comparado com, aproximadamente, 98% da variabilidade nos dados originais, explicado

pelo ajuste do quadrado médio. Esta capacidade preditiva parece satisfatória, a menos que os pontos a serem preditos se assemelhem a um ponto considerado de influência.

Analisando seletivamente a tabela 3.3, faz-se as seguintes escolhas: O polinômio proposto por Pires & Calegario (2004) em todas as classes; o polinômio de Hradetzky nas classes 4 e 8; e o polinômio de 5ºGrau na classe 7. Na classe 5, tanto faz o polinômio de Hradetzky ou o de Pires & Calegario (2004).

A diferença entre os valores de R^2_{predito} é muito pequena e pode ser considerada insignificante na escolha do melhor polinômio, lembrando-se sempre do problema na classe 7 com o polinômio de Hradetzky.

As Figuras 3.1, 3.2 e 3.3 apresentam os valores da distância de Cook para os polinômios de 5ºGrau, de Hradetzky e o de Pires & Calegario (2004), para todas as classes diamétricas.

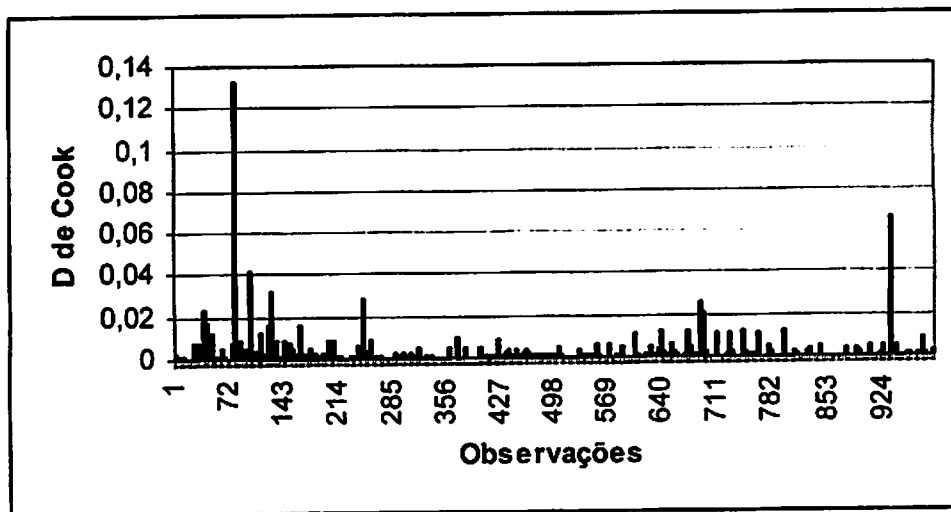


Figura 3.1- Valor da distância de Cook no polinômio de 5ºGrau, para todas as classes diamétricas.

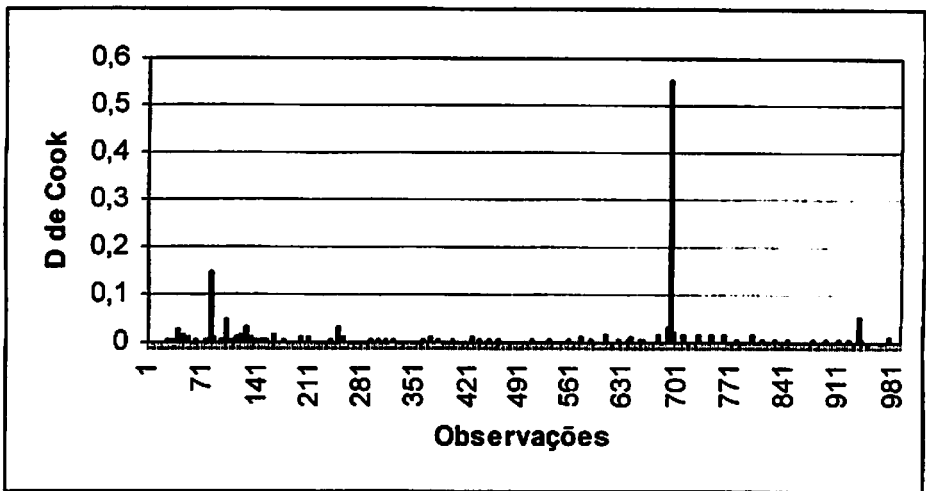


Figura 3.2- Valor da distância de Cook no polinômio de Potências Fracionárias e Inteiras, para todas as classes diamétricas.

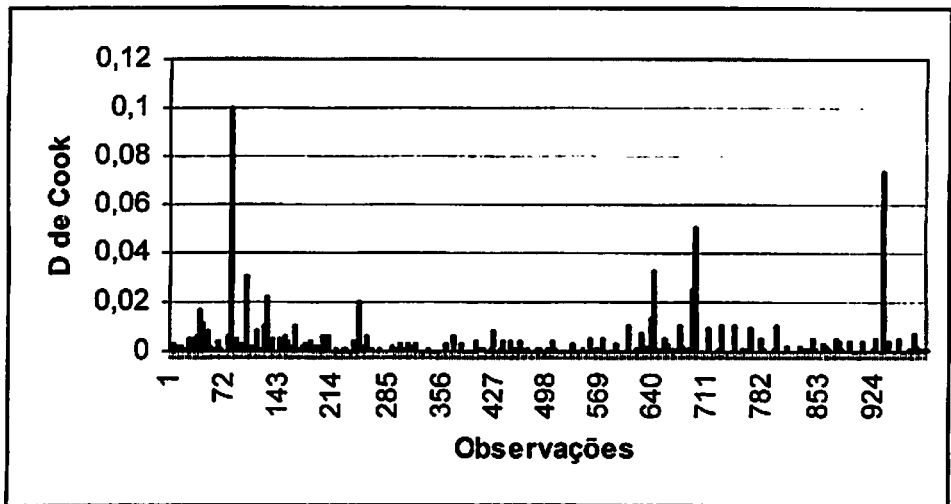


Figura 3.3- Valor da distância de Cook no polinômio proposto por Pires & Calegario (2004), para todas as classes diamétricas.

Nas Figuras 3.1, 3.2 e 3.3 observa-se que quanto maior o valor de distância de Cook (chamado D de Cook) maior será a influência da observação nos coeficientes da regressão. Pode-se notar nas figuras que a observação 693

apresentou um valor muito alto de distância de Cook, figura 3.2, referente ao polinômio de Hradetzky. Os três maiores valores de distância de Cook, para o polinômio de 5ºGrau, estão nas observações 79, 99 e 937. Para o polinômio de potências fracionárias e inteiras e para o polinômio proposto por Pires & Calegario (2004) as observações 79, 693 e 937 se destacam pelo alto valor de D_i . Para os três polinômios, as observações 79 e 937 merecem atenção.

As análises do $DFFITS_i$ e do $COVRATIO_i$ foram feitas para todos os três modelos polinomiais, sem distinção das classes diamétricas. A contagem das observações que influenciam os valores ajustados e a precisão da estimativa são apresentados na tabela 3.4.

TABELA 3.4 – Contagem do número de observações que influenciam nos valores ajustados e na precisão da estimativa, para os três polinômios testados.

| Polinômio | $DFFITS_i$ | $COVRATIO_i$ | Total de observações |
|---------------------------|------------------------------|--------------------------------|-----------------------------|
| 5ºGrau | 807 | 54 | 992 |
| Potências Fracionárias | 810 | 176 | 992 |
| Pires & Calegario | 926 | 94 | 992 |
| Total | 2543 | 324 | 2976 |

Pode ser notada a presença de um número elevado de observações, que influenciam os valores ajustados e um número muito menor de observações, que influenciam na precisão da estimativa, isto se deve ao fato do $DFFITS_i$ ser afetado pela influência e pelo erro de predição. O polinômio de 5ºGrau obteve um número menor de observações de influencia nos valores ajustados e na

precisão da estimativa. O polinômio de Potências Fracionárias e Inteiras de Hradetzky apresentou maior número de observações de influência, na precisão da estimativa e o polinômio Pires & Calegario (2004) apresentou maior número de observações de influência nos valores ajustados.

O calculo e análise do $DFBETAS_{j,i}$ só será realizado para o modelo selecionado para validação, não tendo nenhum peso na seleção do polinômio.

3.4.2- Condicionantes do Resíduo

A acuracidade dos modelos polinomiais testados neste capítulo fora efetuada no capítulo 2, e o seus gráficos de resíduo também. Porém, neste capítulo é abordada a necessidade de verificação das condicionantes do resíduo. Nas tabelas 3.5, 3.6 e 3.7 é apresentada a avaliação da homocedasticidade dos polinômios de 5ºGrau, Hradetzky e Pires & Calegario (2004), respectivamente, e na figura 3.1 um resumo destas tabelas.

TABELA 3.5 - Avaliação da homocedasticidade do Polinômio de 5ºGrau nas diferentes classes diamétricas.

| Classe | β_1 | t tabela | valor t | Pr > t |
|--------|-----------|----------|---------|---------------|
| Todas | 0,00483 | 2,5808 | 5,36 | <,0001 |
| 2 | 0,00288 | 2,642082 | 3,39 | 0,0011 |
| 3 | 0,00572 | 2,634206 | 2,49 | 0,0147 |
| 4 | 0,00215 | 2,624893 | 3,86 | 0,0002 |
| 5 | 0,00205 | 2,614779 | 5,85 | <,0001 |
| 6 | 0,000431 | 2,60814 | 2,28 | 0,024 |
| 7 | -0,00195 | 2,598354 | -2,85 | 0,0048 |
| 8 | -0,00609 | 2,599227 | -2,01 | 0,0458 |

Onde: t tabela = valor de t student tabelado para 1% de probabilidade; valor t = valor do t student calculado; β_1 = parâmetro estimado que deve ser estatisticamente igual a zero; e Pr > |t| = probabilidade encontrada, em módulo, para o valor de β_1 .

TABELA 3.6- Avaliação da homocedasticidade do Polinômio de Potências Fracionárias e Inteiras nas diferentes classes diamétricas.

| Classe | β_1 | t tabela | valor t | Pr > t |
|--------|-----------|----------|---------|---------------|
| Todas | 0,00443 | 2,5808 | 5,02 | <,0001 |
| 2 | 0,00286 | 2,642082 | 3,45 | 0,0009 |
| 3 | 0,0057 | 2,634206 | 2,54 | 0,0129 |
| 4 | 0,00207 | 2,624893 | 3,77 | 0,0003 |
| 5 | 0,00206 | 2,614779 | 5,88 | <,0001 |
| 6 | 0,000325 | 2,60814 | 1,76 | 0,0806 |
| 7 | -0,001 | 2,598354 | -2,07 | 0,0393 |
| 8 | -0,00635 | 2,599227 | -2,06 | 0,0403 |

Onde: t tabela, valor t, β_1 e Pr > |t| = já definidos anteriormente.

