

KALÉO DIAS PEREIRA

**APRENDIZAGEM DE MÁQUINA E TÉCNICAS MULTIVARIADAS NO ESTUDO
DA QUALIDADE DO CARVÃO VEGETAL**

Dissertação apresentada à Universidade Federal de Viçosa como parte das exigências do Programa de Pós-Graduação em Estatística Aplicada e Biometria, para a obtenção do título de Magister Scientiae.

VIÇOSA
MINAS GERAIS - BRASIL
2019

**Ficha catalográfica preparada pela Biblioteca Central da Universidade
Federal de Viçosa - Câmpus Viçosa**

T

P436a
2019
Pereira, Kaléo Dias, 1993-
Aprendizagem de máquina e técnicas multivariadas no
estudo da qualidade do carvão vegetal / Kaléo Dias Pereira. –
Viçosa, MG, 2019.
xii, 53 f. : il. (algumas color.) ; 29 cm.

Inclui apêndice.

Orientador: Antônio Policarpo Souza Carneiro.

Dissertação (mestrado) - Universidade Federal de Viçosa.

Inclui bibliografia.

1. Fisher, Funções de. 2. *Corymbia*. 3. Algoritmos.
4. Variáveis (Matemática). 5. Carvão - Qualidade.
I. Universidade Federal de Viçosa. Departamento de Estatística.
Programa de Pós-Graduação em Estatística Aplicada e
Biometria. II. Título.

CDD 22. ed. 511.326

KALÉO DIAS PEREIRA

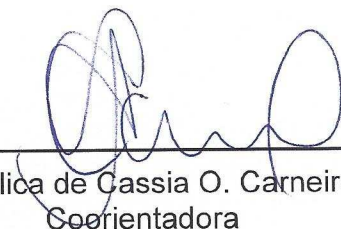
**APRENDIZAGEM DE MÁQUINA E TÉCNICAS MULTIVARIADAS NO
ESTUDO DA QUALIDADE DO CARVÃO VEGETAL**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Estatística Aplicada e Biometria, para obtenção do título de *Magister Scientiae*.

APROVADA: 18 de fevereiro de 2019.



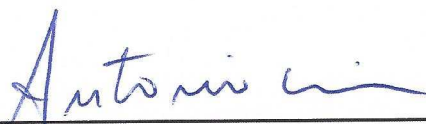
Gerson Rodrigues dos Santos
Coorientador



Angelica de Cassia O. Carneiro
Coorientadora



Helio Garcia Leite



Antonio Policarpo Souza Carneiro
Orientador

Aos meus pais e meus irmãos...

AGRADECIMENTOS

Aos meus pais Sulamita Dias Pereira e Dourival de Souza Pereira pelo apoio, confiança e orientação que sempre me guiam ao caminho certo.

Aos meus irmãos, primos, tios e avós que sempre torcem e emanam boas energias para que eu vença os desafios profissionais e da vida.

Ao meu grande e melhor amigo Yweenns Teixeira Ferraz pelo vínculo de irmandade que sempre se faz presente nos momentos bons e adversos.

Ao meu orientador Antônio Policarpo Souza Carneiro e meus coorientadores Gerson Rodrigues dos Santos e Angélica de Cassia Oliveira Carneiro pelas oportunidades, ensinamentos e suporte ao longo do curso de Mestrado.

Aos meus amigos, colegas, docentes e funcionários do Departamento de Estatística e da pós-graduação em Estatística Aplicada e Biometria pela força, troca de experiências e conhecimentos que me ajudaram a concluir mais esta etapa e evoluir como indivíduo e profissional.

Aos companheiros do Laboratório de Painéis e Energia da Madeira da Universidade Federal de Viçosa que forneceram dados e apoio que viabilizaram este trabalho.

À Universidade Federal de Viçosa pela oportunidade de acesso à infraestrutura e filosofia de fomento do conhecimento diferenciadas, de modo a tornar-se um segundo lar acadêmico.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e à Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) pela concessão da bolsa de mestrado e pelos recursos financeiros disponibilizados para participação de eventos científicos, respectivamente.

BIOGRAFIA

Kaléo Dias Pereira, primogênito fruto do casamento de Sulamita Dias Pereira e Dourival de Souza Pereira, nascido em 28 de Maio de 1993 no Município de Belém, Pará.

Morou praticamente a vida toda em Ananindeua, região metropolitana de Belém, onde completou todo o ensino básico. Em 2011, ingressou no curso de Bacharel em Engenharia Florestal, campus de Belém – PA, da Universidade Federal Rural da Amazônia (UFRA).

No início de 2012, entrou para o Programa Institucional de Bolsas de Iniciação Científica (PIBIC) da UFRA, ficando alocado no laboratório de análise química de solos e participando de pesquisas envolvendo fertilidade de solos amazônicos, nutrição mineral de plantas e estudos relacionados à implantação e manejo de reflorestamentos comerciais de eucalipto.

Até meados de 2014, acompanhou vários trabalhos de campo voltados para o manejo do eucalipto em virtude de projeto de pesquisa em parceria com a empresa Vale Florestar S.A. (Suzano Papel e Celulose S.A.), o que levou ao despertar do interesse pela pesquisa e, em particular, pelos métodos de análise estatística de dados.

Entre 2014 e 2015, passou a dar auxílio, basicamente, nos processos de análise de dados das atividades de projeto de pesquisa financiado pela Vale S.A., que visava o controle de gramíneas exóticas e recomposição da vegetação nativa da FLONA de Carajás, em Parauapebas – PA. Em junho de 2016, recebeu o diploma de Engenheiro Florestal.

Em 2017, iniciou o mestrado acadêmico no Programa de Pós-Graduação em Estatística Aplicada e Biometria da Universidade Federal de Viçosa, em Viçosa – MG. No mestrado aprofundou e ampliou os conhecimentos que já vinham sendo buscados na área de Estatística Aplicada. Em 18 de Fevereiro de 2018 foi aprovado no exame de defesa de dissertação para obtenção do título de Magister Scientiae.

SUMÁRIO

LISTA DE FIGURAS	vii
LISTA DE TABELAS	viii
RESUMO	ix
ABSTRACT	xi
1 INTRODUÇÃO GERAL	1
2 OBJETIVOS	2
2.1 Objetivo geral	2
2.2 Objetivos específicos	2
3 REVISÃO BIBLIOGRÁFICA	2
3.1 Hibridação com espécies do gênero <i>Corymbia</i>	2
3.2 Qualidade do carvão vegetal	3
3.3 Aprendizagem de máquina	5
3.3.1 Aspectos da aprendizagem	5
3.3.2 Paradigmas da aprendizagem	6
3.3.3 Árvores de decisão	7
3.3.4 Random trees e random forest	9
3.3.5 Random forest para regressão	11
3.3.6 Support vector machine – SVM	12
3.4 Análise multivariada	13
3.4.1 Análise de variância multivariada (MANOVA)	14
3.4.2 Função discriminante de Fisher	16
REFERÊNCIAS BIBLIOGRÁFICAS	18
CAPÍTULO 1	24
Resumo	24
1 INTRODUÇÃO	25
2 MATERIAL E MÉTODOS	26
2.1 Propriedades anatômicas e físicas da madeira	27
2.2 Análise da composição química elementar e estrutural da madeira	27
2.3 Rendimento gravimétrico e propriedades do carvão vegetal	27
2.4 Construção dos algoritmos	27
3 RESULTADOS E DISCUSSÃO	28
3.1 Propriedades da madeira e rendimento/qualidade do carvão vegetal	28
3.2 Validação cruzada	32
4 CONCLUSÕES	35
REFERÊNCIAS BIBLIOGRÁFICAS	35

CAPÍTULO 2	40
Resumo	40
1 INTRODUÇÃO	41
2 MATERIAL E MÉTODOS.....	42
2.1 Rendimento gravimétrico e propriedades do carvão vegetal.....	42
2.2 Análises estatísticas.....	43
3 RESULTADOS E DISCUSSÃO.....	45
4 CONCLUSÕES	49
REFERÊNCIAS BIBLIOGRÁFICAS.....	49
APÊNDICE	53

LISTA DE FIGURAS

Figura 1 – Exemplo de árvore de decisão	8.
Figura 2 – Processo splitting de uma árvore de decisão com duas variáveis independentes.....	8.
Figura 3 – Exemplo de funcionamento do Random forest.....	11.
Figura 4 – Ilustração de uma support vector machine linear para classificação.....	12.
Figura 1 – Influência das propriedades da madeira: densidade básica da madeira (Dens.), poder calorífico superior (P.C.S.), teor de extrativos (Extrat.), lignina total (Lignina), holocelulose (Holocel.) e relação cerne/alburno (Cerne.Alburno) sobre o rendimento da carbonização (Rend. da carbonização), teor de materiais voláteis, teor de carbono fixo, friabilidade, densidade relativa aparente (Dens. aparente) e poder calorífico superior (PCS) do carvão vegetal de clones de <i>Corymbia</i> medida pela redução média da impureza do nó (RMIN) das árvores de regressão.....	29.

LISTA DE TABELAS

Tabela 1 – Correlação de cópulas (r') e coeficiente de determinação (R^2) do algoritmo random forest (RF) implementados apenas com as propriedades mais influentes para o rendimento e qualidade do carvão vegetal de clones do gênero <i>Corymbia</i>	30.
Tabela 2 – Médias das propriedades do carvão de clones do gênero <i>Corymbia</i>	30.
Tabela 3 – Coeficiente de determinação do random forest (RF), support vector regression (SVR) e regressão linear múltipla (RLM) em função das seis variáveis preditoras da madeira (1ª estimação) e em função apenas da mais importante (2ª estimação), considerando dados de validação.....	33.
Tabela 4 – Coeficiente de correlação linear entre valores observados e preditos pelo random forest (RF), support vector regression (SVR) e regressão linear múltipla (RLM) em função das seis variáveis preditoras da madeira (1ª estimação) e em função apenas da mais importante (2ª estimação), considerando dados de validação.....	33.
Tabela 5 – Erro médio absoluto do random forest (RF), support vector regression (SVR) e regressão linear múltipla (RLM) utilizando as seis variáveis preditoras da madeira (1ª estimação) e apenas a mais importante (2ª estimação), considerando dados de validação.....	34.
Tabela 6 – Raiz quadrada do erro quadrático médio do random forest (RF), support vector regression (SVR) e regressão linear múltipla (RLM) utilizando as seis variáveis preditoras da madeira (1ª estimação) e apenas a mais importante (2ª estimação), considerando dados de validação.....	34.
Tabela 1 – Resultados dos testes de normalidade e homocedasticidade das variáveis do carvão vegetal.....	45.
Tabela 2 – Resumo da MANOVA das variáveis de rendimento e qualidade do carvão vegetal em função de 12 clones do gênero <i>Corymbia</i>	46.
Tabela 3 – Informação retida pelos respectivos autovalores.....	47.
Tabela 4 – Coeficientes normalizados da primeira função discriminante (FD1) e da segunda função discriminante (FD2) de Fisher.....	47.
Tabela 5 – Médias das variáveis canônicas de 12 clones de eucalipto plantados no Município de Dionísio, MG.....	48.
Tabela 6 – Valores da estatística F, valor-p e acurácia seletiva (AS) das variáveis observadas e variáveis canônicas.....	49.

RESUMO

PEREIRA, Kaléo Dias, M.Sc., Universidade Federal de Viçosa, fevereiro de 2019. **Aprendizagem de máquina e técnicas multivariadas no estudo da qualidade do carvão vegetal.** Orientador: Antônio Policarpo Souza Carneiro. Coorientadores: Gérson Rodrigues dos Santos e Angélica de Cássia Oliveira Carneiro.

Os estudos sobre as variáveis que determinam a qualidade do carvão vegetal e a influência do material de origem norteiam o planejamento de programas de seleção de melhores genótipos para a produção de carvão. O emprego de novos métodos de análise que se adequem ao estudo das propriedades do carvão vegetal, possibilita a avaliação dos dados por ângulos diferentes e amplia as possibilidades das pesquisas na área. Nesse sentido, o objetivo do presente trabalho foi utilizar procedimentos de aprendizagem de máquina e técnicas multivariadas na análise do rendimento e qualidade do carvão vegetal produzido a partir de clones de *Corymbia*. As amostras analisadas foram obtidas a partir de um plantio clonal com sete anos de idade estabelecido no município de Dionísio, MG. No primeiro capítulo são apresentados os resultados da utilização do algoritmo random forest no estudo da influência das propriedades da madeira sobre o rendimento e propriedades de qualidade do carvão vegetal, bem como a comparação da acurácia dos valores preditos pelo random forest com os preditos pelo support vector regression e regressão linear múltipla. As variáveis teor de holocelulose, relação cerne/alburno e densidade básica da madeira foram as mais importantes para a modelagem via aprendizagem de máquina. Quanto a acurácia, o random forest foi superior aos demais métodos considerando o coeficiente de determinação, correlação linear entre valores observados e preditos, erro médio absoluto e raiz quadrada do erro quadrático médio, inclusive mostrando desempenho adequado para que seja viável a utilização do algoritmo para a estimação das propriedades do carvão vegetal. No segundo capítulo relata-se o emprego da função discriminante de Fisher na classificação dos clones de *Corymbia* quanto ao potencial para a produção de carvão vegetal em termos de rendimento e qualidade. Os dados foram inicialmente testados quanto às pressuposições de normalidade multivariada e homogeneidade de matrizes de variâncias/covariâncias, para em seguida aplicar a análise de variância multivariada (MANOVA). Pelos resultados da MANOVA, constatou-se que existe diferença no campo multivariado entre os clones e, a partir das matrizes de soma de quadrados e produtos do resíduo e do efeito de clones, foram estimados os coeficientes das duas primeiras funções discriminantes, que juntas retiveram aproximadamente 80% da informação contida no conjunto de dados. As duas funções discriminantes foram utilizadas para calcular duas variáveis

canônicas que são funções das variáveis observadas do carvão vegetal. Comparando os clones por meio das médias das variáveis canônicas, verificou-se que o genótipo AMF 1119 é o mais indicado para a produção de carvão vegetal.

ABSTRACT

PEREIRA, Kaléo Dias, M.Sc., Universidade Federal de Viçosa, February, 2019. **Machine learning and multivariate techniques in the study of charcoal quality**. Advisor: Antônio Policarpo Souza Carneiro. Co-advisors: Gérson Rodrigues dos Santos and Angélica de Cássia Oliveira Carneiro.

The studies on the variables that determine the charcoal quality and the influence of the source material guide the planning of programs to select the best genotypes for the charcoal production. The use of new methods of analysis that are adequate for the study of the charcoal properties, allows the evaluation of the data by different angles and enlarges the possibilities of the researches in the area. In this sense, the objective of the present work was to use machine learning procedures and multivariate techniques to analyze the yield and quality of charcoal produced from *Corymbia* clones. The analyzed samples were obtained from a clonal plantation with seven years of age established in the city of Dionísio, MG. In the first chapter are presented the results of the use of the random forest algorithm in the study of the influence of the properties of the wood on the yield and quality properties of the charcoal, as well as the comparison of the accuracy of the values predicted by the random forest with those predicted by the support vector regression and multiple linear regression. The variables holocellulose, heartwood/sapwood ratio and basic density of the wood were the most important for modeling through machine learning. As for the accuracy, the random forest was superior to the other methods considering the coefficient of determination, linear correlation between observed and predicted values, absolute mean error and square root of the mean square error, even showing adequate performance so that it is feasible to use the algorithm for the estimation of the charcoal properties. In the second chapter we report the use of Fisher's discriminant function in the classification of *Corymbia* clones as to the potential for charcoal production in terms of yield and quality. Data were initially tested for assumptions of multivariate normality and homogeneity of variance/covariance matrices, followed by multivariate analysis of variance (MANOVA). From the MANOVA results, it was found that there is a difference in the multivariate field between the clones and, from the matrices of sum of squares and products of the residue and the effect of clones, the coefficients of the two first discriminant functions were estimated, which together almost 80% of the information contained in the data set. The two discriminant functions were used to calculate two canonical variables that are functions of the

observed charcoal variables. Comparing the clones by means of the canonical variables, it was verified that the genotype AMF 1119 is the most suitable for the charcoal production.

1 INTRODUÇÃO GERAL

Historicamente, a biomassa florestal sempre teve papel importante no contexto energético do Brasil, sendo, inclusive, a principal fonte de energia até 1972. Apesar da substancial redução da participação da madeira no cenário energético nacional ainda na década de 70, o consumo de carvão vegetal manteve-se expressivo por conta da crise do petróleo, sendo destinado principalmente à siderurgia (BRITO, 1990).

Um ponto importante relacionado à produção de carvão vegetal é a evolução da participação da madeira de florestas equiâneas como fonte de matéria-prima, passando de 49 para 84% do total de carvão consumido no país entre 2008 e 2016, o que representa 3,8 milhões de toneladas no último levantamento (IBÁ, 2018). Assim, torna-se evidente a função ambiental das florestas energéticas na conservação da biodiversidade ao suprir as demandas por carvão, aliviando a pressão sobre as florestas inequiâneas constatada em anos anteriores (SANTOS e HATAKEYAMA, 2012; SIMIONI et al., 2017).

As espécies de eucalipto são as mais utilizadas em reflorestamentos para fins energéticos. Embora possuam naturalmente características desejáveis como crescimento rápido e plasticidade, várias pesquisas têm buscado selecionar genótipos que potencializem a qualidade e quantidade de madeira produzida (BRITO et al., 1983; PEREIRA, 2012; PROTÁSIO et al., 2013; CARNEIRO et al., 2017). Segundo Loureiro (2016), híbridos do gênero *Corymbia* tem atraído as atenções de pesquisadores, visto que o cruzamento das duas espécies tem gerado materiais que apresentam ganhos significativos em crescimento, facilidade de propagação vegetativa e maior tolerância ao frio, doenças e ataques de insetos.

Na siderurgia, a qualidade do processo depende da homogeneidade do carvão, que, por sua vez, depende de fatores relacionados à carbonização da madeira, ressaltando a necessidade do conhecimento acerca da influência das características da madeira sobre as propriedades do carvão (SOARES et al., 2014; SOARES et al.; 2015). Devido a grande quantidade de propriedades da madeira e do carvão usualmente avaliadas, técnicas estatísticas univariadas podem não ser eficientes, por desconsiderarem a dependência entre variáveis e apresentam interpretação global dos resultados mais dispendiosa.

Trabalhos com seleção de clones para a produção de carvão também se deparam com o problema de alta dimensionalidade de dados, ao passo que existe uma gama de variáveis importantes para a eficiência energética. Para lidar adequadamente com análise simultânea de muitas variáveis, pesquisas sobre propriedades de madeira/carvão e seleção de clones para

florestas energéticas tem utilizado técnicas de análise multivariada frequentemente (TRUGILHO et al., 1997; LOBÃO et al., 2010; PROTÁSIO et al., 2012; CASTRO et al., 2013).

