



UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
DEPARTAMENTO DE QUÍMICA
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA

Tese de Doutorado

**Metodologia baseada em imagem digital, espectros
UV-Vis e quimiometria para *screening* de
adulteração de café por cascas e paus**

Urijatan Teixeira de Carvalho Polari Souto

SAPIENTIA AEDIFICAT

João Pessoa – PB – Brasil

Fevereiro/2017



UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
DEPARTAMENTO DE QUÍMICA
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA

Tese de Doutorado

Metodologia baseada em imagem digital, espectros
UV-Vis e quimiometria para *screening* de
adulteração de café por cascas e paus

Urijatan Teixeira de Carvalho Polari Souto*

Tese apresentada ao Programa de Pós-Graduação em Química da Universidade Federal da Paraíba como parte dos requisitos para obtenção do título de Doutor em Química, área de concentração Química Analítica.

Orientador: Prof. Dr. Edvan Cirino da Silva

Co-Orientador: Prof Dr Mario Cesar Ugulino de Araújo

* Bolsista do Conselho Nacional de Desenvolvimento Científico e Tecnológico

João Pessoa – PB – Brasil

Fevereiro/2017

Catálogo na publicação
Seção de Catalogação e Classificação

S728m Souto, Urijatan Teixeira de Carvalho Polari.

Metodologia baseada em imagem digital, espectros UV-Vís e quimiometria para screening de adulteração de café por cascas e paus / Urijatan Teixeira de Carvalho Polari Souto. - João Pessoa, 2019.

107 f.

Orientação: Edvan Cirino da Silva.

Coorientação: Mário César Ugulino de Araújo.

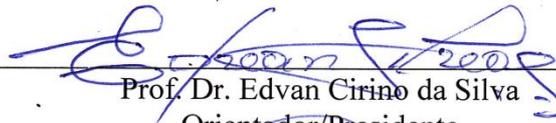
Tese (Doutorado) - UFPB/CCEN.

1. Café-Adulteração. 2. SIMCA. 3. PLS-DA. 4. SPA-LDA.
I. Silva, Edvan Cirino da. II. Araújo, Mário César Ugulino de. III. Título.

UFPB/CCEN

Metodologia baseada em imagem digital, espectros UV-VIS e quimiometria para screening de adulteração de café por cascas e paus.

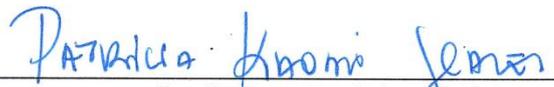
Tese de Doutorado apresentada pelo aluno Urijatan Teixeira de Carvalho Polari e aprovada pela banca examinadora em 17 de fevereiro de 2017.



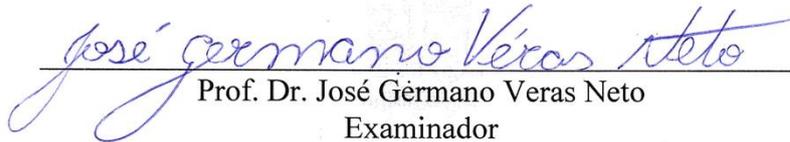
Prof. Dr. Edvan Cirino da Silva
Orientador/Presidente



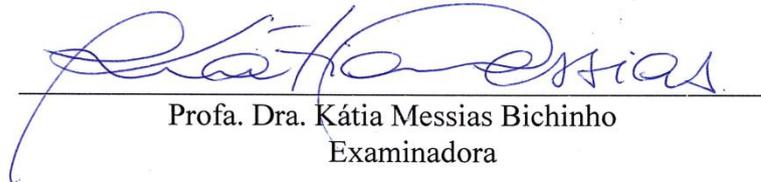
Prof. Dr. Mário César Ugulino de Araújo
2º. Orientador



Profa. Dra. Patricia Kaori Soares
Examinadora



Prof. Dr. José Germano Veras Neto
Examinador



Profa. Dra. Kátia Messias Bichinho
Examinadora



Prof. Dr. Márcio José Coelho de Pontes
Examinador

Dedicatória

Dedico a tese de doutorado aos meus pais: Mariseth Teixeira de Carvalho

e

Fernando Polari Souto

Agradecimentos

- Em primeiro lugar, agradeço ao nosso criador Deus, pelo consentimento da vida a todos os seres que habitam o nosso planeta e as galáxias do Universo.
- Em segundo lugar aos meus queridos bichanos amados: Jose Pitoco, Maria Vitória, Ceguinha, Maria Princesa, Buchãozinha, Buchão, Pretinha, Magrela, Veio, Bebezinha, Pretinho, Magrelo, Galeco, Bradock, Loba, Lobinhas, Estevão, Lobão , Spike e a todos os peludos que amo em meu coração e penso neles todos os dias de minha vida, por me fazerem feliz hoje e sempre.
- Aos meus avós maternos e paternos pelo amor, carinho, respeito, paciência em minha infância.
- Aos meus tios e tias especialmente a tia e mãe Tereza de Carvalho Seixas e ao tio Jose Maria Teixeira de Carvaho.
- Ao meu padrasto Jose Antonio da Silva Junior.
- Aos meus parentes mais próximos.
- A todos os meus alunos do ensino médio que tive a oportunidade de lecionar nas escolas do estado da Paraíba e aos colegas de profissão de ensino.
- Ao Prof Dr Edvan Cirino da Silva, aos esclarecimentos, conselhos, contribuições, e amizade durante todo o doutorado.
- Ao Prof Dr Mario Cesar Ugulino de Araujo, pela oportunidade de trabalho, amizade, respeito e coorientação.
- Aos demais Professores do LAQA pelas contribuições acadêmicas
- A todos que fazem a familia LAQA pela convivência, agradável nestes anos de trabalho em especial a querida Maiara Ferreira Barbosa, Herbeth Vieira (Berbete), ao Prof Dr Paulo Henrique Diniz (Paulinho) e Del Lira Welligton da Silva Lira na finalização do trabalho de tese.
- Ao amigo Sergio Bezerra Ricardo (Serguinho), Taina Kevla, Marcelo, Flaviano Leite, Sofacles Figueredo, David (Negão) e outros.
- Aos pesquisadores do Senai e Nugap especialmente a Dra Lilian Duarte e Dra Giselia Campos pela amizade e doação das amostras de café.
- Ao Sindicafé em especial a Isabela e Camila pelo novo trabalho a ser desenvolvido no LAQA.

SUMÁRIO

	Pág
LISTA DE FIGURAS	x
LISTA DE TABELAS	xiv
LISTA DE ABREVIATURAS	xv
RESUMO	xv
ABSTRACT	xviii
CAPÍTULO 1	
Introdução	1
1-Introdução	1
1.1 Caracterização da problemática e proposta	2
1.2 Objetivos	5
1.2..1 Objetivo geral	5
1.2.2.1 Objetivos específicos	5
CAPÍTULO 2	
Fundamentação Teórica	6
2.0. Histórico, composição química, qualidade e adulteração do café	7
2.1. Histórico	7
2.2 Composição química dos grãos de café cru e torrado	7
2.3 Composição química das cascas e paus do café	7
2.4 Composição química do extrato aquoso do café	9
2.5 Os ácidos clorogênicos, cafeína e trigonelina	10
2.6 Qualidade do café	10
2.7 Determinação da qualidade do extrato aquoso	11
2.8 Adulteração em cafés torrados e moídos	11
2.9. Imagens Digitais	14
2.9.1 Introdução aos modelos de cores	15
2.9.2 Modelos de cores primárias aditivas ou RGB	15
2.9.3 Modelo HSI	17
2.9.4 Modelo HLS	18
2.10 Espectroscopia de absorção molecular na região do ultravioleta e visível	18
2.10.1 Introdução	18

2.10.2 Origem do sinal analítico na região do ultravioleta e visível	19
2.10.3 Características dos espectros de absorção absorção molecular na região do Ultravioleta e Visível	19
2.11 Quimiometria	20
2.11.1 Reconhecimento de padrões	20
2.11.2 Análise por componentes principais PCA	21
2.11.3 Fundamentação teórica do SIMCA	23
2.11.4 Fundamentação teórica da Análise Discriminate Linear- LDA.	25
2.11.5 Algoritmo das Projeções Sucessivas-SPA para classificação usando Análise Discriminante Linear-LDA	26
2.11.6 Fundamentação teórica da Analise Discriminante por Método Mínimos Quadrados parciais PLS-DA.	28
2.11.7 Figuras de mérito para modelos de classificação dos cafés adulterados e não adulterados	30
Sensibilidade	31
Especificidade	31
Acurácia	31
2.12 Revisão bibliográfica	32
2.12.1 Discriminação e classificação de cafés usando espectroscopia de absorção molecular do Ultravioleta-Vísivel (UV-Vis)	32
2.12.2 Discriminação e classificação de cafés com o uso de Imagens Digitais	33
CAPÍTULO 3	
Experimental	36
3.1 Aquisição das amostras de café torrado e moído	43
3.2 Obtenção das imagens digitais	43
3.2.1 Aparato instrumental	43
3.2.2 Procedimento analítico para as imagens digitais	37
3.2.3 Modelagem quimiométrica para as imagens digitais	38
3.2.4 Modelagem quimiométrica para a espectroscopia de absorção molecular do ultravioleta(UV)	38

3.2.5 Algoritmos e Softwares utilizados na construção dos modelos de classificação para as imagens digitais e a espectroscopia de absorção molecular na região do ultravioleta (UV)	39
3.3. Obtenção dos espectros de absorção molecular do ultravioleta e visível (UV-Vis)	40
3.3.1 Aparato instrumental	40
3.3.2 Procedimento analítico para a obtenção dos espectros dos extratos aquosos das amostras do cafés adulterados e nãoa dulterados	40
CAPÍTULO 4	
Resultados e Discussão	43
4.1 Análise baseada em imagens digitais e quimiometria	44
4.1.1 Análise exploratória de adulteração de cafés em imagens digitais	44
4.1.2 Análise screening de adulteração de cafés em imagens digitais	48
4.2 Análise baseada em espectros UV e quimiometria	56
4.2.1 Análise exploratória de adulteração de cafés UV	56
4.2.2 Análise screening de adulteração de cafés do ultravioleta UV	61
CAPÍTULO 5	
Conclusão	68
5.0 Conclusão	69
5.1 Propostas de trabalhos futuros	70
Referências	72
Anêxo	80
Apêndice	90

Lista de Figuras

Figura 1.0 - Fotografia de Grãos de café Arábica(A) e Robusta(B) adaptado de Juliano Ribeiro , 2009.

Figura 2.0 - Fotografia do Pericarpo parte externa da casca de café (exocarpo e mesocarpo) do Atlas de Microscopia.

Figura 2.1 - Fotografia do Pericarpo parte interna da casca de café (endocarpo) e grão do Atlas de Microscopia,2010.

Figura 2.2 - Estrutura molecular do ácido **5-ACQ**.

Figura 2.2.1 - Espectro de absorção **UV** do ácido **5-ACQ**.

Figura 2.3 - Estrutura química da cafeína.

Figura 2.3.1 - Espectro de absorção **UV** da Cafeína.

Figura 2.4 - Estrutura da trigonelina.

Figura 2.4.1 - Espectro eletrônico de absorção **UV** da trigonelina.

Figura 2.9- Cores primárias aditivas: vermelho, verde e azul.

Figura 2.9.1- Representação geométrica do modelo **RGB**.

Figura 2.9.2- Sólido de cor **HSI** Adaptado de Cristiano Bertolini.

Figura 2.9.3- Sólido de cor **HLS** Adaptado de Cristiano Bertolini.

Figura 2.11.1 –Gráfico de **PC1 versus PC2**, onde os círculos representam as amostras - adaptado de Beebe, K.R et al 1988.

Figura 2.11.2 –Ilustração gráfica de modelos **SIMCA** para três classes distintas A, B, e C. As amostras, da classe A ocupam uma linha, da classe B um plano e a classe C um cubo. X e Y são amostras desconhecidas- adaptado de Beebe ,K.R et al 1988.

Figura 2.11.3 - Previsão da amostra Y com um modelo SIMCA para a classe de amostras A. “c” é o resíduo de PCA; “b” é a distância entre a fronteira e a projeção de Y na PC e “a” é a proximidade de Y da caixa A, calculada por: $a^2 = b^2 + c^2$. Adaptado de Beebe, K.R et al 1988.

Figura 3.0- Aparato Instrumental utilizado na obtenção das imagens digitais dos cafés adulterados e não adulterados.

Figura 3.1- Seleção das imagens digitais de cafés adulterados(cascas e paus) pelo programa Delfi 7 desenvolvido no LAQA.

Figura 3.2 --Histogramas das imagens digitais de cafés adulterados (cascas e paus).

Figura 3.3 - Fotografia do espectrofotômetro (modelo HP 8453) utilizado para obtenção dos espectros das 102 amostras de cafés não adulterados(NA) e adulterados(A).

Figura 3.4 Extrações dos extratos aquosos das amostras de cafés.

Figura 4.0 - Histogramas de distribuição de frequências das amostras de cafés torrado e moído não adulterados (**vermelho**) e adulterados (**azul**), com uma matriz de dimensão de (103x3585) amostras e variáveis obtidos pelo Delfi 7.

Figura 4.1. Gráfico dos escores de **PC1 versus PC2** para as amostras de cafés adulterados (**quadrado azul**) e não adulterados (**círculo vermelho**).

Figura 4.2 Gráfico dos escores de **PC1 versus PC3** obtidos para das amostras de cafés adulterados (**quadrado azul**) e não adulterados (**círculo vermelho**).

Figura 4.3 -Gráficos dos pesos do modelo **HLS**.

Figura 4.4 – Histogramas médios com os valores máximo, médio e mínimo do **H L S** correspondentes aos canais (a) **H** (11 a 24), (b) **L** (20 a 90) e (c) **S** (43 a 135) das 103 amostras dos cafés não adulterado (linha preta) e adulterado (**linha azul**).

Figura 4.5 (a) Gráfico do valor do custo de validação G (0,7544) versus número de variáveis selecionadas para o modelo **SPA-LDA** e **Figura 4.5 (b)** Histograma médio das 103 amostras de café torrado e moído indicando as variáveis selecionadas (bolinhas círculo preto), a segunda seta de cor preta em (a) indica o número ótimo de variáveis selecionadas pelo **SPA-LDA** (20 variáveis).

Figura 4.6 – Espectros de absorção **UV** de 102 amostras de café torrado na faixa de 239-380 nm. Em **azul**, são exibidos os espectros das amostras adulteradas em **vermelho**, os espectros das não adulteradas.

Figura 4.7 – Gráfico dos escores de **PC1 versus PC2** obtidos das amostras de cafés adulterados (**quadrado azul**) e não adulterados (**círculo vermelho**).

Figura 4.8 – Gráfico dos escores **PC1 versus PC3** obtidos das amostras de cafés adulterados (**quadrado azul**) e não adulterados (**círculo vermelho**).

Figura 4.9 – Gráfico dos pesos **PC1e PC2** obtidos para as 102 amostras de cafés

Figura 4.9.1 – (a) valor do custo de validação G em função do número de variáveis usadas na modelagem **SPA-LDA** e (b) indicação das 21 variáveis seta preta, selecionadas no espectro médio das 102 amostras de café.

Figura 4.9.2 – Função discriminante de Fisher para as 39 amostras de cafés do conjunto de teste usando as 21 variáveis apresentadas na **Figura 4.9.1 b**.

Lista de Tabelas

Tabela 1.0 – Acurácia dos modelos de cor dos histogramas de frequências.

Tabela 1.1 – Matriz de confusão para *screening* das amostras de cafés pelos modelos **SIMCA, PLS-DA e SPA-LDA** com base nos histogramas das imagens digitais obtidas a partir do café torrado e moído.

Tabela 1.2 – Matriz de confusão para o *screening* das amostras de cafés pelos modelos **SIMCA, PLS-DA e SPA-LDA** com base nos espectros **UV**.

Lista de Siglas e Abreviaturas

ABIC- Associação Brasileira da Indústria e do Café

ANN- Rede Neural Artificial

ANVISA- Agência Nacional de Vigilância Sanitária

Blends - Misturas de vários tipos de café

Fcal -Valor calculado para o teste F

Fcrit -Valor crítico adotado para o teste F

HCA - Análise Hierárquica de Agrupamentos

HLS -Tom, Luminosidade e Saturação

HPLC - Cromatografia Líquida de Alta Performance

HSI - Tom, Luminosidade, Intensidade

ICO - Organização Internacional do Café

KS - Algoritmo de seleção de amostras

LDA- Análise Discriminante Linear

Loadings - Pesos

MAPA- Ministério da Agricultura Pecuária e Abastecimento

NIR- Infravermelho Próximo

PCA- Análise de Componentes Principais

PCs- Componentes principais

PLS-DA - Análise discriminante pelo método dos mínimos quadrados parciais

PQC - Programa de Qualidade do Café

RGB - Vermelho, Verde, azul

Escores- escores

SIMCA - Modelagem independente por analogia de classes

SPA – Algoritmo das Projeções Sucessivas

UV-Vís - Ultravioleta e Visível

Resumo

Título: Metodologia baseada em imagem digital, espectros UV-Vis e quimiometria para screening de adulteração de café por cascas e paus

A qualidade do café depende de diversos fatores relacionados a todas as etapas da produção, desde a escolha da espécie (ou variedades) e das transformações durante a torrefação até o preparo da bebida. Entretanto, a qualidade pode ser alterada pela prática ilícita da adulteração, que consiste na contaminação por cascas e paus e outros adulterantes ao café torrado e moído. Desse modo, a adulteração provoca a alteração das propriedades sensoriais e ocasiona danos onerosos à sociedade devido ao elevado consumo de cafés pelos brasileiros. Em face do exposto, o seguinte trabalho propõe, o desenvolvimento de novas metodologias, simples, rápidas e de baixo custo, que combine o uso das imagens digitais e a espectroscopia de absorção molecular do ultravioleta e visível (**UV- Vis**) com os métodos de reconhecimento de padrão supervisionados como o **SIMCA, PLS-DA e SPA-LDA**, para a construção dos modelos quimiométricos de classificação. De fato, as modelagens construídas discriminaram e classificaram os cafés adulterados dos não adulterados, que podem ser comprovada através dos parâmetros de avaliação de desempenho dos modelos, relacionados a acurácia, sensibilidade e especificidade nos conjuntos de treinamento e teste, apresentados nas **Tabela 1.1 e Tabela 1.2**. Portanto os melhores resultados, foram obtidos para as imagens digitais e o ultravioleta (**UV**) com as modelagens do **PLS-DA e SPA-LDA** em comparação a modelagem do **SIMCA**. Na modelagem das imagens digitais, o **PLS-DA** apresentou os seguintes percentuais de acertos para os conjuntos de treinamento e teste, conforme a **Tabela 1.1: PLS-DA** acurácia (73,0%), sensibilidade (69,0%) e especificidade (76,0%) ; acurácia (97,0%), sensibilidade (100,0%) e especificidade (95,2%) e o **SPA-LDA** acurácia (90,0%), sensibilidade (94,0%) e especificidade (85,0%) ; acurácia (95,0%), sensibilidade (95,0%) e especificidade (95,0%); **SIMCA** acurácia (60,0%), sensibilidade (80,0%) e especificidade (51,1%) ; acurácia (82,5%), sensibilidade (93,3%) e especificidade (76,0%). Na modelagem dos extratos aquosos pelo ultravioleta (**UV**) o **PLS-DA, SPA-LDA e SIMCA** apresentaram os seguintes percentuais de acertos para os conjuntos de treinamento e teste, segundo a **Tabela 1.2: PLS-DA** acurácia (96,8%), sensibilidade (97,3%) e especificidade (96,1%) ; acurácia (97,4%), sensibilidade (100,0%) e especificidade (94,7%); **SPA-LDA** acurácia (100,0%), sensibilidade (100,0%) e especificidade (100,0%) ; acurácia (100,0%), sensibilidade (100,0%) e especificidade (100,0%) ; **SIMCA** acurácia (73,0%), sensibilidade (62,2%) e especificidade (88,5%); acurácia (79,5%), sensibilidade (65,0%) e especificidade (94,7%). Portanto, as propostas de metodologias desenvolvidas, podem assegurar a compra dos cafés comercializados sem a presença do adulterante, cascas e paus. Dessa forma, será possível atenuar danos onerosos a sociedade e auxiliar o controle de qualidade dos cafés aos órgãos de fiscalização e comercialização (**MAPA, ANVISA e ABIC**).

Palavras-chave: café, adulteração, **SIMCA, PLS-DA e SPA-LDA**

Abstract

Title: **Methodology based on digital images, UV-Vis spectra and chemometrics for screening of adulteration of coffee by husks and sticks**

The quality of the coffee depends on several factors related to all stages of production, from the choice of species (or varieties) and from the transformations during the roasting to the preparation of the beverage. However, quality can be altered by the illicit practice of adulteration, which consists of contamination by husks and sticks and other adulterants to roasted and ground coffee. Thus, tampering causes the alteration of sensory properties and causes costly damage to the soil due to the high coffee consumption by Brazilians. In view of the above, the following work proposes the development of new, simple, fast and low-cost methodologies combining the use of digital images and ultraviolet and visible molecular absorption spectroscopy (**UV-Vis**) with the methods of recognition of supervised standards such as **SIMCA**, **PLS-DA** and **SPA-LDA**, for the construction of chemometric classification models. In fact, the constructed models discriminated and classified the adulterated coffees of the non-adulterated coffees, which can be verified through the performance evaluation parameters of the models, related to accuracy, sensitivity and specificity in the training and test sets presented in **Table 1.1** and **Table 1.2**. Therefore, the best results were obtained for the digital and ultraviolet (**UV**) images with the **PLS-DA** and **SPA-LDA** modeling compared to **SIMCA** modeling. In the modeling of the digital images, the **PLS-DA** presented the following percentage of correctness for the training and test sets, according to **Table 1.1**: **PLS-DA** accuracy (73.0%), sensitivity (69.0%) and specificity (76.0%); accuracy (90.0%), sensitivity (94.0%) and specificity (85.0%) and **SPA-LDA** ; accuracy (95.0%), sensitivity (95.0%) and specificity (95.0%); **SIMCA** accuracy (60.0%), sensitivity (80.0%) and specificity (51.1%); accuracy (82.5%), sensitivity (93.3%) and specificity (76.0%). In the modeling of the ultraviolet (**UV**) aqueous extracts the **PLS-DA**, **SPA-LDA** and **SIMCA** presented the following percentage of correct answers for the training and test sets according to **Table 1.2**. **PLS-DA** accuracy (96.8%), sensitivity (97.3%) and specificity (96.1%); accuracy (97.4%), sensitivity (100.0%) and specificity (94.7%); **SPA-LDA** accuracy (100,0%), sensitivity (100,0%) and specificity (100,0%); accuracy (100.0%), sensitivity (100.0%) and specificity (100.0%); **SIMCA** accuracy (73.0%), sensitivity (62.2%) and specificity (88.5%); accuracy (79.5%), sensitivity (65.0%) and specificity (94.7%). Therefore, the proposed methodologies developed can ensure the purchase of coffee marketed without the presence of the adulterant, husks and sticks. In this way, it will be possible to mitigate costly damages to society and help the quality control of the coffee to the inspection and commercialization (**MAPA**, **ANVISA** and **ABIC**).

Keywords: coffee, adulteration, **SIMCA**, **PLS-DA** and **SPA-LDA**

Capítulo 1
Introdução

1.1 Caracterização da problemática e proposta

Segundo a **ICO** (Organização Internacional do Café, 2016) há um destaque do Brasil por ser o maior produtor mundial de café em grãos e o segundo maior consumidor do extrato aquoso. Desse modo, torna-se necessário assegurar a qualidade dos grãos de café que são comercializados. Além disso, o Brasil possui uma considerável posição de país exportador mundial de café arábica (63 %) e café robusta (37%) que totalizam cerca de 34% da produção mundial e 17% do mercado exportador (**ABIC, 2016**). Em 2016, foram produzidas e exportadas 45,49 e 32,0 milhões de sacas com sessenta quilos (**ABIC, 2016**).

Portanto, assegurar a qualidade dos grãos de café é de extrema importância nacional e mundial para o comércio agroexportador, especialmente as espécies de café a *Coffea Arabica* L (Arabica) e *Coffea Canephora Pierre* (Robusta) (**Nebesny & Budryn, 2006; Van der Vossen, 2009**).

Diversos fatores afetam a qualidade dos grãos comercializados, dentre os quais destacam-se as etapas da colheita, beneficiamento, seja por via seca ou úmida e da estocagem e por fim a prática ilícita das fraudes, sendo de caráter intencional ou não o ato de fraudar o café torrado e moído (**MAPA, 2016**).

Dentre as fraudes praticadas ao café torrado e moído, a mais comumente refere-se à contaminação dos resíduos do café, as cascas e paus, resultante das etapas de colheita e beneficiamento dos grãos (**Sano, Assad Cunha, Correa, Rodrigues, 2003 ; Toledo, Hantao, Ho, Augusto, & Anderson, 2014**).

Na tentativa de burlar os órgãos de fiscalização, bem como os consumidores de café, a contaminação por cascas e paus, decorre do processo de torrefação que consiste em torrar os grãos de café e os demais contaminantes (milho, trigoilho, cevada , açai) com o mesmo grau de torra(torra clara, média e escura) ou elevar seu grau de torra, por exemplo, de um torra clara e média para torra escura, a presença destes são mascarados, devido à absorção do

óleo e aderência das partículas mais finas de café torrado e moído as suas superfícies, tornando difícil o reconhecimento sem o auxílio de métodos analíticos adequados (Sano, Assad, Cunha, Correa, & Rodrigues, 2003; Toledo, Hantao, Ho, Augusto, & Anderson, 2014).

A presença de contaminantes proporcionam alterações indesejáveis tanto ao café torrado e moído quanto ao seu extrato, por afetar as características sensoriais como o sabor, corpo, aroma, acidez, amargor que são propriedades percebidas apenas por especialistas da análise sensorial, quando os contaminantes são adicionados em pequenas quantidades.