TABELA 3.7- Avaliação da homocedasticidade do Polinômio proposto por Pires & Calegario (2004) nas diferentes classes diamétricas.

| Classe | β_1 | t tabela | valor t | Pr > t |
|--------|-----------|----------|---------|---------------|
| Todas | 0,0038 | 2,5808 | 4,19 | <,0001 |
| 2 | 0,00302 | 2,642082 | 3,83 | 0,0003 |
| 3 | 0,0059 | 2,634206 | 2,84 | 0,0057 |
| 4 | 0,00215 | 2,624893 | 3,88 | 0,0002 |
| 5 | 0,00202 | 2,614779 | 5,76 | <,0001 |
| 6 | 0,000406 | 2,60814 | 2,16 | 0,0323 |
| 7 | -0,00246 | 2,598354 | -3,09 | 0,0023 |
| 8 | -0,0064 | 2,599227 | -2,1 | 0,0367 |

Onde: t tabela, valor t, β_1 e Pr > |t| = já definidos anteriormente.

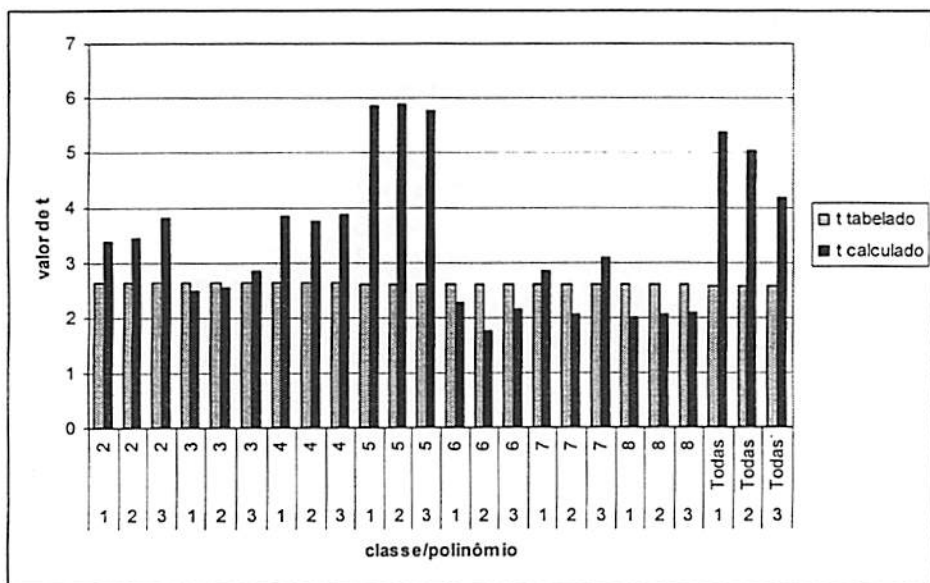


Figura 3.4- Comparação dos valores de t tabelado e calculado, na verificação da homocedasticidade, para os diferentes polinômios nas diferentes classes diamétricas.

Na Figura (3.4), no eixo x, tem-se, de baixo para cima, os polinômios 1,2 e 3, sendo: 1 = o polinômio de 5ºGrau, 2 = o polinômio de Potências Fracionárias e inteiras e 3 = o polinômio Pires & Calegario (2004). Acima observa-se as classes, começando de 2 até todas as classes, no sentido da esquerda para a direita. No eixo y, tem-se os valores de t student. Espera-se que os valores calculados devem ser maiores que os tabelados.

Nota-se uma mudança de sinal nos β_1 dos três polinômios testados, este fato é um dos indicativos da presença de multicolinearidade, suas desvantagens foram discutidas no capítulo 2.

Usando-se os conceitos de Anscombe (1973), nas tabelas 3.5, 3.6 e 3.7 observa-se que as classes e valores destacados em negrito não apresentam homogeneidade de variância, a nível de 1% de probabilidade, no teste t de student. A homogeneidade é aceitável a nível de 5% de probabilidade para todos

os polinômios e em todas as classes diamétricas, exceto o polinômio de Hradetzky na classe 6 que só é aceita a 10%.

Em uma avaliação quantitativa da homogeneidade dos resíduos, para os diferentes polinômios, pode se destacar o polinômio proposto por Pires & Calegario (2004) em todas as classes, exceto nas classes 6 e 8, o polinômio de 5ºGrau em todas as classes, exceto nas classes 3 e 8, e o polinômio de Hradetzky em todas as classes, exceto nas classes 3, 6, 7 e 8. Se considerar esta como uma avaliação primordial, deve-se escolher, para aplicar na descrição do perfil longitudinal de árvores, o polinômio de 5ºGrau ou o polinômio Pires & Calegario.

O resíduo padronizado e o estudentizado são úteis na detecção do afastamento da normalidade. Se os erros são normalmente distribuídos, então, aproximadamente, 68% dos resíduos padronizados deveriam estar entre -1 e +1, e, aproximadamente, 95% deles deveriam estar entre -2 e +2. O desvio substancial para estes limites indicam uma potencial violação da hipótese de normalidade. A tabela 3.8 apresenta o teste de normalidade para os polinômios nas diversas classes diamétricas.

TABELA 3.8- Teste de normalidade para os polinômios de 5ºGrau, Hradetzky e Pires & Calegario, nas diversas classes diamétricas, considerando o resíduo Padronizado e o Estudentizado com a probabilidade de 95%.

| Polinômio | Classe | (%) R.Pad. | (%) R.Estud. |
|-----------|--------|------------|--------------|
| 5ºGrau | Todas | 94,76 | 94,76 |
| | 2 | 94,87 | 94,87 |
| | 3 | 96,59 | 96,59 |
| | 4 | 95,19 | 95,19 |
| | 5 | 93,85 | 91,54 |
| | 6 | 96,15 | 95,51 |
| | 7 | 96,40 | 96,40 |
| | 8 | 99,53 | 99,07 |

Continua...

TABELA 3.11 - VIF para os parâmetros do polinômio proposto por Pires & Calegario nas diferentes classes diamétricas.

| Classe diamétrica | $\hat{\beta}_1$ | $\hat{\beta}_2$ | $\hat{\beta}_3$ | $\hat{\beta}_4$ | $\hat{\beta}_5$ | $\hat{\beta}_6$ | $\hat{\beta}_7$ | $\hat{\beta}_8$ |
|-------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| 2 | 29,313 | 80,979 | 46,902 | 13,858 | | | | |
| 3 | 52,869 | 272,69 | 102,58 | 207,23 | 37,063 | | | |
| 4 | 1,6315 | 1,6315 | | | | | | |
| 5 | 78,383 | 466,32 | 384,85 | 47,913 | | | | |
| 6 | 2969,6 | 11894 | 6074,7 | 121,01 | 1014,4 | 644,49 | 66,674 | |
| 7 | 6,4667 | 21,342 | 9,4719 | | | | | |
| 8 | 28,259 | 8192,3 | 87973 | 13083 | 22899 | | | |
| Todas | 154,21 | 812,71 | 194,82 | 3798,1 | 178,33 | 12131 | 89,698 | 5484,1 |

Pode ser notado nas tabelas acima que os VIF para todos os polinômios foi acima do aceitável ($VIF < 10$) com poucas exceções. Porém, a análise do fator de inflação da variância deve ser observada, de um modo geral, nas classes diamétricas, pois é desta forma que os polinômios serão utilizados. Estes valores para polinômios já eram esperados, pois os mesmos possuem alta correlação linear entre seus parâmetros. Dentre os três polinômios analisados, o que possui menores valores de VIF foi o de Hradetzky (Potências Fracionárias e Inteiras). Mesmo assim, o mesmo apresentou sérios problemas de multicolinearidade nas classes 7, 8 e em todas as classes, menores VIF foram verificados na classe 3. O polinômio proposto por Pires & Calegario (2004) ficou com valores de VIF menores que os valores obtidos para o polinômio de 5º Grau. Outra observação que pode ser feita é sobre a relação direta entre o número de parâmetros selecionados pelo modelo e o valor de VIF. Quanto menor o número de parâmetros menor o valor de VIF.

3.4.4- Teste de Hipótese dos Parâmetros

A seguir, tem-se o teste de hipótese dos coeficientes da regressão visando à verificação da necessidade do controle por classe diamétrica.

TABELA 3.12 – Coeficientes da regressão selecionados pelo teste t, considerando como variável dependente d, para o polinômio de 5º Grau, testando os coeficientes das variáveis selecionadas, o seu valor, o erro padrão, o valor de t calculado, e o valor da probabilidade encontrado para o t calculado, para verificar a necessidade do controle por classe diamétrica.

| | Valor | Erro Padrão | valor t | Pr(> t) |
|----------------------|-----------|-------------|---------|----------|
| Intercepto | 2,1948 | 0,5524 | 3,9732 | 0,0001 |
| dap | 0,9327 | 0,1134 | 8,2238 | 0,0000 |
| Classe | 0,4350 | 0,1841 | 2,3629 | 0,0183 |
| h/H | -6,7986 | 8,7826 | -0,7741 | 0,4391 |
| ((h/H)^2) | 16,9187 | 56,6882 | 0,2985 | 0,7654 |
| ((h/H)^3) | 11,4980 | 157,7287 | 0,0729 | 0,9419 |
| ((h/H)^4) | -75,7733 | 189,0216 | -0,4009 | 0,6886 |
| ((h/H)^5) | 51,4674 | 80,0371 | 0,6430 | 0,5203 |
| dap*Classe | -0,0155 | 0,0067 | -2,3095 | 0,0211 |
| dap*(h/H) | -4,2360 | 2,3796 | -1,7801 | 0,0754 |
| dap*((h/H)^2) | 28,9768 | 16,8135 | 1,7234 | 0,0851 |
| dap*((h/H)^3) | -89,8635 | 46,9184 | -1,9153 | 0,0557 |
| dap*((h/H)^4) | 118,3975 | 55,2071 | 2,1446 | 0,0322 |
| dap*((h/H)^5) | -54,0752 | 22,8712 | -2,3643 | 0,0183 |
| Classe*(h/H) | 0,3300 | 4,6171 | 0,0715 | 0,9430 |
| Classe*((h/H)^2) | -35,9759 | 34,0340 | -1,0571 | 0,2907 |
| Classe*((h/H)^3) | 141,7724 | 95,9057 | 1,4782 | 0,1397 |
| Classe*((h/H)^4) | -200,2127 | 113,4531 | -1,7647 | 0,0779 |
| Classe*((h/H)^5) | 93,4452 | 47,2413 | 1,9780 | 0,0482 |
| Dap*Classe*(h/H) | 0,2364 | 0,0681 | 3,4731 | 0,0005 |
| dap*Classe*((h/H)^2) | -0,4491 | 0,1774 | -2,5321 | 0,0115 |
| dap*Classe*((h/H)^3) | 0,2265 | 0,1185 | 1,9119 | 0,0562 |

Onde: dap = diâmetro a 1,30 metros; Classe = controle das classes diamétricas 2, 3, 4, 5, 6, 7, 8; h = altura das várias seções de d; H = altura total.

