



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



TESE DE DOUTORADO

**Novas estratégias para classificação
simultânea do tipo e origem geográfica de chás**

Paulo Henrique Gonçalves Dias Diniz

João Pessoa – PB - Brasil
Julho/2013



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

TESE DE DOUTORADO

Novas estratégias para classificação simultânea do tipo e origem geográfica de chás

Paulo Henrique Gonçalves Dias Diniz*

Tese de Doutorado 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. Mário César Ugulino de Araújo**

2ª Orientador: **Prof. Dr. Marcelo Fabián Pistonesi**

* Bolsista (Capes)

João Pessoa – PB - Brasil
Julho/2013

D585n Diniz, Paulo Henrique Gonçalves Dias.

Novas estratégias para classificação simultânea do tipo e origem geográfica de chás / Paulo Henrique Gonçalves Dias Diniz. – João Pessoa, 2013.

129. : il.

Orientadores: Mário César Ugulino de Araújo, Marcelo Fabián Pistonesi

Tese (Doutorado) – UFPB / CCEN

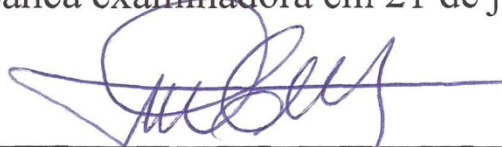
1. Química Analítica. 2. *Camellia sinensis*. 3. Classificação. 4. Imagens digitais. 5. Espectroscopia NIR. 6. Composição química. 7. Quimiometria.

UFPB/BC

CDU: 543(043)

Tese de Doutorado submetida ao Programa de Pós-Graduação em Química do Departamento de Química do Centro de Ciências Exatas e da Natureza da Universidade Federal da Paraíba como parte dos requisitos para a obtenção do grau de Doutor em Química, área de concentração Química Analítica.

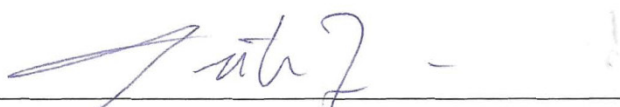
Tese de Doutorado de Paulo Henrique Gonçalves Dias Diniz aprovada pela banca examinadora em 21 de junho de 2013:



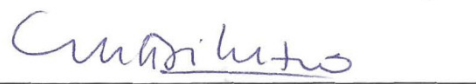
Prof. Dr. Mario César Ugulino de Araújo
Orientador/Presidente



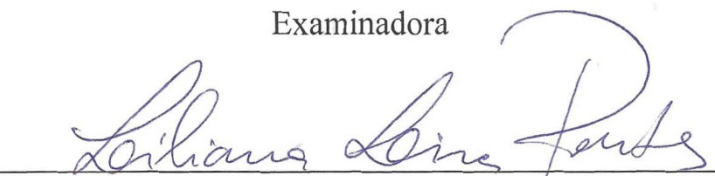
Prof. Dr. Marcelo Fabián Pistonesi
2º Orientador



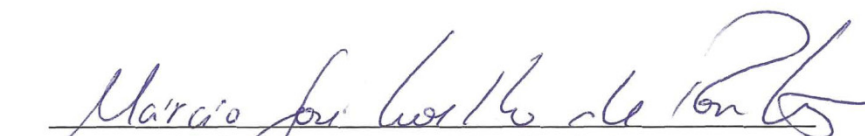
Profa. Dra. Maria Eugenia Centurión
Examinadora



Profa. Dra. Maria Susana Di Nezio
Examinadora



Profa. Dra. Liliansa de Fátima Bezerra Lira de Pontes
Examinadora



Prof. Dr. Márcio José Coelho de Pontes
Examinador

"É muito melhor arriscar coisas grandiosas, alcançar triunfos e glórias, mesmo expondo-se à derrota, do que formar fila com os pobres de espírito que nem gozam muito nem sofrem muito, porque vivem nessa penumbra cinzenta que não conhece vitória nem derrota."

- Theodore Roosevelt -

A Deus Todo-Poderoso pela existência das pessoas que amo.

A minha amada mãe, Miriam, pelo amor incondicional.

A todas as pessoas que me amam verdadeiramente.

Com amor e gratidão,

Dedico.

Ao amigo Prof. Dr. Marcelo Fabián Pistonesi,

Com admiração e gratidão,

Dedico.

AGRADECIMENTOS

- A Deus, que sempre tem me abençoado, fortalecido e presenteado com dádivas e vitórias sucessivas.
- Aos meus amados pais, Mirian Gonçalves e Sebastião Diniz, pelo esforço, empenho, ensinamentos, apoio e incentivo em tudo o que faço.
- A minha vó Maria da Penha e minhas tias Isabel Cristina e Marina Lopes pelos ensinamentos, apoio e incentivo.
- A Adriano Araújo, Anja Sittner, Bruno Brandão, Edilene Dantas, Daniel Jackson, Daniel Sarmiento, Licarion Neto, Sara Regina, Thiago Barros e Wellington Lyra pela amizade, sinceridade, paciência, apoio e participação nos momentos importantes.
- Ao Prof. Dr. Mário César Ugulino de Araújo pela orientação, confiança e ensinamentos.
- Ao amigo Prof. Dr. Marcelo Fabián Pistonesi pela oportunidade de aprendizagem, ensinamentos e sugestões acadêmicas e pessoais desde a época de voluntário no LAQA.
- Às professoras Dr.^a Maria Centurión e Dr.^a Susana Di Nezio pelas relevantes contribuições acadêmico-científicas para esta Tese.
- A Hebertty Vieira, Mayara Barbosa, Karla Daniele, Elaine Cristina, Eugenio Campaña e Mónica Alvarez pela colaboração na parte experimental.
- A Prof.^a Dr.^a Beatriz Susana Fernández Band pela oportunidade de realização do Doutorado-Sanduiche na Universidad Nacional del Sur, Argentina.
- Ao amigo Gastón Pierini pela amizade, viagens e bons momentos vividos na Argentina.
- A todos os amigos que ganhei no LAQA/UFPB e na UNS.
- A Capes pelas bolsas concedidas.

SUMÁRIO

LISTA DE FIGURAS	x
LISTA DE TABELAS	xiii
LISTA DE ABREVIATURAS.....	xv
RESUMO.....	xvii
ABSTRACT.....	xvii
CAPÍTULO 1	
1. Introdução	1
1.1 O chá	2
1.1.1 Tipos de chá	4
1.1.1.1 Chá preto.....	5
1.1.1.1.1 Colheita	5
1.1.1.1.2 Murchamento.....	5
1.1.1.1.3 Enrolamento do chá convencional.....	6
1.1.1.1.4 Chás não murchados	7
1.1.1.1.5 Fermentação	7
1.1.1.1.6 Formação de aroma	8
1.1.1.1.7 Torra.....	9
1.1.1.1.8 Graduação e armazenamento	10
1.1.1.1.8 Nova tecnologia de fabricação	10
1.1.1.2 Chá oolong.....	11
1.1.1.3 Chá verde.....	11
1.1.2 Composição química	12
1.1.2.1 Pigmentos	13
1.1.2.2 Outros polifenóis	14
1.1.2.3 Teaflavinas.....	14
1.1.2.4 Ácidos teaflávicos.....	15
1.1.2.5 Tearubiginas.....	15
1.2 Justificativa	16

1.3	Objetivo	22
1.2.1	Objetivo geral.....	22
1.2.2	Objetivos específicos.....	23
CAPÍTULO 2		
2.	Fundamentação teórica	24
2.1	Imagens digitais.....	25
2.1.1	Revisão da literatura.....	29
2.2	Espectroscopia NIR.....	33
2.2.1	Revisão da literatura.....	35
2.3	Técnicas de reconhecimento de padrões	38
2.3.1	Procedimentos preliminares	40
2.3.2	Pré-tratamento de dados	41
2.3.3	Análise exploratória dos dados.....	42
2.3.4	Redução e seleção de variáveis	44
2.3.5	Técnicas de reconhecimento de padrões supervisionados	46
2.3.5.1	SIMCA.....	46
2.3.5.2	PLS-DA	47
2.3.5.3	LDA	48
2.3.5.3.1	SPA-LDA	50
CAPÍTULO 3		
3.	Classificação de chás usando imagens digitais e quimiometria	52
3.1	Experimental.....	53
3.1.1	Amostras	53
3.1.2	Aquisição de imagens.....	54
3.1.3	Histogramas de cor e análise de dados.....	56
3.2	Resultados e discussão.....	57
3.2.1	Análise de componentes principais	57
3.2.2	Classificação.....	60
3.2.2.1	SIMCA.....	60
3.2.2.2	PLS-DA	61
3.2.2.3	SPA-LDA.....	62

CAPÍTULO 4	
4. Classificação de chás usando espectroscopia NIR e quimiometria	65
4.1 Experimental.....	66
4.1.1 Amostras	66
4.1.2 Medidas dos espectros NIR.....	66
4.1.3 Procedimento quimiométrico	66
4.2 Resultados e discussão.....	67
4.2.1 Análise de componentes principais	69
4.2.2 Classificação.....	71
4.2.2.1 SIMCA.....	71
4.2.2.2 PLS-DA	72
4.2.2.3 SPA-LDA.....	73
CAPÍTULO 5	
5. Classificação de chás usando composição química e quimiometria.....	75
5.1 Experimental.....	76
5.1.1 Amostras e reagentes.....	76
5.1.2 Procedimentos analíticos.....	76
5.1.2.1 Determenações nas folhas de chá	76
5.1.2.1.1 Umidade	76
5.1.2.1.2 Cinzas totais	77
5.1.2.1.3 Cafeína	77
5.1.2.1.4 Metais totais	78
5.1.2.2 Determenações nas infusões de chá	80
5.1.2.2.1 Preparação das infusões	80
5.1.2.2.2 Fluoreto	80
5.1.2.2.3 Polifenóis totais	81
5.1.2.2.4 Metais	81
5.1.3 Procedimento quimiométrico	81
5.2 Resultados e discussão.....	82
5.2.1 Considerações iniciais	82
5.2.2 Apresentação e avaliação preliminar dos dados	85
5.2.3 Análise de componentes principais	86

5.2.4	Classificação.....	88
5.2.4.1	SIMCA.....	88
5.2.4.2	PLS-DA.....	89
5.2.4.3	SPA-LDA.....	89
CAPÍTULO 6.....		
6.	Conclusão.....	93
6.1	Propostas futuras.....	95
REFERÊNCIAS.....		
	Referências.....	96
ANEXOS.....		
Anexo 1: Tabela de resultados da 3ª aplicação.		
Anexo 2: Artigo publicado no periódico <i>Analytical Methods</i> referente à 1ª aplicação desenvolvida nesta Tese de Doutorado.		
Anexo 3: Artigo publicado no periódico <i>Talanta</i> referente às atividades desenvolvidas no Doutorado-Sanduíche.		
Anexo 4: Lista de referências das imagens utilizadas.		

LISTA DE FIGURAS

Figura 1.1. (a) Plantação de arbustos de <i>Camellia sinensis</i> ; (b) broto e folhas superiores empregados na fabricação dos chás verde, oolong e preto	3
Figura 1.2. Processamento das folhas de chá e produção dos chás verde, oolong e preto	4
Figura 1.3. Composição química das folhas de chá. Os teores mínimo e máximo são mostrados em barras verde e azul, respectivamente	12
Figura 1.4. Os flavonóis que ocorrem no chá	13
Figura 1.5. Formação de quinona.....	14
Figura 1.6. Produção total mundial de chá	17
Figura 1.7. Exemplos de variedades comerciais dos chás (a) verde, (b) oolong e (c) preto.	20
Figura 2.1. Sistema de cor RGB.....	27
Figura 2.2. Sistema de cor HSB	29
Figura 3.1. (a) Exemplo das etapas do quarteio de uma amostra de chá; (b) sucessivos quarteios para redução e homogeneização da amostra.....	54
Figura 3.2. Suporte construído para captura das imagens de chá. (a) Câmera digital, (b) suporte metálico, (c) lâmpada circular, (d) chá verde brasileiro, (e) chá verde argentino, (f) chá preto brasileiro, (g) chá preto argentino e (h) chá preto srilanquense.....	55
Figura 3.3. Exemplo da janela de comandos do programa ImageJ e dos histogramas de cor e imagens de uma amostra de chá verde..	56

- Figura 3.4.** Histograma médio de todas as amostras de chá contendo os modelos de cor avaliados58
- Figura 3.5.** Escores de PCA obtidos a partir dos histogramas de (a) RGB, (b) HSB, (c) escala de cinza e (d) RGB + escala de cinza + HSB para todas as amostras de chás. (▼) Chá verde brasileiro, (▲) chá preto brasileiro, (■) chá verde argentino, (◆) chá preto argentino, (●) chá preto srilanquense.....59
- Figura 3.6.** As 17 variáveis selecionadas (●) pelo SPA para o modelo RGB + escala de cinza + HSB63
- Figura 3.7.** Funções discriminantes obtidas por SPA-LDA na classificação das amostras dos conjuntos de validação e teste usando o modelo de cor RGB + escala de cinza + HSB. (▼) Chá verde brasileiro, (▲) chá preto brasileiro, (■) chá verde argentino, (◆) chá preto argentino, (●) chá preto srilanquense.....64
- Figura 4.1.** Espectros NIR (a) brutos e (b) derivados para as 84 amostras de chá estudadas.....68
- Figura 4.2.** Gráfico de escores de PC1 × PC2 para as 84 amostras de chás estudadas usando os espectros (a) brutos e (b) derivados. (▼) Chá verde brasileiro, (▲) chá preto brasileiro, (■) chá verde argentino, (◆) chá preto argentino, (●) chá preto srilanquense. A variância explicada em cada PC é indicada em parênteses...70
- Figura 4.3.** (a) Determinação do número ótimo de variáveis em SPA-LDA e (b) espectro médio derivado das 84 amostras de chá indicando os números de onda selecionados por SPA-LDA72
- Figura 4.4.** Funções discriminantes obtidas por SPA-LDA na classificação das 84 amostras de (▼) chá verde brasileiro, (▲) chá preto brasileiro, (■) chá verde argentino, (◆) chá preto argentino, (●) chá preto srilanquense.....74

Figura 5.1. Gráficos de escores (a–d) e seus respectivos pesos (a’–d’) para as duas primeiras PCs para os quatro conjuntos de dados estudados: (a) “*Todos*”, (b) “*CAA*”, (c) “*Metais totais*” e (d) “*Metais na infusão*”. Chá preto argentino (■), chá preto srilanquense (●) e chá verde argentino (▲).....87

Figura 5.2. Gráficos de escores das funções discriminantes de Fisher para as amostras de chás estudadas usando as três variáveis (K, Al e Mg) selecionadas por SPA-LDA para “*Metais na infusão*”. Chá preto argentino (■), chá preto srilanquense (●) e chá verde argentino (▲)90

LISTA DE TABELAS

Tabela 3.1. Erros de classificação SIMCA para as amostras de chá do conjunto de teste usando quatro diferentes histogramas de cor (RGB, HSB, Escala de cinza e RGB + escala de cinza + HSB) a um nível de significância de 5% para o teste- <i>F</i>	60
Tabela 3.2. Erros de classificação PLS-DA para as amostras de chá do conjunto de teste usando quatro diferentes histogramas de cor (RGB, HSB, Escala de cinza e RGB + escala de cinza + HSB)	61
Tabela 3.3. Erros de classificação SPA-LDA para as amostras de chá do conjunto de teste usando quatro diferentes histogramas de cor (RGB, HSB, Escala de cinza e RGB + escala de cinza + HSB)	62
Tabela 4.1. Número de amostras de treinamento e teste em cada classe estudada	67
Tabela 4.2. Resultados da classificação SIMCA em quatro níveis de significância para o teste- <i>F</i> (1%, 5%, 10% e 25%) para o conjunto de teste usando espectros NIR das amostras de chá	71
Tabela 4.3. Resultados de classificação PLS-DA para o conjunto de teste usando espectros NIR das amostras de chá	72
Tabela 4.4. Resultados de classificação para o conjunto de teste usando as 6 variáveis selecionadas pelo SPA-LDA nos espectros NIR das amostras de chá.....	74
Tabela 5.1. Comprimentos de onda escolhidos para as determinações dos metais e seus respectivos limites de determinação	79

Tabela 5.2. Erros de classificação para o conjunto de teste usando SIMCA, PLS-DA e SPA-LDA para os quatro conjuntos de dados estudados: “ <i>Todos</i> ”, “ <i>CAA</i> ”, “ <i>Metais totais</i> ” e “ <i>Metais na infusão</i> ”	88
--	----

LISTA DE ABREVIATURAS

ANN	<i>Redes neurais artificiais</i>
BLC	<i>Condicionador de folhas Barбора</i>
BP-ANN	<i>Retropropagação em redes neurais artificiais</i>
CART	<i>Árvore de classificação e regressão</i>
CTC	<i>Esmagamento, rasgo, enrolamento</i>
HSI	<i>Matiz, saturação, intensidade</i>
ICP-OES	<i>Espectroscopia de emissão ótica com plasma acoplado indutivamente</i>
kNN	<i>k-Vizinhos mais próximos</i>
LDA	<i>Análise discriminante linear</i>
LS-SVM	<i>Máquina de vetores de suporte com regressão por mínimos quadrados</i>
MCLS	<i>Regressão por mínimos quadrados parciais multiclases</i>
NIR	<i>Espectroscopia no infravermelho próximo</i>
PCA	<i>Análise por componentes principais</i>
PLS	<i>Regressão por mínimos quadrados parciais</i>
PLS-DA	<i>Análise discriminante por mínimos quadrados parciais</i>
PTFE	<i>Politetrafluoretileno</i>
RGB	<i>Vermelho, verde, azul</i>
SIMCA	<i>Modelagem independente e flexível por analogia de classe</i>
SPA	<i>Algoritmo das projeções sucessivas</i>
SVM	<i>Máquina de vetores de suporte</i>
TCR	<i>Enrolador contínuo Tocklai</i>
WT	<i>Transformada wavelet</i>

RESUMO

Título: Novas estratégias para classificação simultânea do tipo e origem geográfica de chás

O chá tem uma importância econômica e cultural, não só para produtores e consumidores, mas também por um interesse científico. A qualidade organoléptica da infusão da *Camellia sinensis* depende da natureza e da quantidade de vários metabólitos secundários (tais como polifenóis, cafeína, aminoácidos, etc.), os quais podem ser relacionados diretamente com a origem geográfica das plantas. Estes componentes são a base do valor econômico do chá e de seus efeitos benéficos sobre a saúde humana. Por isso, há um crescente interesse dos consumidores por chás de alta qualidade com uma clara identidade geográfica. Durante as últimas décadas, as metodologias analíticas que empregam técnicas instrumentais modernas tornaram-se mais sensíveis, confiáveis e rápidas. Entretanto, tais técnicas têm vantagens e limitações para a aplicação da análise da qualidade do chá e de suas origens geográficas. Assim, uma combinação de diferentes técnicas analíticas pode ser mais útil do que depender de um único método. Seguindo estes preceitos, nós propusemos três novas estratégias para a classificação simultânea de chás de acordo com o tipo (verde e preto) e a origem geográfica (Argentina, Brasil e Sri Lanka). As metodologias propostas empregam o uso de (1) imagens digitais, (2) espectroscopia NIR e (3) composição química (umidade, cinza total, cafeína, polifenóis totais, fluoreto e quinze metais (Na, Mg, Al, P, K, Ca, Cr, Mn, Fe, Co, Ni, Cu, Zn, Cd e Pb) nas folhas e infusões dos chás). Uma classificação correta de todas as amostras de chás (100% de acerto) foi sempre obtida utilizando Análise Discriminante Linear associada à técnica de seleção de variáveis feita pelo Algoritmo das Projeções Sucessivas (SPA-LDA). Modelagem Independente e Flexível por Analogia de Classe (SIMCA) e Análise Discriminante por Mínimos Quadrados Parciais (PLS-DA) também foram utilizadas. Tais estratégias podem ser úteis para a elaboração de normas para o controle de qualidade de chás no Brasil, que ainda é inexistente.

Palavras-chaves: *Camellia sinensis*, Classificação, Imagens digitais, Espectroscopia NIR, Composição química, SIMCA, PLS-DA, SPA-LDA.