Algumas técnicas multivariadas também podem ser reunidas dentro de um conceito oriundo da computação conhecido como aprendizagem de máquina, ou machine learning em inglês, a exemplo da análise discriminante (aprendizagem supervisionada para classificação) e análise de agrupamento (aprendizagem não supervisionada). Na computação, o aprendizado de máquina é um ramo da inteligência artificial e é entendido basicamente como o uso associado de técnicas estatísticas e computacionais para criar softwares ou algoritmos autônomos treinados a partir de um banco de dados conhecido, ou seja, algoritmos capazes de analisar e detectar padrões de forma automática em um conjunto de dados (MURPHY, 2012). A aplicação do machine learning na análise de dados amplia as possibilidades de forma tão significativa que tem se tornado comum, em determinadas situações, substituir o termo análise estatística por ciência de dados.

2 OBJETIVOS:

2.1 Objetivo geral: aplicar aprendizagem de máquina e técnicas multivariadas na análise do rendimento e qualidade do carvão vegetal produzido a partir de clones de *Corymbia*.

2.2 Objetivos específicos:

Estudar a relação entre propriedades da madeira e do carvão aplicando aprendizagem de máquina;

Identificar as propriedades da madeira mais influentes no rendimento e qualidade do carvão do vegetal;

Verificar o potencial do uso de aprendizado de máquina na predição das propriedades do carvão vegetal;

Utilizar a função discriminante de Fisher para reduzir a dimensionalidade dos dados de rendimento e qualidade do carvão vegetal;

Classificar os clones de eucalipto quanto ao rendimento e qualidade da produção de carvão vegetal.

3 REVISÃO BIBLIOGRÁFICA

3.1 Híbridaç o com esp cies do g nero *Corymbia*

Atualmente, os gen tipos resultantes da hibrida o entre *Corymbia torelliana* e *Corymbia citriodora* tem recebido aten o especial dos pesquisadores e empresas que atuam

com reflorestamento comercial, embora pouco se saiba sobre as reais potencialidades dos híbridos dessas espécies. Conforme os resultados de Assis (2000), um dos principais ganhos na hibridação é quanto ao enraizamento, visto que o *C. citriodora* tem apenas 2,5% de capacidade de enraizamento, enquanto que *C. torelliana* possui 32,6%, de modo que o cruzamento de *Corymbia* promove uma capacidade de enraizamento de 27,5%.

Além de melhor capacidade de estabelecimento de mudas em campo, a melhoria do enraizamento também favorece a propagação vegetativa, conforme mostra Trueman e Richardson (2008) trabalhando com enraizamento de estacas, no qual as estacas de híbridos apresentaram maior volume de raízes que as de *C. torelliana* e *C. citriodora*, além de maior área foliar.

O potencial energético de híbridos de *C. torelliana* com *C. citriodora* foi avaliado por Loureiro (2016) em seu trabalho com clones plantados em dois municípios de Minas Gerais, no qual verificou que os materiais em geral produzem carvão com poder calorífico superior dentro de valores satisfatórios, inclusive superando híbridos de outras espécies. O autor ainda concluiu que o melhor genótipo, dentre os avaliados, para produção de carvão com destino à siderurgia foi plantado no município de Itamarandiba e é resultado do cruzamento de *C. citriodora* com *C. torelliana*.

3.2 Qualidade do carvão vegetal

A qualidade do processo nas siderurgias depende da qualidade do carvão vegetal utilizado, que por sua vez depende de vários fatores como processo de carbonização, infraestrutura e mão de obra da produção e o tipo de madeira empregada (GALVAN, 2016). Com relação especificamente às características da madeira, suas propriedades físico-químicas influenciam não apenas a qualidade do carvão vegetal, como também são importantes na dinâmica da decomposição térmica, o que é determinante no rendimento do carvão produzido (PROTÁSIO et al., 2012).

Em plantios estabelecidos com a finalidade de produzir carvão vegetal com destino a siderurgia, Brito e Barrichelo (1977) citam que devem ser utilizados, preferencialmente, clones que produzam madeira de alta densidade e com elevada concentração de lignina. Trugilho et al. (1997) afirmam que a madeira para produção de carvão também deve apresentar baixo teor de cinzas e parede celular mais espessa e de menor largura. A escolha da madeira com características favoráveis, somado ao processo de carbonização adequado, são determinantes para uma elevada qualidade e rendimento em carvão vegetal (TRUGILHO et al., 1997).

Entre os aspectos relacionados às propriedades do carvão vegetal que são importantes para a performance dos aparelhos de redução das siderúrgicas, Coutinho e Ferraz (1988) citam a composição química, densidade, reatividade, friabilidade, resistência mecânica, higroscopicidade, além de outras. Os autores ainda se referem em especial à friabilidade, afirmando que, em caso de carvão vegetal altamente friável, a geração de finos desde a fabricação até a utilização pode chegar a 25%.

A resistência do carvão vegetal é um fator de grande importância para a maioria dos usos ao qual pode ser destinado, sendo esta dependente da densidade do material (PEREIRA et al., 2016). Segundo Brito (1993), a utilização de carvão vegetal de alta densidade em comparação ao emprego de insumos com as mesmas características e menor densidade, resulta em maior produtividade no alto forno, ou seja, o carvão de maior densidade reside por mais tempo no interior da zona de reserva térmica do alto forno.

A composição química do carvão vegetal é influenciada pela temperatura de carbonização, além da madeira empregada. A avaliação da composição química do carvão é realizada a partir da análise química imediata, a qual revela os teores de carbono fixo, de voláteis e de cinzas. A análise química imediata se faz necessária para distinguir o carbono total presente no carvão vegetal e a parcela realmente fixa, visto que parte do carbono é perdido em forma de gás CO, CO₂ e hidrocarbonetos que fazem parte dos materiais voláteis (SILVA, 1988) e não são interessantes para o uso do carvão em si. O teor de cinzas representa o produto da combustão total do carvão vegetal e agrega todas impurezas inorgânicas que, em grandes quantidades, prejudicam a eficiência dos altos fornos, especialmente por aumentar o consumo de carvão por unidade de ferro-gusa produzido (BARROSO, 2007). De acordo com Soares et al. (2014), o teor de carbono fixo presente no carvão vegetal é a propriedade de maior influência sobre o seu poder calorífico superior, ou seja, o carbono fixo determina a quantidade de energia gerada por unidade de massa de carvão.

É de suma relevância discorrer também sobre o rendimento gravimétrico da carbonização (RGC), embora não seja uma característica do carvão vegetal e nem da madeira, é talvez a primeira medida a se observar a viabilidade de se utilizar um determinado material para produzir carvão, já que o rendimento gravimétrico mensura a proporção da conversão da madeira em carvão. O aumento do RGC tende a ser obtido, além pelo controle do processo, por meio da utilização de madeiras mais densas e com maior concentração de lignina (BRITO e BARRICHELO, 1977; TRUGILHO et al., 2001; NEVES et al., 2011; PEREIRA et al., 2013). Do ponto de vista prático, o carvão vegetal dito de qualidade deve apresentar ruído metálico ao

ser quebrado, ser inodoro, duro, ter superfície de ruptura curva, lisa e sedosa, queimar sem desprender fumaça, fagulha ou cheiro, além de ter cor negra e brilhante (BRITO, 1993).

3.3 Aprendizagem de máquina

Os resultados gerados pelos algoritmos de aprendizagem de máquina são obtidos a partir do aprendizado por indução, a qual trata-se de uma forma de inferência lógica que possibilita chegar a conclusões generalizadas com base em um conjunto de exemplos específicos. A inferência indutiva é um dos métodos mais utilizados para fornecer novas informações e prever eventos futuros (MONARD e BARANAUSKAS, 2005).

3.3.1 Aspectos da aprendizagem

A construção de um sistema de aprendizagem de máquina é influenciada por quatro questões (RUSSELL e NORVIG, 2004): a) os componentes do sistema que devem ser aprendidos; b) a forma de realimentação que será utilizada no aprendizado dos componentes; c) a representação dos componentes; e d) a disponibilidade de conhecimento anterior.

Os componentes do sistema podem incluir:

Um mapeamento direto de condições no estado atual para ações: a exemplo do treinamento de direção veicular, que quando o instrutor indica “freie!”, o agente poderá aprender uma regra de condição-ação para quando deve frear o veículo.

Uma forma de dedução das propriedades relevantes do mundo a partir da sequência de percepções: ao visualizar várias imagens sobre as quais é informado que mostram um modelo de carro, o agente poderá aprender a reconhecê-lo.

Informações sobre o modo como o mundo evolui e sobre os resultados de ações possíveis que o agente pode executar: testando ações e observando os resultados, por exemplo, observar os efeitos de uma freada brusca numa pista molhada.

Informações de utilidade indicando respostas às suas ações: após uma freada brusca numa pista molhada, o agente percebe algum tipo de dano no veículo, dando-lhe uma informação útil sobre a ação.

Informação de valores de ações indicando a resposta às suas ações.

Metas que descrevem classes de estados cujas realizações maximizam a utilidade do agente.

O tipo de realimentação disponível para a aprendizagem geralmente é o aspecto mais importante na determinação da natureza do problema sobre o qual o algoritmo deve atuar.

Segundo Dantas (2017) os tipos de realimentação ou aprendizagem de máquina podem ser divididos em três vertentes:

1) Aprendizagem supervisionada: é incluído no algoritmo o atributo de saída, que podem ser rótulos ou valores. O treinamento é realizado a partir de exemplos com atributos de entrada e saída conhecidos, sobre os quais o algoritmo estabelece um padrão. O aprendizado supervisionado ainda pode ser dividido em duas categorias:

- Classificação: os atributos de saída são rótulos, por exemplo, sim ou não;
- Regressão: os atributos de saída são numéricos, por exemplo, medida de rendimento da carbonização.

2) Aprendizagem não supervisionada: não existem atributos de saída ou classes pré-estabelecidas, cabendo ao algoritmo identificar padrões e classificar novos elementos dentro de grupos construídos na fase de treinamento.

3) Aprendizagem por reforço: não há classes previamente definidas, porém, existe um ambiente observável que auxilia o algoritmo a assimilar os padrões e tomar decisões após executar uma sequência de testes. Em outras palavras, o algoritmo não recebe informações bem definidas sobre as variáveis resposta, e sim “pistas” fornecidas pelo ambiente sobre as variações nas variáveis resposta.

Outro ponto importante é a representação das informações aprendidas, que também terá papel na determinação do tipo de algoritmo que será empregado. Existem diversos algoritmos para os mais variados fins, a exemplo de redes neurais para funções de utilidade em programas de jogos, sentenças lógicas proposicionais e de primeira ordem para todos os componentes de um agente lógico e inferências probabilísticas como redes bayesianas para teorias de tomada de decisão. A disponibilidade de conhecimento anterior é relevante, porém não é geral como as anteriores, sendo mais importante em estudos iniciais de novas áreas de conhecimento, nas quais as pesquisas com aprendizagem de máquina são sustentadas com informações empíricas (RUSSELL e NORVIG, 2004).

3.3.2 Paradigmas da aprendizagem

A representação das informações aprendidas está relacionada com os paradigmas de aprendizagem, que por sua vez dão origem a algoritmos ou métodos que seguem princípios

diferentes. Alguns dos paradigmas mais conhecidos são o simbólico, estatístico, sistemas baseados em protótipo, conexionista e genético. Os sistemas de aprendizagem simbólicos aprendem construindo representações simbólicas de um conceito analisando exemplos e contraexemplos, dentre os quais os mais estudados são os baseados em árvores e regras de decisão (PRATI, 2006).

O paradigma estatístico consiste em utilizar modelos estatísticos para obter uma boa aproximação do conceito induzido. Muito dos métodos estatísticos são paramétricos, onde são assumidos modelos e estimados os parâmetros a partir dos dados. O aprendizado Bayesiano é um dos sistemas estatísticos mais conhecidos (MONARD e BARANAUSKAS, 2005). Os sistemas baseados em protótipos realizam classificações baseadas em aprendizado obtido a partir de exemplos similares conhecidos.

Os sistemas conexionistas, segundo Braga, Carvalho e Ludermir (2007), são sistemas de processamento paralelo e distribuído não-algorítmicos que lembram o funcionamento dos neurônios humanos. Neste grupo, estão as redes neurais artificiais, que são compostas por unidades de processamento simples (nodos ou neurônios artificiais) que calculam determinadas funções matemáticas (normalmente não lineares). Já os classificadores genéticos, de acordo com Monard e Baranauskas (2005), trabalham com uma população de elementos que competem entre si para realizar a predição, ao passo que os mais fortes se multiplicam em detrimento dos mais fracos que são descartados. Nota-se que o paradigma genético faz uma clara analogia a teoria evolutiva de Darwin.

3.3.3 Árvores de decisão

As árvores de decisão podem ser para classificação (função de valores discretos) ou regressão (função de valores contínuos), daí o termo CART (Classification and Regression Trees) popularmente utilizado na literatura para se referir ao método. Os algoritmos de árvores de decisão são modelos não lineares de aprendizagem, nos quais a inferência sobre a variável-alvo é realizada a partir de regras simples de decisão, definidas com base nos dados de treinamento. Uma árvore de decisão chega a uma conclusão após executar uma sequência de testes. Cada nó interno de uma árvore corresponde ao teste do valor de uma das propriedades, sendo as ramificações que partem do nó referentes aos valores possíveis do teste. Os nós-folha da árvore representam os valores ou a decisão resultante do teste (Figura 1).

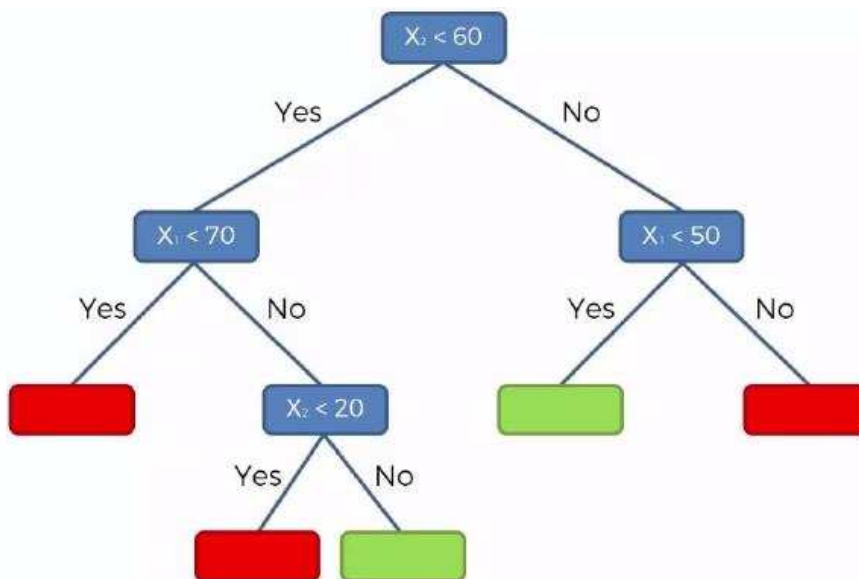


Figura 1 – Exemplo de árvore de decisão.

Fonte: Souza (2017).

Os modelos realizam análises matemáticas das variáveis independentes (valores de features) que resultam na divisão dos dados em subgrupos, processo conhecido como *splitting* (Figura 2). O *splitting* é um processo recursivo, que só termina quando todos os subgrupos estão bem definidos, de modo que os elementos dentro dos subgrupos possuam o mesmo valor da variável-alvo (SOUZA, 2017).

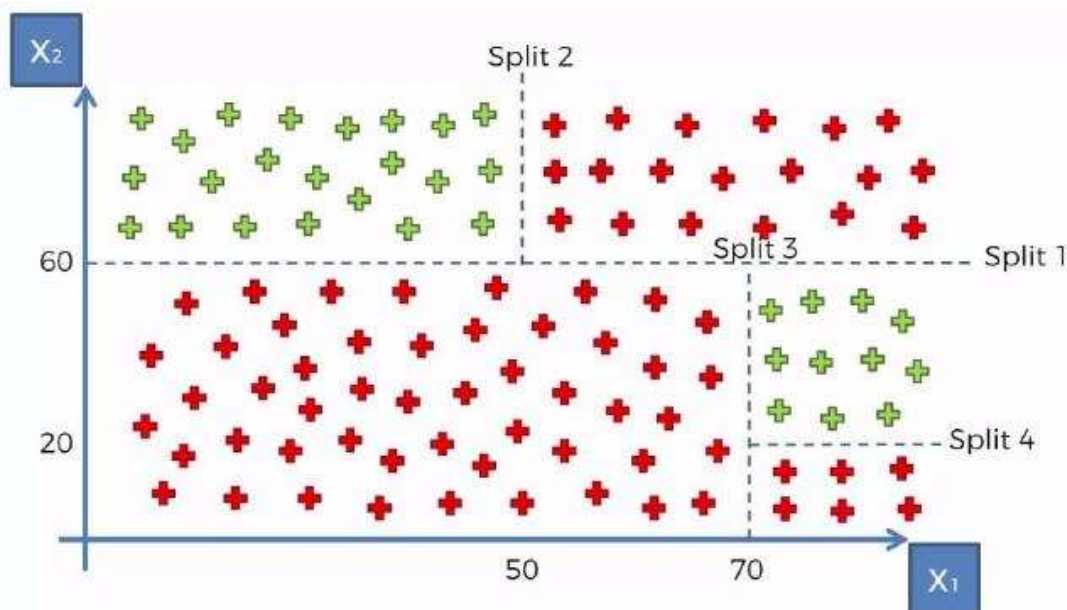


Figura 2 – Processo *splitting* de uma árvore de decisão com duas variáveis independentes.

Fonte: Souza (2017).

O fator mais importante na construção de uma árvore de decisão é a escolha da variável que será utilizada como nó da árvore. A escolha deve ser feita de tal modo que as divisões das árvores sejam as mais “puras” possíveis, ou seja, que os elementos classificados em um mesmo conjunto sejam os mais homogêneos possível. Pode ser usado como critério o cálculo do ganho de informação, conforme descrito por Russell e Norvig (2004):

$$\text{Ganho}(A) = I - \text{Restante}$$

$$\text{Restante}(A) = \sum_{i=1}^n \frac{p_i + n_i}{p + n} I$$

$$I = -\frac{p}{p+n} \log_2 \frac{p}{p+n} - \frac{n}{p+n} \log_2 \frac{n}{p+n}$$

ou

$$I = \sum_{i=1}^n -P(v_i) \log_2 P(v_i),$$

sendo Ganho (A) o ganho de informação com o novo requisito, I a quantidade de informações contidas em uma resposta correta, Restante (A) é a quantidade de informações sem o novo requisito, p é o número de exemplos positivos, n é o número de exemplos negativos e $P(v_i)$ é a probabilidade de respostas possíveis.

Após a elaboração da árvore, Carvalho (2014) cita a necessidade de realizar “podas”, visto que esse tipo de algoritmo é suscetível ao overfitting (super ajuste). Quando há super ajuste, ruídos nos dados de treinamento podem gerar ramos e nós que irão atrapalhar a generalização de novas classificações. Normalmente modelos com overfitting apresentam desempenho muito bom com os dados de treinamento, porém, a acurácia com os dados de teste se torna baixa.