Na Tabela 3.12 algumas variáveis foram consideradas não significativas, porém, interagindo as três variáveis, (dap, classe e (h/H)), tem-se um resultado significativo. Portanto estas não podem ser descartadas. O modelo descrito acima tem como vantagem poder ser usado para todas as classes diamétricas. Outros detalhes deste modelo são as suas medidas de acuracidade: $s_{yx} = 0,4685$,

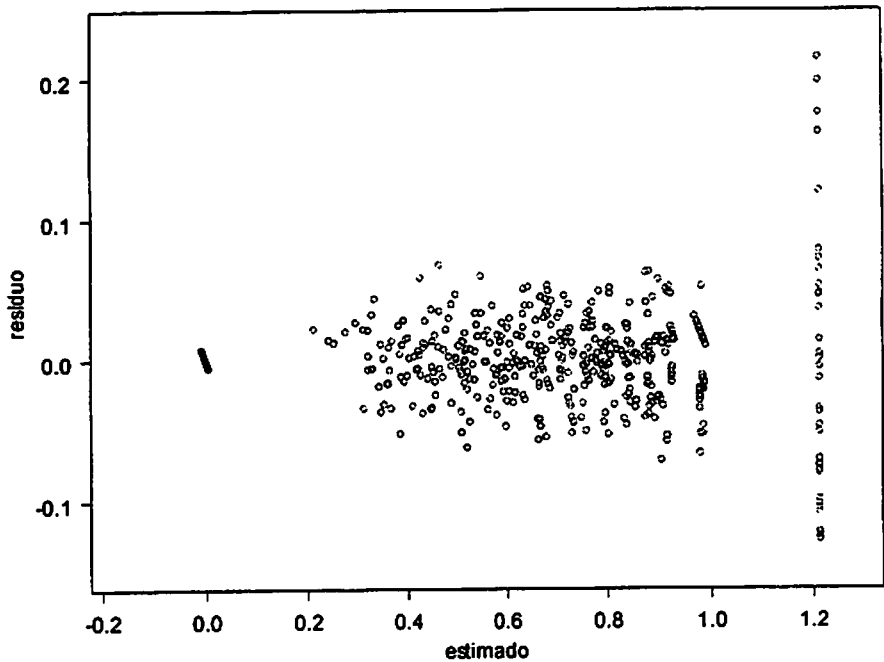


Figura 3.5 - Resíduo versus o valor estimado, para o polinômio proposto por Pires & Calegario (2004), com todos os dados ajustados, na validação.

Pode ser notado, na Figura 3.5, que muitos dos valores discrepantes desapareceram, apresentando uma melhor distribuição dos resíduos. Também o pequeno número de dados utilizado (35% do total) pode ter influenciado nesta distribuição dos resíduos.

Na tabela 3.14 e na figura 3.6 é apresentada a avaliação da homocedasticidade deste polinômio na validação.

Os novos parâmetros estimados as potências e variáveis selecionadas encontram-se anexos. Na tabela 3.13 é apresentada as novas medidas de acuracidade.

TABELA 3.13 – Medidas de Acuracidade do polinômio proposto por Pires & Calegario (2004) para todas as classes diamétricas e para cada classe, na validação dos dados.

| Classe diamétrica | F | SQErro | R ² (%) | R ² _{ajustado} (%) | Syx |
|----------------------|----------------|----------------|--------------------|---|----------------|
| 2 | 2876,31 | 0,04303 | 99,29 | 99,26 | 0,03240 |
| 3 | 2606,51 | 0,03957 | 99,47 | 99,43 | 0,03069 |
| 4 | 2344,57 | 0,07574 | 98,88 | 98,84 | 0,03780 |
| 5 | 3018,49 | 0,03289 | 99,56 | 99,53 | 0,02232 |
| 6 | 4243,83 | 0,04821 | 99,38 | 99,35 | 0,02455 |
| 7 | 10622,4 | 0,04691 | 99,46 | 99,45 | 0,02020 |
| 8 | 4423,73 | 0,05204 | 99,40 | 99,38 | 0,02205 |
| Todas | 5663,85 | 0,70023 | 98,70 | 98,68 | 0,03656 |

Na validação dos dados, o polinômio Pires & Calegario (2004) foi melhor que anteriormente em todas as classes, exceto nas classes 4 e 6. Este resultado era esperado, pois na validação trabalhou-se com um volume de dados menor que no ajuste. Na Figura 3.5 apresenta-se o resíduo, em função do valor estimado, para todas as classes diamétricas.

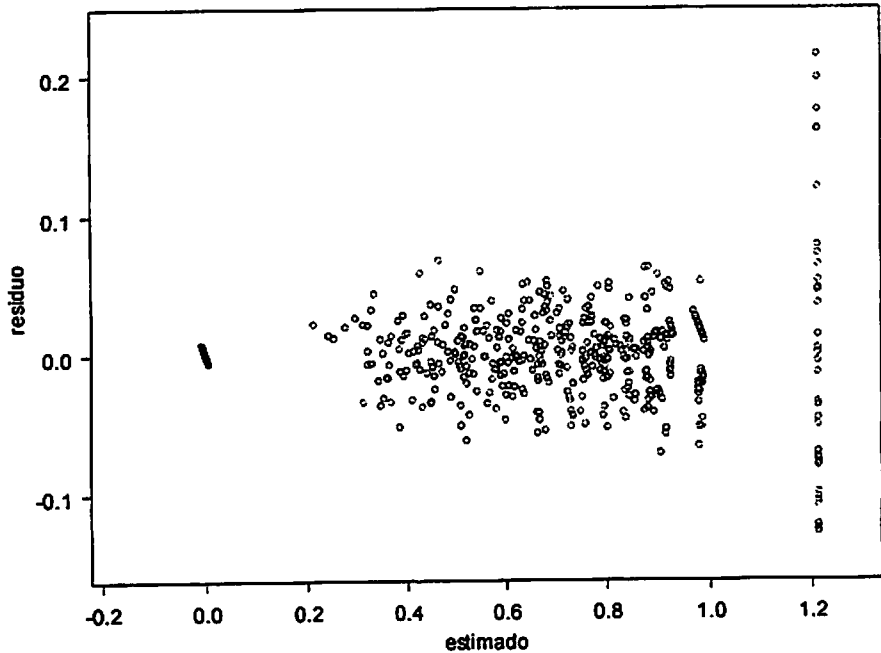


Figura 3.5 - Resíduo versus o valor estimado, para o polinômio proposto por Pires & Calegario (2004), com todos os dados ajustados, na validação.

Pode ser notado, na Figura 3.5, que muitos dos valores discrepantes desapareceram, apresentando uma melhor distribuição dos resíduos. Também o pequeno número de dados utilizado (35% do total) pode ter influenciado nesta distribuição dos resíduos.

Na tabela 3.14 e na figura 3.6 é apresentada a avaliação da homocedasticidade deste polinômio na validação.

TABELA 3.14- Avaliação da homocedasticidade do Polinômio Pires & Calegario (2004) nas diferentes classes diamétricas, na validação dos dados.

| Classe | β_1 | t tabela | | Pr > t |
|--------|-----------|--------------|---------|---------|
| | | (0,01;glres) | valor t | |
| Todas | 0,00426 | 2,585148 | 8,29 | <,0001 |
| 2 | 0,00238 | 2,698071 | 2,84 | 0,007 |
| 3 | 0,0016 | 2,692286 | 3,49 | 0,0011 |
| 4 | 0,0032 | 2,669985 | 2,39 | 0,0203 |
| 5 | 0,000526 | 2,647903 | 2,47 | 0,0161 |
| 6 | 0,000849 | 2,637134 | 3,48 | 0,0008 |
| 7 | 0,000338 | 2,618872 | 1,69 | 0,0937 |
| 8 | 0,000952 | 2,621273 | 1,65 | 0,1026 |

Onde: t tabela, valor t, β_1 e Pr > |t| = já definidos anteriormente.

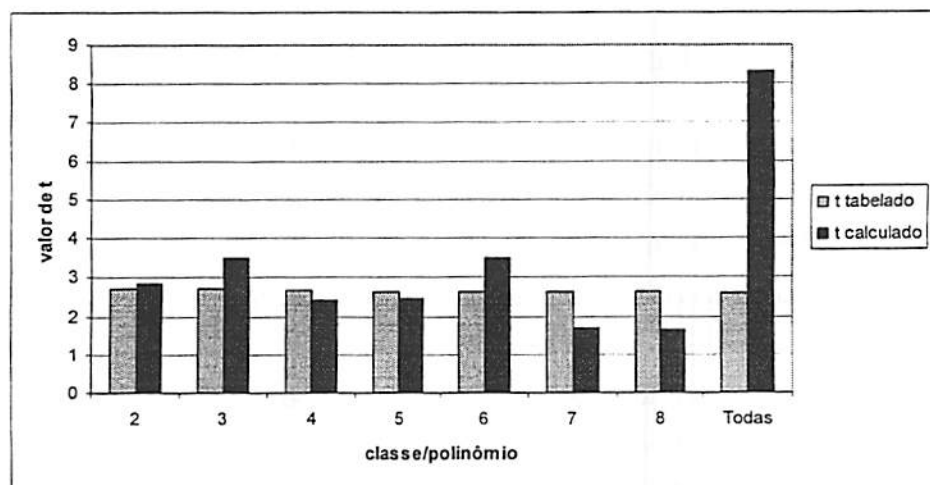


Figura 3.6- Comparação dos valores de t tabelado e o calculado, na verificação da homocedasticidade, para o polinômios Pires & Calegario (2004), nas diferentes classes diamétricas.

A descrição da Figura 3.6 já foi definida anteriormente.

Pode-se verificar que a homocedasticidade foi pior nas classes 4, 5, 7 e 8, e nas demais foi aceitável, sendo que na classe 4 foi aceitável a 5% de probabilidade, na classe 7 somente a 10% de probabilidade, e na classe 8, a mais crítica, acima de 10%.

Na tabela 3.15 é apresentado o teste de normalidade para o polinômio proposto por Pires & Calegario (2004) na validação dos dados.

TABELA 3.15- Teste de normalidade para o polinômio Pires & Calegario (2004), nas diversas classes diamétricas, considerando o residuo Padronizado e o Estudentizado com a probabilidade de 95%, na validação dos dados.

| Polinômio | Classe | (%) R.Pad. | (%) R.Estud. |
|-------------------|--------|------------|--------------|
| Pires & Calegario | Todas | 95,68 | 95,68 |
| | 2 | 95,45 | 95,45 |
| | 3 | 93,48 | 91,30 |
| | 4 | 96,43 | 96,43 |
| | 5 | 97,22 | 95,83 |
| | 6 | 94,05 | 92,86 |
| | 7 | 96,61 | 96,61 |
| | 8 | 97,32 | 97,32 |

Onde: R. Pad.; R.Estud. = já definidos anteriormente.

