

Classificação one-class do café agroflorestal produzido no Maciço de Baturité/CE¹.

Aluno: Monis Neves Baptista Manuel
Orientadora: Profa. Dra. Livia Paulia Dias Ribeiro

Resumo

A espectroscopia NIR associada ao método de Classificação de Uma Classe foi aplicada ao controle de qualidade do café especial agroflorestal. Neste estudo, foram utilizadas 34 amostras. Os dados espectrais foram obtidos por meio de um microNIR e foram avaliadas diferentes estratégias de pré-tratamento para correção da linha de base dos dados. Foram realizadas técnicas de reconhecimento de padrões não supervisionadas (PCA e HCA). Foram utilizadas as construções do modelo de classificação utilizando o algoritmo DD-SIMCA com 19 amostras adquiridas diretamente de produtores reconhecidos pelo melhor controle de qualidade do café tipo especial. Para testar o modelo, foram avaliadas 15 amostras do tipo tradicional obtidas nos mercados locais. O modelo de classificação com maior taxa de classificação correta (RCC) foi de 100% e 87% para os grupos de validação e teste, respectivamente. Os resultados demonstraram uma aplicação bem sucedida dessa estratégia para verificar a autenticidade de amostras especiais de café agroflorestal.

Palavras-chave: NIRS; Café agroflorestal; Classificação

Abstract

The NIR spectroscopy associated with the One-Class Classification method has been applied to quality control of agroforestry special coffee. A total of 34 samples was used in this study. Spectral data were obtained using a microNIR and different pre-treatment strategies for baseline correction of the data were evaluated. Unsupervised pattern recognition (PCA and HCA) techniques were performed. The constructions of the classification model using the DD-SIMCA algorithm with 19 samples acquired directly from producers recognized for the best quality control of the special type coffee were used. In order to test the model, 15 samples of the traditional type obtained in local markets were evaluated. The classification model with the highest correct classification rate (CCR) was 100% and 87% for the validation and test groups respectively. The results demonstrated successful application of this strategy for verifying the authenticity of special type agroforestry coffee samples.

Key words: NIRS; Agroforestry coffee; Classification

¹ Trabalho de Conclusão de Curso apresentado ao Curso de Licenciatura em Química do Instituto de Ciências Exatas e da Natureza da Universidade da Integração Internacional da Lusofonia Afro-brasileira, como requisito para a obtenção do título de Licenciado em Química. Defendido e Aprovado em: 23/03/2021.

SUMÁRIO

1 INTRODUÇÃO.....	3
2 MATERIAS E METODOS	5
2.1 Amostras de café	5
2.2 Obtenções dos espectros NIR.	6
2.3 Pré-processamento e análise dos dados	6
2.4 Modelos de classificação de única classe.....	6
2.5 Softwares	7
3 RESULTADOS E DISCUSSÃO	7
4 CONSIDERAÇÕES FINAIS	11
AGRADECIMENTO	12
REFERÊNCIAS	13