ABSTRACT

Title: New strategies for simultaneous classification of both the variety and geographical origin of teas

Tea has an economic and cultural importance, not only for producers and consumers, but also for a scientific interest. The organoleptic quality of the *Camellia sinensis*' infusion depends on the nature and amount of several secondary metabolites (such as polyphenols, caffeine, amino acids, etc.), which can be directly related to the geographical origin of the tea plants. These components are the basis of the economic value of teas and its beneficial effects on human health. Therefore, there is a growing consumer's interest in high quality teas with a distinct geographical identity. In last decades, the analytical methods employing modern instrumental techniques have become more sensitive, reliable and fast. However, these techniques have advantages and limitations for the application in the analyses of the tea quality and their geographic origins. Thus, a combination of different techniques could be more useful than relying on a single method. Following these principles, we propose three new strategies for simultaneous classification of teas according to both the type (green and black) and geographic origin (Argentina, Brazil and Sri Lanka). The proposed methodologies employ the use of (1) digital images, (2) NIR spectroscopy, and (3) chemical composition (moisture, ash, caffeine, total polyphenols, fluoride and fifteen metals (Na, Mg, Al, P, K, Ca, Cr, Mn, Fe, Co, Ni, Cu, Zn, Cd and Pb) in both tea leaves and infusions). A correct classification of all tea samples (100% of correct classification) was always obtained using the Linear Discriminant Analysis associated with the variable selection technique taken by the Successive Projections Algorithm. Soft Independent Modeling of Class Analogy (SIMCA) and Partial Least Squares Discriminant Analysis (PLS-DA) were also used. The proposed strategies might be useful for the development of legislation for the quality control of teas in Brazil, which is still lacking.

Keywords: *Camellia sinensis*, Classification, Digital images, NIR spectroscopy, Chemical composition, SIMCA, PLS-DA, SPA-LDA.

Capítulo 1

Introdução





1. INTRODUÇÃO

1.1. O chá

O chá é uma infusão preparada a partir das folhas vaporizadas e secas do arbusto perene *Camellia sinensis* (SZYMCZYCHA-MADEJA et al., 2012). Num sentido mais amplo, denominam-se chás ervais, frutais e florais as infusões preparadas a partir de folhas, frutos e flores, respectivamente, como é o caso das infusões de boldo, maçã e camomila, por exemplo.

Botanicamente, a planta do chá pertence ao gênero *Camellia*, um membro da família *Theaceae*. Duas variedades básicas do chá são reconhecidas: a variedade chinesa *C. s. sinensis* e a variedade assamesa *C. s. assamica*. A planta é originária dos países do leste asiático (China, Birmânia, Laos e Vietnam) e chegou à Europa no século XVI. O chá foi inicialmente cultivado na província indiana de Assam e gradualmente se espalhou para Indonésia e Sri Lanka. O cultivo em larga escala difundiu-se também para muitos outros países na Ásia, África e América do Sul (NAGALAKSHMI, 2003).

A árvore da *C. sinensis* pode alcançar até 15 metros de altura na natureza. Entretanto, para fins de cultivo, são feitas podas constantes de modo a evitar que a planta não ultrapasse 1,5 metros (Fig. 1.1a). As folhas do chá são escuras, oblongas, lustrosas, com nervuras bem marcadas nas superfícies e de margem inteiramente denteada. Já as folhas mais novas são cobertas de pequenos tricomas brancos. Para a fabricação dos chás são utilizadas apenas as duas folhas e o broto da parte superior da planta (Fig. 1.1b).

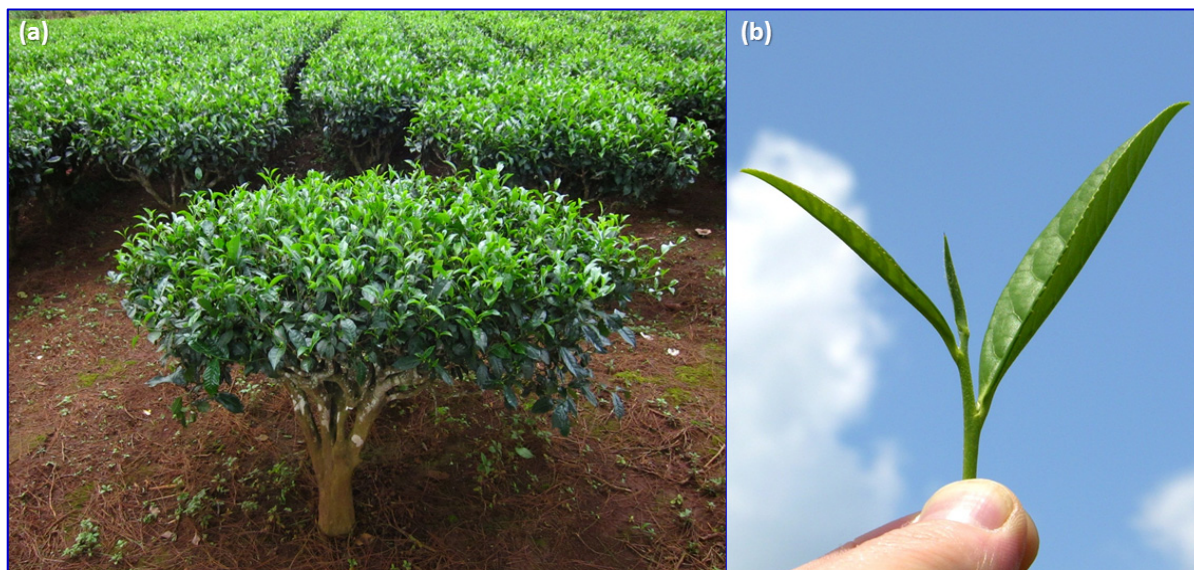


Figura 1.1. (a) Plantação de arbustos de *Camellia sinensis*; (b) broto e folhas superiores empregados na fabricação dos chás verde, oolong e preto.

A história do chá como bebida foi iniciada pelos Chineses por volta de 2700 a.C. na época do Imperador Shen Nung. A primeira menção registrada do chá está em um antigo dicionário chinês, *Erh-Ya*, e data de cerca de 350 a.C. Entretanto, uma pesquisa arqueológica reportada por Jenilek em 1978 sugere que a preparação da infusão de folhas de diferentes plantas selvagens e também do arbusto do chá provavelmente já tenha sido praticada a mais de 500 mil anos (GUTMAN e RYU, 1996).

Desde tempos ancestrais, o chá tem sido bebido com a finalidade de melhorar o fluxo sanguíneo, eliminar toxinas e aumentar a resistência às doenças. Isso foi introduzido progressivamente ao redor do mundo por comerciantes e viajantes. A partir daí, o chá passou a ser associado com o estilo de vida e os hábitos alimentares das pessoas. Evidências científicas sobre os benefícios proporcionados à saúde têm dado ao chá um lugar único entre o consumo de uma simples bebida e uma função social onde o alimento está intimamente associado ao estado de espírito. Benefícios nutricionais e farmacológicos, incluindo propriedades antimicrobiana e antioxidante, também têm sido comprovados. O chá e seus

componentes melhoram não apenas a função cardiovascular, mas também produzem efeitos sobre o cérebro e outros órgãos. Além disso, eles têm efeitos protetores intrínsecos e modulam a bioquímica de vários tipos de células, incluindo os do sistema vascular. Muito interesse também tem sido centrado na função da atividade oxidante/antioxidante no que diz respeito ao processo de envelhecimento e às doenças degenerativas, tais como o câncer, doenças cardiovasculares e diabetes (PINTO, 2013; JAIN et al., 2013; PREEDY, 2012).

1.1.1. Tipos de chá

Mais de 300 tipos de chá são produzidos a partir das folhas da *Camellia sinensis* através de diferentes processos de fabricação. O chá pode ser dividido em seis categorias básicas: branco, amarelo, verde, oolong, preto e pu-erh. Dentre eles, os três principais tipos são diferenciados de acordo com o grau de fermentação e oxidação das folhas: verde (não fermentado), oolong (semifermentado) e preto (fermentado) (SZYMCZYCHA-MADEJA et al., 2012; SANG et al., 2011). Os diferentes passos do processamento destes tipos são sintetizados na **Fig. 1.2**.



Figura 1.2. Processamento das folhas de chá e produção dos chás verde, oolong e preto.



1.1.1.1. Chá preto

O chá preto é feito de folhas jovens e brotos fechados da planta do chá. As principais etapas envolvidas na produção de chá preto são colheita, murchamento, distorção foliar, fermentação, secagem, classificação/graduação, embalagem e armazenamento (NAGALAKSHMI, 2003).

1.1.1.1.1. Colheita (NAGALAKSHMI, 2003)

As folhas frescas de chá verde geralmente são colhidas à mão, em intervalos de 7–14 dias, durante todo o ano. Geralmente, apenas os brotos que crescem rapidamente próximos à segunda ou terceira folha desdobrada são arrancados e usados, conforme pode ser visualizado na **Fig. 1.1b**.

Um longo intervalo de colheita ou o uso de clones que se desviam das folhas-padrão desejáveis resultam em vários parâmetros químicos, tais como a redução dos níveis de teaflavinas, cafeína e compostos voláteis relacionados ao sabor, além de transmitir propriedades sensoriais pobres, como os aromas doce, floral e de grama. A composição química do chá nivelada em percentagem de peso em relação à base seca é a seguinte: 25% de flavonóis, cafeína 4%, aminoácidos 4%, ácidos orgânicos 0,5%, monossacarídeos 4%, polissacarídeos 13%, proteína 15%, celulose 7%, lignina 6%, lipídeos 3%, clorofila e outros pigmentos 0,5%, cinzas 5%, e compostos voláteis 0,1%.

1.1.1.1.2. Murchamento (NAGALAKSHMI, 2003)

A etapa de murchamento faz com que as folhas de chá recém-colhidas suportem certas alterações bioquímicas e fisiológicas que auxiliam nas demais etapas de processamento (enrolamento e fermentação). Estas alterações que



ocorrem na folha viva do chá continuam, mas o murchamento altera o padrão e a extensão dessas mudanças. As seguintes alterações químicas e bioquímicas devidas ao murchamento garantem a qualidade do chá preto: (a) aumento dos níveis de aminoácidos, carboidratos simples, e cafeína; (b) atividade máxima da polifenol oxidase; (c) perda de atividade da pectinase; e (d) quebra da clorofila.

O murchamento é realizado com folhas espalhadas em camadas finas (0,3–0,7 kg/m²) em bandejas de sistemas abertos no andar superior da fábrica de processamento tradicional de chá. O período de murchamento varia de 16 a 20 h, dependendo das condições da folha e da exigência de processamento do chá. O ar quente é soprado a partir do fundo da bandeja do sistema de murchamento. A umidade nas folhas é evaporada pelo ar, o que provoca a secagem. Fábricas localizadas em baixas altitudes usam o ar ambiente e fábricas em alta altitude usam ar quente para a secagem. Observa-se que os períodos de murchamento (12 h) e baixas temperaturas (10–15 °C) resultam em uma boa qualidade do sabor no chá pronto, enquanto que o murchamento mais longo (20–30 h), murchamento forçado e temperaturas elevadas (25–30 °C) têm um efeito positivo sobre a cor, mas um efeito negativo sobre as propriedades químicas e de sabor.

No fim do processo de murchamento, a umidade final é reduzida para 70% e as folhas de chá tornam-se flexíveis o suficiente para serem enroladas. Esta etapa é geralmente omitida para o chá verde, mas é sempre usada para o chá preto.

1.1.1.1.3. Enrolamento do chá convencional (NAGALAKSHMI, 2003)

Após o processo de murchamento, a folha é distorcida por enrolamento ou corte. O processamento convencional das folhas requer enrolamento para produzir chá preto e chá verde. A distorção das folhas imediatamente após a colheita não é



aconselhável. A técnica de enrolamento expulsa o “suco” da folha e a torce. Um enrolador é composto por uma mesa circular, uma caixa ou revestimento cilíndrico e uma tampa para aplicar pressão. A folha é machucada, torcida e então quebrada em pequenos pedaços pelo aumento da pressão e peneirada. O restante do material é enrolado. Cada enrolamento varia de 15 a 60 minutos e, na prática normal, o número de rolos varia de dois a cinco, dependendo do grau de murchamento, do tipo de chá, da carga do rolo, das condições de enrolamento e da temperatura.

1.1.1.1.4. Chás não murchados (NAGALAKSHMI, 2003)

No caso de chás não murchados, máquinas para distorção das folhas são utilizadas isoladamente ou em combinação, tais como *Legg-cut*, CTC (do inglês: *crushing, tearing, curling*) e *Rotorvane*. O objetivo é a maceração intensiva das folhas de chá para assegurar a fermentação rápida e completa. A máquina CTC consiste de dois rolos metálicos cilíndricos funcionando como uma espécie de espremedor, um trabalhando a 70 rpm e o outro a 700 rpm. A folha é então cortada, despedaçada ou enrolada em um pequeno espaço entre as superfícies serrilhadas dos rolos.

Máquinas utilizadas para a distorção da folha incluem o triturador, a máquina de processamento de chá contínuo do Sri Lanka, o enrolador contínuo Tocklai (TCR, do inglês: *Tocklai continuous roller*), o condicionador de folhas Barbora (BLC, do inglês: *Barbora leaf conditioner*) e o enrolador contínuo soviético.

1.1.1.1.5. Fermentação (NAGALAKSHMI, 2003)

A fermentação é o passo mais importante no processamento do chá preto para as alterações químicas e bioquímicas necessárias. O processo inicia-se na



maceração da folha e é permitido continuar sob condições ambientais. A folha verde depois de enrolada e peneirada (no caso do chá convencional) ou a folha macerada (tipo CTC) é espalhada em finas camadas de 5–8 cm de profundidade no chão da fábrica ou em bandejas emprateiradas em uma sala de fermentação. O controle de temperatura e a difusão do ar são facilitados usando-se humidificadores ou ar fresco. O tempo de fermentação varia entre 45 minutos e 3 horas, dependendo da natureza da folha, das técnicas de maceração, da temperatura ambiente e da exigência do tipo de chá. A temperatura varia entre 24–27 °C. Baixas temperaturas (15–25 °C) melhoram o sabor.

No final da fermentação a cor das folhas muda de verde para vermelho acobreado, paralelamente ao desenvolvimento de um aroma característico agradável. O ponto terminal é determinado pela habilidade do fabricante ou por técnicas instrumentais. A fermentação é encerrada pela etapa de torra. A fermentação pode ser avaliada através da medição do teor de teaflavinas e tearubiginas, que são formadas na proporção de 1:10 sob condições ideais de fermentação. Estimar o teor de taninos é outro método usual. Os taninos diminuem durante este período de 20% na folha de chá para 10–12% no chá fermentado. Os desenvolvimentos modernos em tecnologia de fermentação oferecem vantagens como temperaturas ótimas controladas, custo reduzido, menor necessidade de espaço no chão e maior brilho em bebidas de chá.

1.1.1.1.6. Formação de aroma (NAGALAKSHMI, 2003)

As catequinas oxidadas reagem com as moléculas precursoras presentes no chá verde e produzem compostos voláteis no chá preto, que são dependentes principalmente de variáveis genéticas, de cultivo e de fabricação. A oxidação de



aminoácidos, carotenos e lipídeos insaturados produzem compostos formadores do aroma durante o período de fermentação. Os aldeídos de aminoácidos formados a partir da reação de degradação de Strecker (SCHONBERG e MOUBACHER, 1952) também são de importância significativa. Iononas, álcoois terpênicos, aldeídos terpênicos e seus produtos de oxidação, tais como teaspirona e diidroactinidióis, resultam da oxidação dos carotenoides. Mais de 638 compostos formadores do aroma do chá foram identificados. Os mais importantes destes componentes são os terpenos, álcoois terpênicos, lactonas, cetonas, ésteres e compostos espiro. Compostos formadores do aroma de alguns chás de renome mundial com características únicas de sabor foram identificados: o chá Darjeeling da Índia tem linalool, óxidos I e II de linalool, geraniol; o chá Keemun da China contém 2-feniletanol e geraniol; o chá Uva do Sri Lanka contém metilsalicilato, linalool e os óxidos de linalool. As concentrações dos constituintes do aroma e o teor de óleos voláteis são elevados no chá convencionais em comparação com chás tipo CTC. Chás CTC dão bebida fermentada forte e aroma maltado, enquanto chás convencionais produzem uma bebida fermentada leve, com aromas ricos, finos e florais devido a altas concentrações de linalool, óxidos de linalool, metilsalicilato, etanol fenil genaniol, cisjasmona e β -ionona, etc.

1.1.1.1.7. Torra (NAGALAKSHMI, 2003)

A torra é feita num secador especial a uma alta temperatura, entre 90–95 °C, imediatamente após a fermentação. O tempo necessário para a torra é de 20 minutos para reduzir a umidade de 3–4% e inativar a enzima polifenol oxidase, evitando a perda dos polifenóis por oxidação. O ar quente provoca a evaporação, fazendo com que os brotos de chá percam sua coloração vermelho cobreada e marrom e são transformados no chá preto. A torra diminui constituintes do aroma e

aumenta o teor de alguns ácidos carboxílicos alifáticos, o que indica reações de oxidação. A torra também ajuda a prevenir a perda de vitaminas, principalmente no chá verde que possui um maior teor de vitaminas que chás fermentados.

1.1.1.1.8. Graduação e armazenamento (NAGALAKSHMI, 2003)

O chá é frequentemente peneirado para remoção de pedaços de caule e obtenção dos diferentes graus baseados no tamanho de partícula. Mudanças químicas ocorrem durante o armazenamento de produtos finalizados de chá, que perdem o verdor residual em pouco tempo. Em local fresco e protegido de umidade e oxigênio o chá permanece sólido e cheio de sabor por mais de um ano.

1.1.1.1.9. Nova tecnologia de fabricação (NAGALAKSHMI, 2003)

A tecnologia tradicional de fabricação de chá tem algumas desvantagens: (a) os tecidos não danificados das folhas de chá (20–25%) durante o enrolamento ou método CTC, que produz por sua vez uma fração não uniforme de oxidação, resultam em elevadas perdas de polifenóis; (b) perda de 70–80% do óleo essencial na fase de torra, o que enfraquece o aroma; e (c) o envelhecimento rápido do chá, o que resulta na perda da sua alta qualidade. Por isso, o tratamento térmico de chás abaixo da fermentação desejada em uma fábrica ou armazém a temperatura ambiente (até 40°C) é praticada em muitas partes do mundo, particularmente na China e na Índia. O tratamento térmico elimina o odor de grama e o gosto áspero em chás não fermentados, os quais são devido a polifenóis, catequinas e outros constituintes presentes em proporções elevadas. O tratamento térmico resulta em isomerização e epimerização das catequinas, na degradação da clorofila, na síntese de aldeídos e óleos essenciais e na melhoria da qualidade do aroma do chá.

1.1.1.2. Chá oolong

Oolong é um chá chinês tradicional produzido através de um processo único, que inclui secagem ao sol forte e oxidação antes do enrolamento/torção. Muitos chás oolong, especialmente aqueles de boa qualidade, envolvem cultivares de plantas de chá que são utilizadas exclusivamente para variedades específicas. O grau de oxidação pode variar de 8–85% dependendo da variedade e do estilo de produção. O chá oolong é popular principalmente entre os apreciadores de chá do sul da China e expatriados chineses no Sudeste Asiático, como é o processo de preparação de Fujian, conhecida como a cerimônia do chá Gongfu.

O sabor do chá oolong varia muito entre as várias subvariedades. Ele pode ser doce e frutado com aromas de mel, amadeirados e denso com aromas torrados, verde e fresco com aromas buquê, tudo dependendo da horticultura e estilo de produção. Diversas subvariedades de chá oolong, incluindo os produzidos nas Montanhas Wuyi do norte de Fujian, como Da Hong Pao, estão entre os mais famosos chás chineses. As variedades de chá oolong são processadas de forma diferente, mas as folhas são formadas por um dos dois estilos distintos. Algumas são enroladas em folhas curvas longas, enquanto outros são enroladas na forma de pequenas pérolas, cada uma com uma cauda. O primeiro modelo é o mais tradicional dos dois (NAGALAKSHMI, 2003).