Pode se verificar na validação que no teste de normalidade, tabela 3.14, o polinômio Pires & Calegario (2004) foi melhor na classe 5 e considerando todas as classes diamétricas. Para a classe 3 foi pior, e na classe 6 foi semelhante ao ajuste inicial. Nas demais classes foi considerado aceitável.

Na tabela 3.16 ilustra-se o fator de inflação da variância obtido para cada parâmetro e em cada classe diamétrica, para o polinômio proposto por Pires & Calegario (2004) na validação.

TABELA 3.16 - VIF para os parâmetros do polinômio proposto por Pires & Calegario (2004) nas diferentes classes diamétricas, na validação.

| Classe diamétrica | $\hat{\beta}_0$ | $\hat{\beta}_1$ | $\hat{\beta}_2$ | $\hat{\beta}_3$ | $\hat{\beta}_4$ | $\hat{\beta}_5$ | $\hat{\beta}_6$ | $\hat{\beta}_7$ |
|-------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| 2 | 0 | 1,9063 | 1,9063 | | | | | |
| 3 | 0 | 40,033 | 152,06 | 49,388 | | | | |
| 4 | 0 | 1,6398 | 1,6398 | | | | | |

Continua...

Continuação...

| | | | | | | | | |
|--------------|----------|---------------|---------------|---------------|--------------|--------------|---------------|---------------|
| 5 | 0 | 2141,1 | 1729,7 | 273,66 | 206,96 | 24,986 | | |
| 6 | 0 | 18,981 | 28,357 | 3,8756 | | | | |
| 7 | 0 | 2,3677 | 2,3677 | | | | | |
| 8 | 0 | 876,09 | 888,02 | 19,720 | 8,6702 | | | |
| Todas | 0 | 207,65 | 1410,0 | 221,59 | 11656 | 10806 | 17,850 | 936,49 |

O VIF foi melhor, menor valor, nas classes 2, 3, 6, 7 e 8. Foi igual na classe 4 e pior na classe 5 e em todas as classes. Valores de VIF superiores a 10 são considerados como multicolinearidade. Pode ser observado que há uma relação direta entre o valor de VIF e o número de variáveis selecionadas. Somente nas classes 2, 4 e 7 este valor foi inferior a 10, ou seja, aceitável.

Encontra-se na tabela 3.17 a contagem das observações que necessitam de atenção utilizando os seguintes critérios: “outliers”; DFFITS_i; COVRATIO_i. todas estas para o polinômio proposto por Pires & Calegario (2004) na validação dos dados. Ao lado destes critérios encontram-se o número total de observações na classe em questão, a percentagem de observações que são valores estranhos (“outliers”), a percentagem de observações de influência nos valores ajustados, e a percentagem de observações de influência na precisão da estimativa.

TABELA 3.17 – Contagem das observações que necessitam de atenção e sua percentagem, utilizando alguns critérios estatísticos, para o polinômio Pires & Calegario (2004), na validação dos dados.

| Classes | Estatística | | | | Inf.1 (%) | Inf.2 (%) | Inf.3 (%) |
|---------|-------------|--------|----------|-----|-----------|-----------|-----------|
| | outliers | DFFITS | COVRATIO | obs | | | |
| 2 | 2 | 42 | 34 | 44 | 4,55 | 95,45 | 77,27 |
| 3 | 3 | 44 | 26 | 46 | 6,52 | 95,65 | 56,52 |

Continua...

Continuação...

| | | | | | | | |
|--------------|-----------|------------|-----------|------------|-------------|--------------|--------------|
| 4 | 2 | 48 | 40 | 56 | 3,57 | 85,71 | 71,43 |
| 5 | 2 | 64 | 13 | 72 | 2,78 | 88,89 | 18,06 |
| 6 | 6 | 78 | 36 | 84 | 7,14 | 92,86 | 42,86 |
| 7 | 4 | 106 | 95 | 118 | 3,39 | 89,83 | 80,51 |
| 8 | 3 | 102 | 30 | 112 | 2,68 | 91,07 | 26,79 |
| Todas | 24 | 469 | 54 | 532 | 4,51 | 88,16 | 10,15 |

Onde: obs = número total de observações na classe; Inf.1 = corresponde a percentagem de dados que apresenta “outliers”; Inf.2 = corresponde a percentagem de dados que influência nos valores estimados (DFFITSi); Inf.3 = corresponde a percentagem de dados que apresenta influência na precisão da estimativa (COVRATIOi).

De maneira geral, conforme tabela 3.17, o número de valores estranhos (“outliers”) foi menor comparado com observações de influência nos valores ajustados e na precisão da estimativa. Pode-se notar que há um grande número de valores que influenciam nos valores ajustados (uma percentagem muito alta). O COVRATIOi foi muito alto em quase todas as classes, exceto na classe 5 e 8, mas foi, relativamente, baixo considerando todas as classes. O número de “outliers” foi menor que as observações de influência, em todas as classes consideradas. Comparativamente ao ajuste antes da validação, em percentagem, para todas as classes diamétricas, o número de “outliers” (5,55% no ajuste) e de DFFITSi (93,35% no ajuste) foram menores e o número de COVRATIOi (9,48% no ajuste) foi maior. Pode-se considerar que, em termos percentuais, o número de observações pode não ter influenciado neste resultado.

Ao observar o poder preditivo do polinômio proposto por Pires & Calegario (2004), na validação, tabela 3.18, verifica-se que, o resíduo PRESS foi menor em todas as classes na validação dos dados. Já o R^2_{predito} foi maior nas classes 2, 3, 5 e, desconsiderando o controle por classe, igual nas classes 7 e 8, e menor na classe 4. O número de observações na validação pode ter influenciado neste resultado.

TABELA 3.18 – Poder preditivo do polinômio proposto por Pires & Calegario, na validação dos dados.

| Polinômio | Classe | SQRes | PRESS | R ² _{predito} (%) |
|--|--------|---------|---------|---------------------------------------|
| Proposto por Pires & Calegario (2004) | Todas | 0,70023 | 0,72364 | 98,65 |
| | 2 | 0,04303 | 0,05156 | 99,15 |
| | 3 | 0,03957 | 0,04876 | 99,34 |
| | 4 | 0,07574 | 0,08655 | 98,72 |
| | 5 | 0,03289 | 0,03863 | 99,49 |
| | 6 | 0,04821 | 0,05295 | 99,31 |
| | 7 | 0,04691 | 0,04973 | 99,43 |
| | 8 | 0,05204 | 0,06071 | 99,30 |

Os diagnósticos distância de Cook e DFBETAS_{j,i} não foram utilizados na validação do polinômio Pires & Calegario (2004). Pode se atribuir as diferenças encontradas, a grande maioria, ao volume de dados menor que no ajuste do modelo.

Após a escolha do polinômio proposto por Pires & Calegario (2004), é apresentado, na tabela 3.19, um resumo das observações que influenciam nos parâmetros estimados, utilizando o critério DFBETAS_{j,i}, e com maiores detalhes em anexo.

TABELA 3.19 – Observações de influência nos parâmetros ajustados, DFBETAS_{j,i}, para o polinômio proposto por Pires & Calegario (2004), desconsiderando o controle por classe diamétrica.

| Parâmetros | DFBETAS _{j,i} | | | | | | | | |
|--------------|------------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| | β_0 | β_1 | β_2 | β_3 | β_4 | β_5 | β_6 | β_7 | β_8 |
| n° obs. Inf. | 33 | 31 | 20 | 32 | 6 | 16 | 18 | 11 | 9 |

Onde: n° obs. Inf. = número de observações de influência nos parâmetros ajustados.

Observa-se, na tabela 3.19, que há um número maior de observações de influência nos parâmetros β_0 , β_1 e β_3 , e um número menor nos parâmetros β_4

e β_8 . Maiores detalhes das observações que influenciam em cada parâmetro estão anexas.

3.4.6- Retirada das Observações de Influência

Após a escolha do polinômio proposto por Pires & Calegario (2004), foram retiradas 33 observações coincidentes nos critérios de distância de Cook, $DFBETAS_{j,i}$, matriz “hat” (critério de Hoaglin e Welsh), valores estranhos “outliers”, $DFFITs_i$ e $COVRATIO_i$.

Na tabela 3.20 são apresentados os novos valores estimados, as potências selecionadas e suas respectivas variáveis, o valor de F, o VIF, $R^2_{ajustado}$ e o erro padrão residual (syx), para o polinômio proposto por Pires & Calegario, após a retirada das observações de influência, e desconsiderando o controle de classes diamétricas.

TABELA 3.20 – Análises do polinômio proposto por Pires & Calegario (2004), após a retirada das observações de influência, desconsiderando o controle por classes diamétrica.

| Par. | VE | Pot. | Var. | F | VIF | $R^2_{ajustado}$ (%) | syx |
|-----------------|----------|------|------|---------|----------|-------------------------|---------|
| $\hat{\beta}_0$ | 1,19693 | - | - | 9582,76 | 0 | 98,77 | 0,03401 |
| $\hat{\beta}_1$ | -0,22763 | 0,5 | xc | - | 187,3346 | - | - |
| $\hat{\beta}_2$ | 0,42203 | 1 | xd | - | 1264,270 | - | - |
| $\hat{\beta}_3$ | 0,03776 | 1 | xb | - | 243,2914 | - | - |
| $\hat{\beta}_4$ | -2,63453 | 2 | xd | - | 11127 | - | - |
| $\hat{\beta}_5$ | 2,24919 | 3 | xd | - | 11112 | - | - |
| $\hat{\beta}_6$ | -2,12E-5 | 3 | xb | - | 204,9855 | - | - |
| $\hat{\beta}_7$ | 3,945E-7 | 4 | xc | - | 96,70430 | - | - |
| $\hat{\beta}_8$ | -0,85876 | 5 | xd | - | 1056,183 | - | - |

Onde: Par. = parâmetros; VE = valor estimado; Pot. = potências selecionadas pelo processo stepwise; Var. = variáveis selecionadas pelo processo stepwise.

Com a retirada das observações de influência, pode-se notar, na tabela 3.20, que houve melhoria nas medidas de acuracidade, $R^2_{ajustado}$, syx e teste F. Porém, o VIF foi maior para alguns parâmetros, ou seja, excede o valor recomendado, o que indica a presença de multicolinearidade.