3.3.4 Random trees e random forest

Random tree é uma árvore selecionada aleatoriamente a partir de um conjunto de árvores possíveis equiprováveis com m variáveis escolhidas ao acaso em cada nó. O aumento na

acurácia da classificação tende a ser maior com o aumento do número de random trees, que selecionam a classe mais popular e dão origem às random forests (BREIMAN, 2001). Random forest é um algoritmo que combina várias árvores de decisão que dependem de vetores aleatórios com a mesma distribuição para todas as árvores (OSHIRO, 2013). Breiman (2001) define formalmente esse algoritmo como um classificador composto por um conjunto de árvores classificadoras $\{h(\mathbf{x}, \boldsymbol{\theta}_k), k = 1, 2, \dots, n\}$, em que $\boldsymbol{\theta}_k$ são vetores aleatórios independentes e identicamente distribuídos e cada árvore seleciona a classe mais popular para a inclusão de \mathbf{x} .

Em termos práticos, o random forest é uma metodologia de árvores de decisão que fornece classificações ou estimativas baseadas nas médias de todas as árvores. É o método de aprendizagem de máquina mais difundido atualmente, por apresentar boa performance com eventos lineares e não lineares, além de baixa variância, sendo demonstrado que possui taxa de acerto superior à regressão logística e árvore de decisão simples (SOUZA, 2017). Brokamp et al. (2017) também cita a robustez do random forest em situações de amostras pequenas ($20 < n < 40$) e grande conjunto de preditores ($50 < p < 500$), que normalmente geram problemas com multicolinearidade e inviabiliza modelos de regressão convencionais.

De acordo com Oshiro (2013), random forests aplicam o método bootstraps sobre a base de dados de treinamento original para gerar amostras aleatórias de conjuntos de treinamento para cada árvore aleatória (Figura 3). Cada árvore é construída utilizando um subconjunto de treinamento e variáveis selecionados aleatoriamente. As variáveis selecionadas aleatoriamente para cada nó da árvore são avaliadas e comparadas, sendo escolhida a melhor para basear a divisão do nó.

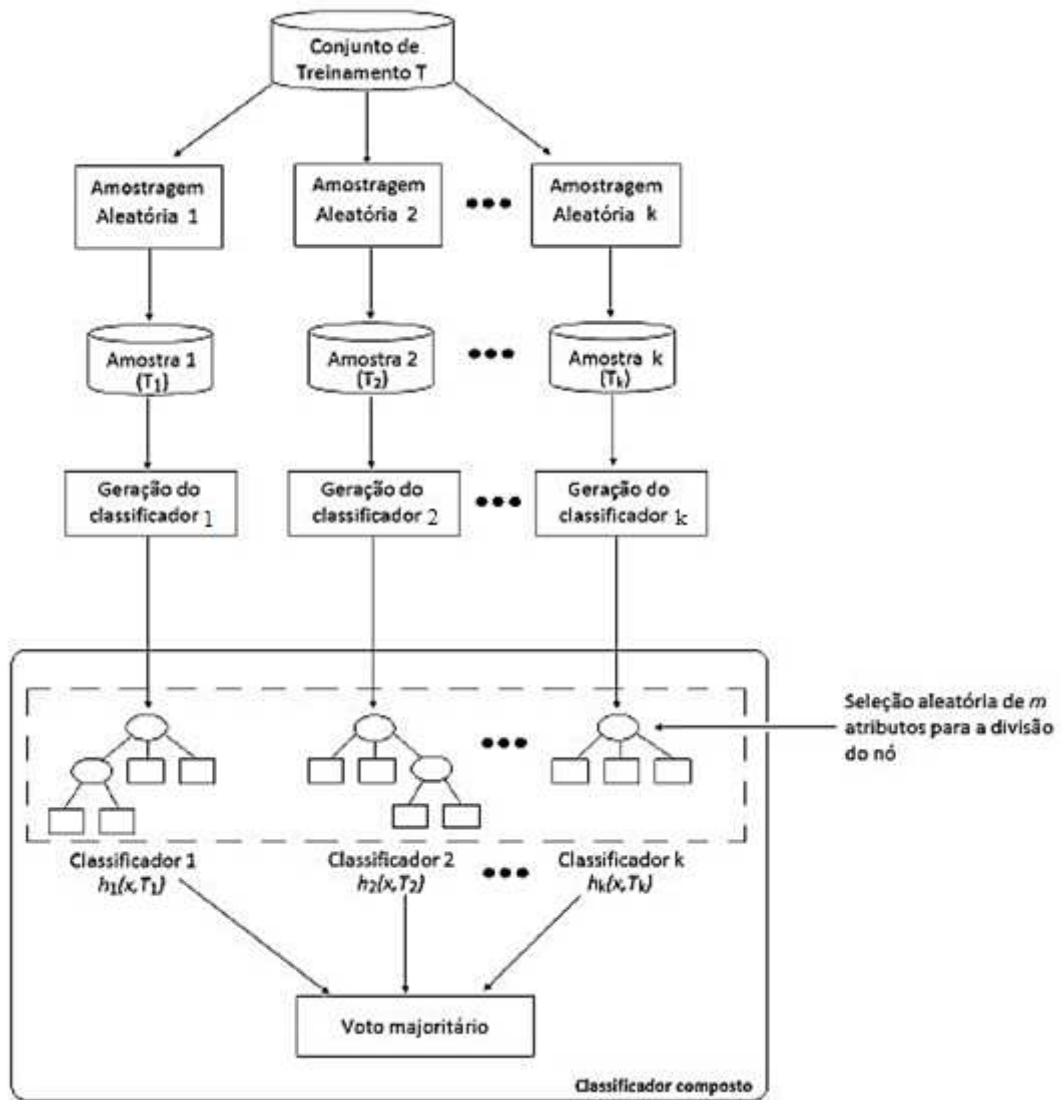


Figura 3 – Exemplo de funcionamento do Random forest.

Fonte: Oshiro (2013).

3.3.5 Random forest para regressão

Os algoritmos de random forests para regressão assumem valores de saída numéricos, ao contrário dos rótulos de classe. Breiman (2001) descreve os teoremas 1 e 2 que baseiam a construção de árvores eficientes. Definindo o erro médio quadrático generalizado de qualquer preditor numérico $h(\mathbf{x})$ como $E_{\mathbf{X},Y}(Y - h(\mathbf{X}))^2$ e assumindo-se que o conjunto de treino é selecionado aleatoriamente a partir do vetor aleatório Y, \mathbf{X} , define-se os teoremas:

Teorema 1 – Como o número de árvores na floresta tende ao infinito, então:

$$E_{X,Y}(Y - av_k h(\mathbf{X}, \theta_k))^2 \rightarrow E_{X,Y}(Y - E_\theta h(\mathbf{X}, \theta))^2.$$

Considerando que $E_{X,Y}(Y - E_\theta h(\mathbf{X}, \theta))^2$ é o erro generalizado da floresta (E(floresta)) e $E_\theta E_{X,Y}(Y - h(\mathbf{X}, \theta))^2$ o erro médio generalizado das árvores (E(árvores)), tem-se:

Teorema 2 – Para todo θ , $EY = E_\theta h(\mathbf{X}, \theta)$, então:

$$E(\text{floresta}) \leq \bar{\rho} E(\text{árvores}),$$

em que $\bar{\rho}$ é a correlação ponderada entre os resíduos $Y - h(\mathbf{X}, \theta)$ e $Y - h(\mathbf{X}, \theta')$, onde θ e θ' são independentes. O teorema 2 indica que a acurácia das florestas de regressão depende da baixa correlação entre os resíduos e de erros reduzidos das árvores. O random forest diminui o erro médio das árvores por meio do fator $\bar{\rho}$.

3.3.6 Support vector machine – SVM

As SVM foram preconizadas por Cortes e Vapnik (1995) e tem sua base na Teoria do Aprendizado Estatístico, ou seja, fazem parte do paradigma de aprendizagem estatístico. A ilustração mais didática da ideia por trás da SVM pode ser observada na Figura 4, onde é definido um hiperplano ideal para a separação de duas classes por meio da maximização da margem entre as classes (MEYER, 2018).

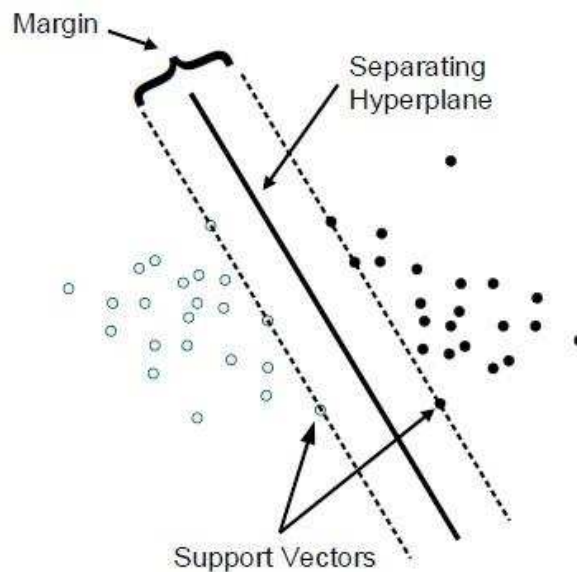


Figura 4 – Ilustração de uma support vector machine linear para classificação.

Fonte: Meyer (2018).

Os pontos que ficam no limite da margem entre as classes são os vetores de suporte e a linha que separa as classes representa a divisão ideal. Os objetos ou fenômeno para o qual se pretende fazer previsões são representados por rótulos ou classes, que, quando são discretos tem-se um problema de classificação, enquanto que no caso de rótulos contínuos trata-se de regressão e a técnica passa a ser chamada de support vector regression (MITCHELL, 1997).

Algumas das principais características da SVM que a torna atrativa para análise de dados são citadas por Smola et al. (1999):

Capacidade de generalização: o algoritmo SVM normalmente geram classificadores com boa capacidade de generalização, ou seja, sua capacidade de classificação ou estimação mantem-se satisfatória mesmo em dados que não fazem parte do conjunto de treinamento (menor ocorrência de overfitting);

Robustez em alta dimensionalidade: a SVM é robusta perante conjuntos de dados com muitos atributos, como no caso de imagens;

Convexidade da função objetivo: implica na otimização de uma função quadrática, a qual possui apenas um ponto de mínimo global;

Base teórica bem definida: a teoria por trás da SVM é bastante consolidada dentro da Matemática e Estatística.

Em revisão sobre a SVM, Lorena e Carvalho (2003), ressalta a capacidade desses algoritmos de obter altas taxas de precisão e generalização, além de apresentarem tempo de treinamento menores que várias técnicas de aprendizagem de máquina. Os autores ainda afirmam que os classificadores da SVM adquirem e processam dados de uma forma que não é facilmente interpretável, o que representa uma desvantagem em relação à técnicas como árvores de decisão, as quais organizam as informações em estruturas diretamente interpretável.

3.4 Análise multivariada

Análise multivariada pode ser entendida como qualquer técnica estatística que analisa de forma simultânea mais de uma variável referente a um objeto de estudo. Vale ressaltar que o conceito correto de análise multivariada reside na combinação de variáveis, ou seja, não apenas na avaliação de múltiplas variáveis ou observações (HAIR JR. et al., 2009).

Em experimentos que envolvem apenas dois grupos ou tratamentos, Hummel e Sligo (1971) citam três alternativas para lidar com dados multivariados com p médias igual a 2. Uma das formas é utilizar análise de variância (ANOVA) e testes univariados. Outra metodologia,

sugerida por Cramer and Bock (1966), baseia-se na análise concomitante de todas as variáveis envolvidas utilizando testes apropriados, a exemplo do T^2 de Hotelling, seguido por testes de médias univariados para identificar em qual variável houve diferença entre os grupos. Em adição à Cramer and Bock (1966), Morrison (1967) propôs utilizar intervalos de confiança simultâneos para verificar especificamente onde ocorrem as diferenças entre os grupos. Entretanto, em pesquisas envolvendo p médias maior que 2 o uso dos procedimentos supracitados não é possível, sendo necessário recorrer a outras metodologias de análise multivariada.

3.4.1 Análise de variância multivariada (MANOVA)

A análise de variância multivariada é uma expansão da análise de variância univariada sob enfoque de modelos lineares, desenvolvida para analisar as diferenças entre duas ou mais medidas métricas dependentes a partir de variáveis categóricas (HAIR JR. et al., 2009). Algebricamente a MANOVA e ANOVA podem ser entendidas da seguinte forma:

Análise de variância univariada

$$Y_1 = X_1 + X_2 + X_3 + \dots + X_n$$

(métrica) (categórica)

Análise de variância multivariada

$$Y_1 + Y_2 + Y_3 + \dots + Y_n = X_1 + X_2 + X_3 + \dots + X_n$$

(métrica) (categórica)

Embora a utilização de testes univariados para cada variável envolvida na pesquisa seja uma opção intuitiva e viável, Haase e Ellis (1987) afirmam que este procedimento pode prejudicar seriamente a interpretação dos resultados ao aumentar a probabilidade de erro Tipo I (rejeitar H_0 quando esta é verdadeira), além de desconsiderar as possíveis correlações entre as variáveis. Hair Jr. et al. (2009) complementa citando que a MANOVA tem a capacidade de detectar diferenças “combinadas” não verificadas em testes univariados, não obstante, quando a dimensão dos dados não ultrapassa cinco variáveis dependentes o poder estatístico dos testes da MANOVA equivale ao da ANOVA.

Fazendo uma breve análise sobre o aumento no erro Tipo I global ao realizar um conjunto de testes univariados, Sartorio (2008) cita o exemplo da realização de dez testes cada

um ao nível de 0,05 de significância, ou seja, individualmente cada teste inclui uma probabilidade de 5% de ocorrência de erro Tipo I, no entanto, ao considerar os dez testes essa probabilidade passa para 40%: $(1 - P[\text{não rejeitar nenhuma hipótese nula}] = 1 - 0,95^{10} = 0,401)$. Embora seja possível ajustar o nível de significância por meio de correções (BONFERRONI, 1936; HOLM, 1979; HOCHBERG, 1988; HOMMEL, 1988; BENJAMINI E HOCHBERG, 1995) para controlar os erros Tipo I e Tipo II globais, considerando todos os aspectos é mais conveniente utilizar técnicas multivariadas.

Para a aplicação da MANOVA faz-se necessário atender pressupostos equivalentes aos da ANOVA para o caso multivariado, ou seja, independência dos erros, distribuição normal multivariada, homogeneidade nas matrizes de variância/covariância e não ocorrência de multicolinearidade elevada entre variáveis dependentes (HAIR JR. et al., 2009).

Na prática, é de comum conhecimento que dados experimentais normalmente não atendem perfeitamente todas as pressuposições, especialmente no caso multivariado. Nesse contexto, Mardia (1971) demonstra que os testes associados à MANOVA são robustos à falta de normalidade, porém, testes de homogeneidade de covariância são sensíveis à não normalidade.

Ao adentrar no campo multivariado todos os procedimentos matemáticos passam pelos princípios do Modelo Linear de Gauss-Markov, sendo assim, o modelo estatístico se torna $\mathbf{Y} = \mathbf{XB} + \mathbf{E}$, em que \mathbf{Y} é a matriz de observações, \mathbf{X} é a matriz de delineamento, \mathbf{B} é a matriz de parâmetros e \mathbf{E} a matriz de erros. Sendo assim, a estimação dos parâmetros é realizada pela minimização da soma de quadrados e produtos do resíduo (\mathbf{E}), ou seja, minimizando $\mathbf{Q} = \mathbf{E}^T \mathbf{E} = (\mathbf{Y} - \mathbf{XB})^T - (\mathbf{Y} - \mathbf{XB})$. A sequência do procedimento gera uma matriz \mathbf{T} e outra \mathbf{H} referente a soma de quadrados e produtos totais e de efeito dos tratamentos, respectivamente, no caso de um delineamento inteiramente casualizado (FERREIRA, 2011). A matriz \mathbf{E} de resíduos é obtida pela diferença entre \mathbf{T} e \mathbf{H} .

Após a obtenção das matrizes de somas de quadrados e produtos, a avaliação da hipótese nula para os vetores de médias é realizada aplicando-se testes específicos, os quais são conhecidos como teste de Pillai, teste de Hotelling Lawey, teste de Roy e o mais usado teste de Wilks (Λ):

$$\Lambda = \frac{\det(\mathbf{H})}{\det(\mathbf{T})},$$

em que $\det(\mathbf{H})$ e $\det(\mathbf{T})$ são os determinantes da matriz \mathbf{H} e \mathbf{T} , respectivamente. Como os testes são calculados por metodologias distintas, os resultados em relação a hipótese testada também podem variar. Os valores calculados das estatísticas podem ser comparados utilizando as tabelas específicas de cada teste, ou, mais convencionalmente, podem ser convertidos à valores aproximados da estatística F de Snedecor (GOMES, 2009).

3.4.2 Função discriminante de Fisher

A MANOVA assim como a ANOVA são procedimentos que possibilitam verificar se ao menos algum dos grupos se diferencia dos demais, de tal modo que há a necessidade da aplicação de metodologias complementares para a comparação de médias e análise mais detalhada dos resultados. No caso da MANOVA, não existe um meio direto para comparação de médias, sendo uma das alternativas a utilização da função discriminante de Fisher (FISHER, 1936), ou ainda chamada de função discriminante canônica.

A aplicação inicial da função discriminante de Fisher, foi para análise discriminante, ou seja, o objetivo é encontrar funções lineares que combinem todas as variáveis de interesse, maximizando a diferença entre as populações pré-definidas, de tal forma que seja possível discriminar novos indivíduos dentro das populações pré-definidas a partir das mesmas variáveis incluídas nas funções. O ponto principal para a discriminação das populações é estabelecer funções das variáveis observadas que expliquem e maximizem as diferenças entre as populações a partir da nova variável criada (CAMPOS, 2012). Em resumo, cria-se uma regra de classificação a partir de um algoritmo de aprendizagem de máquina, definida por McLachlan (2012) como um procedimento de aprendizagem supervisionada.

Para uma melhor compreensão sobre a análise discriminante, Hair Jr. (2009) faz um paralelo com a MANOVA e com a regressão linear. O autor afirma que a função discriminante seria comparável a uma MANOVA “invertida”, visto que na MANOVA tem-se como dependentes as variáveis métricas e como independente(s) variável(eis) não métrica(s), enquanto na análise discriminante tem-se o oposto. Já em relação a regressão linear, assim como numa função discriminante, trata-se de uma combinação linear de variáveis independentes métricas que deverão explicar uma variável dependente, ao passo que a diferença básica entre as duas técnicas é que a variável dependente no contexto de regressão é métrica, enquanto na análise discriminante será categórica.

É possível encontrar na literatura diferentes aplicações das funções discriminantes, a exemplo do trabalho de Araujo, Nascimento e Zucchi (1998) que utilizaram funções

discriminantes para classificar quatro espécies *Anastrepha* do grupo *fraterculus* (moscas-das-frutas) com base nas medidas do acúleo. Outro estudo que utilizou funções discriminantes foi o de Benites et al. (2010), que aplicaram a técnica para discriminar solos sob diferentes usos a partir de atributos da matéria orgânica. Já Nanni, Demattê e Fiorio (2004) aplicaram a análise discriminante para classificar solos por meio de medidas da resposta espectral ao nível do solo.