1 INTRODUÇÃO

O café é um produto que apresenta elevado consumo e alto impacto econômico mundial. Segundo a Organização Internacional do Café (OIC), nos anos de 2019/2020 a safra deve totalizar 168.864 milhões de sacas. O Brasil é considerado como um dos maiores produtores e exportadores do café no mundo, onde 84% de produção das sacas são destinadas para exportação (OIC, 2020; FORCHETTI et al., 2019). A comercialização global de café permite aos países produtores a exportação da maior parte da produção, que vale em torno de US\$ 20 bilhões por ano (OIC, 2019). O café é consumido diariamente por mais de um bilhão de pessoas de diferentes faixas etárias e fornece um sabor considerável agradável além de vários benefícios à saúde física e mental. A cafeína é o maior constituinte do café e pode está associada à prevenção da depressão e redução de casos de infarto do miocárdio, alguns tipos de câncer, diabetes, previne de doenças como: Alzheimer, parkinson, asma, depressão, sistema cardiovascular (MILANI et al., 2019; MEES et al., 2017; ABIC, 2020). O café ainda favorece no aumento da capacidade de concentração, ânimo e diminuição da fadiga e redução índice de obesidade (SOUTO et al., 2015). O Café agroflorestal produzido na região do Maciço de Baturité/ Ceara- Brasil é cultivado de forma sombreada por árvores altas e o processamento dos grãos é realizado de forma artesanal. Essas características permitem a produção de café do tipo especial, sem uso de defensivos agrícolas, com seleção dos melhores grãos e controle de torra, resultando café de excelente qualidade e com valor comercial elevado. O café produzido no Maciço não apresenta um protocolo padrão de qualidade o que resulta em cafés produzidos com alto controle de qualidade até cafés torrados e moídos sem nenhum padrão. O processo de colheita, preparação e torra dos grãos desses cafés agroflorestal do Maciço de Baturité dão origem variações no manejo pós-colheita. A seleção dos grãos, processo retirada das cascas, secagem e a torra são fatores fundamentais para a qualidade do produto final de café torrado e moído. Quando as etapas de produção do café possuem um padrão definido e controlado pelo produtor, com destaque na seleção dos grãos com mínimas imperfeições e controle do tempo e temperatura da torra, o produto final é denominado café especial. Quando o produtor não se utilizada de controle rigoroso das etapas podendo resultar em vários níveis de

qualidade, o produto é denominado café convencional ou não especial. Em decorrência do seu valor comercial e elevado consumo, lamentavelmente, o café é frequentemente alvo de adulterações e fraudes, que é uma questão que recebe constante atenção da mídia internacional (ABIC, 2020; MILANI et al., 2019). Com intuito de baratear o custo de fabricação e produção do café o mesmo é facilmente adulterado a partir da mistura com materiais não comuns ao alimento, geralmente de baixo custo que altera a qualidade original, visando fins lucrativos. Dentre os adulterantes estão: cascas e caules de café, milho, cevada, centeio, caramelo, feijão, açúcar mascavo, soja, e outros (Tramitacao-PL-7664-2017; SOUTO et al., 2015; NÚÑEZ et al., 2020). Também pode ser realizada substituição dos grãos de café original ou misturas com outros semelhantes de menor valor, qualidade e preço (SOUTO et al., 2015).. A adulteração de alimento é dada como ato ilegal, porque alteram a qualidade do produto e causam danos ao consumidor, especialmente, os de ordem econômica e da saúde. Por essa razão, a determinação da veracidade nos alimentos é uma questão muito importante no controle de qualidade e segurança (Tramitacao-PL-7664-2017; SOUTO et al., 2015). O monitoramento de qualidade do café depende das suas características básicas como: aroma, sabor, e acidez (LUISA et al., 2015). Tradicionalmente essas características são avaliadas por analistas competentes, porém este método sensorial humana é imparcial e não confiável (OKUBO et al., 2019; BURATTI et al., 2015). A implementação de técnicas analíticas é vital para garantir o controle de segurança e qualidade de produtos alimentícios ou suplementos à base de compostos naturais (MEES et al., 2017). Para avaliação de autenticidade e rastreabilidade pode ser utilizada a espectroscopia conciliada com métodos quimiométricos, por apresentam uma maior rapidez e precisão, sendo a espectroscopia no infravermelho próximo (NIR) [13], mais frequentemente escolhida para construção de metodologias propostas com fim de identificação da qualidade (MEES et al., 2017; GIRAUDO et al., 2018; CATELANI et al., 2017; ARAÚJO et al., 2020)]. A espectroscopia de infravermelho (NIR) é uma técnica que está relacionada ao sinal obtido pela interação da radiação infravermelha, nos comprimentos de onda de 750 nm a 2500 nm, com espécies químicas que apresentam principalmente ligações do tipo C-H, S-H, H-O e N-H. A energia nesse intervalo de comprimento de onda é suficiente para promover combinações e sobreposições de vibrações das ligações químicas. Essa técnica tem sido amplamente empregada em estudos de identificação e classificação de adulterantes e qualidade de diversos produtos