Assim, as alterações sensoriais não são facilmente detectadas pelos consumidores, o que impossibilita a sua identificação e distinção (Murray, Delahunty, & Baxter, 2001; Sano et al., 2003; Tavares et al., 2012). Portanto os órgãos de fiscalização e comercialização no Brasil, como a Instrução Normativa nº 16 de 24/05/2010 do Ministério da Agricultura, Pecuária e Abastecimento (MAPA, 2016) e a Portaria nº 377, de 26 de Abril de 1999 da Agência Nacional de Vigilância Sanitária (ANVISA, 2016) e a Associação Brasileira da Indústria e do Café (ABIC, 2016) buscam atenuar a prática ilícita das adulterações e contaminações, como exemplo o projeto inicial do Programa de Qualidade do Café (PQC) proposto pela Associação Brasileira da Indústria e do Café (ABIC, 2016).

Nesse sentido destaca-se a Portaria nº 377 da Agência Nacional de Vigilância Sanitária (ANVISA, 2016) que estabelece um padrão de qualidade para o café torrado e/ou moído, em função das características físicas, sensoriais e químicas, a qual fixou os teores máximos de umidade, resíduo mineral fixo (cinzas) e cafeína e teores mínimos de extrato aquoso e o extrato etéreo (ANVISA, 2016). Entretanto, as análises realizadas com esse intuito são invasivas, laboriosas, onerosas e necessitam de uma experiência considerável do analista no diagnóstico dos resultados obtidos em especial a adulteração por cascas e paus (Sano, Assad, Cunha, Correa, Rodrigues, 2003).

Em função da apresentação da problemática, o seguinte trabalho propõe o desenvolvimento de novas metodologias analíticas simples, rápidas e de baixo custo para auxiliar o controle de qualidade dos cafés comercializados, torrados e moídos e seus extratos aquosos, sobre a adulteração por cascas e paus.

1.2 Objetivos

1.2.1 Objetivo geral

Desenvolver novas metodologias simples, rápidas e de baixo custo através da análise screening de cafés adulterados e não adulterados por cascas e paus com a utilização das imagens digitais e espectros **UV-Vis**. Pretende-se que essas ferramentas auxiliem os laboratórios especializados no controle de qualidade de cafés.

1.2.2. Objetivos específicos

Construir modelos quimiométricos para fins de classificação, a partir da aquisição das imagens digitais dos cafés torrados e moídos adulterados e não adulterados por cascas e paus e a utilização das técnicas de reconhecimento de padrão supervisionada, como o **SIMCA, PLS-DA e SPA-LDA**, para atestar a qualidade dos cafés comercializados.

Explorar as regiões do espectro eletromagnético ultravioleta e visível (**UV-Vís**) especialmente a região de trabalho do ultravioleta (**UV**), para a obtenção dos espectros de absorção dos extratos aquosos de cafés adulterados por cascas e paus e não adulterados.

Posteriormente, construir e validar modelos quimiométricos para fins de classificação usando as técnicas de reconhecimento de padrão supervisionadas tais como : modelagem independente por analogia de classe (**SIMCA**), análise discriminante linear (**SPA-LDA**) e análise discriminante pelo método dos mínimos quadrados parciais (**PLS-DA**).

Capítulo 2

Fundamentação Teórica

2.0 Histórico, composição química, qualidade e adulteração do café

2.1. Histórico

Segundo a literatura especializada, o café não é nativo do Brasil, sendo descoberto casualmente na região da Etiópia, pela prática de criação de cabras dos pastores do século XV (Cirilo, M P G 2001 ; Juliano Ribeiro, 2009). Porém, com o consumo de seus extratos aquosos pelos monges, pastores e os árabes, o café se tornou bastante popular no mundo e demais países da Europa, com a chegada ao Brasil no século XVIII (Cirilo, M P G, 2001 ; Juliano Ribeiro, 2009).

Segundo a organização internacional do café (ICO, 2016) o Brasil no século XXI apresenta os melhores índices de exportação e consumo interno de café, fato que causa preocupação em manter um padrão de qualidade dos cafés comercializados e exportados.

Desse modo, é preciso evitar as práticas ilícitas da adulteração dos cafés, para que não haja alterações indesejáveis em sua composição química e com isso, afete os perfis sensoriais de seus extratos aquosos (Cirilo, M P G, 2001; Juliano Ribeiro, 2009).

Em virtude dessa necessidade, há atuação dos órgãos de fiscalização e comercialização dos cafés, como o Ministério da Agricultura Pecuária e Abastecimento (MAPA, 2016) e a Associação Brasileira da Indústria e do Café (ABIC, 2016) em assegurar um padrão de qualidade dos cafés comercializados em função de suas características químicas, físicas e sensoriais, evitando assim as adulterações ocasionadas pela contaminação de impurezas e matérias estranhas.

2.2. Composição química dos grãos de café cru e torrado

As *especies* dos cafés arábica e robusta e suas variedades, apresentam uma grande diversidade de constituintes químicos em sua composição e possuem aspectos morfológicos distintos, sendo ilustrado na **Figura 1.0** - Fotografia de Grãos de café Arábica(A) e Robusta(B) Adaptado de Juliano Ribeiro, 2009.

Os grãos de café cru e torrado apresentam a seguinte composição química: cafeína, trigonelina, os ácidos clorogênicos (ácido caféico) lipídeos, ácidos alifáticos, oligossacarídeos, polissacarídeos, aminoácidos, proteínas, melanoidrinas e minerais (Charrier, A e Berthaud, J et al, 1975 ; Pimenta, C.J, 1985 ; Cirilo, M P G, 2001 ; Juliano Ribeiro, 2009).

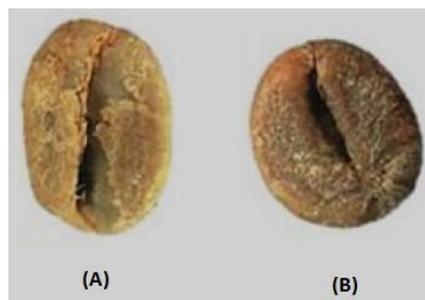


Figura 1.0 - Fotografia de Grãos de café Arábica(A) e Robusta(B) Adaptado de Juliano Ribeiro , 2009.

2.3 Composição química das cascas e paus do café

Dentre as etapas de beneficiamento do café, que correspondem a colheita, secagem dos grãos e estocagem, os resíduos finais são as cascas e paus de café, que foram extraídas através dos processos via seca ou via úmida. Sendo um produto secundário, as cascas e paus de café podem ser utilizadas na alimentação de ruminantes, adubo, cobertura morta e na prática ilícita da adulteração(Barcellos et al 1997 (a) e (b) ;Chalfon, S.M, et al 2008 ; Aretha, 2009; Pereira et al 2014; Aquino et al 2014).

Em relação a composição química das cascas de café, há uma grande diversificação químico bromatológica por parte da literatura especializada (Barcellos et al 1997 (a) e (b) ;Chalfon, S.M, et al 2008 ; Aretha et al, 2009; Pereira et al 2014; Aquino et al 2014).

Entretanto, seja a casca melosa ou a casca de café, a diferença principal está na ausência de pergaminho na casca melosa, que decorre dos processos de beneficiamento, seja por via seca ou úmida, ilustradas nas Figuras 2.0 e Figura 2.1 - Adaptado do Atlas de Microscopia, 2010 -Pericarpo parte externa (exocarpo) e interna(mesocarpo) da casca de café. Portanto, as cascas de café, apresentam os seguintes constituintes químicos: cafeína, proteína bruta, fibras, celulose, hemicelulose, lignina, compostos fenólicos, trigonelina, ácidos orgânicos, carboidratos, amido, nitrogênio, extrato etéreo, lipídeos, matéria seca e minerais (Barcellos et al 1997 (a) e (b), Chalfon, S.M, et al 2008 ;Aretha et al, 2009; Pereira et al 2014; Aquino et al 2014).



Figura 2.0- Fotografia do Pericarpo, parte externa da casca de café e interna (exocarpo e mesocarpo) Adaptado do Atlas de Microscopia,2010.



Figura 2.1 -Fotografia do Pericarpo parte interna da casca de café (endocarpo) e semente de café(grão) Adaptado do Atlas de Microscopia, 2010.

As cascas de café, possuem uma considerável similaridade de composição química com os grãos crus ou torrados e moídos, mas apesar disso há variações nos níveis de concentração da composição química das cascas, quando comparada aos grãos de café. Além disso, os extratos das cascas de café, possuem exclusivamente uma maior concentração de ácido cítrico e ácido quinico protonado, bem como uma baixa concentração de açúcares, ausência de ácidos clorogênicos e ácido cafeico (**5-ACQ**) e menor concentração de cafeína e trigonelina (**Aretha et al, 2009 ;Andrade et al 2009 (a), Aquino et al 2014 ; Pereira et al 2014**).

2.4 Composição química do extrato aquoso do café

Os constituintes majoritários do extrato aquoso correspondem aos ácidos clorogênicos, cafeína , trigonelina, açúcares e melanoidrinas . Segundo a literatura especializada, são facilmente solubilizados em água quente. Assim, a presença e a quantidade dessas substâncias no extrato aquoso, oferece uma indicação da boa qualidade da bebida, bem como uma possível distinção entre os diferentes tipos de cafés adulterados e não adulterados(**Charrier, A e Berthaud, J et al , 1975; Pimenta,C.J ,1985; Barcellos et al 1997 (a) e (b) ; Pereira, R G F.A et al 1997**).

Além disso, os extratos de cafés apresentam sua composição química os seguintes ácidos : o ácido cafeico desidratado, ácido quinico desidratado, ácido cafeico e ácido fenil acetico, não presentes nos extratos das cascas e paus e maior concentração de cafeína e

trigonelina (Cirilo, M P G ,2001; Juliano Ribeiro, 2009; Aretha et al, 2009; Pereira et al 2014; Aquino et al 2014).

2.5 Os ácidos clorogênicos, cafeína e trigonelina

Os ácidos clorogênicos são os principais compostos fenólicos presentes no extrato aquoso do café, sendo dependentes da formulação dos *blends* (mistura de grãos de variedade e/ou espécies de café). Entretanto, eles podem ser degradados durante a torrefação em virtude da sua instabilidade frente ao tratamento térmico. Esses compostos apresentam propriedades sensoriais variadas, entre as quais destacam-se o amargor e a adstringência.

Os principais grupos de ácidos clorogênicos são: cafeoilquínicos, feruloilquínicos e os dicafeoilquínicos, sendo cada um constituído de três isômeros (Pimenta, C.J ,1985; Pereira, R G F.A 1997; Cirilo, M P G ,2001). Contudo, o ácido clorogênico mais relevante presente no extrato aquoso é o ácido caféico (5-ACQ) cuja estrutura é apresentada na Figura 2.2 e na Figura 2.2.1 corresponde ao o espectro de absorção do UV do ácido caféico-5 ACQ

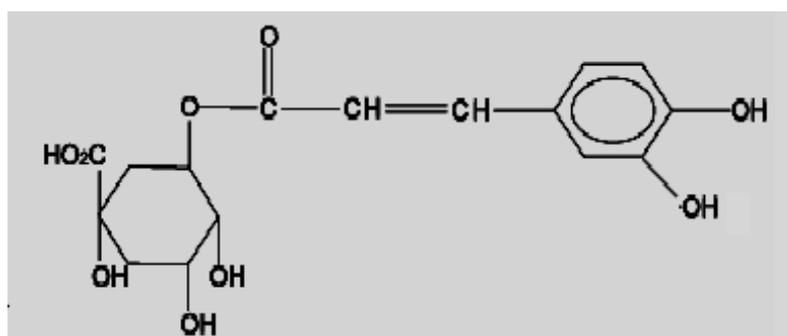


Figura 2.2 - Estrutura molecular do ácido 5-ACQ.

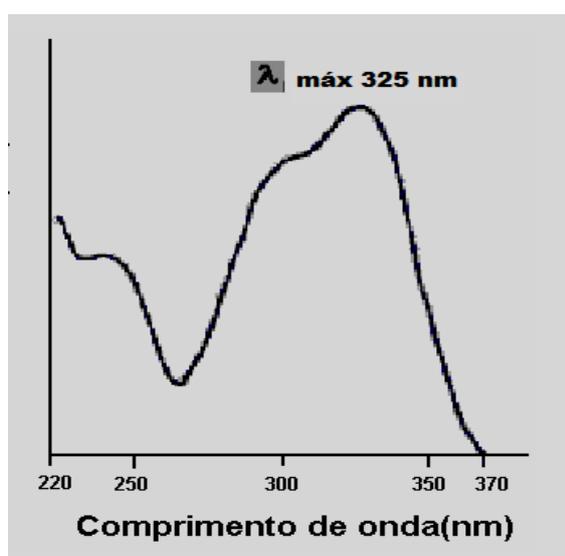


Figura 2.2 .1 - Espectro de absorção UV do ácido 5-ACQ.

A cafeína contribui para o amargor do extrato aquoso, porém sua presença é fundamental em virtude de suas propriedades fisiológicas (Pimenta, C.J, 1985; Pereira, R G F.A et al 1997; Cirilo, M P G ,2001). Os teores de cafeína no extrato aquoso dependem também da elaboração dos *blends* e do tipo de processamento durante a torrefação, em especial o processo de descafeinação. A estrutura química da cafeína encontra-se ilustrada na Figura 2.3 e na Figura 2.3.1 o espectro de absorção da cafeína no UV.

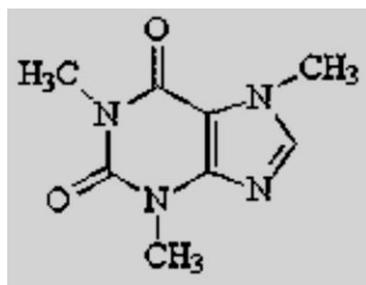


Figura2.3 - Estrutura química da cafeína.

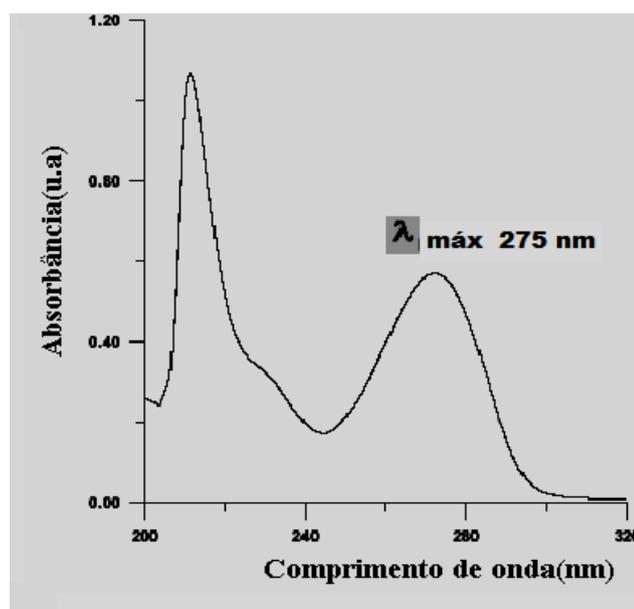


Figura 2.3.1 - Espectro de absorção UV da Cafeína.

A trigonelina (Figura 2.4) não apresenta estabilidade térmica sendo bastante sensível a torrefação. A utilização de *blends* diferentes não afeta significativamente o teor final de trigonelina no produto, pois os cafés do tipo coffea Árabica L e Robusta apresentam quantidades similares de trigonelina, sendo ilustrada a presença da cafeína na Figura 2.4 e na Figura 2.4.1 o espectro de absorção do UV da trigonelina.

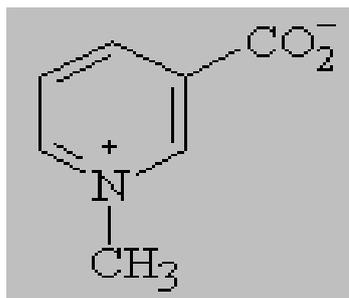


Figura 2.4 - Estrutura da trigonelina.

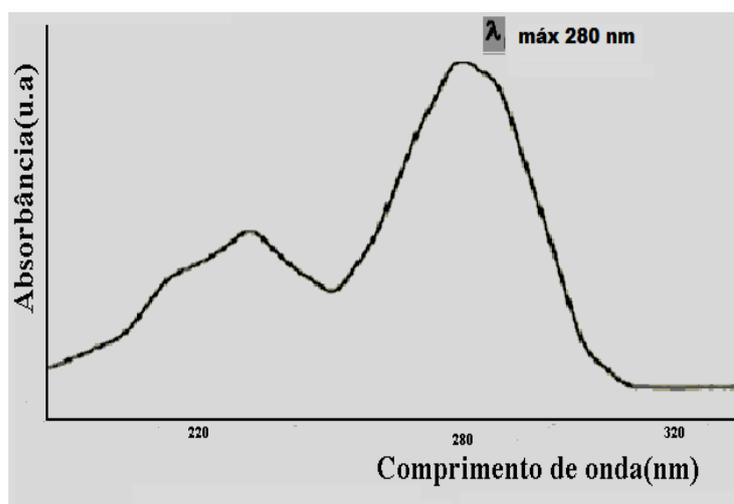


Figura 2.4.1 - Espectro eletrônico de absorção **UV** da trigonelina.

2.6 Qualidade do Café

A qualidade do café depende de diversos fatores relacionados a todas as etapas da produção, ou seja, desde a escolha da espécie (ou variedade) e das transformações durante a torrefação até o preparo da bebida. De fato, a composição química dos grãos, bem como os métodos de colheita, processamento e armazenamento, afetam significativamente a qualidade da bebida do café (Aretha, 2009; Pereira et al 2014; Aquino et al 2014).

Não obstante a influência desses vários fatores, não se pode deixar de destacar a estreita correlação entre a qualidade e a composição química do café que pode ser alterada pela prática ilícita da adulteração dos cafés torrado e moído por cascas e paus, que causam danos a sociedade que podem ser de ordem social, econômica e ambiental (Aretha, 2009; Pereira et al 2014; Aquino et al 2014).

Portanto, a qualidade do café torrado e moído é afetada, especialmente sobre os atributos sensoriais, que podem ser percebidos por especialistas que realizam a análise sensorial.

2.7 Determinação da qualidade do extrato aquoso

A determinação da qualidade da bebida do café é comumente realizada por meio da análise sensorial conhecida como *prova da xícara* (Pimenta, C. J, 1985 ; Pereira et al, R G F.A, 1997). Nesta análise, provadores treinados experimentam o café (a ser degustado) e o classifica de acordo com os diferentes padrões de bebida em função do sabor e aroma (Pimenta,C.J,1985 ;Pereira et al, R G F.A, 1997).Contudo, para adquirir a experiência necessária para distinguir os atributos sensoriais para classificação da bebida, é necessário que o profissional seja treinado por um período de tempo considerável e além do mais a análise é laboriosa e dispendiosa.

2.8 Adulteração em cafés torrados e moídos

Apesar dos órgãos de fiscalização e comercialização como a Agência Nacional de Vigilância Sanitária (ANVISA, 2016), Ministerio da Agricultura Pecuaria e Abastecimento (MAPA, 2016) e a Associação Brasileira da Indústria do Café (ABIC, 2016) atuarem com a finalidade de evitar a fraude, por cascas e paus, a mesma perdura no século XXI.

Destacando-se as impurezas e matérias estranhas, que são contaminates encontrados nas amostras de cafés comercializados, como exemplo: cascas e paus, milho, cevada, triguilhos, açaí, açúcar mascavo, soja, feijão, torrões, que alteram a sua qualidade e causam danos ao consumidor principalmente os de ordem econômica (Sano, Assad ,Cunha, Correa, & Rodrigues, 2003 Toledo ; Chalfon, S.M, et al 2008; Tavares et al., 2012; Hantao, Ho, Augusto, & Anderson, 2014).

Alguns fatores físicos e químicos, contribuem para disfarçar a contaminação por cascas e paus, tais como: ao aspecto exterior granuloso do café, a sua textura oleosa aderente , cor dos grãos que apresentam tonalidades do castanho-avermelhado ao pardo-escuro, com isso se tornam imperceptíveis ao reconhecimento visual humano. Por outro lado quando são submetidas ao processo de torrefação, que envolva o mesmo grau de torra ou a processos de torra distintos, como exemplo, a torra escura, tal fato contribui para mascarar a presença dos contaminates por cascas e paus e outros (Sano, Assad,Cunha, Correa, & Rodrigues, 2003 Toledo ; Chalfon, S.M, et al 2008;Tavares et al., 2012 ; Hantao, Ho, Augusto, & Anderson, 2014).

Desta forma as substâncias adicionadas, são mascaradas pela absorção do óleo e aderências das partículas finas do pó de café a sua superfície, tornando-se difícil o seu reconhecimento sem o auxilio de equipamentos e métodos analíticos especiais, como exemplo, a metodologia de referência que utiliza da análise histológica e gravimétrica para

determinar a contaminação por cascas e paus (Sano, Assad, Cunha, Correa, & Rodrigues, 2003 Toledo ; Tavares et al., 2012; Hantao, Ho, Augusto, & Anderson, 2014).

Para a identificação e quantificação de matérias estranhas e impurezas presentes no café, a análise comumente empregada refere-se a microscopia eletrônica, empregada pelos órgãos de fiscalização (MAPA, ANVISA). De fato, a análise consiste em um exame visual de lâminas da histologia de cada espécie vegetal, quando submetidas ao microscópio pelo analista. A sua quantificação, baseia-se na comparação do percentual do extrato aquoso da amostra analisada, com amostra do café puro (Sano, Assad, Cunha, Correa, & Rodrigues, 2003 Toledo ; Tavares et al., 2012; Hantao, Ho, Augusto, & Anderson, 2014).

2.9. Imagens Digitais

As imagens digitais podem ser caracterizadas por representação de objetos ou cenais reais, tal fato pode ser demonstrado por uma função matemática bidimensional ou tridimensional. Considerando uma função bidimensional do tipo $B = f(X, Y)$ na qual as coordenadas espaciais X e Y descrevem a imagem em termos de valores numéricos, sendo que para cada valor de f , em qualquer ponto da imagem há uma proporcionalidade relacionada ao brilho (Thiago Cesar 2008; Lyra, Wellington 2009, Mariana Godinho 2014, Cristiano Bertolini 2010, Pedro Ivo, 2014).

De fato a imagem registrada pela visão humana ou por qualquer dispositivo, como exemplo câmara fotográfica, possui uma natureza contínua e a partir do processo de digitalização da imagem, há conversão da função contínua em uma função discreta e por fim o seu armazenamento e manipulação nos computadores em uma sequência de bits.

Portanto a imagem contínua é convertida em uma imagem digital, constituída por uma estrutura quadriculada de pequenos quadrados, denominados de pixels e suas respectivas intensidade (Thiago Cesar 2008; Lyra, Wellington 2009; Mariana Godinho 2014; Cristiano Bertolini 2010; Pedro Ivo, 2014).

No caso das imagens coloridas a informação, consideravelmente significativa refere-se a cor, que pode ser compreendida e percebida através de suas propriedades relacionadas a luminância e cromância. Portanto a cor, pode ser representada por sistemas de coordenadas espaciais tridimensionais, que são associadas aos modelos de cores, como exemplo o RGB, HSI, HSV e outros apresentados na seção 2.9.1 (Thiago Cesar 2008; Lyra, Wellington 2009; Mariana Godinho 2014; Cristiano Bertolini 2010; Pedro Ivo, 2014).

2.9. 1 Introdução aos modelos de cores

Os modelos de cores possuem o propósito de especificar as cores a partir de uma padronização específica, que pode tornar aceitável a partir da captação das cores através da visão humana. Desse modo, um modelo de cor corresponde a uma representação de um sistema de coordenadas tridimensionais e possui um subespaço localizado em seu interior, no qual cada cor pode ser representada por um único ponto. Os modelos de cores, podem ser direcionados para diversas finalidades, desde monitores coloridos, escâneres, impressoras, manipulação gráfica e outros (Thiago Cesar 2008; Lyra, Wellington 2009; Mariana Godinho 2014; Cristiano Bertolini 2010; Pedro Ivo, 2014).

2.9.2. Modelos de cores primárias aditivas ou RGB

O tratamento das cores a partir do modelo de cor **RGB** relaciona-se diretamente com a intensidade da percepção humana das cores pela visão humana. Em função disso e a partir da interação das cores que são as primárias aditivas, em diferentes combinações e níveis variados de intensidade. Assim torna-se possível simular todas as cores existentes no espectro da região do visível (Thiago Cesar 2008; Lyra, Wellington 2009; Mariana Godinho 2014; Cristiano Bertolini 2010; Pedro Ivo, 2014).

A combinação entre duas cores primárias aditivas puras resultará em uma cor secundária (ou primária subtrativa) representado pelo diagrama mostrado na **Figura 2.9**. As cores secundárias, ciano, magenta e amarelo, são as cores opostas ao vermelho, verde e azul, respectivamente.

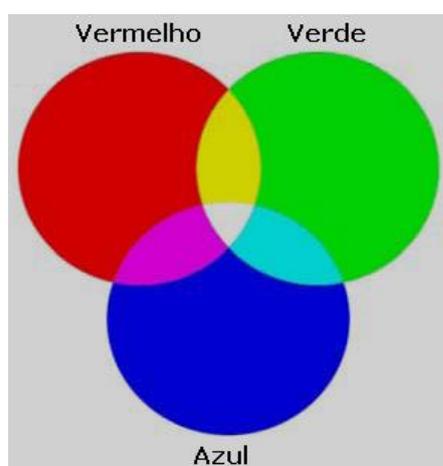


Figura 2.9 -Cores primárias aditivas: vermelho, verde e azul. Adaptado de Lyra, Wellington 2009.

Portanto, devido a sensibilidade de interpretação humana, em detectar as cores através do olho humano, há uma motivação por parte das indústria de eletrodomésticos de produzir

escâneres, monitores de televisão, cameras fotográficas, computadores e outros, que são o resultado de uma cópia e interpretação das cores, resultantes do estímulo das luzes, vermelha, verde e azul da visão humana(Thiago Cesar 2008; Lyra, Wellington 2009; Mariana Godinho 2014; Cristiano Bertolini 2010; Pedro Ivo, 2014).

Podemos representar geometricamente um sistema de cores, em um espaço tridimensional, tal representação resulta em um sólido conhecido como cubo de cor representado pela **Figura 2.9.1**, Adaptado de Thiago Cesar, 2008.