1.1.1.3. Chá verde

Aproximadamente 21% da produção total de chá é consumida como chá verde, que contém grandes quantidades de catequinas e vitaminas. O chá verde possui sabor agradável, aroma floral e coloração verde clara com um tom verde-oliva. O chá verde é produzido a partir de um botão e duas a três folhas da parte superior dos arbustos de chá. Uma vez que o chá verde não é fermentado, o

desenvolvimento do processo de oxidação é considerado como um fator negativo. No método japonês, as enzimas são inativadas por aquecimento a vapor, enquanto que no método chinês as enzimas são inativadas torrando as folhas em uma espécie de frigideira. Os passos de fabricação de chá verde são colheita, aquecimento a vapor/torra, aquecimento e enrolamento primário, enrolamento secundário, secagem, refino, torra, graduação e embalagem. Existem 18 tipos diferentes de chá verde, classificados como *pan fixed* (a citar: *pan dried*, *basket dried*, *cured roasted* e *sun dried*) e *steam fixed* (NAGALAKSHMI, 2003).

1.1.2. Composição química

A composição química das folhas de chá é complexa (ver Fig. 1.3). Carboidratos celulósicos, proteínas e lipídios são os maiores constituintes do chá. Uma vez que são insolúveis, eles não se tornam parte da bebida de chá; apenas componentes de baixo peso molecular, tais como polifenóis, cafeína, teobromina, vitamina C, metais e íons fluoreto passam à infusão (CHU e JUNEJA, 1997).

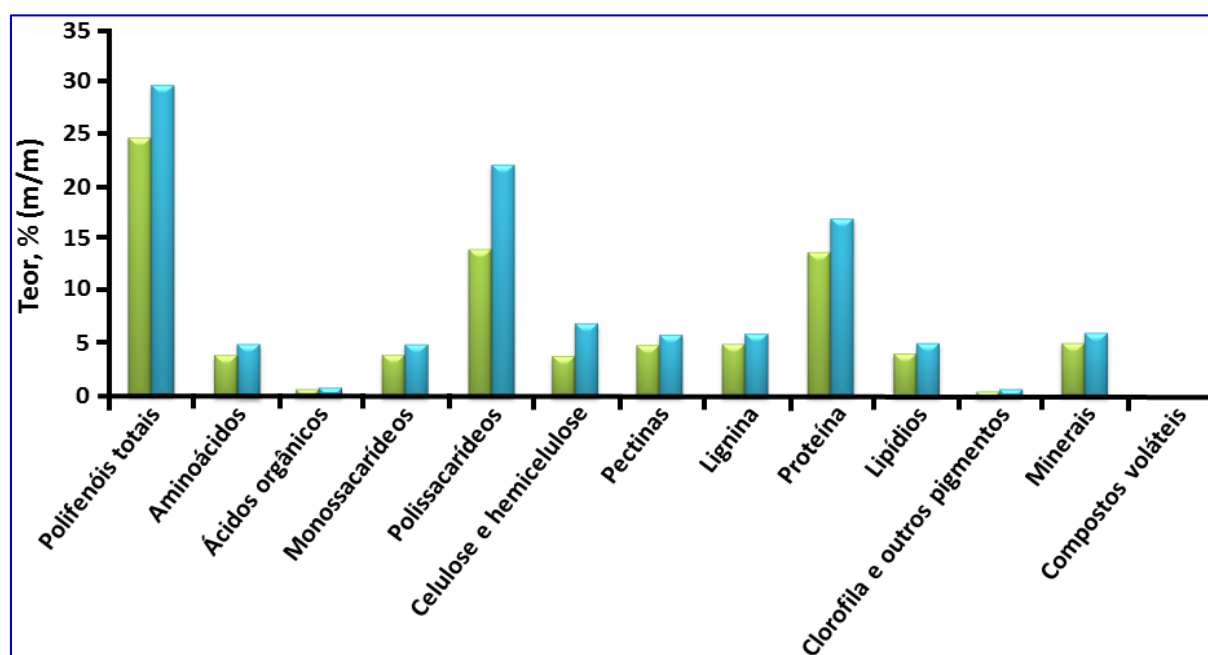


Figura 1.3. Composição química das folhas de chá. Os teores mínimo e máximo são mostrados em barras verde e azul, respectivamente. Adaptado de (SZYMCZYCHA-MADEJA et al., 2012).

A bebida do chá é considerada uma fonte rica de elementos nutricionais essenciais: polifenóis, flavonóis, proteínas, aminoácidos, enzimas, substâncias formadoras de aroma, vitaminas, fibras, óleos voláteis, purinas e carboidratos. Além disso, o chá também possui minerais e elementos-traço em sua composição (SZYMCZYCHA-MADEJA et al., 2012).

1.1.2.1. Pigmentos (NAGALAKSHMI, 2003)

Os flavonóis pertencem ao grupo das catequinas e constituem 25–30% da matéria seca da folha de chá. A qualidade do chá está relacionada principalmente com a concentração de flavonóis da folha fresca. A (–)-epigalocatequina galato constitui uma parte importante dos flavonóis na folha fresca e os níveis de flavonóis galatados diminuem com a idade de folha. As principais catequinas formadoras do pigmento no chá são mostradas na **Fig. 1.4**.

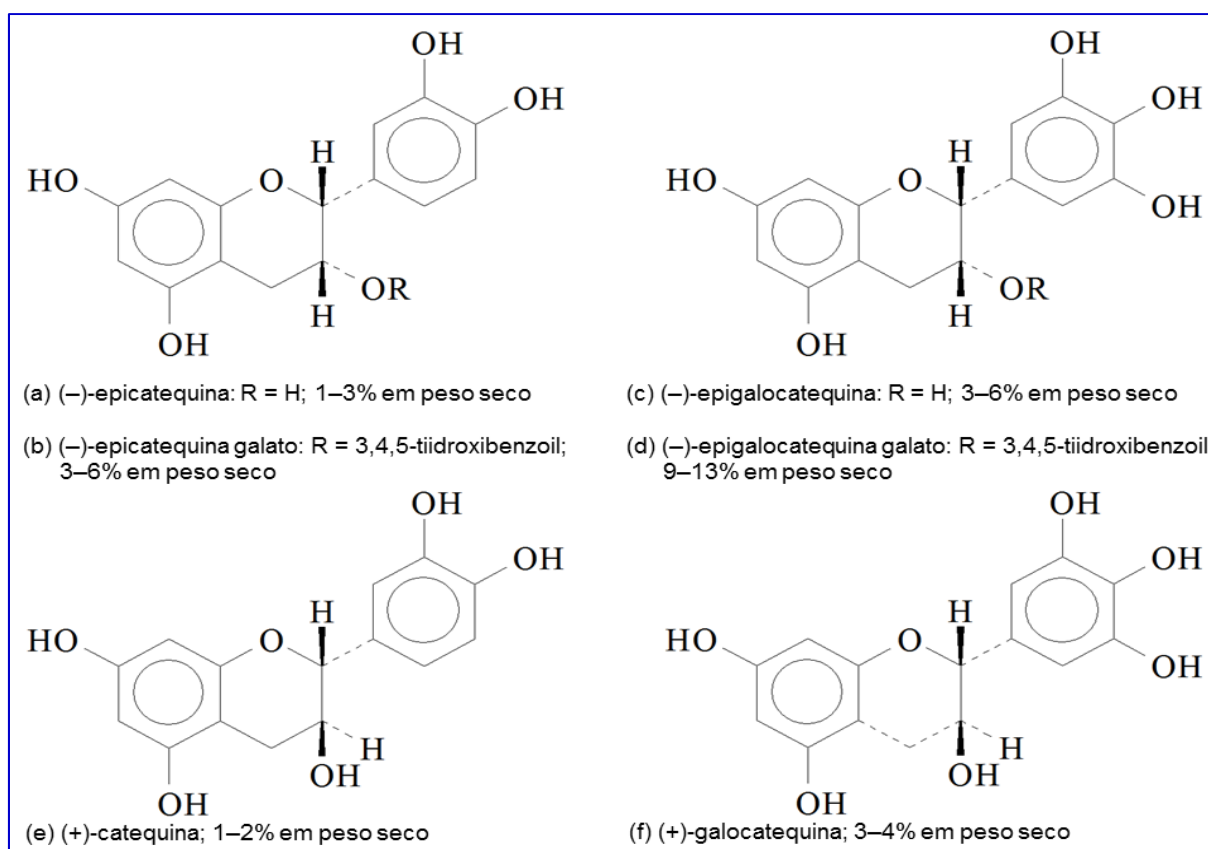


Figura 1.4. Os flavonóis que ocorrem no chá. Adaptado de (NAGALAKSHMI, 2003).

Todas as catequinas sofrem oxidação, principalmente a (-)-epigalocatequina e a (-)-epigalocatequina galato são facilmente oxidadas. A formação de quinona (**Fig. 1.5**) é a principal força motriz da fermentação, que dá cor ao chá preto.

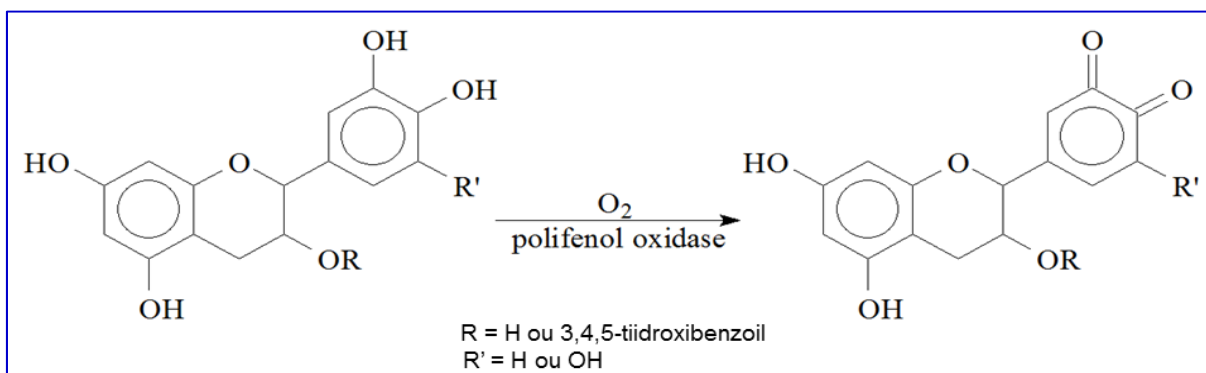


Figura 1.5. Formação de quinona. Adaptado de (NAGALAKSHMI, 2003).

1.1.2.2. Outros polifenóis (NAGALAKSHMI, 2003)

Outros polifenóis incluem o ácido gálico; flavonóis, tais como a quercetina, kaempferol, miricetina e seus glicosídeos, e depsídeos, tais como o ácido clorogênico e ácido quínico p-cumaril. A teogalina (ácido 3-galoilquínico) é exclusiva do chá.

1.1.2.3. Teaflavinas (NAGALAKSHMI, 2003)

A teaflavina é um composto formado a partir da reação entre uma quinona derivada de uma catequina com uma quinona derivada de uma galacatequina. Quatro teaflavinas podem ser isoladas e caracterizadas; outra forma, conhecida como isoteaflavina, também pode ser identificada.

Teaflavinas exibem uma cor vermelho alaranjada brilhante em solução. Elas são importantes na determinação do brilho, que é um atributo desejável de bebidas de chá, e são suficientemente características na cor a ser determinada por

espectrofotometria na fermentação chá. A concentração de teaflavinas totais no chá preto varia de 0,3 a 2,0%. Apenas 10% do teor original de catequina do chá é contabilizado como teaflavinas no chá preto. O prolongamento da fermentação diminui o teor de teaflavinas.

1.1.2.4. Ácidos teaflávicos (NAGALAKSHMI, 2003)

Embora o ácido gálico não seja oxidado diretamente pela polifenol oxidase do chá, ele é convertido em quinona ácido gálico por produtos de oxidação das catequinas. Esta reação com as formas oxidadas das catequinas libera ácidos teaflávicos. Os ácidos teaflávicos são substâncias azedas vermelho-brilhantes. Eles estão presentes apenas em quantidades muito pequenas no chá preto. Produtos oxidados da epigalocatequina acoplado com o ácido gálico forma uma série de substâncias incolores conhecidas como bisflavonóis, que ocorrem em quantidades muito pequenas no chá preto.

1.1.2.5. Tearubiginas (NAGALAKSHMI, 2003)

Durante a formação de chá preto, aproximadamente 15% das catequinas do chá permanecem inalteradas e 10% são consideradas para a formação de teaflavinas, ácidos teaflávicos e bisflavonóis. Aproximadamente 75% das catequinas são convertidas em um complexo, pertencente a um grupo de substâncias conhecidas como tearubiginas devido à sua cor castanho-avermelhada. Determinações de pesos moleculares dessas substâncias estão na faixa de 700–40000. Hidrólise redutiva de tearubiginas produz pequenas quantidades de todas as catequinas encontradas no chá. Existe evidência para a presença de proantocianidina polimérica, mas apenas em uma fração muito pequena.

As tearubiginas constituem o maior grupo de compostos no chá preto (acima de 20% m/m em base seca) e contribuem significativamente para a cor, força e sabor característico da bebida. Os polifenóis oxidados do chá preto são geralmente referidos como taninos, mas não há nenhuma relação com o ácido tânico (éster pentadigaloil de glicose), que não ocorre no chá. As tearubiginas dão um brilho específico e uma coloração vermelha viva para o extrato. A proporção de teaflavinas/tearubiginas é considerada como um índice de força do extrato. O controle destes dois fatores depende do tempo de fermentação e da temperatura.

1.2. Justificativa

Depois da água, o chá é bebida mais consumida, sendo, portanto, cultura de grande importância econômica e social para cerca de metade da população mundial (SZYMCZYCHA-MADEJA et al., 2012). As razões para sua popularidade são devidas ao seu aroma único e sabor característico. Além disso, seus benefícios nutricionais e farmacológicos (incluindo suas propriedades antimicrobianas e antioxidantes) também têm sido confirmados (PINTO, 2013; JAIN et al., 2013; PREEDY, 2012; CHATURVEDULA e PRAKASH, 2011; CHOW e HAKIM, 2011; SHARANGI, 2009).

A *Camellia sinensis* é cultivada comercialmente em mais de 30 países tropicais e subtropicais da Ásia, África e América do Sul. Os maiores produtores incluem China, Índia, Quênia, Sri Lanka e Indonésia (PREEDY, 2012). A **Fig. 1.6** mostra os principais produtores mundiais de chá, segundo a Organização das Nações Unidas para Alimentação e Agricultura. Aproximadamente 76–78% da produção mundial de chá é do tipo preto, enquanto o chá verde e oolong alcançam por volta de 18–20% e 2%, respectivamente (SANG et al., 2011).

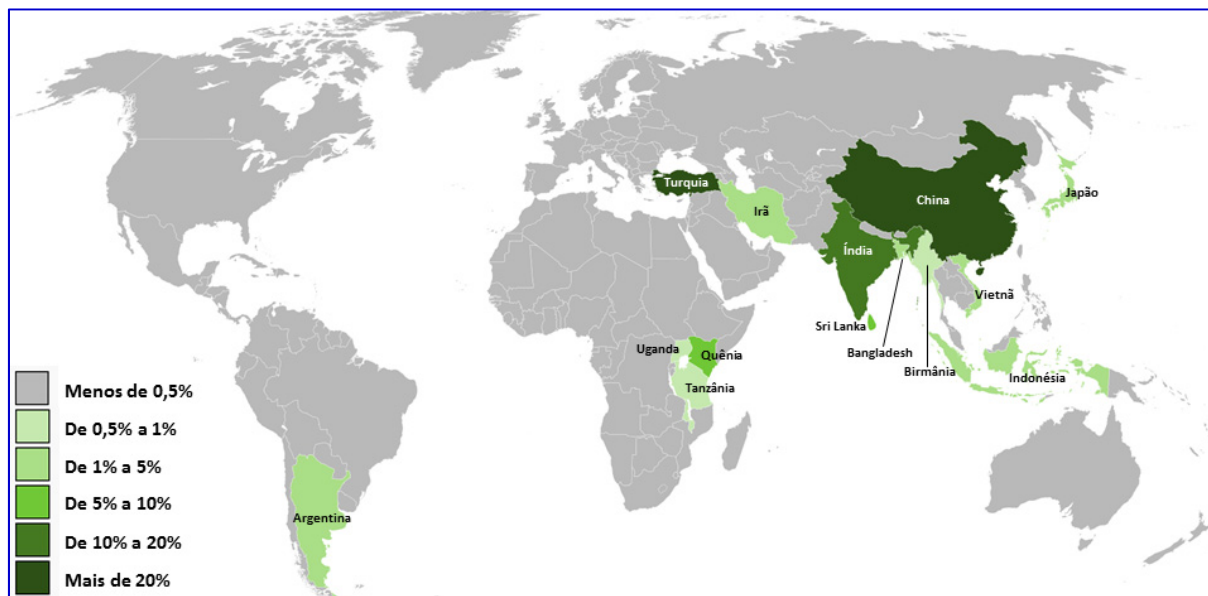


Figura 1.6. Produção total mundial de chá.

Na América do Sul, o principal produtor é a Argentina, que ocupa lugar de destaque como o 9º maior produtor mundial de chá (ONUAA, 2012). Em 2010, a produção argentina de chá foi de 90,7 mil toneladas, enquanto a produção brasileira foi de apenas 7,7 mil toneladas. No Brasil, a principal região produtora de chá é o Vale do Ribeira, no estado de São Paulo, principalmente os municípios de Registro, Pariquera-Açu e Cajati, onde a colonização japonesa foi a grande responsável pela expansão dessa cultura na região. O chá preto produzido no Vale do Ribeira vinha sendo classificado como de baixa qualidade e vendido para formação de mistura (*blend*). Atualmente, o plantio de genótipos melhorados, normalmente híbridos resultantes de cruzamentos do chá chinês (de folha estreita) e do chá assamês (de folha larga), os tratos culturais e o processo de fermentação mais adequados têm sido responsáveis pela possível obtenção de um produto de melhor qualidade. Isso tem favorecido um aumento do valor de mercado e, conseqüentemente, a inserção do produto brasileiro no mercado exportador (LIMA et al., 2012).

Devido à importância global do chá, a Organização das Nações Unidas para Alimentação e Agricultura criou um Grupo Intergovernamental sobre Chá. Em um recente documento gerado na 20ª Sessão desta Comissão, a situação do mercado atual de chá e as perspectivas em médio prazo para 2021 foram descritos. A produção mundial de chá verde e preto aumentou significativamente para 4,1 milhões de toneladas em 2010. A produção de chá preto teve um acréscimo de 5,5% em resposta aos preços recordes, uma vez que a demanda excedeu a oferta. Já a produção de chá verde aumentou 1,9% (ONUAA, 2012). Devido a esse aumento na demanda, a qualidade tem se tornado o parâmetro mais importante para o estabelecimento de mercado cada vez mais competitivo e, conseqüentemente, dos preços.

A qualidade do chá depende principalmente da variedade folha, ambiente de crescimento, técnica de processamento, tamanho de folhas de chá moídas e preparação da infusão. A qualidade é medida com base no licor (brilho, vivacidade, cor, etc), aroma, sabor e a aparência da folha. Em muitos casos, a produção da maioria dos chás envolve a mistura de diversas variedades para manter a consistência do sabor. Assegurar a mistura ideal requer que o provador de chá experimente centenas de amostras (SEETOHUL et al., 2006).

A folha de chá contém enzimas, carboidratos, proteínas, lipídios e elementos estruturais normalmente associados com o crescimento das plantas e a fotossíntese. Além disso, as folhas de chá distinguem-se pelo seu notável conteúdo de metilxantinas e polifenóis. Estes dois grupos de compostos são os principais responsáveis pelas propriedades únicas do chá que contam para a sua popularidade como uma bebida. Os constituintes químicos mais importantes que influenciam o sabor e o aroma de infusões de chá são polifenóis, flavonóis, cafeína, açúcares,

ácidos orgânicos, aminoácidos e compostos de aroma voláteis. Compostos fenólicos dos chás, tais como teaflavinas e tearubiginas, são muito importante do ponto de vista da qualidade intrínseca e constituem 60% dos componentes solúveis totais em água no chá preto. Além disso, eles são responsáveis pelas principais propriedades organolépticas de cor, brilho e adstringência (SEETOHUL et al., 2006). A cafeína é responsável pela vivacidade. Os compostos voláteis de aroma do chá e suas variações na composição devidas a variáveis geográficas e de processamento são de fundamental importância do ponto de vista da qualidade (BORSE et al., 2002).

