



UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
CENTRO DE CIÊNCIAS AGRÁRIAS E ENGENHARIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS FLORESTAIS

LUANDSON ARAÚJO DE SOUZA

**ESTIMAÇÃO DE ALTURA, VOLUME E AFILAMENTO DE ÁRVORES DE
EUCALIPTO UTILIZANDO MÁQUINA DE VETOR DE SUPORTE**

JERÔNIMO MONTEIRO – ES
FEVEREIRO - 2017

LUANDSON ARAÚJO DE SOUZA

**ESTIMAÇÃO DE ALTURA, VOLUME E AFILAMENTO DE ÁRVORES DE
EUCALIPTO UTILIZANDO MÁQUINA DE VETOR DE SUPORTE**

Dissertação apresentada ao Programa de Pós-Graduação em Ciências Florestais do Centro de Ciências Agrárias e Engenharias da Universidade Federal do Espírito Santo, como parte das exigências para obtenção do título de Mestre em Ciências Florestais na área de concentração Ciências Florestais.

Orientador: Adriano Ribeiro de Mendonça
Coorientadora: Mayra Luiza Marques da Silva

JERÔNIMO MONTEIRO – ES

FEVEREIRO – 2017

Dados Internacionais de Catalogação-na-publicação (CIP)
(Biblioteca Setorial de Ciências Agrárias, Universidade Federal do Espírito Santo, ES, Brasil)

S729e Souza, Luandson Araújo de, 1990-
Estimação de altura, volume e afilamento de árvores de eucalipto utilizando máquina de vetor de suporte / Luandson Araújo de Souza. – 2017.
87 f. : il.

Orientador: Adriano Ribeiro de Mendonça.

Coorientadores: Mayra Luiza Marques da Silva.

Dissertação (Mestrado em Ciências Florestais) – Universidade Federal do Espírito Santo, Centro de Ciências Agrárias e Engenharias.

1. Eucalipto. 2. Análise de regressão. 3. Aprendizado de computador. 4. Variáveis dendrométricas. I. Mendonça, Adriano Ribeiro de. II. Silva, Mayra Luiza Marques da. III. Universidade Federal do Espírito Santo. Centro de Ciências Agrárias e Engenharias. IV. Título.

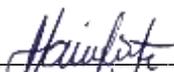
CDU: 630

ESTIMAÇÃO DE ALTURA, VOLUME E AFILAMENTO DE ÁRVORES DE EUCALIPTO UTILIZANDO MÁQUINA DE VETOR DE SUPORTE

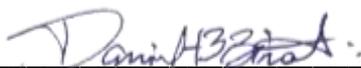
LUANDSON ARAÚJO DE SOUZA

Dissertação apresentada ao Programa de Pós-Graduação em Ciências Florestais do Centro de Ciências Agrárias e Engenharias da Universidade Federal do Espírito Santo, como parte das exigências para obtenção do título de Mestre em Ciências Florestais na área de concentração Ciências Florestais.

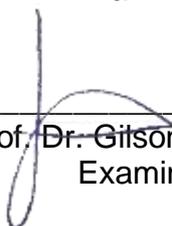
Aprovada em 21 de fevereiro de 2017.



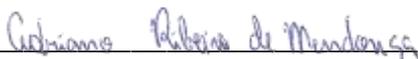
Prof. Dr. Helio Garcia Leite
Examinador externo
UFV



Dr. Daniel Henrique Breda Binoti
Examinador externo
UFES



Prof. Dr. Gilson Fernandes da Silva
Examinador interno
UFES



Prof. Dr. Adriano Ribeiro de Mendonça
Orientador
UFES

DEDICATÓRIA

*Ao meu pai, **Lucilando Celestino de Souza**, do qual tenho muito orgulho, por sempre enfatizar a importância da educação e por ser minha fonte de inspiração e à minha namorada, **Julianne Santos Serrão**, por todo o amor, apoio e conforto em todas as horas.*

DEDICO

AGRADECIMENTOS

A Deus, por sempre me dar saúde e força para superar todas as dificuldades.

Ao meu pai, Lucilando Celestino de Souza, por ser esse homem trabalhador, esforçado e que nunca mediu esforços para dar, a mim e aos meus irmãos, todas as oportunidades de estudo. E à minha mãe, Lucieuda Bezerra de Araújo, por sempre me apoiar em minhas decisões e por sempre estar torcendo por mim.

Aos meus irmãos, Felipe, Renato e Noely por todos os momentos de carinho, amizade e por todo o apoio.

À minha namorada, Julianne Santos Serrão, por ser essa mulher incrível, na qual me faz sempre querer cuidar. Obrigado por ter suportado toda essa distância, pelo amor e apoio incondicional.

Ao meu orientador, Adriano Ribeiro de Mendonça, por todo o conhecimento repassado, no qual foi de grande contribuição para a minha formação profissional. Agradeço pela confiança e amizade.

À minha coorientadora, Mayra Luiza Marques da Silva, por todas as contribuições e aprendizado repassado durante o desenvolvimento da dissertação.

Ao professor Gilson Fernandes da Silva, pela dedicação em repassar o seu conhecimento. E também pelas sugestões repassadas durante a defesa da dissertação.

Ao Daniel Henrique Breda Binoti pela ajuda e ideias durante a etapa de elaboração do projeto e durante a defesa da dissertação. E também pela parceria na elaboração de trabalhos durante o tempo de mestrado.

Ao professor Helio Garcia Leite, por ter aceitado participar da minha defesa de dissertação. As suas contribuições foram de grande ajuda para a melhoria da dissertação.

Aos amigos de graduação e de república, Harliany de Brito Matias e Marcos Alves Nicacio. Foram quatro longos dias para chegar em “Jerominho”, mas valeram a pena. Obrigado por todos os momentos de parceria e amizade e pelos ensinamentos na cozinha.

Aos amigos André Tavares de Jesus e Fábio Lacerda Jucá por toda a amizade e parceria durante os meses de república que passamos juntos.

Aos amigos do laboratório de Manejo (Manejão): Taíse Aozani, Márcia Fernandes, Janiel Silva, Luciana Lorenzoni Sandra Pires, Catherine Leite, Eth Rocha, Julyana Viana, Livia Figueiredo, Isáira Lopes, Anny Ataíde. E aos amigos e fundadores do laboratório de dendrocronologia: Evandro da Silva, Giovanni Vieira, Jeangelis Santos e Jeferson Martins. Agradeço por esse tempo que passamos juntos, por todos os momentos de brincadeiras e de conversas. Todos vocês contribuíram muito para o meu crescimento pessoal e profissional.

Aos amigos de graduação, Cláudio Nunes, Danilo Borges, José Júnior e Rodrigo da Silva, que sempre estiveram torcendo por mim. Obrigado por todos os momentos de amizade.

Ao Programa de Pós-Graduação em Ciências Florestais da Universidade Federal do Espírito Santo (UFES), pela oportunidade de crescimento profissional e à empresa FIBRIA pela bolsa de estudos.

Aos Vigilantes do NEDTEC, Seu Sebastião e Seu Cláudio, pelos momentos de conversa, brincadeiras e amizade. E a todos que, direta ou indiretamente, me ajudaram a concretizar mais esse objetivo.

SOUZA, Luandson Araújo. **Estimação de altura, volume e afilamento de arvores de eucalipto utilizando máquina de vetor de suporte**. 2017. Dissertação (Mestrado em Ciências Florestais) – Universidade Federal do Espírito Santo, Jerônimo Monteiro, ES. Orientador: Adriano Ribeiro de Mendonça. Coorientadora: Mayra Luiza Marques da Silva.

RESUMO

Diante de restrições no uso de florestas inequidêneas, o investimento em florestas equidêneas tem sido cada vez maior. Isso torna crescente o interesse em estudos buscando maximizar o rendimento em volume de madeira dessas florestas. O crescimento desse setor tem impulsionado a busca por métodos mais acurados na estimação de variáveis dendrométricas a fim de obter estimativas confiáveis a respeito do estoque volumétrico dos povoamentos florestais. Este trabalho teve como objetivo avaliar a exatidão da estimação da altura total, volume e afilamento do fuste de árvores de eucalipto, utilizando Máquina de Vetor de Suporte (MVS). Para a realização deste trabalho, foram utilizados dados provenientes de um povoamento de eucalipto. A base de dados foi dividida, aleatoriamente, em dois grupos: 70% para ajuste dos modelos de regressão e treinamento da MVS e 30% para validação. Foram ajustadas diferentes configurações de MVS e diferentes modelos de regressão. A avaliação dos modelos de regressão e das configurações da MVS foi baseada nas estatísticas: coeficiente de correlação, raiz quadrada do erro quadrático médio e viés. Com base nos resultados obtidos, foi observado que a MVS proporcionou maior exatidão nas estimativas da altura total, volume individual e afilamento do fuste em relação aos modelos clássicos de regressão utilizados.

Palavras-chave: Modelos de Regressão, Aprendizagem de Máquinas, Variáveis Dendrométricas.

SOUZA, Luandson Araújo. **Estimation of height, volume and taper of eucalyptus trees using support vector machine**. 2017. Dissertação (Mestrado em Ciências Florestais) – Universidade Federal do Espírito Santo, Jerônimo Monteiro, ES. Orientador: Adriano Ribeiro de Mendonça. Coorientadora: Mayra Luiza Marques da Silva.

ABSTRACT

Faced with restrictions on the use of uneven-aged forests, the investment in even-aged forests has been increasing. This increases the interest in studies seeking to maximize the yield in volume of these forests. The growth of this sector has driven the search for more accurate methods in the estimation of dendrometric variables in order to obtain reliable estimates of the volumetric stock of the forests stands. This work had as objective to evaluate the accuracy of estimate of total height, volume and stem taper of eucalyptus trees, using Support Vector Machine (SVM). For the realization this work, were used data from a eucalyptus stand. The database were randomly divided into two groups: 70% for fit of regression models and training of MVS and 30% for validation. Were fitted different configurations of SVM and different regression models. The evaluation of the regression models and configurations of SVM was based on the statistics: correlation coefficient, root mean square error and bias. Based on the results obtained, it was observed that the SVM provided greater accuracy in the estimates of total height, individual volume and taper in relation to the classic regression models used.

Keywords: Regression Models, Machine Learning, Dendrometric Variables.

SUMÁRIO

1 INTRODUÇÃO	14
2 OBJETIVOS	16
2.1 Objetivo Geral	16
2.2 Objetivos Específicos	16
3 REVISÃO BIBLIOGRÁFICA	17
3.1 Inventário Florestal	17
3.1.1 Relação Hipsométrica	17
3.1.2 Estimação do Volume	19
3.1.3 Afilamento do Fuste	20
3.2 Máquina de Vetor de Suporte	23
3.2.1 Hiperplano de Separação.....	23
3.2.2 Hiperplano de Margem Máxima	24
3.2.3 Margem Suave.....	25
3.2.4 Funções de <i>Kernel</i>	26
3.3 Regressão por Máquina de Vetor de Suporte	27
3.4 Arquitetura da Máquina de Vetor de Suporte Aplicada ao Problema de Regressão	30
4 METODOLOGIA	32
4.1 Descrição dos Dados	32
4.2 Modelos Hipsométricos Avaliados	32
4.3 Modelos Volumétricos Avaliados	34
4.4 Modelos de Afilamento Avaliados	35
4.5 Máquina de Vetor de Suporte	35
4.6 Critérios de Avaliação das Estimativas Obtidas por Regressão e por Máquina de Vetor de Suporte	39
5 RESULTADOS E DISCUSSÃO	41
5.1 Resultados para estimação da altura total de árvores de <i>Eucalyptus</i>	41
5.1.1 Modelos Hipsométricos	41
5.1.2 Máquina de Vetor de Suporte.....	47
5.2 Resultados para estimação do volume total de árvores de <i>Eucalyptus</i>	54
5.2.1 Modelos Volumétricos	54
5.2.2 Máquina de Vetor de Suporte.....	57

5.3 Resultados para estimação do diâmetro ao longo do fuste de árvores de <i>Eucalyptus</i>	62
5.3.1 Modelos de Afilamento	62
5.3.2 Máquina de Vetor de Suporte	67
5.4 Comparação das Configurações de Máquina de Vetor de Suporte Utilizadas nos Três Objetivos de Modelagem	75
6 CONCLUSÕES	78
REFERÊNCIAS	79

LISTA DE FIGURAS

Figura 1 – Forma do tronco de uma árvore da base para o topo.....	21
Figura 2 – Limites de decisões possíveis para um conjunto de dados separável linearmente.	24
Figura 3 – Hiperplano ótimo para padrões linearmente separáveis.....	25
Figura 4 – Funcionamento da máquina de vetor de suporte na transformação de um espaço não linearmente separável (espaço de entrada) em um espaço separável (espaço de características), por meio de funções de kernel (ϕ).	27
Figura 5 – Funcionamento linear da função de custo ϵ -insensível.....	29
Figura 6 – Arquitetura da Máquina de Vetor de Suporte	31
Figura 7 – Alturas estimadas versus alturas observadas, distribuição dos erros percentuais das estimativas de altura e histograma de erros das estimativas de altura, obtidas por análise de regressão, para os dados de ajuste.	42
Figura 8 – Alturas estimadas versus alturas observadas, distribuição dos erros percentuais das estimativas de altura e histograma de erros das estimativas de altura, obtidas por análise de regressão, para os dados de validação.	44
Figura 9 – Alturas estimadas versus alturas observadas, distribuição dos erros percentuais das estimativas de altura e histograma de erros das estimativas de altura, obtidas por análise de regressão com covariável, para os dados de ajuste. .	46
Figura 10 – Alturas estimadas versus alturas observadas, distribuição dos erros percentuais das estimativas de altura e histograma de erros das estimativas de altura, obtidas por análise de regressão com covariável, para os dados de validação.	47
Figura 11 – Alturas estimadas versus alturas observadas, distribuição dos erros percentuais das estimativas de altura e histograma de erros das estimativas de altura, obtidas por Máquina de Vetor de Suporte (MVS), para os dados de treinamento.	49
Figura 12 – Alturas estimadas versus alturas observadas, distribuição dos erros percentuais das estimativas de altura e histograma de erros das estimativas de altura, obtidas por Máquina de Vetor de Suporte (MVS), para os dados de validação.	52

Figura 13 – Volumes estimados versus volumes observados, distribuição dos erros percentuais das estimativas de volume e histograma de erros das estimativas de volume, obtidos por análise de regressão, para os dados de ajuste.	56
Figura 14 – Volumes estimados versus volumes observados, distribuição dos erros percentuais das estimativas de volume e histograma de erros das estimativas de volume, obtidos por análise de regressão, para os dados de validação.	57
Figura 15 – Volumes estimados versus volumes observados, distribuição dos erros percentuais das estimativas de volume e histograma de erros das estimativas de volume, obtidos por Máquina de Vetor de Suporte (MVS), para os dados de treinamento.	59
Figura 16 – Volumes estimados versus volumes observados, distribuição dos erros percentuais das estimativas de volume e histograma de erros das estimativas de volume, obtidos por Máquina de Vetor de Suporte (MVS), para os dados de validação.	61
Figura 17 – Diâmetros estimados versus diâmetros observados, distribuição dos erros percentuais das estimativas de diâmetro e histograma de erros das estimativas de diâmetro, obtidos por análise de regressão, para os dados de ajuste.	64
Figura 18 – Diâmetros estimados versus diâmetros observados, distribuição dos erros percentuais das estimativas de diâmetro e histograma de erros das estimativas de diâmetro, obtidos por análise de regressão, para os dados de validação.	66
Figura 19 – Diâmetros estimados versus diâmetros observados, distribuição dos erros percentuais das estimativas de diâmetro e histograma de erros das estimativas de diâmetro, obtidos por Máquina de Vetor de Suporte (MVS), para os dados de treinamento.	68
Figura 20 – Diâmetros estimados versus diâmetros observados, distribuição dos erros percentuais das estimativas de diâmetro e histograma de erros das estimativas de diâmetro, obtidos por Máquina de Vetor de Suporte (MVS), para os dados de validação.	71
Figura 21 – Perfil real do fuste (valores observados) e perfil estimado (valores estimados) pelo modelo de Ormerod e pela Máquina de Vetor de Suporte (MVS), para os dados de ajuste e treinamento, respectivamente.	74
Figura 22 – Perfil real do fuste (valores observados) e perfil estimado (valores estimados) pelo modelo de Ormerod e pela Máquina de Vetor de Suporte (MVS), para os dados de validação, respectivamente.	74

LISTA DE TABELAS

Tabela 1 – Estatísticas descritivas das variáveis dendrométricas dos povoamentos de <i>Eucalyptus grandis</i> x <i>Eucalyptus urophylla</i> , obtidas por meio da cubagem.	32
Tabela 2 – Funções de <i>kernel</i> testadas na máquina de vetor de suporte.....	37
Tabela 3 – Variáveis utilizadas nas diferentes configurações da Máquina de Vetor de Suporte (MVS)	38
Tabela 4 – Configurações utilizadas da Máquina de Vetor de Suporte (MVS) e suas respectivas denominações.....	39
Tabela 5 – Estatísticas utilizadas para avaliar o desempenho dos modelos de regressão e da máquina de vetor de suporte	39
Tabela 6 – Coeficientes estimados e estatísticas utilizadas para avaliar o desempenho do ajuste dos modelos hipsométricos na estimação da altura total de árvores de <i>Eucalyptus grandis</i> x <i>Eucalyptus urophylla</i>	41
Tabela 7 – Estatísticas utilizadas para avaliar o desempenho dos modelos hipsométricos na estimação da altura total de árvores de <i>Eucalyptus grandis</i> x <i>Eucalyptus urophylla</i> , para os dados de validação	43
Tabela 8 – Coeficientes estimados obtidos a partir do ajuste do modelo hipsométrico de Gompertz com covariável, para árvores de <i>Eucalyptus grandis</i> x <i>Eucalyptus urophylla</i>	45
Tabela 9 – Estatísticas utilizadas para avaliar o desempenho do modelo hipsométrico de Gompertz com covariável na estimação da altura total de árvores de <i>Eucalyptus grandis</i> x <i>Eucalyptus urophylla</i> , para os dados de ajuste e de validação	45
Tabela 10 – Teste da razão de máxima verossimilhança para o modelo hipsométrico de Gompertz sem (2) e com a covariável clone (2.1)	46
Tabela 11 – Estatísticas utilizadas para avaliar o desempenho da Máquina de Vetor de Suporte (MVS) na estimação da altura total de árvores de <i>Eucalyptus grandis</i> x <i>Eucalyptus urophylla</i> , para os dados de treinamento	47
Tabela 12 – Estatísticas utilizadas para avaliar o desempenho da Máquina de Vetor de Suporte (MVS) na estimação da altura total de árvores de <i>Eucalyptus grandis</i> x <i>Eucalyptus urophylla</i> , para os dados de validação	51

Tabela 13 – Coeficientes estimados e estatísticas utilizadas para avaliar o desempenho do ajuste dos modelos volumétricos na estimação do volume individual de árvores de <i>Eucalyptus grandis</i> x <i>Eucalyptus urophylla</i>	55
Tabela 14 – Estatísticas utilizadas para avaliar o desempenho dos modelos volumétricos na estimação do volume individual de árvores de <i>Eucalyptus grandis</i> x <i>Eucalyptus urophylla</i> , para os dados de validação	56
Tabela 15 – Estatísticas utilizadas para avaliar o desempenho da Máquina de Vetor de Suporte (MVS) na estimação do volume individual de árvores <i>Eucalyptus grandis</i> x <i>Eucalyptus urophylla</i> , para os dados de treinamento	58
Tabela 16 – Estatísticas utilizadas para avaliar o desempenho da Máquina de Vetor de Suporte (MVS) na estimação do volume individual de árvores de <i>Eucalyptus grandis</i> x <i>Eucalyptus urophylla</i> , para os dados de validação.....	60
Tabela 17 – Coeficientes estimados obtidos a partir do ajuste dos modelos de afilamento para árvores de <i>Eucalyptus grandis</i> x <i>Eucalyptus urophylla</i>	63
Tabela 18 – Estatísticas utilizadas para avaliar o desempenho do ajuste dos modelos de afilamento na estimação do diâmetro ao longo do fuste de árvores de <i>Eucalyptus grandis</i> x <i>Eucalyptus urophylla</i>	63
Tabela 19 – Estatísticas utilizadas para avaliar o desempenho dos modelos de afilamento na estimação do diâmetro ao longo do fuste de árvores de <i>Eucalyptus grandis</i> x <i>Eucalyptus urophylla</i> , para os dados de validação.....	65
Tabela 20 – Estatísticas utilizadas para avaliar o desempenho da Máquina de Vetor de Suporte (MVS) na estimação do diâmetro ao longo do fuste de árvores de <i>Eucalyptus grandis</i> x <i>Eucalyptus urophylla</i> , para os dados de treinamento.....	67
Tabela 21– Estatísticas utilizadas para avaliar o desempenho da Máquina de Vetor de Suporte (MVS) na estimação do diâmetro ao longo do fuste de árvores de <i>Eucalyptus grandis</i> x <i>Eucalyptus urophylla</i> , para os dados de validação	70

1 INTRODUÇÃO

Visando diminuir o desmatamento de florestas inequiâneas, são crescentes as restrições a respeito do seu uso. Isso tem aumentado a importância no comércio nacional e internacional de madeiras oriundas de reflorestamento visando à produção de papel, celulose, painéis, carvão, móveis e biomassa (SILVA; CORDEIRO, 2014). A área de florestas plantadas vem crescendo de forma acelerada e o Brasil se destaca como referência na produção desse tipo de floresta, composta principalmente por eucalipto (MIRANDA; JUNIOR; GOUVEIA, 2015). Os plantios de eucalipto correspondem a 71,9% do percentual de florestas equiâneas destinadas a fins industriais (IBÁ, 2016).

Pela importância que o eucalipto possui para a economia brasileira, aliada às diversas restrições no uso de florestas inequiâneas, são evidentes as vantagens de investimento e uso de povoamentos oriundos de reflorestamento dessa espécie. Um maior aproveitamento de plantios de eucalipto está relacionado com processos mais exatos de estimação de variáveis dendrométricas, como a altura. Processos de estimação são importantes devido à dificuldade de obtenção dessa variável por métodos diretos, que muitas vezes geram aumento de tempo e custos (LOETSCH; ZÖHRER; HALLER 1973).

A medição da altura, em sua maioria, é realizada por meio de hipsômetros, instrumentos baseados em princípios geométricos ou trigonométricos para a medição indireta. No entanto, em florestas plantadas, há grande densidade de indivíduos por hectare, o que dificulta a visão da copa e torna lenta a medição dessa variável (COUTO; BASTOS, 1988). Uma forma de minimizar essa dificuldade é obter a altura com base na medição de algumas árvores da parcela e, posteriormente, ajustar modelos de regressão para estimação da altura das demais. Tais modelos buscam estimar a altura com base no diâmetro da árvore. A essa relação é atribuído o nome de relação hipsométrica (FINGER, 1992).

Quanto ao volume de madeira, que na maioria das vezes é o principal objetivo de um inventário florestal, o método mais usual de obtenção do volume é por meio de cubagem de árvores-amostra. Essa forma de obtenção do volume também torna dispendiosa a atividade de inventário florestal, pois necessita do abate da árvore. A estimação do volume consiste na amostragem de algumas árvores do povoamento e, por meio de equações volumétricas, estima-se o volume das demais árvores.

Essas equações quantificam o volume total da árvore (CAMPOS; LEITE, 2013; HUSCH; BEERS; KERSHAW JUNIOR, 2003).

Todavia, quando se pensa em destinação de madeira para multiprodutos, faz-se uso de funções de afilamento. Essas funções buscam representar o decréscimo do diâmetro ao longo do fuste, relação denominada *taper* ou afilamento do fuste. A partir de funções de afilamento, pode-se estimar o volume de madeira para diferentes objetivos como a laminação, serraria e celulose (CLUTTER et al., 1983).

Apesar do uso de modelos de regressão ser comum na mensuração florestal, a necessidade de um maior aproveitamento dos recursos florestais vem impulsionando o setor na busca por procedimentos de estimação que proporcionem maior exatidão. Isso ocorre devido à necessidade de investimentos substanciais para a produção de madeira, tornando fundamental a busca e o desenvolvimento de modelos mais exatos para estimação de atributos da árvore e do crescimento de povoamentos (BURKHART; TOMÉ, 2012). Cordeiro et al. (2015) citam que pequenos ganhos em exatidão nas estimativas de volume em relação à árvore representam ganhos significativos quando contabilizados os valores para todo o povoamento.

Uma técnica alternativa ao uso da regressão clássica é a Máquina de Vetor de Suporte (MVS). A MVS é um algoritmo de aprendizagem proposto na década de 90, que tem como objetivo maximizar a capacidade de generalização, baseando-se na teoria do aprendizado estatístico (VAPNIK, 1999). Essa técnica vem se consolidando em diversos campos da ciência como a estimação de constituintes químicos em madeira (NUNES et al., 2011), modelagem de *cross-selling* ou venda cruzada no marketing de empresas (PINHO, 2009) e a previsão de vazões (TEIXEIRA et al., 2014). Quanto à área de mensuração florestal, alguns estudos foram desenvolvidos utilizando a MVS, podendo citar a classificação da capacidade produtiva (COSENZA et al., 2015), prognose do crescimento de árvores de eucalipto (MAGALHÃES, 2014) e predição volumétrica de árvores individuais (CORDEIRO et al., 2015).

Pelo fato de haver poucos trabalhos desenvolvidos no setor florestal utilizando máquina de vetor de suporte, surge a indagação se o uso dessa técnica proporciona um ganho em exatidão na estimação de diferentes variáveis dendrométricas quando comparada aos modelos de regressão.

2 OBJETIVOS

2.1 Objetivo Geral

Avaliar a exatidão da estimação de altura, volume e afilamento do fuste de árvores de eucalipto utilizando Máquina de Vetor de Suporte (MVS).

2.2 Objetivos Específicos

- Avaliar diferentes modelos de regressão na estimação da altura total, volume total e afilamento do fuste de híbridos de *Eucalyptus grandis* x *Eucalyptus urophylla*;
- Avaliar diferentes configurações da MVS buscando a que melhor se adequa à estimação das variáveis dendrométricas alvo desse estudo;
- Comparar a exatidão da MVS com relação aos modelos de regressão.

3 REVISÃO BIBLIOGRÁFICA

3.1 Inventário Florestal

O inventário florestal é um procedimento que visa quantificar e avaliar os povoamentos florestais (SOARES; PAULA NETO; SOUZA, 2011) e é baseado em processos de amostragem e utilizado para estimar características da floresta, que podem ser quantitativas ou qualitativas. Em florestas equiâneas, o uso de procedimentos de amostragem possibilita que a floresta seja estratificada, resultando em estratos mais homogêneos e, conseqüentemente, maior confiabilidade ao elaborar o plano de suprimento madeireiro das empresas (SCOLFORO; MELLO, 2006).