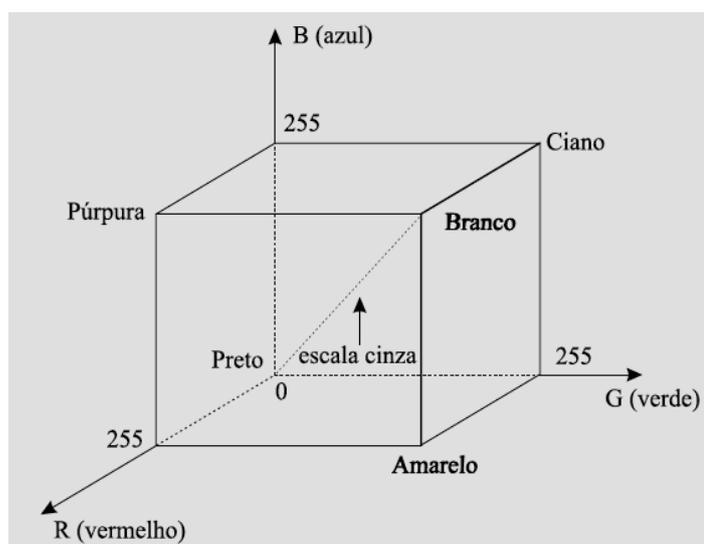


Figura 2.9.1 .- Representação geométrica do modelo **RGB** Adaptado de Thiago Cesar, 2008.

Mediante a representação geométrica do modelo de cor **RGB** na **Figura 2.9.1** pode-se fazer a seguinte interpretação, em função da combinação dos índices de cores. Assim, cada pixel é representado por três bytes de oito bits, que correspondem a um total de 24 bits, como cada byte pode ser expresso, por números inteiros entre 0 e 255 índices de cores e ao efetuar o cálculo de $256 \times 256 \times 256$ dos índices de cores, referentes aos canais **R, G e B**, com isso torna-se possível obter cerca de 17 milhões de combinações de cores(Thiago Cesar 2008; Lyra, Wellington 2009; Mariana Godinho 2014; Cristiano Bertolini 2010; Pedro Ivo, 2014).

Cada par de eixos, gera um plano em que são representadas as cores secundárias (contribuição equitativa de duas cores primárias) no vértice de cada plano. A interseção dos três planos, define a cor preta (não há contribuição de nenhuma das cores primárias) e a cor branca (mistura equitativa das três cores primárias com a intensidade máxima) e a diagonal principal deste cubo, representa a escala de cinza(Thiago Cesar 2008; Lyra, Wellington 2009; Mariana Godinho 2014; Cristiano Bertolini 2010; Pedro Ivo, 2014).

2.9.3 Modelo HSI

Este modelo de cor pode ser representado, por um cone e foi criado com base na combinação de cores desenvolvidas pelos artistas plásticos, sendo muito utilizado para definir as cores, nos programas gráficos de computadores de forma a aproximar, cada vez mais as cores geradas nos dispositivos, com aquelas percebidas pelo sistema visual humano.

O modelo **HSI** utiliza os seguintes parâmetros: **H, S e I** para a definição de uma cor que são a matiz(**H**), saturação(**S**) e intensidade(**I**) associados a tonalidade cromática, a saturação, relacionado com a pureza da cor e a intensidade, correspondente ao brilho, segundo a representação da **Figura 2.9.2** - Sólido de cor do modelo **HSI**, Adaptado de Cristiano Bertonili, 2010(Thiago Cesar 2008; Lyra, Wellington 2009; Mariana Godinho 2014; Cristiano Bertolini 2010; Pedro Ivo, 2014).

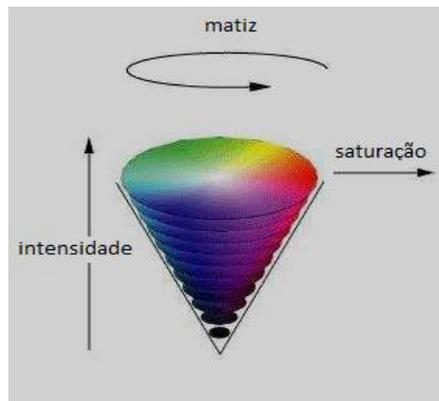


Figura 2.9.2 - Sólido de cor **HSI** Adaptado de Cristiano Bertonili, 2010.

De acordo com a figura acima, o matiz (**H**) é definido como o ângulo ao redor do eixo vertical, a saturação (**S**) e intensidade (**I**) ou brilho (**B**).

2.9.4 Modelo HLS

O modelo de cor **HLS** pode ser reconhecido geometricamente, por um cone duplo, no qual o modelo corresponde há uma transformação na qual cada pixel de uma imagem do modelo **RGB** pode ser representado por coordenadas cartesianas, pode ser transformado em um sistema de coordenadas cilíndricas (r, θ, h) ilustrado segundo a **Figura 2.9.3**.

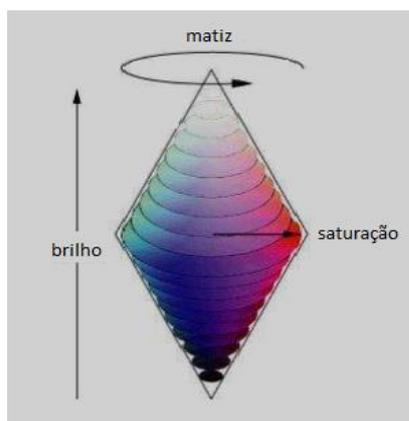


Figura 2.9.3- Sólido de cor **HLS** Adaptado de Cristiano Bertolini, 2010.

Na borda da órbita está o matiz (*hue, H*) que varia de 0 a 360° e no eixo central está a escala de cinza que indica a intensidade, brilho ou luminosidade (*lightness, L*) que varia de 0 a 1. Entretanto, a distância radial do eixo central, no qual o valor de cinza é gradativamente modificado até atingir uma cor pura é definida como saturação (*saturation, S*) e assim como a luminosidade varia de 0 a 1. De fato tanto o modelo de cor **HSI** como o **HLS**, são comumente utilizados na área da geologia especialmente em pedologia, para facilitar as interpretações de classificação das rochas ornamentais (Thiago Cesar 2008; Lyra, Wellington 2009; Mariana Godinho 2014; Cristiano Bertolini 2010; Pedro Ivo, 2014).

2.10 Espectroscopia de absorção molecular na região do ultravioleta e visível- UV-Vís

2.10.1 Introdução

A espectrofotometria de absorção molecular no **UV-Vís** utiliza radiação eletromagnética cujos comprimentos de onda (λ) encontram-se na faixa de 200 a 780 nm (Silverstein, R.M, 1994; Skoog D.A, Holler F.J, Nieman T.A, 2002). Os espectros de absorção **UV-Vís** apresentam geralmente bandas largas resultantes da sobreposição dos sinais provenientes de transições vibracionais e rotacionais ao sinal associado à transição eletrônica. Ademais, nos espectros obtidos com a amostra em fase condensada apresentam bandas lisas, carentes de detalhes e com baixa resolução (Silverstein, R.M, 1994; Skoog D.A, Holler F.J, Nieman T.A, 2002).

2.10.2. Origem do sinal analítico na região UV-Vís

Quando estimulada pela radiação **UV-Vís**, a molécula do composto pode sofrer transições eletrônicas por ocasião da absorção de energia quantizada produzindo um espectro eletrônico de absorção.

A absorção desse tipo de radiação promove uma mudança na estrutura eletrônica da molécula em consequência de transições envolvendo, geralmente, elétrons π de ligações e elétrons de valência n (não ligantes). Isto requer que a molécula contenha pelos menos um grupo funcional insaturado (C=C, C=O, por exemplo) para fornecer os orbitais moleculares π e π^* ou n e π^* (Silverstein, R.M, 1994; Skoog D.A, Holler F.J, Nieman T.A, 2002).

Este centro de absorção, chamado *cromóforo*, responde pelas transições $\pi \rightarrow \pi^*$ e $n \rightarrow \pi^*$, as quais são promovidas por radiações com comprimentos de onda geralmente acima de 200 nm. As bandas de absorção podem ser caracterizadas por dois parâmetros fundamentais: a posição e a intensidade. A posição corresponde normalmente ao “ λ ” da radiação eletromagnética responsável pela transição eletrônica, enquanto a intensidade depende, sobretudo, da energia dos orbitais moleculares e probabilidade de transição. Entretanto, podem ocorrer mudanças significativas na posição e intensidade das bandas em decorrência de interações intramoleculares (cromóforos C=C conjugados) e intermoleculares (resultantes de interações entre moléculas do composto e do solvente (Silverstein, R.M, 1994; Skoog D.A, Holler F.J, Nieman T.A, 2002).

2.10.3 Características dos espectros de Absorção UV-Vís

Os espectros de absorção **UV-Vís** apresentam geralmente bandas largas resultantes da sobreposição dos sinais provenientes de transições vibracionais e rotacionais ao sinal associado à transição eletrônica (Silverstein, R.M, 1994 ; Skoog D.A, Holler F.J, Nieman T.A, 2002).

Os espectros eletrônicos de absorção também podem apresentar problemas de forte sobreposição de bandas associadas a duas ou mais substâncias presentes em uma amostra por causa da natureza alargada das bandas e da modesta correlação entre o espectro e a estrutura molecular (Silverstein, R.M, 1994; Skoog D.A, Holler F.J, Nieman T.A, 2002). De fato, substâncias com estruturas moleculares muito diferentes, mas contendo o(s) mesmo(s) cromóforo(s), podem apresentar espectros **UV-Vís** com perfis similares e bandas localizadas nas mesmas regiões de comprimentos de onda (Silverstein, R.M, 1994; Skoog D.A, Holler F.J, Nieman T.A, 2002). Nesses casos, as pronunciadas sobreposições espectrais dificultam, por exemplo, a classificação ou discriminação de amostras, especialmente quando apresentam diferenças pequenas em suas composições químicas (Silverstein, R.M, 1994; Skoog D.A, Holler F.J, Nieman T.A, 2002).

2.11 Quimiometria

A Quimiometria abrange diversas técnicas estatísticas e matemáticas que possibilitam a resolução de problemas de planejamento e otimização experimental, reconhecimento de padrões e classificação e calibração multivariada (Beebe,K.R et al 1988; Massart,D.L,et al 1988; Pareira, T.F, 2003).

A exploração de técnicas quimiométricas multivariadas cresceu em função da grande quantidade de dados fornecidos por técnicas instrumentais (espectrofotometria UV-Vis, por exemplo) e da disseminação de microcomputadores em laboratórios químicos. As técnicas de análise multivariada podem se enquadrar em duas categorias principais: os métodos de análise exploratória ou reconhecimento de padrões e os de calibração multivariada. (Beebe,K.R et al 1988; Massart,D.L,et al 1988; Pareira, T.F, 2003).

2.11.1 Reconhecimento de Padrões

As técnicas de reconhecimento de padrões podem ser reunidas em duas categorias: (a) supervisionadas e (b) não-supervisionada. As que se enquadram no primeiro grupo são usadas apenas para examinar se existem similaridades e diferenças entre objetos (amostras químicas, por exemplo) de um conjunto de dados. Os métodos supervisionados, por outro lado, são usados quando o objetivo é construir um modelo para classificar amostras futuras.

Para um reconhecimento de padrão não-supervisionado, pode-se recorrer a dois métodos bastante conhecidos que são a *Análise de Componentes Principais (Principal Component Analysis-PCA)* e a *Análise Hierárquica de Agrupamentos (Hierarchical Cluster Analysis-HCA)*(Beebe,K.R et al 1988; Massart,D.L,et al 1988; Pareira, T.F , 2003).

2.11.2 Análise por Componentes Principais PCA

A PCA é uma técnica quimiométrica multivariada que tem por objetivo reduzir a dimensionalidade do espaço dos dados originais, facilitando a visualização de informações relevantes associadas ao conjunto dos dados. Uma descrição sucinta dos conceitos e princípios básicos dessa técnica é realizada a seguir (Beebe,K.R et al 1988; Massart,D.L,et al 1988; Pareira, T.F , 2003).

Considere uma matriz de dados $X_{n \times p}$, na qual “n” linhas correspondem aos objetos (amostras de café, por exemplo) e nas “p” colunas são dispostos os valores das variáveis (medidas espectrais) para cada amostra. A matriz X pode ser representada graficamente em um sistema cartesiano ortogonal cujos eixos são definidos pelas “p” variáveis; as amostras são então localizadas nesse sistema como “n” pontos cujas coordenadas são dadas pelos valores de uma propriedade medidas nas “p” ariáveis. Quando a matriz X contém no máximo 3 colunas (isto é, $p \leq 3$) é possível representá-la geometricamente, o que possibilita a

visualização de eventuais agrupamentos de amostras (Beebe, K.R et al 1988; Massart, D.L, et al 1988; Pareira, T.F , 2003). Naturalmente, isso ocorrerá caso as amostras possuam propriedade(s) similar(s) (por exemplo, composição química), o faz com que elas tenham presença de coordenadas similares. Entretanto, quando $p > 3$ torna-se muito difícil visualizar os dados no domínio geométrico, pois sua dimensionalidade extrapola o espaço tridimensional. Para contornar esse problema, pode-se empregar a técnica PCA (Beebe, K.R et al 1988; Massart, D.L, et al 1988; Pareira, T.F , 2003).

Ao aplicar uma PCA à matriz X , os dados são projetados nas direções (ou eixos) perpendiculares do espaço multidimensional contendo a maior quantidade de informação possível (variância máxima), como ilustrado na Figura 2.11.1. Na Equação 1.0, pode-se constatar que as componentes principais (PCs) são obtidas por intermédio de combinações lineares das variáveis originais. Portanto os valores de V e a correspondem, respectivamente, às variáveis originais e os coeficientes que medem a importâncias de cada variável na componente principal e n é o número de variáveis originais (Beebe, K.R et al 1988; Massart, D.L, et al 1988; Pareira, T.F , 2003).

Esses eixos são denominados componentes principais (Principal Components-PCs), nos quais as amostras assumem novos valores denominados *scores* (Figura 2.11.1)

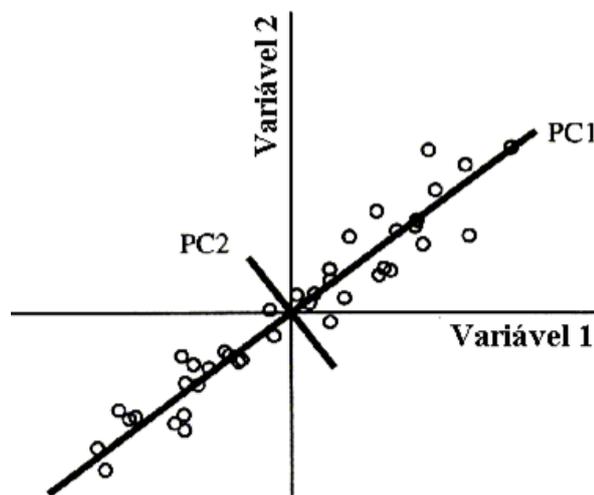


Figura 2.11.1 –Gráfico de PC1 versus PC2, onde os círculos representam as amostras - Adaptado de Beebe, K.R et al 1988.

$$PC_i = a_{i1}V_1 + a_{i2}V_2 + a_{i3}V_3 \dots \dots \dots + a_{in}V_n \quad (1.0)$$

A primeira componente principal (PC1) contém, entre todas as possíveis PCs, a maior variância dos dados originais. A segunda componente PC2, por sua vez, porta a segunda maior quantidade de informações associadas aos dados e assim sucessivamente. Este método

permite reduzir a dimensionalidade da matriz \mathbf{X} , ou seja, todas as amostras podem ser agora representadas em um novo sistema de menor dimensão com pouca perda de informação (Beebe, K.R et al 1988; Massart, D.L, et al 1988; Pareira, T.F, 2003).

Assim, as informações presentes nas variáveis originais são projetadas de modo que 90% ou mais da variância dos dados originais estejam contidas nas primeiras PCs, preferencialmente em PC1 e PC2. Com isto, torna-se possível:

- i. reduzir a dimensionalidade do sistema, ou seja, do espaço geométrico no qual as amostras foram representadas inicialmente;
- ii. reconhecer a formação de agrupamento(s) de amostras (reconhecimento de padrões);
- iii. detectar a presença de amostra(s) anômala(s);
- iv. identificar que variáveis contribuem mais para essas diferenças;
- v. mostrar de que forma uma amostra é diferente da outra, etc.

A importância das próprias variáveis originais na definição das características de uma amostra pode ser avaliada por meio dos pesos. Esses representam o valor do cosseno do ângulo entre cada PC e o eixo correspondente a uma dada variável. A visualização dessa importância pode ser realizada por meio dos gráficos dos pesos versus variáveis. Por outro lado, para cada PC modelada, uma certa quantidade de informação permanece sem ser explicada. Essa parcela é chamada de resíduos, a qual pode ser visualizada nos gráficos de resíduos. (Beebe, K.R et al 1988; Massart, D.L, et al 1988; Pareira, T.F, 2003).

2.11.3 Fundamentação teórica do SIMCA

O SIMCA possibilita a construção de um modelo para cada classe de amostras do conjunto de treinamento. Para isso, aplica-se uma PCA para modelar a forma e a posição do objeto (ou estrutura) formado pelas amostras no espaço multidimensional para definição de uma classe (Beebe, K.R et al 1988; Massart, D.L, et al 1988; Pareira, T.F, 2003).

Como resultado, uma caixa (embalagem) é construída para cada classe e a classificação de novas amostras (previsão) é realizada verificando em qual das embalagens a amostra se enquadra. Diferentes formas para as embalagens podem ser obtidas em uma modelagem SIMCA dependendo do número de PCs necessário para descrevê-las (Beebe, K.R et al 1988; Massart, D.L, et al 1988; Pareira, T.F, 2003).

Assim, uma linha, um plano e um paralelepípedo representam modelos com uma, duas e três componentes principais, respectivamente, como se pode observar na **Figura 2.11.2**

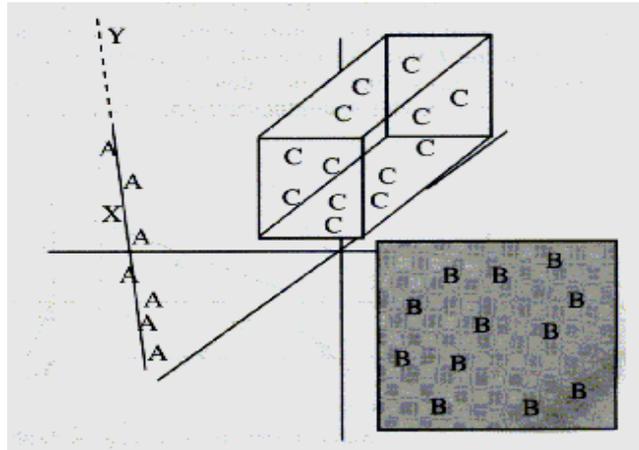


Figura 2.11.2 – Ilustração gráfica de modelos SIMCA para três classes distintas A, B, e C. As amostras, da classe A ocupam uma linha, da classe B um plano e a classe C um cubo. X e Y são amostras desconhecidas- Adaptado de Beebe ,K.R et al 1988

A **Figura 2.11.2** ilustra como é realizada a classificação de uma amostra desconhecida (Y) usando o modelo **SIMCA** baseado em **Equações 2.0, 2.1 e 2.2**.

$$S_i^c = \sqrt{\frac{\sum_{j=1}^p (res_j^c)^2}{p - A_c}} \quad (2.0)$$

$$S_o^c = \sqrt{\frac{\sum_{i=1}^{N_c} \sum_{j=1}^p (res_{ij}^c)^2}{(N_c - A_c - 1) \cdot (p - A_c)}} \quad (2.1)$$

$$F_{cal} = \frac{(S_i^c)^2}{(S_o^c)^2} \cdot \frac{N_c}{N_c - A_c - 1} \quad (2.2)$$

onde:

- N_c = número de amostras do conjunto de treinamento da classe c;
- A_c = número de PCs utilizada pela classe c;
- p = número de variáveis;
- i e j = índices das amostras e variáveis, respectivamente.

N_c é o número de amostras pertencentes ao conjunto de treinamento da classe c ; A_c é o número de componentes principais utilizadas pela classe c ; p representa o número de variáveis, i e j representam os índices das amostras e variáveis, respectivamente. Um teste F é, então, utilizado para verificar a localização da amostra em relação ao(s) modelo(s). Compara-se o valor obtido pela Equação 2.2 (F_{cal}) com um valor crítico (F_{crit}) que pode ser obtido empiricamente ou tabelado para um determinado nível de confiança e graus de liberdade. Se a amostra Y apresentar um valor de F_{cal} menor do que o obtido pelo F_{crit} , a mesma pertencerá à classe em consideração ilustrada na Figura 2.11.2 onde:

$N_c = N^\circ$ de amostras do conj. de treinamento da classe c ;

$A_c =$ número de PCs utilizada pela classe c ;

$p =$ número de variáveis;

i e $j =$ índices das amostras e variáveis, respectivamente.

Para a realização do cálculo da distância da amostra Y ao modelo SIMCA, utilizam-se a variância residual para cada amostra da classe c , S_i (Equação 2.0), e a variância residual total, S_o da (Equação 2.1). Desse modo um teste F é, então, utilizado para verificar a localização da amostra em relação ao(s) modelo(s). Compara-se o valor obtido pela Equação 2.2 (F_{cal}) com um valor crítico (F_{crit}) que pode ser obtido empiricamente ou tabelado para um determinado nível de confiança e graus de liberdade. Se a amostra Y apresentar um valor de F_{cal} menor do que o obtido pelo F_{crit} , a mesma pertencerá à classe em consideração. Para isso, o parâmetro “ a ” é convertido em uma variância e é dividido pela variância residual a fim de determinar um valor para F_{calc} . Um valor para F_{crit} (em geral com 95% de confiança) é escolhido e comparado ao valor de F_{calc} . A amostra Y seria classificada como pertencente à classe das amostras A se o $F_{calc} < F_{crit}$ conforme a ilustração da Figura 2.11.3

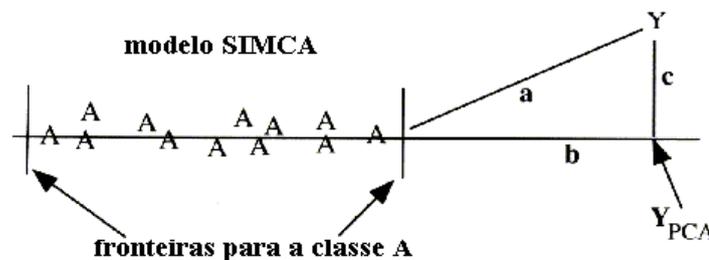


Figura 2.11.3 - Previsão da amostra Y com um modelo SIMCA para a classe de amostras A . “ c ” é o resíduo de PCA; “ b ” é a distância entre a fronteira e a projeção de Y na PC e “ a ” é a proximidade de Y da caixa A , calculada por: $a^2 = b^2 + c^2$. Adaptado de Beebe, K.R et al 1988

2.11.4 Fundamentação teórica da Análise Discriminate Linear- LDA

A análise discriminante foi originalmente desenvolvida na Botânica, e sua aplicação teve como objetivo fazer a distinção de grupos de plantas com base no tamanho e no tipo de folhas, para que, posteriormente, fosse possível classificar as novas espécies encontradas. Mas a aplicação da análise discriminante cedo se generalizou a outras ciências, sempre em situações onde é possível encontrar grupos de indivíduos e conhecer quais as características que os distinguem uns dos outros (Hair, Jr. J.F,2005; Mingoti,A.S, 2013).

A análise discriminante é empregada para descobrir as características que distinguem os membros de um grupo dos de outro, de modo que ao conhecer as características de um novo indivíduo seja possível prever o grupo ao qual ele pertence. Assim, a análise discriminante tem por objetivo estimar uma combinação linear de duas ou mais variáveis independentes que possa, da melhor maneira possível, separar ou discriminar dois ou mais grupos de observações ou casos previamente definidos (Hair, Jr. J.F,2005; Mingoti,A.S, 2013).

A discriminação é realizada determinando-se o conjunto ótimo de pesos para as variáveis independentes de tal maneira que se maximize a variância entre os grupos, relativamente à variância dentro dos grupos (Hair, Jr. J.F,2005; Mingoti,A.S, 2013).

Esta combinação linear de variáveis é chamada de função discriminante, em que cada combinação linear (Z_{jk}) constitui uma função discriminante da seguinte forma que pode ser ilustrada segundo a **Equação 2.3**

$$Z_{jk} = a + W_1 X_{1k} + W_2 X_{2k} + \dots + W_n X_{nk} \quad (2.3)$$

Onde:

Z_{jk} :o escore Z discriminante da função discriminante j para o objeto k

X_{ik} : variável independente i para o objeto k

W_i : peso discriminante para a variável independente i

a: intercepto.

As funções discriminantes são determinadas de modo a maximizar a separação entre diferentes grupos. Uma vez estimadas as funções discriminantes, é possível concretizar os dois objetivos da análise discriminante: classificação dos casos ou observações em grupos distintos e alocação de novos casos ou observações a grupos previamente definidos.

Os pesos discriminantes (W_i) podem ser interpretados como os coeficientes de um modelo de regressão múltipla que servem para identificar as variáveis que mais contribuem

para distinguir os grupos de uma mesma função discriminante Z_{jk} (Hair, Jr. J.F,2005; Mingoti,A.S, 2013).

A análise discriminante é realizada por intermédio de uma ou mais combinações lineares das variáveis discriminantes utilizadas. Deste modo, quando a análise discriminante é usada para separar os grupos de casos ou observações, será preciso apenas uma única função discriminante, no caso de três grupos tem-se duas funções discriminantes e assim sucessivamente (Hair, Jr. J.F,2005; Mingoti,A.S, 2013).

A aplicação da análise discriminante pode ser vista a partir da perspectiva da construção de um modelo em seis estágios: Estágio 1- Problema de Pesquisa; Estágio 2- Desenho do Experimento; Estágio 3- Suposições; Estágio 4- Estimação da Função Discriminante e Avaliação do ajuste da Função Discriminante; Estágio 5- Interpretação dos Resultados; Estágio 6- Validação dos resultados.

2.11.5 Algoritmo das Projeções Sucessivas-SPA para classificação usando Análise Discriminante Linear-LDA

O algoritmo de SPA-LDA tem por objetivo selecionar um subconjunto de variáveis com pequena colinearidade e poder discriminatório adequado para uso em problemas de classificação quando estão envolvidas duas ou mais classes diferentes por exemplo, $C \geq 2$.