No mercado mundial de chá existem algumas disputas comerciais no que diz respeito aos tipos de chás. Em muitos países produtores de chá, como o Japão, Taiwan e Coreia, a tributação de chás importados é feita de forma diferente com base no grau de fermentação. Na Coreia, por exemplo, chás verdes (não fermentados) têm uma tarifa mais de 500%, mas apenas de 40% para chás oolong e preto (fermentados). No entanto, não há nenhum método padrão reconhecido internacionalmente para a classificação de chás. Os consumidores comumente classificam chás de acordo com a cor de infusão, que pode ser um padrão melhor do que a de muitos países que só contam com a aparência das folhas de chá secas, o que, muitas vezes, dificulta a discriminação (WANG et al., 2008). As diferenças entre os chás verde, oolong e preto são devidas principalmente aos processos de murchamento, enrolamento e fermentação (Fig. 1.7). Por isso, os chás são geralmente classificados de acordo com os métodos de processamento, dentre os quais o grau de fermentação é um dos mais importantes (WANG et al., 2008).



Figura 1.7. Exemplos de variedades comerciais dos chás (a) verde, (b) oolong e (c) preto.

É difícil para um não especialista avaliar a qualidade do chá no mercado e, além disso, há também uma abundância de chás falsificados que são de má qualidade, mas vendidos a preços elevados. Por isso, a sua avaliação é um processo sistemático e complicado, no qual os efeitos integrados de atributos, tais como cor, aroma, sabor, bem como a forma (textura) das folhas de chá, devem ser consideradas (MEI et al., 2012). Essa avaliação sensorial é o único método que fornece medições integradas e diretas de intensidades perceptíveis de tais atributos. Eles têm sido tradicionalmente avaliados por provadores de chá especializados que desenvolveram uma linguagem própria, a qual é, muitas vezes, difícil de ser compreendida pelos consumidores para descrever vários atributos de qualidade de uma infusão de chá (HE et al., 2009). O resultado de cada provador depende da aparência (50%), sabor (30%) e aroma (20%); a média dos resultados dos três provadores é então usada para determinar o grau do chá. Contudo, os resultados desse tipo de avaliação não são precisos e podem ser marcadamente imparciais (ZHAO et al., 2009). Além disso, os painéis sensoriais são demorados, dispendiosos e não são adequados para medições em linha. Assim, há uma necessidade, tanto para os consumidores quanto para a indústria do chá, do desenvolvimento de



tecnologias mais eficientes para avaliar objetivamente a qualidade do chá e discriminá-los entre seus diferentes tipos (HE et al., 2009).

Por outro lado, a qualidade organoléptica da infusão da *Camellia sinensis* pode ser relacionada diretamente com a origem geográfica das plantas, visto que ela depende da natureza e da quantidade de vários metabólitos secundários. Estes componentes e seus efeitos benéficos já comprovados sobre a saúde humana são uma das bases do valor econômico do chá. Neste sentido, impulsionado pela reputação de países e plantações particulares, existe no mercado mundial de chá um interesse crescente dos consumidores por chás (e seus derivados) de alta qualidade com uma clara origem geográfica. A produção de chás de alta qualidade significa que os produtores podem exigir um preço significativamente mais caro que a média (YE, 2012). Na Argentina, por exemplo, um chá importado custa de sete a dez vezes mais que um chá nacional.

Uma vez que há uma evidente discordância entre a qualidade do chá e seu preço de mercado, a base teórica para resolver este problema é avaliar a sua qualidade cientificamente. Durante as últimas décadas, as metodologias analíticas que empregam técnicas instrumentais modernas tornaram-se mais sensíveis, confiáveis e rápidas. Após a análise instrumental, a quimiometria é, na maioria das vezes, necessária para analisar a quantidade de dados acumulados e para detectar diferenças sutis que normalmente existem entre amostras de chá de diferentes tipos e origens. Portanto, o desenvolvimento de métodos que podem substituir parcialmente a avaliação sensorial durante análises de rotina, alcançando medidas objetivas em um curto espaço de tempo e de uma forma consistente e de baixo custo relativo é de grande importância para a indústria de alimentos (HE et al., 2009).



Assim, faz-se necessário a realização de pesquisas visando a exploração do máximo potencial produtivo e a produção de chá de melhor qualidade, a exemplo do que tem sido feito para outras culturas agrícolas no Brasil. Portanto, o desenvolvimento de estratégias analíticas para classificação de chás de acordo com o tipo e a origem geográfica pode fornecer segurança aos compradores e consumidores, além de poder prevenir a rotulagem fraudulenta de chás. Tais metodologias devem ser de fácil implementação nas indústrias alimentícia, farmacêutica e de cosméticos, além de serem úteis para a elaboração de normas para o controle de qualidade de chás, ainda inexistentes no Brasil.

1.3. Objetivo

1.3.1. Objetivo geral

Esta Tese de Doutorado tem como objetivo o desenvolvimento de estratégias analíticas para a classificação simultânea de chás (*Camellia sinensis*) comerciais de acordo com o tipo (chá verde ou preto) e origem geográfica (Brasil, Argentina e Sri Lanka), utilizando imagens digitais, espectroscopia no infravermelho próximo e composição química. A avaliação quimiométrica foi realizada usando três diferentes técnicas de reconhecimento de padrões multivariado: (a) Modelagem Independente e Flexível por Analogia de Classe (SIMCA), (b) Análise Discriminante por Mínimos Quadrados Parciais (PLS-DA) e (c) Análise Discriminante Linear associada à técnica de seleção de variáveis feita pelo Algoritmo das Projeções Sucessivas (SPA-LDA).

1.3.2. Objetivos específicos

- **1ª Aplicação:** Classificação simultânea do tipo e origem geográfica de chás utilizando imagens capturadas com uma câmera digital. Histogramas de cor gerados a partir das imagens digitais foram usados como informação analítica. A classificação multivariada foi realizada usando SIMCA, PLS-DA e SPA-LDA.

- **2ª Aplicação:** Classificação simultânea do tipo e origem geográfica de chás empregando espectroscopia NIR associada a SIMCA, PLS-DA e SPA-LDA.

- **3ª Aplicação:** Classificação simultânea do tipo e origem geográfica de chás empregando composição química associada a SIMCA, PLS-DA e SPA-LDA. Para este propósito, foram quantificados os teores de umidade, cinza total, cafeína, polifenóis totais e fluoreto, além de 15 metais (Na, Mg, Al, P, K, Ca, Cr, Mn, Fe, Co, Ni, Cu, Zn, Cd e Pb) nas folhas e infusões dos chás. Diferentes abordagens analíticas foram exploradas e os resultados obtidos foram analisados usando SIMCA, PLS-DA e SPA-LDA.

Capítulo 2

Fundamentação teórica



2. FUNDAMENTAÇÃO TEÓRICA

2.1. Imagens digitais

Para entender a imagem digital, devemos primeiro rever os conceitos básicos de uma imagem. Uma imagem é a representação visual de um objeto. As imagens vão desde desenhos simples até pinturas e fotografias. A imagem é uma representação de um objeto tridimensional sobre uma superfície plana, isto é, uma representação bidimensional. Se isolarmos uma parte da imagem para uma faixa vertical ou horizontal, a intensidade ou a cor variam dependendo se você olha para baixo ou através da faixa. Isoladamente, uma única faixa de dados de imagem é bastante insignificante, mas se combinarmos as faixas juntas temos uma imagem. A imagem digital é, portanto, uma matriz bidimensional de dados com valores em cada elemento da matriz exibida como uma intensidade ou uma cor. A matriz de dados é uma representação digital de uma faixa horizontal ou vertical de uma imagem. Matematicamente, a faixa horizontal da imagem pode ser definida como $f(x)$, em que f representa a intensidade ou cor num determinado local x . A faixa vertical da imagem pode ser definida como $f(y)$, em que f representa a intensidade ou cor numa dada localização y . A combinação de todas as faixas horizontais e verticais de dados permite definir a imagem matematicamente como uma função bidimensional:

$$f(x,y) \tag{2.1}$$

onde x e y são as coordenadas espaciais que identificam qualquer localização na imagem e o valor de f é representado como uma cor ou brilho no ponto (x, y) . Em uma imagem digital, cada ponto da imagem e nível de brilho é um valor discreto (DANIEL, 2009).

A matriz bidimensional da imagem digital é constituída por um número específico de linhas e colunas. A matriz bidimensional representa uma matriz de

números. Cada célula da matriz tem coordenadas espaciais discretas ou um endereço específico que descreve a sua localização na imagem. O endereço de cada ponto pode ser definido usando um sistema de coordenadas cartesianas, que é formado por duas linhas perpendiculares que se intersectam na origem. Representando esse referencial sob a forma de um gráfico, obtemos a abscissa (x) medindo a distância do ponto observado ao eixo das ordenadas (y), paralelamente ao eixo das abscissas. A intensidade da luz em cada um destes pontos também é dada por um valor numérico discreto. Cada endereço ou coordenada espacial é chamado de elemento de imagem ou pixel (DANIEL, 2009).

O número de linhas e colunas (tamanho da matriz) irá definir a resolução espacial da imagem digital. Quanto maior for o tamanho da matriz, melhor será a resolução da imagem digital. O aumento do tamanho da matriz irá diminuir o tamanho de cada pixel. O menor objeto representado na imagem digital ocupa o espaço de 1 pixel e, portanto, a capacidade de resolver esses pequenos objetos depende do tamanho da matriz da imagem digital. Esses conceitos são a base teórica empregada em câmeras digitais (DANIEL, 2009).

A profundidade da imagem refere-se à quantidade de memória do computador atribuída a cada pixel. Computadores armazenam informações em um código binário chamado de "bit". Cada bit pode ser ligado ou desligado e pode ser atribuído um valor de 1 ou 0. O número de bits atribuídos para a matriz da imagem determina a quantidade máxima de informação que pode ser armazenada em um pixel individual. Em uma imagem em preto e branco, o número de bits determina o número de tons de cinza que podem ser exibidos (DANIEL, 2009).

Uma imagem de 10 bits contém 1024 tons de cinza, enquanto que uma imagem de 12 bits contém 4.096 tons de cinza. A imagem em preto e branco ou em

tons de cinza é considerado uma imagem de canal único. Uma imagem colorida pode ter mais do que um canal e, quando estes canais são combinados, eles são capazes de fornecer uma grande variedade de cores. Por exemplo, uma imagem RBG (vermelho, verde e azul) tem três canais. Cada canal terá diferentes intensidades ou tons de vermelho, azul e verde, respectivamente. Como tal, uma imagem RBG exigiria três vezes a quantidade de memória do computador do que a mesma imagem armazenada em escala de cinza. A imagem padrão RGB é de 24 bits, sendo 8 bits para cada canal. A imagem é composta pelos canais das três cores, a cada cor tendo intensidades de brilho entre 0 e 255 (**Fig. 2.1**). Se a imagem de RGB é de 48 bits, cada um dos três canais tem uma escala de cor de 16 bits (**DANIEL, 2009**).

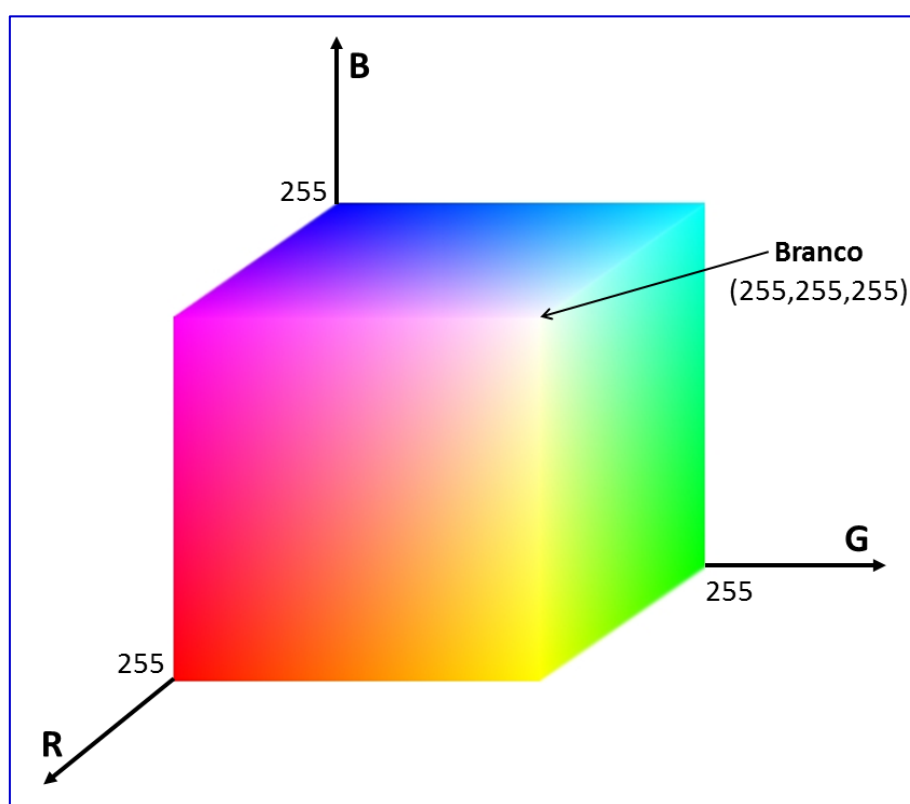


Figura 2.1. Sistema de cor RGB. Adaptado de <http://www.mathworks.com/help/images/reducing-the-number-of-colors-in-an-image.html>

O sistema visual humano pode distinguir centenas de milhares de tons e intensidades de cor diferentes, mas apenas cerca de 100 tons de cinza. Portanto, em uma imagem uma grande quantidade de informação extra pode estar contida na cor, e esta informação adicional pode então ser usada para simplificar a análise da imagem, por exemplo, identificação de objetos e extração com base na cor. Três quantidades independentes são utilizadas para descrever qualquer cor particular. O *matiz* é determinado pelo comprimento de onda dominante. A *saturação* é determinada pela pureza da excitação e depende da quantidade de luz branca misturada com o matiz. O matiz puro é totalmente saturado, ou seja, nenhuma luz branca está misturada. Matiz e saturação em conjunto determinam a cromaticidade de uma determinada cor. Finalmente, a *intensidade* é determinada pela quantidade de luz, com mais luz correspondendo a cores mais intensas (GONZALES e WOODS, 1992).

Luz acromática não tem cor - seu único atributo é a quantidade ou intensidade. A escala de cinza é uma medida de intensidade. A intensidade é determinada pela energia e é, por conseguinte, uma grandeza física. Por outro lado, o *brilho* (ou *luminosidade*) é determinado pela percepção da cor e é, portanto, psicológico. Dados azul e verde igualmente intensos, o azul é percebido como sendo muito mais escuro do que o verde. Note-se também que a percepção da intensidade não é linear, com mudanças de intensidade normalizada de 0,1–0,11 e 0,5–0,55 sendo percebida como mudanças iguais no brilho (FOLEY et al., 1990).

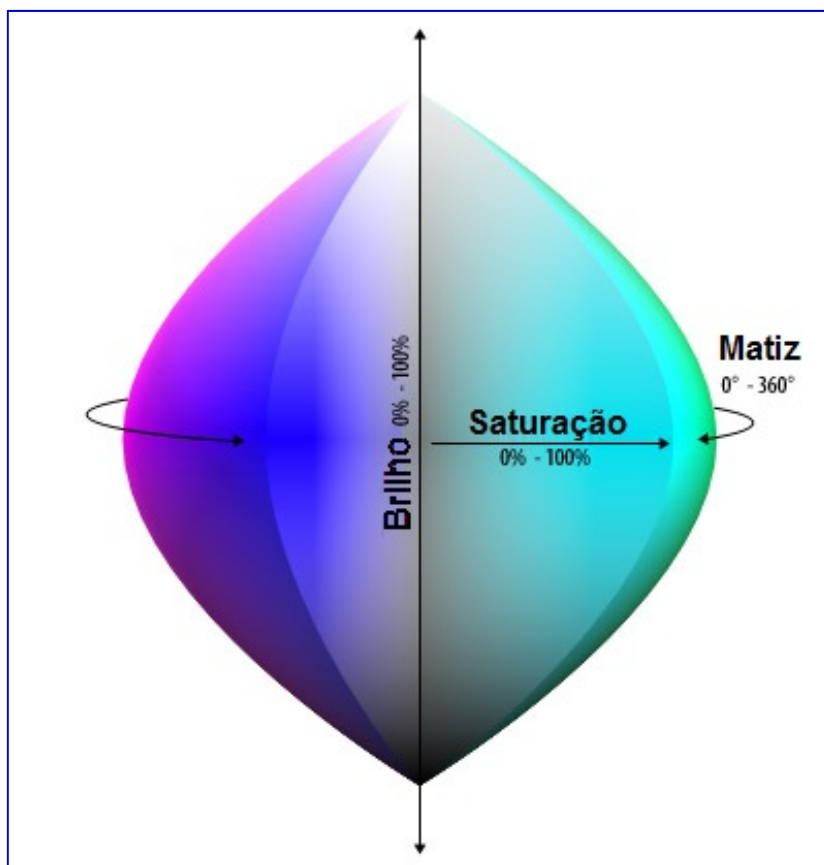


Figura 2.2. Sistema de cor HSB. Adaptado de http://dba.med.sc.edu/price/irf/Adobe_tg/models/hsb.html

A cor depende principalmente das propriedades de reflectância de um objeto. Vemos os raios de luz que são refletidos, enquanto outros são absorvidos. No entanto, também é preciso considerar a cor da fonte luminosa e a natureza do sistema visual humano. Por exemplo, um objeto que reflete vermelho e verde aparecerá verde quando nenhuma luz vermelha o iluminar e, inversamente, ele vai aparecer vermelho na ausência de luz verde. Em uma luz branca pura, ele aparecerá amarelo (= vermelho + verde) (FOLEY et al., 1990).

2.1.1. Revisão da literatura

Tradicionalmente, a avaliação sensorial humana feita por um provador profissional é empregada para estimar a qualidade do chá. O resultado de cada



provador é dependente da aparência (50%), sabor (30%) e aroma (20%); a média dos resultados dos três provadores é então usada para determinar a grau do chá. Contudo, os resultados desse tipo de avaliação não são precisos e podem ser marcadamente imparciais (ZHAO et al., 2009). A imagem digital, por outro lado, é uma ferramenta analítica com grande potencial para uso em análises qualitativa e quantitativa (GAIÃO et al., 2006; WU et al., 2008; LYRA et al., 2009; ZHAO et al., 2009; LOPEZ-MOLINERO et al., 2010; LYRA et al., 2011; GILL et al., 2011; TÔRRES et al., 2011; LI et al., 2011; SOUZA et al., 2012; LADDI et al., 2012; O'DRISCOLL et al., 2013; ANDRADE et al., 2013; LADDI et al., 2013; WU e SUN, 2013). Esta é uma das fronteiras emergentes de pesquisa avançada, e lida com o processo de captura, acondicionamento e processamento da imagem digital usando avançados algoritmos de computação leve que extraem informações e características importantes das imagens adquiridas (JÄHNE, 2005). A técnica não é invasiva, facilita a aquisição de dados e, em muitos casos, diminui custos. Também é possível reproduzir uma avaliação humana de maneira padronizada, imparcial e robusta (GILL et al., 2011).

Na indústria de chá, a cor e as alterações de cor devem ser medidas e monitoradas para avaliar a qualidade do chá durante o processo de fermentação para correlacionar informações essenciais sobre as alterações físicas e químicas que estão ocorrendo. A triagem de chá preto em diferentes graus, de acordo com as variações do tamanho dos grânulos é um processo crítico em indústrias de chá para avaliação da qualidade. O chá passa através de uma série de peneiras que vibram de diferentes tamanhos de malha. Os vários graus de chá são obtidos em diferentes pontos destas peneiras. Estes graus de chá são geralmente classificados em quatro classes: *leaf*, *brokens*, *fannings* e *dust*, em ordem decrescente de seu tamanho de

partícula. Alguns dos termos mais usados para descrever o chá preto são *attractive* ou *well-made* que representa uma amostra bem feita de grânulos com cor e tamanho uniformes; *even* descrevendo uma amostra contendo grânulos de chá de tamanho uniforme; *mixed* que representa a presença de diferentes graus juntos em uma amostra; *bold* indicando a presença de pedaços de folhas; e *stalky* contendo a presença indevida de caule. Estes procedimentos não destrutivos têm sido realizados empregando-se técnicas de visão computacional e análise de imagens para triagem de chás com base em seus parâmetros físicos: granulagem, cor, forma, tamanho e textura (GILL et al., 2011; LADDI et al., 2012; LADDI et al., 2013; WU e SUN, 2013).