Para a realização da análise discriminante, em tese, deve-se assumir pressupostos similares aos da MANOVA, conforme comentam Wahl e Kronmal (1977) sobre o frequente uso da técnica devido a sua relativa simplicidade, embora seja necessário assumir homogeneidade de matrizes de variância/covariância ou, quando não há homocedasticidade, recorrer à análise discriminante quadrática. Os autores ainda afirmam que quando as diferenças nas covariâncias são pequenas, os dois tipos de análise discriminante são equivalentes, porém, quando há forte heterocedasticidade, a análise discriminante quadrática tem desempenho largamente melhor.

Comparando a forma linear e quadrática da análise discriminante em condição de matrizes de covariâncias heterogêneas e normalidade multivariada preservada, Gilbert (1969) constatou que a forma quadrática consegue utilizar discrepâncias na variância para reduzir a probabilidade de erros de classificação. Assim sendo, o autor demonstra que a análise discriminante quadrática se torna melhor que a forma linear proporcionalmente ao aumento no número de variáveis, visto que poderão existir mais variáveis discrepantes para contribuir com o desempenho da forma quadrática.

Quanto à normalidade, estudos mostraram que para populações não normais o resultado da análise discriminante linear é significativamente afetado, sendo indicada a transformação dos dados para aproximar da normalidade e melhorar a acurácia da análise (KRZANOWSKI, 1977). Vale destacar que as pesquisas encontradas sobre a importância dos pressupostos estatísticos nas funções discriminantes referem-se ao seu uso em classificações.

Uma metodologia alternativa de função discriminante é apresentada por Gomes (2009) definida pelo autor como variável canônica. Nesta metodologia a função discriminante não é usada para classificações e sim para transformar as variáveis observadas em uma ou mais variáveis canônicas que irão reter a maior parte da variabilidade dos dados (p variáveis canônicas $\leq p$ variáveis observadas), ou seja, é uma redução dimensional:

$$Z = b_1X_1 + b_2X_2 + \dots + b_pX_p,$$

em que Z representa a variável canônica, b_1 , b_2 e b_p os coeficientes canônicos e X_1 , X_2 e X_p as variáveis observadas. Aplicando a função discriminante descrita por Gomes (2009) gera-se n valores da variável canônica referentes a cada parcela ou observação do experimento.

Os coeficientes da função discriminante são determinados, inicialmente, obtendo-se as raízes características (autovalores) da matriz resultante do produto entre as matrizes $\mathbf{E}^{-1}\mathbf{H}$. De posse dos autovalores, toma-se o maior para estabelecer o seguinte sistema de equações:

$$(H - \lambda_1 E) \begin{bmatrix} b_1 \\ b_2 \\ b_p \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$

sendo λ_1 o maior autovalor. Ao resolver o sistema de equações serão obtidos os autovetores referentes ao maior autovalor, os quais são os coeficientes da função discriminante canônica. Em adição, Gomes (2009) cita que é indicado normalizar os coeficientes para que a soma dos quadrados destes seja igual a um.

Após gerar a variável canônica para cada observação, os tratamentos podem ser comparados por suas respectivas médias canônicas aplicando técnicas univariadas convencionais. É importante ressaltar que a função discriminante definida a partir do maior autovalor tem o objetivo de maximizar a diferença entre os grupos, em outras palavras, ao aplicar a ANOVA na variável canônica o valor da estatística F será máximo (CAMPOS, 2012).

REFERÊNCIAS BIBLIOGRÁFICAS

ARAUJO, E.L.; NASCIMENTO, F.M.; ZUCCHI, R.A. Utilização da análise discriminante em estudos taxonômicos de moscas-das-frutas do gênero *Anastrepha* Schiner, 1868 (Diptera: Tephritidae). **Scientia Agricola**, v.55, n.1, p.105-110, 1998.

ASSIS, T. F. Production and use of Eucalyptus hybrids for industrial purposes. In: QFRI/CRC-SPF SYMPOSIUM, 2000, Noosa. **Proceedings...** Brisbane: Department of Primary Industries, 2000. p. 63-75.

BARROSO, R.C. **Redução do teor de cinzas dos finos de carvão vegetal por concentração gravítica a seco**. 110f. 2007. Dissertação (Mestrado em Engenharia Metalúrgica e de Minas) – Universidade Federal de Minas Gerais, Belo Horizonte, 2007.

BENITES, V. de M.; MOUTTA, R. de O.; COUTINHO, H.L. da C.; BALIEIRO, F. de C. Análise discriminante de solos sob diferentes usos em área de mata atlântica a partir de atributos da matéria orgânica. **Revista Árvore**, v.34, n.4, p.685-690, 2010.

BENJAMINI, Y.; HOCHBERG, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. **Journal of the Royal Statistical Society: Series B**, v.57, p.289-300, 1995.

BONFERRONI, C.E. Teoria statistica delle classi e calcolo delle probabilità. **Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze**, v. 8, p.3-62, 1936.

BORTOLETTO JÚNIOR, G. Produção de compensados com 11 espécies do gênero *Eucalyptus*, avaliação das suas propriedades físico-mecânicas e indicações para utilização. **Scientia Forestalis**, n.63, p.65-78, 2003.

BRAGA, A. de P.; CARVALHO, A.C.P. de L.F.; LUDERMIR, T.B. **Redes neurais artificiais: teoria e aplicações**. 2. ed. Rio de Janeiro: LTC Ed., 2007. 226p.

BREIMAN, L. Random forests. **Machine Learning**, v.45, p.5–32, 2001.

BRITO, L.O.; BARRICHELO, L.E.G. Correlações entre características físicas e químicas da madeira e a produção de carvão vegetal: I. Densidade e teor de lignina da madeira de eucalipto. **IPEF**, n.14, p.9-20, 1977.

BRITO, J.O.; BARRICHELO, L.E.G.; SEIXAS, F.; MIGLIORINI, A.J.; MURAMOTO, M.C. Análise da produção energética e de carvão vegetal de espécies de eucalipto. **IPEF**, n.23, p.53-56, 1983.

BRITO, J.O. Carvão vegetal no Brasil: gestões econômicas e ambientais. **Estudos Avançados**, v.4, n.9, p.221-227, 1990.

BRITO, J.O. Reflexões sobre qualidade do carvão vegetal para uso siderúrgico. **Circular Técnica IPEF**, n.181, 1993.

BROKAMP, C.; JANDAROV, R.; RAO, M.B.; LEMASTERS, G.; RYAN, P. Exposure assessment models for elemental components of particulate matter in an urban environment: A comparison of regression and random forest approaches. **Atmospheric Environment**, v.151, p.1–11, 2017.

CAMPOS, K.A. **Função discriminante de Fisher como alternativa à análise de variância multivariada**. 106f. 2012. Tese (Doutorado em Estatística e Experimentação Agropecuária) – Universidade Federal de Lavras, Lavras, 2012.

CARNEIRO, A. de C.O.; VITAL, B.R.; FREDERICO, P.G.U.; FIGUEIRÓ, C.G.; FIALHO, L. de F.; SILVA, C.M.S. da S. Caracterização energética das madeiras de clones de *Eucalyptus* cultivados em diferentes localidades. **Ciência da Madeira (Brazilian Journal of Wood Science)**, v.8, n.3, p.127-135, 2017.

CARVALHO, H.M. **Aprendizado de máquina voltado para mineração de dados: árvores de decisão**. 2014. 68f. Monografia (Graduação em Engenharia de Software) – Universidade de Brasília, Brasília, 2014.

CASTRO, A.F.N.M.; CASTRO, R.V.O.; CARNEIRO, A. de C.O.; LIMA, J.E. de; SANTOS, R.C. dos; PEREIRA, B.L.C.; ALVES, I.C.N. Análise multivariada para seleção de clones de eucalipto destinados à produção de carvão vegetal. **Pesquisa Agropecuária Brasileira**, v.48, n.6, p.627-635, 2013.

CORTES, C.; VAPNIK, V. Support-vector network. **Machine Learning**, v.20, p.1–25, 1995.

COUTINHO, A.R.; FERRAZ, E.S.B. Determinação da friabilidade do carvão vegetal em função do diâmetro das árvores e temperatura de carbonização. **IPEF**, n.38, p.33-37, 1988.

CRAMER, E.M.; BOCK, R.D. Multivariate analysis. **Review of Educational Research**, v.36, p.604-617, 1966.

DANTAS, P.B. **Utilização da aprendizagem de máquina e seleção de atributos para o diagnóstico de ceratocone a partir de parâmetros biomecânicos**. 2017. 64f. Dissertação (Mestrado Modelagem Computacional do Conhecimento) – Universidade Federal de Alagoas, Maceió, 2017.

FERREIRA, D.F. **Estatística multivariada**. 2 ed. rev. e ampl. Lavras: Ed. UFLA, 2011. 675p.

FISHER, R.A. The use of multiple measurements in taxonomic problems. **Annals of Eugenics**, v.7, n.2, p.179-188, 1936.

GALVAN, J. **Influência do tempo de secagem e estratificação da madeira na obtenção do carvão vegetal**. 51f. 2016. Monografia (Graduação em Engenharia Florestal) – Universidade Federal do Paraná, Curitiba, 2016.

GILBERT, E.S. The Effect of Unequal Variance-Covariance Matrices on Fisher's Linear Discriminant Function. **Biometrics**, v.25, n.3, p. 505-515, 1969.

GOMES, F.P. **Curso de estatística experimental**. 15 ed. Piracicaba: ESALQ, 2009. 451p.

HAASE, R.F.; ELLIS, M.V. Multivariate Analysis of Variance. **Journal of Counseling Psychology**, v. 34, n.4. p.404-413, 1987.

HAIR JR., J.F.; BLACK, W.C.; BABIN, B.J.; ANDERSON, R.E.; TATHAM, R.L. **Multivariate Data Analysis**. 7 ed. Upper Saddle River: Prentice Hall, 2009. 785p.

HOCHBERG, Y. A sharper Bonferroni procedure for multiple tests of significance. **Biometrika**, v.75, p.800-803, 1988.

HOLM, S. A simple sequentially rejective multiple test procedure. **Scandinavian Journal of Statistics**, v.6, p.65-70, 1979.

HOMMEL, G. A stagewise rejective multiple test procedure based on a modified Bonferroni test. **Biometrika**, v.75, p383-386, 1988.

HUMMEL, T.J.; SLIGO, J.R. Empirical comparison of univariate and multivariate analysis of variance procedures. **Psychological Bulletin**, v.76, n.1, p.49-57, 1971.

IBÁ. **Relatório** **2017**. Disponível em:<
http://iba.org/images/shared/Biblioteca/IBA_RelatorioAnual2017.pdf> Acesso: 11 de maio de 2018.

KRZANOWSKI, W.J. The Performance of Fisher's linear discriminant function under non-optimal conditions. **Technometrics**, v.19, n.2,p. 191-200, 1977.

LOBÃO, M. S. L.; CHAGAS, M. P.; COSTA, D. de S. P.; FERREIRA, A. T. B.; SETTE JR., C. R.; CARVALHO, I. L.; FO, M. T. Agrupamento de espécies florestais pela similaridade das características físico-anatômicas e usos da madeira. **Cerne**, v. 16, Suplemento, p. 97-105, 2010.

LORENA, A.C.; CARVALHO, A.C.P.L.F. de. **Introdução às Máquinas de Vetores Suporte (Support Vector Machines)**. São Carlos: Instituto de Ciências Matemáticas e de Computação, 2003. (Relatório Técnico N° 192)

LOUREIRO, B. A. **Análise energética da madeira e do carvão de híbridos entre *Corymbia citriodora* (Hook.) K.D. Hill & L.A.S. Johnson e *Corymbia torelliana* (F. Muell.) K.D. Hill & L.A.S. Johnson**. 2016. 87f. Dissertação (Mestrado em Ciência e Tecnologia da Madeira) – Universidade Federal de Lavras, Lavras, MG, 2016.

MARDIA, K.V. The effect of nonnormality on some multivariate tests and robustness to nonnormality in the linear model. **Biometrika**, v.58, n.1, p.105-121, 1971.

MCLACHLAN, G.J. Discriminant analysis. **Wiley Interdisciplinary Reviews: Computational Statistics**, v.4, p.421–431, 2012.

MEYER, D. **Support vector machine**. 2018. Disponível em:<
<ftp://ftp.cse.yzu.edu.tw/CRAN/web/packages/e1071/vignettes/svmdoc.pdf>> Acesso em: 29 de nov. de 2018.

MITCHELL, T. M. **Machine Learning**. McGraw Hill Science/Engineering/Math, 1997. 432p.

MONARD, M.C.; BARANAUSKAS, J.A. Conceitos sobre aprendizado de máquina. In: REZENDE, S.O. (Org.). **Sistemas inteligentes: fundamentos e aplicações**. Barueri: Manole, 2005. cap. 4, p.39-56.

MORRISON, D. F. **Multivariate statistical methods**. New York: McGraw-Hill, 1967.

MURPHY, K. **Machine Learning: a probabilistic perspective**. Cambridge: The MIT Press, 2012.

NANNI, M.R.; DEMATTÊ, J.A.M.; FIORIO, P.R. Análise discriminante dos solos por meio da resposta espectral no nível terrestre. **Pesquisa Agropecuária Brasileira**,v.39, n.10, p.995-1006, 2004.

NEVES, T.A.; PROTÁSIO, T. de P.; COUTO, A.M.; TRUGILHO, P.F.; SILVA, V.O.; VIEIRA, C.M.M. Avaliação de clones de Eucalyptus em diferentes locais visando a produção de carvão vegetal. **Pesquisa Florestal Brasileira**, v.31,n.68,p.319-330, 2011.

OSHIRO, T.M. **Uma abordagem para a construção de uma única árvore a partir de uma Random Forest para a classificação de bases de expressão gênica**. 2013. 110f. Dissertação (Mestrado em Bioinformática) – Universidade de São Paulo, Ribeirão Preto, 2013.

PEREIRA, B.L.C. **Qualidade da madeira de Eucalyptus para a produção de carvão vegetal**. 2012. 93f. Dissertação (Mestrado em Ciência Florestal) – Universidade Federal de Viçosa, Viçosa, MG, 2012.

PEREIRA, B.L.C.; CARNEIRO, A.C.O.; CARVALHO, A.M.M.L; TRUGILHO, P.F; MELO, I.C.N.A.; OLIVEIRA, A.C. Estudo da degradação térmica da madeira de Eucalyptus através de termogravimetria e calorimetria. **Revista Árvore**, v.37, n.3, p.567-576, 2013.

PEREIRA, B.L.C.; CARVALHO, A.M.M.L; OLIVEIRA, A.C.; SANTOS, L.C.; CARNEIRO, A.C.O.; MAGALHÃES, M.A. Efeito da carbonização da madeira na estrutura anatômica e densidade do carvão vegetal de Eucalyptus. **Ciência Florestal**, v. 26, n. 2, p. 545-557, 2016.

PRATI, R.C. **Nova abordagens em aprendizado de máquina para a geração de regras, classes desbalanceadas e ordenação de casos**. 2006. 189f. Tese (Doutorado em Ciências de Computação e Matemática Computacional) – Universidade de São Paulo, São Carlos, 2006.

PROTÁSIO, T. de P.; TRUGILHO, P. F.; NEVES, T. A.; VIEIRA, C. M. M. Análise de correlação canônica entre características da madeira e do carvão vegetal de Eucalyptus. **Scientia Forestalis**, v.40, n.95, p.317-326, 2012.

PROTÁSIO, T. de P.; COUTO, A. M.; REIS, A. A. dos; TRUGILHO, P. F. Seleção de clones de Eucalyptus para a produção de carvão vegetal e bioenergia por meio de técnicas univariadas e multivariadas. **Scientia Forestalis**, v.41, n.97, p. 015-028, 2013.

RUSSEL, S.; NORVIG, P. **Inteligência artificial**. Tradução de PubliCare Consultoria. 2. ed., 4. reimpr. Rio de Janeiro: Elsevier, 2004. 1021p.

SANTOS, S. de F. de O. M.; HATAKEYAMA, K. Processo sustentável de produção de carvão vegetal quanto aos aspectos: ambiental, econômico, social e cultural. **Produção**, v.22, n.2, p. 309-321, 2012.

SARTORIO, S.D. **Aplicações de técnicas de análise multivariada em experimentos agropecuários usando o software R**. 2008. 130f. Dissertação (Mestrado em Agronomia) – Universidade de São Paulo, Escola Superior de Agricultura “Luiz de Queiroz”, Piracicaba, 2008.

SILVA, D.A. da. Qualidade do carvão vegetal produzido com madeiras da região de Manaus em fornos de alvenaria. **Acta Amazônica**, v.18, n.1-2, p.163-178, 1988.

SIMIONI, F. J.; MOREIRA, J. M. M. Á. P.; FACHINELLO, A. L.; BUSCHINELLI, C. C. de A.; MATSUURA, M. I. da S. F. Evolução e concentração da produção de lenha e carvão vegetal da silvicultura no Brasil. **Ciência Florestal**, v.27, n.2, p.731-742, 2017.

SMOLA, A.J.; BARLETT, P.; SCHÖLKOPF, B.; SCHUURMANS, D. (Eds.) **Advances in large margin classifiers**. Cambridge: MIT Press, 1999. 422p.

SOARES, V.C.; BIANCHI, M.L; TRUGILHO, P.F.; PEREIRA, A.J.; HÖFLER, J. Correlações entre as propriedades da madeira e do carvão vegetal de híbridos de eucalipto. **Revista Árvore**, v.38, n.3, p.543-549, 2014.

SOARES, V.C.; BIANCHI, M.L; TRUGILHO, P.F.; HÖFLER, J.; PEREIRA, A. J. Análise das propriedades da madeira e do carvão vegetal de híbridos de eucalipto em três idades. **Cerne**, v.2, n.2, p.191-197, 2015.

SOUZA, L. A. M. de. **Aplicação de aprendizado de máquina para predição de prioridade em gestão de incidentes**. 2017. 61f. Monografia (Graduação em Sistemas de Informação) – Universidade Federal do Estado do Rio de Janeiro, Rio de Janeiro, 2017.

TRUEMAN, S.J.; RICHARDSON, D.M. Relationships between indole-3-butyric acid, photoinhibition and adventitious rooting of *Corymbia torelliana*, *C. citriodora* and F1 hybrid cuttings. **Tree and Forestry Science and Biotechnology**, v.2, n.1, p.26-33, 2008.

TRUGILHO, P. F.; VITAL, B.R.; REGAZZI, A.J.; GOMIDE, J.L. Aplicação da análise de correlação canônica na identificação de índices de qualidade da madeira de eucalipto para a produção de carvão vegetal. **Revista Árvore**, v.21, n.2, p.259-267, 1997.