Segundo Meunier, Silva e Ferreira (2001), durante o inventário florestal, empregam-se técnicas de mapeamento, mensuração florestal e amostragem, visando obter informações precisas e confiáveis, mas com custos compatíveis. O inventário é composto por três fases: mapeamento da área, amostragem, coleta e análise de dados, que inclui o cálculo e estimativas dos parâmetros e análise de decisões de manejo (CAMPOS; LEITE, 2013).

A maioria dos inventários florestais tem como objetivo a estimativa de madeira produzida. No entanto, um inventário completo da floresta também fornece informações que ajudam no processo de tomada de decisão, sendo elas: estimativas da área, descrição da topografia, facilidades de transporte, estimativas da quantidade e qualidade da madeira e também estimativas do crescimento da floresta (HUSCH; BEERS; KERSHAW JUNIOR, 2003).

3.1.1 Relação Hipsométrica

A altura, juntamente com o diâmetro, é uma característica importante obtida por meio de um inventário florestal. Essa variável serve principalmente para quantificar o volume, incremento em volume e, juntamente com a idade, avaliar a capacidade produtiva de um determinado local. A altura também é utilizada em modelos de predição visando estimar o volume total ou volume para multiprodutos (LOETSCH; ZÖHRER; HALLER 1973).

A altura das árvores pode ser obtida por meio de medidas diretas e indiretas. As medidas diretas podem ser realizadas tanto em árvores derrubadas como em árvores em pé. Para a medição direta, usam-se trenas em árvores derrubadas e é comum o uso de réguas ou varas telescópicas em árvores em pé. Esse tipo de medida, apesar de mais acurada, torna a atividade de inventário mais demorada (SOARES; PAULA NETO; SOUZA, 2012). Já as medições indiretas se dão por meio de aparelhos conhecidos como hipsômetros ou estimação por modelos de regressão. A medição com o uso de hipsômetros pode ser prejudicada pela densidade do povoamento, pois a proximidade entre os indivíduos dificulta a visualização do topo da árvore. Outros fatores como a inclinação da árvore e a presença de ventos também contribuem para a ocorrência de erros na medição indireta da altura (COUTO; BASTOS, 1988).

Conforme Machado e Figueiredo Filho (2009), a forma mais usual de medir a altura é por meio dos hipsômetros, no entanto a dificuldade de visualização da copa torna inviável a medição de todas as árvores do povoamento. Isso torna a altura uma variável de difícil obtenção e o tempo gasto com essa atividade gera maiores custos ao inventário florestal (MENDONÇA; CARVALHO; CALEGARIO, 2015).

Retslaff et al. (2015) citam que ao medir a altura de todas as árvores, é possível que haja mais erros do que ao estimá-la. Para a estimação da altura, realiza-se a medição dessa variável em apenas algumas árvores da parcela e, por meio de análise de regressão, são geradas equações para estimar a altura das demais árvores (LEITE; ANDRADE, 2003).

O uso de modelos de regressão para estimar a altura de árvores busca expressar a relação entre a altura total e o *DAP* das árvores, sendo essa relação denominada de relação hipsométrica. No entanto, muitas vezes esses modelos apresentam valores insatisfatórios para estatísticas de exatidão, subestimando ou superestimando os resultados. Isso gera dúvidas quanto à confiabilidade do seu uso (HESS et al., 2014; SCHNEIDER; SCHNEIDER, 2008).

Alguns estudos têm sido desenvolvidos para estimação da altura utilizando outras técnicas de modelagem, como: modelos de regressão com covariáveis que têm influência na altura como o índice de sítio, a idade e a espécie (SENA et al., 2015); o uso de modelos hipsométricos generalizados mistos (MENDONÇA; CARVALHO; CALEGARIO, 2015); modelos não lineares ajustados por métodos

bayesianos empíricos em vez do ajuste pelo método dos mínimos quadrados ordinários (MOREIRA et al., 2015).

Outra técnica que vem sendo utilizada e ganhando notoriedade no setor florestal é a aprendizagem de máquinas voltadas para a predição, mais especificamente com a aplicação de Redes Neurais Artificiais (RNA). O uso desse tipo de técnica tem sido comum em estudos de relação hipsométrica e vem superando modelos clássicos de regressão em exatidão nas estimativas obtidas (BINOTI; BINOTI; LEITE, 2013; VENDRUSCOLO et al., 2015). Uma técnica de aprendizagem de máquina com alta capacidade de predição e generalização é a máquina de vetor de suporte. Essa técnica tem superado as RNAs (GARCÍA NIETO et al., 2012), no entanto ainda não foi testada a capacidade da MVS na estimação da altura.

3.1.2 Estimação do Volume

A obtenção do volume é, na maioria das vezes, a principal finalidade do inventário florestal. A estimação do volume de espécies comerciais é importante para que as empresas tenham conhecimento a respeito do potencial produtivo de suas florestas e conseqüentemente contribui para um planejamento adequado da colheita florestal (RIBEIRO; GAMA; MELO, 2014).

A forma mais usual para obtenção do volume é por meio da cubagem de árvores-amostra. No entanto, realizar a cubagem de todas as árvores do povoamento é uma atividade demorada e de alto custo, sendo necessários métodos de estimação (MACHADO; FIGUEIREDO FILHO, 2009; SCHIKOWSKI; CORTE; SANQUETTA, 2015).

Visando proporcionar maior exatidão no processo de quantificação volumétrica, aliada à diminuição dos custos de inventário oriundos do processo de cubagem, pode-se fazer o uso de equações de volume. Tais equações demandam dados de cubagem de apenas algumas árvores-amostra de cada parcela, buscando contemplar diferentes classes de diâmetro, a fim de representar bem a população (CAMPOS; LEITE, 2013).

As equações de volume são obtidas a partir do ajuste de modelos de regressão e são utilizadas quando o objetivo é a obtenção do volume total de cada árvore. Entretanto, as equações geradas por diferentes modelos podem não estimar

com exatidão o volume para diferentes espécies, idades, espaçamentos, locais e rotação, necessitando de equações específicas para determinados estratos (COUTO; BASTOS, 1987; MCTAGUE; BATISTA; STEINER, 1989). Conseqüentemente, recomenda-se o ajuste de diferentes modelos e, com base em estatísticas de exatidão, faz-se a escolha do melhor modelo para cada caso (THOMAS et al., 2006). Estudos foram desenvolvidos buscando modelos de regressão que melhor estimem o volume de populações florestais de espécies do gênero *Eucalyptus* (AZEVEDO et al., 2011; CARVALHO et al., 2011; COUTO; BASTOS, 1987; LEITE; ANDRADE, 2003; SCHRÖDER et al., 2013).

O uso de modelos de regressão para estimação volumétrica tem como desvantagem a dificuldade em inserir variáveis qualitativas na modelagem, como o espaçamento, a classe de solo e a forma do fuste, pois geralmente apresentam relações complexas e não lineares com o volume. Essas variáveis influenciam o volume da árvore e, portanto, há a necessidade de técnicas que tornem factível a inserção dessas variáveis como variáveis qualitativas (BINOTI; BINOTI; LEITE, 2014).

Alguns estudos têm sido desenvolvidos buscando a aplicação de outros métodos que tornem possível a inserção de variáveis qualitativas que exerçam influência no volume. Entre esses métodos, pode-se destacar a aplicação as redes neurais artificiais (BINOTI; BINOTI; LEITE, 2014; GORGENS et al., 2014).

Outra técnica que possui tal característica é a MVS. Em relação aos modelos de regressão, essa técnica apresenta como principal vantagem, além de permitir a inserção de variáveis qualitativas, a independência do conhecimento e suposições entre as variáveis utilizadas no ajuste (CORDEIRO et al., 2015). No trabalho de Cordeiro et al. (2015), foi testada apenas uma configuração da MVS e os resultados obtidos superaram os modelos de regressão, demonstrando que o uso de aprendizagem de máquinas no setor florestal pode gerar ganho em exatidão melhorando a estimação do estoque volumétrico.

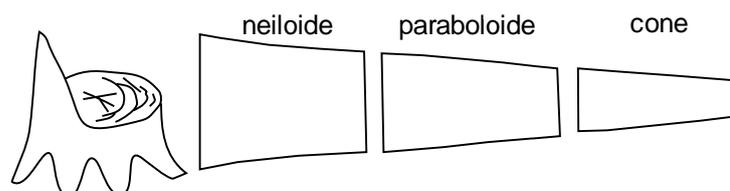
3.1.3 Afilamento do Fuste

Árvores com mesmo diâmetro e mesma altura podem apresentar volumes e rendimentos diferentes para determinados sortimentos (multiprodutos), caso as formas dos fustes sejam diferentes (SCOLFORO, 2005). Vários fatores podem afetar

a forma da árvore, entre esses, destacam-se: sítio, comprimento e largura da copa, idade, posição fitossociológica e o espaçamento (QUEIROZ et al., 2008).

Do ponto de vista geométrico, a forma do fuste de uma árvore se assemelha a um sólido de revolução (paraboloide). No entanto, numa análise mais detalhada, é mais realista considerar três partes distintas do fuste que podem ser aproximadas por diferentes sólidos de revolução (Figura 1). Esses sólidos são: um neiloide na parte inferior, paraboloide na parte central e cone na parte superior (BURKHART; TOMÉ, 2012).

Figura 1 – Forma do tronco de uma árvore da base para o topo.



Fonte: Adaptado de Row e Guttenberg (1966, citado por LOETSCH; ZÖHRER; HALLER, 1973).

Por meio da Figura 1, observa-se que o diâmetro diminui com o aumento da altura da árvore. À essa relação de decréscimo do diâmetro da base de um tronco para o topo é dado o nome de afilamento ou *taper* (CLUTTER et al., 1983). Para caracterizar essa relação de decréscimo de diâmetro, são empregados modelos de afilamento, que são utilizados para estimar o volume de sortimento da madeira de árvores em pé (ANDRADE, 2014).

A aplicação de modelos de afilamento tem como ideia básica o uso de equações que descrevam a forma e o afilamento das árvores, em que a representação da variação diamétrica ao longo do fuste é realizada em função do *DAP* e da altura total (KÖHLER; KÖHLER; FIGUEIREDO FILHO, 2013).

A partir dos modelos de afilamento, é possível estimar, por meio transformações algébricas, além do diâmetro numa certa altura, a altura que ocorre um determinado sortimento e, com a obtenção da integral desses modelos, é possível estimar o volume de diferentes porções do fuste. Esses modelos buscam representar o perfil do fuste, estimando diferentes tipos de volume requeridos, como o volume para laminação, serraria, celulose e energia (TÉO et al., 2013). São diversos os modelos estatísticos destinados para a quantificação dos sortimentos florestais. Tais modelos possuem grande variação quanto ao grau de complexidade

dos ajustes e quanto à qualidade das informações geradas (ASSIS et al., 2001; FIGUEIREDO; SCOLFORO; OLIVEIRA, 2006).

Os modelos de afilamento podem ser divididos em modelos não segmentados e modelos segmentados. Os modelos não segmentados são aqueles em que uma única função representa a forma do fuste desde a base até o ápice. Esses modelos têm como vantagem modelar o fuste numa única expressão, sendo fáceis de ajustar e empregar (CAMPOS; LEITE, 2013). Já os modelos segmentados consistem no ajuste de submodelos, em que cada um descreve uma parte do fuste e são unidos com a condição de que cada função se conecte corretamente aos pontos de união. Isso torna esses modelos mais difíceis de serem empregados, no entanto proporcionam estimativas mais acuradas quando comparados aos modelos não segmentados (SOUZA et al., 2008a).

Em busca de estimativas mais acuradas, diferentes modelos de afilamento já foram empregados em estudos realizados no Brasil (FISCHER et al., 2001; HORLE et al., 2010; MENON, 2006; PIRES; CALEGARIO, 2007; SCOLFORO et al., 1998; SOARES; LEITE; VALE, 2004; SOUZA et al., 2008b). No entanto, a proposta de inclusão de novas metodologias que estimem o estoque volumétrico com maior exatidão ainda é necessária, tendo em vista o custo que essa etapa representa para o inventário florestal (BINOTI et al., 2014).

A quantidade de informações proporcionadas pelo estudo do afilamento do fuste e pela crescente necessidade em destinar a madeira para diversos fins, diferentes técnicas de modelagem têm sido empregadas para descrever o perfil do fuste de espécies florestais (SOUZA et al., 2008c). Entre essas técnicas, pode-se citar o uso de modelos não lineares mistos (CARVALHO et al., 2014) e o uso de Inteligência Artificial (IA), com a aplicação de RNAs em estudos da forma do fuste (CAMPOS, 2014; DIAMANTOPOULOU, 2005; LEITE et al., 2011; SOARES et al., 2013).

A aplicação de técnicas de IA tem se mostrado satisfatória para descrever o afilamento do fuste, superando modelos usuais de regressão (SCHIKOWSKI; CORTE; SANQUETTA, 2015). Na IA existem outros algoritmos semelhantes às RNAs, como a MVS. A MVS também é utilizada em problemas de predição e seus resultados são comparáveis e muitas vezes superiores aos obtidos pelas RNAs (LORENA; CARVALHO, 2007).

3.2 Máquina de Vetor de Suporte

A MVS é um tipo de algoritmo de aprendizagem proposto na década de 90, baseado na teoria do aprendizado estatístico que visa controlar a capacidade de generalização (VAPNIK, 1999). Foi inicialmente desenvolvido para problemas de classificação binária e se diferencia por apresentar algumas propriedades na superfície de decisão que garantem alta capacidade de generalização (CORTES; VAPNIK, 1995).

A generalização é a capacidade do algoritmo em aprender por meio de um conjunto reduzido de exemplos e fornecer bons resultados em dados que não tiveram participação no treinamento. A capacidade de aprender a partir de exemplos e generalizar as informações aprendidas indica que o algoritmo não memorizou os dados de treinamento, processo denominado como *overfitting* (BRAGA; LUDEMIR; CARVALHO, 2000).

A MVS tem como ideia principal, levando em conta os problemas de classificação, a construção de um hiperplano como superfície de decisão (HAYKIN, 2001). A MVS implementa a seguinte ideia: ela mapeia os dados de entrada para um espaço de alta dimensão chamado espaço de características por meio de um mapeamento não linear, escolhido a priori, que são as funções de *kernel*. Nesse espaço, um hiperplano de separação máxima é construído e é gerado um preditor com máxima capacidade de generalização (VAPNIK, 1998).

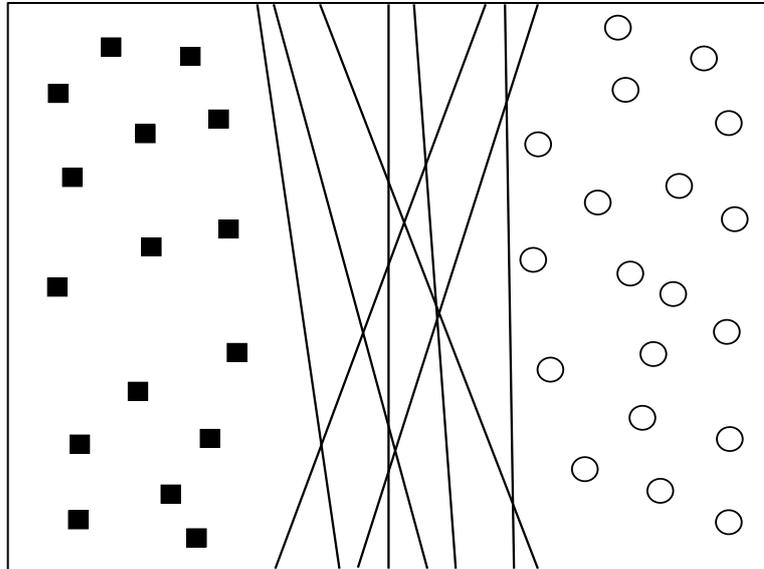
O objetivo de uma MVS é encontrar um hiperplano de separação ideal, que maximiza a margem. A margem pode ser definida como a distância entre o hiperplano separador e o vetor mais próximo de cada uma das classes (BELTRAMI, 2009). Para um melhor entendimento da essência da MVS, é necessário compreender quatro conceitos: (1) hiperplano de separação, (2) hiperplano de margem máxima, (3) margem suave e (4) funções de *kernel* (NOBLE, 2006).

3.2.1 Hiperplano de Separação

Para dados linearmente separáveis, a ideia principal de uma MVS é a construção de um hiperplano que realize a separação de duas classes distintas. É possível encontrar vários hiperplanos que satisfaçam o problema, porém deve-se

escolher aquele que permita a melhor separação, conforme pode ser observado na Figura 2.

Figura 2 – Limites de decisões possíveis para um conjunto de dados separável linearmente.



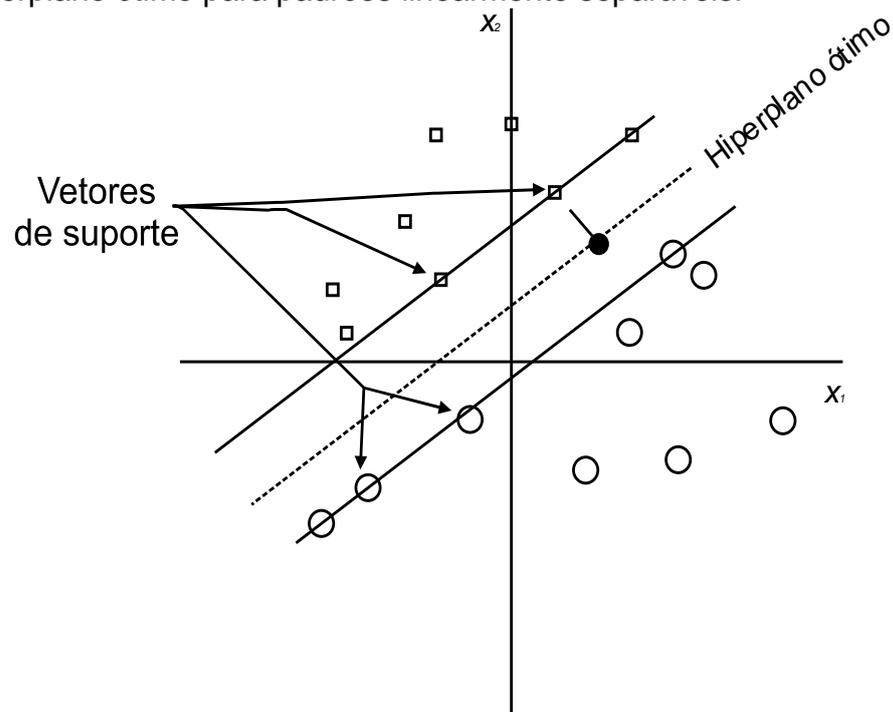
Fonte: adaptado de Tan, Steinbach e Kumar (2009).

3.2.2 Hiperplano de Margem Máxima

Com base na Figura 2, é possível observar que todos os hiperplanos satisfazem o problema em questão, conseguindo separar os dados em dois grupos distintos. É preciso escolher um hiperplano que maximize a margem, que é a distância entre o hiperplano e os vetores de suporte. Embora esses hiperplanos não apresentem erros de classificação, não há certeza de que tais hiperplanos conseguirão classificar corretamente exemplos não vistos anteriormente (TAN; STEINBACK; KUMAR, 2009).

A escolha do hiperplano ótimo ou hiperplano de margem máxima se dá durante o treinamento da MVS, em que são escolhidos os vetores de treinamento (pares de entrada e saída), que são mais importantes para a definição dos limites de cada classe. Esses vetores definem o hiperplano ótimo, que é o hiperplano com menor erro de classificação e são conhecidos como vetores de suporte (ALMEIDA; CUNHA, 2015). O hiperplano ideal é aquele que maximiza a margem, pois isso implicará num maior poder de generalização da MVS (Figura 3).

Figura 3 – Hiperplano ótimo para padrões linearmente separáveis.



Fonte: adaptado de Haykin (2001).

3.2.3 Margem Suave

Os conceitos expostos anteriormente sobre o hiperplano de margem máxima contemplam apenas o caso de conjunto de dados separáveis linearmente, também chamados de MVS de margem rígida. Em condições reais, é difícil encontrar dados que possam ser linearmente separáveis. Isso se deve à presença de ruídos nos dados, ou à própria natureza do problema, como é o caso da maioria dos problemas do mundo real (LORENA; CARVALHO, 2007).

Nesse tipo de problema, não é possível encontrar um hiperplano que maximize a margem sem que ocorram erros de classificação. Isso torna a MVS de margem rígida raramente utilizada na prática, pois forçar uma máquina de aprendizado a classificar corretamente todos os dados de treinamento, sem a permissão de ocorrência de erros, pode levar a MVS ao *overfitting* (IGEL, 2004).

A aceitação de ocorrência de erros de classificação é o que confere à MVS o título de MVS de margem suave. Nesse caso, o que se quer é separar o conjunto de treinamento com um número mínimo de erros. A solução encontrada para esse tipo de problema é a inserção de variáveis de folga, que fazem com que a MVS seja mais flexível, aceitando a ocorrência de erros (CORTES; VAPNIK, 1995).

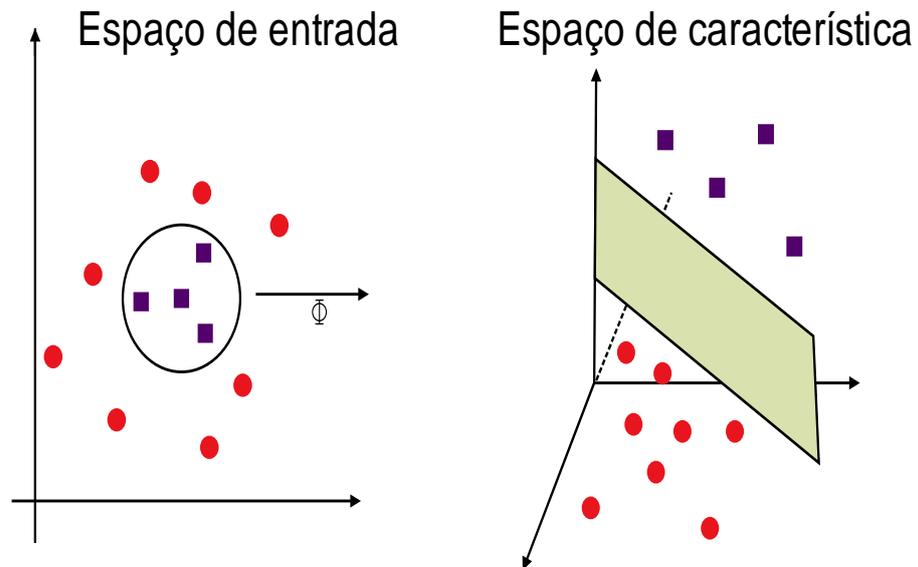
3.2.4 Funções de *Kernel*

A MVS de margem suave ainda apresenta limitações para o tratamento de problemas de separação de dados não lineares, mesmo inserindo variáveis de folga, flexibilizando o processo de aprendizagem (ALMEIDA, 2014). Para solucionar esse tipo de problema, as funções de *kernel*, ou funções de núcleo, fornecem uma solução por meio da adição de uma nova dimensão para os dados. Mais especificamente, as funções de *kernel* podem ser consideradas um artifício matemático que faz com que a MVS consiga executar uma classificação bidimensional de um conjunto de dados originalmente unidimensional (NOBLE, 2006).

Funções de *kernel* realizam o mapeamento dos dados de entrada (*input space*) para um espaço dimensional mais elevado chamado espaço de característica (*feature space*). O aumento das dimensões do espaço possibilita a separação dos dados por um hiperplano linear que minimiza os erros de classificação (KOETZ et al., 2008).

O processo de transformação de um domínio não linearmente separável em um problema linearmente separável pode ser observado na Figura 4. Essa mudança do espaço de entrada para o espaço de características representa o efeito da função de *kernel* na aplicação da MVS em problemas não lineares.

Figura 4 – Funcionamento da máquina de vetor de suporte na transformação de um espaço não linearmente separável (espaço de entrada) em um espaço separável (espaço de características), por meio de funções de kernel (ϕ).



Fonte: adaptado de Tahim (2010).

O processo de separação linear no espaço de característica é o mesmo que uma função de decisão não linear no espaço de entrada. As funções de *kernel* possibilitam o cálculo de um hiperplano sem necessariamente mapeá-lo no espaço de características. Isso faz com que a MVS consiga gerar um hiperplano de separação com margem máxima no novo espaço (TAHIM, 2010). As funções de *kernel* podem ser utilizadas para construir um hiperplano ótimo no espaço de características sem ter que considerar o próprio espaço de características de forma explícita (HAYKIN, 2001).

Essa é justamente uma grande vantagem do uso de funções de *kernel* em algoritmos de aprendizado, pois sabe-se que, conforme há aumento do espaço dimensional, os cálculos para encontrar o hiperplano de margem máxima se tornam mais complexos e as funções de *kernel* facilitam a resolução desses cálculos (GOLLAPUDI, 2016).

3.3 Regressão por Máquina de Vetor de Suporte

A MVS, quando aplicada ao problema de regressão, possui os mesmos princípios do problema de classificação, com algumas diferenças. Em problemas de regressão, busca-se um hiperplano que force as amostras a se posicionarem o mais

próximo possível do hiperplano ótimo, não importando o lado do hiperplano que as amostra se posicionem (LIMA, 2004). O uso de máquina de vetor de suporte em problemas de regressão foi inicialmente proposto por Drucker et al. (1997). Essa técnica é conhecida como regressão por vetores de suporte (*Support Vector Regression* ou SVR).

A SVR tem como objetivo otimizar uma função a partir de uma amostra dos dados de treinamento (vetores de suporte), buscando o ajuste de uma curva que explique o comportamento desses dados. Para isso, define-se que o erro cometido por uma função ou preditor é uma maneira de avaliar o seu desempenho. Uma forma de avaliação desse desempenho é a inserção de funções de perda, que medem o erro que um preditor comete no conjunto de dados da amostra (VERDÉRIO, 2015). A inserção de uma função de perda é o que faz com que a SVR consiga realizar problemas voltados para a análise de regressão e o que a diferencia da SVC (*Support vector classification*), que é a máquina de vetor de suporte para problemas de classificação.