Poucas metodologias empregando análise de imagens e técnicas quimiométricas de reconhecimento de padrões têm sido reportadas na literatura para a classificação de chás. Neste contexto, a maior parte das publicações envolvem estudos que utilizam imagens espectrais associadas com algoritmos de máquinas de vetores de suporte (SVM) (WU et al., 2008; ZHAO et al., 2009; LI et al., 2011).

Wu et al. (2008) empregaram imagens multiespectrais para classificação de quatro variedades de chás verdes produzidos na China. Valores de entropia foram obtidos a partir da textura das imagens e Máquinas de Vetores de Suporte baseadas em Mínimos Quadrados (LS-SVM), alcançando 99,58% de classificação correta. Este resultado foi melhor que àqueles obtidos para PLS-DA (32,5%) e ANN com função de base radial (92,92%).

Zhao et al. (2009) utilizaram imagens hiperespectrais para classificar cinco variedades de chá verde produzidos na China. A fim de extrair variáveis características das imagens foi realizada uma PCA seguida de uma análise de

textura. O modelo de classificação foi construído usando SVM, que obteve 98 e 95% de classificação correta para os conjuntos de treinamento e teste, respectivamente.

[Li et al. \(2009\)](#) classificaram oito variedades de chás chineses famosos (sete chás verde e um chá oolong). Para extrair características da textura das imagens multiespectrais foi utilizada uma transformada wavelet seguida da construção de um modelo que usa SVM associado à regressão por mínimos quadrados multiclasse. O modelo MCLS-SVM obtido conseguiu 96,82% de classificação correta para o conjunto de teste.

Apesar das imagens espectrais permitirem a construção de modelos espectro-espaciais mais elaborados para uma segmentação mais precisa e classificação da imagem, há uma série de desvantagens para o uso desta técnica. Computadores muito rápidos, detectores altamente sensíveis e grande capacidade de armazenamento dos dados, por exemplo, aumentam o custo de aquisição e processamento de dados espectrais. Além disso, há a necessidade de utilização de algoritmos mais complexos para o tratamento das imagens e dados.

Neste contexto, uma alternativa muito simples e barata foi proposta por [Diniz et al. \(2012\)](#) para a classificação simultânea do tipo (verde e preto) e da origem geográfica (Argentina, Brasil e Sri Lanka) usando imagens digitais obtidas com uma câmera digital e classificação multivariada usando Modelagem Independente e Flexível por Analogia de Classe (SIMCA) e o Algoritmo das Projeções Sucessivas (SPA) para seleção de variáveis combinada com LDA (SPA-LDA). A metodologia usou histogramas de cor (RGB, HSI e escala de cinza) gerados a partir de cada imagem e posterior construção dos modelos de classificação. SPA-LDA classificou corretamente todas as amostras de chás estudadas, enquanto SIMCA não demonstrou ser uma ferramenta útil para esta mesma finalidade. Isso demonstra a

vantagem de se utilizar uma técnica de seleção de variáveis para extrair a informação analítica relevante (neste caso, SPA-LDA usou apenas 17 variáveis).

2.2. Espectroscopia NIR

A Sociedade Americana de Testes e Materiais (ASTM) define a região do infravermelho próximo (NIR) do espectro eletromagnético como o intervalo de comprimentos de onda de 780 – 2526 nm, que correspondem à faixa de números de onda de 12820 – 3959 cm^{-1} . As bandas de absorção mais importantes que ocorrem na região NIR são relacionadas com *sobretons* e *combinações de vibrações fundamentais* dos grupos funcionais –CH, –NH, –OH e –SH. Os principais fatores que determinam a ocorrência e propriedades espectrais, isto é, frequência e intensidade das bandas de absorção NIR são a *anarmonicidade* e a *ressonância de Fermi*, que são a base física do que iremos descrever brevemente nesta seção. (PASQUINI, 2003; REICH, 2005; SUN, 2009; WEHLING, 2010).

Uma vez que a curva de energia de uma molécula que oscila é afetada por interações intramoleculares, as vibrações em torno da posição de equilíbrio não são simétricas e os espaçamentos entre os níveis de energia em que a molécula pode alcançar não são idênticos, diminuindo com o aumento da energia. Esta situação assemelha-se ao modelo mecânico-quântico de um *oscilador anarmônico*. Como as regras de seleção mecânico-quânticas não excluem rigorosamente as transições com $\Delta v > 1$ para sistemas anarmônicos, as transições entre os estados vibracionais de $\Delta v = 2$ ou 3 são possíveis, embora a sua probabilidade diminua com um aumento no número quântico vibracional v . Estas transições de energia multinível são a origem das bandas NIR de sobretons que ocorrem em múltiplos das frequências vibracionais fundamentais. Para a maioria das ligações químicas, os números de

sobretons podem ser estimados a partir de suas vibrações fundamentais com uma constante de anarmonicidade χ de 0,01– 0,05 pela seguinte equação (REICH, 2005):

$$\nu_x = \Delta\nu \nu_0 (1 - \Delta\nu \chi) \quad (2.1)$$

onde ν_x é o número de onda do sobreton x , ν_0 é o número de onda da vibração fundamental e χ é a constante de anarmonicidade.

Bandas de combinação que aparecem entre 1900 nm e 2500 nm são o resultado de interações vibracionais, ou seja, as suas frequências são as somas dos múltiplos de cada frequência de interação. Um tipo especial de interação de configuração, chamado *ressonância de Fermi*, leva a característica de que duas bandas de absorção NIR de uma molécula poliatômica com a mesma frequência não se sobrepõem simplesmente e se somam, mas dividem-se em dois picos de frequências um pouco mais altas ou mais baixas do que a esperada para a posição não perturbada. Além disso, ligações de hidrogênio intermoleculares e interações de dipolo devem ser consideradas, uma vez que alteram os estados de energia vibracional, deslocando, assim, bandas de absorção existentes e/ou dando origem a novas. Este efeito permite que substâncias cristalizadas, por exemplo, sejam determinadas por espectroscopia NIR (PASQUINI, 2003; REICH, 2005; SUN, 2009; WEHLING, 2010).

Em conclusão, as bandas de absorção NIR são tipicamente largas, sobrepostas e de 10 a 100 vezes mais fracas do que suas bandas fundamentais de absorção correspondentes no infravermelho médio. Estas características restringem severamente a sensibilidade no sentido espectroscópico clássico e requerem um tratamento quimiométrico dos dados para relacionar a informação espectral com as propriedades da amostra. O baixo coeficiente de absorção, no entanto, permite maior profundidade de penetração e, conseqüentemente, um ajuste da espessura da

amostra. Este aspecto é, na verdade, uma vantagem analítica, uma vez que permite a análise direta de amostras que absorvem fortemente e até mesmo com alto espalhamento de luz, tais como líquidos ou sólidos turvos em quaisquer dos modos de reflectância ou transmitância sem pré-tratamentos adicionais (PASQUINI, 2003; REICH, 2005; SUN, 2009; WEHLING, 2010).

A dupla dependência do sinal analítico sobre as propriedades físicas e químicas da amostra, resultantes de efeitos de absorção e de espalhamento de luz, pode ser usada favoravelmente para realizar análises químicas e físicas a partir de uma única medição. No entanto, se este não for o alvo analítico, os efeitos de dispersão em espectros NIR, resultantes de variações físicas nas amostras, também podem representar problemas analíticos mais ou menos graves. Nestas situações, eles devem ser considerados no processo de calibração como “parâmetros interferentes” (OLINGER et al., 2001; CIURKIZAC, 2001; BOKOBZA, 2002; REICH, 2005), como serão discutidos na Seção 3.

2.2.1. Revisão da literatura

Desde que Hall et al. (1988) usaram a espectroscopia NIR na predição da qualidade de chá preto, a aplicação desta técnica para quantificar componentes químicos e avaliar a qualidade de chás tem sido tema de muitas pesquisas. Entretanto, poucos estudos envolvendo o uso da espectroscopia NIR para a classificação de chás de acordo com a variedade e/ou da origem geográfica têm sido reportados na literatura.

Chen et al. (2005; 2006) estudaram três variedades de chás produzidos em quatro províncias chinesas usando espectroscopia NIR e Modelagem Independente e Flexível por Analogia de Classe (SIMCA). No total, foram estudadas quatro classes

de chás, sendo duas de chá verde, uma de chá oolong e uma de chá preto. Todas as amostras de chá foram classificadas com sucesso segundo a sua variedade a um nível de confiança de 99%. Entretanto, erros de classificação foram obtidos para os conjuntos de treinamento e teste para as duas classes de chá verde, que eram de diferentes origens geográficas.

[Zhao et al. \(2006\)](#) e [Chen et al. \(2007\)](#) utilizaram espectroscopia NIR, Máquina de Vetores de Suporte (SVM) e Retropropagação em Redes Neurais Artificiais (BP-ANN) para classificar chás verde, oolong e preto produzidos na China. Foram obtidos percentuais de classificação correta de 75 e 90% para as diferentes variedades de chá usando BP-ANN e SVM, respectivamente.

[He, Li & Deng \(2007\)](#) classificaram oito diferentes variedades de chá verde produzidos na China usando espectroscopia Vis-NIR (325–1075 nm). Os dados foram pré-processados usando transformada *wavelet* (WT) e, em seguida, aplicou-se Análise de Componentes Principais (PCA) aos dados pré-processados. Os escores das oito primeiras PC's foram usados como entradas para a construção de um modelo BP-ANN, que obteve 100% de classificação correta para as oito diferentes variedades de chá verde estudadas.

[Li & He \(2008\)](#) combinaram a utilização de espectroscopia Vis-NIR, WT, PCA e BP-ANN para classificar diferentes variedades de *Camellia sinensis* produzidos três diferentes jardins da China. Os experimentos foram conduzidos em campo (nos jardins de chá) e os espectros de reflectância foram medidos em tempo real. Foram levadas em consideração muitas fontes óbvias de introdução de erros causadas por ruído de fundo, ângulo e intensidade de incidência solar e do vento. O percentual médio de classificação para o modelagem proposta foi de 73,3%.

Chen et al. (2008) propuseram a utilização de espectroscopia NIR para classificar seis variedades de chá verde e uma de chá oolong provenientes da China. Eles empregaram Análise Discriminante Linear (LDA) e ANN, que obtiveram 93,65 e 100% de classificação correta, respectivamente.

Chen et al. (2009) também estudaram a classificação de chás verdes segundo sua origem geográfica empregando espectroscopia NIR e diferentes técnicas de reconhecimento de padrões. As amostras foram produzidas em quatro diferentes províncias chinesas: Anhui, Henan, Jiangsu e Zhejiang. Os percentuais de classificação correta foram de 92,3% para LDA, 96,3% para *k*NN (*K*-Vizinhos mais Próximos), 96,3% para BP-ANN e 100% para SVM.

Liu, Tsai & Ou (2010) investigaram a classificação de chás parcialmente fermentados produzidos em Taiwan utilizando de espectroscopia NIR. As amostras foram coletadas e analisadas separadamente segundo suas seis diferentes variedades, seis diferentes áreas de produção e duas diferentes estações do ano. Foi empregada Análise Discriminante por Mínimos Quadrados Parciais (PLS-DA), que classificou corretamente 96,3% das amostras de acordo com a variedade, 94,1% segundo a área de produção e 99,2% de acordo com a estação do ano.

Tan et al. (2012) classificaram chás verde, oolong, preto e amarelo produzidos na China usando espectroscopia NIR, PLS-DA e Árvore de Regressão e Classificação (CART). A metodologia proposta obteve uma classificação correta de 83 e 94% usando PLS-DA e CART, respectivamente.

Ren et al. (2013) empregaram espectroscopia NIR e PLS-DA para classificar chás pretos de acordo com a sua origem geográfica, que incluíam diferentes países (China, Índia, Quênia, Sri Lanka e Birmânia) e diferentes províncias chinesas (Anhui, Hubei e Yunnan). Foram obtidos apenas dois erros de classificação (uma amostra

de Hubei e uma amostra do Sri Lanka), totalizando uma média de 94,3% de classificação correta.

2.3. Técnicas de reconhecimento de padrões

Hoje em dia, os instrumentos de análise modernos permitem produzir grandes quantidades de informação (variáveis ou características) para um grande número de amostras (objetos) que podem ser analisados em um período de tempo relativamente curto. Isto conduz à existência de uma matriz de dados multivariados, que requerem o uso de procedimentos matemáticos e estatísticos, a fim de extrair de forma eficiente o máximo de informação útil a partir dos dados (BERRUETA et al., 2007).

Técnicas de reconhecimento de padrões supervisionadas utilizam a informação sobre a associação de classe das amostras para um determinado grupo (classe ou categoria) de modo a classificar novas amostras desconhecidas em uma das classes conhecidas (MASSART et al., 1997; LAVINE, 2000). Para isso, elas utilizam uma estratégia comum de acordo com o algoritmo aplicado, que consiste nos passos seguintes (MASSART et al., 1997; BRERETON, 2003):

- (1) Seleção dos conjuntos de treinamento, validação e teste, que consistem em objetos de classe conhecida para os quais variáveis são medidas.
- (2) Seleção de variáveis. As variáveis que contêm informações para a classificação destinada são mantidas, enquanto que aquelas que codificam o ruído e/ou não possuem poder de classificação são eliminadas.

- (3) Construção de um modelo usando o conjunto de treinamento. A modelagem matemática baseia-se em um determinado número de variáveis medidas nas amostras que constituem o conjunto de treinamento e as suas categorias conhecidas.
- (4) Validação do modelo utilizando as amostras de um conjunto de teste independente, a fim de avaliar a confiabilidade da classificação obtida.

Vários tipos de métodos de reconhecimento de padrões foram aplicados em ciência dos alimentos, mas eles diferem essencialmente na forma como alcançam a classificação (MASSART et al., 1997). Dois tipos de métodos são comumente distinguidos em uma primeira abordagem: aqueles focados na discriminação entre classes, como Análise Discriminante Linear (LDA), k -Vizinhos mais Próximos (k NN), Árvores de Regressão e Classificação (CART), a Análise Discriminante por Mínimos Quadrados Parciais (PLS-DA), Redes Neurais Artificiais (ANN) e Máquina de Vetor de Suporte (SVM); e aqueles orientados através da modelagem de classes, como Modelagem Independente Flexível por Analogia de Classe (SIMCA) e Classes Dispersas Desiguais (UNEQ). Técnicas de discriminação são usadas para construir modelos baseados em todas as categorias envolvidas no estudo, enquanto que os métodos de modelagem de classe criam um modelo separado para cada categoria. Uma das desvantagens dos métodos de discriminação é que amostras são sempre classificadas em uma das categorias apresentadas, mesmo que eles não pertencem a nenhuma delas. Métodos de modelagem de classe consideram os objetos que se encaixam no modelo para uma categoria como parte do modelo, e classificam como não membros aqueles que não o fazem (BERRUETA et al., 2007).

Técnicas de reconhecimento de padrões supervisionado também podem ser agrupadas como métodos paramétricos/não paramétricos (MASSART et al., 1997), determinísticos/ probabilísticos ou linear/não linear. Técnicas paramétricas como LDA, PLS-DA, SIMCA e UNEQ usam parâmetros estatísticos da distribuição dos objetos na derivação da função de decisão (geralmente uma distribuição normal multivariada é assumida). A estatística de métodos não paramétricos, como *k*NN e CART não se baseiam na suposição de distribuição, o que torna o cálculo das probabilidades de classificação correta mais difícil. Já a classificação linear/não linear baseia-se na natureza das funções discriminante linear/não linear, respectivamente, usadas para classificação entre as classes (BERRUETA et al., 2007).

2.3.1. Procedimentos preliminares

Os dados analíticos são assumidos ser obtidos por métodos analíticos validados em termos de precisão, exatidão, sensibilidade, especificidade, incerteza, robustez e rastreabilidade. Além disso, a quantidade de dados utilizados para treinamento deve ser suficientemente grande para cobrir uma possível variação conhecida no domínio do problema, de modo que a generalização dos modelos para dados ainda não medidos seja possível (EURACHEM GUIDE, 1998).

Reconhecimento de padrões supervisionado requer um conjunto de treinamento com objetos de categorias conhecidas para se obter um modelo capaz de identificar amostras desconhecidas. Portanto, em primeiro lugar é obrigatório estabelecer se as medições químicas são realmente boas o suficiente para pertencer às classes pré-determinadas, já que as técnicas de reconhecimento de padrões não podem compensar planejamentos experimentais mal projetados ou

dados experimentais inadequados (BRERETON, 2003). Esta tarefa pode ser difícil devido ao fato de que as técnicas analíticas modernas são capazes de gerar tantos dados que a informação útil pode não ser facilmente evidenciada. A análise exploratória de dados usando PCA é comumente usada para simplificar e adquirir um melhor conhecimento sobre os conjuntos de dados. A próxima etapa é, portanto, eliminar o ruído e a redundância, mantendo a informação significativa (SILBERT, 2001).

Neste contexto, regiões espectrais selecionadas, sinais ou comprimentos de onda específicos selecionados costumam oferecer melhores resultados nas classificações do que o uso de toda a informação espectral. Além disso, elas podem ser selecionadas com base no conhecimento de que os sinais analíticos indicam a presença de grupos químicos específicos ou com base em tratamentos quimiométricos (SOHN et al., 2005).

2.3.2. Pré-tratamento de dados

Em geral, o pré-tratamento de dados é necessário antes da aplicação de técnicas de análise de dados multivariados. Dados digitais (como espectros, integrais de pico, posições e mudanças de pico, por exemplo) muitas vezes dependem de como a informação é processada e, por isso, uma especial atenção deve ser dada a sua manipulação. Espectros e cromatogramas são sinais sequenciais no tempo ou frequência, que podem ser processados utilizando métodos diferentes: (i) *suavização*, que elimina o ruído sem alargar excessivamente os picos, (ii) *derivação*, para ressaltar a informação (picos e vales), (iii) *convolução* para o aumento do sinal, (iv) *transformada de Fourier* (FT) para converter dados brutos no "domínio do tempo" para o "domínio da frequência", (v) *transformada*

wavelet, para alisamento ou compressão de dados, (iv) *métodos bayesianos*, (vii) *ortogonalização*, para a remoção de variações sistemáticas dos dados que não estão correlacionadas com o membro de classe. Além disso, diferentes combinações destes métodos também são usadas para o processamento do sinal (BRERETON, 2003; WOLD et al., 1998).

Em química de alimentos, dados brutos usados em tratamentos quimiométricos são geralmente dados químicos ou físicos. Regiões espectrais selecionadas, sinais ou comprimentos de onda específicos frequentemente fornecem melhores resultados em classificação que o uso de toda a informação espectral. Regiões espectrais ou sinais específicos podem ser selecionados com base no conhecimento de sinais que indicam a presença de grupos químicos específicos ou baseados em tratamentos quimiométricos (BRERETON, 2003; BERRUETA et al., 2007).

2.3.3. Análise exploratória dos dados

A técnica principal para análise exploratória dos dados é a Análise de Componentes Principais (PCA), que é frequentemente usada para detectar padrões nos dados medidos (LAVINE, 2006). Uma vez que PCA é uma técnica que reduz a dimensionalidade dos dados, ela permite, tanto quanto possível, a visualização das informações presentes nos dados originais. Então, PCA transforma as variáveis originais medidas em novas variáveis não correlacionadas chamadas *componentes principais*. Cada componente principal é uma combinação linear das variáveis originais medidas. Esta técnica dá origem a um conjunto de eixos ortogonais que representam as direções de maior variação nos dados. A primeira componente principal (PC1) representa o máximo da variância total, enquanto a segunda (PC2)

não está correlacionada com a primeira e representa o máximo da variância residual, e assim por diante, até que a variância total seja contabilizada. Por razões práticas, é suficiente para reter apenas os componentes que representam uma grande porcentagem da variância total. Os coeficientes lineares da relação inversa das combinações lineares são chamados pesos (*loadings*), ou seja, os coeficientes de correlação entre as variáveis originais e as componentes principais. Os valores que representam as amostras no espaço definido pelas componentes principais são os escores (*scores*). Os escores podem ser utilizados como entrada para outras técnicas de análise multivariada, em vez das variáveis originais medidas (KEMSLEY, 1996).