No gráfico de resíduo, Figura 3.7, observa-se melhora em todos as partes do resíduo, porém na Figura 3.8, gráfico do resíduo padronizado para detecção de "outliers", percebe-se que continua havendo presença dos mesmos. Esta presença de "outliers" era esperada, pois nem todos os valores estranhos foram removidos para o ajuste do modelo.

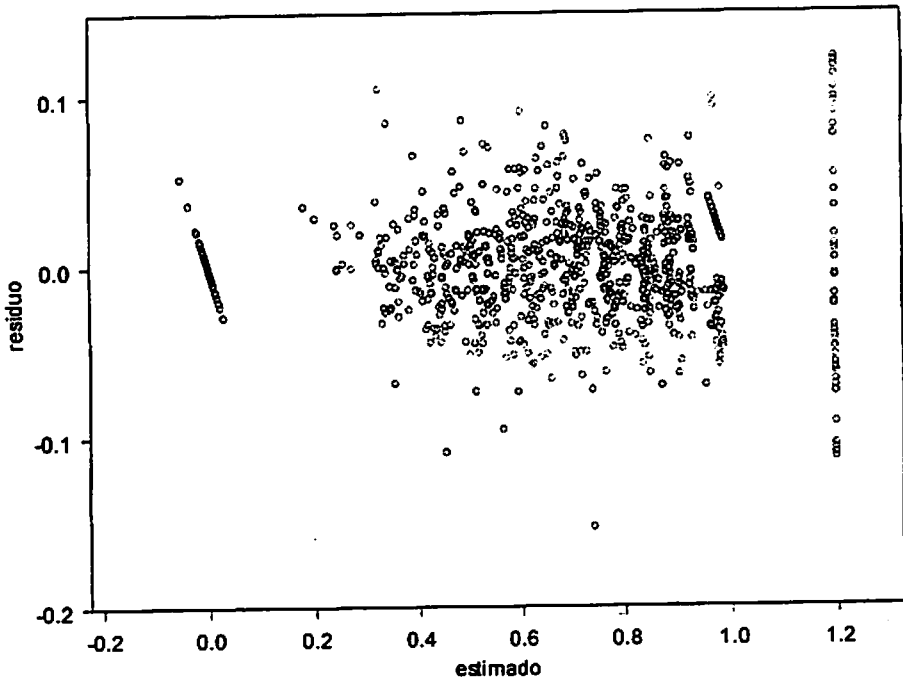


Figura 3.7- Gráfico de Resíduo, para o polinômio proposto por Pires & Calegario (2004), após a retirada das observações de influência, desconsiderando as classes diamétricas.

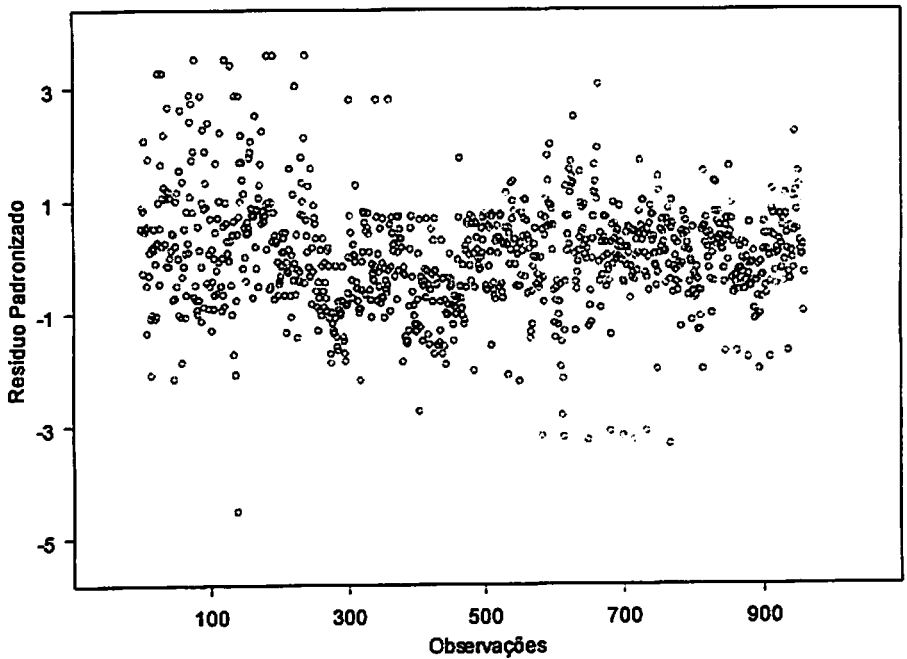


Figura 3.8- Resíduo Padronizado para detecção de “outliers”, após a retirada das observações de influência, para o polinômio proposto por Pires & Calegario (2004), desconsiderando as classes diamétricas.

De acordo com a Figura 3.8, se se utilizar o valor de t tabelado 1,96 para detectar “outliers”, verifica-se que há um grande número destes no polinômio proposto por Pires & Calegario.

A tabela 3.21 ilustra as condicionantes do resíduo, homocedasticidade e normalidade, para o polinômio proposto por Pires & Calegario (2004) após a retirada das observações de influência, desconsiderando as classes diamétricas.

TABELA 3.21 – Avaliação do resíduo segundo as pressuposições básicas de homogeneidade e normalidade, para o polinômio Pires & Calegario (2004) após a retirada das observações de influência, desconsiderando as classes diamétricas.

| Homocedasticidade | Valores obtidos | Normalidade | Valores obtidos |
|--------------------------|------------------------|--------------------|------------------------|
| $\hat{\beta}_1$ | 0,00225 | Resíduo | 94,26 |
| t tabela (0,01; 957) | 2,580982 | Padronizado (%) | |
| t calculado | 9,69 | Resíduo | 94,26 |
| Pr > t | <,0001 | Estudentizado (%) | |
| | | - | - |
| | | - | - |

Onde: t tabela, t calculado, $\hat{\beta}_1$ e Pr > |t| = já definidos anteriormente.

A normalidade foi um pouco melhor no ajuste sem a retirada das observações de influência. A homocedasticidade obteve um t calculado maior, mas ambos foram considerados homocedásticos (variância homogênea dos resíduos).

O poder de predição do polinômio proposto por Pires & Calegario (2004) desconsiderando o controle por classe, após a retirada das observações de influência, foram: PRESS = 1,120058 e o $R^2_{\text{predito}} (\%) = 99,88321$. Observando este resultado, pode-se verificar que a predição melhorou, pois o PRESS foi menor e o R^2_{predito} maior, concluindo-se, assim, que a retirada das observações de influência aumentaram o poder preditivo do modelo.

3.5 CONCLUSÕES

- De maneira geral, as condicionantes do resíduo testadas foram satisfeitas, com raras exceções.
- Foi observado a presença de valores estranhos nos três polinômios testados.
- Mesmo com a retirada dos valores estranhos (“outliers”), o polinômio proposto por Pires & Calegario (2004) continuou com a presença destes.
- O diagnóstico de influência forneceu informações coincidentes em suas várias faces. Na verificação da influência, nos coeficientes da regressão, a distância de Cook e o $DFBETAS_{j_i}$ foram claros ao determinar as observações que afetavam os coeficientes. $COVRATIO_{j_i}$ apresentou as observações que influenciavam na precisão da estimativa dos polinômios testados. Por fim, o $DFFITS_{j_i}$ apresentou as observações que influenciavam nos valores ajustados, e este, foi realmente alto em comparação aos demais diagnósticos.
- Na validação, o polinômio proposto por Pires & Calegario (2004) foi o selecionado, devido a suas medidas de acuracidade (apresentadas no capítulo 2) e a nova proposta polinomial feita pelo autor. Houve várias diferenças com relação ao ajuste, parte desta diferença é atribuída ao pequeno volume de dados utilizado (35% do dados totais).
- Após a retirada das observações de influência, diagnosticadas anteriormente, e dos valores estranhos, o polinômio proposto por Pires & Calegario (2004) apresentou melhor acuracidade e melhor distribuição dos resíduos.
- O teste de hipóteses, t de student, indica que o polinômio de 5ºGrau, sem o controle por classes diamétricas, pode ser usado com alta acuracidade.
- A Multicolinearidade estava presente em todos os polinômios testados.

3.6 REFERÊNCIAS BIBLIOGRÁFICAS

AFIFI, A.A.; AZEN, S.P. **Statistical Analysis**. Nova York. John Wiley. 1979. 442p.

ALLEN, D.M. **The prediction sum of squares as a criterion for selecting prediction variables**. Technical Report of University of Kentucky, Kentucky, n.23, 1971

ALLEN, D.M. **The relationship between variable selection and data augmentation and a method for prediction**. Alexandria, v.16, n.1, p.125-127, feb. 1974.

ANSCOMBE, F.J. **Graphs in statistical analysis**. Alexandria, 27(1), p.17-21, feb. 1973.

BEHNKEN, D.W.; DRAPER, N.R. **Residuals and their variance patterns**. Alexandria, n1., p.101-111, feb. 1972.

BELSLEY, D.A.; KUH, E.; WELSCH, R.E. **Regression diagnostics**. Nova York: John Wiley. 1980. 292p.

COOK, R.D. **Detection of influential observation in linear regression**. Alexandria, n.1, p.15-18, feb. 1977.

COOK, R.D. **Influential observations in linear regression**. Washnigton, n.365, p.169-174, mar. 1979.

DAVIES, R.B.; HUTTON, B. **The effects of errors in the independent variables in linear regression**. London, n.2, p. 383-391, aug. 1975.

DRAPER, N.R.; SMITH, H. **Applied regression analysis**. Nova York. John Wiley. 1966. 407p.

HRADETZKY, J. **Analyse und interpretation statistischer abhängigerkeiten (Biometrische Beiträge zu aktuellen forschungs projekten)**. Baden: Württemberg Mitteilungen der FVA, 1976. 146p.. (Abt. Biometric und Informatik, 21).

HOAGLIN, D.C.; WELSCH, R.E. **The hat matrix in regression and ANOVA.** Alexandria, 32(1), p.17-22., feb. 1978.

HUBER, P.J. **Robustness and designs.** A survey of Statistical Design and Linear Models. North- Holland, Amsterdam. 1975. 308p.

MATTIOLLI, C.H. **Métodos de seleção de equações de regressão linear múltipla.** Piracicaba: ESALQ-USP, 1983. 105p. Tese (Mestrado em ciência florestal) – Escola Superior de Agronomia de Piracicaba, 1983.

MONTGOMERY, D.C.; PECK, E.A. **Introduction to linear regression analysis.** Nova York: John Wiley 2 Ed. 1992. 526p.

MOSTELLER, F.; TUKEY, J.W. **Data analysis and regression: A second course in statistics.** Addison-Wesley, Mass. 1977. 428p.

PIRES, L.M.; CALEGARIO N. **Modelos Estocásticos Lineares para a Descrição do Perfil Longitudinal de Árvores.** 49ºRBRAS. Uberlândia, 2004. P.1-5.