Para isso, deve-se assumir o propósito que há um conjunto de treinamento de N objetos com rótulos de classes conhecidas encontra-se disponível na orientação do processo de seleção de variáveis (Pontes, M J C , 2009 ;Soares, S.C.F et al ,2014).

Considerando os dados espectroscópicos ou os dados das imagens digitais, para cada objeto (amostra), por exemplo, consiste de um espectro registrado em K números de onda (comprimentos de onda) e uma distribuição de frequências de histogramas nos quais os pixels são representados pelos seus espaços de cor , por exemplo, RGB, HSL,HSI e outros.

Quando há seleção de variáveis o algoritmo possui a seguinte rotina, que pode ser dividida em duas fases principais. Na primeira, fase 1 os N objetos de treinamento estão centrados na média de suas respectivas classes e dispostas sob a forma de uma matriz \mathbf{X} ($N \times K$). Cada coluna de \mathbf{X} está associada a uma variável (isto é, um número de onda no caso de dados espectroscópicos). Operações de projeção que envolvem as colunas de \mathbf{X} são então realizadas para formar K cadeias de L variáveis, em que $L = \min (K, N - C)$. Cada cadeia é iniciada com uma das variáveis K disponíveis (Pontes, M J C , 2009 ;Soares, S.C.F et al, 2014). . As variáveis subsequentes da cadeia são selecionadas de modo a apresentar a menor

colinearidade (avaliada pelas operações de projeção) com as variáveis precedente (Pontes, M J C et al, 2009 ;Soares, S.C.F et al, 2014).

Na fase 2, diferentes subconjuntos de variáveis são extraídos das cadeias e em seguida, avaliados em termos de uma função de custo J_{Cost} , utilizada quando há um número reduzido de amostras nos conjuntos de treinamento e teste (previsão), que será definida pelas Equações 2.4 e 2.5

$$J_{Cost} = \frac{1}{N_{Train} - L - C} \sum_{n=1}^{N_{Train}} g_n \quad (2.4)$$

Onde :

J_{Cost} : Função de Custo

N_{Train} : Número de amostras do conjunto de Treinamento.

L : Número de variáveis da cadeia

C : Número de Classes

g_n : risco de classificação incorreta

Sendo g_n , corresponde ao risco de uma classificação incorreta, do objeto \mathbf{x}_k da k -ésima amostra de validação e definido segundo a equação 2.5

$$g_n = \frac{r^2[\mathbf{x}_n, \bar{\mathbf{x}}(I_n)]}{\min_{I_j \neq I_n} r^2[\mathbf{x}_n, \bar{\mathbf{x}}(I_j)]} \quad (2.5)$$

Na equação. (2.5) o numerador $r^2[\mathbf{x}_n, \bar{\mathbf{x}}(I_n)]$ é a distância de Mahalanobis ao quadrado entre o objeto \mathbf{x}_n (de índice de classe I_n) e a média da amostra $\bar{\mathbf{x}}(I_n)$ de sua verdadeira classe (ambos são vetores de linha). Esta distância é calculada pela a seguir Equação 2.6 :

$$r^2[\mathbf{x}_n, \bar{\mathbf{x}}(I_n)] = [\mathbf{x}_n - \bar{\mathbf{x}}(I_n)]\mathbf{S}^{-1}[\mathbf{x}_n - \bar{\mathbf{x}}(I_n)]^T \quad (2.6)$$

onde \mathbf{S} é uma matriz de covariância conjunta, a qual é calculada de acordo com o procedimento LDA padrão . O denominador na Equação. (2.5) corresponde à distância de

Mahalanobis ao quadrado entre o objeto \mathbf{x}_n e o centro da classe errada mais próxima. Desse modo, um valor pequeno de \mathbf{g}_n indica que \mathbf{x}_n está perto do centro da sua classe verdadeira e distante dos centros das classes restantes. Portanto, minimizando os resultados do custo G há uma melhor separação dos objetos de acordo com suas verdadeiras classes (Pontes, M J C 2009 ;Soares, S.C.F et al, 2014).

2.11.6 Fundamentação teórica da Análise Discriminante por Método Mínimo Quadrado Parciais - PLS-DA

A fundamentação teórica do **PLS-DA** consiste em encontrar variáveis latentes no espaço multivariado da matrix \mathbf{X} que corresponde as variáveis independentes, neste caso associadas com as medidas espectroscópicas na região do ultravioleta e visível ou as imagens digitais e com isso correlacionar com as variáveis dependentes da matrix \mathbf{Y} que representam os índices das classes, através da utilização do método dos mínimos quadrados parciais (**PLS**). Dessa forma é estabelecido um limite de confiança para cada classe, a partir dos valores preditos de \mathbf{Y} do modelo **PLS-DA** e posteriormente, calcula-se a probabilidade de um dado valor de \mathbf{Y} pertence a classe previamente definida (Juliano Ribeiro 2009;Jaqueline Martins 2013).

Portanto os valores preditos, serão o número um ou zero, sendo que o número um indica que a amostra predita pertence a classe e o zero não pertence. Dessa forma, será possível discriminar e classificar novas amostras apesar da possibilidade de ocorrer os erros de classificação que correspondem aos falsos positivos ou falsos negativos (Juliano Ribeiro 2009;Jaqueline Martins 2013). A seguir são apresentadas, as condições detalhadas do modelo **PLS-DA** para fins de classificação.

Conidere a matriz \mathbf{X} de dimensão $\mathbf{N} \times \mathbf{K}$ e um vetor \mathbf{Y} de dimensão $\mathbf{N} \times \mathbf{1}$, em que \mathbf{K} representa os comprimentos de onda relacionados aos extratos aquosos dos cafés adulterados e não adulterados e \mathbf{N} corresponde, ao número de amostras das classes analisadas através de seus espectros dos extratos aquosos (A.Ferrer et al 2008;M Anzanelo et al, 2012).

As amostras podem ser representadas pelo vetor $\mathbf{x}_i(\mathbf{x}_{i1}, \mathbf{x}_{i2}, \mathbf{x}_{i3} \dots \dots \mathbf{x}_{ik})$ que corresponde a i -ésima amostra para cada comprimento de onda \mathbf{K} do **UV-Vis** e o escalar \mathbf{y}_i é uma variável categórica que representa a classe da i -ésima amostra. Considerando as duas classes de cafés e seus índices que assumem os valores de 0 para a classe adulterada e 1 para a classe dos não adulterados(A.Ferrer et al 2008;M Anzanelo et al, 2012).

Desse modo o **PLS-DA**, gera A , que corresponde as combinações lineares independentes $t_a (a = 1, 2, 3 \dots A)$ a partir dos comprimentos de onda originais e seus respectivos coeficientes $w_a = (w_{1a}, w_{2a} \dots \dots \dots w_{ka})$ " segundo a **Equação 2.7** :

$$t_i a = w_1 a x_{i1} + w_2 a x_{i2} + w_3 a x_{i3} \qquad w_{ka} x_{ik} = w' a x_i \quad (2.7)$$

Sendo que w_{ka} corresponde ao peso do comprimento de onda de processo k na componente a e cada componente responde por uma fração de variância explicada em Y , denotado por R_{ya}^2 explicada pela componente $a = 1, \dots \dots \dots A$ (**A.Ferrer et al 2008; M Anzanelo et al, 2012**).

O **PLS-DA** proporciona também a obtenção de variáveis latentes $u_a = (a, 1, 2, \dots \dots \dots A)$ a partir dos comprimentos de onda das variáveis categoricas, segundo a **Equação 2.8**, em que $c_1 a$ representa o peso de y na componente a :

$$u_{ia} = c_{1a} y_{1i} \quad (2.8)$$

Dessa forma o número de variáveis latentes é reduzido e são ortogonais umas as outras, tal fato reduz os problemas de colinearidade relacionados aos dados originais, associados as variáveis espectrais dos extratos aquosos .

Segundo **A.Ferrer et al 2008 e M Anzanelo et al, 2012**, os vetores w_a e c_a são calculados para maximizar a covariância entre componentes t_a e u_a e também afirmam que w_a e c_a fornecer informações sobre os comprimentos de onda que se combinam para fornecer a relação quantitativa entre X e Y , permitindo a identificação dos comprimentos de onda de x mais relevantes ou seja os com maiores w_{ka} . Portanto os coeficientes de regressão pelo **PLS-DA** podem ser obtidos a partir da **Equação 2.9**.

$$b_k = \sum_z c_a w_{ka} \quad (2.9)$$

Em que quanto maior o valor absoluto de b_k , mais relevante será o comprimento de onda k na previsão do variável categórica y (**A.Ferrer et al 2008, M Anzanelo et al, 2012**).

Portanto o **PLS-DA** é um método de análise qualitativa para fins de classificação que tem por finalidade discriminar e classificar as amostras em suas classes previamente definidas (**Juliano Ribeiro, 2009 ,Jaqueline Martins 2013**).

2.11.7 Figuras de mérito para modelos de classificação dos cafés adulterados e não adulterados

Dentre as características principais para verificar a performance dos modelos de classificação em função do número de acertos, destacam-se a sensibilidade, especificidade, acurácia e seus complementos a taxa de falsos positivos (especificidade) e taxa de falsos negativos (sensibilidade) (Tavares 2012; Liska 2012). De fato, as figuras de mérito referidas são ferramentas de diagnóstico para verificar o ajuste do modelo nas etapas de validação e predição. Nesse sentido, um modelo pode ser considerado ajustado quando apresenta um percentual de classificação elevado na etapa de validação e predição referente a sensibilidade, especificidade e acurácia discutidas a seguir.

Sensibilidade

Sensibilidade pode ser definida como a probabilidade do modelo em avaliar a proporção de verdadeiros positivos preditos, ou seja, no caso da adulteração, o modelo pode determinar quais são os cafés, que estão adulterados por cascas e paus através da **Equação 2.10**.

$$\text{Sensibilidade} = VP / (VP + FN) \quad (2.10)$$

onde:

- VP verdadeiros positivos
- FN falsos negativos

Especificidade

Especificidade pode ser conceituada como a probabilidade do modelo em avaliar a proporção de verdadeiros negativos preditos, em especial no caso da adulteração o modelo determina quais são os cafés, que não estão adulterados por cascas e paus mediante a utilização da **Equação 2.11**.

$$\text{Especificidade} = VN / (VN + FP) \quad (2.11)$$

- VN verdadeiros negativos
- FP falsos positivos

Acurácia

Acurácia pode ser conceituada como a probabilidade do modelo em avaliar a proporção de verdadeiros positivos e verdadeiros negativos na predição ou seja determinar quais são os cafés adulterados e não adulterados, podendo ser calculado segundo a **Equação 2.12**

$$\text{Acurácia} = \frac{VP + VN}{VP + FN + VN + FP} \quad (2.12)$$

Onde:

- *VP* verdadeiros positivos
- *VN* verdadeiros negativos
- *FP* falsos positivos
- *FN* falsos negativos

Segundo a literatura especializada, os falsos positivos *FP* e os falsos negativos *FN* correspondem ao complemento da especificidade e sensibilidade, que no caso dos cafés podem ser associados respectivamente a proporção incorreta das predições positivas do modelo em relação ao total dos negativos (**erro do tipo I**) e a proporção incorreta das predições negativas do modelo em relação ao total de positivos (**erro do tipo II**) (Tavares 2012;Liska 2012)

2.12 Revisão Bibliográfica

Mediante a revisão bibliográfica da literatura especializada, um pequeno número de trabalhos foram encontrados, sobre a discriminação e classificação de cafés, com a utilização da espectroscopia de absorção molecular na região do Ultravioleta-Vísivel (**UV-Vís**) e Imagens Digitais com a utilização da Análise Multivariada. Entretanto, não foram encontrados trabalhos na literatura especializada que contemple a discriminação e classificação de cafés torrados e moídos e seus extratos aquosos, entre os cafés adulterados dos não adulterados com a utilização das técnicas referidas.

Desse modo, há uma notável motivação para o desenvolvimento de novas metodologias com a finalidade de discriminar e classificar cafés adulterados por cascas e paus dos não adulterados. A seguir nas seções **2.12.1** e **2.12.2** serão apresentados sucintamente alguns trabalhos na literatura, sobre a discriminação e classificação de cafés.

2.12.1 Discriminação e classificação de cafés usando espectroscopia de absorção molecular do Ultravioleta-Vísivel (UV-Vís)

No trabalho desenvolvido por (**Souto et al 2010**) foi realizada a classificação de cafés cafeinados, descafeinados e avaliação do estado de conservação através do envelhecimento. Com isso, foi proposta uma metodologia alternativa para avaliar a qualidade dos cafés comercializados. Para o seu desenvolvimento utilizou-se a espectroscopia de absorção molecular **UV-Vís** com modelagem quimiométrica, com a utilização das técnicas de reconhecimento de padrão supervisionada (**SIMCA, SPA-LDA**)

A partir de resultados reportados por (**Souto et al 2010**)conclui-se que a metodologia utilizada é satisfatória no que se refere a classificação dos cafés, sendo destacado o algoritmo das projeções sucessivas por Analise Discriminante Linear (**SPA-LDA**) que apresentou um satisfatório resultado de discriminação e classificação dos cafés cafeinados, descafeinados e envelhecidos de não envelhecidos. Dessa forma a metodologia desenvolvida satisfaz a discriminação e classificação de cafés quando são submetidos ao processo de descafeinação e pode ser útil também para avaliar o estado de conservação dos cafés comercializados.

2.12.2 Discriminação e Classificação de cafés com o uso de Imagens Digitais

No trabalho desenvolvido por **(Sano, Assad, Cunha, Correa & Rodrigues, 2003)** foi proposta uma metodologia para a classificação e quantificação de principais adulterantes, impurezas e misturas no café torrado. Para isso, foi realizada a aquisição das imagens multiespectrais das amostras de café, adulterantes e impurezas, por meio de uma lupa acoplada a uma câmara CCD (dispositivo de câmara acoplada) após as etapas de limpeza, secagem e homogeneização

A captação e transferência das imagens para um analisador de imagens, foi realizada por uma medida de reflectância, com um acoplamento de uma lupa a uma câmara CCD. Desse modo, houve a aquisição das imagens digitais no espaço de cor **RGB** e seus respectivos canais; o azul (**B**), verde (**G**) e vermelho (**R**) do espectro visível, utilizado para fins de classificação. Portanto, a classificação é realizada a partir da identificação da quantidade de pixels que apresentam a mesma resposta espectral. Com isso, os pixels que correspondem aos componentes de uma mesma classe, ou seja, a classe de cascas e paus, classe de milho, classe de café puro (pó), puderam ser classificados através da identificação das classes por contagem de pixels **(Sano, Assad, Cunha, Correa & Rodrigues, 2003)**.

Assim, foi possível reconhecer a área de cada classe na imagem e a área de cada componente, as quais foram utilizadas na construção das curvas de calibração, que correlaciona a porcentagem da área na imagem com a porcentagem em peso da amostra fraudada artificialmente, para cada tipo de impureza e mistura.

A partir dos resultados obtidos por **Sano, Assad, Cunha, Correa, & Rodrigues, 2003**, foi possível concluir que a metodologia utilizada, apresentou resultados satisfatórios na classificação e em sua quantificação de adulterantes em cafés torrados, bem como permite agilidade da resposta, ausência de subjetividade nos resultados, não-destruição das amostras e assegura um patamar de no mínimo de detecção de 95% das impurezas do produto

No trabalho apresentado por **(Aycheh et al em 2008)** foi desenvolvida uma metodologia para a classificação de grãos de cafés arábica e robusta, a partir dos aspectos morfológicos e de cor das diversas variedades cultivadas na Etiópia. Para o desenvolvimento da metodologia, foram utilizadas as dez características morfológicas, sendo que seis foram extraídas de cada imagem dos grãos de cafés, com a utilização dos espaços de cor RGB e HS e realizou-se a classificação por Bayes e Redes Neurais, as quais foram comparadas posteriormente, através dos parâmetros de morfologia, cor e a sua combinação.

A partir dos resultados de **(Aycheh em 2008)** a classificação por Redes Neurais foi consideravelmente mais satisfatória em relação as características da morfologia, quando comparado as características de cor. Entretanto, ao se utilizar as características da morfologia e cor em conjunto, os resultados sofreram uma melhora mais significativa para a classificação. Portanto, conclui-se que os melhores resultados de classificação, foram por meio das Redes Neurais para as características de morfologia e cor.

No trabalho desenvolvido por **(Pedro Ivo Oyama em 2014)** foi proposta uma metodologia para a discriminação e classificação de cafés em grãos por imagens digitais, pelo uso de atributos selecionados de morfologia, textura e cor. A metodologia desenvolvida utilizou as Redes Neurais Artificiais e uma visão computacional para identificação das vinte e uma classes de grãos de cafés crus em amostras. Desse modo, foram utilizados cerca de quatrocentos e vinte e um atributos de três diferentes naturezas; morfologia (211), textura (126) e cor (84) associadas as classes de grãos defeituosos; ardido, barrento, branco, brocado, concha, marinheiro, choco dentre outras; grãos perfeitos e impurezas (casca, casca de marinheiro, coco, paus e pedra).

O sistema utilizou algoritmos de processamento de imagens, para extrair os atributos de morfologia, textura e cor. Mediante a aquisição dos atributos, houve a divisão do conjunto em cinco subconjuntos; morfologia, cor, textura, cor-textura e morfologia-cor-textura cada um contendo diferentes combinações das distintas naturezas e sendo associados a uma análise **(Pedro Ivo Oyama em 2014)**.

Em função disso, houve a seleção dos melhores atributos que foram associados a cor e textura nos processos classificatórios, pelas técnicas chi-quadrado, Análise de Componentes Principais (PCA) e ganho de informação utilizados, para a discriminação entre as classes.

A partir dos resultados da seleção, determinou-se as entradas para os três processos classificatórios que são relacionados a morfologia, cor e textura, morfologia mais cor e mais textura, morfologia mais cor e textura, a fim de averiguar o mais efetivo que apresentou 85% de acurácia para a validação cruzada e com recall das classes variando de 62,8 a 95,4% para a morfologia, cor e textura. Contudo, a metodologia desenvolvida apresentou alguma dificuldade em identificar as nuances para a característica morfológica da forma que separam as classes **(Pedro Ivo Oyama em 2014)**.

Capítulo 3

Experimental

3.0 Experimental

3.1. Aquisição das amostras de café torrado e moído

As amostras de café utilizadas nos estudos realizados neste trabalho, foram fornecidas pelos laboratórios de pesquisa pelo Núcleo de Apóio Global (**NUGAP**) em Minas Gerais e pelo Serviço Nacional de Aprendizagem Industrial (**SENAI**) de São Paulo, credenciados pela Associação Brasileira da Indústria e do café (**ABIC**).

Além disso, as amostras são de procedência de diversos estabelecimentos comerciais e de indústrias de torrefação dos diversos estados brasileiros e seus teores cascas e paus foram determinados pelos laboratórios referidos, ressalta-se que as amostras adquiriram o laudo técnico da certificação e não foram adulteradas em laboratório.

Destaca-se que, os cafés enviados pelos laboratórios apresentam também uma considerável variabilidade em função dos seus diferentes blends, graus e ponto de torra, moagens (fina, média e grossa) devido a diferentes processos industriais de torrefação.

No total foram 103 amostras adquiridas distribuídas em duas classes, com 46 amostras não-adulteradas e 57 amostras adulteradas com cascas e paus .

As amostras adulteradas por cascas e paus apresentaram os seguintes percentuais médio e desvio padrão de adulteração na faixa de 1,08-8,49 g/100 g e desvio padrão de $4,05 \pm 1,66$ g/100g respectivamente. De fato, a amostra será considerada adulterada quando apresentar o percentual de valor máximo limite de 1,0% massa/massa de cascas e paus.

3.2 Obtenção das imagens digitais

3.2.1 Aparato instrumental

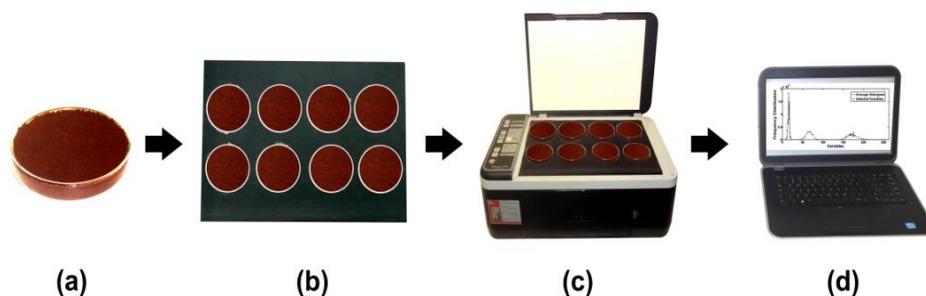


Figura 3.0- Aparato Instrumental utilizado na obtenção das imagens digitais dos cafés adulterados e não adulterados

- Disco de Petri com a amostra de café não adulterado ou adulterado;
- Disco de petri e suporte com oito orifícios para colocar as amostras de cafés;
- Scanner
- Notebook

3.2.2 Procedimento analítico para as imagens digitais

- Inicialmente foram pesados em uma balança analítica 1,0000 g de cafés não adulterados e cafés adulterados respectivamente e após suas massas foram transferidas para os discos de petri;
- Discos de petri de 5,5 cm de diâmetro por 1,1 de altura e suporte circular de 21 cm x 29,7cm x 0,5 cm de comprimento, largura e altura com oito orifícios circulares no qual foram colocadas as amostras de cafés não adulterados e adulterados, se faz necessário realçar que tanto as pesagens e aquisição das imagens digitais foram realizadas em triplicatas. Além disso, foram testadas e avaliadas as mudanças de posições dos discos de petri no scanner, mas nada afetou a qualidade das imagens digitais dos cafés adulterados e não adulterados.
- A aquisição das imagens digitais foi realizada pelo scanner e seu processamento através do programa Delfi 7. Desse modo, ocorreu a seleção das regiões centrais dos discos de petri, com uma área de 253x254 pixels e resolução de 300 dpi que resultou em histogramas dos cafés, segundo as [Figura 3.1](#) e [Figura 3.2](#) respectivamente



Figura 3.1- Seleção das imagens digitais de cafés adulterados (cascas e paus) pelo programa Delfi 7 desenvolvido no LAQA

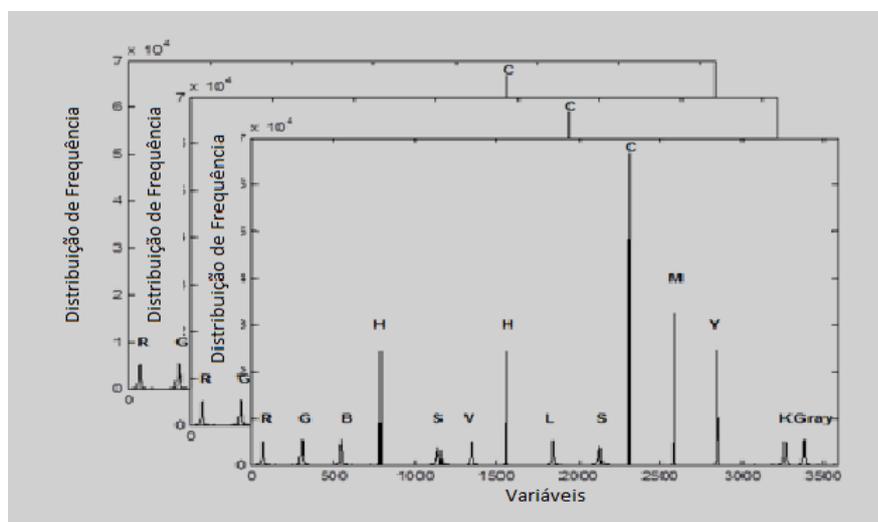


Figura 3.2- Histogramas das imagens digitais de cafés adulterados (cascas e paus)

- d) Na etapa final os dados foram transferidos para o Notebook e realizado o tratamento quimiométrico dos histogramas de distribuição de frequências, que ocasionou a construção dos modelos quimiométricos para fins de discriminação e classificação dos café.

3.2.3 Modelagem quimiométricas para as imagens digitais

A modelagem quimiométrica foi realizada com a seleção das amostras dos conjuntos de treinamento e teste pelo algoritmo Kennard Stone (KS), utilizado para extrair um subconjunto representativo de objetos ou seja as amostras de um determinado conjunto de dados (KENNARD, R.W.; STONE, L.A, 1969).

Neste caso, foram selecionados um total de 63 amostras, sendo 37 amostras de cafés adulteradas e 26 para amostras de café não adulteradas , que corresponde a 65% do conjunto de treinamento e 40 amostras para o conjunto teste, com as 20 amostras de cafés adulteradas e 20 de cafés não adulterados relativo em um total de 35%.

3.2.4 Modelagem quimiométrica para a espectroscopia de absorção molecular na região do ultravioleta (UV)

A modelagem quimiométrica foi realizada com a seleção das amostras dos conjuntos de treinamento e teste pelo KS com um total de 63 amostras. Para o conjunto de treinamento foram selecionadas um total de 37 amostras de cafés adulteradas e 26 para amostras de café não adulteradas com um percentual de 65% e 35% para o conjunto teste, no qual as 39 amostras dos cafés selecionadas, foram divididas em 20 amostras de cafés adulteradas e 19 amostras de café não adulterados .(KENNARD, R.W.; STONE, L.A, 1969).

Cabe ressaltar que, foram analisados outros percentuais de distribuição das amostras para os conjuntos de treinamento de 60% a 65% e conjunto teste de 35% a 40% , mas não houve mudanças nos percentuais de acertos das classes dos cafés, permanecendo a mesma para os modelos construídos pelo SIMCA, PLS-DA e SPA-LDA.

3.2.5 Algoritmos e Softwares utilizados na construção dos modelos de classificação para as imagens digitais e a espectroscopia de absorção molecular na região do ultravioleta (UV)

Para a seleção das variáveis foram utilizado o algoritmo das progressões sucessivas para a análise discriminante linear (SPA-LDA) com a utilização do software Matlab® versão 2009 e para os modelos de classificação, construídos com o uso do SIMCA e PLS-DA, o toolbox de classificação por Milano Quimiometria e QSAR grupo de pesquisa e a realização

da Análise de Componentes Principais (**PCA**) pelo Unscrambler versão 9.7 da **CAMOS**, as modelagens foram construídas, com a atribuição do nível de significância de 5%.