Outras técnicas de reconhecimento de padrões podem ser utilizadas para avaliação preliminar da informação contida nas matrizes de dados, tais como a análise de agrupamento hierárquico (HCA) (MØLLER et al., 2005). Em HCA, as amostras são agrupadas com base em semelhanças, sem levar em conta as informações sobre a categoria da classe (LAVINE, 2006). A HCA agrupa as amostras de acordo com uma métrica de similaridade, a qual pode ser a distância, a correlação ou uma combinação de ambas. Esta técnica baseia-se na ideia de que a similaridade é inversamente proporcional à distância entre as amostras. Assim, HCA calcula as distâncias (ou correlação) entre todas as amostras utilizando uma métrica definida como a distância Euclidiana, distância de Mahalanobis, etc. O agrupamento das amostras pode ser realizado por diferentes algoritmos de agrupamento, dependendo dos critérios considerados para definir a distância entre dois grupos (regra de ligação): ligação simples (vizinho mais próximo), completa (vizinho mais distante) ou média, método centroide, método de Ward, etc. Um pré-tratamento dos dados é necessário, de modo a evitar o efeito de diferentes escalas das variáveis

([BERRUETA et al., 2007](#)). Outras informações sobre técnicas de análise exploratória de dados (por exemplo, o algoritmo de PCA, a definição das medidas de distância e algoritmos de agrupamento) podem ser encontradas em ([MASSART et al., 1997](#); [OTTO, 1999](#)).

2.3.4. Redução e seleção de variáveis

A seleção de variáveis é um passo preliminar usado na análise multivariada de dados. Se o número de objetos é relativamente pequeno, o número de variáveis é grande e muitas destas variáveis contêm informação redundante ou ruidosa. Nestes casos, um procedimento de seleção de variáveis ou características é necessário, a fim de não incorrer no problema de sobreajuste (*overfitting*). O sobreajuste ocorre quando o ruído é modelado e então o modelo perde a sua capacidade de generalização ([BERRUETA et al., 2007](#)). A seleção de variáveis pode ter várias vantagens, tais como a remoção de não linearidade e de ruído, quando comparado com o uso do espectro completo ([ANDERSEN e BRO, 2010](#)).

A seleção de variáveis para análise discriminante seleciona um subconjunto de variáveis que são as mais discriminatórias. Para esta finalidade, é necessária uma técnica de reconhecimento de padrões que forneça um modelo matemático e uma função objetivo que guia a busca. A função objetivo avalia a qualidade dos subconjuntos de variáveis provisoriamente selecionados durante a busca do melhor modelo, sendo absolutamente fundamental para se obter soluções adequadas. Além disso, deve ser calculada uma estimativa de erro de predição. Em quimiometria, as funções objetivo mais utilizadas são baseadas em métodos de validação cruzada ([BERRUETA et al., 2007](#)).

Os algoritmos mais comumente descritos na literatura para a seleção de variáveis em dados químicos incluem Algoritmo Genético (HOLLAND, 1975), Recozimento Simulado (HORCHNER e KALIVAS, 1995), Busca Tabu (HAGEMAN et al., 2003), Colônia de Formigas (SHAMSIPUR et al., 2003), Algoritmo de Projeções Sucessivas (SPA) (PONTES et al., 2005), dentre outros.

Em redução de variáveis, o número de variáveis é reduzido através da combinação das variáveis originais alcançando um número menor de variáveis latentes (que são variáveis derivadas, em vez de medidas), como é feito, por exemplo, em Mínimos Quadrados Parciais (PLS) (MASSART et al., 1997). Esta abordagem é útil quando o número de variáveis excede o número de objetos. Um determinado número de componentes principais é extraído, suprimindo os de ordem mais elevada, reduzindo assim o ruído, em certa medida, e, em seguida, um reconhecimento de padrões supervisionado é aplicado. A eliminação de componentes principais tem que ser feita com cuidado para que informações importantes para a classificação não sejam perdidas. A abordagem mais simples para determinar o número de componentes significativos é através da medição do erro de predição e avaliando, em seguida, a soma dos quadrados dos erros residuais de previsão (PRESS) ou a raiz do erro quadrático médio (RMSE), que também pode ser expressa como uma percentagem de variância. Este erro pode ser utilizado para decidir quantas componentes devem ser incluídas no modelo. A decisão pode ser tomada: por meio de um erro percentual de corte padrão; ignorando as componentes a partir das quais o erro caiu para o nível de ruído; considerando as componentes até que o erro atinja um patamar; selecionando PCs cujos autovalores são iguais ou maiores que 1; ou procurando por uma ruptura em um gráfico de autovalor versus o número de componentes (BRERETON, 2003).

2.3.5. Técnicas de reconhecimento de padrões supervisionadas

Técnicas de reconhecimento de padrões supervisionadas são aplicadas a uma ampla variedade de dados químicos com diversas finalidades, tais como identificar perfis, impressões digitais, autenticação, detecção de falsificação, avaliação da qualidade dos alimentos, interpretação de dados, etc. (LAVINE, 2006). Nas próximas seções serão abordadas apenas as três técnicas empregadas nesta Tese de Doutorado.

2.3.5.1. SIMCA

A Modelagem Independente e Flexível por Analogia de Classe (SIMCA) é a mais utilizada das técnicas de modelagem de classe. Em SIMCA, cada categoria é modelada de forma independente utilizando PCA e pode ser descrita por um número diferente de componentes principais (PCs). O número de PCs para cada classe no conjunto de treinamento é determinado por validação cruzada. Desta forma, um número suficiente de componentes principais é mantido de modo a ser responsável pela maior parte da variação dentro de cada classe, embora assegurando uma elevada relação sinal-ruído sem incluir no modelo de classe as chamadas componentes principais secundárias ou que contenham apenas ruído (LAVINE, 2000).

SIMCA determina a distância de classe e os poderes de modelagem e de discriminação (BRERETON, 2003). A distância de classe pode ser calculada como a distância geométrica a partir dos modelos de componentes principais. Outra abordagem considera que cada classe é delimitada por uma região do espaço, o que representa uma percentagem do nível de confiança (geralmente 95%) que um determinado objeto pertence a uma classe. Assim, os erros podem ser de dois tipos:

erro tipo I, que consiste de uma amostra não incluída na sua própria classe; e erro tipo-II, quando uma amostra é incluída em uma classe incorreta. Para este trabalho, nós consideramos se a amostra apresentava algum erro, independente de ser tipo I ou tipo II. O poder de classificação mede o quão bem uma variável discrimina entre duas classes. Isso difere do poder de modelagem no sentido em que se uma variável é capaz de modelar bem uma classe, não implica necessariamente que ela seja capaz de discriminar dois grupos de forma eficaz. Assim, os gráficos de escores e pesos da PCA realizada no conjunto de treinamento fornecem informações sobre os *outliers*, subagrupamentos e estruturas dentro da classe (MASSART et al., 1997).

2.3.5.2.PLS-DA

A modelagem por Mínimos Quadrados Parciais (PLS) é um método de projeção multivariada para modelar uma relação entre variáveis dependentes (Y) e variáveis independentes (X). O princípio de PLS é encontrar as componentes na matriz de entrada (X) que descrevem tanto quanto possível as variações significativas nas variáveis de entrada e, ao mesmo tempo, têm correlação máxima com o valor alvo em Y , dando menos importância às variações que são irrelevantes ou ruidosas. Assim, os modelos PLS encontram simultaneamente as variáveis latentes em X que irão prever as variáveis latentes em Y . PLS maximiza a covariância entre as matrizes X e Y . A Análise Discriminante por Mínimos Quadrados Parciais, também conhecida como PLS Discriminante (DPLS), tem como objetivo encontrar as variáveis e direções no espaço multivariado que discriminam as classes estabelecidas no conjunto de calibração. O número ótimo de variáveis latentes pode ser estimado por meio de validação cruzada ou conjunto de teste externo (BERRUETA et al., 2007).

Em PLS-DA, uma matriz Y "simulada" é construída com zeros e uns. A matriz consiste em quantas colunas existem classes e uma observação tem o valor 1 para a classe a que pertence e 0 para o resto. A matriz X é composta dos dados originais (pré-processados). As matrizes X e Y são decompostas num produto de duas outras matrizes de escores e pesos. Em contraste com a PCA que só utiliza a informação de matriz X , PLS também leva em conta a informação na matriz Y . Assim, os pesos de X são calculados a partir dos escores de Y , enquanto que os pesos de Y são determinados a partir dos escores de X . As decomposições não são independentes, mas relacionados com os escores de X e Y . Então, um modelo é desenvolvido para cada classe. Quanto mais próximo um elemento de uma determinada coluna em Y está de 1 e os elementos das outras colunas está de 0, maior a probabilidade de um objeto ser um membro de uma classe particular (BERRUETA et al., 2007).

Uma característica importante de PLS é que ele leva em conta erros em ambas as matrizes, X e Y , e assume que eles são igualmente distribuídos. Além disso, PLS é adequado para conjuntos de dados com menos objetos do que variáveis e quando há um alto grau de intercorrelação entre as variáveis independentes (BERRUETA et al., 2007).

2.3.5.3.LDA

A Análise Discriminante Linear (LDA) (FISHER, 1936) é provavelmente o método de reconhecimento de padrões supervisionado mais utilizado e melhor estudado. LDA baseia-se na determinação de funções discriminantes lineares, que maximizam a razão da variância entreclasses e minimiza a razão da variância intraclasse. Em LDA, as classes são ditas seguir uma distribuição normal

multivariada e ser linearmente separadas. LDA e PCA podem ser consideradas como métodos de redução de variáveis/características, no sentido em que ambas determinam um hiperplano de dimensão menor no qual os pontos serão projetados a partir da dimensão mais elevada. No entanto, enquanto PCA seleciona uma direção que retém a estrutura máxima entre os dados em uma dimensão inferior, LDA seleciona uma direção que consegue a separação máxima entre as classes dadas. A variável latente obtida em LDA é uma combinação linear das variáveis originais. Esta função é chamada variável canônica e os seus valores são as raízes. Dadas k classes, $k - 1$ variáveis canônicas podem ser determinadas se o número de variáveis é maior do que k (MASSART et al., 1997; BERRUETA et al., 2007).

LDA é uma variante de análise discriminante em que as fronteiras de classificação são lineares. LDA exige que as matrizes de variância–covariância das classes estabelecidas possam ser agrupadas. Isto só é possível quando estas matrizes podem ser consideradas iguais, o que significa que 95% dos elipsoides de confiança têm um volume igual (variância) e orientação no espaço (covariância) (MASSART et al., 1997).

Uma vez que LDA está sujeita a uma restrição de que o número de variáveis não deve exceder o número de objetos, uma solução para este problema consiste em aplicar métodos de redução de variáveis, tais como PLS-DA, ou a seleção de variáveis (MASSART et al., 1997). Neste contexto, o Algoritmo de Projeções Sucessivas (SPA) (PONTES et al., 2005) tem sido aplicado com sucesso como técnica de seleção de variáveis associado à Análise Discriminante Linear para classificar óleos vegetais e diesel (PONTES et al., 2005), solos (PONTES et al., 2009), óleos vegetais comestíveis (GAMBARRA-NETO et al., 2009; SOARES et al., 2010), cigarros (MOREIRA et al., 2009), cafés (SOUTO et al., 2010), misturas de

biodiesel/diesel (PONTES et al., 2011), cervejas (GHASEMI-VARNAMKHASTI et al., 2012), tintas de canetas azuis (SILVA et al., 2012a), etanol combustível (SILVA et al., 2012b), chás (DINIZ et al., 2012) e sucos cítricos (FREITAS et al., 2013).

2.3.5.3.1. SPA-LDA

O algoritmo de SPA-LDA (SOARES et al., 2013) visa selecionar um subconjunto de variáveis com pequena colinearidade e poder discriminatório adequado para uso em problemas de classificação envolvendo $C \geq 2$ classes diferentes. Para este propósito, assume-se que um conjunto de treinamento de N objetos com rótulos de classes conhecidas está disponível para orientar o processo de seleção de variáveis. No caso de dados espectroscópicos, por exemplo, cada objeto consiste de um espectro registado em K números de onda (ou comprimentos de onda) (GHASEMI-VARNAMKHASTI et al., 2012).

O algoritmo SPA-LDA pode ser dividido em duas fases principais. Na 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. As variáveis subsequentes da cadeia são selecionadas de modo a apresentar a menor colinearidade (avaliada pelas operações de projeção) com as variáveis precedentes (GHASEMI-VARNAMKHASTI et al., 2012).

Na fase 2, diferentes subconjuntos de variáveis são extraídos das cadeias e, em seguida, avaliados em termos de uma função de custo G definida como

$$G = \frac{1}{N} \sum_{n=1}^N g_n \quad (2.2)$$

onde

$$g_n = \frac{r^2[\mathbf{x}_n, \bar{\mathbf{x}}(I_n)]}{\min_{I_j \neq I_n} r^2[\mathbf{x}_n, \bar{\mathbf{x}}(I_j)]} \quad (2.3)$$

Na Eq. (2.3), o numerador $r^2[\mathbf{x}_n, \bar{\mathbf{x}}(I_n)]$ é a distância de Mahalanobis (MAESSCHALCK et al., 2000) 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 como

$$r^2[\mathbf{x}_n, \bar{\mathbf{x}}(I_n)] = [\mathbf{x}_n - \bar{\mathbf{x}}(I_n)] \mathbf{S}^{-1} [\mathbf{x}_n - \bar{\mathbf{x}}(I_n)]^T \quad (2.4)$$

onde \mathbf{S} é uma matriz de covariância conjunta, a qual é calculada de acordo com o procedimento LDA padrão (WU et al., 1996). O denominador na Eq. (2.3) corresponde à distância de Mahalanobis ao quadrado entre o objeto \mathbf{x}_n e o centro da classe errada mais próxima. Um valor pequeno de 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 (GHASEMI-VARNAMKHAZI et al., 2012).

Capítulo 3

Classificação de chás usando imagens digitais e quimiometria





CLASSIFICAÇÃO DE CHÁS USANDO IMAGENS DIGITAIS E QUIMIOMETRIA

3.1. EXPERIMENTAL

3.1.1. Amostras

Foram adquiridas 100 amostras de chás em supermercados das cidades de João Pessoa (Brasil) e Bahía Blanca (Argentina), das quais eram 40 amostras de chá verde (sendo 20 brasileiras e 20 argentinas) e 60 amostras de chá preto (sendo 20 brasileiras, 20 argentinas e 20 srilanquenses). O sistema de embalagem das amostras de chás é composto por um filme polimérico, uma caixa de papelão e sachês (opcionais) de papel ou de alumínio termolacrados contendo os saquinhos.

Foram adquiridos cinco lotes diferentes de cada marca, de modo assegurar a representatividade das amostras. Então, uma etapa de quarteio das amostras foi realizada de modo a reduzir a massa total do material a uma porção representativa e homogênea para ser analisada no laboratório (**Fig. 3.1a**). Para isso, para cada lote o conteúdo de 100 saquinhos de chá, de aproximadamente 2 g cada, foi reunido em uma bandeja e, em seguida, homogeneizado. A partir dos 200 g de amostras de chás do lote selecionado iniciou-se a processo de quarteio, onde a amostra foi acomodada numa pilha uniforme sobre uma superfície limpa e seca. A espessura foi a mesma em todos os quarteios realizados. Em seguida, a amostra foi dividida em quatro partes aproximadamente iguais. Dois quartos opostos são escolhidos e seus conteúdos são então misturados e novamente quarteados de maneira análoga à anterior. O procedimento é repetido até que a amostra final contenha 25 g de chá (**Fig. 3.1b**), que é armazenada em sacolas plásticas com vedação hermética (lacre) para conservá-las de forma adequada, evitando contaminação e/ou adulteração.

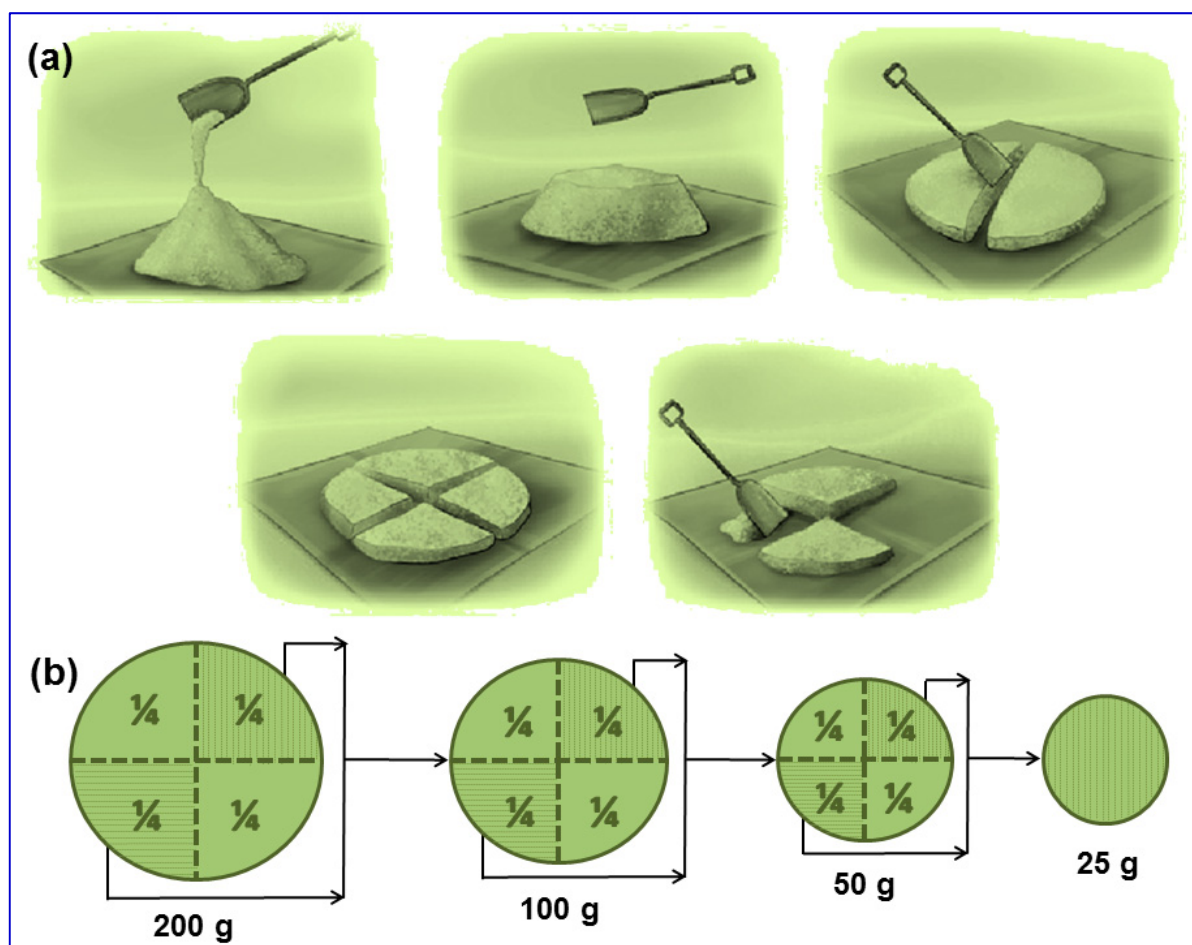


Figura 3.1. (a) Exemplo das etapas do quarterio de uma amostra de chá; (b) sucessivos quarterios para redução e homogeneização da amostra. Adaptado de: <http://dc392.4shared.com/doc/mIHB4ime/preview004.png>

3.1.2. Aquisição de imagens

O método de aquisição de imagens leva em consideração as características visuais globais da superfície da amostra. Todas as amostras foram fotografadas usando uma câmera digital, modelo Olympus SP-510 UZ, com 7.1 megapixels de resolução. Para alcançar uma padronização das imagens, os fabricantes resolvem questões de balanço das cores vermelho-verde-azul (RGB), matiz (H), saturação (S) e brilho (B), bem como a intensidade de tons de cinza, de forma particular para cada modelo de câmera vendido. Estas considerações dos fabricantes não afetam o método global, mas os escores individuais para cada valor RGB, HSB e tons de

cinza. As imagens de cada amostra foram obtidas em triplicata, a partir das quais foram obtidos histogramas médios de cada amostra.