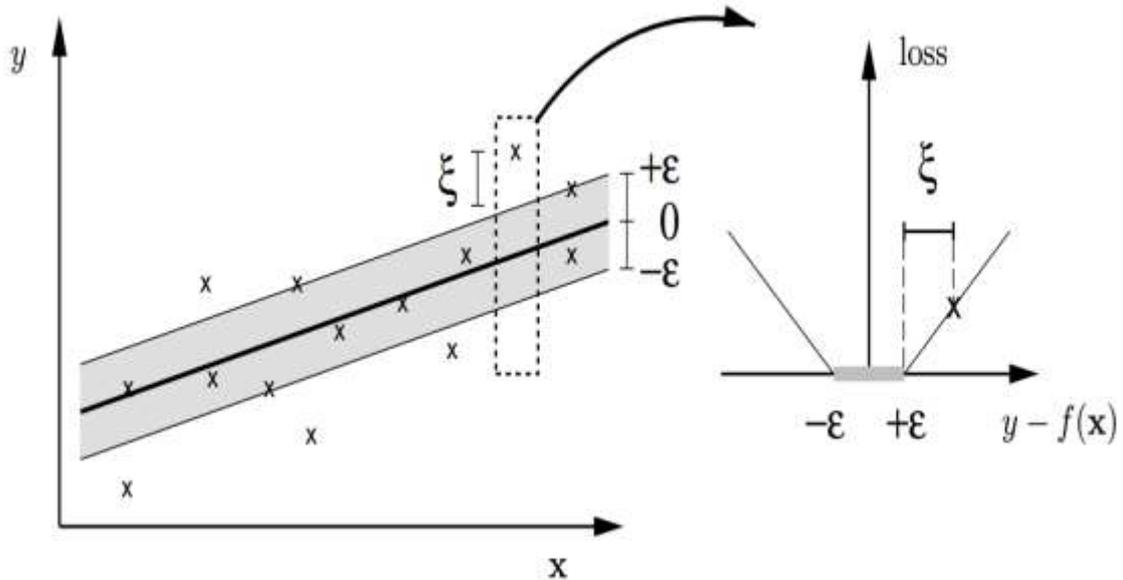
Em problemas de predição, a SVR busca encontrar uma função f que relacione os dados de entrada com dados de saída ($x \rightarrow f(x)$). A diferença obtida entre o valor dado por $f(x_i)$ e o valor de y_i é chamada de erro e é um indicador da precisão do ajuste de x_i (DIAS, 2007).

Entre os diferentes tipos de máquina de vetor de suporte aplicados ao problema de regressão, destacam-se as técnicas conhecidas como ε -SVR e ν -SVR, também denominadas como funções do tipo *epsilon-SVR* e *nu-SVR*, respectivamente. Com o uso da função de perda conhecida como *ε -insensitive loss function* ou função de perda insensível, em problemas de regressão, tem-se uma variação da MVS que é ε -SVR. Nessa função, pode-se controlar um parâmetro que é equivalente a um parâmetro de margem (ε) para separação do hiperplano (VAPNIK, 1998).

Na regressão por ε -SVR, o objetivo é encontrar uma função $f(x)$ que tenha no máximo um desvio ε dos alvos de entrada y_i para todos os dados de treinamento. Em outras palavras, não há preocupação com o erro enquanto ele estiver menor do que ε , no entanto, não se aceita um desvio maior que isso

(SMOLA; SCHÖLKOPF, 2004). O funcionamento dessa função de perda em regressão por vetores de suporte pode ser observado na Figura 5.

Figura 5 – Funcionamento linear da função de custo ε -insensível.



Fonte: Schölkopf e Smola (2002).

A Figura 5 demonstra graficamente a solução obtida para um problema de regressão linear, em que os pontos fora da região sombreada contribuem para a geração do hiperplano ótimo. Nesse exemplo, os desvios são penalizados de forma linear (SMOLA; SCHÖLKOPF, 2004).

A função de perda insensível introduz variáveis de folga não negativas ξ_i, ξ_i^* que consideram os pontos situados fora da margem $|f(x) - y_i| \leq \varepsilon$ como uma certa penalidade, separando os dados de forma a minimizar os erros de treinamento (DIAS, 2007). Isso faz com que os resultados da SVR sejam influenciados apenas por rótulos estimados que estejam sobre ou fora do limite do hiperplano ótimo (RODRIGUEZ-GALIANO et al., 2015). A margem de erro dessa função é caracterizada pelo intervalo $[y_i - \varepsilon, y_i + \varepsilon]$, em que desvios são permitidos desde que não ultrapassem a margem especificada (BELTRAMI, 2009).

Fica evidente que o parâmetro ε tem influência direta na capacidade da SVR em generalizar, pois quanto menor o valor desse parâmetro, menor será a região de insensibilidade. Isso ocasiona maior número de vetores de suporte, havendo perda da capacidade de generalização (AMARAL JÚNIOR, 2012).

Uma dificuldade da ε -SVR é a escolha do parâmetro ε . Sem um conhecimento prévio a respeito dos ruídos contidos nos dados de treinamento, a seleção desse parâmetro se torna um processo de adivinhação (SHILTON; PALANISWAMI, 2004). Pelo fato de o erro ter de ser definido a priori pelo usuário, o controle do erro é, ao mesmo tempo, uma vantagem e uma desvantagem dessa variante da máquina de vetor de suporte.

O parâmetro ε pode ser útil caso a precisão requerida possa ser especificada previamente. No entanto, em alguns casos, o que se quer é um processo de estimação mais preciso possível, sem haver determinação de certo nível de precisão (SCHÖLKOPF et al., 1999).

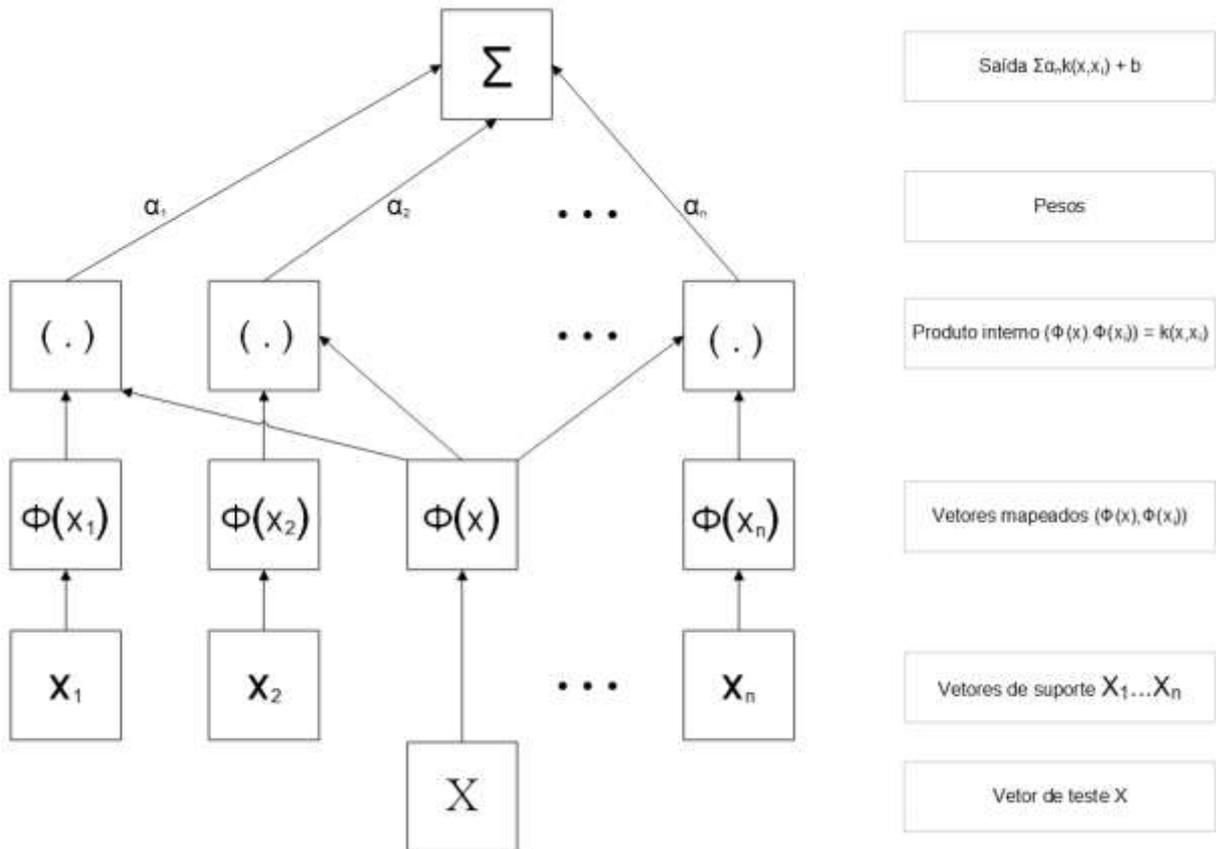
Uma alternativa a esse processo de adivinhação, visando uma melhor escolha de ε , é a utilização da ν -SVR, descrita por Schölkopf et al. (1999). Essa técnica tem como vantagem a capacidade de ajustar automaticamente o parâmetro ε (LIANG; XILONG, 2009). A ν -SVR visa minimizar o número de vetores de suporte, passando a ser definido, a priori, o parâmetro ν . Esse parâmetro delimita a fração de pontos determinados fora do tubo ou hiperplano, buscando melhorar a capacidade de generalização da SVR gerada (BISHOP, 2006).

A seleção apropriada dos parâmetros de uma MVS é fundamental para a obtenção de um bom desempenho. A combinação dos dois tipos de regressão por máquina de vetor de suporte apresentados com as funções de *kernel* mais usuais, anteriormente citadas, pode proporcionar um ganho em exatidão do preditor gerado.

3.4 Arquitetura da Máquina de Vetor de Suporte Aplicada ao Problema de Regressão

As etapas de processamento da MVS aplicada à análise de regressão podem ser observadas na Figura 6.

Figura 6 – Arquitetura da Máquina de Vetor de Suporte



Fonte: adaptado de Smola e Schölkopf (2004).

Inicialmente, os dados de entrada (X), nos quais as previsões devem ser feitas, e os vetores de suporte ($x_i, i = 1, \dots, n$), são mapeados não linearmente no espaço de características pela função de *kernel* (Φ) e então os produtos internos são calculados. Posteriormente, os produtos internos são somados utilizando os pesos (α_n). Isso, mais o termo constante b , gera a predição final de x_i . Mais detalhes sobre a arquitetura da MVS aplicadas à análise de regressão podem ser encontrados em de Smola e Schölkopf (2004).

4 METODOLOGIA

4.1 Descrição dos Dados

Os dados utilizados nesse estudo foram obtidos por meio da cubagem de 1.300 árvores-amostra oriundas de povoamentos híbridos de *Eucalyptus grandis* x *Eucalyptus urophylla*, totalizando 21 clones. Os povoamentos foram estabelecidos em espaçamento de 3 x 3 m e localizados na região centro-oeste do Estado de Minas Gerais.

As árvores-amostra foram cubadas em seções de 1 m, sendo o volume total com casca obtido pelo método de Smalian (CAMPOS; LEITE, 2013). Durante a cubagem, também foram coletados o diâmetro na altura de 1,30 m do solo (*DAP*) e a altura total (*Ht*). Para os três objetivos de modelagem, foram utilizadas as medidas coletadas apenas nas árvores cubadas. Essas árvores tinham, em média, sete anos.

Os dados foram separados, aleatoriamente, em dois grupos: 70% para ajuste dos modelos de regressão e treinamento da MVS e 30% para validação dos modelos e da MVS. As estatísticas descritivas das variáveis dendrométricas dos povoamentos estão apresentadas na Tabela 1.

Tabela 1 – Estatísticas descritivas das variáveis dendrométricas dos povoamentos de *Eucalyptus grandis* x *Eucalyptus urophylla*, obtidas por meio da cubagem.

Variável	Ajuste/Treinamento				Validação			
	Mín.	Máx.	Média	CV (%)	Mín.	Máx.	Média	CV (%)
<i>DAP</i> (cm)	4,14	28,33	13,77	42,31	4,46	28,01	13,99	41,17
<i>Ht</i> (m)	5,00	33,20	18,69	32,53	6,00	33,80	19,03	31,25
<i>V</i> (m ³)	0,002	0,93	0,18	94,35	0,003	0,79	0,18	92,45

DAP = diâmetro à altura de 1,30 m; *Ht* = altura total; *V* = volume; CV = coeficiente de variação.

4.2 Modelos Hipsométricos Avaliados

Para descrever a relação hipsométrica do povoamento em estudo, foram utilizados dados de altura total como variável dependente e o *DAP* como variável independente. Os modelos testados foram:

Curtis (1967)

$$Ht_j = \exp^{\beta_0 + \beta_1 \cdot \left(\frac{1}{DAP_j}\right)} + \varepsilon_j \quad (1)$$

Gompertz

$$Ht_j = \beta_0 \cdot \exp^{\left[-\beta_1 \cdot \exp(-\beta_2 \cdot DAP_j)\right]} + \varepsilon_j \quad (2)$$

Logístico

$$Ht_j = \left[\frac{\beta_0}{1 + \exp\left(\frac{\beta_1 - DAP_j}{\beta_2}\right)} \right] + \varepsilon_j \quad (3)$$

em que: Ht_j = altura total da j-ésima árvore (m), DAP_j = diâmetro, com casca, à uma altura de 1,30 m, da j-ésima árvore (m), \exp = exponencial; β_k = parâmetros dos modelos e ε_j = erro aleatório da estimativa da j-ésima árvore.

O ajuste dos modelos de regressão foi realizado utilizando o software estatístico R (R CORE TEAM, 2015). O ajuste dos modelos foi feito por meio da função *nls* para modelos não lineares, que faz uso do algoritmo Gauss-Newton.

Pelo fato de a altura sofrer influência de outras variáveis, além do diâmetro, o modelo de melhor desempenho teve seus parâmetros decompostos, associando a esses parâmetros a covariável clone. O ajuste do modelo com a inclusão da covariável clone foi realizado por meio do pacote *nlme*, do software R, que utiliza o método dos mínimos quadrados generalizados (função *gnls*).

A significância do tipo de clone na modelagem foi verificada por meio do Teste da Razão de Máxima Verossimilhança (TRMV) a 1% de significância. O TRMV

foi realizado para a relação hipsométrica antes e após a inclusão da covariável, utilizando-se a expressão (5):

$$TRMV = 2[\ln(mv_2) - \ln(mv_1)] \quad (5)$$

em que: \ln = logaritmo natural; mv_2 = máxima verossimilhança do modelo com covariável; mv_1 = máxima verossimilhança do modelo sem covariável.

O valor calculado é testado com distribuição de qui-quadrado (χ^2), com grau de liberdade obtido pela diferença entre o número de parâmetros do modelo com covariável e sem a covariável. O critério de informação de Akaike (AIC) também foi utilizado como método de avaliação do ajuste após a inserção da covariável. O AIC permite avaliar o quanto de informação é perdida pelo modelo utilizado, sendo o melhor modelo aquele que apresentar o menor valor (HOBBS; HILBORN, 2006).

4.3 Modelos Volumétricos Avaliados

Para estimação do volume foram utilizados o modelo não linear de Schumacher e Hall (1933) (6) e o modelo de Spurr (1952) (7). Ambos os modelos são comumente empregados na predição volumétrica e estão apresentados a seguir:

$$V_j = \beta_0 \cdot DAP_j^{\beta_1} \cdot Ht_j^{\beta_2} \cdot \varepsilon_j \quad (6)$$

$$V_j = \beta_0 + \beta_1 \cdot (DAP_j^2 \cdot Ht_j) + \varepsilon_j \quad (7)$$

em que: V_j = volume total, com casca, da j-ésima árvore (m^3).

Os modelos 6 e 7 foram ajustados, respectivamente, pelas funções *lm* e *nls* do software R

4.4 Modelos de Afilamento Avaliados

Para obtenção das estimativas de diâmetro ao longo do fuste foram utilizados os seguintes modelos encontrados em Campos e Leite (2013):

Demaerschalk (1972)

$$d_{ij} = 10^{\beta_0} \cdot DAP_j^{\beta_1} \cdot Ht_j^{\beta_2} \cdot (Ht_j - h_{ij})^{\beta_3} + \varepsilon_{ij} \quad (8)$$

Ormerod (1973)

$$d_{ij} = DAP_j \left[\frac{(Ht_j - h_{ij})}{Ht_j - 1,3} \right]^{\beta_1} + \varepsilon_{ij} \quad (9)$$

Schöepfer (1966) - Polinômio de quinto grau

$$d_{ij} = DAP_j \left[\beta_0 + \beta_1 \cdot \left(\frac{h_{ij}}{Ht_j} \right) + \beta_2 \cdot \left(\frac{h_{ij}}{Ht_j} \right)^2 + \beta_3 \cdot \left(\frac{h_{ij}}{Ht_j} \right)^3 + \beta_4 \cdot \left(\frac{h_{ij}}{Ht_j} \right)^4 + \beta_5 \cdot \left(\frac{h_{ij}}{Ht_j} \right)^5 \right] + \varepsilon_{ij} \quad (10)$$

em que: d_{ij} = diâmetro (cm), com casca, na i -ésima posição h_{ij} , da j -ésima árvore;
 h_{ij} = altura (m) em que foi medido o diâmetro d_{ij} ; ε_{ij} = erro obtido na estimativa do diâmetro (cm), com casca, na i -ésima posição h_{ij} .

O ajuste desses modelos foi realizado no software R, conforme procedimento descrito em 4.2 para modelos não lineares.

4.5 Máquina de Vetor de Suporte

Para o treinamento da MVS foram utilizadas oito configurações, formadas a partir da combinação de duas funções de erro e quatro funções de *kernel*. Foram

aplicadas as mesmas configurações para os diferentes objetivos de modelagem (relação hipsométrica, volume total e afilamento do fuste).

As funções de erro a serem otimizadas são: função do tipo I e do tipo II. Essas funções visam minimizar o erro de estimação da altura total, do volume total e do diâmetro ao longo do fuste de árvores de *Eucalyptus grandis* x *Eucalyptus urophylla*. Essas funções são conhecidas respectivamente como *epsilon*-MVS e *nu*-MVS e são descritas a seguir:

Função do tipo I

$$\text{Minimizar } \frac{1}{2} \|w\|^2 + C \cdot \sum_{i=1}^N (\xi_i + \xi_i^*) \quad (11)$$

Sujeito às seguintes restrições:

$$y_i - [w \cdot \phi(x_i) - b] \leq \varepsilon + \xi_i \quad (13)$$

$$[w \cdot \phi(x_i) + b] - y_i \leq \varepsilon + \xi_i^* \quad (12)$$

$$\xi_i, \xi_i^* \geq 0, \quad \forall i = 1, \dots, N \quad (14)$$

em que: w = vetor de coeficientes; C = constante de regularização; ξ_i, ξ_i^* = variáveis de folga que caracterizam, respectivamente, o erro acima e abaixo do ε - tubo; i = casos de treinamento; N = número total de casos de treinamento; $\phi(x_i)$ = *kernel* utilizado; b = *bias*; y_i = dados de saída e ε = erro máximo admitido.

Função do tipo II

$$\text{Minimizar } \frac{1}{2} \|w^2\| + C \left[\nu \cdot \varepsilon + \frac{1}{N} \cdot \sum_{i=1}^N (\xi_i + \xi_i^*) \right] \quad (15)$$

Sujeito às seguintes restrições:

$$[w \cdot \phi(x_i) + b] - y_i \leq \varepsilon + \xi_i \quad (16)$$

$$y_i - [w \cdot \phi(x_i) + b] \leq \varepsilon + \xi_i^* \quad (17)$$

$$\xi_i, \xi_i^* \geq 0, \quad \varepsilon \geq 0 \quad \forall i = 1, \dots, N \quad (18)$$

em que: ν = parâmetro que regula o número de vetores de suporte.

Devido à necessidade de que os parâmetros da MVS e das funções de *kernel* serem definidos de forma empírica, foi utilizado um processo iterativo para conseguir uma combinação ótima desses parâmetros que minimize o erro de estimação. Ao processo de escolha da melhor combinação dos parâmetros da MVS em conjunto com os parâmetros da função de *kernel* é dado o nome de otimização dos hiperparâmetros¹. O método utilizado para a seleção dos hiperparâmetros foi o *grid search* (pesquisa em grade), por meio da função *tune*, no software R.

Foram testadas quatro funções de *kernel*, sendo elas do tipo: linear, polinomial, função de base radial (RBF) e sigmoidal (Tabela 2).

Tabela 2 – Funções de *kernel* testadas na máquina de vetor de suporte

Tipo de <i>Kernel</i>	Função	Parâmetros	Nº
Linear	$K(X_i, X_j)$	-	(19)
Polinomial	$K(X_i, X_j) = (\gamma \cdot X_i \cdot X_j + c)^d$	γ, d, c	(20)
RBF	$K(X_i, X_j) = \exp\left(-\gamma \cdot \ X_i - X_j\ ^2\right)$	γ	(21)
Sigmoidal	$K(X_i, X_j) = \tanh(\gamma \cdot X_i \cdot X_j + c)$	γ, c	(22)

em que: $K(X_i, X_j) = \langle \phi(X_i), \phi(X_j) \rangle$ representa a função de *kernel* aplicada aos dados de entrada; γ = parâmetro de forma; d = grau do polinômio; c = constante de interceptação; \tanh = tangente hiperbólica.

Para o ajuste da MVS na predição da altura total, volume individual e diâmetro ao longo do fuste, foram utilizadas as variáveis de entrada e saída apresentadas na

¹ Parâmetros que são definidos em função de outros parâmetros (LIMA, 2004).

Tabela 3. Para fins de estratificação, o genótipo (clone) foi utilizado como variável qualitativa de entrada. A variável clone foi representada de forma binária (0 ou 1) por meio da transformação 1-de-N.

Tabela 3 – Variáveis utilizadas nas diferentes configurações da Máquina de Vetor de Suporte (MVS)

Funções MVS	Variável Estimada	Entradas	Saídas
	Ht_j	$DAP_j, Clone$	Ht_j
Tipo I e Tipo II	V_j	$DAP_j, Ht_j, Clone$	V_j
	d_{ij}	$DAP_j, h_{ij}, Ht_j, Clone$	d_{ij}

em que: Ht_j = altura total da j-ésima árvore (m); DAP_j = diâmetro (cm), com casca, à uma altura de 1,30 m, da j-ésima árvore; V_j = volume total (m³), com casca, da j-ésima árvore; d_{ij} = diâmetro (cm), com casca, na i-ésima posição h_{ij} , da j-ésima árvore; h_{ij} = altura (m) em que foi medido o diâmetro d_{ij} .

O processamento dos dados foi realizado com as variáveis de entrada e saída na sua forma padronizada, utilizando-se o método Z-score de padronização. Neste método, as variáveis passam a obter média 0 e variância 1, conforme a expressão (23):

$$Z = \frac{X_i - \mu}{\sigma} \quad (23)$$

em que: Z_i = variável padronizada; X_i = valor que se deseja padronizar; μ = média da variável a ser padronizada; σ = desvio padrão da variável a ser padronizada.

As oito configurações de MVS utilizadas, bem como suas respectivas denominações, estão apresentadas na Tabela 4.

Tabela 4 – Configurações utilizadas da Máquina de Vetor de Suporte (MVS) e suas respectivas denominações

MVS	Função	Kernel
1	Tipo I	Linear
2		RBF
3		Polinomial
4		Sigmoidal
5	Tipo II	Linear
6		RBF
7		Polinomial
8		Sigmoidal

Para o ajuste das diferentes parametrizações da MVS, foi utilizado o software R, por meio do pacote e1071 (MEYER et al., 2015), que possui funções para a regressão por máquina de vetor de suporte.

4.6 Critérios de Avaliação das Estimativas Obtidas por Regressão e por Máquina de Vetor de Suporte

Para a avaliação das estimativas geradas pelos modelos estatísticos e pela MVS, para ajuste e validação, foram obtidas as seguintes estatísticas: coeficiente de correlação linear entre os valores observados e estimados ($r_{y\hat{y}}$), viés (V%), raiz quadrada do erro quadrático médio (RQME%), apresentadas na Tabela 5.

Tabela 5 – Estatísticas utilizadas para avaliar o desempenho dos modelos de regressão e da máquina de vetor de suporte

Estatísticas	Fórmulas	Nº
Coeficiente de correlação	$r_{y\hat{y}} = \frac{\sum_{i=1}^n (Y_i - \bar{Y})(\hat{Y}_i - \hat{Y}_m)}{\sqrt{\left[\sum_{i=1}^n (Y_i - \bar{Y})^2 \right] \left[\sum_{i=1}^n (\hat{Y}_i - \hat{Y}_m)^2 \right]}}$	(24)
Viés (%)	$V(\%) = \frac{100}{\bar{Y}} \cdot \frac{\sum_{i=1}^n Y_i - \sum_{i=1}^n \hat{Y}_i}{n}$	(25)
Raiz quadrada do erro quadrático médio (%)	$RQME(\%) = \frac{100}{\bar{Y}} \cdot \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n}}$	(26)

em que: Y_i = variável dependente observada, \hat{Y}_i = variável dependente estimada, \bar{Y} = média da variável dependente observada, \hat{Y}_m = média da variável dependente estimada e n = número de observações.

Para complementar a avaliação, foi realizada a análise gráfica entre os dados observados e estimados, distribuição dos erros percentuais e histograma de erros percentuais. O erro de cada observação foi calculado em porcentagem, da seguinte forma:

$$E_i(\%) = 100 \cdot \frac{Y_i - \hat{Y}_i}{Y_i} \quad (27)$$

em que: $E_i(\%)$ = erro de cada observação.

Para verificar a significância dos parâmetros estimados dos modelos de regressão, utilizou-se o teste t a 1% de significância.

Para a avaliação das estimativas de diâmetro ao longo do fuste, também foram gerados gráficos do perfil real do fuste e do perfil estimado pelo melhor modelo de regressão e pela melhor configuração da MVS para os dados de ajuste e treinamento e para os dados de validação.

5 RESULTADOS E DISCUSSÃO

5.1 Resultados para estimação da altura total de árvores de *Eucalyptus*

5.1.1 Modelos Hipsométricos

As estatísticas de avaliação do ajuste dos modelos hipsométricos, bem como os coeficientes estimados, estão na Tabela 6.