agrícolas (ARAÚJO et al., 2020; MEES et al., 2017; BURATTI et al., 2015; CATELANI et al., 2017; OKUBO et al., 2019), assim como avaliar várias características do café (BURATTI et al., 2015; OKUBO et al., 2019). Os espectros obtidos em estudos com NIR, no geral, apresentam uma grande quantidade de informações, que são analisados por modelos matemáticos e estatísticos denominados pertencentes à Quimiometria. O uso de ferramentas quimiométricas facilitam a identificação da impressão digital e caracterização de espécies químicas. Ainda, proporcionam o desenvolvimento modelos capazes de correlacionar os espectros das espécies de referência e sua composição, possibilitando a detecção e quantificação de maneira eficiente (FORCHETTI et al., 2019). Embora os espectros NIR dos produtos de café sejam complexos, devido à forte sobreposição de picos originários de muitas espécies químicas, estudos demonstraram a viabilidade do uso de NIR e Quimiometria para detecção de adulteração em café torrado (GIRAUDO et al., 2018; TOCI et al., 2015).

Este trabalho objetivou-se no desenvolvimento de modelos de classificação capazes de identificar o padrão de qualidade do café agroflorestal do tipo especial produzido na região do Maciço de Baturité, usando a espectroscopia próximo NIR com auxílio de ferramentas Quimiométrica.

2 MATERIAS E METODOS

2.1 Amostragens de café

Foram coletadas 34 amostras de café, nas formas de torrado e moído e grãos torrados, com manejos do tipo agroflorestal especial (18 amostras) e não especial (16 amostras), no período de setembro de 2019 a janeiro de 2020. As coletas foram realizadas nas cidades de Pacoti, Baturité, Mulungu e Guaramiranga, que produtoras de café agroflorestal no Maciço de Baturité. As amostras do tipo grão foram moídas usando moedor (marca Cadence, modelo di grano). A peneira mesh nº30, 28, com abertura de 0,60 mm foi usada para padronização da granulometria de todas as amostras de café.

2.2 Obtenções dos espectros NIR.

Os espectros de refletância difusa na região do infravermelho próximo foram obtidos usando o espectro NIR DLP® NIRscan™ Nano Evaluation Module, Texas Instruments, na faixa de 937 nm a 1655 nm, com um incremento de 2 nm. Frascos de vidro 5 mL foram utilizados como porta amostra. Os espectros das amostras foram registrados, contra espectro de referência de Teflon, em triplicata. A média dos espectros foi utilizada para a realização do modelo de classificação.

2.3 Pré-processamento e análise dos dados

Para correção das variações de linha de base presentes nos espectros originais, foi aplicada uma segunda derivada Savitzky–Golay (SAVITZKY e GOLAY, 1964), polinômio de segunda ordem e janela de 11 pontos. Em seguida, um estudo exploratório para reconhecimento de padrões não supervisionado usando PCA (Análise por Componentes Principais) (JOLLIFFE e CADIMA, 2016), com validação cruzada completa e HCA (Análise Hierárquica de Clusters) (JOHNSON, 1967) com distância euclidiana e métodos Ward's foram aplicados para identificação de similaridades e diferenças entre espectros de infravermelho próximo de café tipo especial e não especial.

Para o estudo de classificação o conjunto de dados foi dividido em dois grupos: Amostras de treinamento: amostras de café especial; Previsão: amostras de cafés especial e não especial. As amostras de treinamento foram selecionadas utilizando algoritmo de seleção de amostras KS (Kennard-Stone) (KENNARD e STONE, 1969) [20]. A identificação e remoção de amostras do tipo especial anômalas e outliers foi realizada utilizando estatística T^2 de Hotelling para determinar os limites da classe a 95% de confiança.

2.4 Modelos de classificação de única classe

Considerando a especificidade das amostras de café especiais, um modelo de classificação de classe única foi realizado usando one class SIMCA (BRERETON, 2011). Neste caso uma PCA foi aplicada ao conjunto de amostras alvo (amostras de café especial de treinamento), sendo determinada a distância de scores (DS) e a

distância ortogonal (DO) de cada amostra alvo com relação aos resultados obtidos com a PCA. Com base nessas distâncias um valor limite ou limiar da área de classificação usando chi-quadrado é determinado para previsão de novas amostras. A quantidade de componentes principais (PCs) selecionadas na modelagem SIMCA deve apresentar um número menor de amostras alvo extremas (falsos negativos) ou outliers (amostras fora do limiar, a uma probabilidade de 95%) (ZONTOV et al., 2019).