Para garantir a reprodutibilidade no processo de aquisição de imagens, um suporte metálico (**Fig. 3.2**) que abriga a câmera digital e uma lâmpada fluorescente circular comercial de luz branca (40 W) foi montado de tal forma que foram mantidos fixos o posicionamento, a luminosidade, a distância amostra-câmera e o foco. A câmera digital foi fixada verticalmente acima do suporte onde as amostras de chá foram acomodadas. A distância entre a lente e a amostra foi de 12 cm e entre a lâmpada e a amostra foi de 10 cm. A câmera digital foi posicionada no centro da lâmpada fluorescente circular. Um porta-amostras de politetrafluoroetileno (PTFE) foi utilizado de modo a minimizar os efeitos de espalhamento de luz e fluorescência e, conseqüentemente, seus efeitos sobre os histogramas de cor das imagens.

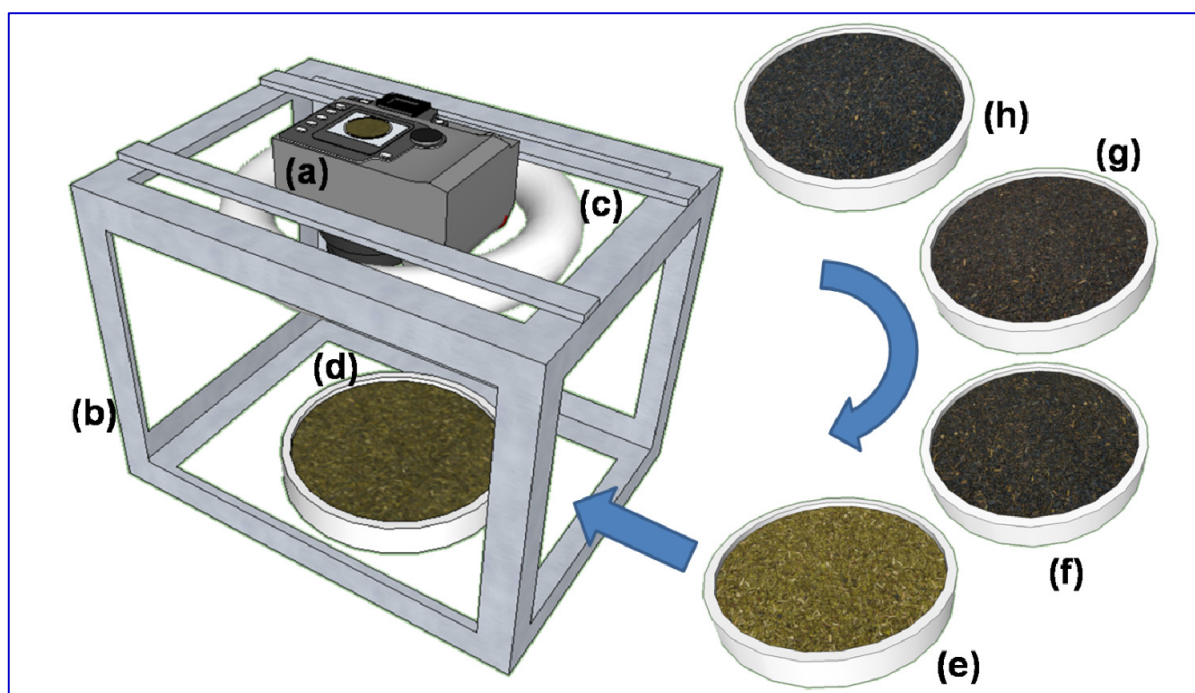


Figura 3.2. Suporte construído para captura das imagens de chá. (a) Câmera digital, (b) suporte metálico, (c) lâmpada circular, (d) chá verde brasileiro, (e) chá verde argentino, (f) chá preto brasileiro, (g) chá preto argentino e (h) chá preto srilanquense.

3.1.3. Histogramas de cor e análise de dados

Histogramas de cor (RGB, HSB e escala de cinza) foram obtidos a partir de cada imagem digital usando o software gratuito ImageJ 1.44p. A **Fig. 3.3** mostra um exemplo de cada tipo de histograma (e sua imagem correspondente) obtido para uma amostra de chá verde.

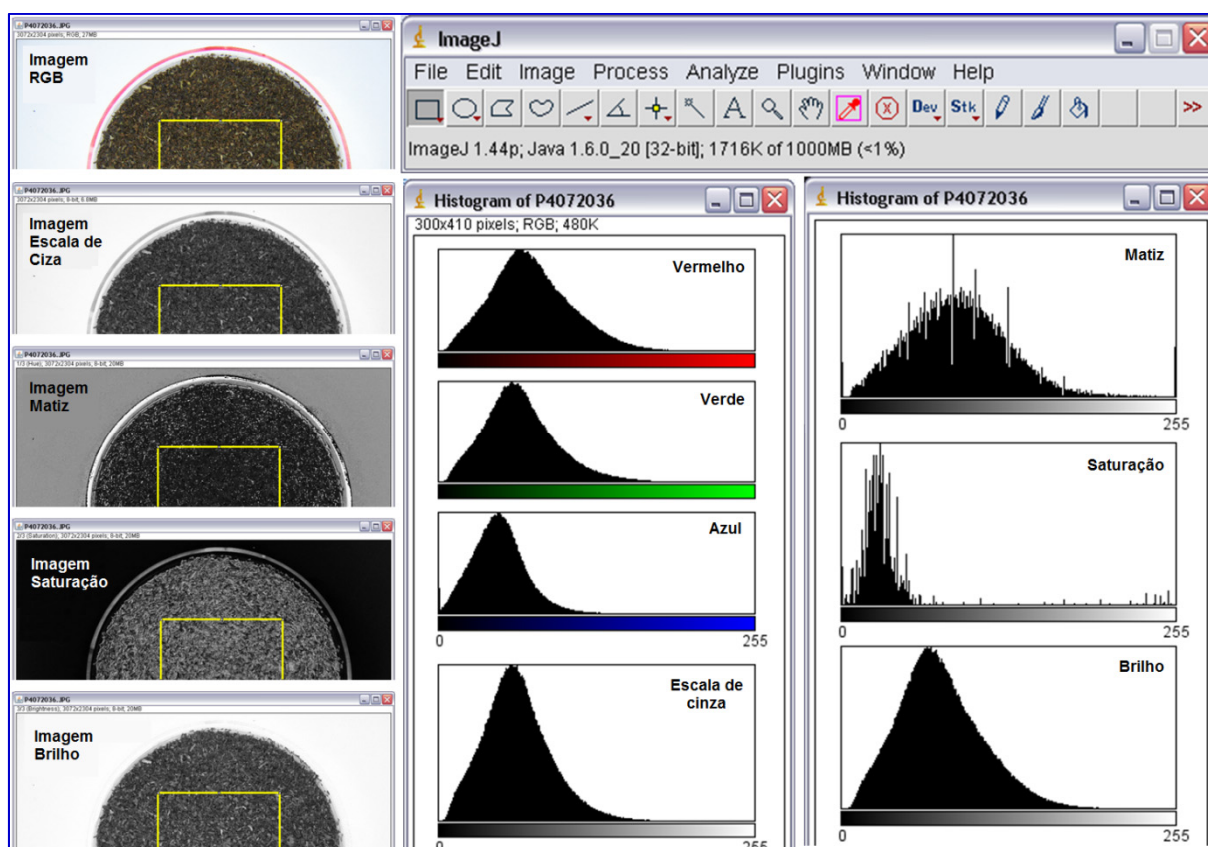


Figura 3.3. Exemplo da janela de comandos do programa ImageJ e dos histogramas de cor e imagens de uma amostra de chá verde.

Para o tratamento de dados foi definida uma área quadrada de 1000 x 1000 pixels a partir do centro de cada imagem, o que representa aproximadamente 14% da área total da imagem. Usando apenas a região selecionada das imagens, foram construídos histogramas empregando vermelho, verde, azul, matiz, saturação, brilho e tons de cinza. Cada componente de cor dos modelos é composto de 256 tons, que são utilizados como informação analítica. A fim de verificar se há influência relativa



de cada cor, foram selecionados quatro diferentes modelos de cor empregando (a) RGB, (b) HSB, (c) em tons de cinza, (d) RGB + cinza + HSB. Os modelos foram compostos de (a) 3 x 256, (b) 3 x 256, (c) 1 x 256 e d) 768 + 256 + 768 variáveis, respectivamente. Histogramas médios foram obtidos a partir de três fotos para cada amostra.

A informação analítica extraída dos histogramas supracitados foi utilizada para a construção de modelos quimiométricos de classificação utilizando SIMCA, PLS-DA e SPA-LDA. Foi avaliada a classificação das amostras de chás em cinco diferentes classes: (a) chá preto argentino, (b) chá preto brasileiro, (c) chá preto srilanquense, (d) chá verde argentino e (d) chá verde brasileiro. Os dados obtidos a partir de cada histograma foram separados em conjuntos de treinamento (60), validação (20) e teste (20) usando o algoritmo Kennard-Stone (KS) (KENNARD e STONE, 1969), que foi aplicado separadamente para cada classe. O algoritmo KS é um método clássico usado para extrair um conjunto representativo de objetos a partir de um determinado conjunto de dados maximizando a distância euclidiana mínima entre os objetos já selecionados e os demais objetos. Portanto, cada classe foi composta de 20 histogramas de chá, separados em conjuntos de treinamento (12), validação (4) e teste (4). PCA, SIMCA e PLS-DA foram calculados usando o programa The Unscrambler 9.7 (CAMO S/A) e os algoritmos KS e SPA-LDA foram executados usando o programa Matlab 2009b (Mathworks Inc.).

3.2. RESULTADOS E DISCUSSÃO

3.2.1. Análise de componentes principais

Uma vez que há variação nas cores das partículas individuais do chá, as imagens são caracterizadas por uma textura estocástica natural, isto é, elas nos

permitem extrair informações sobre o arranjo espacial de cores e intensidades. Histogramas de cor descrevem a distribuição estatística dos pixels como uma função da componente de cor registrada e não as propriedades físico-químicas diretamente. A **Fig. 3.4** mostra um histograma médio de todas as amostras de chá contendo os modelos de cor avaliados.

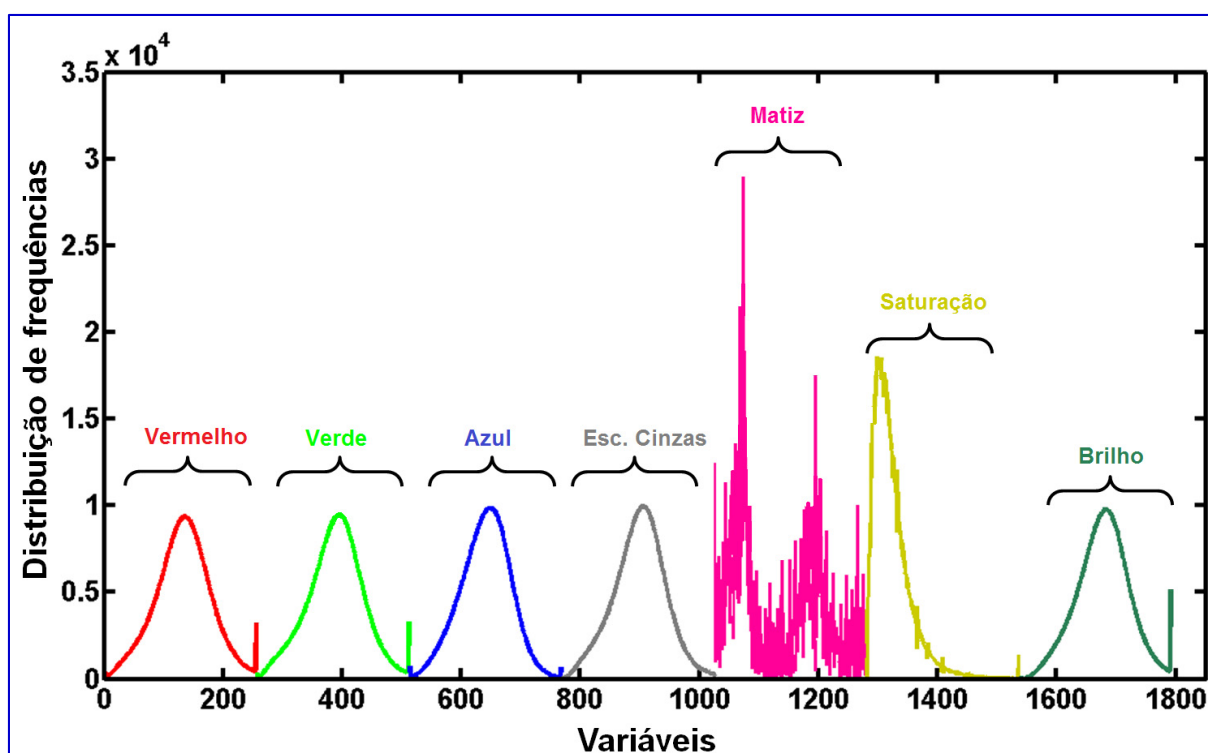


Figura 3.4. Histogramas médio de todas as amostras de chá contendo os modelos de cor avaliados.

Uma análise exploratória dos dados foi realizada utilizando Análise de Componentes Principais para os modelos de cor RGB, HSB e escala de cinza individualmente e um modelo utilizando todos os histogramas juntos (RGB + escala de cinza + HSB). A **Fig. 3.5** mostra os gráficos de escores das duas primeiras PCs utilizando os quatro diferentes modelos de cor.

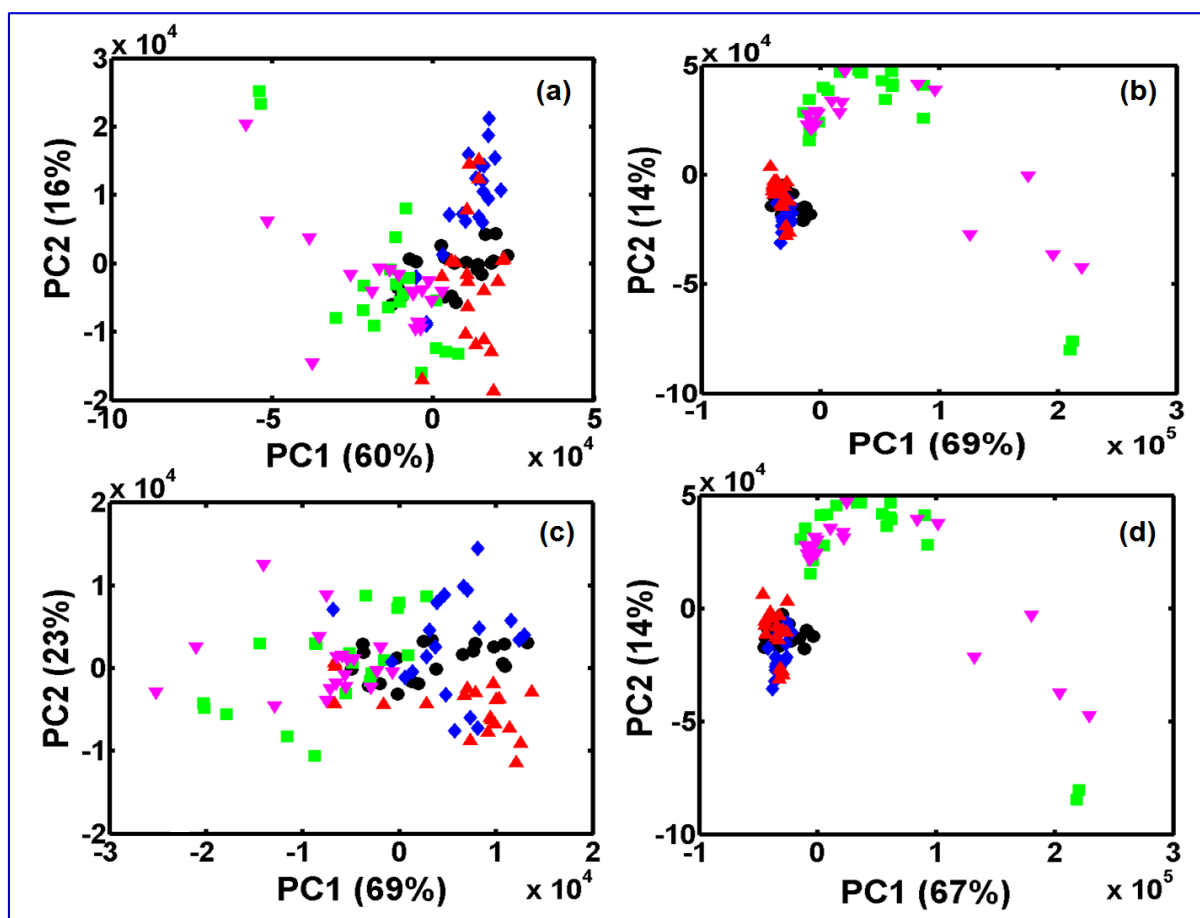


Figura 3.5. Escores de PCA obtidos a partir dos histogramas de (a) RGB, (b) HSB, (c) escala de cinza e (d) RGB + escala de cinza + HSB para todas as amostras de chás. (▼) Chá verde brasileiro, (▲) chá preto brasileiro, (■) chá verde argentino, (◆) chá preto argentino, (●) chá preto srilanquense.

Os gráficos de escores para os modelos RGB (Fig. 3.5a) e escala de cinza (Fig. 3.5c) apresentam forte sobreposição entre as amostras e nenhuma tendência de separação. Já os gráficos de escores para os modelos HSB (Fig. 3.5b) e RGB + escala de cinza + HSB (Fig. 3.5d) apresentam resultados semelhantes, separando as amostras em dois grandes conjuntos de chá preto e verde, mas sobreposições ainda permanecem. Isso era esperado uma vez que os chás verde e preto são muito diferentes em cor, textura da superfície, tamanho de partículas, etc. As três classes de chás pretos se sobrepõem mais que as duas de chás verdes. Isso ocorre porque há repetição das informações nas imagens, o que requer seleção de variáveis para

encontrar quais tons de cor conduzem a um melhor modelo de classificação. Neste caso, SPA-LDA foi usado para resolver este problema e, conseqüentemente, classificar as amostras de chá.

3.2.2. Classificação

3.2.2.1. SIMCA

Modelos SIMCA foram construídos individualmente para cada classe de chá estudada utilizando a série de teste como técnica de validação. A **Tab. 3.1** apresenta os resultados da classificação SIMCA para as amostras de chá do conjunto de teste a um nível de significância para o teste-*F* de 5%. O número de PCs adotado para cada modelo é indicado entre parênteses.

Tabela 3.1. Erros de classificação SIMCA para as amostras de chá do conjunto de teste usando quatro diferentes histogramas de cor (RGB, HSB, Escala de cinza e RGB + escala de cinza + HSB) a um nível de significância de 5% para o teste-*F*.

Classe	ArP				BrP				SrP				ArV				BrV			
Modelo de cor	RGB (7)	HSB (7)	Cinza (8)	Todos (8)	RGB (5)	HSB (6)	Cinza (7)	Todos (6)	RGB (8)	HSB (7)	Cinza (7)	Todos (7)	RGB (7)	HSB (9)	Cinza (6)	Todos (6)	RGB (5)	HSB (6)	Cinza (8)	Todos (7)
ArP	4	4	4	4	3	1	4	4	4	4	3	3	-	-	-	-	-	-	-	-
BrP	4	3	3	3	4	4	4	4	4	3	3	3	-	-	-	-	-	-	-	-
SrP	-	-	-	-	-	-	-	-	4	4	4	4	-	-	-	-	-	-	-	-
ArV	-	-	-	-	-	-	-	-	-	-	-	-	4	4	4	4	4	4	4	4
BrV	-	-	-	-	-	-	-	-	-	-	-	-	3	4	2	2	4	4	4	4

ArP: chá preto argentino; BrP: chá preto brasileiro; SrP: chá preto srilanquense; ArV: chá verde argentino; BrV: chá verde brasileiro. O número de PCs adotado para cada modelo é indicado entre parênteses.

Para SIMCA, todas as amostras de chás do conjunto de teste foram classificadas corretamente dentro de suas respectivas classes (erros tipo-I);

entretanto, a maioria das amostras foi classificada incorretamente em outras classes (erros tipo-II). Isto pode ser atribuído à elevada sobreposição entre as classes, atingindo erros de classificação acima de 75% para todas as classes estudadas independentemente do modelo de cor utilizado. Por outro lado, observa-se que SIMCA foi capaz de classificar as amostras de acordo com o tipo de chá (verde ou preto), mas não segundo a origem geográfica (Argentina, Brasil e Sri Lanka).