No capítulo 4, são apresentados os limiares para as classes dos cafés adulterados e não adulterados pelo **PLS-DA**

3.3 Obtenção dos espectros de absorção UV-Vís

3.3.1. Aparato instrumental

Um espectrofotômetro UV-Visível com arranjo de fotodiodos Hewlett Packard, modelo HP 8453, resolução de 1nm, foi utilizado para obtenção dos espectros na faixa de 200 a 800 nm conforme a ilustração da **Figura 3.3**. Este equipamento encontra-se disponível no Laboratório de Instrumentação e Automação em Química Analítica/Quimiometria (LAQA)-DQ-UFPB.



Figura 3.3 - Fotografia do espectrofotômetro (modelo HP 8453) utilizado para obtenção dos espectros das 102 amostras de cafés não adulterados(NA) e adulterados(A)

3.3.2 Procedimento analítico para a obtenção dos espectros dos extratos aquosos das amostras do cafés adulterados e não adulterados

- (1) Pesou-se 1,0000 g de cada amostra de café não adulterado e adulterado e realizou-se a extração em uma etapa, com a adição de 50 mL de água destilada por uma proveta graduada de 50 mL em quatro etapas sucessivas, com um total de 200 mL e mantendo-se a temperatura na faixa de 90–98°C com a finalidade de assegurar, uma melhor extração dos parâmetros de qualidade dos extratos aquosos dos cafés (**Vitorino, M. D. et al, 2001, Souto et al, 2010**).

(2) . A faixa escolhida para as extrações correspondeu a 90–98°C, porque assegura a condição de extração dos constituintes majoritários dos extratos aquosos: cafeína, trigonelina, ácidos clorogênicos e ácido caféico (Vitorino, M. D. et al, 2001, Souto et al, 2010).

(3) O volume do extrato aquoso obtido para cada amostra foi de aproximadamente 200 ml, sendo recolhido em erlemmeyer de 250 mL segundo a Figura 3.4 ;



Figura 3.4 - Extrações dos extratos aquosos das amostras de cafés

(4) Após o resfriamento, alíquotas de 2,5 mL do extrato aquoso foram transferidos para balões volumétricos de 50 ml usando uma pipeta volumétrica e realizada a diluição com água destilada na proporção de 1:20 (v/v) para todos os extratos e por fim o registro dos espectros com a utilização de uma cubeta de quartzo de caminho óptico de 10 mm.

Capítulo 4

Resultados e

Discussão

4.0 Resultados e Discussão

4.1 Análise baseada em imagens digitais e quimiometria

A inspeção visual das amostras de café torrado e moído não permite constatar que houve adulteração, promovida pela adição de cascas e paus, como pode ser observado nas imagens apresentadas na **Figura 3.1** e corroboradas com a **Figura 4.0.**, tal similaridade pode ser apresentada na **Figura 4.0** das distribuição de frequências, na qual evidência uma considerável sobreposição dos histogramas das amostras de cafés torrados e moídos não adulterados (vermelho) e os cafés adulterados (azul) que parecem conter pixels com o mesmo tom de marron ou seja a tonalidade ou matiz (H) dos cafés para os modelos de cor **HSV e HLS**, indpendemente de estarem ou não adulteradas Esse fato motiva o uso de ferramentas da quimiometria disponíveis para a análise multivariada, conforme descrição na Seção **2.11**

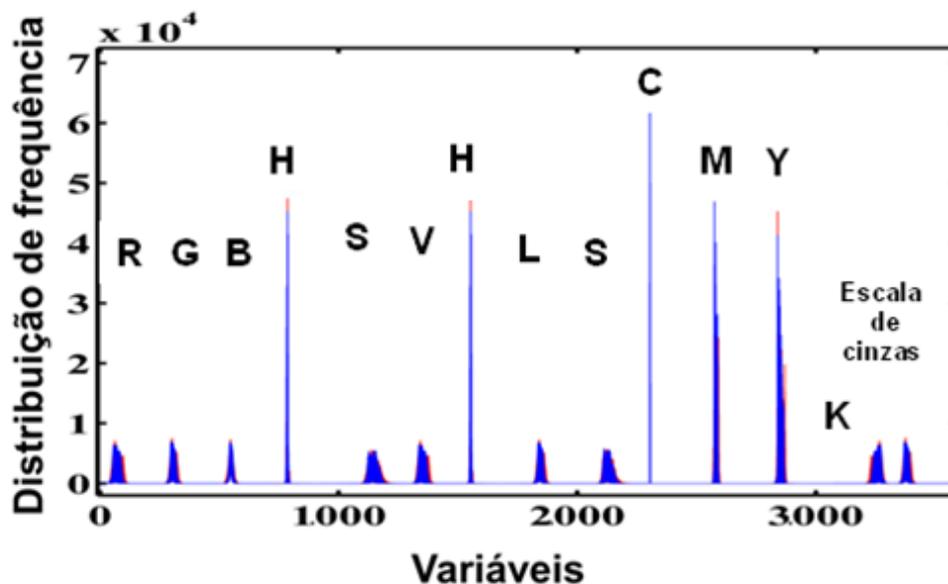


Figura4.0 - Histogramas de distribuição de frequências das amostras de cafés torrado e moído não adulterados (vermelho) e adulterados (azul), com uma matriz de dimensão de (103x3585) relacionados as amostras e variáveis, obtidos pelo Delfi 7.

4.1.1 Análise exploratória de adulteração de cafés

Para averiguar a possibilidade de uma possível discriminação entre as amostras de cafés não adulterados (círculo vermelho) e adulterados (quadrado azul) foi realizada a Análise de Componentes Principais (PCA) individualmente para cada modelo de cor (RGB,

HSV,HLS,CMYK) e suas combinações (**RGB-HSV**, **RGB-HLS,HSV-HLS**). Contudo, não foram vislumbrados discriminação entre as amostras de cafés, quando foi realizada a análise de componentes principais, nas combinações dos modelos de cores (**RGB-HSV**, **RGB-HLS,HSV-HLS**) e alguns modelos individuais (**RGB,HSV e CMYK**).

Além disso, não houve discriminação nas demais componentes principais, fato que possibilitou a explicação da variância, apenas as atribuições das três primeiras componentes principais (**PCs**). Não foi realizado nenhum pré-processamento nas variáveis, apenas a centralização dos dados na média, foi necessário as amostras.

Entretanto, o melhor resultado para fins de classificação dos cafés adulterados e não adulterados, foi obtido pelo modelo de cor **HLS**, conforme a apresentação dos resultados de classificação na **Tabelas 1.0 e Tabela 1.1** nas quais são apresentadas as figuras de mérito (acurácia, sensibilidade, especificidade) que evidenciaram um melhor resultado para o modelo de cor **HLS** em termos dos percentuais de acerto, para a acurácia, sensibilidade e especificidade nos conjuntos de treinamento e teste.

A seguir, são apresentados os gráficos dos escores de **PC1 versus PC2 e PC1 versus PC3** e o gráfico dos pesos na **Figura 4.3** para o modelo de cor **HLS**, com as (103x264) que correspondem as amostras e variáveis.

Ao analisar-se o gráfico dos escores para o modelo de cor **HLS** de **PC1 versus PC2** e **PC1 versus PC3** apresentados nas figuras **Figura 4.1. e Figura 4.2**, nota-se que não foi possível a discriminação por Análise de Componentes Principais (**PCA**).

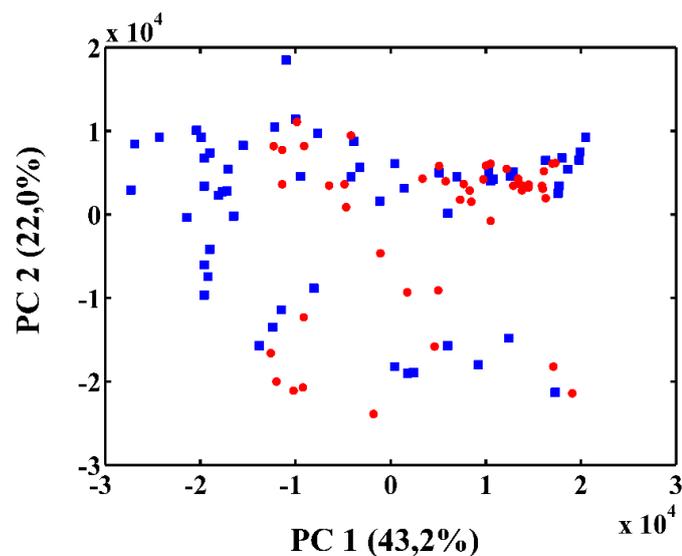


Figura 4.1 - Gráfico dos escores de **PC1 versus PC2** para as amostras de cafés adulterados (quadrado azul) e não adulterados (círculo vermelho).

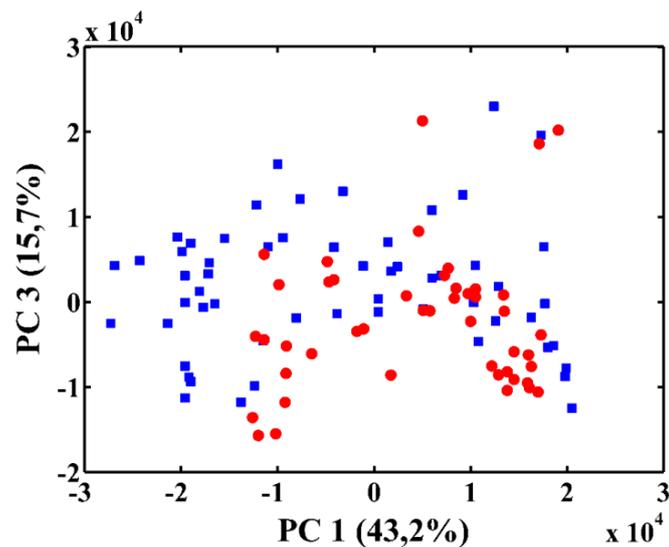


Figura 4.2 - Gráfico dos escores de **PC1 versus PC3** obtidos para das amostras de cafés adulterados (quadrado azul) e não adulterados (círculo vermelho).

A análise do gráfico dos escores apresentadas nas **Figura 4.1 e Figura 4.2** reflete a grande similaridade entre os cafés não adulterados (círculo vermelho) e adulterados (quadrado azul), não sendo possível a discriminação entre as amostras de cafés, apesar dos cafés não adulterados apresentarem propriedades distintas de tonalidade, luminosidade (brilho, intensidade) e saturação em relação aos cafés adulterados, que serão discutidas na seção **4.1.2**. Portanto há considerável similaridade dos tons entre os cafés que pode ser realçada através dos gráficos dos pesos apresentados a seguir na **Figura 4.3**.

De fato, o peso mais relevante que evidencia que há uma considerável similaridade entre as amostras de cafés, seja em relação a **PC1 versus PC2** ou **PC1 versus PC3** é referente à matiz, tonalidade (**H**) o qual exerceu maior influência na similaridade entre os cafés, não desconsiderando também as influências dos pesos referentes a saturação (**S**) e a luminosidade (**L**) com os seus respectivos intervalos, vislumbrados na **Figura 4.4** da seção **4.1.2** para o modelo **HLS** que possui uma matriz de dimensão (**103x264**) que correspondem as amostras e variáveis, respectivamente.

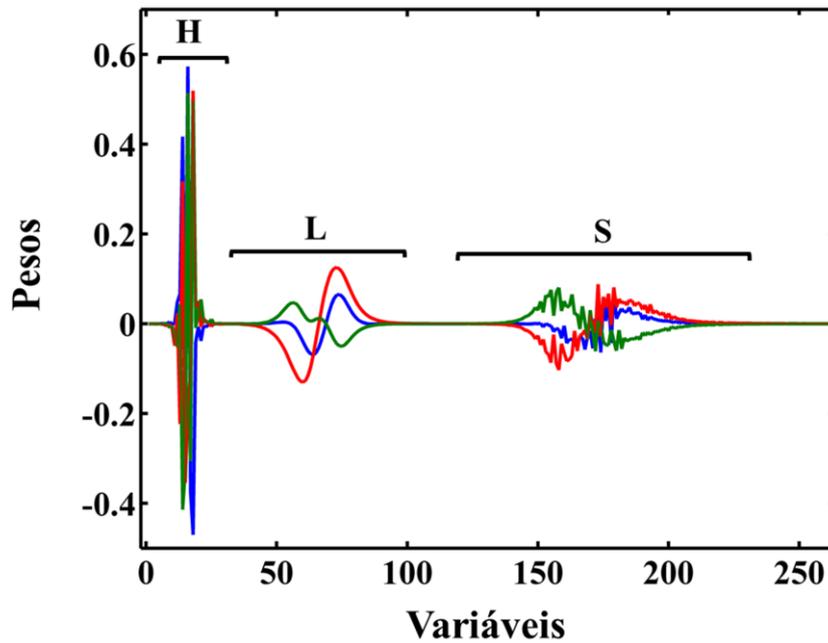


Figura 4.3 - Gráficos dos pesos do modelo **HLS**. Pesos de PC 1 (linha azul), pesos de PC 2 (linha vermelha) e pesos de PC 3 (linha verde).

Entretanto, foram avaliados diversos modelos de cores a partir de seus histogramas de distribuição de frequências e na **Tabela 1.0**, são apresentados os valores dos percentuais de classificação de acurácia.

Mediante a análise dos resultados de classificação da **Tabela 1.0 e Tabela 1.1**, torna-se possível concluir que os melhores resultados foram obtidos, com o **SPA-LDA** para o modelo de cor **HLS**, destacado em negrito com a validação cruzada.

Tabela 1.0- Acurácia dos modelos de cor dos histogramas de frequências

	Acurácia (%) Validação cruzada					
	SPA-LDA		PLS-DA		SIMCA	
	Validação (%)	Predição (%)	Validação (%)	Predição (%)	Validação (%)	Predição (%)
HLS	90,5	95,0	73,0	98,0	60,0	82,5
HSV	93,7	90,0	78,0	83,0	60,0	78,0
RGB	87,3	80,0	82,1	95,0	70,0	75,1
CMYK	92,1	87,5	74,3	88,0	69,1	88,0
Escala de cinza	82,5	72,5	73,0	63,0	76,2	78,1
RGB-HLS	85,7	92,5	76,2	93,0	65,0	80,0
RGB-HSV	92,1	92,5	79,1	85,0	69,0	70,0
HSV-HLS	82,7	87,1	72,0	98,0	65,0	73,0

Ao analisar os resultados apresentados dos modelos de cores na **Tabela 1.0** observa-se, que há pequenas diferenças em relação aos percentuais de classificação de acurácia nas etapas de validação e predição entre os modelos. Contudo, o modelo de cor **HLS**, com a seleção de variáveis e a análise discriminante linear (**SPA-LDA**), apresentou melhor resultado de classificação para os conjuntos de treinamento e teste, quando comparado aos modelos de cores por **HLS-PLS-DA** e **HLS-SIMCA** observados na **Tabela 1.1**.

Em função disso, pode-se justificar que o modelo de cor **HLS** separa a informação relativa a cor, associada à matiz e saturação do brilho, tal fato contribui de forma decisiva, para um melhor resultado de classificação nas etapas de validação e predição, apesar das variações de distribuição de luz que ocorre no sistema de iluminação do *scanner* (**Thiago Cesar 2008; Mariana Godinho 2014, Cristiano Bertolini 2010, Pedro Ivo, 2014**).

A seguir na Seção **4.1.2** são apresentadas as discussões e justificativas da análise screening dos cafés adulterados e não adulterados para as imagens digitais.

4.1.2 *Análise screening de adulteração de cafés*

Para aplicação dessas ferramentas, é necessário obter um conjunto de dados multivariados resultantes das medidas de propriedades que portem informações distintivas entre os cafés não adulterados e adulterados, nesse caso por cascas e paus. Para esse propósito as imagens digitais, obtidas das amostras de café não adulterado e adulterado, foram transformadas em um conjunto de dados multivariados usando o sistema de cor **HLS** (hue, luminosity, and saturation) (**Gonzalez e Woods 2000**). Esse modelo mostrou-se adequado para a distinção entre as classes de cafés pois, de acordo com a literatura, os adulterados têm alto valor de **H**, valores baixo de **L** e **S** quando comparados com cafés não adulterados (**Gonzalez e Woods 1992 ; Franca et al 2005 a ,b; Mendonça, 2008; Pauli 2010; Bicho et al 2012, 2013; Nadia Reis, 2012; Tavares 2012**).

De fato os grãos de café verde, quando são submetidos ao processo de torrefação, sofrem uma diminuição em sua tonalidade, saturação e brilho quando comparado ao grão de café torrado, variações que podem ser mais pronunciadas de acordo com o grau de torrefação (clara, média e escura). Assim, quando os grãos crus são submetidos aos processo de torrefação há uma perda em suas tonalidades de esverdeada e ou amarela – esverdeada para a tonalidade marrom, decorrente da ocorrência da reação de Maillard. Segundo **Bekdan et al 2006**, a reação de Maillard é uma reação de escurecimento não enzimático, entre a combinação do grupamento da carbonila de carboidratos redutores com o grupo amínico de aminoácidos, proteínas e peptídeos, na qual um de seus produtos corresponde as melanoidrinas, que proporcionam um tom de marron aos cafés torrados.

Além disso, a saturação é maior em torras clara e média e menor em torra escura, assim como o brilho (**Gonzalez e Woods 1992 ; Franca et al 2005 a,b ; Mendonça, 2008; Pauli 2010; Bicho et al 2012, 2013; Nadia Reis, 2012; Tavares 2012**).

As cascas e paus , quando são submetidas aos diferentes graus de torra (clara, media e escura) suas fibras sofrem uma maior degradação, porque apresentam uma estrutura fisica mais fina, quando comparada aos grãos de café, conseqüentemente alcançam mais rapidamente o grau de torrefação devido à maior exposição de sua superfície interna ou externa durante o processo de torrefação, o que pode ocasionar sua carbonização(**Gonzalez e Woods 1992 ; Franca et al 2005 a,b ; Mendonça, 2008; Pauli 2010; Bicho et al 2012, 2013; Nadia Reis, 2012; Tavares 2012**).

Portanto os parâmetros de cor com a tonalidade, o brilho e a saturação sofrem alterações, como resultado, as cascas e paus apresentam propriedades distintas em seus parâmetros de cor. Dessa forma há um menor brilho, saturação e maior tonalidade para as cascas, que pode

ser justificada devido à presença de uma menor quantidade de pigmentos avermelhados ou amarronzados e amarelados quando comparados aos grãos de cafés (Gonzalez e Woods 1992 ; Franca et al 2005 a ,b; Mendonça, 2008; Pauli 2010; Bicho et al 2012, 2013; Nadia Reis, 2012; Tavares 2012).

Vale ressaltar, que o modelo **HLS** é apropriado para o presente estudo visto que possibilita a separação entre a cromaticidade (definida como o grau de pureza da cor, estando relacionada com o **H** e **S**) e a luminosidade **L** (brilho, intensidade). Essa característica evita problemas com variações na distribuição da radiação durante a aquisição das imagens empregando um scanner (Gonzalez e Woods 1992 ; Franca et al 2005 a ,b; Mendonça, 2008; Pauli 2010; Bicho et al 2012, 2013; Nadia Reis, 2012; Tavares 2012).

Portanto as imagens digitais que foram obtidas através de um scanner, apresentaram uma melhor qualidade em sua resolução, com utilização modelo de cor (**HLS**). De fato, a consequência dessa característica, possibilitou uma melhor discriminação e classificação entre as classes dos cafés, cujos os resultados de classificação são apresentados nas **Tabela 1.0 e Tabela 1.1**. Ademais, eventuais mudanças na posição dos discos Petri não prejudica os resultados, pois esses dependem apenas da distribuição estatística dos pixels (histogramas de cor) oriundos dos componentes de cor das imagens digitais. (Diniz, 2012).

Em face dos argumentos discutidos nas **Seções 4.1.1 e 4.1.2**, foram gerados os histogramas **HLS** (máximo, mínimo e médio), apresentados na **Figura 4.4**, para as amostras de café não adulterados e adulterados, com cascas e paus. Como se pode observar, os histogramas indicam uma acentuada dispersão entre as classes estudadas. Esse comportamento pode ser imputado à forma e coloração das amostras de café, as quais são diretamente influenciadas pelo tipo de moagem e processo de torra.

A característica mais notável é que ambas as classes de cafés apresentam perfis similares nos histogramas médios obtidos para os três canais **H, L e S** conforme apresentado na **Figura 4.4**. Essa semelhança suscita o uso das técnicas quimiométricas para a discriminação entre as classes estudadas, cujos resultados são apresentados e discutidos na **Seção 4.1.1 e Seção 4.1.2**.

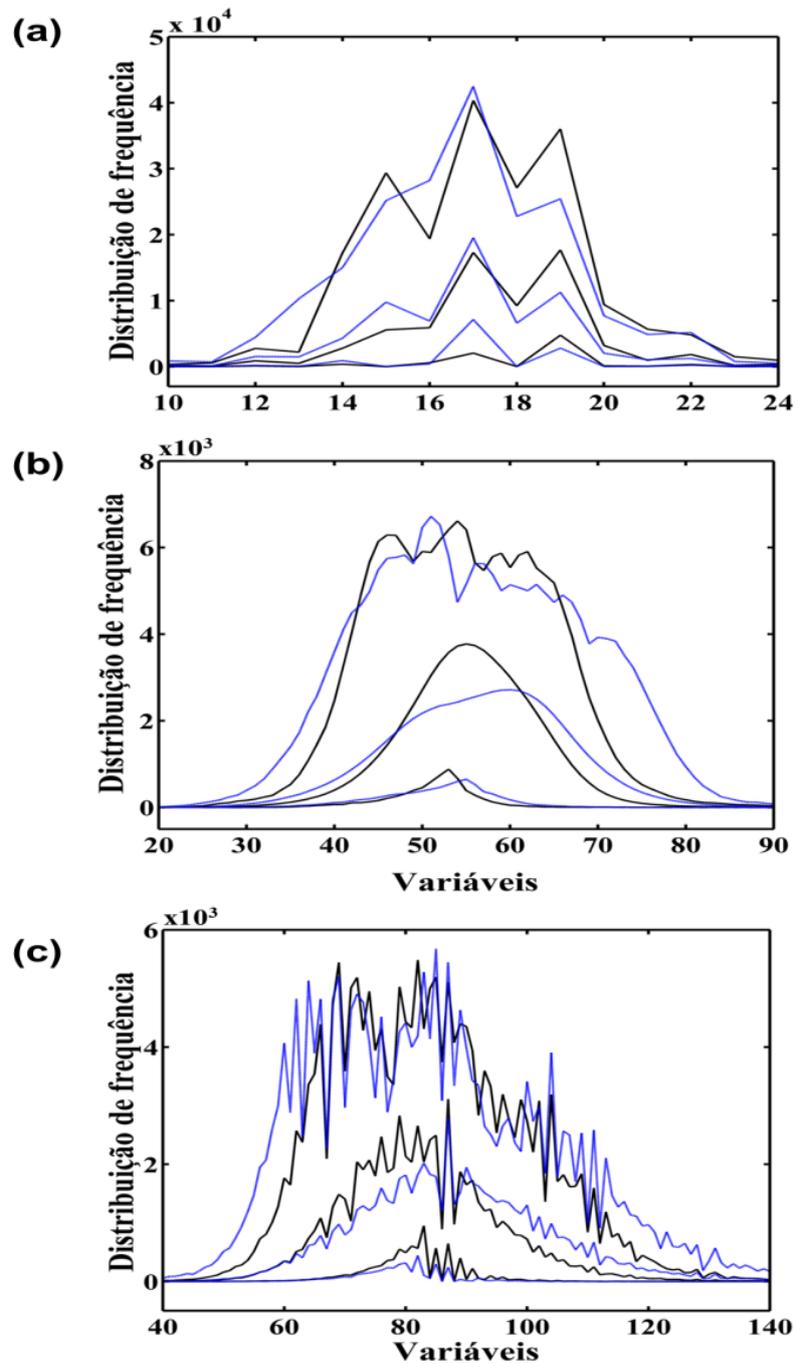


Figura 4.4 – Histogramas médios com os valores máximo, médio e mínimo do **H L S** correspondentes aos canais (a) **H** (11 a 24), (b) **L** (20 a 90) e (c) **S** (43 a 135) das 103 amostras dos cafés não adulterado (linha preta) e adulterado (linha azul).

◆ Modelagem **SIMCA**

Inicialmente, realizou-se uma modelagem SIMCA dos dados de histogramas gerados nos três canais a partir das imagens obtidas para cada classe de amostras de cafés não

adulterados (NA) e adulterados (A). Os modelos resultantes foram otimizados sendo que houve a necessidade de, respectivamente, 01 (uma) e 08 (oito) componentes principais para descrever as classes de amostras de café não adulterado e adulterado. Os resultados do *screening* de adulteração encontram-se reunidos na **Tabela 1.1** que é denominada matriz de confusão.

Tabela 1.1 – Matriz de confusão e métrica de desempenho para *screening* das amostras de cafés pelos modelos **SIMCA**, **PLS-DA**(limiar de **0,062**) e **SPA-LDA**, com nível de significância de 5% ,com base nos histogramas das imagens digitais obtidas a partir dos cafés torrado e moídos.