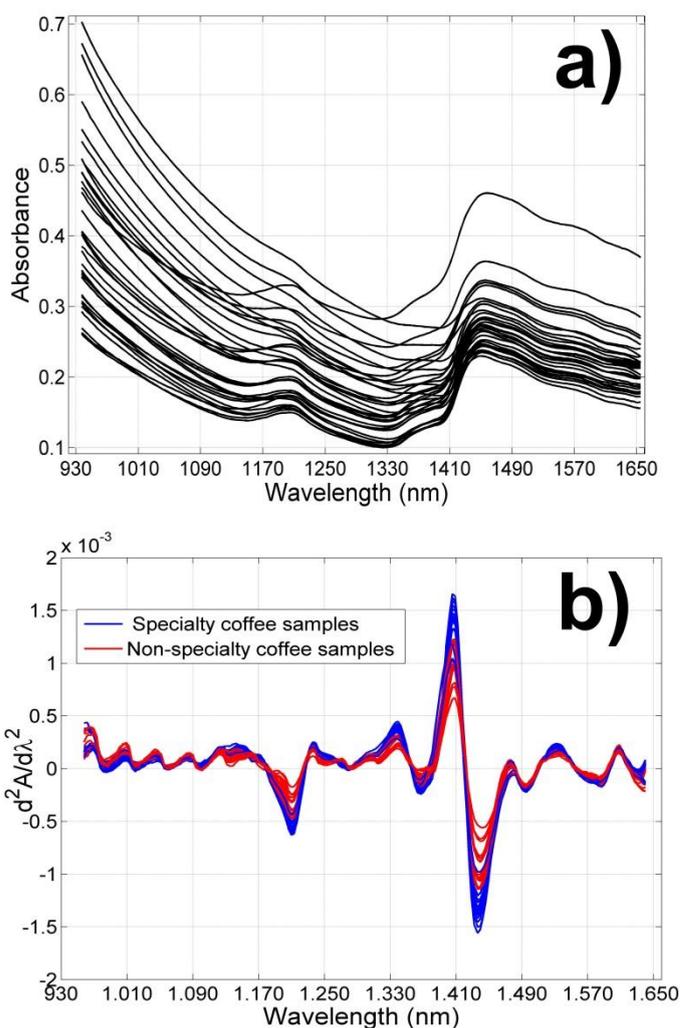
2.5 Softwares

O pré-processamento e análise exploratória (PCA e HCA) foram realizados no The Unscrambler X 10.4. (Camo Analytics, Oslo, Norway). A seleção de amostras Kennard- Stone e a modelagem de classificação de uma classe (interface DD-SIMCA) foram realizadas em software Matlab ® (The Mathworks, Natick, MA, USA)

3 RESULTADOS E DISCUSSÃO

Na Figura 1a são apresentados os espectros brutos infravermelhos próximos das amostras de café especial e não especial estudado. A variação de linha de base presente nos dados foi removida usando a estratégia de pré-processamento Savitzky–Golay, como apresentados na Figura 1b. As bandas mais intensas visualizadas dos dados originais e pré-processados são relacionadas às bandas: 1.100 – 1250 nm (segundo sobreton de C-H, C-H₂ e CH₃ - oriundos de proteínas, lipídios, cafeína e ácidos orgânicos), e 1300- 1490 nm (primeiro sobreton de RN-H das proteínas, primeiro sobreton de O-H de água e ácidos), (ARAÚJO et al., 2020; MEES et al., 2017; BURATTI et al., 2015; CATELANI et al., 2017; OKUBO et al., 2019).