Tabela 6 – Coeficientes estimados e estatísticas utilizadas para avaliar o desempenho do ajuste dos modelos hipsométricos na estimação da altura total de árvores de *Eucalyptus grandis* x *Eucalyptus urophylla*

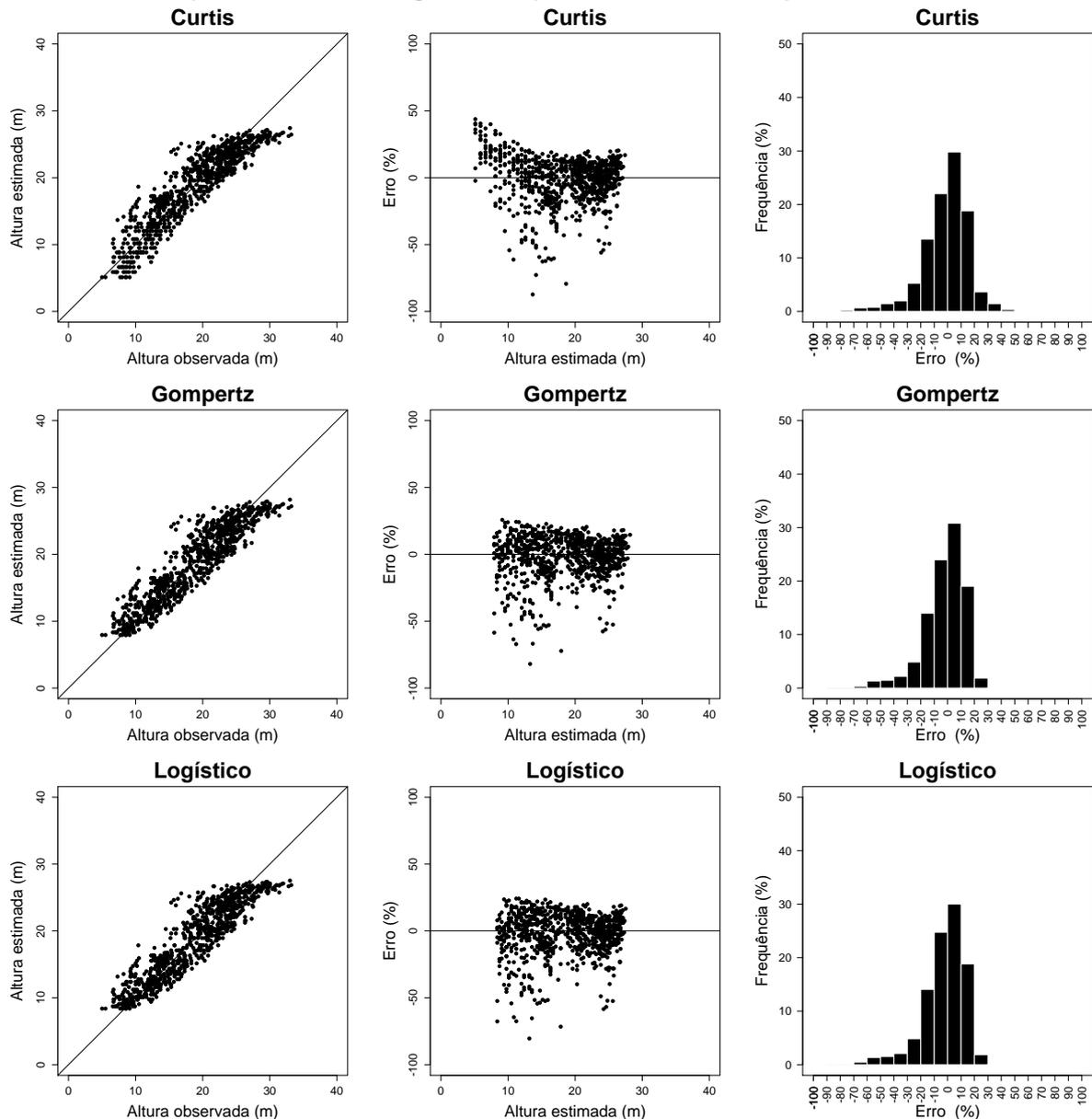
Modelo	Coeficientes			$r_{\hat{y}y}$	V (%)	RQME (%)
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$			
Curtis	3,5989**	-8,1429**	-	0,9149	0,24	13,15
Gompertz	30,5493**	2,1843**	0,1165**	0,9210	-0,01	12,66
Logístico	28,4898**	9,1434**	5,7201**	0,9199	-0,04	12,75

em que: $\hat{\beta}_k$ = estimativa dos parâmetros; $r_{\hat{y}y}$ = coeficiente de correlação entre as alturas observadas e estimadas; V (%) = viés; RQME (%) = raiz do erro quadrático médio; ** = significativo a 1% pelo teste t, ^{ns} = não significativo a 1% pelo teste t.

É possível observar, na Tabela 6, que os modelos tiveram todos os parâmetros significativos ($p < 0,01$) pelo teste *t*. Quanto às estatísticas, todos os modelos apresentaram valores de coeficiente de correlação ($r_{\hat{y}y}$) acima de 0,90. Os modelos geraram estimativas livres de viés, com valores semelhantes para todos os modelos. Em relação à estatística RQME, destaca-se o modelo de Gompertz, no entanto todos os modelos apresentaram valores semelhantes.

As análises gráficas da correlação entre as alturas observadas e estimadas pelos modelos, distribuição dos erros percentuais e histograma dos erros percentuais, são apresentadas na Figura 7.

Figura 7 – Alturas estimadas versus alturas observadas, distribuição dos erros percentuais das estimativas de altura e histograma de erros das estimativas de altura, obtidas por análise de regressão, para os dados de ajuste.



Fonte: o autor.

Em todos os modelos, o gráfico da correlação entre as alturas observadas e estimadas demonstrou que está havendo tendência mais visível nas maiores alturas. Segundo Campos e Leite (2013), nesse tipo de gráfico verifica-se a distribuição dos dados ao longo de uma linha que parte da origem e que possui inclinação de 45° , sendo ideal que os dados fiquem o mais próximo possível dessa linha.

Observa-se, com base na análise da distribuição dos erros, que os modelos de Gompertz e Logístico tiveram comportamento semelhante na distribuição dos

erros, com valores tendendo à superestimação por toda a amplitude de dados. O modelo de Curtis apresentou maior dispersão dos erros, com tendência à subestimação da altura de árvores com até dez metros e superestimação das alturas de árvores de dez a trinta metros, obtendo o pior desempenho entre os modelos avaliados. Também é possível observar a presença de possíveis *outliers* (observações discrepantes) no ajuste de todos os modelos.

Os histogramas de frequência dos erros percentuais foram semelhantes para todos os modelos, em que as maiores frequências foram nas classes de -20% a 20%. Pela análise dos histogramas, é possível notar que a distribuição dos erros é assimétrica à esquerda. Segundo Campos e Leite (2013), esse tipo de análise é importante, pois permite avaliar a exatidão e, ao mesmo tempo, verificar se a pressuposição de normalidade dos erros foi atendida.

Considerando todos os critérios estatísticos de avaliação, e também as análises gráficas realizadas, o modelo de Gompertz pode ser indicado como o de maior exatidão em estimar a altura total para os povoamentos em estudo, seguido pelo modelo Logístico. Resultados semelhantes de melhor ajuste do modelo de Gompertz já foram constatados no trabalho de SENA et al. (2015).

Na Tabela 7, estão apresentadas as estatísticas das estimativas dos modelos aplicados aos dados de validação.

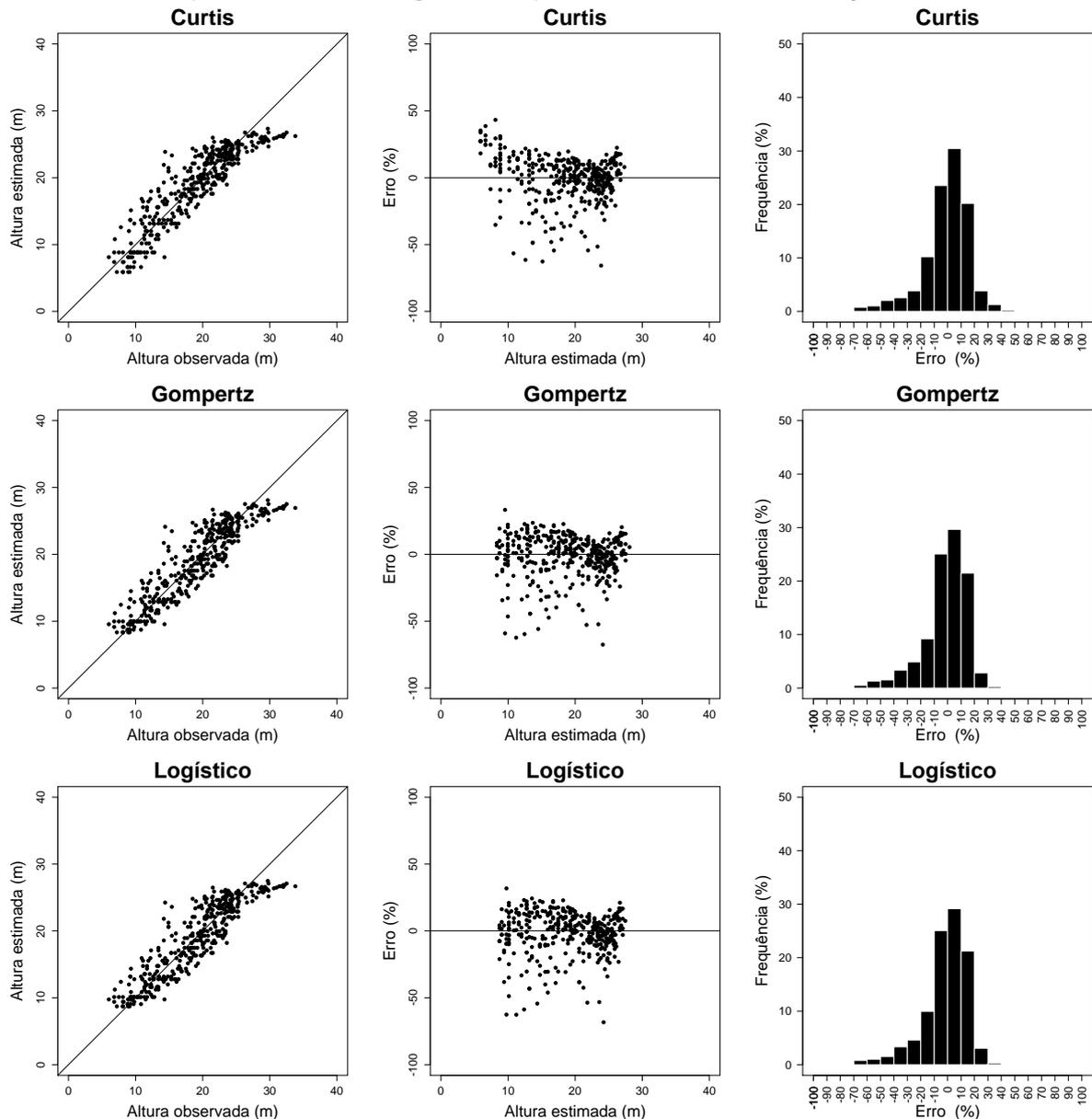
Tabela 7 – Estatísticas utilizadas para avaliar o desempenho dos modelos hipsométricos na estimação da altura total de árvores de *Eucalyptus grandis* x *Eucalyptus urophylla*, para os dados de validação

Modelo	$r_{\hat{y}y}$	V (%)	RQME (%)
Curtis	0,9018	0,69	13,57
Gompertz	0,9068	0,48	13,18
Logístico	0,9043	0,54	13,34

em que: $r_{\hat{y}y}$ = coeficiente de correlação entre as alturas observadas e estimadas; V (%) = viés; RQME (%) = raiz do erro quadrático médio.

Nota-se que todos os modelos obtiveram valores semelhantes para as estatísticas analisadas, com superioridade do modelo de Gompertz. Os resultados para a correlação entre as alturas observadas e estimadas, distribuição dos erros percentuais e histogramas dos erros percentuais se repetiu para os dados de validação, tendo o modelo de Curtis o pior desempenho (Figura 8).

Figura 8 – Alturas estimadas versus alturas observadas, distribuição dos erros percentuais das estimativas de altura e histograma de erros das estimativas de altura, obtidas por análise de regressão, para os dados de validação.



Fonte: o autor

De modo geral, deseja-se um modelo que obtenha maior exatidão não só para os dados de ajuste, como também para os dados de validação. Um modelo ajustado que possua melhor desempenho para dados desconhecidos implica em maior poder de generalização da equação gerada. Portanto, o modelo de Gompertz pode ser considerado mais consistente em relação aos resultados de ajuste e validação, apesar de os resultados estatísticos dos modelos serem semelhantes.

O modelo de melhor desempenho (Gompertz), considerando os dados de ajuste e validação, teve seus parâmetros decompostos, inserindo o tipo de clone como covariável. No modelo, a covariável clone foi associada aos parâmetros β_1 e β_2 . No modelo de Gompertz com covariável, todos os parâmetros foram significativos ($p < 0,01$) pelo teste t , como pode ser observado na Tabela 8, indicando que há variação da relação hipsométrica nos diferentes clones.

Tabela 8 – Coeficientes estimados obtidos a partir do ajuste do modelo hipsométrico de Gompertz com covariável, para árvores de *Eucalyptus grandis* x *Eucalyptus urophylla*

Coeficiente	Estimativa
$\hat{\beta}_0$	31,93036**
$\hat{\beta}_1$	1,73008**
$\hat{\beta}_{1.Clone}$	0,03925**
$\hat{\beta}_2$	0,09181**
$\hat{\beta}_{2.Clone}$	0,00145**

em que: $\hat{\beta}_k$ = estimativa dos parâmetros; ** = significativo a 1% pelo teste t .

Na Tabela 9 estão apresentadas as estatísticas de avaliação do modelo para os dados de ajuste e de validação.

Tabela 9 – Estatísticas utilizadas para avaliar o desempenho do modelo hipsométrico de Gompertz com covariável na estimação da altura total de árvores de *Eucalyptus grandis* x *Eucalyptus urophylla*, para os dados de ajuste e de validação

Dados	Estatísticas		
	$r_{y\hat{y}}$	V (%)	RQME (%)
Ajuste	0,9241	-0,02	12,42
Validação	0,9108	0,45	12,91

em que: $r_{y\hat{y}}$ = coeficiente de correlação entre as alturas observadas e estimadas; V (%) = viés; RQME (%) = raiz do erro quadrático médio.

Nota-se que houve ganho de aproximadamente 2% em exatidão (RQME) no modelo de Gompertz com a inserção da covariável clone. Também houve aumento da correlação e diminuição do viés. Na Tabela 10 está apresentado o resultado do teste de razão de máxima verossimilhança entre o modelo de Gompertz com inserção da covariável (2.1) e o modelo na forma original (2).

Tabela 10 – Teste da razão de máxima verossimilhança para o modelo hipsométrico de Gompertz sem (2) e com a covariável clone (2.1)

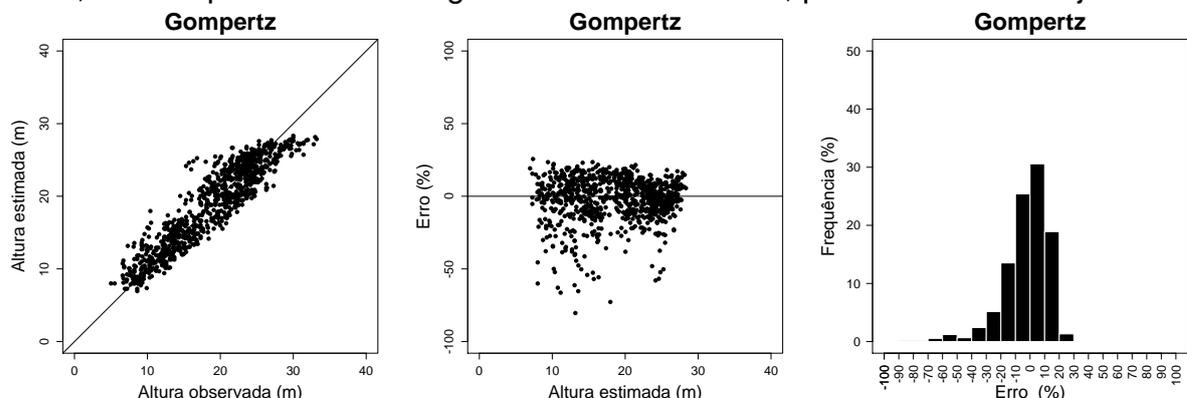
Modelo	GL	AIC	InLik	Teste	Razão	p-valor
2	4	4.153,97	-2.072,99	-	-	-
2.1	6	4.123,07	-2.055,54	2 vs 2.1	34,90	<0,0001

em que: GL = graus de liberdade; AIC = Critério de Informação de Akaike; InLik = logaritmo natural da função de máxima verossimilhança.

Observa-se que a inserção da covariável clone foi significativa ($p < 0,01$) pelo TRMV. Isso demonstra que o modelo de Gompertz foi mais exato para descrever a relação hipsométrica de *Eucalyptus grandis* x *Eucalyptus urophylla* quando foi incluído o material genético como covariável. A modelagem da relação hipsométrica com a inclusão de covariáveis resulta, na maioria das vezes, na redução de erro nas estimativas, conforme já constatada nos trabalhos de Calegario et al. (2005), Guimarães et al. (2009), Mendonça et al. (2011) e Sena et al. (2015). Para os dados de validação, a inserção da covariável clone também melhorou os resultados das estatísticas de avaliação em relação ao modelo sem covariável (Tabela 8). O ganho em exatidão foi de aproximadamente 1,5%.

Ao analisar a Figura 9, percebe-se que ainda há um comportamento semelhante dos erros, em que o modelo não conseguiu reduzir a tendência à superestimção das alturas por toda a amplitude de dados.

Figura 9 – Alturas estimadas versus alturas observadas, distribuição dos erros percentuais das estimativas de altura e histograma de erros das estimativas de altura, obtidas por análise de regressão com covariável, para os dados de ajuste.

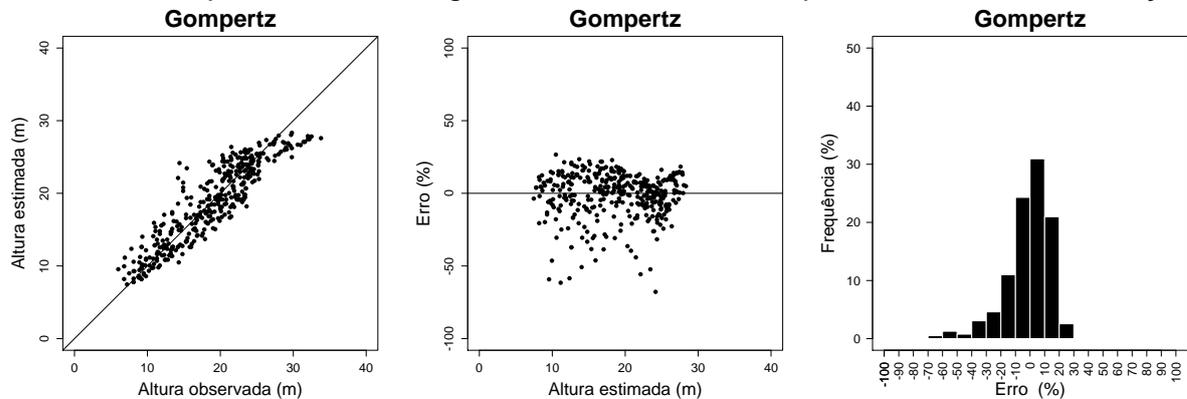


Fonte: o autor

Com base na Figura 10 pode ser observado que a inserção da covariável clone no modelo não conseguiu diminuir a tendenciosidade nas estimativas,

mantendo-se semelhantes à tendenciosidade observada no modelo na forma original, tanto para os dados de ajuste como de validação.

Figura 10 – Alturas estimadas versus alturas observadas, distribuição dos erros percentuais das estimativas de altura e histograma de erros das estimativas de altura, obtidas por análise de regressão com covariável, para os dados de validação.



Fonte: o autor

5.1.2 Máquina de Vetor de Suporte

Na Tabela 11 estão apresentados os resultados das estatísticas de avaliação do treinamento das configurações da MVS.

Tabela 11 – Estatísticas utilizadas para avaliar o desempenho da Máquina de Vetor de Suporte (MVS) na estimacão da altura total de árvores de *Eucalyptus grandis* x *Eucalyptus urophylla*, para os dados de treinamento

MVS	$r_{\hat{y}y}$	V (%)	RQME (%)
1	0,9366	2,40	11,96
2	0,9743	-0,27	7,34
3	0,9623	-0,26	8,85
4	0,9440	-0,37	10,74
5	0,9437	-0,22	10,76
6	0,9733	-0,29	7,48
7	0,9628	-0,19	8,79
8	0,9436	-0,32	10,77

em que: $r_{\hat{y}y}$ = coeficiente de correlação entre as alturas observadas e estimadas; V (%) = viés; RQME (%) = raiz do erro quadrático médio.

Todas as configurações da MVS apresentaram forte correlação ($r_{\hat{y}y}$) entre as alturas observadas e estimadas, sendo os maiores valores encontrados para a MVS 2 e MVS 6. Com exceção da MVS 1, todas as configurações geraram modelos com baixo viés, com os menores valores encontrados para as duas funções de erro combinadas com a função de *kernel* polinomial (MVS 3 e 7) e para a MVS 5.

A RQME encontrada para todas as configurações foi inferior à dos modelos de regressão utilizados. As MVSs 2 e 6 geraram, respectivamente, ganho em exatidão de aproximadamente 41% e 40% quando comparadas ao ajuste do modelo de Gompertz com covariável. Isso demonstra o alto poder preditivo da MVS em relação aos modelos de regressão. Estudos confrontando as duas técnicas, em que a MVS apresentou maior capacidade preditiva, são encontrados nos trabalhos de Binoti et al. (2016), Cordeiro et al. (2015), García Nieto et al. (2012), Görgens, Montaghi e Rodriguez (2015), Monnet, Chanussot e Berger (2011).

Ao observar a Figura 11, fica mais evidente a superioridade da MVS em relação aos modelos de regressão. É possível notar que a MVS conseguiu reduzir a dispersão dos erros, gerando estimativas de maior exatidão.

Figura 11 – Alturas estimadas versus alturas observadas, distribuição dos erros percentuais das estimativas de altura e histograma de erros das estimativas de altura, obtidas por Máquina de Vetor de Suporte (MVS), para os dados de treinamento.

(Continua)

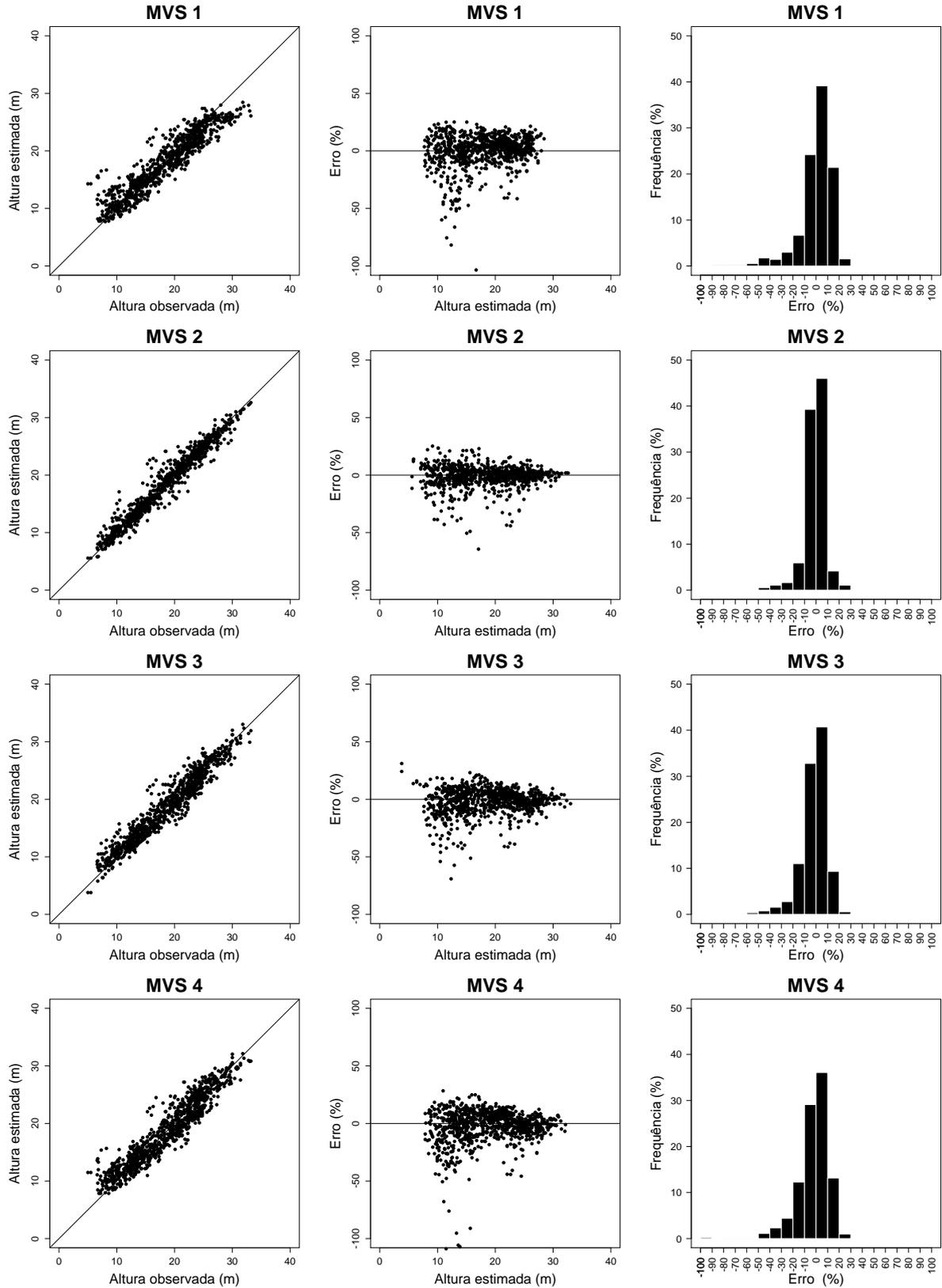
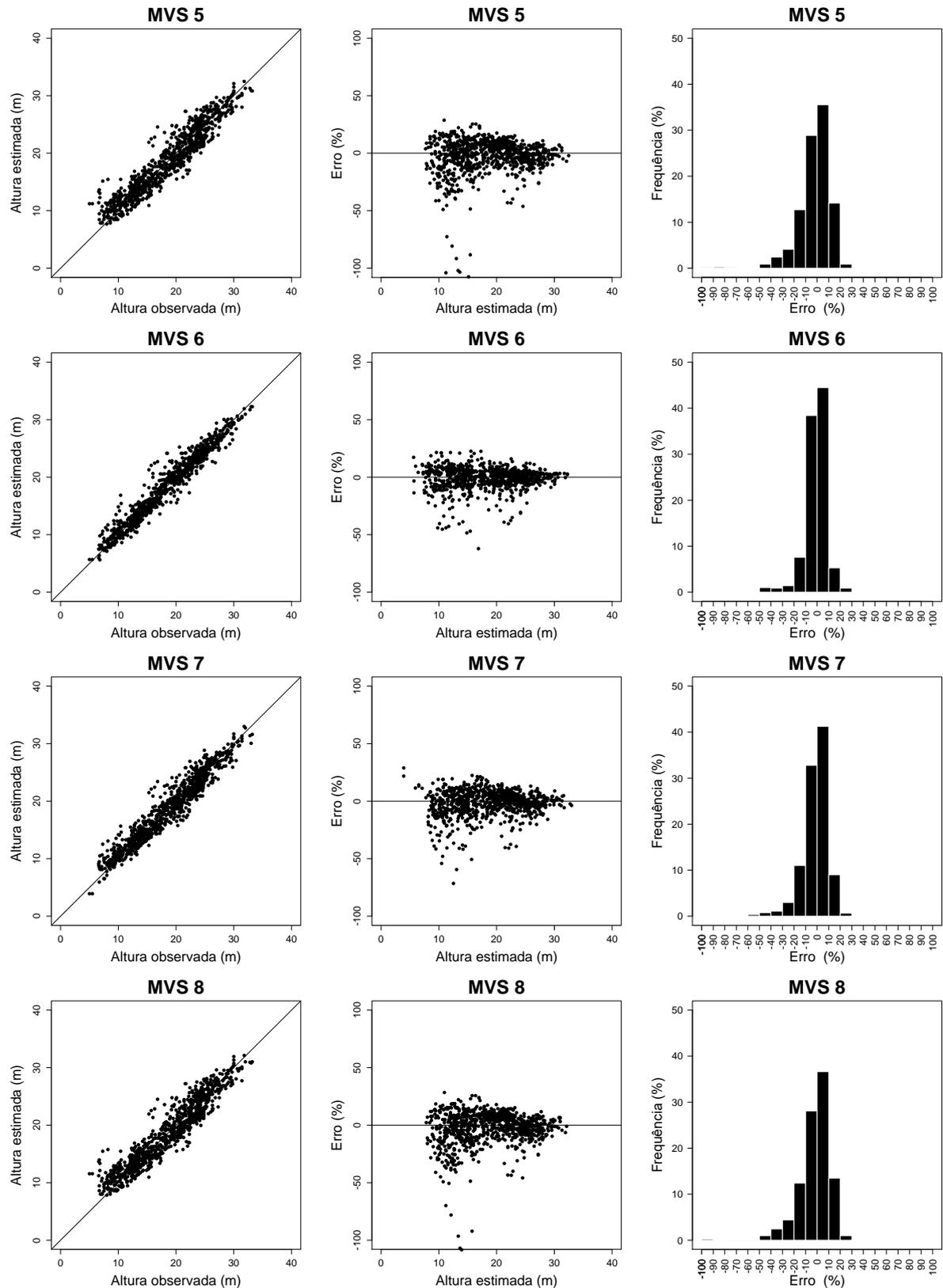


Figura 11– Alturas estimadas versus alturas observadas, distribuição dos erros percentuais das estimativas de altura e histograma de erros das estimativas de altura, obtidas por Máquina de Vetor de Suporte (MVS), para os dados de treinamento.