TRUGILHO, P.F; LIMA, J.T. MORI, F.A.; LINO, A.L.L. Avaliação de clones de *Eucalyptus* para produção de carvão vegetal. **Cerne**, v.7, n.2, p.104-114, 2001.

WAHL, P.W.; KRONMAL, R.A. Discriminant functions when covariances are unequal and sample sizes are moderate. **Biometrics**, v.33, n.3, p. 479-484, 1977.

CAPÍTULO 1

APRENDIZAGEM DE MÁQUINA NO ESTUDO DA INFLUÊNCIA DAS PROPRIEDADES DA MADEIRA SOBRE A PRODUÇÃO DE CARVÃO VEGETAL

Resumo – O entendimento das relações entre as propriedades da madeira e do carvão vegetal possibilita a melhoria da produção de carvão, sendo assim, objetivou-se utilizar o algoritmo random forest na análise da influência das propriedades da madeira sobre a qualidade do carvão, além de avaliar a acurácia dos valores preditos em relação ao support vector regression e regressão linear múltipla. Foram mensuradas seis propriedades da madeira e seis do carvão de 13 clones de *Corymbia* com sete anos de idade. O estudo da influência das propriedades da madeira sobre o rendimento e qualidade do carvão foi realizado utilizando a medida de redução média da impureza do nó (soma de quadrados residual) do random forest e a correlação de cópulas. O random forest foi comparado ao support vector regression e regressão linear múltipla por meio do coeficiente de determinação, correlação linear entre valores observados e preditos, erro médio absoluto e raiz quadrada do erro quadrático médio. A acurácia do random forest foi maior que a do support vector regression e da regressão linear múltipla, principalmente em termos de coeficiente de determinação e correlação linear entre valores observados e preditos. O rendimento e qualidade do carvão vegetal de clones de *Corymbia* foram influenciados, principalmente, pelo teor de holocelulose, relação cerne/alburno e densidade básica da madeira. A densidade relativa aparente do carvão foi a variável em que o random forest atingiu o melhor nível de explicação da variabilidade em função das propriedades da madeira, porém, foi para o teor de carbono fixo que o algoritmo forneceu predições com menor erro

Palavras-chave: aprendizagem supervisionada, *Corymbia*, random forest, support vector regression, regressão linear múltipla.

MACHINE LEARNING IN STUDY OF THE INFLUENCE OF WOOD PROPERTIES ON PRODUCTION OF CHARCOAL

Abstract – The understanding of the relationships between the properties of wood and charcoal makes it possible to improve the production of coal. Thus, the objective was to use the random forest algorithm in analysis of the influence of wood properties on the quality of the charcoal,

besides evaluating the accuracy of predicted values in relation to support vector regression and multiple linear regression. Six wood properties and six of the charcoal of 13 *Corymbia* clones at seven years of age were measured. The study of the influence of the wood properties on the yield and quality of the charcoal was carried out using the mean reduction of the node impurity (residual squares sum) of the random forest and the copula correlation. The random forest was compared to the support vector regression and multiple linear regression by means of the determination coefficient, linear correlation between observed and predicted values, absolute mean error and square root of the mean square error. The accuracy of the random forest was greater than that of the support vector regression and the multiple linear regression, mainly in terms of determination coefficient and linear correlation between observed and predicted values. The charcoal yield and quality of *Corymbia* clones was influenced mainly by holocellulose content, heartwood/sapwood ratio and basic density of the wood. The charcoal apparent relative density was the variable in which the random forest reached the best level of explanation of the variability as a function of the wood properties, however, it was for the fixed carbon content that the algorithm provided predictions with less error.

Keywords: supervised learning, *Corymbia*, random forest, support vector regression, multiple linear regression.

1 INTRODUÇÃO

A heterogeneidade da qualidade do carvão vegetal é um dos principais problemas enfrentados pelas indústrias siderúrgicas no controle do processo de redução do minério de ferro. As propriedades do carvão vegetal são influenciadas pelo processo de carbonização e pelas características do material de origem, tais como espécie ou material genético, características químicas da madeira e idade das árvores (PROTÁSIO et al., 2012; SOARES et al., 2015).

Dentre as características da madeira que influenciam na qualidade do carvão, Oliveira et al. (2010) destaca a densidade básica, poder calorífico, constituição química e teor de umidade como as mais importantes para a seleção de materiais adequados à produção de carvão. Soares et al. (2014) ressaltam que ainda é necessário aprofundar o conhecimento sobre as relações das propriedades da madeira com as do carvão vegetal, bem como compreender melhor as reações químicas que ocorrem na madeira durante o processo de carbonização.

Existe na literatura uma considerável diversidade de pesquisas que estudaram as relações entre as características da madeira e do carvão, trabalhos que utilizaram metodologias

simples como correlação linear de Pearson (VITAL et al., 1994; MEDEIROS NETO, OLIVEIRA e PAES, 2014), outros via regressão linear simples (BRITO e BARRICHELO, 1977; BRITO e BARRICHELO, 1980) e metodologias mais complexas como correlação canônica também já foram empregadas (TRUGILHO et al., 1997; TRUGILHO, LIMA e MORI, 2003; PROTÁSIO et al., 2012).

Um novo conceito de análise que associa princípios estatísticos com programação computacional vem crescendo e se estabelecendo em relação às demandas das pesquisas científicas, conhecido como aprendizagem de máquina. A ideia de aprendizagem de máquina pode ser traduzida simplesmente como a implementação de algoritmos que permitem ao computador reconhecer padrões, aprender e executar análises de forma automatizada (DANTAS, 2017), o que não é tão inovador quando se pensa que metodologias usuais a exemplo de análise de regressão e algumas técnicas multivariadas podem ser incluídas nesse contexto (BIAMONTE et al., 2017). A grande contribuição da aprendizagem de máquina está no incremento sistemático de programação, gerando algoritmos avançados e robustos como o random forest.

O algoritmo random forest é baseado em árvores de decisão/regressão, que, segundo Breiman (2001), fornece estimativas numéricas que são médias de todas as k árvores quando utilizado para regressão. O random forest tem se mostrado mais preciso que outros algoritmos como redes neurais e support vector machine (CARUANA, KARAMPATZIAKIS e YESSINALINA, 2008), fato que tem levado ao crescente uso da técnica nas mais diversas áreas de aplicação, como em pesquisas com sensoriamento remoto (GIROLAMO NETO et al., 2015), em estudos da física do solo (CARVALHO JÚNIOR et al., 2016), em dados da área da saúde (LENTO, 2017), no campo da energia eletromecânica (LOPES, 2017), além de outras.

Portanto, objetivou-se utilizar o algoritmo random forest (RF) na análise da influência das propriedades da madeira sobre a qualidade do carvão, além de avaliar a acurácia dos valores preditos em relação ao support vector regression (SVR) e regressão linear múltipla (RLM).

2 MATERIAL E MÉTODOS

A presente pesquisa foi realizada na ArcelorMittal BioFlorestas, empresa pertencente ao Grupo ArcelorMittal, no Município de Dionísio, Minas Gerais. O clima da região é classificado como subtropical úmido, cuja temperatura média anual varia entre 20 a 23° C e a precipitação média é de 1100 a 1400 mm (Motta et al., 1996).

O plantio de 13 genótipos do gênero *Corymbia* foi realizado em 2008, em área de Latossolo Vermelho-Amarelo com espaçamento entre plantas de 3 x 2,5 m. Os plantios tinham sete anos de idade quando foram coletadas as amostras para as análises das propriedades da madeira (relação cerne/alburno, densidade básica, poder calorífico superior, lignina total, teor de holocelulose e teor de extrativos) e do carvão (poder calorífico superior, rendimento gravimétrico da carbonização, friabilidade, densidade relativa aparente, teor de materiais voláteis e teor de carbono fixo).

2.1 Propriedades anatômicas e físicas da madeira

A relação cerne/alburno foi mensurada conforme realizado no trabalho de Castro (2011). A densidade básica foi mensurada conforme a norma ABNT NBR 11941 (ABNT, 2003) e o poder calorífico superior foi determinado segundo a NBR 8633 (ABNT, 1984).

2.2 Análise da composição química elementar e estrutural da madeira

O teor de extrativos totais foi determinado de acordo com a norma TAPPI 204 cm-97 (TAPPI, 1997); a lignina total foi quantificada conforme Gomide e Demuner (1986) e Goldshimid (1971), respectivamente; enquanto que o teor de holocelulose foi calculado subtraindo-se de 100% os percentuais de lignina total e de extrativos.

2.3 Rendimento gravimétrico e propriedades do carvão vegetal

O rendimento gravimétrico da carbonização foi determinado pela relação entre massa de carvão vegetal e massa seca de madeira; a densidade relativa aparente do carvão foi determinada utilizando o método hidrostático, conforme Vital (1984); a friabilidade foi determinada de acordo com a metodologia proposta por Oliveira et al. (1982); os teores de materiais voláteis e cinzas foram determinados seguindo a NBR 8112 (ABNT, 1986); o teor de carbono fixo foi calculado subtraindo-se de 100% os teores de materiais voláteis e cinzas e o poder calorífico superior foi mensurado segundo metodologia da NBR 8633 (ABNT, 1984).

2.4 Construção dos algoritmos

As árvores de regressão do RF foram construídas utilizando o pacote randomForest do R (R CORE TEAM, 2018), para o qual definiu-se três parâmetros básicos (BREIMAN, 2002): número de árvores (ntree), variáveis selecionadas aleatoriamente por nó (mtry) e número de nós terminais (nodesize), optando-se por manter o default do pacote, ou seja, 500 árvores, quatro

variáveis por nó e cinco nós terminais. Para o treinamento do SVR foi utilizado o kernel da função de base radial, por meio da função `train` do pacote `caret` (R CORE TEAM, 2018).

As propriedades da madeira foram ranqueadas quanto a importância da sua influência sobre as propriedades do carvão utilizando a medida de redução média da impureza dos nós das árvores no algoritmo RF, calculada pela soma de quadrados residual. Adicionalmente, calculou-se o coeficiente de correlação de cópulas (r^*) entre propriedades da madeira e do carvão, o qual é um equivalente ao coeficiente linear de Pearson com o diferencial de mensurar também associações não lineares (DING e LI, 2013). Repetiu-se o treinamento das árvores em função apenas das variáveis mais importantes para dimensionar a influência dessas preditoras sobre o rendimento e qualidade do carvão por meio do coeficiente de determinação (R^2):

$$R^2 = \left[1 - \frac{n^{-1} \sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n y_i^2 - \frac{(\sum_{i=1}^n y_i)^2}{n}} \right] \cdot 100 ,$$

em que \hat{y}_i é a estimativa para a i -ésima observação (y_i).

Para comparar a acurácia do RF com o SVR e RLM, os dados foram divididos aleatoriamente em 75% para treinamento e 25% para a validação cruzada. A comparação das metodologias foi baseada na validação cruzada seguindo os critérios usados nos trabalhos de Hallak e Pereira Filho (2011), Aitkenhead e Coull (2016), Carvalho Junior et al. (2016) e Malone et al. (2016), em que foram calculados o coeficiente de determinação (R^2), a correlação linear de Pearson entre valores observados e preditos (r), o erro médio absoluto (EMA) e a raiz quadrada do erro quadrático médio (REQM):

$$EMA = \left(\frac{n^{-1} \sum_{i=1}^n |y_i - \hat{y}_i|}{\bar{Y}} \right) \cdot 100$$

$$REQM = \left\{ \frac{[n^{-1} \sum_{i=1}^n (y_i - \hat{y}_i)^2]^{1/2}}{\bar{Y}} \right\} \cdot 100 ,$$

em que \bar{Y} é a média da variável.

Para o cálculo dos critérios de acurácia utilizou-se a função `postResample` do pacote `caret` (R CORE TEAM, 2018).

3 RESULTADOS E DISCUSSÃO

3.1 Propriedades da madeira e rendimento/qualidade do carvão vegetal

Verificou-se que o teor de holocelulose da madeira foi a variável mais influente sobre o rendimento da carbonização, teor de materiais voláteis e friabilidade do carvão. A relação

cerne/alburno foi a variável mais importante para o teor de carbono fixo do carvão. No caso da densidade relativa aparente e poder calorífico superior do carvão, a densidade básica da madeira foi a propriedade de maior influência (Figura 1).

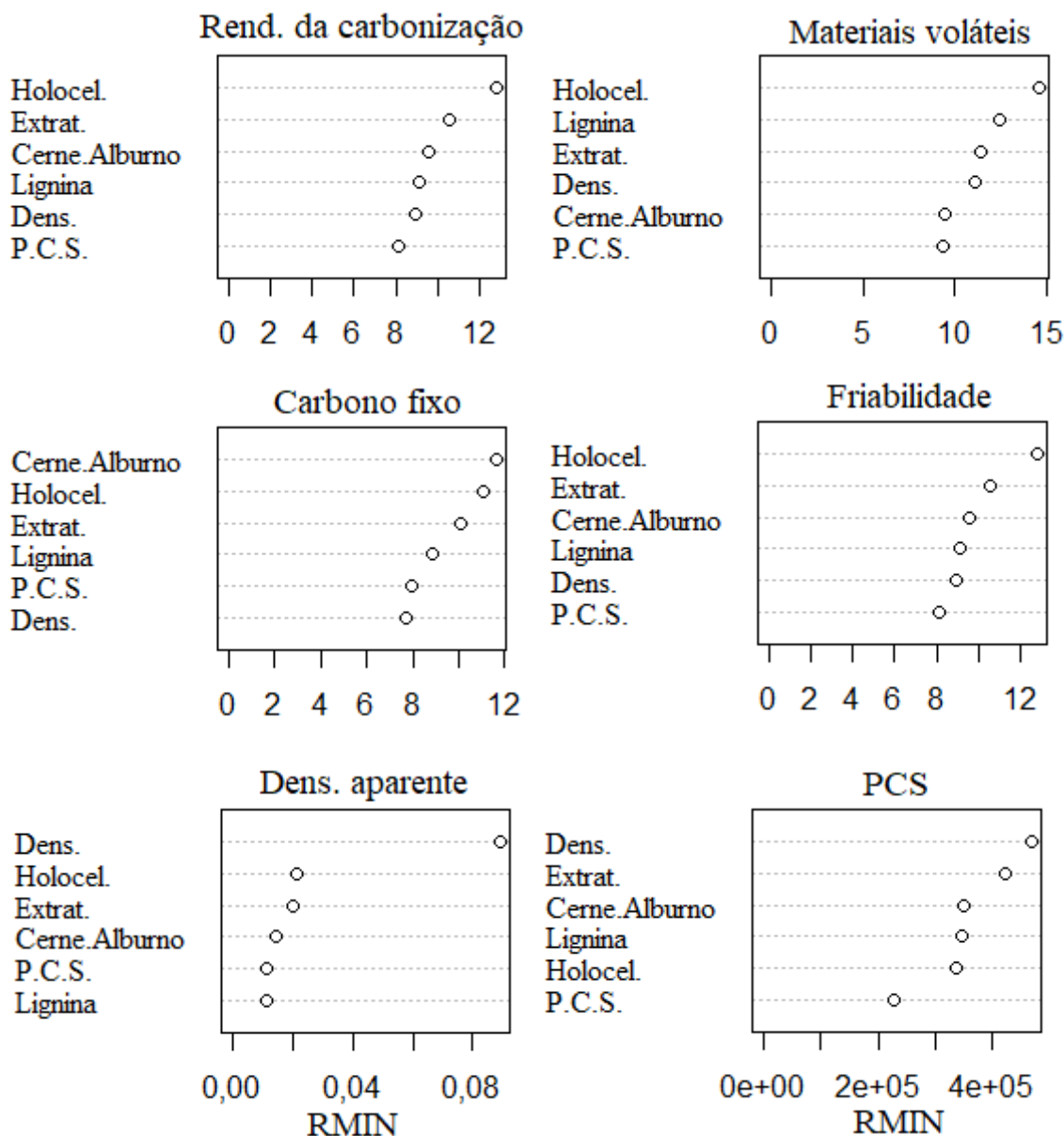


Figura 1 – Influência das propriedades da madeira: densidade básica da madeira (Dens.), poder calorífico superior (P.C.S.), teor de extrativos (Extrat.), lignina total (Lignina), holocelulose (Holocel.) e relação cerne/alburno (Cerne.Alburno) sobre o rendimento da carbonização (Rend. da carbonização), teor de materiais voláteis, teor de carbono fixo, friabilidade, densidade relativa aparente (Dens. aparente) e poder calorífico superior (PCS) do carvão vegetal de clones de eucalipto medida pela redução média da impureza do nó (RMIN) das árvores de regressão.

Verifica-se na Tabela 1 que com exceção das relações da holocelulose e densidade básica da madeira com friabilidade e densidade relativa aparente, respectivamente, todas as demais relações foram inversamente proporcionais. Analisando os valores de coeficiente de determinação, constata-se que foi possível explicar a maior parte da variabilidade das propriedades do carvão a partir apenas das variáveis da madeira mais importantes do ponto de vista da modelagem com random forest (RF).

Tabela 1 – Correlação de cópulas (r') e coeficiente de determinação (R^2) do algoritmo random forest (RF) implementados apenas com as propriedades mais influentes para o rendimento e qualidade do carvão vegetal de clones do gênero *Corymbia*.

Variável dependente*	Preditor**	r' (%)	R^2 (%)***
Rendimento da carbonização	Holocelulose	-55,19	61,65
Teor de materiais voláteis	Holocelulose	-37,90	62,84
Teor de carbono fixo	Cerne/Alburno	-12,92	61,74
Friabilidade	Holocelulose	41,59	77,48
Densidade relativa aparente	Densidade básica	84,84	85,23
Poder calorífico superior	Densidade básica	-33,07	80,45

*Referentes ao carvão; **Referentes à madeira; ***Sobre dados de validação.

Diferente da forte relação entre rendimento da carbonização e teor de holocelulose da madeira considerando o R^2 do RF, Santos (2008) verificou baixa correlação linear (-0,15) entre as duas variáveis, bem como Soares et al. (2014), que observaram a maior dependência linear do rendimento da carbonização com o poder calorífico superior da madeira. O aumento do rendimento da carbonização normalmente tende a ocorrer havendo maior teor de lignina na madeira, visto que essa é a macromolécula mais estável à degradação térmica (TRUGILHO et al., 2001; PEREIRA et al., 2013). A média geral do rendimento da carbonização foi 36,47% (Tabela 2), valor inferior aos observados para híbridos do gênero *Corymbia* no trabalho de Moutinho et al. (2017).

Tabela 2 – Médias das propriedades do carvão de 13 clones do gênero *Corymbia*.