SCOLFORO, J.R.S. “**Biometria florestal 2: Técnica de regressão aplicada para estimar: volume, biomassa, relação hipsométrica e múltiplos produtos de madeira.**” Lavras: UFLA/FAEPE, 1997. 292p.

SEBER, G.A.F. **Linear regression analysis.** Nova York: John Wiley. 1977. 465p.

WEISBERG, S. **Applied linear regression.** Nova York: John Wiley. 1980. 283p.

ANEXOS

TABELA a1 – Parâmetros estimados do polinômio de 5ºGrau para todas as classes e para cada classe.

| Polinômio 5ºGrau | $\hat{\beta}_0$ | $\hat{\beta}_1$ | $\hat{\beta}_2$ | $\hat{\beta}_3$ | $\hat{\beta}_4$ | $\hat{\beta}_5$ |
|---------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Classe 2 | 1,26603 | -3,6988 | 14,93 | -34,871 | 37,81235 | -15,438 |
| Classe 3 | 1,35404 | -5,3901 | 24,83905 | -59,128 | 63,37536 | -25,050 |
| Classe 4 | 1,27049 | -4,1476 | 18,55534 | -45,893 | 51,58059 | -21,366 |
| Classe 5 | 1,22203 | -3,9948 | 17,20782 | -40,421 | 42,87352 | -16,888 |
| Classe 6 | 1,15734 | -2,3696 | 6,60862 | -13,807 | 14,03 | -5,6201 |
| Classe 7 | 1,10494 | -2,8116 | 12,82878 | -31,591 | 33,78896 | -13,307 |
| Classe 8 | 1,13305 | -2,8040 | 12,20991 | -30,333 | 32,80891 | -13,013 |
| Todas as classes | 1,20994 | -3,7289 | 16,07902 | -37,612 | 39,24094 | -15,1874 |

TABELA a2 – Parâmetros estimados do polinômio de Potências Fracionárias e Inteiras (Hradetzky, 1976) para todas as classes e para cada classe.

| Polinômio de Potências Fracionárias | $\hat{\beta}_0$ | $\hat{\beta}_1$ | $\hat{\beta}_2$ | $\hat{\beta}_3$ | $\hat{\beta}_4$ | $\hat{\beta}_5$ | $\hat{\beta}_6$ |
|--|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Classe 2 | 1,2665 | -0,1708 | -0,7924 | -0,3033 | | | |
| Potências selecionadas | | 0,00001 | 0,9 | 30 | | | |
| Classe 3 | 1,3557 | -0,2755 | -0,5738 | -0,3691 | -0,1374 | | |
| Potências selecionadas | | 0,00001 | 0,8 | 2 | 90 | | |
| Classe 4 | 1,2722 | -0,2038 | -0,8209 | -0,2476 | | | |
| Potências selecionadas | | 0,00001 | 1 | 90 | | | |
| Classe 5 | 1,2243 | -0,1768 | -0,7689 | -0,2779 | | | |
| Potências selecionadas | | 0,00001 | 0,9 | 5 | | | |
| Classe 6 | 1,1623 | -0,1842 | -0,8099 | -0,1684 | | | |
| Potências selecionadas | | 0,2 | 1 | 15 | | | |
| Classe 7 | 1,1109 | 159,91 | -160,23 | -0,7080 | 0,5273 | -7,8068 | 7,1988 |
| Potências selecionadas | | 0,00001 | 0,0004 | 2 | 20 | 90 | 95 |
| Classe 8 | 1,1408 | -0,1252 | -0,6180 | -0,2928 | 0,4117 | -0,5164 | |
| Potências selecionadas | | 0,00001 | 1 | 2 | 15 | 20 | |
| Todas as classes | 1,2189 | -0,1887 | -0,7366 | -0,1874 | 0,2988 | -0,4052 | |
| Potências selecionadas | | 0,00001 | 1 | 3 | 25 | 95 | |

Continuação...

| | | | | | | | | |
|-------|-----|----------|---|-----|----------|---|-----|----------|
| todas | 973 | 0,012316 | 4 | 23 | 0,149697 | 8 | 178 | 0,058267 |
| todas | 543 | 0,012948 | 5 | 69 | 0,109487 | 8 | 196 | 0,058267 |
| todas | 531 | 0,013273 | 7 | 205 | 0,056327 | 8 | 1 | 0,067592 |
| todas | 655 | 0,013461 | 7 | 49 | 0,057254 | 8 | 19 | 0,067592 |
| todas | 811 | 0,013461 | 7 | 117 | 0,057694 | 8 | 35 | 0,067592 |
| todas | 555 | 0,013604 | 7 | 135 | 0,057694 | 8 | 51 | 0,067592 |
| todas | 907 | 0,013939 | 7 | 171 | 0,057694 | 8 | 67 | 0,067592 |
| todas | 991 | 0,013939 | 7 | 67 | 0,058116 | 8 | 83 | 0,067592 |
| todas | 760 | 0,01428 | 7 | 136 | 0,058577 | 8 | 99 | 0,067592 |
| todas | 604 | 0,01474 | 7 | 1 | 0,06677 | 8 | 115 | 0,067592 |
| todas | 672 | 0,014973 | 7 | 17 | 0,06677 | 8 | 131 | 0,067592 |
| todas | 690 | 0,014973 | 7 | 33 | 0,06677 | 8 | 147 | 0,067592 |
| todas | 726 | 0,014973 | 7 | 51 | 0,06677 | 8 | 163 | 0,067592 |
| todas | 622 | 0,015207 | 7 | 69 | 0,06677 | 8 | 181 | 0,067592 |
| todas | 794 | 0,018757 | 7 | 85 | 0,06677 | 8 | 199 | 0,067592 |
| todas | 692 | 0,019968 | 7 | 101 | 0,06677 | 8 | 179 | 0,067926 |
| todas | 956 | 0,020277 | 7 | 119 | 0,06677 | 8 | 197 | 0,067926 |
| todas | 974 | 0,020277 | 7 | 139 | 0,06677 | 8 | 17 | 0,068578 |
| todas | 761 | 0,021421 | 7 | 155 | 0,06677 | 8 | 18 | 0,0753 |
| todas | 957 | 0,021681 | 7 | 173 | 0,06677 | 8 | 34 | 0,0753 |
| todas | 975 | 0,021681 | 7 | 189 | 0,06677 | 8 | 50 | 0,0753 |

Continua...

Continuação...

| | | | | | | | | |
|-------|-----|----------|---|-----|----------|---|-----|--------|
| todas | 605 | 0,021707 | 7 | 207 | 0,06677 | 8 | 66 | 0,0753 |
| todas | 673 | 0,021839 | 7 | 16 | 0,071993 | 8 | 82 | 0,0753 |
| todas | 691 | 0,021839 | 7 | 32 | 0,071993 | 8 | 98 | 0,0753 |
| todas | 727 | 0,021839 | 7 | 50 | 0,071993 | 8 | 114 | 0,0753 |
| todas | 623 | 0,021962 | 7 | 68 | 0,071993 | 8 | 130 | 0,0753 |
| todas | 795 | 0,022509 | 7 | 84 | 0,071993 | 8 | 146 | 0,0753 |
| 2 | 65 | 0,170915 | 7 | 100 | 0,071993 | 8 | 162 | 0,0753 |
| 2 | 59 | 0,428523 | 7 | 118 | 0,071993 | 8 | 180 | 0,0753 |
| 3 | 41 | 0,155328 | 7 | 138 | 0,071993 | 8 | 198 | 0,0753 |
| 3 | 61 | 0,155328 | 7 | 154 | 0,071993 | 8 | 214 | 0,0753 |

TABELA a5 – Observações de influência do polinômio de Potências Fracionárias e Inteiras, utilizando o critério de Hoaglin e welsch (1978).

| classe | obs | hii | classe | obs | hii | classe | obs | hii |
|--------|-----|----------|--------|-----|----------|--------|-----|----------|
| todas | 760 | 0,012216 | 5 | 1 | 0,076923 | 7 | 154 | 0,076923 |
| todas | 604 | 0,01261 | 5 | 11 | 0,076923 | 7 | 172 | 0,076923 |
| todas | 672 | 0,012811 | 5 | 21 | 0,076923 | 7 | 188 | 0,076923 |
| todas | 690 | 0,012811 | 5 | 31 | 0,076923 | 7 | 206 | 0,076923 |
| todas | 726 | 0,012811 | 5 | 41 | 0,076923 | 7 | 222 | 0,076923 |
| todas | 622 | 0,013015 | 5 | 51 | 0,076923 | 7 | 1 | 0,076923 |
| todas | 794 | 0,016396 | 5 | 61 | 0,076923 | 7 | 17 | 0,076923 |
| todas | 956 | 0,018273 | 5 | 71 | 0,076923 | 7 | 33 | 0,076923 |
| todas | 974 | 0,018273 | 5 | 81 | 0,076923 | 7 | 51 | 0,076923 |
| todas | 761 | 0,020325 | 5 | 91 | 0,076923 | 7 | 69 | 0,076923 |
| todas | 605 | 0,021068 | 5 | 101 | 0,076923 | 7 | 85 | 0,076923 |
| todas | 673 | 0,021471 | 5 | 111 | 0,076923 | 7 | 101 | 0,076923 |
| todas | 691 | 0,021471 | 5 | 121 | 0,076923 | 7 | 119 | 0,076923 |
| todas | 727 | 0,021471 | 6 | 1 | 0,07563 | 7 | 139 | 0,076923 |
| todas | 623 | 0,021899 | 6 | 13 | 0,07563 | 7 | 155 | 0,076923 |
| todas | 795 | 0,03514 | 6 | 25 | 0,07563 | 7 | 173 | 0,076923 |

Continua...