(Conclusão)



Fonte: o autor

Observa-se, nos histogramas dos erros percentuais da MVS 2 e 6, que aproximadamente 80% dos erros estão agrupados na amplitude das classes de $\pm 10\%$, corroborando a diminuição da dispersão dos erros em relação aos modelos de regressão.

Para avaliar a exatidão da generalização da MVS, na Tabela 12 estão apresentadas as estatísticas de avaliação dos modelos gerados aplicados aos dados de validação.

Tabela 12 – Estatísticas utilizadas para avaliar o desempenho da Máquina de Vetor de Suporte (MVS) na estimação da altura total de árvores de *Eucalyptus grandis* x *Eucalyptus urophylla*, para os dados de validação

MVS	$r_{\hat{y}y}$	V (%)	RQME (%)
1	0,9348	-0,05	11,12
2	0,9617	-0,13	8,66
3	0,9559	-0,05	9,20
4	0,9348	-0,06	11,09
5	0,9353	0,01	11,07
6	0,9615	-0,15	8,65
7	0,9557	-0,06	9,22
8	0,9345	-0,02	11,12

em que: $r_{\hat{y}y}$ = coeficiente de correlação entre as alturas observadas e estimadas; V (%) = viés; RQME (%) = raiz do erro quadrático médio.

Nota-se que a MVS também apresentou maior exatidão nas estimativas de altura para os dados de validação, com destaque para as configurações 2 e 6, seguidas pelas configurações 3 e 7, que utilizaram a função de *kernel* polinomial. Isso demonstra que a MVS conseguiu aprender a relação biológica em que a altura pode ser explicada pelo diâmetro e que não ocorreu *overfitting* durante o treinamento da MVS.

Os valores de RQME encontrados para as configurações 2, 3, 6 e 7, quanto à generalização da altura, podem ser considerados baixos. Vendruscolo et al. (2015), ao confrontar a exatidão da generalização de modelos de regressão e redes neurais artificiais na estimação da altura de árvores de eucalipto, encontraram, respectivamente, RQME de 9,57% e de 9,47%. Isso demonstra a viabilidade em utilizar a MVS para predição da altura de espécies do gênero *Eucalyptus*.

A maior exatidão da generalização da MVS em relação ao modelo de Gompertz com covariável foi evidenciada na análise dos gráficos de distribuição dos erros na Figura 12.

Figura 12 – Alturas estimadas versus alturas observadas, distribuição dos erros percentuais das estimativas de altura e histograma de erros das estimativas de altura, obtidas por Máquina de Vetor de Suporte (MVS), para os dados de validação.
(Continua)

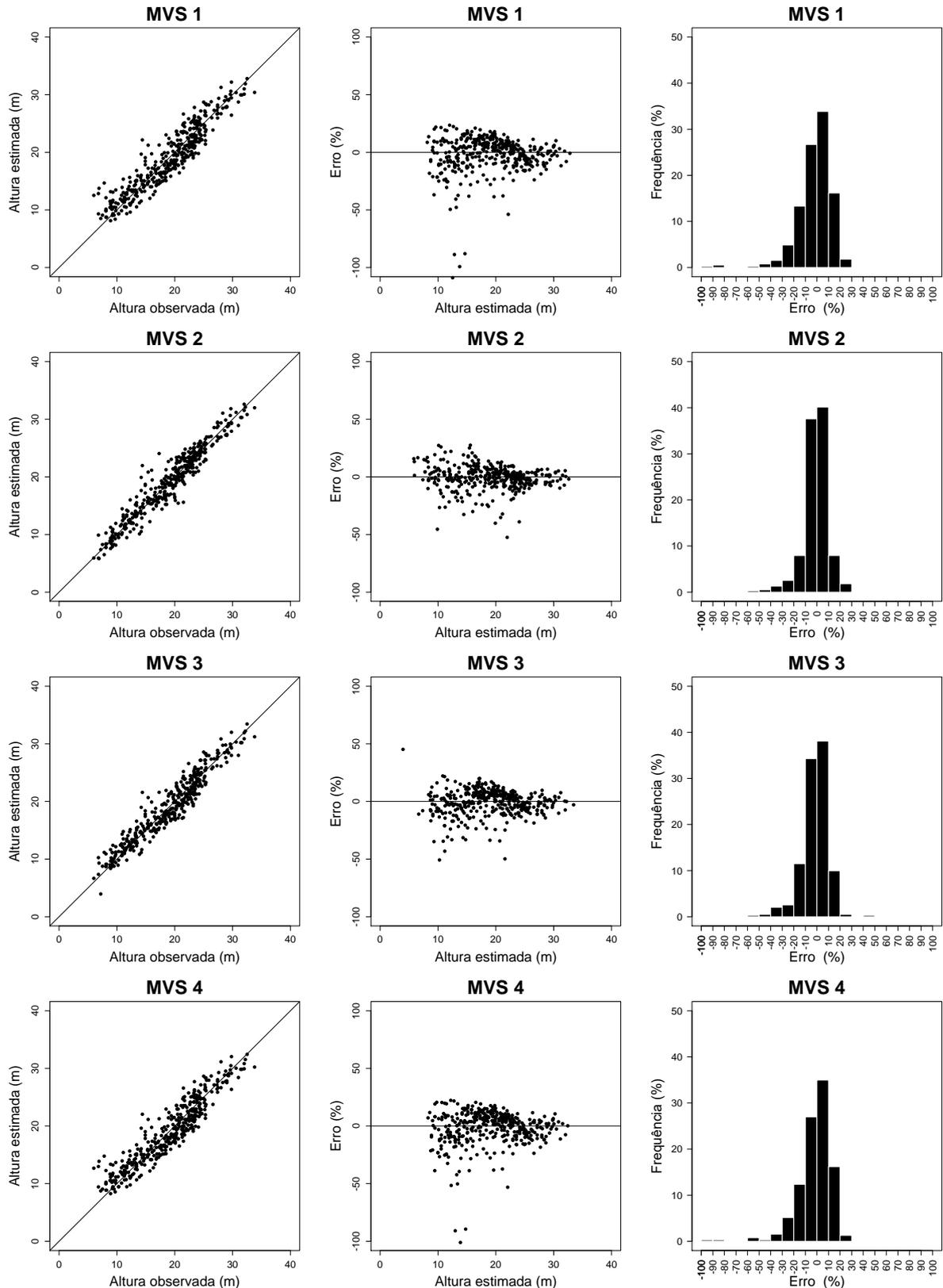
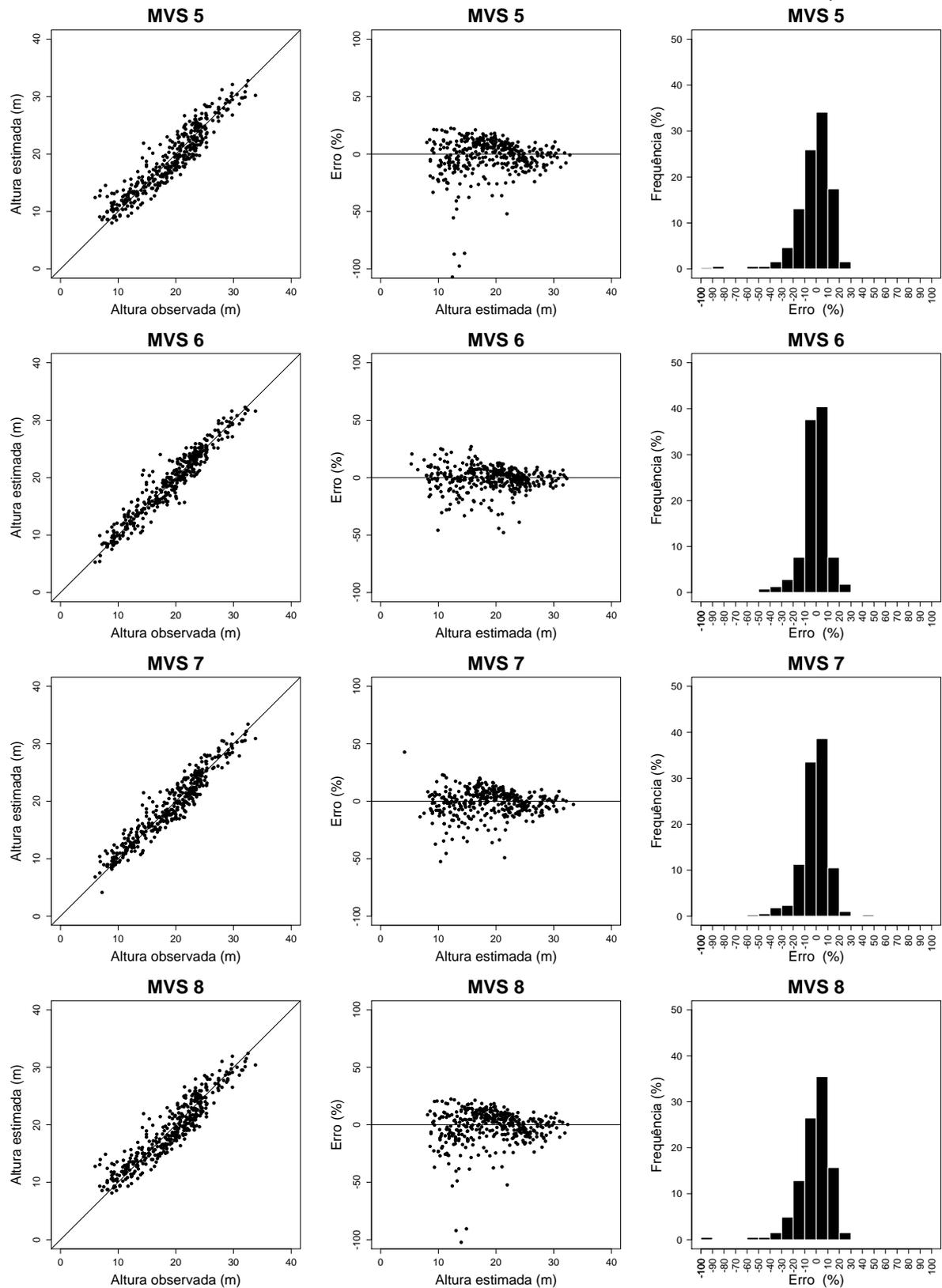


Figura 12 – Alturas estimadas versus alturas observadas, distribuição dos erros percentuais das estimativas de altura e histograma de erros das estimativas de altura, obtidas por Máquina de Vetor de Suporte (MVS), para os dados de validação.
(Conclusão)



Fonte: o autor

Com base na Figura 12, nota-se que as MVSs 2 e 6 apresentaram menor dispersão dos erros, com maior concentração de erros nas classes de $\pm 10\%$. Isso indica que é possível utilizar essa técnica para generalizar a relação hipsométrica, mesmo com diferentes tipos de clone, e obter resultados melhores do que no uso de modelos clássicos de regressão.

Segundo Couto e Bastos (1987), em situações em que há diferentes microclimas, a capacidade de generalização de modelos hipsométricos pode ser reduzida. E como solução para o inventário florestal, os autores relatam que é aconselhável gerar equações hipsométricas para cada estrato por parcela. Ou ainda, pode-se utilizar modelos com covariáveis ou modelos genéricos, tendo em vista que a altura é dependente também da capacidade produtiva do local e de outras variáveis como o material genético (CAMPOS; LEITE, 2013).

Em relação aos modelos de regressão utilizados nesse trabalho, uma vantagem no uso de técnicas de inteligência artificial é o fato de haver maior facilidade na modelagem de variáveis qualitativas, como o material genético, tipo de solo e a qualidade do fuste. Segundo Binoti, Binoti, Leite (2013), nesses modelos, esse tipo de variável pode ser incluído apenas como variável binária e ocasionam maior complexidade na modelagem. Com isso, fica evidente a vantagem de se utilizar técnicas de inteligência artificial tendo em vista maior exatidão nas estimativas de altura total geradas pela MVS, tanto para os dados de ajuste e treinamento, como para os dados de validação.

5.2 Resultados para estimação do volume total de árvores de *Eucalyptus*

5.2.1 Modelos Volumétricos

Os coeficientes estimados e as estatísticas para avaliar os modelos volumétricos utilizados estão apresentados na Tabela 13. Todos os parâmetros foram significativos ($p < 0,01$) pelo teste t , demonstrando que as variáveis utilizadas foram relevantes para explicar as variações do volume.

Tabela 13 – Coeficientes estimados e estatísticas utilizadas para avaliar o desempenho do ajuste dos modelos volumétricos na estimação do volume individual de árvores de *Eucalyptus grandis* x *Eucalyptus urophylla*

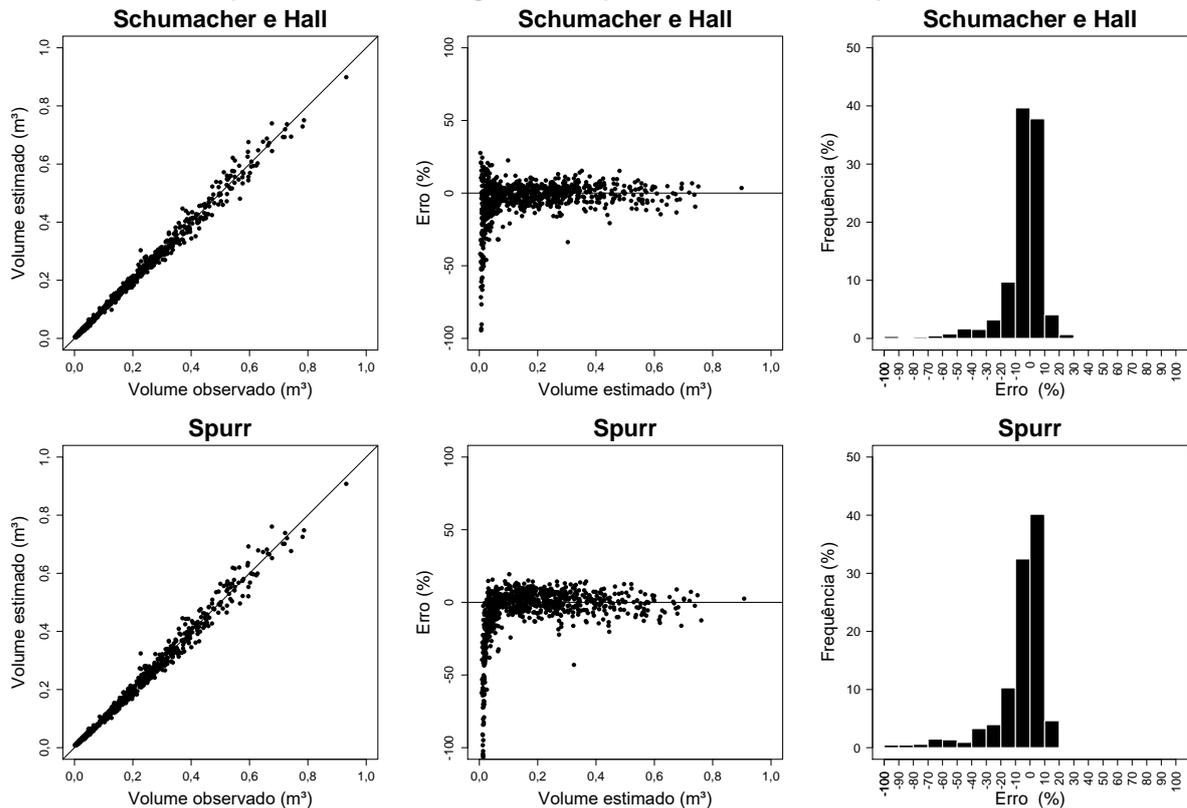
Modelo	Coeficiente			$r_{y\hat{y}}$	V (%)	RQME (%)
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$			
Schumacher e Hall	0,00003605**	1,730**	1,241**	0,9961	-0,27	8,29
Spurr	0,004613**	0,0000341**	-	0,9951	0,00	9,34

em que: $\hat{\beta}_k$ = estimativa dos parâmetros; $r_{y\hat{y}}$ = coeficiente de correlação entre os volumes observados e estimados; V (%) = viés; RQME (%) = raiz do erro quadrático médio; ** = significativo a 1% pelo teste *t*.

Os modelos analisados apresentaram valores semelhantes para o coeficiente de correlação. O modelo de Schumacher e Hall foi o que apresentou estimativas de maior exatidão quando se avalia o valor de RQME. Segundo Campos e Leite (2013), o modelo de Schumacher e Hall tem sido mais difundido por resultar quase sempre em estimativas não tendenciosas. Como já observado em outros estudos, o modelo de Schumacher e Hall apresenta maior exatidão na estimação do volume individual de árvores de *Eucalyptus*, conforme pode ser constatado nos trabalhos de Azevedo et al. (2011), Leal et al. (2015), Müller et al. (2014).

A distribuição dos valores observados e estimados, distribuição dos erros percentuais e histogramas dos erros foram semelhantes para os dois modelos utilizados, com superestimação dos menores volumes (Figura 13).

Figura 13 – Volumes estimados versus volumes observados, distribuição dos erros percentuais das estimativas de volume e histograma de erros das estimativas de volume, obtidos por análise de regressão, para os dados de ajuste.



Fonte: o autor

Com base no histograma dos erros percentuais, nota-se que modelo de Schumacher e Hall apresentou maior porcentagem de erros nas classes de $\pm 10\%$, corroborando os resultados encontrados para as estatísticas de ajuste.

Na Tabela 14 estão apresentados os resultados das estatísticas dos modelos volumétricos para os dados de validação.

Tabela 14 – Estatísticas utilizadas para avaliar o desempenho dos modelos volumétricos na estimação do volume individual de árvores de *Eucalyptus grandis* x *Eucalyptus urophylla*, para os dados de validação

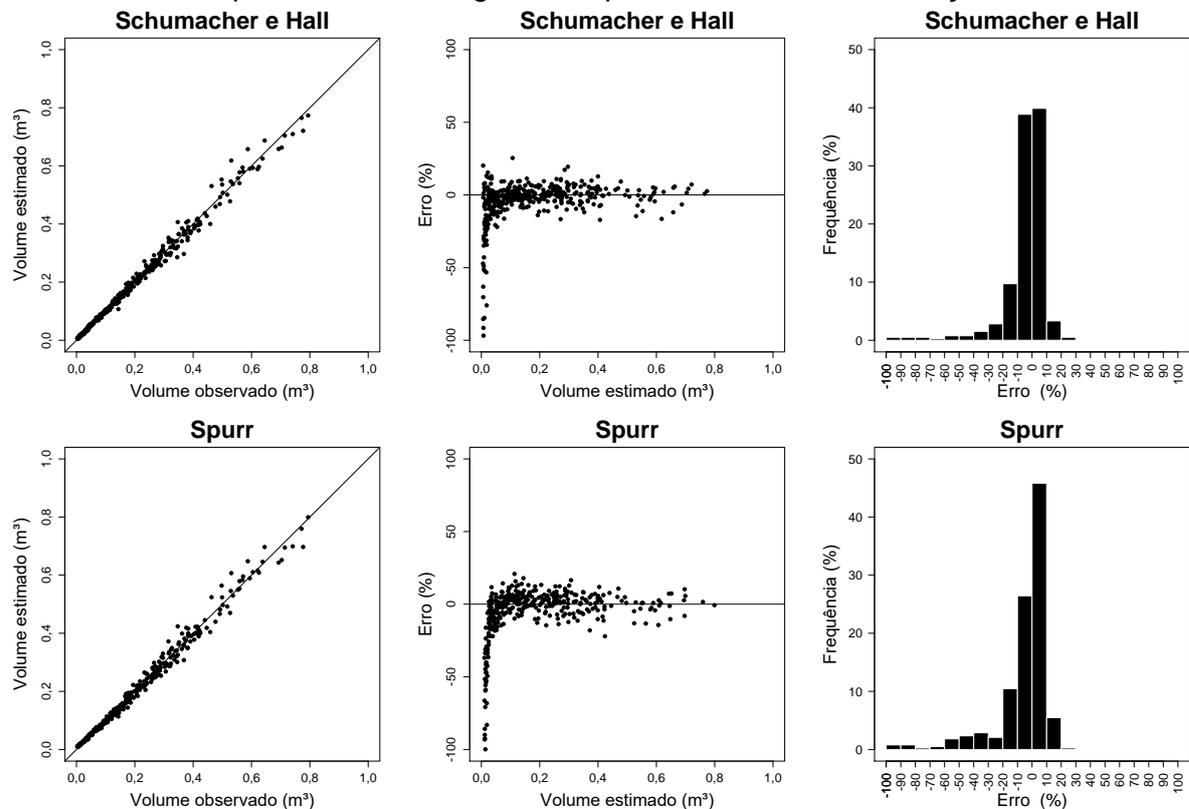
Modelo	$r_{\hat{y}y}$	V (%)	RQME (%)
Schumacher e Hall	0,9959	0,31	8,34
Spurr	0,9951	0,71	9,14

em que: $r_{\hat{y}y}$ = coeficiente de correlação entre os volumes observados e estimados; V (%) = viés; RQME (%) = raiz do erro quadrático médio.

Nota-se generalização de maior exatidão para o modelo de Schumacher e Hall, superando o modelo de Spurr em todas as estatísticas avaliadas. Conforme

pode ser observado na Figura 14, os dois modelos mantiveram a tendência em superestimar os menores volumes, com maior uniformidade nos erros para o modelo de Schumacher e Hall, como apresentado nos dados de ajuste (Figura 13).

Figura 14 – Volumes estimados versus volumes observados, distribuição dos erros percentuais das estimativas de volume e histograma de erros das estimativas de volume, obtidos por análise de regressão, para os dados de validação.



Fonte: o autor

5.2.2 Máquina de Vetor de Suporte

Na Tabela 15 estão apresentadas as estatísticas de avaliação da MVS para os dados de treinamento. As configurações da MVS que fizeram uso das funções de *kernel* linear e sigmoidal não se ajustaram aos dados, gerando estimativas negativas de volume, não sendo incluídas nos resultados. Isso pode ter ocorrido pelo fato de que esses *kernels* não conseguiram extrair a relação entre as variáveis de entrada e a variável de saída.

Tabela 15 – Estatísticas utilizadas para avaliar o desempenho da Máquina de Vetor de Suporte (MVS) na estimação do volume individual de árvores *Eucalyptus grandis* x *Eucalyptus urophylla*, para os dados de treinamento

MVS	$r_{\hat{y}y}$	V (%)	RQME (%)
2	0,9968	-0,11	7,60
3	0,9966	0,44	7,78
6	0,9967	-0,32	7,68
7	0,9969	0,10	7,43

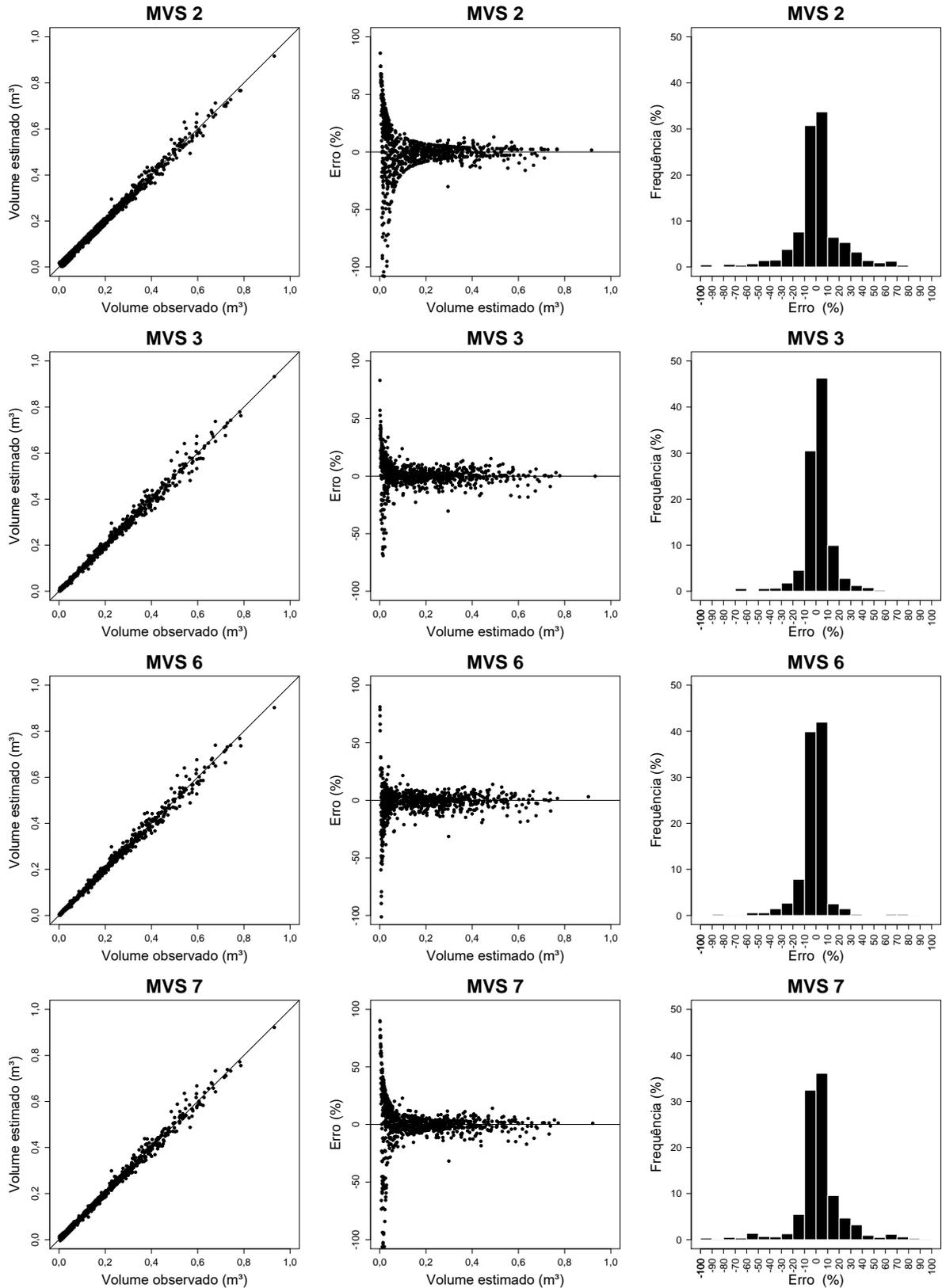
em que: $r_{\hat{y}y}$ = coeficiente de correlação entre os volumes observados e estimados; V (%) = viés; RQME (%) = raiz do erro quadrático médio.