		Treinamento		Teste	
		Classe A	Classe NA	Classe A	Classe NA
Classificação SIMCA	Classe A(8 PCs) ^a	16	21	14	6
	Classe NA(1PC) ^a	4	22	1	19
Métrica SIMCA	Acurácia (%)	60,0		82,5	
	Sensibilidade (%)	80,0		93,3	
	Especificidade (%)	51,1		76,0	
Classificação PLS-DA (5 variáveis latentes) ^b	Classe A	28	9	19	1
	Classe NA	8	18	0	20
Métrica PLS-DA	Acurácia (%)	73,0		97,5	
	Sensibilidade (%)	69,0		100	
	Especificidade (%)	76,0		95,2	
Classificação SPA-LDA	Classe A	33	4	19	1
	Classe NA	2	24	1	19
Métrica SPA-LDA	Acurácia (%)	90,0		95,0	
	Sensibilidade (%)	94,2		95,0	
	Especificidade (%)	85,7		95,0	

A = amostra adulterada e NA = amostra não adulterada ;^a Ótimo número de PCs para cada classe indicado em parenteses ;^b Ótimo número de variáveis latentes indicada em parenteses

Pode-se observar na **Tabela 1.1** que o modelo **SIMCA** classificou corretamente 38 e 33 amostras de café (não adulterado e adulterado) o que representa uma taxa de acurácia de,

respectivamente, 60 e 83% para os conjuntos de treinamento e teste. Esse desempenho pode ser expresso também com base nas taxas de sensibilidade e especificidade, as quais representam uma medida da eficiência em relação às decisões positivas corretas (classe A) e negativas corretas (classe NA). No primeiro caso, o modelo **SIMCA** classificou 80 e 93% das amostras de café da classe (A), respectivamente, dos conjuntos de treinamento e teste. A taxa de especificidade de 51% foi obtida na classificação das amostras do conjunto de treinamento, enquanto a classificação das amostras de teste alcançou uma taxa maior de 76%.

A matriz de confusão da **Tabela 1.1** também apresenta os erros de classificação, expressos em termos de falsos negativos e falsos positivos. O **SIMCA** produziu 4 e 1 falsos negativos (20 e 7%) na classificação de amostras adulteradas como puras nos conjuntos de treinamento e teste. Quanto aos falsos positivos, foram 21 (49%) amostras puras de treinamento e 6 (24%) de teste classificadas como adulteradas.

◆ Modelagem **PLS-DA**

O modelo **PLS-DA** ótimo foi obtido usando 5 variáveis latentes o que permitiu uma acurácia na classificação de 73 e 98% (**Tabela 1.1**) das amostras de café dos conjuntos de, respectivamente, treinamento e teste. Com o desdobramento da acurácia, pode-se avaliar a eficiência da classificação com relação às decisões positivas corretas (sensibilidade) e negativas corretas (especificidade). O **PLS-DA** classificou corretamente as amostras da classe (A) cujas taxas de sensibilidade foram, respectivamente, de 69 e 100% para as amostras dos conjuntos de treinamento e teste. Quanto à especificidade, a taxa de acertos obtida na classificação das amostras da classe (NA) foi de 76 % no conjunto de treinamento e 95% no de teste, sendo adotado o limiar de valor **0,06** para inclusão de uma amostra em uma dada classe.

Na **Tabela 1.1**, pode-se observar ainda que o modelo **PLS-DA** cometeu os seguintes erros: 8 (31%) e 0 (0%) de falsos negativos foram obtidos quando as amostras adulteradas foram classificadas como puras, respectivamente, nos conjuntos de treinamento e teste. Por outro lado, o modelo **PLS-DA** produziu 9 (24%) e 1 (5%) de falsos positivos ao classificar amostras não adulteradas (classe NA) como adulteradas (classe A) nos conjuntos de treinamento e teste.

◆ *Modelagem SPA-LDA*

O modelo **SPA-LDA** foi construído e otimizado com base no valor mínimo da função-custo G obtido no conjunto de validação, conforme mostrado na **Figura 4.5a**.

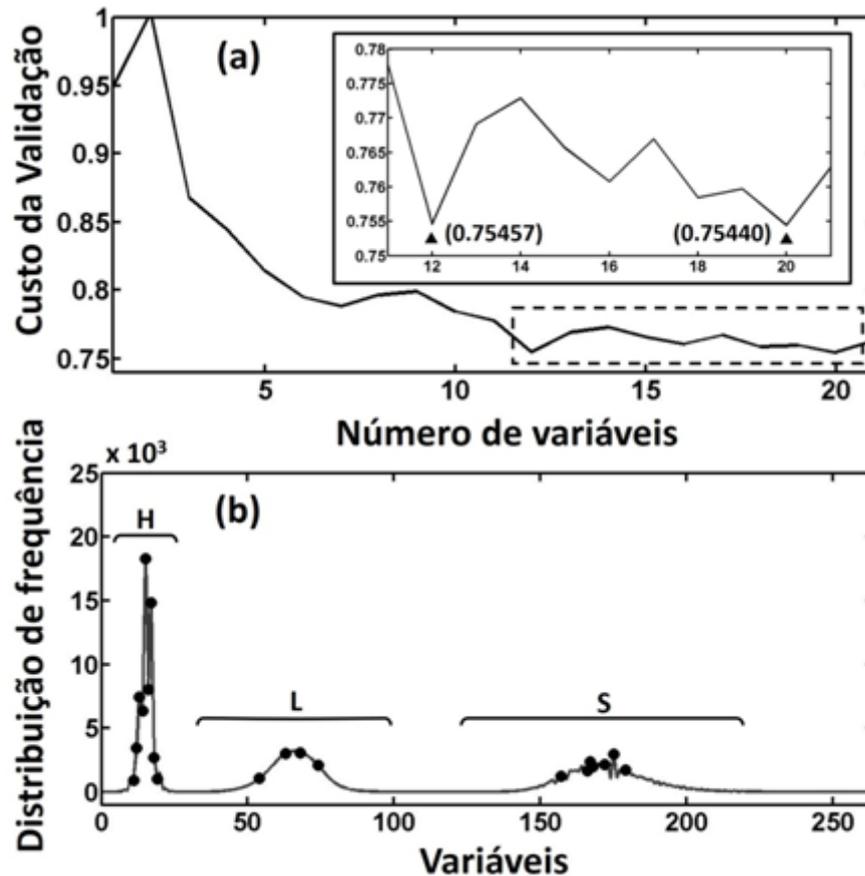


Figura 4.5 - (a) Gráfico do valor do custo de validação G (0,7544) versus número de variáveis selecionadas para o modelo **SPA-LDA** e **Figura 4.5 - (b)** Histograma médio das 103 amostras de café torrado e moído indicando as variáveis selecionadas (**circulo preto**), a segunda seta de cor preta em (a) indica o número ótimo de variáveis selecionadas pelo **SPA-LDA** (20 variáveis).

A **Figura 4.5 b** mostra que as variáveis selecionadas pelo **SPA-LDA** se distribuem ao longo do histograma para cada canal **H**, **L** e **S**. Essa resultado sugere que os três canais contribuem para habilidade do modelo **SPA-LDA** em discriminar as amostras puras das adulteradas por cascas e paus.

O modelo **SPA-LDA** foi, em seguida, aplicado às amostras dos conjuntos de treinamento e teste com o intuito de classificá-las como adulteradas (classe A) ou classe não adulterada (NA), os resultados são apresentados na **Tabela 1.1**. Ao contrário dos modelos

SIMCA e PLS-DA, o **SPA-LDA** apresentou alta taxa de acurácia na classificação das amostras dos conjuntos de treinamento (90%) e teste (95%).

Do ponto de vista da sensibilidade e especificidade, o modelo **SPA-LDA** apresentou também, um considerável desempenho na classificação. Com efeito, foram classificadas corretamente como adulteradas 94 e 95% das amostras da classe A (adulterada) nos conjuntos de, respectivamente, treinamento e teste. Analogamente, o modelo **SPA-LDA** classificou corretamente as amostras da classe NA (não adulterada) com alta taxa de especificidade, sendo 86% para as amostras de treinamento e 95% para as de teste.

O modelo **SPA-LDA** apresentou baixas taxas de erro de classificação, sendo que 2 (6%) e 1 (5%) de falsos negativos foram obtidos para as amostras dos conjuntos de, respectivamente, treinamento e teste. Quanto aos falsos positivos, o **SPA-LDA** cometeu 4 (14%) de erros ao classificar as amostras de treinamento e 1 (5%) na classificação das de teste.

São observados, conforme a **Tabela 1.1**, os resultados apresentados dos diferentes modelos de classificação para o café torrado e moído, em função dos percentuais de acurácia, sensibilidade e especificidade para o conjunto de treinamento e teste. Desse modo, nota-se que os valores elevados da especificidade principalmente para a modelagem **PLS-DA** (76,0%) e **SPA-LDA** (85,7%) quando comparado com os resultados de especificidade para a modelagem do **SIMCA** (51,1%) no conjunto de treinamento, ocasiona uma redução do número de falsos positivos nas etapas de treinamento e teste; **PLS-DA** (95,2%) e **SPA-LDA** (95%) e **SIMCA** (76,0%).

Entretanto, a modelagem **SIMCA** apresentou (**Tabela 1.1**) resultados inferiores aos dos modelos **PLS-DA e SPA-LDA** em termos dos percentuais de acurácia, sensibilidade e especificidade, referentes as etapas de treinamento e teste. Quando comparados o resultados dos modelos **PLS-DA e SPA-LDA** entre si, constata-se que o **SPA-LDA** apresentou melhor desempenho no conjunto de treinamento ao passo que no conjunto teste **PLS-DA** foi mais eficiente. Portanto, uma possível justificativa decorre do processo de seleção de variáveis pelo **SPA-LDA**, o qual seleciona variáveis com pequena multicolinearidade melhorando poder discriminatório conforme ilustrado na **Figura 4.5(b)**.

Portanto a modelagem **SIMCA** é baseada em função da **PCA**, que não possibilitou a discriminação das classes de cafés adulterados e não adulterados. Desse modo há uma possível explicação para este fato que pode ser justificado pela influência indireta dos processos físicos e químicos tais como: complexidade de composição química entres os cafés adulterados e não adulterados e os processos de torrefação que foram refletidos nas imagens

digitais, ocasionando uma considerável similaridade para os parâmetros de cor (tonalidade, saturação e luminosidade) especialmente o tom, conforme apresentado no gráfico dos pesos na **Figura 4.3** que são refletidos indiretamente na composição química das amostras dos cafés, que podem ser observados na **Figura 4.4**.

Os resultados dos percentuais apresentados de acurácia, sensibilidade e especificidade na etapa de treinamento e teste segundo a **Tabela 1.1** respectivamente, para a modelagem **SIMCA** conjunto de treinamento : (60,0%), (80,0%), (51,1%) e teste: (82,0%), (93,0%), (76,1%) consideravelmente inferior ao resultado do **SPA-LDA** para os conjuntos de treinamento: (90%), (94,2%) e (85,7%) e teste; (95,0%), (95,0%), (95,0%) .

Os resultados dos percentuais de acurácia, sensibilidade e especificidade do **PLS-DA** na etapa de treinamento, respectivamente são apresentados na **Tabela 1.1** e correspondem a : (73,0%) , (69,0%) , (76,0%) e para o conjunto teste ; (97,5%) , (100,0%) e (95,2%) são similares com a performance do modelo **SPA-LDA** para os conjunto de treinamento **SPA-LDA**; (90%), (94,2%) e (85,7%) e teste **SPA-LDA** (95,0%),(95,0%) e (95,0%).

De fato, o desempenho similar de classificação do **PLS-DA** em relação ao **SPA-LDA** para o conjunto teste, pode ser justificado através da correlação pelo método dos mínimos quadrados parciais na qual o **PLS-DA** estabelece entre os escores das classes dos cafés adulterados e não adulterados. Neste caso, a matrix (X) foi obtida a partir dos espectros de absorção **UV** das amostras dos cafés e sua correlação foi estabelecida, entre seus escores e com os escores da matrix (Y), que correspondem aos índices das classes. Desse modo, a covariância entre as matrizes foi maximizada e tal fato, resultou na obtenção das variáveis latentes, que são relacionadas à variabilidade nas matrizes (X) e (Y) quando comparada a **PCA**, que caracterizada apenas a variabilidade da matriz (X).

Em função disso, houve um melhor desempenho de classificação do **PLS-DA**, que pode ser observado na **Tabela 1.1**, através dos valores de percentuais de acurácia, sensibilidade e especificidade na etapa de treinamento e teste, quando comparado à modelagem **SIMCA**.

Portanto, as características que discriminam as categorias, por exemplo a classe dos cafés adulterados segundo a literatura especializada, correspondem a uma menor intensidade de brilho (intensidade,luminosidade) e saturação e uma maior tonalidade (tom) para as cascas, quando comparado ao café não adulterado (**Gonzalez e Woods 1992 ; Franca et al 2005 a ,b ; Mendonça, 2008; Pauli 2010; Bicho et al 2012, 2013; Nadia Reis, 2012; Tavares 2012**).

Em relação as amostras que foram mal classificadas, conforme a observação na **Tabela 1.1**, através dos resultados da acurácia, sensibilidade especificidade é possível analisar

que, os modelos que apresentaram um bom desempenho dessa mal classificação, correspondem as modelagem do **PLS-DA** , **SPA-LDA** quando comparados com o **SIMCA** nas etapas de treinamento : **SIMCA** (21,4), **PLS-DA** (9,8), **SPA-LDA** (4,2) e teste : **SIMCA**(6,1) , **PLS-DA** (1,0) e **SPA-LDA** (1,1).

Nota-se que, há uma redução do números de erros dos **FP** (especificidade) e **FN** (sensibilidade) para a mal classificação nas etapas de treinamento e teste, que podem ser decorrentes das características distintas de tonalidade, brilho e saturação entre os cafés adulterados(A) e não adulterados (NA), provavelmente pela influenciada de diversos fatores, dentre os quais como : plantação, colheita, beneficiamento e processos de torrefação(**Gonzalez e Woods 1992 ; Franca et al 2005 a ,b; Mendonça, 2008; Pauli 2010; Bicho et al 2012, 2013; Nadia Reis, 2012; Tavares 2012**).

4.2 Análise baseada em espectros UV e quimiometria

4.2.1 Análise exploratória de adulteração de cafés

Os espectros de absorção molecular **UV-Vís** foram obtidos a partir dos extratos aquosos das 102 amostras de cafés torrados e moídos (não adulterados e adulterados). Na **Figura 4.6**, são apresentados os espectros na faixa de 239 – 380 nm. Observe que a região dos espectros mais informativa, encontra-se relacionada com a banda de C=O associada à transição $n \rightarrow \pi^*$. Esse cromóforo encontra-se nas moléculas de trigonelina, cafeína, ácido caféico, ácidos clorogênicos e seus isômeros e em alguns ácidos orgânicos como ácido cítrico, ácido quinico protonado, ácido fenil acetico dentre outros e melanoidrinas (**Kogan et al ,1953 ; Vitorino et al , 2001; Martinez Lopez et al 2003; Bekdan et al 2006; Fujioka,K et al 2008; Moreira et al 2012, Moreira et al 2014**).

No entanto, a posição dessa banda pode mudar em função de interações intramoleculares (devido às olefinas conjugadas) e intermoleculares (devido ao efeito da polaridade da água usada nas extrações).A consequência é o deslocamento do máximo da banda no espectro da trigonelina de 278 nm para 280 nm, da cafeína de 274 nm para 275 nm, ácido caféico de 320 para 325 nm (**Kogan et al ,1953 ; Vitorino et al , 2001; Martinez Lopez et al 2003; Bekdan et al 2006; Fujioka,K et al 2008; Moreira et al 2012, Moreira et al 2014**).

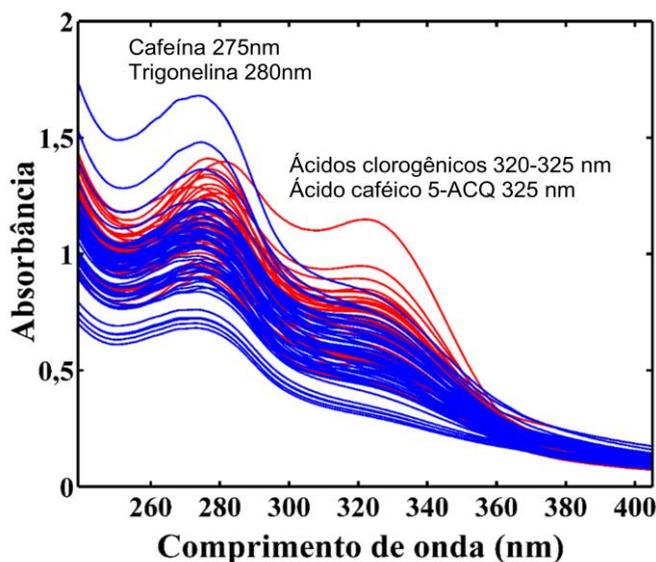


Figura 4.6 – Espectros de absorção UV de 102 amostras de café torrado na faixa de 239-380 nm. Em azul, são exibidos os espectros das amostras adulteradas em vermelho, os espectros das não adulteradas.

Como pode observar na **Figura 4.6**, os espectros das amostras de café adulterado são muito parecidos com os espectros de café não adulterado, o que dificulta a discriminação ou diferenciação visual. Em face do exposto, para superar essa dificuldade aplicou-se uma **PCA**, à matriz dos dados espectrais. A finalidade era avaliar a capacidade discriminante dos espectros, obtidos a partir do extrato aquoso das amostras de café, aliada ao uso da **PCA**, salienta-se que as demais componentes principais e suas combinações, não apresentaram poder discriminatório entre os cafés, apenas as duas primeiras **PCs**, que apresentaram uma tendência de discriminação.

Na **Figura 4.7**, são apresentados os gráficos dos escores de **PC1 versus PC2** e **PC1 versus PC3** das 102 amostras dos cafés, com uma dimensão de matriz de 102 amostras por 163 variáveis (**102x163**). Nota-se que, apesar da significativa superposição entre os agrupamentos, há uma tendência à separação das duas classes principalmente no gráfico dos escores de **PC1 versus PC2**. Entretanto a superposição pode ser atribuída à semelhança dos espectros, decorrente da similaridade entre as composições química das amostras de cafés.

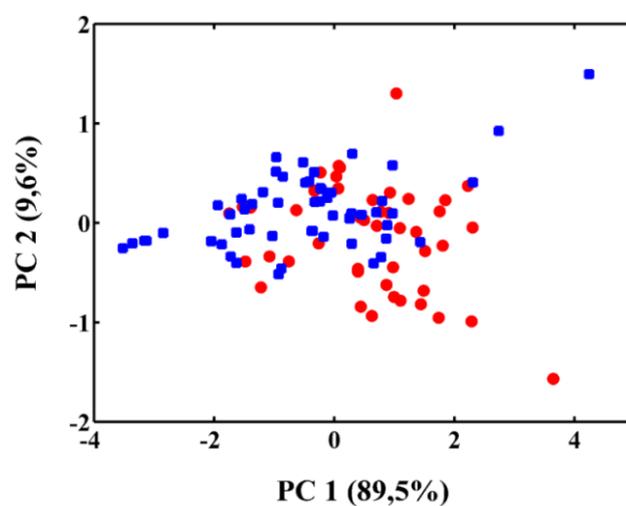


Figura 4.7 – Gráfico dos escores de **PC1 versus PC2** obtidos das amostras de cafés adulterados (quadrado azul) e não adulterados (círculo vermelho).

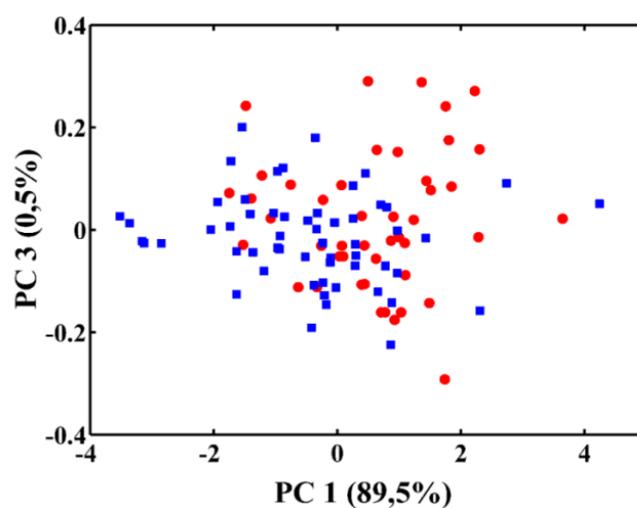


Figura 4.8 – Gráfico dos escores **PC1 versus PC3** obtidos das amostras de cafés adulterados (quadrado azul) e não adulterados (círculo vermelho).

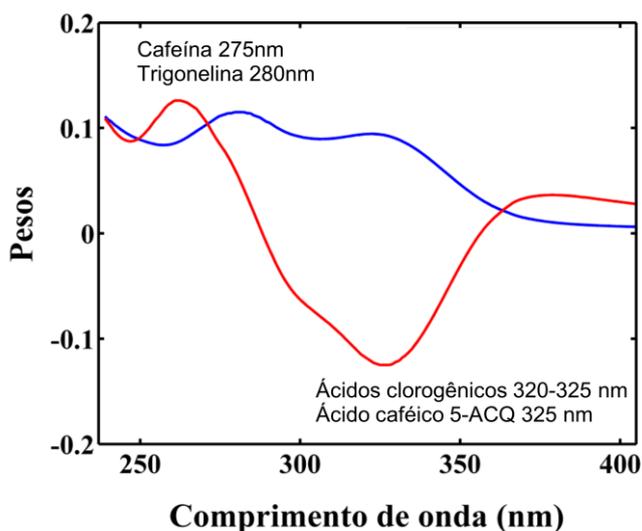


Figura 4.9 – Gráfico dos pesos **PC1** e **PC2** obtidos para as 102 amostras de cafés. Pesos de PC 1 (linha azul) e pesos de PC 2 (linha vermelha)

Pode-se observar na **Figura 4.9** do gráfico dos pesos que, os pesos mais significativos podem ser associados, às bandas de absorção da trigonelina com máximo de absorção a 280 nm, a cafeína em 275 nm, os ácidos clorogênicos na faixa de 320-325nm e o ácido caféico (5-ACQ) em 325nm e outros ácidos como o cítrico, quínico na faixa de 290 e 325 nm (**Kogan et al ,1953 ; Vitorino et al , 2001; Martínez Lopez et al 2003; Bekdan et al 2006; Fujioka,K et al 2008; Moreira et al 2012, Moreira et al 2014**).

Além disso, o gráfico dos escores na **Figuras 4.7 e Figura 4.8** e o gráfico dos pesos na **Figura 4.9**, observa-se que a uma propensão de discriminação mais pronunciada entre os escores das amostras de cafés não adulterados(**círculo vermelho**) e adulterados(**quadrado azul**), ao longo de **PC1 versus PC2** do que **PC1 versus PC3**.

Nota-se que, ao longo de **PC1**, os cafés não adulterados(**círculo vermelho**), apresentam em sua maioria, valores positivos de escores em relação aos adulterados (**quadrado azul**), que possuem escores negativos. Em relação ao gráfico dos escores de **PC1 versus PC3**, as amostras dos cafés adulterados(**quadrado azul**) e não adulterados(**círculo vermelho**), possuem valores de escores positivos e negativos, tanto em **PC1** quanto em **PC3**, observados na **Figura 4.8**. Portanto, há uma tendência de discriminação entre os cafés por **PC1 versus PC2** do que **PC1 versus PC3**.

Observa-se na **Figuras 4.7**, que os escores dos cafés adulterados(**quadrado azul**) apresentam uma dispersão menor do que os cafés não adulterados(**círculo vermelho**) com uma dispersão maior, ao longo de **PC1**. Uma possível explicação, para o fenômeno observado decorre do fato que, as amostras dos cafés adulterados(**quadrado azul**) possuem em sua

maioria, níveis de concentração de cascas e paus de 2,42% a 2,96% m/m ; 3,30 a 3,81% m/m ; 4,00 a 4,92% m/m e 5,06 a 5,70% m/m em **PC1**.

Ao analisar o gráfico dos pesos na **Figura 4.8**, os pesos que mais contribuíram significativamente para essa tendência de discriminação em **PC1**, correspondem às bandas de absorção da cafeína (275nm) e trigonelina (280nm) com os seus respectivos valores de pesos de 0,112 e 0,114, quando comparado aos pesos das bandas de absorção dos ácidos clorogênicos (320-325nm) e caféico **5-ACQ** (325nm) com valores de 0,094 a 0,093 e 0,093 respectivamente.

Em **PC2**, nota-se que os pesos referentes às bandas de absorções da cafeína (275nm) e trigonelina (280nm) possuem, os seguintes valores de pesos ; 0,077 e 0,004, os ácidos clorogênicos (320-325nm) de 0,119 a 0,125 e para o ácido caféico, **5-ACQ** (325nm) o valor de 0,125, conforme observação da **Figura 4.9**. Apesar dos ácidos clorogênicos e caféico (**5-ACQ**) apresentarem maiores valores de pesos, quando comparados aos pesos da cafeína e trigonelina em **PC2**, não foi possível vislumbrar discriminação, entre as classes dos cafés ao longo de **PC2**.

Ao analisar os gráfico dos escores na **Figura 4.7** e o gráfico dos pesos na **Figura 4.9** há uma possível explicação para esse fato, os extratos dos cafés não adulterados (**círculo vermelho**) que apresentam em sua composição química, uma maior concentração de cafeína e trigonelina, quando comparado aos extratos das cascas de café dos adulterados (**quadrado azul**) com uma menor concentração de cafeína e trigonelina, tal fato contribui para a tendência de discriminação ao longo de **PC1** (**Barcellos et al 1997 a, 1997 b; Aretha et al, 2009; Pereira et al 2014; Aquino et al 2014**).

Além disso, os extratos das cascas de café dos adulterados(**quadrado azul**) são ausentes ou apresentam baixíssimas concentrações do ácido caféico (**5-ACQ**) um dos constituintes majoritários dos extratos aquosos dos cafés não adulterados(**círculo vermelho**) (**Barcellos et al 1997 a, 1997 b; Aretha et al, 2009; Pereira et al 2014; Aquino et al 2014**).

De fato, nota-se pela **Figura 4.6** que as amostras dos cafés adulterados(**espectro azul**), apresentam bandas de absorção de menor intensidade na região dos ácidos clorogênicos, de 320 nm a 325nm e do ácido caféico (**5-ACQ**) em 325 nm, quando comparados aos cafés não adulterados(**espectro vermelho**). Portanto, as absorções dos extratos dos cafés adulterados são de menor intensidade em relação as dos não adulterados.

4.2.2 Análise *screening* de adulteração de cafés

Uma vez realizada a análise exploratória preliminar, construiu-se os modelos **SIMCA**, **PLS-DA** e **SPA-LDA** a partir dos espectros do extrato aquoso dos cafés para *screening* de adulteração por cascas e paus. A **Tabela 1.2** apresenta os resultados da discriminação entre os cafés adulterados e não adulterados. Para isso, o desempenho dos modelos utilizados é expresso em termos de acurácia, sensibilidade e especificidade.