Continuação...

| | | | | | | | | |
|-------|-----|----------|---|-----|----------|---|-----|----------|
| todas | 957 | 0,05284 | 6 | 37 | 0,07563 | 7 | 189 | 0,076923 |
| todas | 975 | 0,05284 | 6 | 49 | 0,07563 | 7 | 207 | 0,076923 |
| todas | 692 | 0,092994 | 6 | 61 | 0,07563 | 7 | 136 | 0,691475 |
| todas | 693 | 0,816485 | 6 | 73 | 0,07563 | 7 | 137 | 0,999863 |
| 3 | 41 | 0,133425 | 6 | 85 | 0,07563 | 8 | 178 | 0,069613 |
| 3 | 61 | 0,133425 | 6 | 97 | 0,07563 | 8 | 196 | 0,069613 |
| 3 | 33 | 0,142024 | 6 | 109 | 0,07563 | 8 | 18 | 0,076888 |
| 3 | 81 | 0,150321 | 6 | 121 | 0,07563 | 8 | 34 | 0,076888 |
| 3 | 13 | 0,195853 | 6 | 133 | 0,07563 | 8 | 50 | 0,076888 |
| 4 | 8 | 0,076923 | 6 | 145 | 0,07563 | 8 | 66 | 0,076888 |
| 4 | 16 | 0,076923 | 6 | 12 | 0,076883 | 8 | 82 | 0,076888 |
| 4 | 24 | 0,076923 | 6 | 24 | 0,076883 | 8 | 98 | 0,076888 |
| 4 | 32 | 0,076923 | 6 | 36 | 0,076883 | 8 | 114 | 0,076888 |
| 4 | 40 | 0,076923 | 6 | 48 | 0,076883 | 8 | 130 | 0,076888 |
| 4 | 48 | 0,076923 | 6 | 60 | 0,076883 | 8 | 146 | 0,076888 |
| 4 | 56 | 0,076923 | 6 | 72 | 0,076883 | 8 | 162 | 0,076888 |
| 4 | 64 | 0,076923 | 6 | 84 | 0,076883 | 8 | 180 | 0,076888 |
| 4 | 72 | 0,076923 | 6 | 96 | 0,076883 | 8 | 198 | 0,076888 |
| 4 | 80 | 0,076923 | 6 | 108 | 0,076883 | 8 | 214 | 0,076888 |
| 4 | 88 | 0,076923 | 6 | 120 | 0,076883 | 8 | 1 | 0,076923 |
| 4 | 96 | 0,076923 | 6 | 132 | 0,076883 | 8 | 19 | 0,076923 |
| 4 | 104 | 0,076923 | 6 | 144 | 0,076883 | 8 | 35 | 0,076923 |
| 5 | 10 | 0,075889 | 6 | 156 | 0,076883 | 8 | 51 | 0,076923 |
| 5 | 20 | 0,075889 | 7 | 117 | 0,065646 | 8 | 67 | 0,076923 |
| 5 | 30 | 0,075889 | 7 | 135 | 0,065646 | 8 | 83 | 0,076923 |
| 5 | 40 | 0,075889 | 7 | 171 | 0,065646 | 8 | 99 | 0,076923 |
| 5 | 50 | 0,075889 | 7 | 67 | 0,069788 | 8 | 115 | 0,076923 |
| 5 | 60 | 0,075889 | 7 | 16 | 0,076923 | 8 | 131 | 0,076923 |
| 5 | 70 | 0,075889 | 7 | 32 | 0,076923 | 8 | 147 | 0,076923 |
| 5 | 80 | 0,075889 | 7 | 50 | 0,076923 | 8 | 163 | 0,076923 |
| 5 | 90 | 0,075889 | 7 | 68 | 0,076923 | 8 | 181 | 0,076923 |
| 5 | 100 | 0,075889 | 7 | 84 | 0,076923 | 8 | 199 | 0,076923 |
| 5 | 110 | 0,075889 | 7 | 100 | 0,076923 | 8 | 17 | 0,208923 |
| 5 | 120 | 0,075889 | 7 | 118 | 0,076923 | 8 | 179 | 0,270918 |
| 5 | 130 | 0,075889 | 7 | 138 | 0,076923 | 8 | 197 | 0,270918 |

TABELA a6 – Observações de influência do polinômio proposto por Pires & Calegario (2004), utilizando o critério de Hoaglin e welsch (1978).

| classe | obs | hii | classe | obs | hii | classe | obs | hii |
|--------|-----|----------|--------|-----|----------|--------|-----|----------|
| todas | 555 | 0,018201 | todas | 728 | 0,041697 | 7 | 137 | 0,050841 |
| todas | 246 | 0,018216 | todas | 762 | 0,041708 | 7 | 1 | 0,056827 |
| todas | 380 | 0,018227 | todas | 744 | 0,043453 | 7 | 17 | 0,056827 |
| todas | 18 | 0,018228 | todas | 958 | 0,045396 | 7 | 33 | 0,056827 |
| todas | 190 | 0,018242 | todas | 605 | 0,047349 | 7 | 51 | 0,056827 |
| todas | 160 | 0,018365 | todas | 674 | 0,048895 | 7 | 69 | 0,056827 |
| todas | 602 | 0,018649 | todas | 710 | 0,049562 | 7 | 85 | 0,056827 |
| todas | 761 | 0,018708 | todas | 860 | 0,051087 | 7 | 101 | 0,056827 |
| todas | 24 | 0,018719 | todas | 892 | 0,054142 | 7 | 119 | 0,056827 |
| todas | 30 | 0,018774 | todas | 778 | 0,066043 | 7 | 139 | 0,056827 |
| todas | 543 | 0,018980 | todas | 572 | 0,067725 | 7 | 155 | 0,056827 |
| todas | 262 | 0,019088 | todas | 844 | 0,068550 | 7 | 173 | 0,056827 |
| todas | 436 | 0,019179 | todas | 876 | 0,068640 | 7 | 189 | 0,056827 |
| todas | 812 | 0,019503 | todas | 624 | 0,077715 | 7 | 207 | 0,056827 |
| todas | 112 | 0,019608 | todas | 606 | 0,104809 | 7 | 68 | 0,064588 |
| todas | 72 | 0,019618 | todas | 640 | 0,114103 | 7 | 154 | 0,064701 |
| todas | 6 | 0,019641 | 2 | 42 | 0,163969 | 7 | 222 | 0,064810 |
| todas | 59 | 0,019757 | 2 | 36 | 0,171867 | 7 | 16 | 0,064876 |
| todas | 270 | 0,019772 | 2 | 66 | 0,186856 | 7 | 100 | 0,065126 |
| todas | 12 | 0,019868 | 2 | 54 | 0,294827 | 7 | 206 | 0,065153 |
| todas | 692 | 0,020034 | 2 | 48 | 0,434005 | 7 | 188 | 0,065259 |
| todas | 42 | 0,020044 | 3 | 62 | 0,165733 | 7 | 32 | 0,065513 |
| todas | 104 | 0,020106 | 3 | 82 | 0,168873 | 7 | 138 | 0,065703 |
| todas | 126 | 0,020190 | 3 | 48 | 0,169899 | 7 | 172 | 0,065927 |
| todas | 66 | 0,020219 | 3 | 42 | 0,192308 | 7 | 118 | 0,066602 |
| todas | 975 | 0,020303 | 3 | 26 | 0,195133 | 7 | 50 | 0,067615 |
| todas | 36 | 0,020320 | 3 | 88 | 0,201726 | 7 | 84 | 0,067699 |
| todas | 727 | 0,020735 | 3 | 54 | 0,202758 | 8 | 178 | 0,058035 |
| todas | 254 | 0,020743 | 3 | 74 | 0,232378 | 8 | 197 | 0,060380 |
| todas | 390 | 0,020763 | 3 | 68 | 0,273317 | 8 | 17 | 0,06257 |
| todas | 638 | 0,020799 | 3 | 6 | 0,314687 | 8 | 179 | 0,066111 |
| todas | 673 | 0,021453 | 3 | 20 | 0,364841 | 8 | 1 | 0,072776 |
| todas | 622 | 0,021521 | 4 | 8 | 0,076425 | 8 | 19 | 0,072776 |
| todas | 78 | 0,021549 | 4 | 16 | 0,076425 | 8 | 35 | 0,072776 |
| todas | 843 | 0,021717 | 4 | 24 | 0,076425 | 8 | 51 | 0,072776 |
| todas | 120 | 0,022392 | 4 | 32 | 0,076425 | 8 | 67 | 0,072776 |
| todas | 98 | 0,023053 | 4 | 40 | 0,076425 | 8 | 83 | 0,072776 |
| todas | 956 | 0,023155 | 4 | 48 | 0,076425 | 8 | 99 | 0,072776 |

Continua...

Continuação...

| | | | | | | | | |
|-------|-----|----------|---|-----|----------|---|-----|----------|
| todas | 84 | 0,024055 | 4 | 56 | 0,076425 | 8 | 115 | 0,072776 |
| todas | 60 | 0,025191 | 4 | 64 | 0,076425 | 8 | 131 | 0,072776 |
| todas | 132 | 0,025249 | 4 | 72 | 0,076425 | 8 | 147 | 0,072776 |
| todas | 908 | 0,025911 | 4 | 80 | 0,07642 | 8 | 163 | 0,072776 |
| todas | 957 | 0,026197 | 4 | 88 | 0,076425 | 8 | 181 | 0,072776 |
| todas | 603 | 0,026243 | 4 | 96 | 0,076425 | 8 | 199 | 0,072776 |
| todas | 992 | 0,026873 | 4 | 104 | 0,076425 | 8 | 130 | 0,075004 |
| todas | 828 | 0,028248 | 6 | 132 | 0,107773 | 8 | 180 | 0,075005 |
| todas | 693 | 0,029153 | 6 | 24 | 0,111049 | 8 | 50 | 0,075134 |
| todas | 924 | 0,029284 | 6 | 12 | 0,131031 | 8 | 214 | 0,075143 |
| todas | 639 | 0,029558 | 6 | 48 | 0,136092 | 8 | 146 | 0,075184 |
| todas | 656 | 0,031227 | 6 | 60 | 0,136092 | 8 | 34 | 0,075250 |
| todas | 796 | 0,031532 | 6 | 72 | 0,150314 | 8 | 114 | 0,075286 |
| todas | 623 | 0,033153 | 6 | 108 | 0,157837 | 8 | 82 | 0,075294 |
| todas | 940 | 0,034177 | 6 | 144 | 0,166436 | 8 | 18 | 0,075349 |
| todas | 604 | 0,036317 | 6 | 96 | 0,189828 | 8 | 162 | 0,075551 |
| todas | 588 | 0,038187 | 6 | 120 | 0,217776 | 8 | 66 | 0,075779 |
| todas | 694 | 0,038658 | 6 | 36 | 0,238377 | 8 | 198 | 0,075911 |
| todas | 976 | 0,039442 | 6 | 84 | 0,354759 | 8 | 98 | 0,07709 |

TABELA a7 –Parâmetros estimados do polinômio proposto por Pires & Calegario (2004), para todas as classes e para cada classe, na validação dos dados.

| CLD | $\hat{\beta}_0$ | $\hat{\beta}_1$ | $\hat{\beta}_2$ | $\hat{\beta}_3$ | $\hat{\beta}_4$ | $\hat{\beta}_5$ | $\hat{\beta}_6$ | $\hat{\beta}_7$ |
|------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| 2 | 1,22399 | -0,7299 | -0,4948 | | | | | |
| PS e | | 0,5 (xd) | 5 (xd) | | | | | |
| Var | | | | | | | | |
| 3 | 1,35305 | -1,3222 | 1,03188 | -1,0640 | | | | |
| PS e | | 0,5 (xd) | 1 (xd) | 2 (xd) | | | | |
| Var | | | | | | | | |
| 4 | 1,26607 | -0,8491 | -0,4184 | | | | | |
| PS e | | 0,5 (xd) | 5 (xd) | | | | | |
| Var | | | | | | | | |
| 5 | 1,23839 | -2,5194 | 0,36970 | 0,22167 | -9,2E-4 | -0,2212 | | |
| PS e | | 0,5 (xd) | 0,5 (xb) | 1 (xd) | 2 (xb) | 5 (xd) | | |
| Var | | | | | | | | |
| 6 | 1,15807 | -0,0975 | -0,0203 | -0,2847 | | | | |
| PS e | | 0,5 (xb) | 1 (xb) | 5 (xd) | | | | |
| Var | | | | | | | | |

Continua...