Variável	Média	Unidade
Rendimento da carbonização	36,47	%
Teor de materiais voláteis	27,89	%
Carbono fixo	70,36	%
Friabilidade	6,14	%
Densidade aparente	0,36	g.cm^{-3}
Poder calorífico superior	7044,74	kcal.kg^{-1}

A relação entre holocelulose e teor de materiais voláteis em termos de correlação linear foi quase nula no trabalho de Soares et al. (2014), de modo que a principal influência relatada pelos autores sobre os materiais voláteis do carvão foi da relação carbono/hidrogênio da madeira. Segundo Carmo (1988) e Oliveira et al. (2010), o teor de voláteis também depende da temperatura de carbonização e taxa de aquecimento. Vale ressaltar que madeiras apropriadas para a produção de carvão devem apresentar baixos teores de materiais voláteis, sendo que a média observada nesta pesquisa fica acima do padrão de valores do gênero *Eucalyptus* (BRITO e BARRICHELO, 1977).

Brito e Barrichelo (1977) verificaram que madeiras de eucalipto mais lignificadas produziram carvão com maior teor de carbono fixo, o que se deve, segundo os autores, à composição da lignina que reúne aproximadamente 65% do carbono elementar. Embora os resultados sugerirem uma relação inversa entre carbono fixo e cerne/alburno, madeiras com maior relação cerne/alburno podem elevar o teor de carbono fixo no carvão, visto que o cerne tende a ter teor de lignina maior ou igual que o alburno (KLITZKE et al., 2008; COSTA et al., 2017; FONTE et al., 2017). De todo modo, o fator que mais interfere no teor de carbono fixo do carvão vegetal é o processo de produção, conforme demonstram Róz et al. (2015), que apontam uma temperatura ótima de carbonização em 300° C, considerando carbono fixo e rendimento gravimétrico do carvão produzido a partir de *Eucalyptus grandis* x *Eucalyptus urophylla*. A média de 70,36% do carbono fixo do *Corymbia* ficou abaixo da média do *Eucalyptus* sp. na pesquisa de Silva (2014) e superior às médias da catingueira (*Poincianella pyramidalis*) e pau-d'arco (*Handroanthus impetiginosus*) verificadas por Medeiros Neto, Oliveira e Paes (2014).

Trabalhos que avaliaram a friabilidade do carvão vegetal de diferentes espécies relacionam uma maior geração de finos, principalmente, ao diâmetro das toras, tempo de carbonização e umidade da madeira (COUTINHO e FERRAZ, 1988; SILVA, 1988; PINHEIRO, 2013). Coutinho e Ferraz (1988) explicam que a geração de finos é causada pela formação de tensões internas durante o processo de perda de umidade, estas, por sua vez, são diretamente influenciadas pela área do cerne e pela variação na densidade entre medula e casca da madeira. A qualidade do carvão do *Corymbia* mostrou-se melhor quanto a friabilidade em comparação ao produzido com madeira de *Eucalyptus saligna* (COUTINHO e FERRAZ, 1988) e de espécies amazônicas (SILVA, 1988).

A relação densidade aparente do carvão com densidade básica da madeira talvez seja a mais bem conhecida, conforme demonstrado por Brito e Barrichelo (1980) com uma associação

de mais de 97% pelo modelo de regressão linear simples. De acordo com Pereira (2012), quanto maior a densidade da madeira, maior será a densidade e a resistência mecânica do carvão vegetal, que, junto com a granulometria, Brito (1993) afirma serem as características mais importantes para a operação do alto forno na siderurgia. A média observada da densidade aparente do carvão para o *Corymbia* ficou na mesma faixa dos valores verificados por Oliveira et al. (2010) para *Eucalyptus pellita*.

O poder calorífico superior, segundo Couto (2014), é dependente da composição química elementar do material. O aumento do poder calorífico pode ser obtido por meio da elevação da temperatura de carbonização, que resulta no aumento do teor de carbono fixo pela eliminação de materiais voláteis (FIGUEIREDO et al., 2018). A média geral de poder calorífico superior dos materiais avaliados ficou dentro do intervalo de valores verificados por Soares et al. (2015) para híbridos de *Eucalyptus grandis* x *Eucalyptus urophylla* em três diferentes idades.

3.2 Validação cruzada

O desempenho do RF quanto às medidas de coeficiente de determinação (R^2), correlação linear entre observados e preditos (r), erro médio absoluto (EMA) e raiz quadrada do erro quadrático médio (REQM) foi superior aos demais algoritmos testados para todas as variáveis do carvão, tanto para estimativas com todas as variáveis da madeira, como utilizando apenas as selecionadas como mais importantes para a predição de cada propriedade do carvão. Vale ressaltar que a validação cruzada foi realizada sobre os dados destinados à validação, que nesse caso contemplam 12 observações, ou seja, trata-se de uma pequena amostra. Os valores de R^2 e r referentes ao RF foram elevados, a maioria acima ou próximo de 90% no cenário de estimação com todas as variáveis, e mantendo-se cima de 60% e 70% mesmo na estimação com apenas as variáveis da madeira mais influentes, para R^2 e r , respectivamente (Tabela 3 e 4). Em ambos os cenários de estimação, os valores de r para o RF foram significativos, o que não ocorreu na mesma medida para o SVR e RLM (Tabela 4).

Tabela 3 – Coeficiente de determinação do random forest (RF), support vector regression (SVR) e regressão linear múltipla (RLM) em função das seis variáveis predictoras da madeira (1ª estimação) e em função apenas da mais importante (2ª estimação), considerando dados de validação.

Variável dependente	RF	SVM	RLM	RF	SVM	RLM
	1ª estimação (%)			2ª estimação (%)		
Rendimento da carbonização	87,10	49,50	55,94	61,65	29,88	49,34
Teor de voláteis	88,26	33,88	20,00	62,84	27,52	0,22
Carbono fixo	89,12	0,36	0,09	61,74	22,92	13,18
Friabilidade	89,62	42,76	30,66	77,48	9,51	4,07
Densidade aparente	93,70	64,95	68,56	85,23	70,55	59,93
Poder calorífico superior	90,30	34,17	22,89	80,45	16,17	24,01

Tabela 4 – Coeficiente de correlação linear entre valores observados e preditos pelo random forest (RF), support vector regression (SVR) e regressão linear múltipla (RLM) em função das seis variáveis predictoras da madeira (1ª estimação) e em função apenas da mais importante (2ª estimação), considerando dados de validação.

Variável dependente	RF	SVM	RLM	RF	SVM	RLM
	1ª estimação (%)			2ª estimação (%)		
Rendimento da carbonização	93,65*	70,35*	74,80*	78,52*	54,67 ^{ns}	70,24*
Teor de voláteis	93,94*	62,54*	43,58 ^{ns}	79,27*	52,46 ^{ns}	-4,65 ^{ns}
Carbono fixo	94,40*	5,99 ^{ns}	3,01 ^{ns}	78,57*	47,87 ^{ns}	-36,31 ^{ns}
Friabilidade	94,67*	65,39*	55,37 ^{ns}	88,02*	30,94 ^{ns}	20,18 ^{ns}
Densidade aparente	96,80*	80,59*	82,80*	92,32*	83,99*	77,41*
Poder calorífico superior	95,03*	58,46*	21,53 ^{ns}	89,69*	40,83 ^{ns}	49,00 ^{ns}

*Significativo de acordo com o teste t a 5% de significância; ns: não significativo.

Os melhores resultados em termos de R^2 e r ocorreram na estimação da densidade relativa aparente do carvão vegetal. Por outro lado, considerando as medidas de erro EMA e REQM (Tabela 5 e 6), constata-se que a acurácia foi mais elevada na estimação do carbono fixo. O EMA e REQM também sugerem diferenças mais estreitas de acurácia do RF para o SVR e RLM, principalmente na predição do carbono fixo e do rendimento da carbonização. O contraste entre o que se pode inferir a partir do R^2 , r , EMA e REQM, acaba levantando a dúvida sobre qual parâmetro reflete a real precisão dos métodos de predição. Nesse caso, sabendo-se que o EMA e REQM mensuraram diretamente a diferença entre valores observados e preditos, estas medidas se tornam mais apropriadas para avaliar os métodos RF, SVR e RLM.

Tabela 5 – Erro médio absoluto do random forest (RF), support vector regression (SVR) e regressão linear múltipla (RLM) utilizando as seis variáveis preditoras da madeira (1ª estimacão) e apenas a mais importante (2ª estimacão), considerando dados de validacão.

Variável dependente	RF	SVM	RLM	RF	SVM	RLM
	1ª estimacão (%)			2ª estimacão (%)		
Rendimento da carbonizacão	0,88*	1,16	1,27	1,27	1,63	1,46
Teor de voláteis	1,72	2,23	3,20	2,37	2,91	3,67
Carbano fixo	0,64	1,18	1,18	0,81	1,09	1,26
Friabilidade	7,06	11,09	12,77	7,73	12,27	13,61
Densidade aparente	2,69	8,07	8,07	5,38	5,38	8,07
Poder calorífico superior	1,17	2,12	2,32	1,39	2,28	2,36

*Valores em porcentagem da média.

Tabela 6 – Raiz quadrada do erro quadrático médio do random forest (RF), support vector regression (SVR) e regressão linear múltipla (RLM) utilizando as seis variáveis preditoras da madeira (1ª estimacão) e apenas a mais importante (2ª estimacão), considerando dados de validacão.

Variável dependente	RF	SVM	RLM	RF	SVM	RLM
	1ª estimacão (%)			2ª estimacão (%)		
Rendimento da carbonizacão	1,60*	1,71	1,68	1,52	1,98	1,76
Teor de voláteis	2,30	3,45	3,88	2,73	3,85	4,46
Carbano fixo	0,78	1,42	1,49	0,97	1,29	1,46
Friabilidade	8,74	16,13	15,80	9,91	18,15	18,15
Densidade aparente	2,69	8,07	8,07	5,38	8,07	8,07
Poder calorífico superior	1,36	2,68	2,98	1,69	2,93	2,93

*Valores em porcentagem da média.

Diferentemente do que foi observado no presente trabalho, Montaño (2016) obteve melhores resultados na predição do volume de pinus, da biomassa e da altura da acácia negra em função do DAP (diâmetro a altura do peito) e altura total com o SVR em comparacão com o RF. No estudo de Carvalho Junior et al. (2016) foi verificado, em geral, maior R^2 e menor REQM para modelos de RLM em relacão ao RF na estimacão da densidade do solo em função de propriedades físico-químicas das amostras. Rodríguez-Lado et al. (2015), por outro lado, ajustando modelos de RLM, redes neurais e RF para a estimacão da densidade do solo em função da matéria orgânica e textura, verificaram melhores resultados de R^2 , MAE e REQM com o uso do RF.

4 CONCLUSÕES

A acurácia do algoritmo *random forest* foi maior que a do *support vector regression* e da regressão linear múltipla, sendo suficiente para considerar viável a estimação do rendimento e propriedades de qualidade do carvão vegetal em função das propriedades da madeira por meio do aprendizado de máquina utilizando o *random forest*. Não obstante, considerando as medidas de erro, a utilização do *support vector regression* e da regressão linear múltipla também possibilita previsões acuradas.

Utilizando o algoritmo *random forest* verifica-se que o rendimento e a qualidade do carvão vegetal de clones de *Corymbia* são influenciados, principalmente, pelo teor de holocelulose, relação cerne/alburno e densidade básica da madeira. A densidade relativa aparente do carvão foi a variável em que o *random forest* atingiu o melhor nível de explicação da variabilidade em função das propriedades da madeira, porém, foi para o teor de carbono fixo que o algoritmo forneceu previsões com menor erro.

REFERÊNCIAS BIBLIOGRÁFICAS

- ABNT. **NBR 11941: Madeira: determinação da densidade básica**. Rio de Janeiro, 2003. 6p.
- ABNT. **NBR 8112: Carvão vegetal: análise imediata**. Rio de Janeiro, 1986. 8p.
- ABNT. **NBR 8633: Carvão vegetal: determinação do poder calorífico**. Rio de Janeiro, 1984. 13 p.
- AITKENHEAD, M.J.; COULL, M.C. Mapping soil carbon stocks across Scotland using a neural network model. *Geoderma*, v.262, p.187–198, 2016.
- BIAMONTE, J.B; WITTEK, P.W.; PANCOTTI, N.P.; REBENTROST, P.R.; WIEBE, N.; LLOYD, S. Quantum machine learning. *Nature*, v.549, p.195-202, 2017.
- BREIMAN, L. Random forests. *Machine Learning*, v.45, p.5–32, 2001.
- BREIMAN, L. 2002. **Manual on setting up, using, and understanding random forests V3.1**. Disponível em:<https://www.stat.berkeley.edu/~breiman/Using_random_forests_V3.1.pdf> Acesso em: 26 de set. de 2018.
- BRITO, J.O.; BARRICHELO, L.E.G. Correlações entre características físicas e químicas da madeira e a produção de carvão vegetal: I. densidade e teor de lignina da madeira de eucalipto. *IPEF*, n.14, p.9-20, 1977.
- BRITO, J.O.; BARRICHELO, L.E.G. Correlações entre características físicas e químicas da madeira e a produção de carvão vegetal: 2. densidade da madeira x densidade do carvão. *IPEF*, n.20, p.101-113, 1980.

BRITO, J.O. Reflexões sobre qualidade do carvão vegetal para uso siderúrgico. **IPEF**, n.181, 1993.

CARMO, J.S. **Propriedades físicas e químicas do carvão vegetal destinado à siderurgia e metalurgia**. 1988. 36f. Monografia (Graduação em Ciências Florestais) – Universidade Federal de Viçosa, Viçosa, 1988.

CARUANA, R.; KARAMPATZIAKIS, N.; YESSINALINA, A. An empirical evaluation of supervised learning in high dimensions. In: International Conference on Machine Learning, 25., 2008, Helsínquia. **Proceedings...** Helsínquia: ACM, 2008. p. 96-103.

CARVALHO JUNIOR, W. de; CALDERANO FILHO, B.; CHAGAS, C. da S.; BHERING, S.B.; PEREIRA, N.R.; PINHEIRO, H.S.K. Regressão linear múltipla e modelo Random Forest para estimar a densidade do solo em áreas montanhosas. **Pesquisa Agropecuária Brasileira**, v.51, n.9, p.1428-1437, 2016.

CASTRO, A.F.N.M. **Efeito da idade e de materiais genéticos de Eucalyptus sp. na madeira e carvão vegetal**. 2011. 97p. Dissertação (Mestrado) – Universidade Federal de Viçosa, Viçosa, 2011.

COSTA, A.C.S.; LEAL, C.S.; SANTOS, L.C.; CARVALHO, A.M.M.L.; OLIVEIRA, A.C.; PEREIRA, B.L.C. Propriedades da madeira de cerne e alburno de Eucalyptus camaldulensis. **Ciência da Madeira (Brazilian Journal of Wood Science)**, v.8, n.1, p.10-20, 2017.

COUTINHO, A. dos R.; FERRAZ, E.S.B. Determinação da friabilidade do carvão vegetal em função do diâmetro das árvores e temperatura de carbonização. **IPEF**, n.38, p.33-37, 1988.

COUTO, A.M. **Influência das propriedades anatômicas, químicas e físicas da madeira de Eucalyptus e Corymbia na qualidade do carvão para uso siderúrgico**. 2014. 173f. Tese (Doutorado em Ciência e Tecnologia da Madeira) – Universidade Federal de Lavras, Lavras, 2014.

DANTAS, P.B. **Utilização da aprendizagem de máquina e seleção de atributos para o diagnóstico de ceratocone a partir de parâmetros biomecânicos**. 2017. 63f. Dissertação (Mestrado em Modelagem Computacional do Conhecimento) – Universidade Federal de Alagoas, Maceió, 2017.

DING, A.A.; LI, Y. Copula correlation: An equitable dependence measure and extension of pearson's correlation. **arXiv preprint arXiv:1312.7214**, 2013.

FONTE, A.P.N.; TRIANOSKI, R.; IWAKIRI, S.; ANJOS, R.A.M. dos. Propriedades físicas e químicas da madeira de cerne e alburno de Cryptomeria japonica. **Revista de Ciências Agroveterinárias**, v.16, n.3, p.277-285, 2017.

GIROLAMO NETO, C.D.; FONSECA, L.M.G.; KORTING, T.S.; SANCHES, I. D.; EBERHARDT, I.D.R.; BENDINI, H. do N.; MARUJO, R. de F.B.; TRANBAQUINI, K. Classificação automática de áreas cafeeiras utilizando imagens de sensoriamento remoto e técnicas de mineração de dados. In: Simpósio Brasileiro de Sensoriamento Remoto, 17., 2015, João Pessoa, **Anais...** João Pessoa: INPE, 2015. p. 1609-1616.

FIGUEIREDO, M.E.O.; LONGUE JÚNIOR, D.; PEREIRA, A.K.S.; CARNEIRO, A.C.O.; SILVA, C.M.S. Potencial da madeira de *Pterogyne nitens* Tul. (madeira-nova) para produção de carvão vegetal. **Ciência Florestal**, v.28, n.1, p.420-431, 2018.

GOLDSCHIMID, O. Ultraviolet spectra. In: SARKANEN, K.V.; LUDWING, C.H. (Eds). **Lignins**. New York: Wiley Interscience. 1971, p.241-266.

GOMIDE, J.L.; DEMUNER, B.J. Determinação do teor de lignina em material lenhoso: método Klason modificado. **O Papel**, v.47, p.36-38, 1986.

HALLAK, R.; PEREIRA FILHO, A.J. Metodologia para análise de desempenho de simulações de sistemas convectivos na região metropolitana de São Paulo com o modelo ARPS: sensibilidade a variações com os esquemas de advecção e assimilação de dados. **Revista Brasileira de Meteorologia**, v.26, n.4, p.591-608, 2011.

KLITZKE, R.J.; SAVIOLI, D.L.; MUÑIZ, G.I.B. de; BATISTA, D.C. Caracterização dos lenhos de cerne, alburno e transição de jatobá (*Hymenaea* sp.) visando ao agrupamento para fins de secagem convencional. **Scientia Forestalis**, v.36, n.80, p.279-284, 2008.

LENTO, G.C. **Random forest em dados desbalanceados: uma aplicação na modelagem de churn em seguro saúde**. 2017. 43f. Dissertação (Mestrado em Modelagem Matemática da Informação) – Fundação Getulio Vargas, Rio de Janeiro, 2017.

LOPES, T.D.; GOEDEL, A.; PALÁCIOS, R.H.C.; GODOY, W.F. Aplicação do algoritmo random forest como classificador de padrões de falhas em rolamentos de motores de indução. In: Simpósio Brasileiro de Automação Inteligente, 13., 2017, Porto Alegre, **Anais...** Porto Alegre: SBA, 2017. p.263-268.

MALONE, B.P.; JHA, S.K.; MINASNY, B.; MCBRATNEY, A.B. Comparing regression-based digital soil mapping and multiple-point geostatistics for the spatial extrapolation of soil data. **Geoderma**, v.262, p.243–253, 2016.

MEDEIROS NETO, P.N. de; OLIVEIRA, Elisabeth de; PAES, J.B. Relações entre as características da madeira e do carvão vegetal de duas espécies da Caatinga. **Floresta e Ambiente**, v.21, n.4, p.484-493, 2014.

MONTAÑO, R.A.N.R. **Aplicação de técnicas de aprendizado de máquina na mensuração florestal**. 2016. 102f. Tese (Doutorado em Informática) – Universidade Federal do Paraná, Curitiba, 2016.