Tabela 1.2 – Matriz de confusão e métrica de desempenho para o *screening* das amostras de cafés pelos modelos **SIMCA**, **PLS-DA** (limiar de 0,060) e **SPA-LDA** com nível de significância de 5% ,com base nos espectros **UV**

		Treinamento		Teste	
		Classe A	Classe NA	Classe A	Classe NA
Classificação SIMCA	Classe A(6 PCs) ^a	23	14	13	7
	Classe NA(3PC) ^a	3	23	1	18
Métrica SIMCA	Acurácia (%)	73.0		79.5	
	Sensibilidade (%)	62.2		65.0	
	Especificidade (%)	88.5		94.7	
Classificação PLS-DA (11 variáveis latentes) ^b	Classe A	28	1	20	0
	Classe NA	1	18	1	18
Métrica PLS-DA	Acurácia (%)	96,8		97.4	
	Sensibilidade (%)	97,3		100	
	Especificidade (%)	96,1		94.7	
Classificação SPA-LDA	Classe A	37	0	20	0
	Classe NA	0	26	0	19
Métrica SPA-LDA	Acurácia (%)	100		100	
	Sensibilidade (%)	100		100	
	Especificidade (%)	100		100	

^a Ótimo número de PCs para cada classe indicado em parenteses ^b Ótimo número de variáveis latentes indicada em parenteses

A matriz de confusão da **Tabela 1.2** mostra que, de um modo geral, o modelo **PLS-DA** permitiu realizar uma excelente classificação, cujo desempenho foi superior a 94% para as três métricas adotadas. Não obstante, o modelo **SPA-LDA** obteve 100% de acerto na classificação de todas as amostras dos conjuntos de treinamento e teste.

Os resultados apresentados nos diferentes modelos de classificação para o extrato aquoso, são observados conforme a **Tabela 1.2** em função dos percentuais de acurácia, sensibilidade e especificidade. Desse modo, nota-se que os valores elevados da especificidade pode ser refletido na redução do número de falsos positivos nas etapas de treinamento ; **SIMCA** (88,5%), **PLS-DA** (96,1%) , **SPA-LDA**(100%) e teste; **SIMCA** (94,7%), **PLS-DA** (97,4%) e **SPA-LDA**(100%).

Entretanto as modelagens do **SIMCA** e do **PLS-DA** apresentaram resultados de classificação inferior nos percentuais de acurácia, sensibilidade e especificidade, nas etapas de treinamento e teste, quando comparados ao **SPA-LDA** conforme a **Tabela 1.2**. Portanto, uma possível justificativa decorre do processo de seleção de variáveis pelo **SPA-LDA**, o qual minimizou o problema da colinearidade e com isso selecionou as 21 variáveis, que são visualizadas na **Figura 4.9.1 (b)**.

Segundo a **Tabela 1.2**, os resultados dos percentuais em termos de respectivamente acurácia, sensibilidade e especificidade na etapa de treinamento para a modelagem **SIMCA** (73,2%, 62,2% e 88,5%) foram inferiores aos dos demais modelos **PLS-DA** (96,8%, 97,3%, e 96,1%) e **SPA-LDA** (100%, 100% e 100%). De fato, a modelagem **SIMCA** é realizada a partir da Análise de Componentes Principais **PCA**, apresentado na **Figura 4.7**, a qual não evidenciou uma considerável discriminação entre as classes dos cafés, mas há tendência de discriminação ao longo de **PC1 versus PC2**. Portanto, uma possível explicação para esse fato decorre da complexidade da composição química dos cafés adulterados e não adulterados. Entretanto, há características distintas entre os cafés adulterados dos não adulterados, que foram referidas na seção **4.2** e que podem ser corroboradas, com os resultados de classificação da **Tabela 1.2**.

Os resultados dos percentuais de acurácia, sensibilidade e especificidade do **PLS-DA** na etapa de treinamento são, respectivamente, 96,8% , 97,3% e 96,1% conforme observado na **Tabela 1.2**. Para o conjunto teste, **PLS-DA** apresentou os valores 97,4% , 100% e 94,7% para essas métricas de desempenho, os quais são praticamente similares aos valores do **SPA-LDA** (100%, 100% e 100%). A seguir são discutidas as possíveis causas que ocasionaram a discriminação e classificação entre os cafés adulterados e não adulterado, conforme apresentados na **Tabela 1.2**.

Segundo a literatura especializada, as características que podem ser consideradas como discriminante e classificatória entre os extratos aquosos dos cafés não adulterados dos adulterados, resultam da presença dos ácidos clorogênicos e o ácido caféico (**5-ACQ**) que não se encontram nas cascas e paus, bem como as variações entre as concentrações de trigonelina

e cafeína (Aretha et al, 2009 ;Andrade et al 2009 (a), Aquino et al 2014). Nas cascas de café, há uma menor concentração de trigonelina e cafeína, quando comparado ao café torrado não adulterado, que apresenta uma maior concentração de cafeína e trigonelina. Além disso, nas cascas de café há ausência dos ácidos clorogênicos e cafeico (Aretha et al, 2009 ;Andrade et al 2009 (a), Aquino et al 2014 ; Pereira et al 2014).

Aretha et al 2009, pela análise da cromatografia líquida de alta eficiência (CLAE) para o café torrado, foram detectados e quantificados os picos característicos da cafeína, trigonelina e ácidos clorogênicos no cromatograma, que foram confirmados, tanto pelos tempos de retenção e quanto pelo espectro de absorção molecular na região do ultravioleta (UV). Contudo, no cromatograma das cascas foram apenas identificados, os picos de absorção molecular no ultravioleta (UV) da trigonelina e cafeína e não detectados os picos característicos dos ácidos clorogênicos e do ácido cafeico (5-ACQ).

Andrade et al 2009, pela análise da cromatografia líquida de alta eficiência (CLAE) dos constituintes bioativos do café torrado e da casca, foram detectados e quantificados no cromatograma, há presença da cafeína, trigonelina, ácidos clorogênicos e ácido cafeico (5-ACQ) nos cafés torrados e para as cascas, não foram detectados os picos característicos dos ácidos clorogênicos e o do ácido cafeico (5-ACQ). Além disso, os níveis de concentração de cafeína e trigonelina são distintos, o café torrado apresentou maior concentração de cafeína cerca de três vezes maior quando comparado as cascas.

Aquino et al 2014, com a utilização da espectrometria de massa (m/z) foi possível detectar, os picos característicos dos constituintes bioativos no café torrado e das cascas, mas as cascas de café, não apresentaram a presença dos picos dos ácidos clorogênicos e do ácido cafeico (5-ACQ) que são os constituintes majoritários dos extratos aquosos dos cafés torrados.

Para os níveis de concentração da trigonelina e cafeína há variações, as cascas de café apresentaram picos de menor intensidade de trigonelina e cafeína.

Pereira et al 2014, através da análise da cromatografia líquida de alta eficiência dos constituintes bioativos do café e da casca crus, foi possível detectar e quantificar no cromatograma a presença dos constituintes bioativos. Entretanto, as cascas de café apresentaram concentrações inferiores de cafeína, trigonelina e ácido cafeico em relação aos grãos de cafés.

Para as cascas de café, os níveis de concentração da cafeína correspondeu a metade (0,8g/100g) quando comparado ao café (1,61g/100g), o ácido cafeico (5-ACQ) cerca de oito vezes menor (0,55g/100g) em relação ao café (4,82g/100g) e a trigonelina nas cascas apresentou concentração inferior (1,31g/100g) ao café (1,36g/100g).

Como observado na **Tabela 1.2**, os modelos de classificação do **PLS-DA, SPA-LDA** apresentaram, melhor resultado de classificação em função dos percentuais de acurácia, sensibilidade e especificidade nos conjuntos de treinamento e teste do que o modelo **SIMCA**, respectivamente: **SIMCA** (73,2%, 62,2% e 88,5%), **PLS-DA** (96,8%, 97,3%, e 96,1%) e **SPA-LDA** (100%, 100% e 100%); **SIMCA** (79,5%, 65,5%, 94,7%), **PLS-DA** (97,4%, 100% e 94,7%) e **SPA-LDA** (100%, 100% e 100%).

Em relação as amostras que apresentaram uma mal classificação, vislumbradas na **Tabela 1.2** e suas figuras de mérito, como a acurácia, sensibilidade e especificidade, os modelos que apresentaram um bom desempenho para essa mal classificação, correspondem as modelagem do **PLS-DA e SPA-LDA** quando comparados com o **SIMCA**, nas etapas de treinamento : **SIMCA** (14,3), **PLS-DA** (1,1) , **SPA-LDA** (0,0) e teste: **SIMCA**(7,1) , **PLS-DA** (0,1) e **SPA-LDA** (0,0).

Observa-se que, há uma redução do números de erros dos **FP** (especificidade) e **FN** (sensibilidade) para a mal classificação das amostras adulteradas, sendo classificadas como não adulteradas e não adulteradas classificadas como adulteradas, nas etapas de treinamento e teste apresentados na **Tabela 1.2**.

No caso das amostras adulteradas classificadas como não adulteradas na modelagem **SIMCA**, as amostras adulteradas apresentaram níveis de concentração dos contaminantes por cascas e paus, para ao conjunto de treinamento um valor mínimo de 1,08 % m/m e máximo de 8,49% m/m e para o conjunto teste valor mínimo de 2,22% m/m e máximo de 6,46% m/m e para o **SPA-LDA**, não houve nenhum erro de classificação, nos conjuntos de treinamento (0,0) e teste (0,0). Para o **PLS-DA**, apenas uma amostras adulterada foi classificada como não adulterada no conjunto de treinamento e seu nível de concentração por cascas e paus foi de 4,30% m/m.

Em virtude dessa mal classificação, entre as amostras adulteradas e não adulteradas há diversos fatores, que contribuíram para a mal classificação entres as amostras dos cafés, como exemplo: a variabilidade da composição química entre os cafés , a presença dos níveis de concentração dos contaminantes (cascas e paus) e os processos de torrefação com torra clara, média e escura que são alteradas e ocasionam modificação física e química nas cascas e paus, com isso possibilitou a mal classificação entre os cafés. (**Aretha et al, 2009 ; Andrade et al 2009 (a) ; Tavares et al, 2012 ; Aquino et al 2014**).

A seguir há uma discussão sobre o resultado obtido pelo **SPA-LDA**, em função do número de variáveis selecionadas. Conforme ilustrado na **Figura 4.9.1**, o valor mínimo de custo G para a classificação, corresponde a **0,5693** que foi determinado em função do número mínimo de variáveis para etapa de validação, **(a)**. A seta indica o número de variáveis (21) utilizadas no processo de validação, que correspondente ao valor mínimo do custo assinaladas na parte **(b)**.

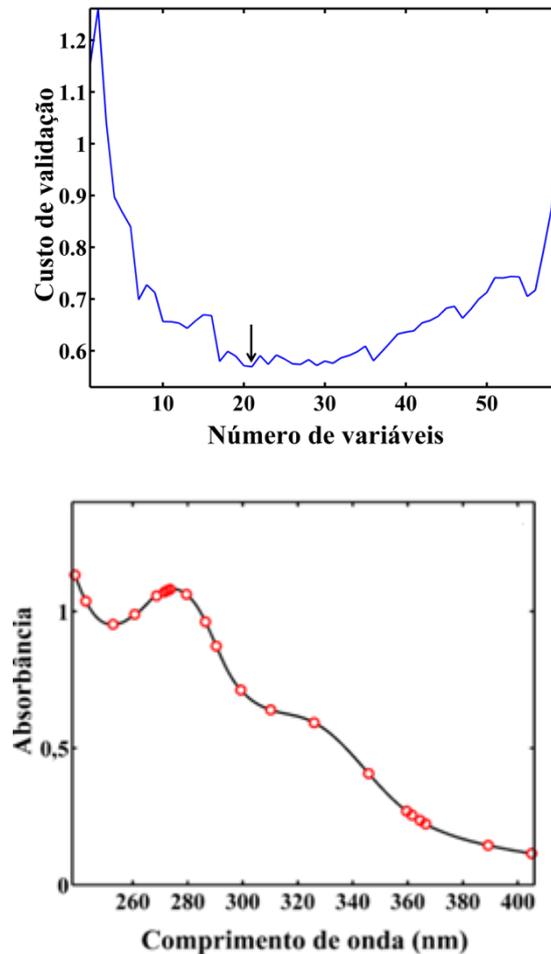


Figura 4.9.1 – (a) valor do custo de validação G (0.5693) em função do número de variáveis usadas na modelagem **SPA-LDA** e (b) indicação das 21 variáveis selecionadas (**círculo vermelho**) no espectro médio (**linha preta**) pela seta preta em (a) para as 102 amostras de café.

Com as variáveis assinaladas na **Figura 4.9.1 (b)** e segundo a ilustração dos escores da função discriminante de Fisher, apresentada na **Figura 4.9** dos extratos aquosos das 39 amostras de cafés do conjunto de teste, nota-se que há uma clara e eficiente separação entre as amostras não-adulteradas (**círculo vermelho**) e as adulteradas (**quadrado azul**) por cascas e paus, que pode ser atribuída à capacidade do modelo **SPA-LDA** de selecionar variáveis com a

mínima colinearidade ou seja os comprimentos de onda que mais contribuíram para a discriminação e classificação dos cafés (Kogan et al ,1953 ; Vitorino et al , 2001; Martinez Lopez et al 2003; Bekdan et al 2006; Fujioka,K et al 2008; Moreira et al 2012, Moreira et al 2014).

Segundo a literatura especializada, os extratos aquosos das cascas de café possuem exclusivamente uma maior concentração de ácido cítrico e ácido quínico protonado. Além disso, têm-se uma baixíssima concentração ou ausência de açúcares e ácidos clorogênicos quando comparados aos extratos aquosos dos cafés não adulterados, com os seguintes ácidos: o ácido cafeico desidratado, ácido quínico desidratado, ácido cafeico e ácido fenil acetico que, não são encontrados nos extratos aquosos das cascas e paus (Aretha et al, 2009 ;Pereira et al 2014; Aquino et al 2014).

Mediante a apresentação do gráfico dos escores da função discriminante, apresentado na **Figura 4.9**. Nota-se que, as amostras dos cafés adulterados(quadrado azul) que mais se aproximam do limiar entre as classes dos cafés não adulterados(círculo vermelho) correspondem as amostras, que possuem as concentrações de cascas e paus de 2,19g/100g e 2,22g/100g em relação as demais de 2,46 a 6,53 g/100g.

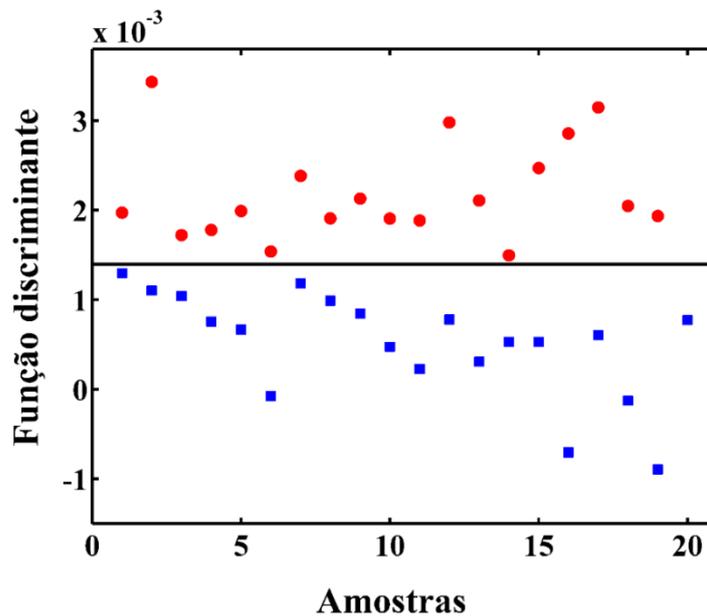


Figura 4.9.2 – Função discriminante de Fisher para as 39 amostras de cafés do conjunto de teste usando as 21 variáveis apresentadas na **Figura 4.9.1 b**.

Portanto, os fatores apresentados e discutidos nesta seção, foram os responsáveis pela satisfatória discriminação e classificação dos extratos aquosos dos cafés adulterados e não adulterados.

Capítulo 5

Conclusão

5.0 Conclusões

Neste trabalho foram desenvolvidas novas metodologias com a finalidade de averiguar a conformidade de cafés comercializados sujeitos à adulteração, particularmente utilizando-se cascas e paus. Para esse propósito utilizou-se, em uma das estratégias, dados multivariados de absorção molecular na região do ultra violeta (**UV**) de amostras de extratos aquosos de cafés com as técnicas de reconhecimento de padrão supervisionada, como o: **SIMCA PLS-DA e SPA-LDA**. Na outra estratégia, usou-se imagens digitais capturadas dos cafés torrados e moídos e modelos quimiométricos foram construídos com as técnicas referidas anteriormente.

Na metodologia baseada em imagens digitais, o modelo **SPA-LDA** classificou as amostras de ambos os conjuntos de treinamento e teste com 92,5% de taxa média de acurácia enquanto os modelos **SIMCA e PLS-DA** com, respectivamente, a média de 71,5% e 85,5% para os referidos conjuntos.

Por outro lado, o modelo **SPA-LDA** apresentou um desempenho ainda melhor quando usado para classificação das amostras dos conjuntos de treinamento e teste, com base nos dados espectrais **UV** de seus extratos aquosos. De fato, todas as amostras de ambos os conjuntos foram classificadas com 100% de acurácia ao passo que os modelos **PLS-DA e SIMCA** apresentaram taxa média de, respectivamente, 97,1% e 76,3%.

Na classificação baseada nos espectros **UV** dos extratos aquosos, os modelos **SPA-LDA e PLS-DA** apresentaram um desempenho melhor nas etapas de treinamento e teste. Tal fato pode ser atribuído, à informação físico-química dos constituintes majoritários dos extratos aquosos, sobretudo os ácidos clorogênicos e o ácido cafeico (**5-ACQ**) em comparação às imagens digitais obtidas dos pós de cafés, apesar desta técnica ser menos invasiva e destrutiva.

Os modelos de classificação do **SIMCA** das imagens digitais e o ultravioleta **UV** apresentaram diferenças significativas entre os percentuais de acurácia, sensibilidade e especificidade, conforme os resultados apresentados nas tabelas **Tabela 1.1** e **Tabela 1.2** que há uma diferença considerável na sensibilidade que para as imagens digitais, apresentou um valor de 93,0% de sensibilidade, quando comparado ao **UV** com 63,0% nos conjuntos de teste, desse modo o modelo pode determinar, quais são os cafés que realmente estão adulterados por cascas e paus.

Além disso, há um ganho considerável de especificidade de 76,0% das imagens digitais para 94,7% do **UV**, nos conjuntos de teste, tal fato contribuiu para o modelo determinar quais são os cafés, que não estão adulterados por cascas e paus.

Os modelos de classificação do **PLS-DA** das imagens digitais e o ultravioleta **UV** não apresentaram diferenças significativas entre os percentuais de acurácia, sensibilidade e especificidade para o conjunto teste, mesmo havendo diferenças técnicas entre as imagens digitais e o **UV**, tal fato justifica a necessidade para a utilização das novas metodologias desenvolvidas nesse trabalho.

O uso das imagens digitais adquiridas com um scanner torna o processo de verificação da conformidade mais rápido e menos trabalhoso e dispendioso. Essas vantagens são particularmente interessantes para a inspeção de amostras de cafés em um estágio de controle de qualidade antes da comercialização, ao passo que o uso dos espectros ultravioleta (**UV**) dos extratos, torna a metodologia proposta mais adequada e atrativa para a inspeção da bebida final.

Portanto, ambas as estratégias da metodologia proposta apresentam características complementares entre si auxiliando o controle de qualidade e conseqüentemente, proporcionam uma maior confiabilidade ao consumidor na escolha dos cafés. Além disso, podem auxiliar os órgãos de fiscalização em uma inspeção para a verificação de adulteração dos cafés por cascas e paus e possibilitar a utilização das metodologia desenvolvidas, através da Associação Brasileira da Indústria e do Café (**ABIC, 2016**) aos laboratórios credenciados, tais como : o **NUGAP** (núcleo de apoio global), **ITAL** (Instituto tecnológico de alimentos), **GAG** (Grupo de ação global).

5.1 Propostas de trabalhos futuros

Propõe-se, como trabalho(s) futuro(s), o uso da espectroscopia no infravermelho próximo (**NIR**) aliada a técnicas de reconhecimento de padrão supervisionadas (**SIMCA, SPA-LDA e PLS-DA**) para construção de novos modelos de classificação.

Além disso, propõe-se o uso de imagens hiperespectrais no **NIR** para aquisição direta nos pós de amostras de cafés a fim de verificar também a presença de outros adulterantes.

Espera-se generalizar a metodologia desenvolvida neste trabalho de modo a aplicá-la à verificação da presença de outros adulterantes, a exemplo do trigoilho, milho, arroz, centeio, entre outros

Referências

<http://www.anvisa.com.br>. Acessada, em outubro de 2016.

<http://www.abic.com.br>. Acessada, em outubro de 2016.

<http://www.agricultura.gov.br>. Acessada, em outubro de 2016.

ANDRADE, A. P.S. **Análise química e avaliação do potencial aleopático da casca do café (Coffea Arábica)**,Uberlândia, Minas Gerais -M G.(2009). p. 107: il (Dissertação de Mestrado – Programa de Pós-graduação em Química).

ANDRADE, A. P.S, MELO,P.G.AQUINO,F.J.T;CHANG,R; MORAIS,S.A.L;SANTOS,D.Q; HERNANDES, TERRONES,M.G. **Avaliação do potencial aleopático e efeitos fitoquímico obtidos da casca de café(Coffea arábica)**. In: XXXII REUNIÃO DA SOCIEDADE BRASILEIRA DE QUÍMICA, Fortaleza, Ceará, 2009.

ANZANELLO, M.; ALBIN, S. W.; **Chaovalitwongse, Multicriteria variable selection for classification of production batches**, Eur. J. Oper. Res.v 218. (2012).p. 97–105.

AQUINO, J.T.; et al, **Direct infusion electrospray ionization coffees adulterated with their husks**. *Microchemical Journal*.v.117. (2014).p.127-132.

ATLAS DE MICROSCOPIA DO CAFÉ TORRADO E MOÍDO (Coffea sp)-Funed (Fundação Ezequiel Dias). MATOS, E. C.; MATOSINHOS, F. C. L.; MARIA ELOIZA SAMPAIO VIRGINIA DEL CARMEN TRONCOSO VALENZUELA; MOREIRA, W. A.; volume único. Belo Horizonte Minas Gerais.M.G.(2016).p.1-48.

AYCHEH, H. M. **Image Analysis for Ethiopian Coffee Classification**.Addis Ababa **University School of Graduate Studies**. Faculty of Informatics. Department of Computer Science, January.(2008).p.99 : il (A Thesis Submitted to the School of Graduate Studies of Addis Ababa University in Partial Fulfillment for the Degree of Master of Science in Computer Science).

BARCELOS, A.F; ANDRADE, I.F; TIENSENHAUSEN, I. M. E. V. **Aproveitamento da casa de café na alimentação de novilhos confinados I. Resultados do primeiro ano.** Revista Brasileira de Zootecnia, v.26, n.6,(1997).p.1215-1221.

BARCELOS, A.F; ANDRADE, I.F; TIENSENHAUSEN, I.M.E.V. **Aproveitamento da casa de café na alimentação de novilhos confinados I. Resultados do primeiro ano.** Revista Brasileira de Zootecnia, v.26, n.6, (1997).p 1208-1214.

BEEBE, K. R., PELL, R. J.; SEASHOLTZ, M. B., **Chemometrics- A Pratical Guide**, Wiley, New York, (1998).

BEKEDAM, E. K.; LOOTS, M. J.; SCHOLS, H. A.; VAN BOEKEL, M. A. J. S.; SMIT, G. **Roasting effects on formation mechanisms of coffee brew melanoidins.** Journal Agricultural and Food Chemistry, v.106. (2008).p.7138-7145.

BERTOLINI, C.. **Sistema para medição de cores usando um espectrofotômetro.** Trabalho de Conclusão de Curso de Bacharelado II- Curso Ciência da Computação Universidade Regional de Blumenau. Santa Catarina, SC. p.87, 2010.

BICHO N. C.; LIDON F. C.; RAMALHO J. C.; LEITÃO A. E. **Quality assessment of Arabica and Robusta green and roasted coffees-a review.** Emir J Food Agric .v.25.(2013).p. 945–950.

BICHO N.C.; LEITÃO A. E.; RAMALHO J.C.; LIDON F. C. **Use of colour parameters for roasted coffee assessment.** Ciência Tecnologia dos Alimentos.v 32. (2012).p.436–442.

CHARRIER, A; BERTHAUD, J. **Variation de la teneur en caféine dans le genre coffea.** Paris, v.11, n°.4.(1975).p.251-264.

CIRILO, M P G. **Influência da adubação potássica e da torra nos teores de aminos bioativas em café.** Belo Horizonte: UFMG, Faculdade de Farmácia. (2001). p.100:il (Dissertação de Mestrado em Ciências de Alimentos).

CHALFON, S.M. **Glossário de Termos Utilizados na Cafeicultura**. 1 ed. Lavras: EPAMIG-CTSM, 2008. 305p

DINIZ, P. H. G. D. **Novas estratégias para classificação simultânea do tipo e origem geográfica de chás** - Universidade Federal da Paraíba. (2013). p.148 : il (Tese de Doutorado em Química Analítica).

FERRER, D. AGUADO, S. VIDAL-PUIG, J. PRATS, M. ZARZO, **PLS: a versatile tool for industrial process improvement and optimization**, Appl. Stochastic Models Bus v. 24 (2008).p.551–567.

FRANCA, A. S; OLIVEIRA, L. S; MENDINÇA, J. C. F; SILVA, X. A. **Composition of green and roasted coffee beans**. Food Chemistry v.90.(2005).p.89-94.

FRANCA, A. S; OLIVEIRA, L. S; MENDINÇA, J. C. F; SILVA, X. A. **Physical and chemical attributes of defective crude and roasted coffee beans**. Food Chemistry, v.90.(2005).p.89-94.