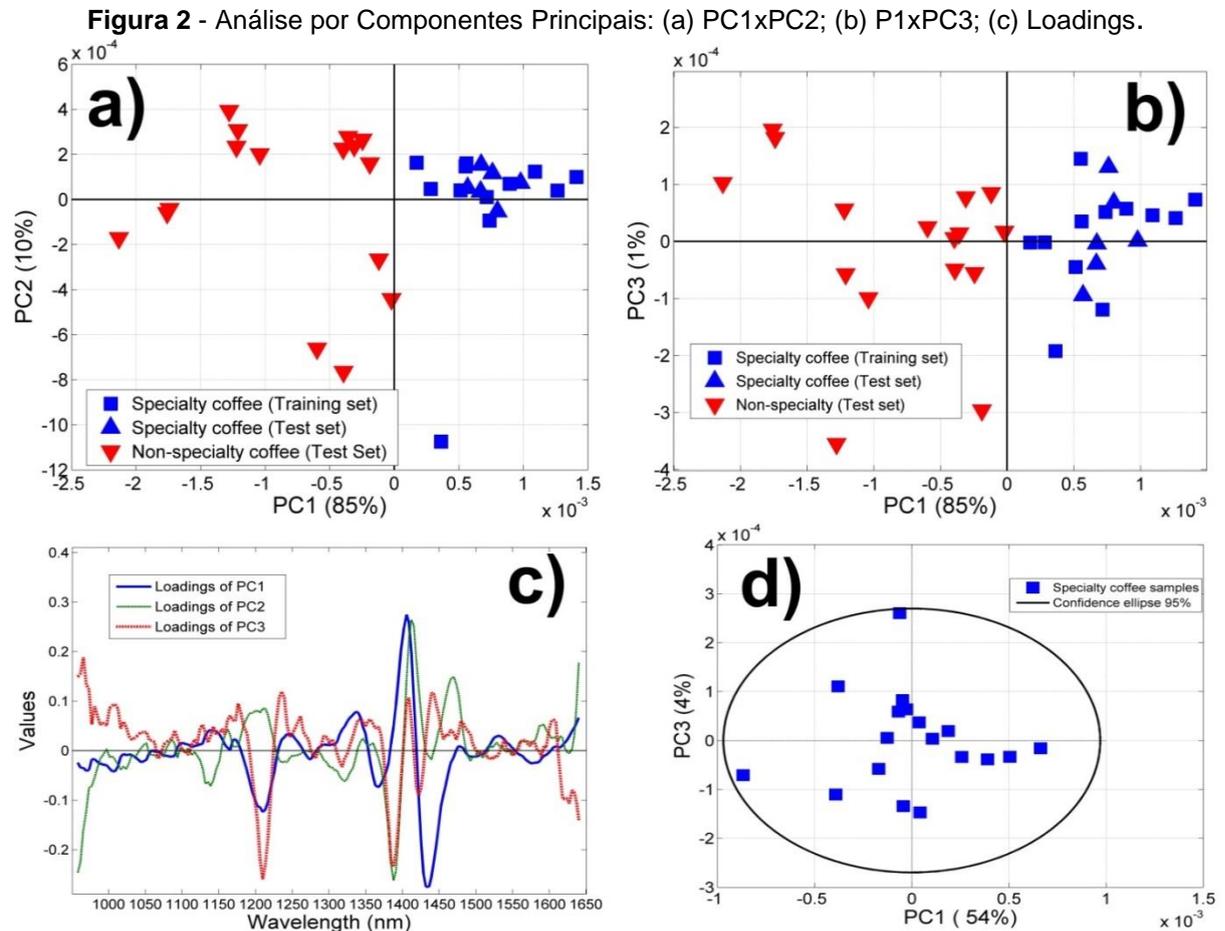
Figura 1 – Dados espectrais: (a) Espectros brutos; (b) Espectros pré-processados usando segunda derivada Savizty-Goley.



Fonte: autor (2020)

Os escores do estudo exploratório usando PCA são representados na Figura 2. Aproximadamente 96% da variância presente nos dados foram explicadas pelas três primeiras componentes principais (PCs): PC1XPC2 (figura 2a) e PC1XPC3 (figura 2b), nas quais se verificou uma diferenciação entre os escores das amostras de café especiais (escores positivos) e não especiais (escores negativos) no componente principal 1 (PC1). Essa diferenciação se deu principalmente por contribuições das faixas espectrais mais significativas e intensas presentes nos dados originais de acordo com o gráfico de loadings (Figura 2c). Na Figura 2d é possível identificar as amostras especiais de previsão que foram selecionadas com Kennard Stones e foram usadas para validação. Para construção do modelo one class SIMCA foram utilizadas 3 componentes principais, as quais permitiram um