Continuação...

| | | | | | | | | | |
|-------|---------|----------|----------|---------|---------|---------|---------|---------|--|
| 7 | 1,10811 | -0,1005 | -0,5628 | | | | | | |
| PS e | | 0,5 (xb) | 3 (xd) | | | | | | |
| Var | | | | | | | | | |
| 8 | 1,14418 | 0,23163 | -0,3452 | -0,2956 | -0,2433 | | | | |
| PS e | | 0,5 (xc) | 0,5 (xb) | 2 (xd) | 5 (xd) | | | | |
| Var | | | | | | | | | |
| Todas | 1,21417 | -0,2434 | 0,50681 | 0,03792 | -2,5950 | 1,92846 | -7,4E-6 | -0,6616 | |
| PS e | | 0,5 (xc) | 1 (xd) | 1 (xc) | 2 (xd) | 3 (xd) | 3 (xc) | 5 (xd) | |
| Var | | | | | | | | | |

Onde: CLD, PS, Var = já foram definidas anteriormente. O significado das variáveis se encontra no modelo proposto por Pires & Calegario(2004).

TABELA a8- Observações de influência nos parâmetros estimados, $DFBETAS_{j,i}$ para o polinômio proposto por Pires & Calegario (2004), desconsiderando o controle por classe diamétrica.

| beta | obs | dfbetas | beta | obs | dfbetas | beta | obs | dfbetas |
|------|-----|---------|------|-----|---------|------|-----|---------|
| 0 | 79 | 0,9896 | 1 | 461 | 0,0709 | 4 | 862 | 0,0919 |
| 0 | 99 | 0,5336 | 1 | 781 | 0,07 | 4 | 507 | 0,0731 |
| 0 | 127 | 0,4549 | 1 | 877 | 0,07 | 4 | 77 | 0,0699 |
| 0 | 247 | 0,4268 | 1 | 893 | 0,07 | 4 | 35 | 0,0688 |
| 0 | 37 | 0,3901 | 1 | 533 | 0,0665 | 5 | 624 | 0,18 |
| 0 | 43 | 0,3251 | 2 | 691 | 0,1902 | 5 | 640 | 0,1491 |
| 0 | 121 | 0,3152 | 2 | 692 | 0,1672 | 5 | 572 | 0,1378 |
| 0 | 161 | 0,3152 | 2 | 75 | 0,1544 | 5 | 778 | 0,1349 |
| 0 | 49 | 0,269 | 2 | 626 | 0,113 | 5 | 957 | 0,1218 |
| 0 | 113 | 0,269 | 2 | 33 | 0,1108 | 5 | 606 | 0,1205 |
| 0 | 371 | 0,2377 | 2 | 712 | 0,1102 | 5 | 876 | 0,1101 |
| 0 | 199 | 0,2335 | 2 | 590 | 0,1066 | 5 | 892 | 0,1048 |
| 0 | 207 | 0,2335 | 2 | 676 | 0,1004 | 5 | 860 | 0,0937 |
| 0 | 255 | 0,2335 | 2 | 937 | 0,1001 | 5 | 710 | 0,0866 |
| 0 | 85 | 0,2279 | 2 | 690 | 0,0988 | 5 | 77 | 0,0783 |
| 0 | 133 | 0,2279 | 2 | 830 | 0,0867 | 5 | 507 | 0,0735 |
| 0 | 141 | 0,2202 | 2 | 101 | 0,0836 | 5 | 605 | 0,0732 |
| 0 | 25 | 0,2089 | 2 | 81 | 0,0834 | 5 | 76 | 0,0652 |
| 0 | 31 | 0,2089 | 2 | 926 | 0,0808 | 5 | 695 | 0,0646 |
| 0 | 239 | 0,1909 | 2 | 655 | 0,0793 | 5 | 35 | 0,0643 |
| 0 | 93 | 0,1772 | 2 | 155 | 0,0762 | 6 | 691 | 0,258 |
| 0 | 147 | 0,1772 | 2 | 79 | 0,0756 | 6 | 692 | 0,2047 |

Continua...

Continuação...

| | | | | | | | | |
|---|-----|--------|---|-----|--------|---|-----|--------|
| 0 | 153 | 0,1772 | 2 | 957 | 0,0728 | 6 | 690 | 0,1462 |
| 0 | 321 | 0,1714 | 2 | 249 | 0,0712 | 6 | 75 | 0,1225 |
| 0 | 361 | 0,1714 | 2 | 76 | 0,0685 | 6 | 655 | 0,1112 |
| 0 | 381 | 0,1714 | 3 | 79 | 0,3309 | 6 | 151 | 0,0972 |
| 0 | 61 | 0,1576 | 3 | 99 | 0,1784 | 6 | 626 | 0,0816 |
| 0 | 175 | 0,1576 | 3 | 127 | 0,1521 | 6 | 712 | 0,0802 |
| 0 | 105 | 0,139 | 3 | 247 | 0,1427 | 6 | 33 | 0,0789 |
| 0 | 191 | 0,1273 | 3 | 37 | 0,1304 | 6 | 590 | 0,0772 |
| 0 | 167 | 0,0996 | 3 | 639 | 0,1091 | 6 | 101 | 0,0744 |
| 0 | 231 | 0,0743 | 3 | 43 | 0,1087 | 6 | 957 | 0,0737 |
| 0 | 263 | 0,0743 | 3 | 161 | 0,1054 | 6 | 676 | 0,0733 |
| 1 | 695 | 0,1416 | 3 | 121 | 0,1054 | 6 | 81 | 0,0731 |
| 1 | 797 | 0,1196 | 3 | 151 | 0,1017 | 6 | 249 | 0,067 |
| 1 | 675 | 0,1174 | 3 | 626 | 0,095 | 6 | 555 | 0,066 |
| 1 | 745 | 0,1174 | 3 | 712 | 0,0944 | 6 | 155 | 0,0657 |
| 1 | 607 | 0,1152 | 3 | 691 | 0,0928 | 6 | 145 | 0,0643 |
| 1 | 729 | 0,1152 | 3 | 113 | 0,0899 | 7 | 639 | 0,4069 |
| 1 | 711 | 0,1128 | 3 | 49 | 0,0899 | 7 | 937 | 0,3413 |
| 1 | 763 | 0,1128 | 3 | 590 | 0,0897 | 7 | 638 | 0,2691 |
| 1 | 862 | 0,1052 | 3 | 676 | 0,0865 | 7 | 693 | 0,1584 |
| 1 | 979 | 0,1026 | 3 | 371 | 0,0795 | 7 | 636 | 0,1445 |
| 1 | 425 | 0,1016 | 3 | 255 | 0,0781 | 7 | 637 | 0,1443 |
| 1 | 977 | 0,0974 | 3 | 207 | 0,0781 | 7 | 79 | 0,1268 |
| 1 | 63 | 0,0904 | 3 | 199 | 0,0781 | 7 | 692 | 0,0962 |
| 1 | 573 | 0,0857 | 3 | 59 | 0,0768 | 7 | 99 | 0,0684 |
| 1 | 557 | 0,0824 | 3 | 133 | 0,0762 | 7 | 841 | 0,0646 |
| 1 | 779 | 0,0795 | 3 | 85 | 0,0762 | 7 | 842 | 0,0642 |
| 1 | 845 | 0,0795 | 3 | 830 | 0,0738 | 8 | 507 | 0,1193 |
| 1 | 925 | 0,0795 | 3 | 145 | 0,0738 | 8 | 15 | 0,0836 |
| 1 | 847 | 0,0794 | 3 | 141 | 0,0736 | 8 | 63 | 0,0811 |
| 1 | 401 | 0,075 | 3 | 926 | 0,0698 | 8 | 77 | 0,0811 |
| 1 | 909 | 0,0734 | 3 | 692 | 0,0698 | 8 | 640 | 0,0802 |
| 1 | 941 | 0,0734 | 3 | 31 | 0,0698 | 8 | 35 | 0,0716 |
| 1 | 15 | 0,0727 | 3 | 25 | 0,0698 | 8 | 991 | 0,0703 |
| 1 | 60 | 0,0727 | 3 | 239 | 0,0638 | 8 | 862 | 0,0697 |
| 1 | 437 | 0,0709 | 4 | 63 | 0,1046 | 8 | 606 | 0,0667 |
| 1 | 449 | 0,0709 | 4 | 15 | 0,0993 | | | |

Onde: beta = parâmetro estimado pelo polinômio Pires & Calegario (2004); obs = número da observação que influência no respectivo parâmetro.

TABELA a9- Variáveis e potências selecionadas do polinômio Pires & Calegario (2004), e do polinômio de Hradetzky, agrupando as classes 2, 3 e 4, 5 e 6, e 7 e 8.

| Modelos | Classe 2, 3 e 4 | | | Classe 5 e 6 | | Classe 7 e 8 | |
|----------------------|-----------------|---------|------|--------------|------|--------------|------|
| | par. | Pot. | Var. | Pot. | Var. | Pot. | Var. |
| Hradetzky | β_0 | - | - | - | - | - | - |
| | β_1 | 0,00001 | - | 0,00001 | - | 0,00001 | - |
| | β_2 | 0,004 | - | 1 | - | 0,8 | - |
| | β_3 | 0,9 | - | 5 | - | 0,2 | - |
| | β_4 | 0,2 | - | - | - | 1 | - |
| | β_5 | - | - | - | - | 15 | - |
| | β_6 | - | - | - | - | 25 | - |
| | β_7 | - | - | - | - | 30 | - |
| Pires e Calegario | β_0 | - | - | - | - | - | - |
| | β_1 | 0,5 | xc | 1 | xd | 0,5 | xd |
| | β_2 | 1 | xd | 1 | xc | 1 | xc |
| | β_3 | 1 | xb | 1 | xb | 2 | xc |
| | β_4 | 2 | xb | 2 | xc | 4 | xd |
| | β_5 | 4 | xd | 4 | xd | 5 | xd |
| | β_6 | 5 | xb | - | - | 5 | xb |

Onde: par. = parâmetros; Pot. = Potências selecionadas pelo processo "stepwise"; Var. = Variáveis selecionadas pelo processo "stepwise".