FUJIOKA, K.; SHIBAMOTO, T. **Chlorogenic acid and caffeine contents in various commercial brewed coffees**. Food Chemistry. v.106 .(2008).p. 217-221.

GODINHO, M. DA S.. **Determinação da tensão interfacial de óleos isolantes usando Análise de Imagens e Espectroscopia no infravermelho próximo**. Unidade Universitária de Ciências Exatas e Tecnológicas. Universidade Estadual de Goiás.(2009).p.96: il (Dissertação de Mestrado em Ciências Moleculares).

GONZALES, R.C; WOODS, R.E. **Processamento de imagens digitais**. São Paulo: Ed. Blucher, (2000).

HAIR, JR. J.F. **Análise Multivariada de dados**, tradução da 5ª ed, São Paulo, Editora Bookam, 593, (2005).

INTERNATIONAL COFFEE ORGANIZATION. Disponível em: http://www.ico.org/improving_quality.asp. Acessado em outubro de 2016.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION; ISO 11292 **Instant coffee -Determination of free and total carbohydrate contents Method using high performance anion-exchange chromatography**, (1995).

KOGAN, L. C.; MAYNARD, W. E.; **Determination of caffeine and trigonelline in coffee by paper chromatography**. Analytical Chemistry.v. 25.(1953).p.1118 -1120.

KENNARD, R.W.; STONE, L.A. **Computer aided desing of experiments**. Technometrics, v. 11, n. 1.(1969).p. 137-148.

LISKA, G. R.; **Classificação de dados em modelos com resposta binária via algoritimo boosting e regressão logística Universidade Federal de Lavras, M G: Programa de Pós Graduação em Estatística e Experimentação Agropecuária - MG.(2012).p.148 : il (Dissertação de Mestrado).**

LOPEZ-MARTINEZ, L.; LOPEZ-DE-ALBA, P. L., GARCIA C. R.; DE LEON-RODRÍGUEZ, L. M. **Simultaneous determination of methylxanthines in coffees and teas by UV-Vis spectrophotometry and parcial least squares**. Analytica Chimica Acta. v.493, (2003).p.83-94.

LYRA, W. S.. **Espectrometria de emissão em chama baseada em imagens digitais**. Universidade Federal da Paraíba- UFPB- Departamento de Química - Pós Graduação. João Pessoa, Paraíba, PB. (2009). p.120: il (Dissertação de mestrado em Química Analítica).

MASSART, D. L; VANDEGINSTE, B.C. M; DEMING, S. N. **Chemometrics: a Textbook** Elsevier, Amsterdam, (1988).

MENDONÇA, F. C. J. N.; **Caracterização física e química por espectrometria de massas com ionização por electrospray dos grãos defeituosos e sadios de cafés**. Programa de Pós Graduação em Ciências de Alimentos, Faculdae de Farmácia da Universidade Federal de Minas Gerais. M.G. (2008). p.160: il (Tese de Doutorado em Ciências de Alimentos).

MINGOTI, A.S.; **Análise de Dados Através de Métodos de Estatística Multivariada. Uma abordagem aplicada**, v.2, 2ª ed, p. 200- Editora UFMG, LTDA. (2013).

MOREIRA, A. S. P.; NUNES, F. M.; DOMINGUES, M. R.; COIMBRA, M. A. **Coffee melanoidins: structures, mechanisms of formation and potential health impacts**. Food & Function. (2012).p.903-915.

MOREIRA, I., SCHEEL, G. L., HATUMURA, P. H., & SCARMINIO, I. S. **Solvent effects on extraction of chlorogenic acids, caffeine and trigonelline in coffea arabica**. Química Nova, v.37. (2014).p.39-43.

MURRAY, J. M.; DELAHUNTY, C. M.; BAXTER, I. A. **Descriptive, sensory analysis: past, present and future** Food Research International v.34.(2001).p.461- 471.

NEBESNY, E, BUDRYN, G. **Evaluation. of sensory attributes of coffee brews from robusta coffee roasted under different conditions**. European Food Research and Technology.v.224.(2006).p.159-165.

OYAMA, P. I. DE C. **Método para classificação de cafés em grãos por imagens digitais pelo uso de atributos selecionados morfologia, cor e textura**. Programa de Pós Graduação em Engenharia Elétrica. São Carlos, São Paulo, SP. (2014). p.110 : il (Dissertação de Mestrado).

PARREIRA, T. F. **Utilização de métodos quimimométricos em dados de natureza multivariada Campinas**; São Paulo: UNICAMP. (2003). p.120: il (Dissertação de Mestrado em Química).

PAULI, E. D.. **Desenvolvimento de metodologia para determinação de adulterantes em café empregando HPAEC-PAD e quimiometria**. Instituto de Química - Centro de Ciências Exatas, Universidade Estadual de Londrina, L.O. Parana, PR. (2010). p.118: il (Dissertação de Mestrado em Química dos Recursos Naturais).

PAULO, J. M.. **Determinação de parâmetros de qualidade de gasolinas automotivas utilizando a espectroscopia de emissão por chama e métodos quimiométricos.** Universidade Federal de Minas Gerais, MG. (2013). p.110: il (Tese de Doutorado em Ciências).

PEREIRA R. G. F. A ; GRANDA, P.P.P F; OLIVEIRA, E.M ABRAÃO, F.R; LIM A.R; ANGELIS, M.C. **Conteúdo de compostos bioativos de resíduos de beneficiamento do café.** In 40º Congresso Brasileiro de Pesquisas Cafeeiras. (2014).p.102 . - Serra Negra S P.

PEREIRA, R G. F.A **.Efeito da inclusão de grãos defeituosos na comparação Química e qualidade do café (coffea arábica L) Estritamente Mole”.** Lavras UFLA. (1997). p. 96 :il (Tese de Doutorado em Ciências dos Alimentos).

PIMENTA, C.J. **Qualidade do café (coffea arábica L) originado de frutos colhidos de quatro estádios de maturação.** Lavras; UFLA. (1985).p.150 : il (Dissertação de Mestrado em Ciências dos Alimentos).

PONTES, M.J.C. **Algoritmo das Projeções Sucessivas para Seleção de Variáveis Espectrais em Problemas de Classificação.** João Pessoa, Programa de Pós-Graduação em Química, UFPB. (2009). p. 123: il (Tese de Doutorado).

SOARES,S.F.C ; PONTES, M. J. C; GALVÃO,R.K.H; ARAÚJO, M.C.U. **A Guinding the Seçection of Variables by the Sucessive Projections Algorithm is Classification Problems.** J. Braz. Chem. Soc. Vol 25.(2014).p.176-181.

REIS, N. **Detecção de Adulteração em café torrado e moído com cascas de café e milho por espectroscopia no infravermelho.** Programa de Pós Graduação em Ciências de Alimentos, Faculdade de Farmácia da Universidade Federal de Minas Gerais. (2012). p.124 : il (Dissertação de Mestrado em Ciências de Alimentos).

RIBEIRO, J. S. **Previsão de diferentes atributos sensoriais ligados à qualidade do café arábica brasileiro utilizando-se métodos analíticos distintos e ferramentas quimiométricas.** Campinas, S P: UNICAMP. (2009). p.264: il (Tese de Doutorado em Química Analítica).

SANO, E. E.; ASSAD, E. D.; CUNHA, S. A. R.; CORREA, T. B. S.; RODRIGUES, H. R. **Quantifying adulteration in roast coffee powders by digital image processing.** Journal of Food Quality. v.26. (2003).p.123-134.

SILVA, O. C. T. DE I. DE C. **Classificação de cervejas por análise de imagens e PCA.** Programa de Pós Graduação Química - Universidade Federal de Goiás. Goiás, G.O. (2014). p. 81: il (Dissertação de Mestrado).

SILVERSTEIN, R.M. **Identificação Espectrométrica de Compostos Orgânicos.** Rio de Janeiro. Editora Guanabara Koogan S.A-1994, v.4, 5ª edição, p.387. 1994.

SKOOG, D. A.; HOLLER F. J.; NIEMAN T. A. **Princípios de Análise Instrumental,** tradução da 5ª ed, Porto Alegre, Editora Bookman, p.836. 2002.

SOARES, S.C.F. **Um novo critério para seleção de variáveis usando o Algoritmo das Projeções Sucessivas.** Universidade Federal da Paraíba-UFPB. (2010). p.91: il (Dissertação de Mestrado em Química Analítica).

SOUTO, U. T. C .P ; PONTES, M. J. C.; SILVA, E. C.; GALVÃO, R. K. H.; ARAUJO, M. C. U.; SANCHES, F. A. C.; CUNHA, F. A. DA S.; OLIVEIRA, M. DO S. R. **UV Vis Spectrometric Classification of Coffees by SPALDA.** Food .Chemistry. v. 119. (2005). p.368-371.

TAVARES, K. M. **Determinação de adulteração em café riado torrado e moído por espectroscopia no infravermelho e análise sensorial.** Universidade Federal de Lavras- Minas Gerais, MG. (2012),p.109: il (Dissertação de Mestrado em Ciências de Alimentos).

TAVARES, K. M.; PEREIRA, R. G. F. A.; NUNES, C. A.; PINHEIRO, A. C. M.; RODARTE, M. P. GUERREIRO, M. C. **Mid-infrared spectroscopy and chemometrics were used to identify adulteration in roasted and ground coffee by addition of coffee husks.** Quimica Nova, v.35.(2012).p.1164-1168.

TOLEDO, B. R.; HANTAO, L. W.; HO, T. D.; AUGUSTO, F.; ANDERSON, J. L; **A adulteration by solid-phase microextraction using polymeric ionic liquid chemometric sorbent approach toward the detection and quantification of coffee coatings.** Journal of Chromatography A. v. 1346.(2014).p.1-7.

VAN DER VOSSEN, H. A. M. **The cup quality of disease-resistant cultivars of arabica coffee (*Coffea arabica*).** Experimental Agriculture, v. 45. (2009).p. 323-332.

VITORINO, M. D ; FRANÇA, A. S.; OLIVEIRA, L. S.; BORGES, M. L. A. **Metodologias de obtenção de extrato de café visando a dosagem de compostos não voláteis.** Revista Brasileira de Armazenamento, v.3.(2001). p.17-24.

Anêxo

Metodologia para a Análise de impurezas cascas e paus e materias estranhas

A seguir são apresentadas as etapas da análise de referência para a identificação de impurezas e matérias estranhas:

Análise de impurezas de cascas e paus

1. Homogeneização da amostra por quarteamo

1.1 Colocar a amostra em um tabuleiro, espalhar e misturar bem com a espátula



Figura2.6-Fotografia da homogeneização da amostra por quarteamo. Atlas de Microscopia do café torrado, 2010.

1.2 Homogeneizar dividindo a amostra em 4 partes



Figura2.7 - Fotografia da homogeneização da amostra por quarteamo. Atlas de Microscopia do café torrado, 2010.

1.3- Misturar os quadrantes opostos. Repetir o quarteamo e misturar os quadrantes novamente

1.4 Espalhar uniformemente a amostra no tabuleiro



Figura2.8 - Fotografia do espalhamento da amostra no tabuleiro. Atlas de Microscopia do café torrado, 2010.

2. Desengorduramento da amostra

2.1 Retirar porções aleatórias da amostra e pesar em béquer 2,0 gramas, em balança analítica



Figura2.9 - Fotografia Retirar porções da amostra. Atlas de Microscopia do café torrado 2010.



Figura2.10 - Fotografia da pesagem da amostra. Atlas de Microscopia do café torrado, 2010.

2.1.1 Em um cálice cônico colocar 60 ml de clorofórmio



Figura2.11 - Fotografia da Adição de clorofórmio na amostra. Atlas de Microscopia do café torrado, 2010.

2.2 Espalhar levemente a amostra sobre a superfície do clorofórmio sem deixar romper a tensão superficial.



Figura2.12- Fotografia do espalhamento da amostra sobre o clorofórmio. Atlas de Microscopia do café torrado, 2010.

2.3 - Revolver, vagorosamente, a camada do pó de café com um bastão de vidro e observar se há precipitação de sedimento.



Figura2.13 - Fotografia do Revolvimento da camada de pó de café. Atlas de Microscopia do café torrado, 2010.

2.4 Agitar com um bastão de vidro o conteúdo do cálice e deixar em repouso por 20 minutos



Figura 2.14 - Fotografia da -Agitação da amostra e repouso. Atlas de Microscopia do café torrado, 2010.

2.5 -Filtrar em papel de filtro qualitativo recolhendo o filtrado em erlenmeyer de 250 ml



Figura2.15 - Fotografia da Filtragem da amostra em um erlemmeyer. Atlas de Microscopia do café torrado , 2010.

2.6 Deixar o resíduo secar em capela de exaustão



Figura2.16 - Fotografia da Transferência do resíduo da amostra. Atlas de Microscopia do café torrado, 2010.

2.9 Tamizar com auxílio de pincel nº 16, até que não seja perceptível a passagem do pó ao bater o tamis sobre uma folha de papel branca



Figura2.17 - Fotografia da Transferência do resíduo da amostra. Atlas de Microscopia do café torrado, 2010.



Figura2.18 - Fotografia da Tamização do resíduo da amostra. Atlas de Microscopia do café torrado, 2010.

2.10 Colocar em placa de Petri o resíduo retido no tamis



Figura2.19 - Fotografia da Transferência do resíduo para o petri. Atlas de Microscopia do café torrado 2010.

2.11 Avaliar a amostra ao estereomicroscópio quanto à presença de matérias estranhas, quantificar as impurezas (cascas e paus), se necessário, ou finalizar o ensaio



Figura2.20- Fotografia da Avaliação da amostra ao estereomicroscópio. Atlas de Microscopia do café torrado, 2010.

3 Determinação quantitativa de cascas e paus

3.1 Pesar todo o resíduo retido no tamis (observado ao estereomicroscópio), em balança analítica usando placa de Petri e anotar o valor (**R**)



Figura2.21 -Fotografia da Pesagem do resíduo retido no tamis . Atlas de Microscopia do café torrado,2010.

3.2 Deste resíduo (**R**), pesar 0,1g ou 0,2g em balança analítica usando outra placa de Petri e anotar o valor (**P**)

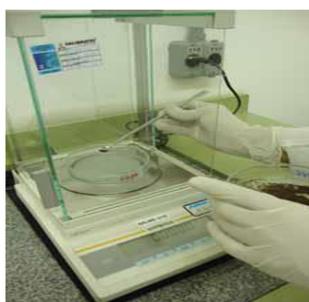


Figura2.22-Fotografia da Pesagem de uma parte do resíduo retido no tamis. Atlas de Microscopia do café torrado, 2010.

3.3 Colocar uma gota de água destilada sobre uma lâmina de vidro



Figura2.23 - Fotografia da Adição de uma gota de água sobre a lâmina de vidro . Atlas de Microscopia do café torrado, 2010.

3.4 Com o auxílio de um estilete umedecido e estereomicroscópio, separar todas as cascas e paus do resíduo transferindo- as para a lâmina com água

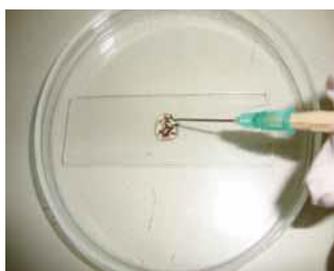


Figura2.24 - Fotografia da Separação das cascas e paus do resíduo e transferência para a lâmina. Atlas da Microscopia do café torrado.

3.5 Transferir as cascas e paus da lâmina para um pesa filtro com tampa, previamente aquecido por 1 hora em estufa a 105°C ($\pm 5^{\circ}\text{C}$), esfriado em dessecador e pesado.



Figura2.25 - Fotografia da Transferência das cascas e paus da lâmina para um pesa filtro. Atlas da Microscopia do café torrado, 2010.

3.6 Levar o pesa-filtro semiaberto à estufa a 105°C ($\pm 5^{\circ}\text{C}$) por uma hora



Figura2.26 - Fotografia da Estufa para colocar o pesa filtro. Atlas da Microscopia do café torrado, 2010.

3.7- Retirar o pesa filtro tampado com auxílio de pinça adequada, esfriar em dessecador por 30 minutos e pesar.



Figura2.27 - Fotografia do Pesa filtro colocado no dessecador por 30 minutos. Atlas da Microscopia do café torrado, 2010.

3.8 Repetir as pesagens até peso constante (**Pfc**) e proceder conforme cálculo abaixo



Figura2.28- Fotografia da Pesagem do pesa filtro até peso constante. Atlas da Microscopia do café torrado, 2010.

4. Cálculo do percentual de cascas e paus determinado pelas equações (a) e (b)

$$P_{tc} = P_{fc} - P_{fv} \quad (a)$$

$$P_{tc} \times R \times 50 = x \% \text{ cascas e paus} / P \quad (b)$$

onde:

P_{fc} = Peso do pesa filtro + cascas e paus

P_{fv} = Peso do pesa filtro vazio

P_{tc} = Peso total de cascas e paus

P = Peso do resíduo para catação (0,1 a 0,2g)

R = Peso do resíduo retido no tamis

Método de identificação histológica para a Análise das matérias estranhas

Trata-se de um método que detecta detritos vegetais não oriundos do cafeeiro, grãos e sementes de outras espécies. Essas matérias estranhas se apresentam moídas e geralmente torradas junto ao café.

Portanto a identificação do material ao microscópico deve considerar as alterações ocorridas nos tecidos vegetais após tratamento térmico e químico.

A seguir segue a descrição sucinta para a identificação de matéria estranhas: Com o auxílio do estereomicroscópio, estilete e ou pinças, retirar amostras do material estranho da placa contendo o resíduo de café (R)

- Colocá-lo em vidro de relógio com hipoclorito de sódio 10% ou hidróxido de sódio 3%, aquecido ou frio;
- Deixar em contato até o clareamento do material;
- Lavar as amostras com água destilada e montar lâminas desse material utilizando água destilada ou água glicerinada 2%;
- Se necessário, preparar lâminas utilizando lugol ou utilizar luz polarizada para auxiliar na identificação de amidos;
- Analisar as lâminas em microscópico e identificar os elementos histológicos, pesquisar e identificar as matérias estranhas ou as fraudes;
- Utilizar para comparação atlas, material de referência, banco de imagens, bibliografias e coleções fotográficas.

Registro dos resultados

- Especificar todos os elementos histológicos e amidos característicos dos vegetais identificados, utilizando o nome comum seguido do nome científico, entre parênteses, de acordo com as normas internacionais;
- Expressar também no resultado os elementos histológicos não característicos do produto e outras estruturas não identificadas;
- Se possível, realizar a captura de imagens para acompanhar o resultado da análise.

Apêndice

Screening for Coffee Adulteration Using Digital Images and SPA-LDA

Urijatan Teixeira de Carvalho Polari Souto · Mayara Ferreira Barbosa · Heberthy Vieira Dantas · Aline Santos de Pontes · Wellington da Silva Lyra · Paulo Henrique Gonçalves Dias Diniz · Mário César Ugulino de Araújo · Edvan Cirino da Silva

Received: 24 July 2014 / Accepted: 7 October 2014 / Published online: 8 November 2014
© Springer Science+Business Media New York 2014

Abstract In this paper, we propose a new methodology to identify adulterations in ground roasted coffees (due to the presence of husks and sticks) using digital images and the successive projections algorithm for variable selection in association with linear discriminant analysis (SPA-LDA). A simple document scanner was used for capturing the images, and a Petri dish support with eight circular holes (one for each sample) to be scanned was employed. Color histograms in the hue-luminosity-saturation (HLS) channels extracted from the digital images were used as input data and statistically evaluated using supervised pattern recognition techniques. For comparison with SPA-LDA, soft independent modeling by class analogy (SIMCA) and partial least squares discriminant analysis (PLS-DA) were also used. In general, SPA-LDA provided significantly better performance than the other classification models, reaching a mean accuracy of 92.5 % for both the training and test sets, while SIMCA and PLS-DA attained only 71.5 and 85.5 %, respectively. More specifically, all of the models presented high rates (above 90 %) for sensitivity and specificity (in the test set samples classification), except SIMCA, which presented a specificity rate of 76 %. Moreover, the SPA-LDA model generally showed the lowest classification error rates. As such, it is a more adequate chemometric tool for discriminating pure coffee samples and adulterated by husks and sticks. The proposed strategy

avoided laborious sample preparation, and additional operational costs, assessing coffee adulteration by husks and sticks.

Keywords Coffee · Adulteration · Digital images · Color histograms · Successive projections algorithm

Introduction

Coffee is one of the most widely traded agricultural commodities in the world. Brazil is the largest coffee-producing country and responded for approximately 34 % of the total world production in the crop year of 2012/2013 (ICO 2014). Both Arabica (*Coffea arabica*) and Robusta (*Coffea canephora*) are cultivated in Brazil, depending on the local climate, soil quality, and altitude of the coffee-growing region. Because of high global production, export, and consumption, there is a great need for the national coffee industry to ensure competitiveness, by maintaining the quality standards of the beans. Recently, authentication is a major topic of interest for industry, regulatory agencies, and consumers alike, with certification being a common practice that ensures quality and enhances market value (Toledo et al. 2014; Souto et al. 2010).

For these reasons, the Brazilian Coffee Industry Association (ABIC) in 2004 created the Coffee Quality Program (PQC), a certification and food safety program for ground roasted coffees. This has provided the product with the highest standard of quality by means of good manufacturing practices, the maintenance of the taste over time, and fair price assurance (ABIC 2014). In 2010, the Ministry of Agriculture, Livestock, and Food Supply (MAPA) issued the Normative Instruction No. 16, regulating ground roasted coffee quality standards, where a maximum limit of 1 % (w/w) for a set of impurities (husks and sticks), sediments (stones, clods and sand), and foreign matter (corn, rye, sugar, barley, açai seeds,

U. T. d. C. P. Souto · M. F. Barbosa · H. V. Dantas · A. S. de Pontes · W. d. S. Lyra · M. C. U. de Araújo · E. C. da Silva (✉)
Departamento de Química, Universidade Federal da Paraíba,
P.O. Box 5093, 58051-970 João Pessoa, PB, Brazil
e-mail: edvan@quimica.ufpb.br

P. H. G. D. Diniz
Centro de Ciência e Tecnologia, Departamento de Química,
Universidade Estadual da Paraíba, Campina Grande, PB, Brazil



Contents lists available at ScienceDirect

LWT - Food Science and Technology

journal homepage: www.elsevier.com/locate/lwt

Identification of adulteration in ground roasted coffees using UV–Vis spectroscopy and SPA-LDA



Urijatan Teixeira de Carvalho Polari Souto ^a, Mayara Ferreira Barbosa ^a,
 Hebertty Vieira Dantas ^a, Aline Santos de Pontes ^a, Wellington da Silva Lyra ^a,
 Paulo Henrique Gonçalves Dias Diniz ^b, Mário César Ugulino de Araújo ^a,
 Edvan Cirino da Silva ^{a,*}

^a Universidade Federal da Paraíba, Departamento de Química, P.O. Box 5093, Zip Code 58051-970, João Pessoa, PB, Brazil

^b Universidade Estadual da Paraíba, Centro de Ciência e Tecnologia, Departamento de Química, Zip Code 58.429-500, Campina Grande, PB, Brazil

ARTICLE INFO

Article history:

Received 12 December 2014

Received in revised form

25 March 2015

Accepted 2 April 2015

Available online 14 April 2015

Keywords:

Ground roasted coffee

Adulteration

UV–Vis spectroscopy

Successive Projections Algorithm

Linear Discriminant Analysis

ABSTRACT

The pleasant taste and flavor of coffee are responsible for making it one of the most preferred and consumed beverages around the world. Industrial coffee plants usually roast both coffee and adulterants together, which affects directly the quality of the product, specially the sensorial characteristics of the drink. Therefore, this work aims the identification of adulterations in ground roasted coffees (due to the presence of husks and sticks) using ultraviolet–visible spectroscopy and the Successive Projections Algorithm for variable selection in association with Linear Discriminant Analysis (SPA-LDA). For this purpose, extracts of ground roasted coffees prepared in hot water alone (representing the final product as ingested by the consumers) were then analyzed. The proposed methodology obtained a 100% classification in both training and test sets, providing a simple and fast analysis of the aqueous extracts of ground roasted coffees. Moreover, it can also afford security to coffee consumers and regulatory agencies, preventing the fraudulent labeling.

© 2015 Published by Elsevier Ltd.

1. Introduction

Coffee is one of the most preferred and consumed beverages around the world due to its pleasant taste and flavor. In order to stimulate the growth in demand for coffee through the establishment of a better overall standard of quality to the global market, the International Coffee Organization (ICO) implemented in 2002 the Coffee Quality Improvement Programme (CQP), subsequently modified by Resolution No. 420 of May 2004. This Programme consists of target standards for exportable coffee, restricting exports of coffee below a minimum accepted quality level and diverting low-grade coffee to alternative uses. This improvement of coffee quality and the consequent increase in its demand took the coffee to be one of the world's top commodities, being Brazil the largest coffee-producing country (more than a third of the total world production in the crop year of 2012/2013 (Nebesny & Budryn, 2006; Van der Vossen, 2009, Souto et al., 2014; ICO, 2014)).

As the largest exporting member of ICO, the Brazilian Coffee Industry Association (ABIC) also implemented the CQP, aiming supply the highest standard of quality to its ground roasted coffees. Moreover, in 2010 the Brazilian Ministry of Agriculture, Livestock, and Food Supply (MAPA) issued the Normative Instruction No. 16, establishing a maximum limit of 1 g/100 g for a set of impurities (husks and sticks), sediments (stones, clods and sand), and foreign matter (corn, rye, sugar, barley, *açaí* seeds, etc.) in ground roasted coffees in order to avoid an illegal practice of adulteration by coffee producers (MAPA, 2010).

Nevertheless, there is an intentional or unintentional character of defrauding the ground roasted coffee by some sectors of the coffee production chain, encompassing some cooperatives and industries roasting. Industrial coffee plants usually roast both coffee and adulterants together, in increasing degrees of roasting levels (whether light, medium, medium-dark, or dark), favoring the adhesion of fine roasted coffee particles onto the surface of the adulterants due to oil absorption. Thus, the presence of these impurities affects directly the quality of the ground roasted coffee, specially the sensorial characteristics of the drink (Sano, Assad,

* Corresponding author. Tel./fax: +55 83 3216 7438.
 E-mail address: edvan@quimica.ufpb.br (E.C. da Silva).