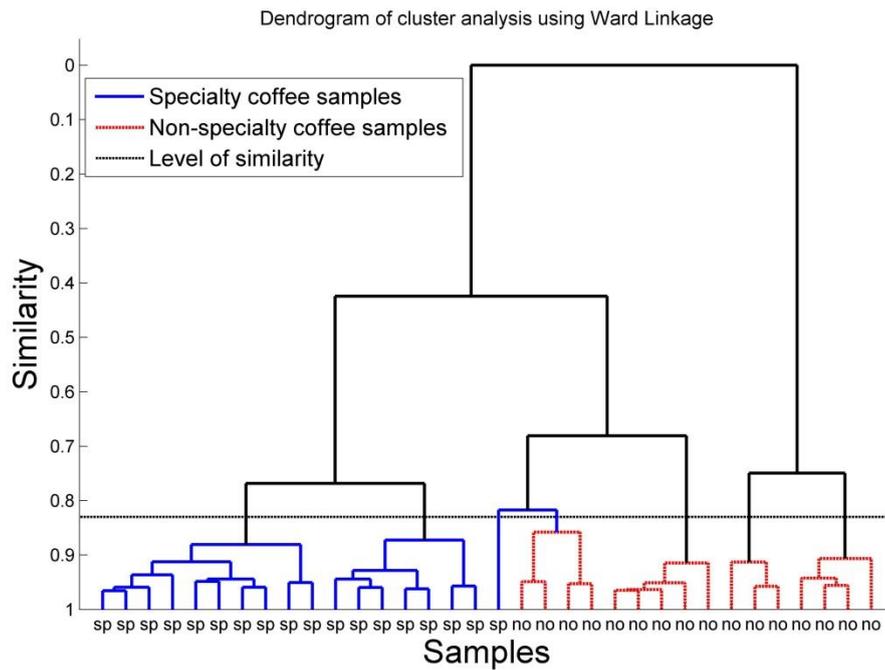
resultado de 0 erros do tipo I (todas amostras do grupo alvo e de validação foram corretamente classificadas) à 0.05, nenhuma outlier foi detectada.



Fonte: autor (2020)

Na Figura 3 é apresentado o dendrograma da análise hierárquica de clusters (HCA). A um nível de similaridade de 0.83 (para similaridade variando de 0 a 1, sendo 0 a maior distância de ligação e 1 a menor distância de ligação possível entre as amostras) foi verificada a formação de clusters apenas com amostras especiais reforçando a possibilidade do uso de espectroscopia NIR para diferenciação entre os dois tipos de café estudados. Dentre os três clusters de café especial formados, um se destacou por apresentar apenas uma amostra. No entanto, nenhuma outlier foi verificada ao avaliar o resultado do t^2 de Hotelling à 95% para as três primeiras componentes principais da PCA para a classe especial.

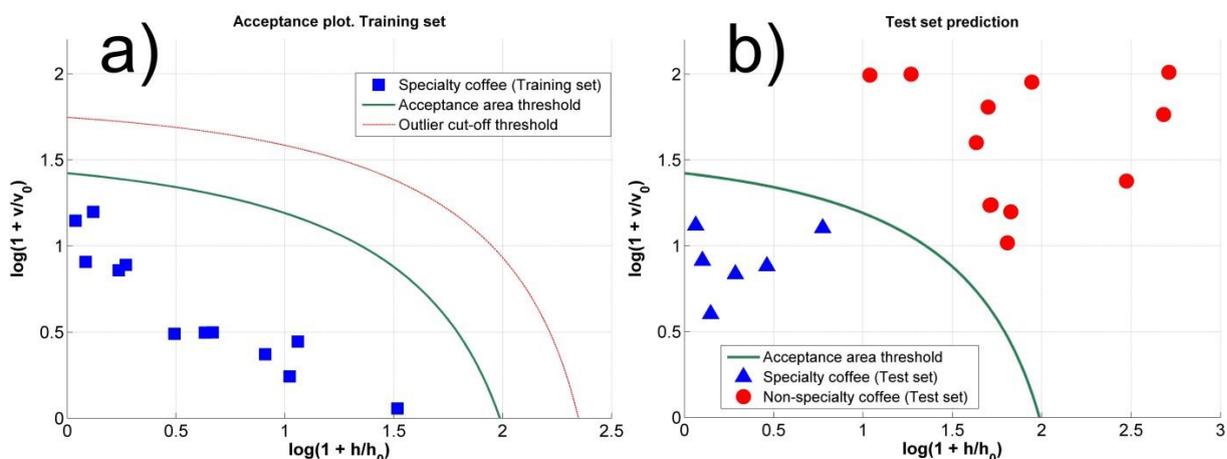
Figura 3 – Análise Hierárquica de Clusters – HCA