Ao observar as estatísticas apresentadas na Tabela 15, nota-se que a MVS 7 foi a que apresentou maior exatidão nas estimativas, seguida pela MVS 2. A MVS 7 apresentou um ganho considerável em exatidão em relação ao modelo de Schumacher e Hall, diminuindo a RQME em aproximadamente 10%.

Resultados de maior exatidão utilizando a MVS para a estimação do volume também foram encontrados por Binoti et al. (2016) e Cordeiro et al. (2015). Esses autores concluíram que a MVS pode ser utilizada para a predição volumétrica de espécies do gênero *Eucalyptus* e *Acacia mangium*, respectivamente. Em ambos os trabalhos, a MVS conseguiu maior exatidão nas estimativas, superando o modelo de Schumacher e Hall. Vale ressaltar que no trabalho desses autores foi utilizada apenas uma configuração da MVS, composta pela função do tipo II e pela função de *kernel* RBF.

Na Figura 15, em que estão ilustrados os gráficos de análise dos erros para os dados de treinamento, é possível notar que houve tendência de apresentar erros com valores maiores para árvores de menor volume em todas as configurações testadas.

Figura 15 – Volumes estimados versus volumes observados, distribuição dos erros percentuais das estimativas de volume e histograma de erros das estimativas de volume, obtidos por Máquina de Vetor de Suporte (MVS), para os dados de treinamento.



Fonte: o autor

Apesar de a MVS 7 apresentar os resultados mais exatos nas estatísticas de avaliação, na análise dos erros é possível observar, conforme demonstrado no histograma dos erros percentuais, que essa configuração teve menor concentração dos erros nas classes de $\pm 10\%$ em comparação com a MVS 6. Portanto, a MVS 6 é mais recomendada por apresentar maior uniformidade na distribuição dos erros.

García Nieto et al. (2012), visando estimar o volume sem casca de árvores de *Eucalyptus globulus*, confrontaram duas técnicas de inteligência artificial como alternativa aos modelos estatísticos convencionais de dupla entrada. Esses autores concluíram que a MVS foi a melhor técnica para estimar o volume sem casca para os dados em questão em relação às RNAs e aos modelos de regressão. No trabalho desses autores, foram testadas três funções de *kernel* (Polinomial, sigmoidal e RBF), com melhores estimativas geradas pela função de *kernel* RBF.

O fato de a MVS 6 ter apresentado maior concentração de erros nas classes de $\pm 10\%$ pode ter contribuído para o seu melhor desempenho de generalização, como pode ser observado na Tabela 16.

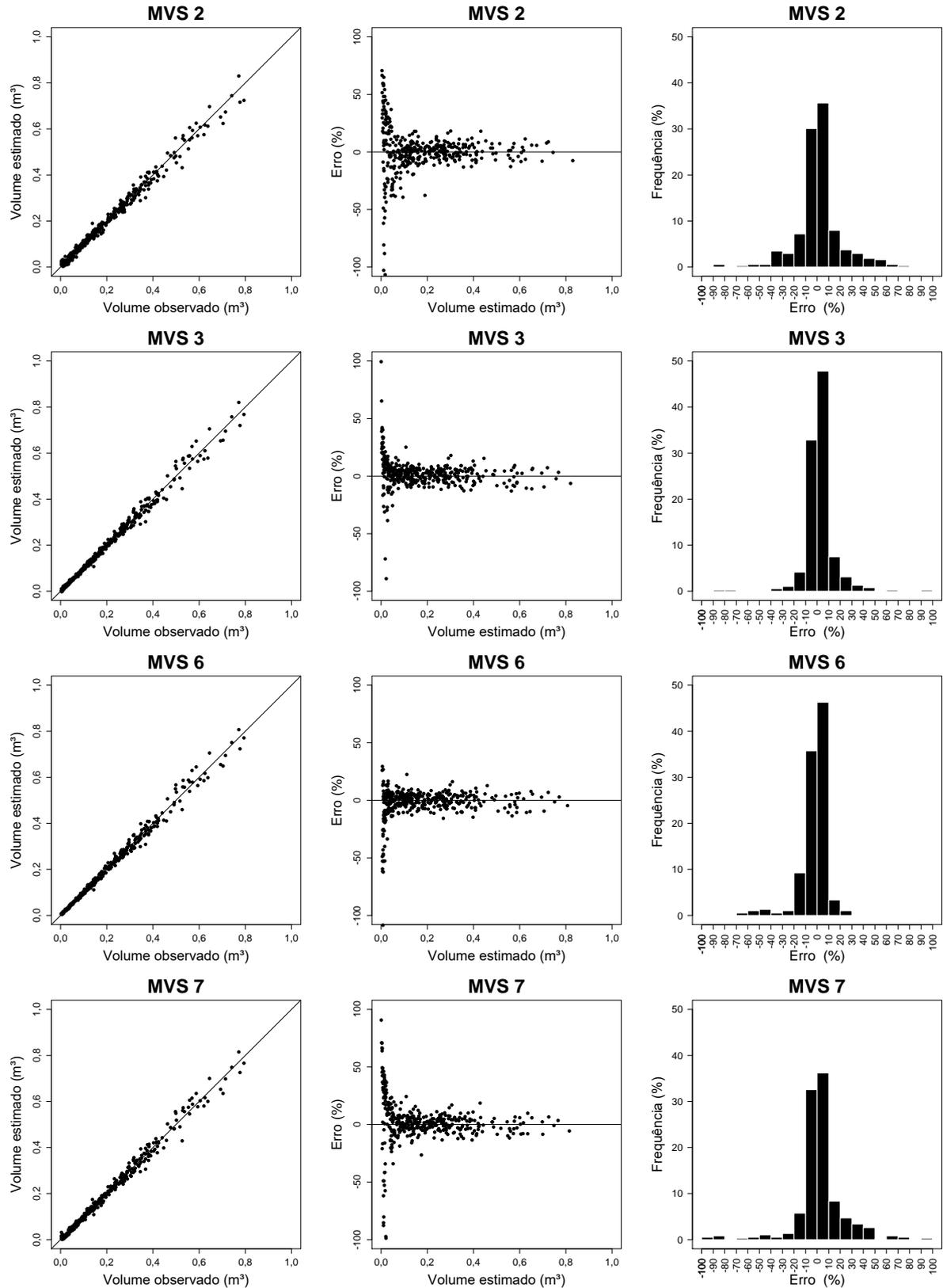
Tabela 16 – Estatísticas utilizadas para avaliar o desempenho da Máquina de Vetor de Suporte (MVS) na estimação do volume individual de árvores de *Eucalyptus grandis* x *Eucalyptus urophylla*, para os dados de validação

MVS	$r_{\hat{y}y}$	V (%)	RQME (%)
2	0,9943	0,76	9,88
3	0,9954	0,87	8,86
6	0,9960	0,01	8,31
7	0,9953	0,65	8,86

em que: $r_{\hat{y}y}$ = coeficiente de correlação entre os volumes observados e estimados; V (%) = viés; RQME (%) = raiz do erro quadrático médio.

A MVS 6, formada a partir da função do tipo II e da função de *kernel* RBF, obteve maior exatidão nas estimativas de volume para os dados de validação. Isso demonstra que essa configuração foi mais eficiente em extrair características importantes dos dados em estudo, gerando um modelo com maior exatidão na generalização. A maior eficiência dessa configuração para os dados de validação pode ser corroborada pela análise da distribuição dos erros (Figura 16).

Figura 16 – Volumes estimados versus volumes observados, distribuição dos erros percentuais das estimativas de volume e histograma de erros das estimativas de volume, obtidos por Máquina de Vetor de Suporte (MVS), para os dados de validação.



Fonte: o autor

Com base na análise da distribuição dos erros, a MVS 6 manteve estimativas mais homogêneas entre as configurações testadas, no entanto todas as configurações apresentaram maiores erros para as árvores menores. Isso demonstra a importância de se testar diferentes configurações da MVS, combinando as variantes existentes (função do tipo I e II) com diferentes funções de *kernel*, buscando aquela que melhor se adeque a determinado problema e que obtenha maior exatidão na generalização.

A escolha da função de *kernel* é fundamental para que a MVS consiga modelar a não linearidade existente na relação entre diversas variáveis com o volume. Segundo Binoti, Binoti e Leite (2014), variáveis como espaçamento, classe de solo, forma do tronco e material genético, geralmente apresentam relações complexas e não lineares com o volume individual de árvores e são difíceis de inserir em modelos clássicos de regressão, sendo necessárias técnicas que obtenham maior facilidade de relacionar esse tipo de variável.

De forma geral, a MVS foi capaz de gerar estimativas de maior exatidão para o volume individual de híbridos de *Eucalyptus grandis* x *Eucalyptus urophylla* em relação a modelos de regressão comumente utilizados. Como também de compreender as relações existentes entre as variáveis utilizadas e gerar estimativas de maior exatidão para dados que não tiveram participação no seu treinamento.

5.3 Resultados para estimação do diâmetro ao longo do fuste de árvores de *Eucalyptus*

5.3.1 Modelos de Afilamento

Na Tabela 17 estão apresentados os coeficientes estimados para os modelos de afilamento avaliados e na Tabela 18 estão apresentadas as estatísticas avaliadas.

Tabela 17 – Coeficientes estimados obtidos a partir do ajuste dos modelos de afilamento para árvores de *Eucalyptus grandis* x *Eucalyptus urophylla*

Modelo	Coeficiente					
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$
Demaerschalk	0,07373**	0,87718**	-0.63497**	0,71102**	-	-
Ormerod	-	0.70020**	-	-	-	-
Schöepfer	1.17472**	-3.24927**	-4.07564**	9.24176**	-9.37755**	3.57893**

em que: $\hat{\beta}_k$ = estimativa dos parâmetros; ** = significativo a 1% pelo teste *t*.

Tabela 18 – Estatísticas utilizadas para avaliar o desempenho do ajuste dos modelos de afilamento na estimação do diâmetro ao longo do fuste de árvores de *Eucalyptus grandis* x *Eucalyptus urophylla*

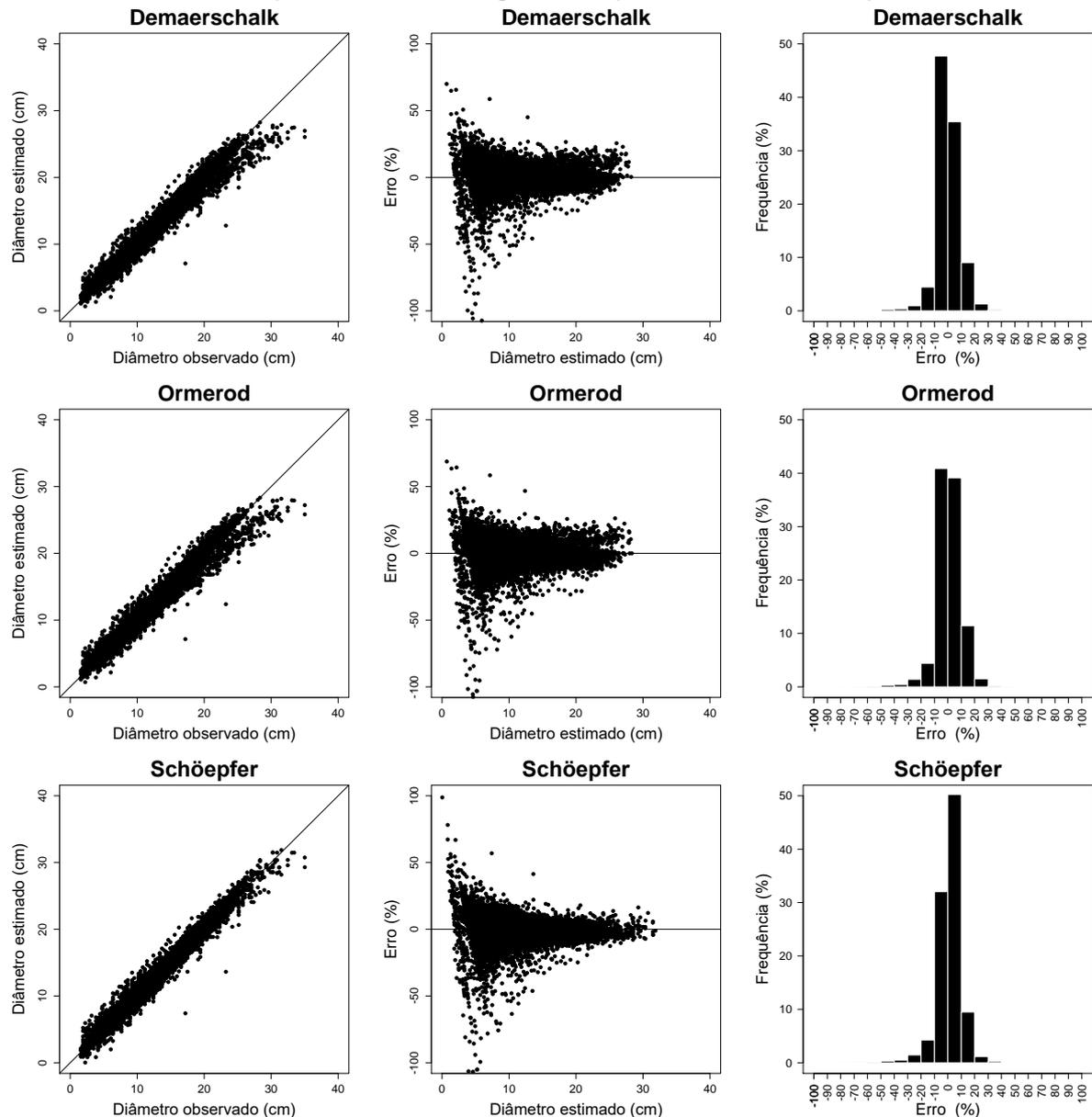
Modelo	$r_{y\hat{y}}$	V (%)	RQME (%)
Demaerschalk	0,9868	0,16	7,69
Ormerod	0,9860	0,97	8,01
Schöepfer	0,9907	0,76	6,70

em que: $r_{y\hat{y}}$ = coeficiente de correlação entre os diâmetros observados e estimados; V (%) = viés; RQME (%) = raiz do erro quadrático médio.

Nota-se que todas as estimativas dos parâmetros foram significativas ($p < 0,01$) pelo teste *t*. É possível observar que o modelo de Schöepfer foi o de maior capacidade em estimar o diâmetro ao longo do fuste de árvores de eucalipto, com menor valor de RQME e com estimativas de maior correlação com os diâmetros observados em relação ao modelo de Demaerschalk e de Ormerod.

Na Figura 17 estão ilustrados os gráficos referentes às análises dos erros para os dados de ajuste.

Figura 17 – Diâmetros estimados versus diâmetros observados, distribuição dos erros percentuais das estimativas de diâmetro e histograma de erros das estimativas de diâmetro, obtidos por análise de regressão, para os dados de ajuste.



Fonte: o autor

Como pode ser observado, os modelos analisados apresentaram erros distribuídos de forma semelhante, com maiores erros nas estimativas de diâmetro da parte superior da árvore (0 a 10 cm). O modelo de Schöpfer foi o que apresentou menor dispersão dos erros, o que é confirmado pela maior porcentagem de erros nas classes de $\pm 10\%$ observada no histograma de erros percentuais.

Na Tabela 19 estão apresentadas as estatísticas dos modelos de afilamento para os dados de validação.

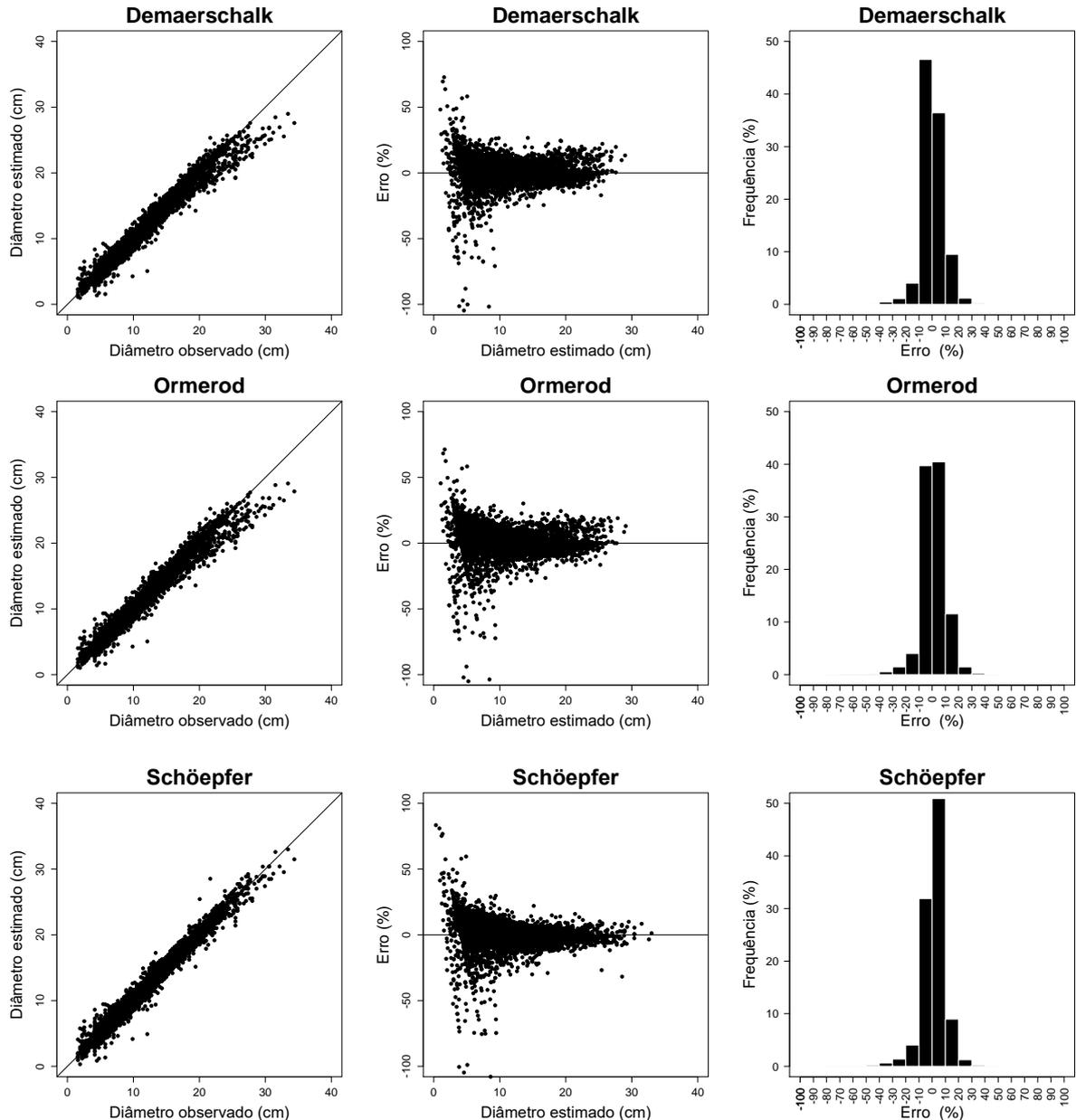
Tabela 19 – Estatísticas utilizadas para avaliar o desempenho dos modelos de afilamento na estimação do diâmetro ao longo do fuste de árvores de *Eucalyptus grandis* x *Eucalyptus urophylla*, para os dados de validação

Modelo	$r_{\hat{y}y}$	V (%)	RQME (%)
Demaerschalk	0,9862	0,38	7,84
Ormerod	0,9853	1,15	8,16
Schöepfer	0,9902	0,72	6,80

em que: $r_{\hat{y}y}$ = coeficiente de correlação entre os diâmetros observados e estimados; V (%) = viés; RQME (%) = raiz do erro quadrático médio.

Observa-se que o modelo de Schöepfer apresentou maior exatidão na generalização entre os modelos avaliados. Comparando com as estatísticas avaliadas para o ajuste, não houve perda de exatidão para as estimativas geradas pelo modelo de Schöepfer a partir dos dados de validação (Tabela 18). Os modelos analisados obtiveram distribuição dos erros semelhante à encontrada para os dados de ajuste (Figura 17), como pode ser observado na Figura 18.

Figura 18 – Diâmetros estimados versus diâmetros observados, distribuição dos erros percentuais das estimativas de diâmetro e histograma de erros das estimativas de diâmetro, obtidos por análise de regressão, para os dados de validação.



Fonte: o autor

Nota-se que para diâmetros maiores que 10 cm, como também foi observado nos dados de ajuste (Figura 17), os modelos apresentaram estimativas mais uniformes, evidenciando a dificuldade desses modelos de regressão em gerar estimativas de maior exatidão para os menores diâmetros.

Miguel et al. (2011), avaliando modelos de afilamento na descrição do fuste de *Eucalyptus urophylla*, também concluíram que o modelo de Schöpfer foi o de maior exatidão juntamente com um modelo de potências inteiras e fracionárias,

podendo ser utilizado para representar o perfil de árvores dessa espécie. No entanto, deve-se ter cautela ao se utilizar o modelo de Schöepfer, pois, ao aumentar o grau de um polinômio, tem-se melhoria no ajuste sob o ponto de vista estatístico, mas é possível que haja perda da consistência biológica do modelo (MURTA JÚNIOR et al., 2015).

5.3.2 Máquina de Vetor de Suporte

Na Tabela 20 estão apresentadas as estatísticas de avaliação da MVS na descrição do diâmetro ao longo do fuste de árvores de *Eucalyptus grandis* x *Eucalyptus urophylla*. A MVS 4, formada a partir da função do tipo I e da função de *kernel* sigmoidal, não se ajustou aos dados, portanto seus resultados não foram apresentados.

Tabela 20 – Estatísticas utilizadas para avaliar o desempenho da Máquina de Vetor de Suporte (MVS) na estimação do diâmetro ao longo do fuste de árvores de *Eucalyptus grandis* x *Eucalyptus urophylla*, para os dados de treinamento

MVS	$r_{\hat{y}y}$	V (%)	RQME (%)
1	0,9757	0,72	10,45
2	0,9945	0,01	4,97
3	0,9879	0,53	7,43
5	0,9763	0,48	10,29
6	0,9947	0,01	4,87
7	0,9879	0,38	7,38
8	0,9763	0,22	10,31

em que: $r_{\hat{y}y}$ = coeficiente de correlação entre os diâmetros observados e estimados; V (%) = viés; RQME (%) = raiz do erro quadrático médio.

Com exceção das MVSs 2 e 6, todas as configurações apresentaram resultados inferiores nas estatísticas avaliadas quando comparadas com o modelo de Schöepfer. As MVSs 1, 5 e 8, formadas a partir das funções de *kernel* linear e sigmoidal com as funções do tipo I e II, apresentaram resultados inferiores aos de todos os modelos de regressão avaliados, o que evidencia menor eficiência dessas funções de *kernel* em descrever o diâmetro do fuste de árvores de eucalipto.

As MVSs 2 e 6 apresentaram valores semelhantes das estatísticas avaliadas e tiveram ganho em exatidão de aproximadamente 30% em relação ao modelo de Schöepfer. É possível inferir que a MVS, quando utiliza a função de *kernel* RBF, é mais eficaz em aprender, a partir dos exemplos fornecidos, a relação entre as

variáveis utilizadas. Isso também ocorreu para a estimativa da altura total e do volume individual.

Na Figura 19 estão ilustrados os gráficos de distribuição dos erros para as configurações de MVS utilizadas, com exceção da MVS 4.

Figura 19 – Diâmetros estimados versus diâmetros observados, distribuição dos erros percentuais das estimativas de diâmetro e histograma de erros das estimativas de diâmetro, obtidos por Máquina de Vetor de Suporte (MVS), para os dados de treinamento.