MORAIS, S.A.L. de; NASCIMENTO, E.A. do; MELO, D.C. de. Análise da madeira de *Pinus oocarpa* parte I – estudo dos constituintes macromoleculares e extrativos voláteis. **Revista Árvore**, v.29, n.3, p.461-470, 2005.

MOTTA, L. P.; MACHADO, C. C.; SOARES, V. P.; RIBEIRO, J. C.; RIBEIRO, C. A. A. S. Utilização de sistema de informações e da distância virtual na otimização do transporte florestal rodoviário. **Revista Árvore**, v.20, n.3, p.381-394, 1996.

MOUTINHO, V.H.P.; TOMAZELLO FILHO, M.; BRITO, J.O.; BALLARIN, A.W.; ANDRADE, F.W.C.; CARDOSO, C. da C. Characterization and statistical correlation between

charcoal's physical and mechanical properties of Eucalyptus and Corymbia clones. **Ciência Florestal**, v. 27, n. 3, p. 1095-1103, 2017.

OLIVEIRA, A.C.; CARNEIRO, A. de C.O.; VITAL, B.R.; ALMEIDA, W.; PEREIRA, B.L.C.; CARDOSO, M.T. Parâmetros de qualidade da madeira e do carvão vegetal de Eucalyptus pellita F. Muell. **Scientia Forestalis**, v.38, n.87, p.431-439, 2010.

OLIVEIRA, J.B.; VIVACQUA FILHO, A.; MENDES, M.G.; GOMES, P.A. Produção de carvão vegetal – aspectos técnicos. In: CETEC. **Produção e Utilização de Carvão Vegetal**. Belo Horizonte: CETEC. 1982, p.59-72.

PEREIRA, B.L.C. **Qualidade da madeira de Eucalyptus para a produção de carvão vegetal**. 2012. 93f. Dissertação (Mestrado em Ciência Florestal) – Universidade Federal de Viçosa, Viçosa, 2012.

PEREIRA, B.L.C.; CARNEIRO, A.C.O.; CARVALHO, A.M.M.L; TRUGILHO, P.F; MELO, I.C.N.A.; OLIVEIRA, A.C. Estudo da degradação térmica da madeira de Eucalyptus através de termogravimetria e calorimetria. **Revista Árvore**, v.37, n.3, p.567-576, 2013.

PINHEIRO, M.A. **Influência das dimensões da madeira na secagem e nas propriedades do carvão vegetal**. 2013. 69f. Dissertação (Mestrado em Ciência Florestal) – Universidade Federal de Viçosa, Viçosa, 2013.

PROTÁSIO, T. de P.; TRUGILHO, P. F.; NEVES, T. A.; VIEIRA, C. M. M. Análise de correlação canônica entre características da madeira e do carvão vegetal de Eucalyptus. **Scientia Forestalis**, v.40, n.95, p.317-326, 2012.

R CORE TEAM. **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, 2018.

Rodríguez-Lado, L.; RIAL, M.; TABOADA, T.; CORTIZAS, A.M. A pedotransfer function to map soil bulk density from limited data. **Procedia Environmental Sciences**, v.27, p.45-48, 2015.

RÓZ, A.L. da; RICARDO, J.F. C.; NAKASHIMA, Gabriela T.; SANTOS, L.R.O.; YAMAJI, F.M. Maximização do teor de carbono fixo em biocarvão aplicado ao sequestro de carbono. **R. Bras. Eng. Agríc. Ambiental**, v.19, n.8, p.810–814, 2015.

SANTOS, I.D. **Influência dos teores de lignina, holocelulose e extrativos na densidade básica e contração da madeira e nos rendimentos e densidade do carvão vegetal de cinco espécies lenhosas do Cerrado**. 2008. 92f. Dissertação (Mestrado em Ciências Florestais) – Universidade de Brasília, Brasília, 2008.

SILVA, D.A. de. Qualidade do carvão vegetal produzido com madeiras da região de Manaus em fornos de alvenaria. **Acta Amazonica**, v.18, n.1-2, p.163-178, 1988.

SILVA, J.G.B. **Propriedade do carvão da madeira e da casca de Eucalyptus sp. plantado na região do Recôncavo da Bahia**. 2014. 23f. Monografia (Graduação em Engenharia Florestal) – Universidade Federal do Recôncavo da Bahia, Cruz das Almas, 2014.

SOARES, V.C.; BIANCHI, M.L.; TRUGILHO, P.F.; PEREIRA, A.J.; HÖFLER, J. Correlações entre as propriedades da madeira e do carvão vegetal de híbridos de eucalipto. **Revista Árvore**, v.38, n.3, p.543-549, 2014.

SOARES, V.C; BIANCHI, M.L.; TRUGILHO, P.F.; HÖFLER, J.; PEREIRA, A.J. Análise das propriedades da madeira e do carvão vegetal de híbridos de eucalipto em três idades. **Cerne**, v.21, n.2, p.191-197, 2015.

TAPPI T 264 cm-97. **Preparation of wood for chemical analysis**. Atlanta: Tappi Press, 1997.

TRUGILHO, P. F.; REGAZZI, A. J.; VITAL, B. R.; GOMIDE, J. L. Aplicação de algumas técnicas multivariadas na avaliação da qualidade da madeira de Eucalyptus e seleção de genótipos superiores para a produção de carvão vegetal. **Revista Árvore**, v.21, n.1, p. 113-130, 1997.

TRUGILHO, P.F.; VITAL, B.R.; REGAZZI, A.J.; GOMIDE, J.L. Aplicação da análise de correlação canônica na identificação de índices de qualidade da madeira de eucalipto para a produção de carvão vegetal. **Revista Árvore**, v.21, n.2, p.259-267, 1997.

TRUGILHO, P.F; LIMA, J.T. MORI, F.A.; LINO, A.L.L. Avaliação de clones de Eucalyptus para produção de carvão vegetal. **Cerne**, v.7, n.2, p.104-114, 2001.

TRUGILHO, P.F.; LIMA, J.T.; MORI, F.A. Correlação canônica das características químicas e físicas da madeira de clones de Eucalyptus grandis e Eucalyptus saligna. **Cerne**, v.9, n.1, p.66-80, 2003.

VITAL, B.R. **Métodos de determinação da densidade da madeira**. Viçosa, MG: Sociedade de Investigações Florestais, 1984. 21p.

VITAL, B.R.; ALMEIDA, J. de; VALENTE, O.F.; PIRES, I.E. Características de crescimento das árvores e de qualidade da madeira de Eucalyptus camaldulensis para a produção de carvão. **IPEF**, n.47, p.22-28, 1994.

CAPÍTULO 2

FUNÇÃO DISCRIMINANTE DE FISHER APLICADA NA CLASSIFICAÇÃO DO CARVÃO VEGETAL DE CLONES DE CORYMBIA

Resumo – Objetivou-se utilizar a função discriminante de Fisher para sumarizar as propriedades do carvão vegetal e classificar os clones de Corymbia quanto ao rendimento e qualidade da produção de carvão. Foram consideradas seis variáveis: rendimento da carbonização, teor de materiais voláteis, carbono fixo, friabilidade, densidade relativa aparente e poder calorífico superior do carvão vegetal. Os dados foram previamente testados quanto às pressuposições de normalidade multivariada e homogeneidade de variâncias/covariâncias. Posteriormente foi realizada a análise multivariada de variância e aplicada a função discriminante de Fisher para transformação das variáveis originais em duas variáveis canônicas. A comparação dos clones quanto ao rendimento e qualidade do carvão vegetal foi realizada aplicando-se o teste de Scott-Knott a 1% de significância sobre as médias das variáveis canônicas. Em comparação às análises univariadas, o uso da função discriminante de Fisher para gerar novas variáveis apresenta vantagens quanto a simplificação da interpretação dos resultados e, principalmente, na diferenciação dos clones em relação à qualidade global do carvão vegetal. Comparando os clones por meio das variáveis canônicas constatou-se que o genótipo AMF1119 é o mais indicado para a produção de carvão vegetal.

Palavras-chave: análise multivariada, redução de dimensionalidade, variável canônica.

FISHER'S DISCRIMINANT FUNCTION APPLIED IN THE CLASSIFICATION OF THE CHARCOAL OF CORYMBIA CLONES

Abstract – The objective of this study was to use Fisher's discriminant function to summarize charcoal properties and to classify Corymbia clones for yield and quality of charcoal production. Six variables were considered: carbonization yield, volatile matter content, fixed carbon, friability, apparent relative density and higher calorific value of charcoal. The data were previously tested for multivariate normality and homogeneity of variances/covariates. Subsequently, the multivariate analysis of variance was performed and Fisher's discriminant function was applied to transform the original variables into two canonical variables. The comparison of the clones regarding the charcoal yield and quality was carried out by applying

the Scott-Knott' test at 1% significance on the averages of the canonical variables. In contrast to the univariate analysis, the use of the Fisher's discriminant function to generate new variables presents advantages in terms of simplifying the interpretation of the results and, mainly, in the differentiation of the clones in relation to the overall charcoal quality. Comparing the clones by means of the canonical variables it is verified that the AMF1119 genotype is the most suitable for the charcoal production.

Keywords: multivariate analysis, dimensionality reduction, canonical variable.

1 INTRODUÇÃO

De acordo com o último relatório elaborado pela FAO (FAO, 2017), o Brasil é o maior produtor de carvão vegetal do mundo, produzindo 6,2 milhões de toneladas em 2015, o que representou cerca de 12% da produção global daquele ano. O carvão vegetal consumido dentro do país é destinado principalmente às indústrias de ferro-gusa, ferro-liga e aço, chegando a um consumo interno de 4,5 milhões de toneladas de carvão em 2017, fazendo do Brasil líder mundial na produção de aço a partir de carvão vegetal, o qual é de melhor qualidade e resulta em menor impacto ambiental (ABRAF, 2013; IBÁ, 2019).

As florestas equiâneas de eucalipto têm sido as mais utilizadas para a produção de carvão, visto que apresentam alta produtividade volumétrica e madeira com características apropriadas para este fim (GOULART et al., 2003). Devido à elevada variabilidade da qualidade do carvão vegetal e à exigência do mercado, é importante que as empresas de reflorestamento busquem selecionar materiais genéticos mais produtivos e com propriedades da madeira favoráveis à qualidade do carvão (TRUGILHO et al., 2001; OLIVEIRA et al., 2010; CASTRO et al., 2013).

Os estudos com qualidade do carvão vegetal se deparam com um amplo conjunto de variáveis, o que direciona frequentemente tais pesquisas às análises no campo multivariado (PROTÁSIO et al., 2012; PROTÁSIO et al., 2013; CASTRO et al., 2013). A vantagem da utilização de técnicas multivariadas, segundo Mingoti (2005), dar-se às correlações que normalmente existem entre as variáveis, fato que torna as análises por vias tradicionais complexas e trabalhosas.

Quando se trata de selecionar clones, tem-se um fator qualitativo para ser analisado em que os níveis são os diferentes genótipos, ou seja, o objetivo é comparar médias dos diferentes clones. As análises multivariadas apresentam algumas alternativas que podem simplificar a comparação de médias partindo de um grande conjunto de variáveis observadas, como análise

de componentes principais, análise de fatores e análise multivariada de variância (MANOVA). Gomes (2009) descreve uma proposta de aplicação da função discriminante de Fisher ou primeira variável canônica que, diferente do uso convencional da análise discriminante para classificação de um indivíduo em uma população predefinida (FISHER, 1936), possibilita transformar todas as variáveis originais em uma única variável que retém a maior parte da informação dos dados.

A utilização da função discriminante como complemento da MANOVA para transformar as variáveis sem grande perda de informação resulta numa simplificação eficiente dos resultados (CAMPOS, 2012). No contexto da avaliação das propriedades do carvão, a função discriminante conforme a metodologia descrita por Gomes (2009) é um meio de resumir toda a informação em um único indicador de qualidade global do carvão vegetal que é uma função das variáveis observadas. Sendo assim, objetivou-se utilizar a função discriminante de Fisher para sumarizar as propriedades do carvão vegetal e classificar os clones de *Corymbia* quanto ao rendimento e qualidade da produção de carvão.

2 MATERIAL E MÉTODOS

A presente pesquisa foi realizada na ArcelorMittal BioFlorestas, empresa pertencente ao Grupo ArcelorMittal, no Município de Dionísio, Minas Gerais. O clima da região é classificado como subtropical úmido, cuja temperatura média anual varia entre 20 a 23° C e a precipitação média é de 1100 a 1400 mm (Motta et al., 1996).

O plantio de 13 genótipos do gênero *Corymbia* foi realizado em 2008, em área de Latossolo Vermelho-Amarelo com espaçamento entre plantas de 3 x 2,5 m. Os plantios tinham 7 anos de idade quando as amostras para as análises das propriedades da madeira e carvão foram coletadas.

2.1 Rendimento gravimétrico e propriedades do carvão vegetal

O rendimento gravimétrico foi determinado pela relação entre massa de carvão vegetal e massa seca de madeira; a densidade relativa aparente do carvão vegetal foi determinada utilizando o método hidrostático, conforme Vital (1984); a friabilidade foi determinada de acordo com a metodologia proposta Oliveira et al. (1982) para carvão vegetal; os teores de materiais voláteis e cinzas do carvão vegetal foram determinados seguindo a NBR 8112 (ABNT, 1986); o teor de carbono fixo foi calculado subtraindo-se de 100% os teores de

materiais voláteis e cinzas; por fim, o poder calorífico superior foi mensurado utilizando-se uma bomba calorimétrica adiabática IKA300, segundo metodologia da NBR 8633 (ABNT, 1984).

2.2 Análises estatísticas

A avaliação das diferenças entre os atributos de qualidade do carvão entre os diferentes clones de *Corymbia* foi realizada por meio da MANOVA. Precedendo a realização da MANOVA, os dados foram submetidos a análise de normalidade e homocedasticidade de matrizes de variâncias/covariâncias.

A normalidade dos dados foi avaliada inicialmente pelo teste de Shapiro-Wilk a 1% de significância aplicado à cada variável individualmente. A normalidade multivariada foi avaliada por meio dos testes de assimetria e curtose propostos por Mardia (1970; 1974; 1975) a 1% de significância. Para a avaliação univariada das variáveis quanto a homocedasticidade de variâncias foi aplicado o teste de Bartlett a 1% de significância. A generalização do teste de Levene, o método PERMDIST (ANDERSON, 2006), foi utilizado para analisar a homogeneidade de matrizes de variância/covariância. Também foi realizada uma análise para detecção de outliers multivariados pelo método da distância de Mahalanobis robusta.

A análise multivariada de variância (MANOVA) foi realizada conforme Gomes (2009), utilizando o critério lambda de Wicks (Λ) com aproximação para a distribuição F para testar a hipótese de igualdade dos vetores de médias dos clones ($H_0: \mu_{v1} = \mu_{v2} = \dots = \mu_{v39}$), conforme a transformação indicada para casos gerais de graus de liberdade (HARRIS, 1975):

$$F'_{[pk_1, ab-c]} = \left[\left(\frac{1}{\Lambda} \right)^{\frac{1}{b}} - 1 \right] \frac{ab - c}{pk_1},$$

em que:

$$\Lambda = \frac{\det(\mathbf{H})}{\det(\mathbf{T})}$$

$$a = k_2 - \frac{1}{2} (p - k_1 + 1)$$

$$b = \sqrt{\frac{p^2 k_1^2 - 4}{p^2 + k_1^2 - 5}}$$

$$c = \frac{1}{2}(pk_1 - 2),$$

sendo F' o valor aproximado da estatística F , p igual ao número de variáveis, k_1 e k_2 são os graus de liberdade de tratamento e do resíduo, respectivamente; Λ é a estatística do teste de Wilks, $\det(\mathbf{H})$ e $\det(\mathbf{T})$ são os determinantes das matrizes de soma de quadrados e produtos de efeito dos tratamentos e do resíduo, na ordem. Em adição, também foram realizados os testes de Pillai, de Hotelling-Lawley e de Roy para efeito de ratificação.

A aplicação da função discriminante para gerar uma variável canônica em função das variáveis originais foi realizada conforme descreveu Gomes (2009), o que indica a utilização da primeira função discriminante baseado na ideia de reter a maior parte da informação contida nos dados e maximizar a discriminação dos tratamentos, visto que esta é associada ao maior autovalor. A retenção de informação foi mensurada pela proporção do autovalor em relação a soma de todos os autovalores:

$$\delta = \frac{\lambda_p}{\sum_{p=1}^6 \lambda_p},$$

em que δ é a proporção de informação retida pelo autovalor λ_p .

Os coeficientes da função discriminante foram estimados a partir da decomposição espectral da matriz $\mathbf{E}^{-1}\mathbf{H}$, em outras palavras, os coeficientes não normalizados da função são os autovetores da matriz de autovetores. A normalização de cada coeficiente foi realizada conforme a equação:

$$b_p^* = \frac{b_p}{\sqrt{\sum_{p=1}^6 (b_p)^2}},$$

em que b_p^* é o coeficiente normalizado e b_p o não normalizado. Estimada a variável canônica, submeteu-se a nova variável ao teste de normalidade de Shapiro-Wilk e ao teste de homocedasticidade de Bartlett, ambos a 1% de significância. Em seguida, a variável canônica foi submetida à ANOVA e os clones de *Corymbia* foram comparados por meio da variável canônica aplicando-se o teste de Scott-Knott a 1% de significância.

Para validar e comparar a qualidade da função discriminante em relação às análises univariadas tradicionais, foram utilizados critérios similares ao de Campos (2012). O principal critério de comparação foi a magnitude da estatística F e os respectivos valores-p, além do cálculo da acurácia seletiva (RESENDE, 2002):

$$AS = \left(1 - \frac{1}{F}\right)^{1/2}$$

em que AS é a acurácia seletiva, F é a razão entre quadrado médio do efeito de clone e quadrado médio do resíduo. Para a realização das análises foram utilizados os pacotes vegan, MVN, car e ScottKnott do R (R CORE TEAM, 2018).

3 RESULTADOS E DISCUSSÃO

Apesar da análise de normalidade multivariada ter indicado rejeição da hipótese nula de simetria, para a curtose houve aproximação da normal multivariada. Analisando a normalidade das variáveis individualmente, constatou-se distribuição normal em quase todos os casos, apenas o teor de materiais voláteis não apresentou distribuição normal. Quanto a homocedasticidade, o pressuposto de homogeneidade de variâncias/covariâncias foi atendido tanto para as variáveis individualmente como pelo teste multivariado (Tabela 1).

Tabela 1 – Resultados dos testes de normalidade e homocedasticidade das variáveis do carvão vegetal.