Fonte: autor (2020).

Os resultados da modelagem one class SIMCA são apresentados na figura 4a. Para obtenção da área de aceitação de amostras de café especial (linha verde) foram usados 3 componentes principais. De acordo com a figura 4a nenhum erro do tipo I (amostra de treinamento foi classificado como não especial) ou outliers à 95% de confiança (threshold de corte de outliers) foi detectado no grupo de treinamento.

Figura 4 – Modelo one-class SIMCA de classificação: (a) amostras de café especial; (b) Classificação das amostras de café especial e café não especial.



Fonte: autor (2020).

Na figura 4b é apresentada a curva de aceitação do modelo SIMCA (linha verde) e o resultado da classificação das amostras grupo de teste. Todas as amostras de teste foram corretamente classificadas e nenhum erro do tipo II (uma amostra não alvo classificado como especial) á 95% de confiança foi verificado. Também é possível verificar uma não homogeneidade nos escores das amostras classificadas como não especiais com relação ao modelo SIMCA, que pode indicar que existam outras fontes de variabilidade nos espectros NIR que estão relacionadas ás amostras não especiais.

4 CONSIDERAÇÕES FINAIS

O trabalho apresenta uma metodologia baseada no uso de espectroscopia no infravermelho próximo e classificação one class SIMCA para identificação de café agroflorestal especial produzido no Maciço de Baturité, Ceará - Brasil. Estudo de reconhecimento de padrões não supervisionado usando análise por componentes principais (PCA) e análise hierárquica de cluster (HCA) foi aplicado, e foi possível verificar formação de agrupamentos que permitiram de padrões que diferenciam amostras de café especiais e não tradicionais. O estudo classificatório one class SIMCA não apresentou erros do tipo I ou do tipo II a 95% de confiança. Nesse sentido, todas as amostras especiais (alvo) e do grupo de teste foram corretamente identificadas. Metodologia apresentada nesse trabalho é simples, rápida e pode ser aplicada no local e permitiu a obtenção de resultados promissores para identificação

e diferenciação de amostras de café agroflorestal especial obtida no Maciço de Baturité. Os resultados obtidos demonstraram que o uso de espectroscopia NIR associada à estratégia de classificação SIMCA se caracteriza como uma alternativa promissora para verificação da autenticidade de amostras de café agroflorestal especiais.

AGRADECIMENTO

Este trabalho teve o apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e do Instituto Nacional de Ciências e Tecnologias Analíticas Avançadas (INCTAA, processo CNPq 465768 / 20148). A. C. Silva agradece a bolsa do programa DCR (Processo: 300621 / 20205) pela Fundação Cearense de Apoio ao Desenvolvimento Científico e Tecnológico (FUNCAP) e CNPq.

REFERÊNCIAS

ABIC-Brazilian Coffee Industry Association (2020). Retrieved from <https://www.abic.com.br/o-cafe/cafe-e-saude/>. Accessed January 23, 2020.

ARAÚJO, C. DA S. et al. Determination of pH and acidity in green coffee using near-infrared spectroscopy and multivariate regression. **Journal of the Science of Food and Agriculture**, v. 100, n. 6, p. 2488–2493, 12 abr. 2020.

BRERETON, R. G. One-class classifiers. **Journal of Chemometrics**, v. 25, n. 5, p. 225–246, maio 2011.