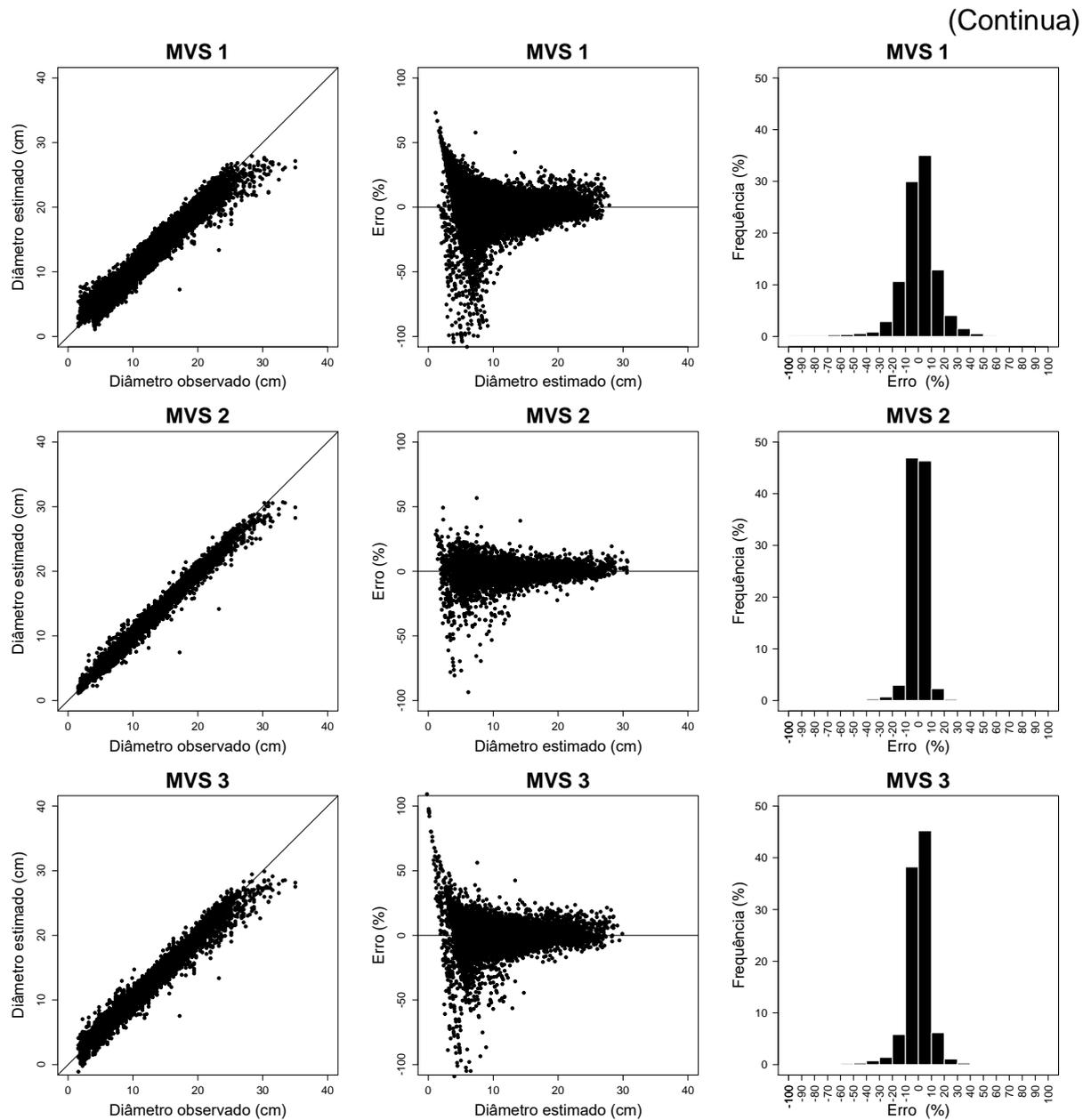
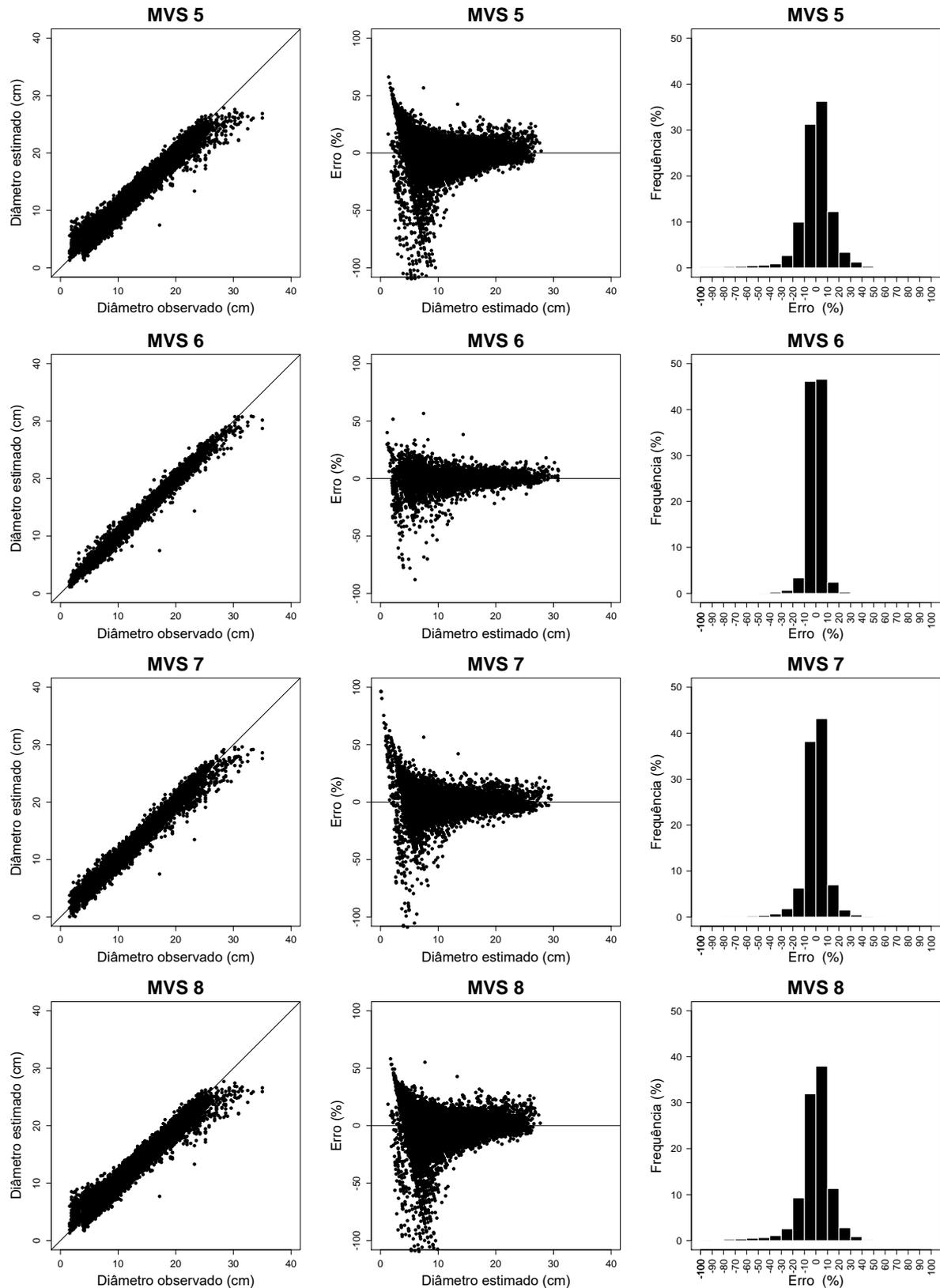


Figura 19 – Diâmetros estimados versus diâmetros observados, distribuição dos erros percentuais das estimativas de diâmetro e histograma de erros das estimativas de diâmetro, obtidos por Máquina de Vetor de Suporte (MVS), para os dados de treinamento.

(Conclusão)



Fonte: o autor

Ao se analisar a Figura 19, fica ainda mais evidente que as MVSs 2 e 6 geraram estimativas de maior exatidão em relação ao modelo de Schöpfer. Essa técnica conseguiu diminuir a dispersão dos erros ao longo de toda a amplitude de dados. Vale ressaltar que nas estimativas geradas pelas MVSs 2 e 6 também houve maiores erros nos diâmetros menores que 10 cm. Com base no histograma dos erros percentuais, nota-se que 90% dos erros estão agrupados nas classes de $\pm 10\%$, enquanto que para o modelo de Schöpfer, essa proporção foi de aproximadamente 80%.

Na Tabela 21 estão apresentadas as estatísticas avaliadas das configurações de MVS aplicadas aos dados de validação.

Tabela 21– Estatísticas utilizadas para avaliar o desempenho da Máquina de Vetor de Suporte (MVS) na estimação do diâmetro ao longo do fuste de árvores de *Eucalyptus grandis* x *Eucalyptus urophylla*, para os dados de validação

MVS	$r_{\hat{y}y}$	V (%)	RQME (%)
1	0,9752	1,08	10,50
2	0,9934	0,13	5,43
3	0,9872	0,71	7,61
5	0,9758	0,80	10,35
6	0,9935	0,12	5,37
7	0,9872	0,56	7,56
8	0,9758	0,53	10,38

em que: $r_{\hat{y}y}$ = coeficiente de correlação entre os diâmetros observados e estimados; V (%) = viés; RQME (%) = raiz do erro quadrático médio.

Pode ser observado que as MVSs 2 e 6 mantiveram-se como as configurações de maior exatidão também para os dados de validação, evidenciando que essas configurações foram mais eficientes em extrair características importantes dos dados de treinamento e que não houve *overfitting* durante o treinamento. Para os dados de validação, o ganho de exatidão das MVSs 2 e 6 em relação ao modelo de Schöpfer foi de aproximadamente 23%. Na Figura 20 estão ilustrados os gráficos de distribuição dos erros da MVS para os dados de validação.

Figura 20 – Diâmetros estimados versus diâmetros observados, distribuição dos erros percentuais das estimativas de diâmetro e histograma de erros das estimativas de diâmetro, obtidos por Máquina de Vetor de Suporte (MVS), para os dados de validação.

(Continua)

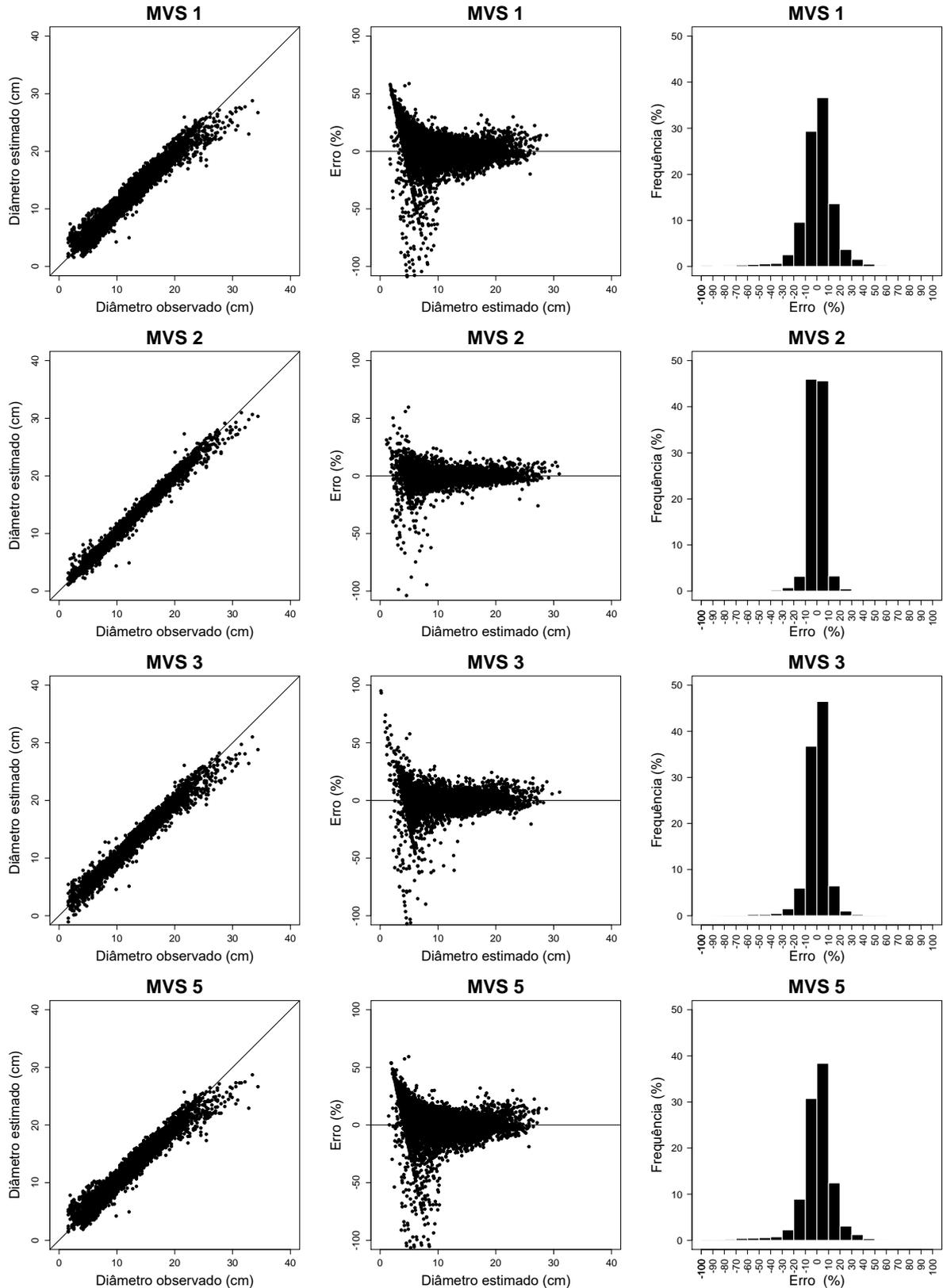
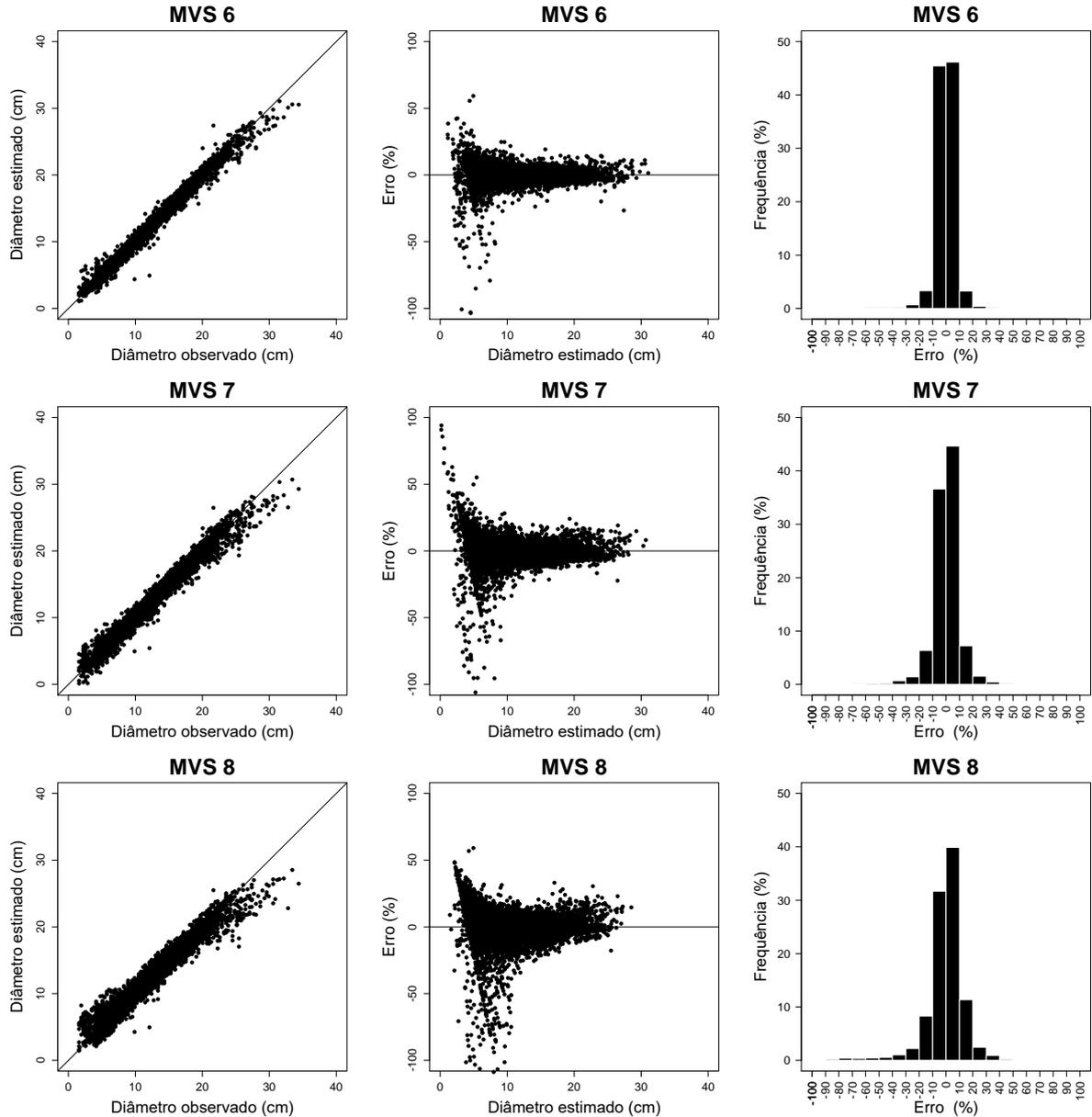


Figura 20 – Diâmetros estimados versus diâmetros observados, distribuição dos erros percentuais das estimativas de diâmetro e histograma de erros das estimativas de diâmetro, obtidos por Máquina de Vetor de Suporte (MVS), para os dados de validação.

(Conclusão)



Fonte: o autor

Nota-se que as MVSs 2 e 6 mantiveram distribuição dos erros mais uniforme também para os dados de validação, com aproximadamente 90% dos erros encontrados nas classes de $\pm 10\%$. A menor dispersão dos erros das estimativas geradas pela MVS ressalta sua capacidade de aprendizagem a partir de exemplos e de gerar estimativas de alta exatidão mesmo em dados que não tiveram participação no seu treinamento. O fato de a MVS se basear nos princípios da teoria do

aprendizado estatístico, que visa principalmente que o preditor gerado possua máxima capacidade de generalização, faz com que a mesma seja eficiente em compreender diversas relações entre as variáveis, conferindo grande poder preditivo.

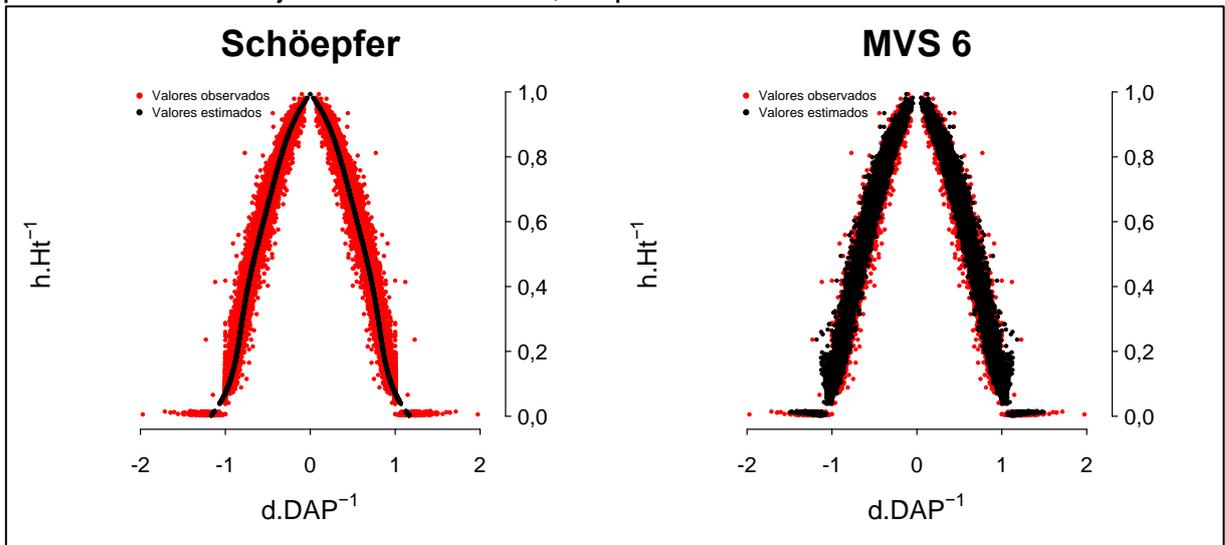
Com base na análise gráfica e nas estatísticas avaliadas para os dados de validação, a MVS pode ser considerada uma técnica promissora para aplicação em outros estudos no setor florestal, obtendo estimativas de alta exatidão, comparáveis às de RNAs, como pode ser observado no trabalho de Martins et al (2016). Esses autores encontraram RQME de aproximadamente 5,47% nas estimativas de diâmetro para o gênero *Eucalyptus*. Vale ressaltar que no trabalho desses autores foram utilizadas também como variáveis de entrada a idade e o projeto, o que contribui para melhor estratificação e estimativas de maior exatidão.

A forma do tronco de uma árvore tem sido objeto de muitos estudos pelo fato de essa variável ter relação com o volume de madeira. A forma do tronco varia da base até o topo da árvore, havendo uma relação de decréscimo do diâmetro denominada de *taper*. Essa relação de decréscimo pode variar de acordo com a espécie, idade das árvores e capacidade produtiva e é a razão fundamental para a variação no volume (HUSCH; BEERS; KERSHAW JUNIOR, 2003).

Fica evidente a importância em se utilizar métodos de estimação que descrevam o perfil do fuste de forma fidedigna, contribuindo para maior aproveitamento da madeira gerada, podendo destiná-la para multiprodutos. Segundo Pires e Calegario (2007), a avaliação dos perfis do fuste de árvores é uma forma de verificar a exatidão das estimativas geradas, confrontando com o valor real e permite verificar em qual posição o modelo é mais ou menos exato.

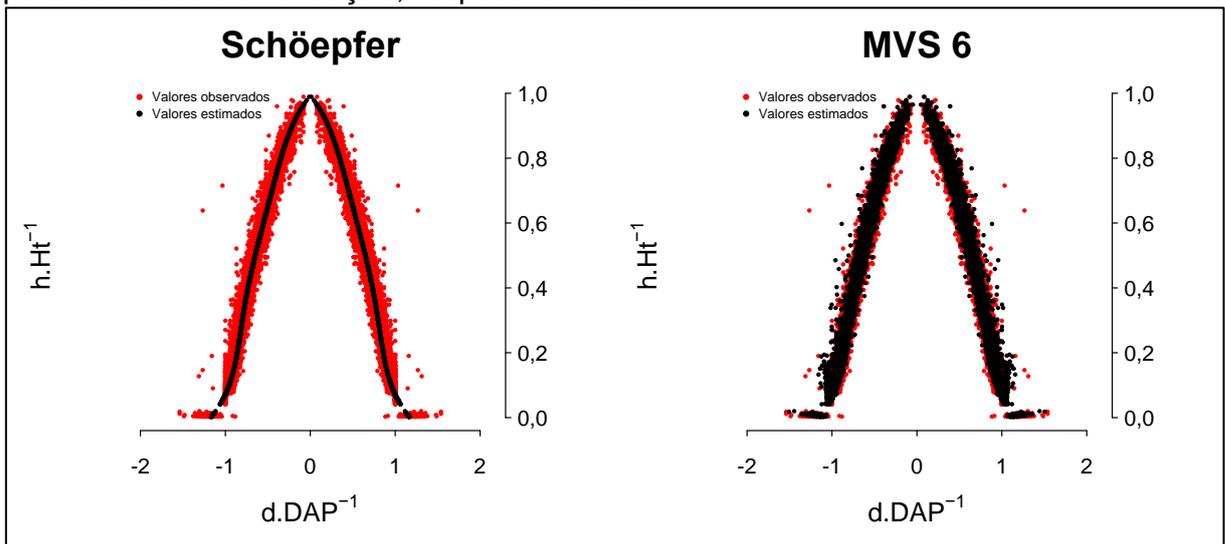
Na Figura 21 e 22 estão ilustrados os perfis observados e estimados pelo modelo de Schöepfer e pela MVS 6 para os dados de ajuste e treinamento e para os dados de validação, respectivamente.

Figura 21 – Perfil real do fuste (valores observados) e perfil estimado (valores estimados) pelo modelo de Ormerod e pela Máquina de Vetor de Suporte (MVS), para os dados de ajuste e treinamento, respectivamente.



Fonte: o autor

Figura 22 – Perfil real do fuste (valores observados) e perfil estimado (valores estimados) pelo modelo de Ormerod e pela Máquina de Vetor de Suporte (MVS), para os dados de validação, respectivamente.



Fonte: o autor

Tanto em relação aos dados de ajuste e treinamento como para os dados de validação, observa-se que a MVS 6 foi mais eficiente em descrever o diâmetro ao longo de todo o fuste, com estimativas mais próximas das medições observadas, o que corrobora a menor dispersão dos erros encontrada pela MVS 6 (Figuras 18 e 19).

O modelo de Schöepfer apresentou estimativas de menor exatidão, mais visíveis na altura da base das árvores. Isso pode ter ocorrido pelo fato de que modelos polinomiais não explicam com exatidão as deformações existentes na base da árvore (SCOLFORO, 2005). Portanto, fica evidente o potencial da MVS, sendo indicada para descrever o perfil do fuste de árvores de eucalipto.

5.4 Comparação das Configurações de Máquina de Vetor de Suporte Utilizadas nos Três Objetivos de Modelagem

Para a variável altura total, foi possível notar que houve semelhança nos valores de RQME para todas as configurações de MVS, quando se compara a sua exatidão no conjunto de treinamento e de validação (Tabelas 11 e 12). Nos dois conjuntos de dados, é possível observar que as funções do tipo I e do tipo II, combinadas com as mesmas funções de *kernel*, geraram resultados semelhantes para as estatísticas avaliadas. Isso se torna mais evidente no conjunto de validação.

A semelhança entre os valores das estatísticas avaliadas no conjunto de treinamento também se repetiu para a estimação do volume individual. No entanto, houve maior diferença entre as funções de erro utilizadas quando se avaliou o conjunto de validação. A MVS 2, embora tenha apresentado maior exatidão para o conjunto de treinamento, foi mais viesada que a MVS 6 quando se avaliou o conjunto de validação (Tabelas 15 e 16), o que pode evidenciar maior capacidade de generalização quando se combina a função de *kernel* RBF com a função de erro do tipo II.

Para o afilamento do fuste de árvores de *Eucalyptus grandis* x *Eucalyptus urophylla*, houve maior exatidão nas estimativas geradas pelas MVSs 2 e 6 e semelhança nas estatísticas avaliadas entre essas duas configurações. Essa semelhança das estatísticas avaliadas entre as duas funções de erro utilizadas também ocorreu para as outras funções de *kernel* tanto para os dados de treinamento como para os dados de validação (Tabelas 20 e 21).

Com base nas estimativas obtidas para as variáveis altura total, volume individual e afilamento do fuste de árvores de *Eucalyptus grandis* x *Eucalyptus urophylla*, é possível inferir que a exatidão das configurações avaliadas está mais relacionada com a função de *kernel* utilizada do que com as duas funções de erro avaliadas (função do tipo I e tipo II).

A função de *kernel* é fundamental para que a MVS consiga aprender relações não lineares entre variáveis. Üstün, Melssen e Buydens (2007) ressaltam que avaliar diferentes funções de *kernel* é necessário, visto que durante o mapeamento dos dados é importante que não haja perda de informações sobre a propriedade que se deseja estimar, o que pode comprometer a generalização do preditor gerado.

A altura total, o volume individual e afilamento do fuste sofrem influência de diferentes variáveis qualitativas como o espaçamento e o material genético. Variáveis qualitativas geralmente apresentam relação não linear com as variáveis dendrométricas e isso torna a função de *kernel* fundamental para relacionar esse tipo de variável.

As estimativas de pior desempenho foram geradas pelas configurações compostas pelas funções de *kernel* linear e sigmoidal. Vale ressaltar que as essas duas funções geraram estimativas negativas para o volume individual. Na função de *kernel* linear, o produto interno entre os vetores de entrada é calculado diretamente, sem a adição de não linearidades aos pesos gerados, o que implica em um modelo mais difícil de tratar problemas em que exista relação não linear entre as variáveis. Isso pode explicar o menor desempenho das configurações compostas por essa função, sendo mais recomendada quando não há necessidade de gerar uma MVS não linear.