Variável	Teste	Valor-p
Rendimento da carbonização	Shapiro-Wilk	0,0313 ^{ns}
Teor de materiais voláteis	Shapiro-Wilk	0,0068*
Carbono fixo	Shapiro-Wilk	0,1432 ^{ns}
Friabilidade	Shapiro-Wilk	0,3305 ^{ns}
Densidade relativa aparente	Shapiro-Wilk	0,0208 ^{ns}
Poder calorífico superior	Shapiro-Wilk	0,1403 ^{ns}
Rendimento da carbonização	Bartlett	0,2279 ^{ns}
Teor de materiais voláteis	Bartlett	0,0135 ^{ns}
Carbono fixo	Bartlett	0,3621 ^{ns}
Friabilidade	Bartlett	0,6243 ^{ns}
Densidade relativa aparente	Bartlett	0,5391 ^{ns}
Poder calorífico superior	Bartlett	0,3437 ^{ns}
	Assimetria multivariada	0,0009*
	Curtose multivariada	0,2671 ^{ns}
	Homocedasticidade multivariada	0,6099 ^{ns}

*Significativo a 1% de significância; ns: não significativo.

Ainda na análise preliminar, foi constatada a presença de sete outliers multivariados considerando o quantil ajustado de 16,64 da distribuição qui-quadrado. A presença de outliers pode resultar na rejeição da suposição de normalidade multivariada e gerar uma distribuição normal multivariada contaminada (BARBOSA, PEREIRA e OLIVEIRA, 2018), o que afeta todas inferências posteriores baseadas no pressuposto de normalidade (SCHWAGER e MARGOLIN, 1982). Entretanto, a maior parte dos outliers englobam todas as observações de um dos clones, ou seja, provavelmente é indício de um material genético superior em termos de propriedades do carvão.

Após a realização da análise multivariada de variância (MANOVA), verificou-se diferença significativa no campo multivariado com relação à qualidade do carvão vegetal produzido a partir da madeira dos diferentes clones de eucalipto (Tabela 2). O valor aproximado da estatística F para o teste de Wilks, assim como para os demais testes, foi altamente significativo (valor-p < 0,001), evidenciando a diferença de potencial entre os clones para a produção de carvão.

Tabela 2 – Resumo da MANOVA das variáveis de rendimento e qualidade do carvão vegetal em função de 12 clones do gênero *Corymbia*.

Teste	Estatística	F'	GL(num)	GL(den)	Valor-p
Wilks	0,0177	3,7450	66	219,4900	< 0,0001***
Pillai	2,5058	2,9277	66	270,0000	< 0,0001***
Hotelling-Lawley	7,8345	7,8345	66	230,0000	< 0,0001***
Roy	3,9556	16,1820	11	45,0000	< 0,0001***

F': valor aproximado da estatística F; GL(num): graus de liberdade do numerador; GL(den): graus de liberdade do denominador;***Significativo a 0,1% pelo teste F.

Diferenças na qualidade do carvão vegetal de clones de *Corymbia* já haviam sido constadas por Costa (2016) ao avaliar mecanicamente a qualidade do carvão produzido a partir de híbridos sob temperaturas de carbonização variando de 350 a 700° C. Já Couto (2014) verificou que existem variações nas propriedades do carvão de clones de *Corymbia* em função de amostras de madeira coletadas em posições radiais distintas.

Procedendo a decomposição espectral da matriz resultante de $E^{-1}H$, foi constatado que o maior autovalor reteve aproximadamente 49% da informação contida nos dados (Tabela 3). Gomes (2009) afirma que é possível utilizar apenas primeira função discriminante (associada ao maior autovalor) para transformar as variáveis e explicar a variabilidade dos dados, desde que a retenção da informação pela primeira função seja suficiente. Fazendo um paralelo com

análise fatorial e de componentes principais, para redução eficiente da dimensionalidade de dados Hair Jr. et al. (2009) sugere retenção de ao menos 60% da variância total, já Ferreira (2011) indica 70% ou mais. Como a primeira função discriminante reteve menos da metade da informação dos dados, foi utilizada também a segunda função discriminante para analisar os dados, agregando, assim, quase 80% da informação total.

Tabela 3 – Informação retida pelos respectivos autovalores.

Autovalor (λ)	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6
Informação retida (δ)	3,8648	2,4469	0,9430	0,3183	0,2152	0,1111
Informação acumulada	0,4894	0,3099	0,1191	0,0403	0,0272	0,0140
	0,4894	0,7993	0,9184	0,9587	0,9859	1,0000

Selecionando-se e normalizando os autovetores da primeira e segunda coluna da matriz de autovetores referentes a matriz $\mathbf{E}^{-1}\mathbf{H}$, obteve-se os coeficientes normalizados da primeira e da segunda função discriminante de Fisher (Tabela 4). Conforme supracitado, por estarem associadas aos dois maiores autovalores, as duas primeiras funções discriminantes, juntas, explicam aproximadamente 80% da variabilidade dos dados, bem como maximizam a diferenciação dos tratamentos, neste caso os clones de *Corymbia*. Nota-se, ainda, que o maior coeficiente, em valor absoluto, é relacionado a variável carbono fixo no caso da primeira função, enquanto que na segunda função o maior foi da densidade relativa aparente.

Tabela 4 – Coeficientes normalizados da primeira função discriminante (FD1) e da segunda função discriminante (FD2) de Fisher.

Variável	Coeficiente FD1	Coeficiente FD2
X ₁ Rendimento da carbonização	b ₁ 0,4570	b ₁ 0,2924.10 ⁻¹
X ₂ Teor de materiais voláteis	b ₂ 0,5438	b ₂ 0,4180
X ₃ Carbono fixo	b ₃ 0,6762	b ₃ 0,4251
X ₄ Friabilidade	b ₄ -0,1842	b ₄ 0,2318
X ₅ Densidade relativa aparente	b ₅ -0,0654	b ₅ -0,7681
X ₆ Poder calorífico superior	b ₆ -0,9279.10 ⁻³	b ₆ 0,3263.10 ⁻³
$Z = X_1b_1 + X_2b_2 + X_3b_3 + X_4b_4 + X_5b_5 + X_6b_6$		

A magnitude em módulo dos coeficientes da função pode ser utilizada como indicador de quais variáveis são mais importantes para a discriminação dos grupos em questão (SIMEÃO e PADOVANI, 2008; SOUZA et al., 2009). Neste caso, a variável carbono fixo e densidade relativa aparente apresentam-se como os principais atributos de qualidade do carvão vegetal para diferenciar os clones de *Corymbia*.

O carbono fixo do carvão vegetal é considerado um dos principais parâmetros para a avaliação de qualidade, sendo influenciada pelo material de origem e pelo processo de carbonização (NEVES, 2012). Destaca-se que o teor de materiais voláteis, segundo Oliveira et al. (2010), é normalmente inversamente proporcional ao índice de carbono fixo do carvão, o que corrobora e sustenta a coerência dos resultados constatados, uma vez que os coeficientes dessas variáveis apresentaram magnitudes semelhantes em ambas as funções discriminantes.

A maximização da diferença dos clones fica evidente ao analisar a tabela 5, na qual nota-se a formação de 4 “grupos” distintos para as duas variáveis canônicas, de acordo com o teste de Scott-Knott, de modo que a qualidade global do carvão vegetal do clone AMF1119 foi superior aos demais. Interessante destacar que parte dos outliers multivariados detectados englobam justamente as observações referentes ao clone AMF1119, confirmando a superioridade desse genótipo.

Tabela 5 – Médias das variáveis canônicas de 12 clones de eucalipto plantados no Município de Dionísio, MG.

Clone	Variável canônica 1	Clone	Variável canônica 2
AMF1119	73,0840a*	AMF1119	46,6979a
M02	72,3399b	M43	46,4865a
M04	72,0753b	M22	46,3313b
M01	71,9164b	M06	46,1968b
M06	71,8650b	M08	46,0822c
M03	71,8545b	M09	46,0304c
M08	71,8506b	M-S/N	46,0255c
M09	71,5499c	M07	46,0225c
M07	71,5338c	M04	45,8170d
M-S/N	71,0537d	M02	45,7817d
M43	71,0478d	M01	45,7817d
M22	70,7581d	M03	45,6111d

*Clones com médias seguidas de mesma letra pertencem ao mesmo grupo, pelo teste Scott-Knott a 1% de significância.

Ao comparar as análises univariadas com o resultado final da função discriminante (Tabela 6), constata-se que as variáveis canônicas representaram de forma muito mais eficiente a variabilidade em função dos clones, ao passo que o valor da estatística F destas variáveis é consideravelmente maior que os valores para as demais variáveis, especialmente da primeira variável canônica. A acurácia seletiva das variáveis canônicas também superou as demais, ratificando que a função discriminante é melhor para a discriminação dos genótipos avaliados.

Tabela 6 – Valores da estatística F, valor-p e acurácia seletiva (AS) das variáveis observadas em análise univariada e variáveis canônicas.

Variável	F	valor-p	AS (%)
Rendimento da carbonização	5,165	$3,45 \cdot 10^{-5}$	89,80
Teor de materiais voláteis	1,375	0,22	52,22
Carbono fixo	1,110	0,38	31,48
Friabilidade	5,956	$7,01 \cdot 10^{-6}$	91,22
Densidade relativa aparente	4,124	$32,80 \cdot 10^{-4}$	87,04
Poder calorífico superior	2,677	$98,00 \cdot 10^{-3}$	79,15
Variável canônica 1	16,040	$3,95 \cdot 10^{-12}$	97,34
Variável canônica 2	9,459	$1,66 \cdot 10^{-8}$	94,57

4 CONCLUSÕES

A função discriminante de Fisher foi eficiente na sumarização dos dados quando utilizadas as duas primeiras variáveis canônicas, retendo aproximadamente 80% da informação contida no conjunto de seis variáveis. Em comparação às análises univariadas, o uso da função discriminante para gerar novas variáveis que são funções das variáveis originais apresenta vantagens quanto a simplificação da interpretação dos resultados e, principalmente, na diferenciação dos clones em relação à qualidade global do carvão vegetal.

As variáveis de maior influência na discriminação do carvão produzido pelos diferentes clones foram carbono fixo e densidade relativa aparente do carvão. Comparando os clones por meio da variável canônica constata-se que o clone AMF1119 é o mais indicado para a produção de carvão vegetal.

REFERÊNCIAS BIBLIOGRÁFICAS

- ABNT. **NBR 8633: Carvão vegetal: determinação do poder calorífico**. Rio de Janeiro, 1984. 13 p.
- ABNT. **NBR 8112: Carvão vegetal: análise imediata**. Rio de Janeiro, 1986. 8p.
- ABRAF. **Anuário estatístico ABRAF 2013 ano base 2012**. Brasília: ABRAF, 2013. 148p.
- ANDERSON, M.J. (2006) Distance-based tests for homogeneity of multivariate dispersions. **Biometrics**, v.62, p.245–253, 2006.
- BARBOSA, J.J; PEREIRA, T.M.; OLIVEIRA, F.L.P de. Uma proposta para identificação de outliers multivariados. **Ciência e Natura**, v.40, 2018.

CAMPOS, K.A. **Função discriminante de Fisher como alternativa à análise de variância multivariada**. 106f. 2012. Tese (Doutorado em Estatística e Experimentação Agropecuária) – Universidade Federal de Lavras, Lavras, 2012.

CASTRO, A.F.N.M.; CASTRO, R.V.O.; CARNEIRO, A. de C.O.; LIMA, J.E. de; SANTOS, R.C. dos; PEREIRA, B.L.C.; ALVES, I.C.N. Análise multivariada para seleção de clones de eucalipto destinados à produção de carvão vegetal. **Pesquisa Agropecuária Brasileira**, v.48, n.6, p.627-635, 2013.

COSTA, L.J. **Caracterização mecânica do carvão vegetal de clones de Corymbia**. 2016. 90f. Dissertação (Mestrado em Ciência e Tecnologia da Madeira) – Universidade Federal de Lavras, Lavras, 2016.

COUTO, A.M. **Influência das características anatômicas, químicas e físicas da madeira de Eucalyptus e Corymbia na qualidade do carvão para o uso siderúrgico**. 2014. 173f. Tese (Doutorado em Ciência e Tecnologia da Madeira) – Universidade Federal de Lavras, Lavras, 2014.

FAO. **The charcoal transition: greening the charcoal value chain to mitigate climate change and improve local livelihoods**. Roma: FAO, 2017

FERREIRA, D.F. **Estatística multivariada**. 2 ed. rev. e ampl. Lavras: Ed. UFLA, 2011. 675p.

FISHER, R.A. The use of multiple measurements in taxonomic problems. **Annals of Eugenics**, v.7, n.2, p.179-188, 1936.

GOMES, F.P. **Curso de estatística experimental**. 15 ed. Piracicaba: ESALQ, 2009. 451p.

GOULART, M.; HASELEIN, C.R.; HOPPE, J.M.; FARIAS, J.A.; PAULESKI, D.T. Massa específica básica e massa seca de madeira de Eucalyptus grandis sob o efeito do espaçamento de plantio e da posição axial no tronco. **Ciência Florestal**, v.13, p.167-175, 2003.

HAIR JR., J.F.; BLACK, W.C.; BABIN, B.J.; ANDERSON, R.E.; TATHAM, R.L. **Multivariate Data Analysis**. 7 ed. Upper Saddle River: Prentice Hall, 2009. 785p.

HARRIS, R.J. **A primer of multivariate statistics**. Nova York: Academic Press, 1975.

IBÁ. **Dados do Relatório Ibá 2018 – Ano base 2017**. Disponível em: <<https://www.iba.org/datafiles/publicacoes/relatorios/digital-sumarioexecutivo-2018.pdf>> Acesso em: 7 de março de 2019.

MARDIA, K. V. Measures of multivariate skewness and kurtosis with applications. **Biometrika**, London, v. 57, n. 3, p. 519-530, 1970.

MARDIA, K.V. The effect of nonnormality on some multivariate tests and robustness to nonnormality in the linear model. **Biometrika**, v.58, n.1, p.105-121, 1971.

MARDIA, K. V. Applications of some measures of multivariate skewness and kurtosis for testing normality and robustness studies. **Sankhyā: The Indian Journal of Statistics**, v.36, n.2, p. 115-128, 1974.

MARDIA, K. V. Assessment of multinormality and the robustness of Hotelling's T^2 test. **Applied Statistics**, v. 24, n. 2, p. 163-171, 1975.

MINGOTI, S.A. **Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada**. Belo Horizonte: Editora UFMG, 2005.

MOTTA, L. P.; MACHADO, C. C.; SOARES, V. P.; RIBEIRO, J. C.; RIBEIRO, C. A. A. S. Utilização de sistema de informações e da distância virtual na otimização do transporte florestal rodoviário. **Revista Árvore**, v.20, n.3, p.381-394, 1996.

NEVES, T.A. **Qualidade da madeira e do carvão vegetal de clones de Eucalyptus cultivados no Sul de Minas Gerais**. 2012. 97f. Dissertação (Mestrado em Ciência e Tecnologia da Madeira) – Universidade Federal de Lavras, Lavras, 2012.

OLIVEIRA, A.C.; CARNEIRO, A. de C.O.; VITAL, B.R.; ALMEIDA, W.; PEREIRA, B.L.C.; CARDOSO, M.T. Parâmetros de qualidade da madeira e do carvão vegetal de *Eucalyptus pellita* F. Muell. **Scientia Forestalis**, v.38, n.87, p.431-439, 2010.

OLIVEIRA, J.B.; VIVACQUA FILHO, A.; MENDES, M.G.; GOMES, P.A. Produção de carvão vegetal – aspectos técnicos. In: CETEC. **Produção e Utilização de Carvão Vegetal**. Belo Horizonte: CETEC. 1982, p.59-72.

PEREIRA, P.V. da C. **Uma revisão da análise de variância multivariada com aplicações em Biologia**. 2014. 132f. Monografia (Graduação em Estatística) – Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2014.

PROTÁSIO, T. de P.; TRUGILHO, P. F.; NEVES, T. A.; VIEIRA, C. M. M. Análise de correlação canônica entre características da madeira e do carvão vegetal de *Eucalyptus*. **Scientia Forestalis**, v.40, n.95, p.317-326, 2012.

PROTÁSIO, T. de P.; COUTO, A. M.; REIS, A. A. dos; TRUGILHO, P. F. Seleção de clones de *Eucalyptus* para a produção de carvão vegetal e bioenergia por meio de técnicas univariadas e multivariadas. **Scientia Forestalis**, v.41, n.97, p. 015-028, 2013.

R CORE TEAM. **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, 2018.

RESENDE, M.D.V. de. **Genética biométrica e estatística no melhoramento de plantas perenes**. Brasília: Embrapa Informação Tecnológica, 2002. 975 p.

SCHWAGER, S.J.; MARGOLIN, B.H. Detection of multivariate normal outliers. **The Annals of Statistics**, v.10, n.3, p.943-954, 1982.

SIMEÃO, S.F.A.P.; PADOVANI, C.R. Utilização da função discriminante quadrática em ciências experimentais. **Revista Energia na Agricultura**, v.23, n.1, p.116-134, 2008.

SOUZA, U.R. de; SILVA, F.L. da; GRIFFITH, J.J.; LIMA, J.E. de; QUINTELA, M.C.A.; COSTA, E.C.V. Determinantes dos novos contratos de fomento florestal na Mesorregião do Vale do Rio Doce, Minas Gerais. **Revista Árvore**, v.33, n.2, p.377-386, 2009

TRUGILHO, P.F.; LIMA, J.T.; MORI, F.A.; LINO, A.L. Avaliação de clones de Eucalyptus para produção de carvão vegetal. **Cerne**, v.7, n.2, p.104-114, 2001.

VITAL, B.R. **Métodos de determinação da densidade da madeira**. Viçosa, MG: Sociedade de Investigações Florestais, 1984. 21p.

APÊNDICE

Apêndice 1 – Médias do rendimento da carbonização (Rend), teor de materiais voláteis (Volát), carbono fixo (Carb), friabilidade (Friab), densidade aparente (Dens) e poder calorífico superior (PCS).

Clone	Rend(%)	Volát(%)	Carb(%)	Friab(%)	Dens(g cm ⁻³)	PCS(kcal g ⁻¹)
AMF 1119	38,21	28,79	70,61	6,40	0,35	7084,40
M01	36,23	28,67	69,70	5,02	0,41	6910,60
M02	37,50	28,68	69,49	4,93	0,33	6952,00
M03	36,82	28,24	69,66	4,77	0,37	7032,75
M04	36,80	27,81	70,33	5,68	0,43	6840,00
M05	36,21	28,09	70,05	6,55	0,40	6857,00
M06	36,84	27,79	70,50	6,59	0,38	7021,40
M07	36,35	28,07	70,15	5,98	0,41	7160,40
M08	36,35	27,79	70,53	6,24	0,36	6884,40
M09	35,85	26,93	71,28	5,72	0,30	7118,80
M22	35,40	27,85	70,28	7,17	0,33	7264,60
M43	35,72	27,18	71,11	7,60	0,35	7234,25
M-S/N	36,05	27,20	70,45	6,89	0,33	7065,17
Média	36,49	27,93	70,32	6,12	0,37	7032,75
CV	0,021	0,021	0,007	0,143	0,109	0,020
Máximo	38,21	28,79	71,28	7,60	0,43	7264,60
Mínimo	35,40	26,93	69,49	4,77	0,30	6840,00