BURATTI, S. et al. Discrimination between washed Arabica, natural Arabica and Robusta coffees by using near infrared spectroscopy, electronic nose and electronic tongue analysis. **Journal of the Science of Food and Agriculture**, v. 95, n. 11, p. 2192–2200, 30 ago. 2015.

Chamber of Deputies. PL-7664-2017, processing of bills Source. Disponível em: <https://www.camara.leg.br/proposicoesWeb/fichadetramitacao?idProposicao=2137930>. Acesso em: 18 de Janeiro de 2020.

CATELANI, T. A. et al. A Non-invasive Real-Time Methodology for the Quantification of Antioxidant Properties in Coffee During the Roasting Process Based on Near-Infrared Spectroscopy. **Food and Bioprocess Technology**, v. 10, n. 4, p. 630–638, 13 abr. 2017.

FORCHETTI, D. A. P.; POPPI, R. J. Detection and Quantification of Adulterants in Roasted and Ground Coffee by NIR Hyperspectral Imaging and Multivariate Curve Resolution. **Food Analytical Methods**, v. 13, n. 1, p. 44–49, 25 jan. 2020.

GIRAUDO, A. et al. Determination of the geographical origin of green coffee beans using NIR spectroscopy and multivariate data analysis. **Food Control**, v. 99, p. 137–145, maio 2019.

ICO.Statistical Database. International Coffee Organization. Disponível em: through the website: http://www.ico.org/pt/about_statistics_p.asp?section=Estat%EDstica. Acesso em 18 de Janeiro de 2020.

International Coffee Organization. Statistical Database. International Coffee Organization. Disponível em: http://www.ico.org/pt/about_statistics_p.asp?section=Estat%EDstica. Acesso

em 18 de Janeiro de 2020.

JOHNSON, S. C. Hierarchical clustering schemes. **Psychometrika**, v. 32, n. 3, p. 241–254, set. 1967.

JOLLIFFE, I. T.; CADIMA, J. Principal component analysis: a review and recent developments. **Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences**, v. 374, n. 2065, p. 20150202, 13 abr. 2016.

KENNARD, R. W.; STONE, L. A. Computer Aided Design of Experiments. **Technometrics**, v. 11, n. 1, p. 137, fev. 1969.

LUIZA, P. F. et al. Fatty acid profiles and parameters of quality of specialty coffees produced in different Brazilian regions. **African Journal of Agricultural Research**, v. 10, n. 35, p. 3484–3493, 27 ago. 2015.

MEES, C. et al. Identification of coffee leaves using FT-NIR spectroscopy and SIMCA. **Talanta**, v. 177, p. 4–11, jan. 2018.

MILANI, M. I. et al. Authentication of roasted and ground coffee samples containing multiple adulterants using NMR and a chemometric approach. **Food Control**, v. 112, p. 107104, jun. 2020.

NÚÑEZ, N. et al. Authentication of the Origin, Variety and Roasting Degree of Coffee Samples by Non-Targeted HPLC-UV Fingerprinting and Chemometrics. Application to the Detection and Quantitation of Adulterated Coffee Samples. **Foods**, v. 9, n. 3, p. 378, 24 mar. 2020.

OKUBO, N.; KURATA, Y. Nondestructive Classification Analysis of Green Coffee Beans by Using Near-Infrared Spectroscopy. **Foods**, v. 8, n. 2, p. 82, 22 fev. 2019.

TOCI, A. T. et al. Coffee Adulteration: More than Two Decades of Research. **Critical Reviews in Analytical Chemistry**, v. 46, n. 2, p. 83–92, 3 mar. 2016.

WORKMAN JR., J.; WEYER, L. **Practical Guide and Spectral Atlas for Interpretive Near-Infrared Spectroscopy**. [s.l.] CRC Press, 2012.

SAVITZKY, A.; GOLAY, M. J. E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. **Analytical Chemistry**, v. 36, n. 8, p. 1627–1639, 1 jul. 1964.

ZONTOV, Y. V. et al. DD-SIMCA – A MATLAB GUI tool for data driven SIMCA approach. **Chemometrics and Intelligent Laboratory Systems**, v. 167, p. 23–28, ago. 2017.