Uma característica importante que uma função de *kernel* deve possuir é satisfazer condições do teorema de Mercer. De forma geral, um kernel satisfaz as condições do teorema de Mercer se der origem à matriz positiva semidefinida (LORENA; CARVALHO, 2007). Mais detalhes sobre o teorema de Mercer podem ser encontrados em Burges (1998). A função de *kernel* sigmoidal, apesar de ser amplamente utilizada, não satisfaz esse teorema para alguns valores de seus parâmetros, o que pode ocasionar em um preditor de baixo desempenho, podendo corroborar estatísticas inferiores encontradas para as configurações compostas por essa função de *kernel* (AMAMI; AYED; ELLOUZE, 2013).

O baixo desempenho da função de *kernel* sigmoidal já foi constatado por Ma, Zhang e Wang (2015). Esses autores, avaliando o desempenho de diversas funções de *kernel*, encontraram a função sigmoidal como a de pior desempenho. Lin e Lin (2003), em seu trabalho, não recomendam a utilização da função de *kernel* sigmoidal, pelo fato de haver dificuldade de solução mínima local para os parâmetros dessa função.

Para a estimação da altura, do volume individual e do afilamento do fuste de árvores de *Eucalyptus grandis* x *Eucalyptus urophylla*, as configurações compostas pela função de *kernel* polinomial apresentaram estatísticas de maior exatidão em relação às estatísticas das funções de *kernel* linear e sigmoideal para os dados de treinamento e validação. A função de *kernel* polinomial possui três parâmetros (Tabela 2), o que causa maior dificuldade no treinamento da MVS. Isso faz com que essa função de *kernel* seja menos utilizada que a RBF e pode estar relacionado com o fato de que, sob treinamento semelhante e custo de teste, uma função de *kernel* polinomial pode não gerar um preditor com maior exatidão, fato esse que ocorreu para as variáveis analisadas nesse trabalho (HSIEH; CHANG, 2010).

Considerando o conjunto de treinamento e de validação, houve maior exatidão das configurações compostas pela função de *kernel* RBF. Para esse estudo, outra vantagem da função de *kernel* RBF, evidenciada nas estimativas de volume geradas para os dados de treinamento, é que a MVS 6, embora com RQME maior do que a MVS 7, apresentou erros mais concentrados nas amplitudes de $\pm 10\%$ (Figura 15), o que pode ter contribuído para maior exatidão da MVS 6 para os dados de validação (Tabela 16).

Com base nos resultados de treinamento e validação, é possível inferir que a combinação das duas funções de erro com a função de *kernel* RBF gerou as configurações mais eficazes em extrair características importantes das variáveis de entrada e saída e, conseqüentemente, com maior exatidão na generalização.

6 CONCLUSÕES

O modelo de Gompertz apresentou o melhor ajuste para a estimação da altura de árvores de eucalipto. O material genético exerceu influência na variável altura total. Os modelos de Schumacher e Hall e de Schöepfer foram os de maior exatidão nas estimativas de volume individual e afilamento do fuste, respectivamente.

As funções de erro testadas apresentaram resultados estatísticos semelhantes quando utilizadas com as mesmas funções de *kernel*. As configurações da máquina de vetor de suporte que fizeram uso da função de *kernel* RBF apresentaram maior exatidão nas estimativas geradas para as variáveis analisadas nesse estudo.

A máquina de vetor de suporte se mostrou uma técnica eficiente na predição da altura total, volume individual e afilamento do fuste de árvores de eucalipto, gerando estimativas de maior ou de igual exatidão em relação aos modelos de regressão comumente utilizados.

A máquina de vetor de suporte apresentou generalização de maior exatidão que os modelos de regressão.

REFERÊNCIAS

ALMEIDA, M. B. de; CUNHA, M. J. da. **Support vector machine**: uma introdução. 2015. 84 p.

ALMEIDA, O. C. P. de. **Classificação de tábuas de madeira usando processamento de imagens digitais e aprendizado de máquina**. 2014. 107 p. Tese (Doutorado em Agronomia) – Universidade Estadual de São Paulo, Botucatu, 2014.

AMAMI, R.; AYED, D. B.; ELLOUZE, N. Practical selection of svm supervised parameters with different feature representations for vowel recognition. **International Journal of Digital Content Technology and its Applications (JDCTA)**, v. 7, n. 9, 418-424, 2013.

AMARAL JÚNIOR, J. B. **Estudo comparativo de previsão entre redes neurais artificiais, máquina de suporte vetorial e modelos lineares: uma aplicação à estrutura a termo das taxas de juros**. 2012. 88 p. Dissertação (Mestrado em Economia) – Universidade Federal de Pernambuco, Recife, 2012.

ANDRADE, V. C. L. de. et al. Amostragem e agrupamento de dados de relação hipsométrica em inventários florestais de cerrado Tocantinense. **Pesquisa Florestal Brasileira**, v. 35, n. 83, p. 227-238, 2015.

ANDRADE, V. C. L. de. Modelos de taper do tipo expoente-forma para descrever o perfil do fuste de árvores. **Pesquisa Florestal Brasileira**, v. 34, n. 80, p. 271-283, 2014.

ASSIS, A. L. de. et al. Comparação de modelos polinomiais segmentados e não-segmentados na estimativa de diâmetros e volumes ao longo do fuste de *Pinus taeda*. **Cerne**, v. 7, n. 1, p. 20-40, 2001.

AZEVEDO, T. L. et al. Equações hipsométricas e volumétricas para um povoamento de *Eucalyptus* sp. localizado na FLONA do Ibura, Sergipe. **Agrária**, v. 6, n. 1, p. 105-112, 2011.

BELTRAMI, M. **Precificação de opções sobre ações por modelos de support vector regression**. 2009. 125 p. Dissertação (Mestrado em Métodos Numéricos em Engenharia) – Universidade Federal do Paraná, Curitiba, 2009.

BINOTI, D. H. B.; BINOTI, M. L. M. da S.; LEITE, H. G. Configuração de redes neurais artificiais para estimação do volume de árvores. **Ciência da Madeira**, v. 5, n. 1, p. 58-67, 2014.

BINOTI, D. H. B. et al. Support vector machine to estimate volume of eucalypt trees. **Revista Árvore**, v. 40, n. 4, p. 689-693, 2016.

BINOTI, M. L. M. da S.; BINOTI, D. H. B.; LEITE, H. G. Aplicação de redes neurais artificiais para estimação da altura de povoamentos equiâneos de eucalipto. **Revista Árvore**, v. 37, n. 4, p. 639-645, 2013.

BINOTI, M. L. M. da S. et al. Redes neurais artificiais para estimação do volume de árvores. **Revista Árvore**, v. 38, n. 2, p. 283-288, 2014.

BISHOP, C. M. **Pattern Recognition and Machine Learning**. Cambridge: Springer, 2006. 738 p.

BRAGA, A. de P.; LUDEMIR, T. B.; CARVALHO, A. C. P. de L. F. **Redes neurais artificiais: teoria e aplicações**. Rio de Janeiro: LTC, 2000. 262 p.

BURGES, C. J. C. A tutorial on support vector machines for pattern recognition. **Data Mining and Knowledge Discovery**, v. 2, n. 2, p. 121-167, 1998.

BURKHART, H. E.; TOMÉ, M. **Modeling forest trees and stands**. New York: Springer, 2012. 457 p.

CALEGARIO, N. et al. Melhoria da qualidade de ajuste de modelos biométricos florestais pelo emprego da teoria dos modelos não lineares generalizados. **Scientia Forestalis**, n. 69, p. 38-50, 2005.

CAMPOS, B. P. F. **Utilização de redes neurais artificiais em inventário de florestas comerciais**. 2014. 79 p. Dissertação (Mestrado em Ciências Florestais) – Universidade Federal do Espírito Santo, Jerônimo Monteiro, 2014.

CAMPOS, J. C. C; LEITE, H. G. **Mensuração Florestal: perguntas e respostas**. 4. ed. Viçosa, MG: UFV, 2013. 605 p.

CARVALHO, S. de P. C. E. et al. Modelagem não linear mista para descrever o afilamento de árvores clonais de *Eucalyptus* sp. **Scientia Forestalis**, v. 42, n. 104, p. 605-614, 2014.

CARVALHO, S. de P. C. E. et al. Modelos não lineares generalizados aplicados na predição da área basal e volume de *Eucalyptus* clonal. **Cerne**, v. 17, n. 4, p. 541-548, 2011.

CHANG, Y-W.; HSIEH, C-J.; CHANG, K-W. Training and testing low-degree polynomial data mappings via linear SVM. **Journal of Machine Learning Research**, v. 11, p. 1471-1490, 2010.

CHEN, G.; HAY, G. J.; ZHOU, Y. Estimation of forest height, biomass and volume using support vector regression and segmentation from Lidar transects and Quickbird imagery. In: International Conference on Geoinformatics, 18, 2010. Beijing, China. **Anais... IEEE**, 2010.

CHEN, G.; HAY, G. J. Support vector regression approach to estimate forest biophysical parameters at the object level using airborne lidar transects and quickbird data. **Photogrammetric Engineering & Remote Sensing**, v. 7, n. 7, p. 733-741, 2011.

CLUTTER, J. L. et al. **Timber management: a quantitative approach**. New York: John Wiley and Sons, 1983. 333 p.

CORDEIRO, M. A. et al. Estimativa do volume de *Acacia mangium* utilizando técnicas de redes neurais artificiais e máquinas vetor de suporte. **Pesquisa Florestal Brasileira**, v. 35, n. 83, p. 255-261, 2015.

CORTES, C.; VAPNIK, V. Support-vector networks. **Machine Learning**, v. 20, p. 273-297, 1995.

COSENZA, D. N. et al. Classificação da capacidade produtiva de sítios florestais utilizando máquina de vetor de suporte e rede neural artificial. **Scientia Forestalis**, v. 43, n. 108, p. 955-963, 2015.

COUTO, H. T. Z. do; BASTOS, N. L. M. Erros de medição de altura em povoamentos de *Eucalyptus* em região plana. **Ipef**, v. 1, n. 39, p. 21-31, 1988.

COUTO, H. T. Z. do.; BASTOS, N. L. M. Modelos de equações de volume e relações hipsométricas para plantações de *Eucalyptus* no Estado de São Paulo. **Ipef**, n. 37, p. 33-44, 1987.

DIAMANTOPOULOU, M. J. Predicting fir trees stem diameters using Artificial Neural Network models. **Southern African Forestry Journal**, v. 205, n. 1, p. 39-44, 2005.

DIAS, M. S. **O uso de máquina de suporte vetorial para regressão (SVR) na estimação da estrutura a termo de taxa de juros do Brasil**. 2007. 77 p. Dissertação (Mestrado em Matemática) – Pontifícia Universidade Católica, Rio de Janeiro, 2007.

DRUCKER, H. et al. Support vector regression machines. **Advances in neural information processing systems**, v. 9, p. 155-161, 1997.

FIGUEIREDO, E. O.; SCOLFORO, J. R. S.; OLIVEIRA, A. D. de. Seleção de modelos polinomiais para representar o perfil e volume do fuste de *Tectona grandis* L.f. **Acta Amazonica**, v. 36, n. 4, p. 465-482, 2006.

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

FISCHER, F. et al. Exatidão dos modelos polinomiais não-segmentados e das razões entre volumes para representar o perfil do tronco de *Pinus taeda*. **Ciência Florestal**, v. 11, n. 1, p. 167-188, 2001.

GARCÍA NIETO, P. J. et al. Support vector machines and neural networks used to evaluate paper manufactured using *Eucalyptus globulus*. **Applied Mathematical Modelling**, v. 36, n. 12, p. 6137-6145, 2012.

GOLLAPUDI, S. **Practical machine learning**. Packt , 2016. 468 p.

GÖRGENS, E. B.; MONTAGHI, A.; RODRIGUEZ, L. C. E. A performance comparison of machine learning methods to estimate the fast-growing forest plantation yield based on laser scanning metrics. **Computers and Electronics in Agriculture**, v. 116, p. 504-5012, 2015.

HAYKIN, S. **Redes neurais: princípios e prática**. 2. ed. Porto Alegre: Bookman, 2001. 900 p.

GORGENS, E. B. et al. Influência da arquitetura na estimativa de volume de árvores individuais por meio de redes neurais artificiais. **Revista Árvore**, v. 38, n. 2, p. 289-295, 2014.

GUIMARÃES, M. A. M. et al. Height-diameter models in forestry with inclusion of covariates. **Cerne**, v. 15, n. 3, p. 313-321, 2009.

HESS, A. F. et al. Ajuste de relação hipsométrica para espécies da floresta amazônica. **Ambiência**, v. 10, n. 1, p. 21-29, 2014.

HOBBS, T.; HILBORN, R. Alternatives to statistical hypothesis testing in ecology: a guide to self teaching. **Ecological Application**, v. 16, n. 1, p. 5-19, 2006.

HORLE, D. C. et al. Modelagem não linear do perfil longitudinal de fustes de *Pinus oocarpa*. **Cerne**, v. 16, n. 2, p. 177-184, 2010.

HUSCH, B.; BEERS, T. W.; KERSHAW JUNIOR, J. A. **Forest mensuration**. 4. ed. New Jersey, Hoboken: John Wiley & Sons, Inc, 2003. 443 p.

IGEL, C. **Machine learning: kernel-based methods**. Copenhagen: University of Copenhagen, 2004. 61 p.

INDÚSTRIA BRASILEIRA DE ÁRVORES (IBÁ). **Relatório IBÁ 2015**. Brasília, 2015. 64 p.

KOETZ, B. et al. Multi-source land cover classification for forest fire management based on imaging spectrometry and LiDAR data. **Forest Ecology and Management**, v. 256, n. 3, p. 263-271, jul. 2008.

KOHLER, S. V.; KOEHLER, H. S.; FIGUEIREDO FILHO, A. Modelos de afilamento para *Pinus taeda* por classes de idade. **Floresta e Ambiente**, v. 20, n. 4, p. 470-479, 2013.

LEAL, F. A. et al. Amostragem de árvores de *Eucalyptus* na cubagem rigorosa para estimativa de modelos volumétricos. **Revista Brasileira de Biometria**, v. 33, n. 1, p. 91-103, 2015.

LEITE, H. G. et al. Estimation of inside-bark diameter and heartwood diameter for *Tectona grandis* Linn. trees using artificial neural networks. **European Journal of Forest Research**, v. 130, n. 2, p. 263-269, 2011.

LEITE, H. G.; ANDRADE, V. C. L. de. Importância das variáveis altura dominante e altura total em equações hipsométricas e volumétricas. **Revista Árvore**, v. 27, n. 3, p. 301-310, 2003.

LIANG, H.; XILONG, C. Parameter selection in time series prediction bases on nu-support vector regression. **High Technology Letters**, v. 15, n. 4, p. 337-342, 2009.

LIMA, C. A. M. **Comitê de máquinas: uma abordagem unificada empregando máquina de vetores-suporte**. 2004. 378 p. Tese (Doutorado em Engenharia Elétrica) – Universidade Estadual de Campinas, Campinas, 2004.

LIN, H-T.; LIN, C-J. A study on sigmoid kernels for SVM and the training of non-PSD kernels by SMO-type methods. **Department of Computer Science and Information, Engineering National Taiwan University Taipei**, 2003.

LOETSCH, F.; ZOHRER, F. HALLER, K. E. **Forest inventory**. Munchen: BLV Verlagsgesellschaft, 1973. v. 2, 469 p.

LORENA, A. C.; CARVALHO, C. P. L. F. de. Uma Introdução às Support Vector Machines. **Revista de Informática Teórica e Aplicada**, v. 14, n. 2, p. 43-67, 2007.

MA, X.; ZHANG, Y.; WANG, Y. Performance evaluation of kernel functions based on grid search for Support Vector Regression. In: 7 th IEEE Conference on Cybernetics and Intelligent Systems (CIS) and Robotics, Automation and Mechatronics (RAM), 2015. **Anais...IEEE**, 2015. p. 283-288.

MACHADO, S. do A.; FIGUEIREDO FILHO, A. **Dendrometria**. 2. ed. Guarapuava, PR: UNICENTRO, 2009. 316 p.

MAGALHÃES, J. G. de S. **Prognose e regulação de florestas de eucalipto utilizando modelo de crescimento e inteligência computacional**. 2014. 66 p. Dissertação (Mestrado em Ciência Florestal) – Universidade Federal de Viçosa, Viçosa, 2014.

MARTINS, E. dos R. et al. Configuração de redes neurais artificiais para estimação do afilamento do fuste de árvores de eucalipto. **Agrária**, v. 11, n. 1, p. 33-38, 2016.

MCTAGUE, J. P.; BATISTA, J. L. F.; STEINER, L. H. Equações de volume total, volume comercial e forma do tronco para plantações de *Eucalyptus* nos Estados de São Paulo e Rio de Janeiro. **Ipef**, n. 41/42, p. 56-63, 1989.

MENDONÇA, A. R. de et al. Modelos hipsométricos e de crescimento em altura das árvores dominantes e codominantes para *Pinus caribaea* var . hondurensis. **Scientia Forestalis**, v. 39, n. 90, p. 151-160, 2011.

MENDONÇA, A. R. de; CARVALHO, S. de P. C. E; CALEGARIO, N. Modelos hipsométricos generalizados mistos na predição da altura de *Eucalyptus* sp. **Cerne**, v. 21, n. 1, p. 107-115, 2015.

MENON, M. U. Ajuste dos modelos de Kozak e do sistema Burkhart e Cao para plantações de *Pinus oocarpa*. **Ambiência**, v. 2, n. 2, p. 197-205, 2006.

MEUNIER, I. M. J.; SILVA, J. A. A.; FERREIRA, R. L. C. **Inventário florestal: programas de estudo**. Recife: Imprensa Universitária da UFRPE, 189 p. 2001.

MEYER, D. et al. **e1071**: Misc functions of the department of statistics, probability theory group, TU Wien. R package version 1.6-7, 2015.

- MIGUEL, E. P. et al. Modelos polinomiais para representar o perfil e o volume do fuste de *Eucalyptus urophylla* na região norte do Estado do Goiás. **Floresta**, v. 41, n. 2, p. 355-368, 2011.
- MIRANDA, D. L. C. de; JUNIOR, V. B.; GOUVEIA, D. M. Fator de forma e equações de volume para estimativa volumétrica de árvores em plantio de *Eucalyptus urograndis*. **Scientia Plena**, v. 11, n. 3, p. 1-8, 2015.
- MONNET, J-M.; BERGER, F.; CHANUSSOT, J. Support vector regression for the estimation of forest stand parameters using airborne laser scanning. **IEEE Geoscience and Remote Sensing Letters**, v. 8, n. 3, p. 580-584.
- MOREIRA, M. F. B. et al. Estimativa da relação hipsométrica com modelos não lineares ajustados por métodos bayesianos empíricos. **Cerne**, v. 21, n. 3, p. 405-411, 2015.
- MÜLLER, M. D. et al. Equações de altura, volume e afilamento para eucalipto e acácia estabelecidos em sistema silvipastoril. **Floresta**, v. 44, n. 3, p. 473-484, 2014.
- MURTA JÚNIOR, L. S. et al. Avaliação do método da similaridade de perfis na estimação do volume de árvores. **Scientia Forestalis**, v. 43, n. 106, p. 435-444, 2015.
- NOBLE, W. S. What is a support vector machine? **Nature Biotechnology**, v. 24, n. 12, p. 1565-1567, 2006.
- NUNES, C. A. et al. Determinação de constituintes químicos em madeira de eucalipto por PI-CG/EM e calibração multivariada: comparação entre redes neurais artificiais e máquina de vetor suporte. **Química Nova**, v. 34, n. 2, p. 279-283, 2011.
- PINHO, A. G. de. Modelagem de cross-selling no marketing de relacionamento por máquinas de vetor de suporte. **Revista Pensamento Contemporâneo em Administração**, v. 3, n. 1, p. 66-79, 2009.
- PIRES, L. M.; CALEGARIO, N. Ajuste de modelos estocásticos lineares e não-lineares para a descrição do perfil longitudinal de árvores. **Árvore**, v. 31, n. 5, p. 845-852, 2007.
- QUEIROZ, D. et al. Identidade de modelos em funções de afilamento para *Mimosa scabrella* Benth em povoamentos nativos da região metropolitana de Curitiba/PR. **Floresta**, v. 38, n. 2, p. 339-349, 2008.
- R CORE TEAM. **R: a language and environment for statistical computing, version 3.2.4**. Vienna, Austria: R Foundation for Statistical Computing, 2015.
- RETSLAFF, F. A. S. et al. Curvas de sítio e relações hipsométricas para *Eucalyptus grandis* na região dos Campos Gerais, Paraná. **Cerne**, v. 21, n. 2, p. 219-225, 2015.
- RIBEIRO, A. et al. Estratégias e metodologias de ajuste de modelos hipsométricos em plantios de *Eucalyptus* sp. **Cerne**, v. 16, n. 1, p. 22-31, 2010.

RIBEIRO, R. B. da S.; GAMA, J. R. V.; MELO, L. de O. Seccionamento para cubagem e escolha de equações de volume para a floresta nacional do Tapajós. **Cerne**, v. 20, n. 4, p. 605-612, 2014.

RODRIGUEZ-GALIANO, V. et al. Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. **Ore Geology Reviews**, v. 71, p. 804-818, 2015.

SCHIKOWSKI, A. B.; CORTE, A. P. D.; SANQUETTA, C. R. Estudo da forma do fuste utilizando redes neurais artificiais e funções de afilamento. **Pesquisa Florestal Brasileira**, v. 35, n. 82, p. 119-127, 2015.

SCHNEIDER, P. R.; SCHNEIDER, P. S. P. **Introdução ao manejo florestal**. 2. ed. Santa Maria, RS: FACOS-UFSM, 2008. 556 p.

SCHÖLKOPF, B. et al. Shrinking the tube: a new support vector regression algorithm. **Advances in Neural Information Processing Systems**, v. 11, p. 330-336, 1999.

SCHÖLKOPF, B.; SMOLA, A. J. **Learning with kernels**. Cambridge, MA: MIT Press, 2002.

SCHRÖDER, T. et al. Métodos de estimativa de volume comercial para *Eucalyptus grandis*: especificidades e recomendações. **Pesquisa Florestal Brasileira**, v. 33, n. 73, p. 1-7, 2013.

SCOLFORO, J. R. S. **Biometria florestal**: parte I: modelos de regressão linear e não-linear; parte II modelos para relação hipsométrica, volume, afilamento, e peso de matéria seca. Lavras: UFLA/FAEPE, 2005. 352 p.

SCOLFORO, J. R. S.; MELLO, J. M de. **Inventário florestal**. Lavras: UFLA/FAEPE, 2006. 561 p.

SCOLFORO, J. R. S. et al. Acuracidade de equações de afilamento para representar o perfil do fuste de *Pinus elliottii*. **Cerne**, v. 4, n. 1, p. 100-122, 1998.

SENA, A. L. M. et al. Modelos lineares e não lineares com uso de covariantes para relação hipsométrica de duas espécies de *Pinus* tropicais. **Ciência Florestal**, v. 25, n. 4, p. 969-980, 2015.

SHILTON, A.; PALANISWAMI, M. A modified v-SV method for simplified regression. In: International Conference on Intelligent Sensing and Information Processing, 2004. **Anais...** IEEE, 2004. p. 422-427.

SILVA, M. L.; CORDEIRO, S. A. Evolução dos estudos econômicos relacionados à eucaliptocultura. In: VALE, A. B. do. et al. **Eucaliptocultura no Brasil**: silvicultura, manejo e ambiência. Viçosa, MG: SIF, 2014. p. 413-420.

SMOLA, A. J.; SCHÖLKOPF, B. A tutorial on support vector regression. **Statistics and Computing**, v. 14, p. 199-222, 2004.

SOARES, C. P. B.; PAULA NETO, F.; SOUZA, A. L. S. **Dendrometria e inventário florestal**. 2. ed. Viçosa, MG: UFV, 2011. 272 p.

SOARES, F. A. A. M. N. et al. Recursive diameter prediction for calculating merchantable volume of *Eucalyptus* clones using Multilayer Perceptron. **Neural Computing and Applications**, v. 22, n. 7-8, p. 1407-1418, 2013.

SOARES, T. S.; LEITE, H. G.; VALE, A. B. do. Um modelo alternativo para a quantificação de multiprodutos em árvores individuais. **Revista Árvore**, v. 28, n. 6, p. 875-882, 2004.

SOUZA, G. T. O. et al. Relações hipsométricas para *Eucalyptus urophylla* conduzidos sob regime de alto fuste e talhadia no sudoeste da Bahia. **Scientia Plena**, v. 9, n. 4, p. 1-7, 2013.

SOUZA, C. A. M. de. et al. Avaliação de modelos de afilamento segmentados na estimação da altura e volume comercial de *Eucalyptus* sp. **Revista Árvore**, v. 32, n. 3, p. 453-463, 2008a.

SOUZA, C. A. M. de. et al. Modelos de afilamento para o sortimento do fuste de *Pinus taeda* L. **Ciência Rural**, v. 38, n. 9, p. 2506-2511, 2008b.

SOUZA, C. A. M. de. et al. Avaliação de modelos de afilamento não-segmentados na estimação da altura e volume comercial de *Eucalyptus* sp. **Ciência Florestal**, v. 18, n. 3, p. 387-399, 2008c.

TAHIM, A. P. N. **Máquinas de vetores de suporte (SVM)**. Florianópolis, 2010.

TAN, P.-N.; STEINBACH, M.; KUMAR, V.; **Introdução ao data mining**: mineração de dados. Rio de Janeiro: Ciência Moderna, 2009. 900 p.

TEIXEIRA, L. L. et al. Comparativo entre arima, redes neurais artificiais, máquinas de vetores suporte para regressão e combinação linear de previsões de vazões mensais no posto 266 – Itaipu. In: Simpósio de Engenharia de produção, 21., 2014, Bauru, SP. **Anais...** Bauru, SP: UNESP, 2014. p. 1-15.

TÉO, S. J. et al. Modelos de afilamento para *Pinus elliottii* em diferentes idades, na região de Caçador, SC. **Floresta**, v. 43, n. 3, p. 439-452, 2013.

THOMAS, C. et al. Comparação de equações volumétricas ajustadas com dados de cubagem e análise de tronco. **Ciência Florestal**, v. 16, n. 3, p. 319-327, 2006.

ÜSTÜN, B.; MELSSSEN, W. J.; BUYDENS, L. M. C. Visualisation and interpretation of support vector regression models. **Analytica Chimica Acta**, v. 595, p. 299-309, 2007.

VAPNIK, V. N. An overview of statistical learning theory. **Ieee Transactions on Neural Networks**, v. 10, n. 5, p. 988-999, 1999.

VAPNIK, V. N. **Statistical learning theory**. New York: John Wiley and Sons, 1998. 732 p.

VENDRUSCOLO, D. G. S. et al. Estimativa da altura de eucalipto por meio de regressão não linear e redes neurais. **Revista Brasileira de Biometria**, v. 33, n. 4, p. 556-569, 2015.

VERDÉRIO, A. **Sobre o uso de regressão por vetores suporte para a construção de modelos em um método de região de confiança sem derivadas**. 2015. 115 p. Tese (Doutorado em Matemática Aplicada) – Universidade Federal do Paraná, Curitiba, 2015.