3.2.2.2. PLS-DA

Um único modelo PLS foi construído para as cinco classes de chá e o número de fatores foi selecionado empregando um conjunto externo de validação. Um valor de corte de 0,5 foi adotado para inclusão de uma amostra em uma dada classe. Na **Tab. 3.2** são apresentados os erros de classificação PLS-DA para as amostras de chá do conjunto de teste usando quatro diferentes histogramas de cor. O número de variáveis latentes é indicado em parênteses.

Tabela 3.2. Erros de classificação PLS-DA para as amostras de chá do conjunto de teste usando quatro diferentes histogramas de cor (RGB, HSB, Escala de cinza e RGB + escala de cinza + HSB).

Classe	ArP				BrP				SrP				ArV				BrV			
	RGB (7)	HSB (9)	Cinza (8)	Todos (8)	RGB (7)	HSB (9)	Cinza (8)	Todos (8)	RGB (7)	HSB (9)	Cinza (8)	Todos (8)	RGB (7)	HSB (9)	Cinza (8)	Todos (8)	RGB (7)	HSB (9)	Cinza (8)	Todos (8)
ArP	1	2	-	1	-	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-
BrP	-	1	-	-	3	2	-	2	1	1	1	-	-	-	-	-	-	-	-	-
SrP	1	1	-	1	-	-	-	-	1	2	-	2	-	-	-	-	-	-	-	-
ArV	-	-	-	-	-	-	-	-	-	-	-	-	3	3	3	4	1	-	1	-
BrV	-	1	-	-	-	-	-	-	-	1	-	-	-	-	-	-	3	3	2	4

ArP: chá preto argentino; BrP: chá preto brasileiro; SrP: chá preto srilanquense; ArV: chá verde argentino; BrV: chá verde brasileiro. O número de variáveis latentes é indicado em parênteses.

Com base na **Tab. 3.2** pode-se verificar que as amostras de chás verdes foram mais bem classificadas do que as amostras de chás pretos dentro de suas respectivas classes independentemente do modelo de histogramas de cor utilizado. Comparando-se os resultados em relação aos modelos de histogramas de cor, o modelo RGB + escala de cinza + HSB obteve o melhor resultado, classificando corretamente 13 das 20 amostras do conjunto de teste dentro de suas respectivas classes e apresentando apenas 1 erro do tipo II. Considerando-se conjuntamente os erros tipo I e tipo II, uma média de 80% de classificação correta foi obtida para o conjunto de teste usando o modelo que contém todos os histogramas de cor.

3.2.2.3. SPA-LDA

Na **Tab. 3.3** são apresentados os erros de classificação SPA-LDA para as amostras de chá do conjunto de teste usando os quatro diferentes histogramas de cor estudados.

Tabela 3.3. Erros de classificação SPA-LDA para as amostras de chá do conjunto de teste usando quatro diferentes histogramas de cor (RGB, HSB, Escala de cinza e RGB + escala de cinza + HSB).

Classe	ArP				BrP				SrP				ArV				BrV			
	RGB (20)	HSB (16)	Cinza 1(2)	Todos (17)	RGB (20)	HSB (16)	Cinza 1(2)	Todos (17)	RGB (20)	HSB (16)	Cinza 1(2)	Todos (17)	RGB (20)	HSB (16)	Cinza 1(2)	Todos (17)	RGB (20)	HSB (16)	Cinza 1(2)	Todos (17)
ArP	4	4	3	4	1	-	2	-	-	-	-	-	-	-	-	-	-	-	1	-
BrP	-	-	1	-	3	4	2	4	-	-	-	-	-	-	-	-	-	-	1	-
SrP	-	-	-	-	-	-	-	-	4	4	3	4	-	-	-	-	-	-	1	-
ArV	-	-	-	-	-	-	-	-	-	-	-	-	4	3	4	4	2	-	-	-
BrV	-	-	-	-	-	-	-	-	-	-	1	-	-	1	-	-	2	4	1	4

ArP: chá preto argentino; BrP: chá preto brasileiro; SrP: chá preto srilanquense; ArV: chá verde argentino; BrV: chá verde brasileiro. Em parênteses é indicado o número de variáveis selecionadas pelo SPA-LDA.

Como se pode observar na **Tab. 3.3**, SPA-LDA melhorou significativamente a capacidade de classificação usando os quatro diferentes modelos de cor estudados quando comparados com os resultados obtidos por SIMCA e PLS-DA. O melhor resultado foi obtido para o modelo de cor RGB + escala de cinza + HSB, que classificou corretamente todas as amostras de chás estudadas usando apenas 17 variáveis, conforme indicado na **Fig. 3.6**.

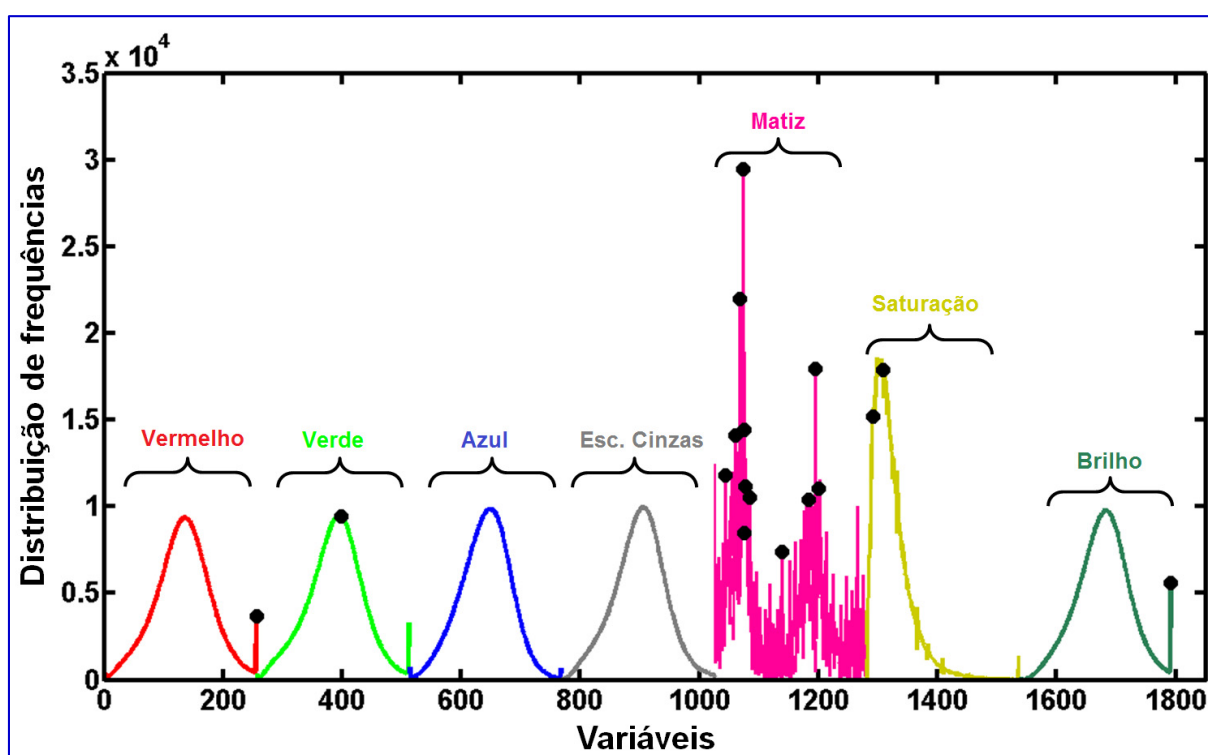


Figura 3.6. As 17 variáveis selecionadas (●) pelo SPA para o modelo RGB + escala de cinza + HSB.

A **Fig. 3.7** mostra os gráficos das funções discriminantes obtidas pelo SPA-LDA usando o modelo de cor RGB + escala de cinza + HSB. Observa-se uma excelente classificação entre os chás verde em dois agrupamentos (brasileiros e argentinos), e em relação aos chás pretos, a classe de chá srilanquense foi separada dos grupos de chás preto brasileiro e argentino em DFs superiores.

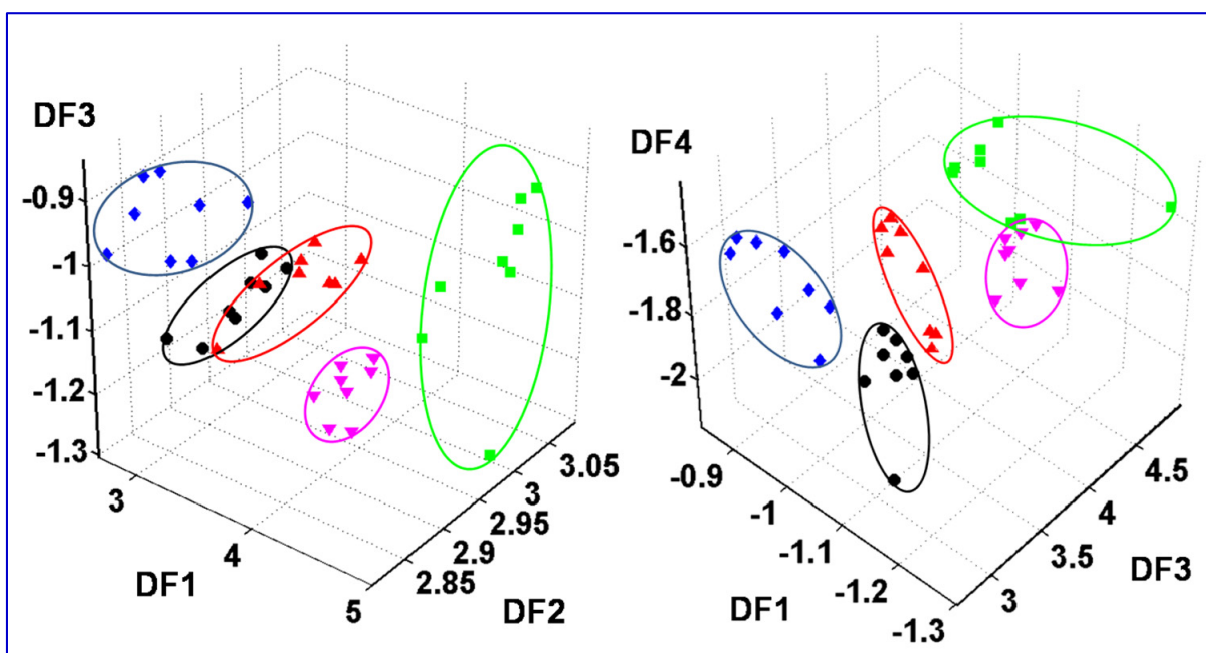


Figura 3.7. Funções discriminantes obtidas por SPA-LDA na classificação das amostras dos conjuntos de validação e teste usando o modelo de cor RGB + escala de cinza + HSB. (▼) Chá verde brasileiro, (▲) chá preto brasileiro, (■) chá verde argentino, (◆) chá preto argentino, (●) chá preto srilanquense.

Capítulo 4

Classificação de chás usando espectroscopia NIR e quimiometria



CLASSIFICAÇÃO DE CHÁS USANDO ESPECTROSCOPIA NIR E QUIMIOMETRIA

4.1. EXPERIMENTAL

4.1.1. Amostras

Foram adquiridas 82 amostras de chás em supermercados das cidades de João Pessoa (Brasil) e Bahía Blanca (Argentina), das quais eram 31 amostras de chá verde (sendo 16 brasileiras e 15 argentinas) e 51 amostras de chá preto (sendo 15 brasileiras, 20 argentinas e 16 srilanqueses). Foram adquiridos cinco lotes diferentes de cada marca, de modo assegurar a representatividade das amostras. O sistema de embalagem e os procedimentos de quarteio das amostras foram análogos aos descritos na [Seção 3.3.1](#). Após o quarteio, as amostras de chá foram moídas adequadamente em um moinho manual de mesa a fim de homogeneizar o tamanho das partículas.

4.1.2. Medidas dos espectros NIR

Os espectros NIR (3700 – 14000 cm^{-1}) das amostras de chás estudadas foram adquiridos em duplicata usando um espectrofotômetro FTIR da Bomem[®], modelo FTLA2000-160, a temperatura ambiente (23 ± 1 °C). Cada espectro foi obtido a partir de uma média de 32 varreduras a uma resolução de 8 cm^{-1} . O espectro de uma pastilha de PTFE foi usado como branco.

4.1.3. Procedimento quimiométrico

As amostras foram divididas em conjuntos de treinamento e teste aplicando o algoritmo Kennard-Stone (KS) ([Kennard & Stone, 1969](#)). O número de amostras

em cada conjunto é apresentado na **Tab. 4.1**. As amostras de treinamento foram utilizadas para os procedimentos de modelagem (incluindo seleção de variáveis para o modelo SPA-LDA e determinação das componentes principais em cada modelo SIMCA) e validadas usando validação cruzada completa. As amostras teste foram utilizadas apenas na avaliação final para comparação dos resultados da classificação (Soares et al., 2013). Derivação Savitzky-Golay dos espectros, PCA, PLS-DA e SIMCA foram realizadas usando o programa The Unscrambler® 9.6 (CAMO S.A.). Quatro diferentes níveis de significância (1%, 5%, 10% e 25%) para o teste-*F* da classificação SIMCA foram testados. Os algoritmos KS e SPA-LDA foram calculados em Matlab ® 6.5.

Tabela 4.1. Número de amostras de treinamento e teste em cada classe estudada.

Classe	Conjunto	
	Treinamento	Teste
ArP	15	5
BrP	10	5
SrP	12	4
ArV	10	5
BrV	11	5
Total	58	24

ArV: chá verde argentino; BrV: chá verde brasileiro; ArP: chá preto argentino; BrP: chá preto brasileiro; SrP: chá preto srilanquense.

4.2. RESULTADOS E DISCUSSÃO

4.2.1. Espectros NIR

Na **Fig. 4.1a** são apresentados os espectros NIR brutos das 82 amostras de chás na faixa de 3794 a 9581 cm^{-1} . Como se pode observar, os espectros são ruidosos e exibem variações sistemáticas na linha base. Estes problemas foram

contornados aplicando o procedimento Savitzky-Golay (SAVITZKY e GOLAY, 1964) de primeira derivada com um polinômio de segunda ordem e uma janela de 75 pontos, como mostrado na Fig. 4.1b. Cada espectro resultante tinha 1426 variáveis.

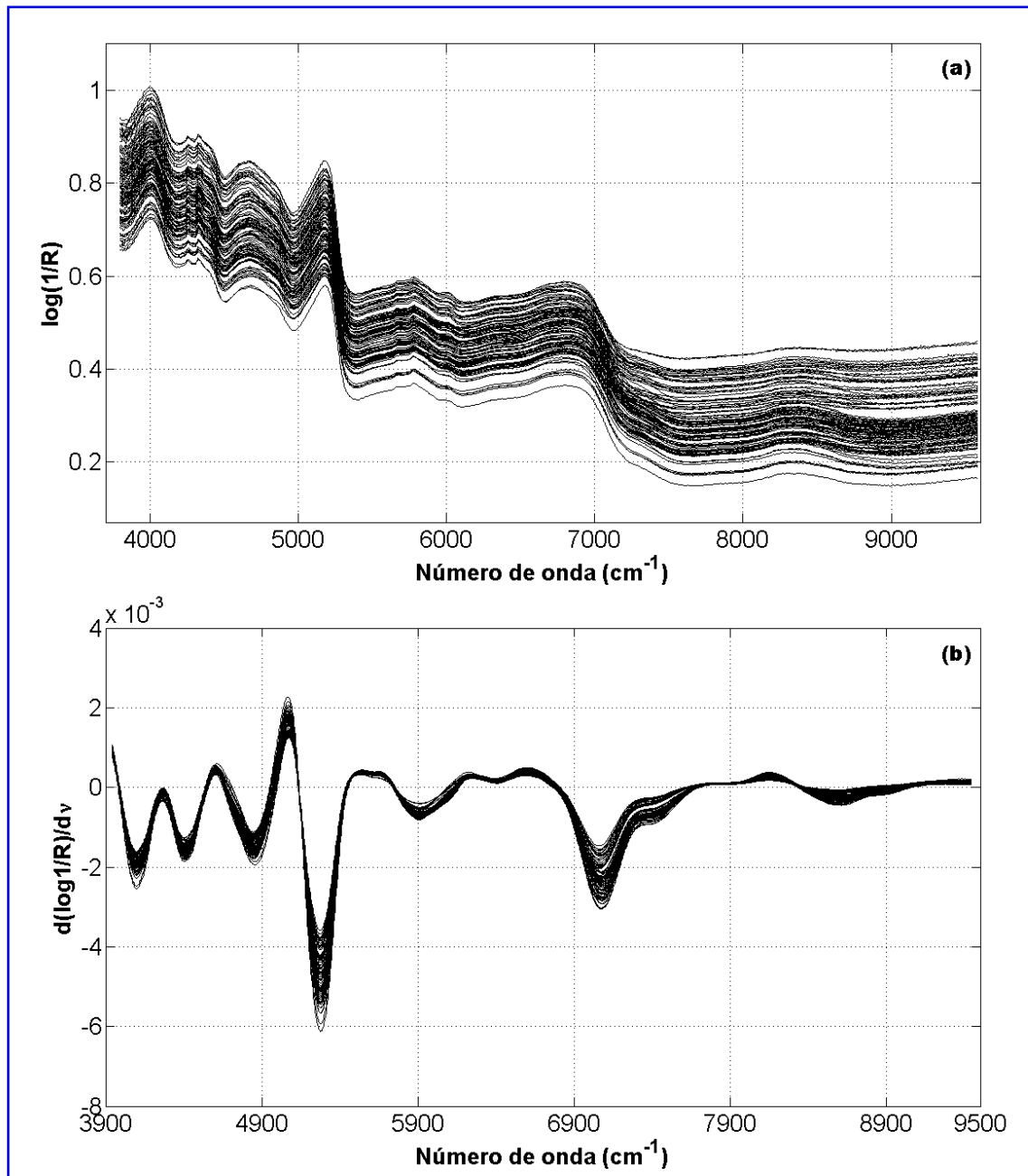


Figura 4.1. Espectros NIR (a) brutos e (b) derivados para as 84 amostras de chá estudadas.

Investigando os espectros apresentados na Fig. 4.1a podemos verificar que existem bandas de absorção de água em torno de 5155 e 7000 cm^{-1} referentes às

duas vibrações de estiramento e deformação de O–H. As bandas em aproximadamente 4000 cm^{-1} e $4300\text{--}4350\text{ cm}^{-1}$ podem ser atribuídos à combinação de absorções de estiramento de C–H e C–C. Outras designações são as seguintes: $\sim 4750\text{ cm}^{-1}$ (combinação de estiramento de C–O e deformação de O–H), $\sim 5200\text{ cm}^{-1}$ (combinação da banda de estiramento de O–H e do primeiro sobretom de deformação de C–O), a banda larga em $5200\text{--}6000\text{ cm}^{-1}$ (primeiros sobretoms de estiramento de C–H em vários grupos), $6000\text{--}7000\text{ cm}^{-1}$ (primeiro sobretom de estiramento de O–H e N–H) e 8300 cm^{-1} (segundo sobretoms de estiramento C–H em vários grupos). Os espectros de NIR brutos na faixa de acima de 9000 cm^{-1} têm informações químicas pobres e/ou estão apresentando apenas ruído.

A banda mais intensa no espectro pertence à vibração do segundo sobretom do grupo carbonila (5350 cm^{-1}), seguido pelas vibrações de estiramento e deformação C–H ($\sim 7200\text{ cm}^{-1}$) e sobretoms de --CH_2 ($\sim 5750\text{ cm}^{-1}$) e --CH_3 ($\sim 5800\text{ cm}^{-1}$). As vibrações do grupo carbonila, C–H e --CH_2 são causadas por alguns ingredientes, tais como: polifenóis, alcaloides, proteínas, ácidos voláteis e não voláteis e alguns compostos de aroma. Em geral, o teor de água nas folhas de chás secas é de aproximadamente 7% (m/m) e, portanto, o efeito da água deve ser considerado (CHEN et al., 2009).

4.2.2. Análise de componentes principais

Para avaliar a capacidade discriminante dos espectros NIR em relação ao tipo e à origem geográfica das amostras de chá, uma análise por componentes principais foi realizada. A Fig. 4.2 apresenta o gráfico de escores de PC1 \times PC2 resultante da aplicação de PCA aos espectros brutos e derivados das amostras de chás estudadas.

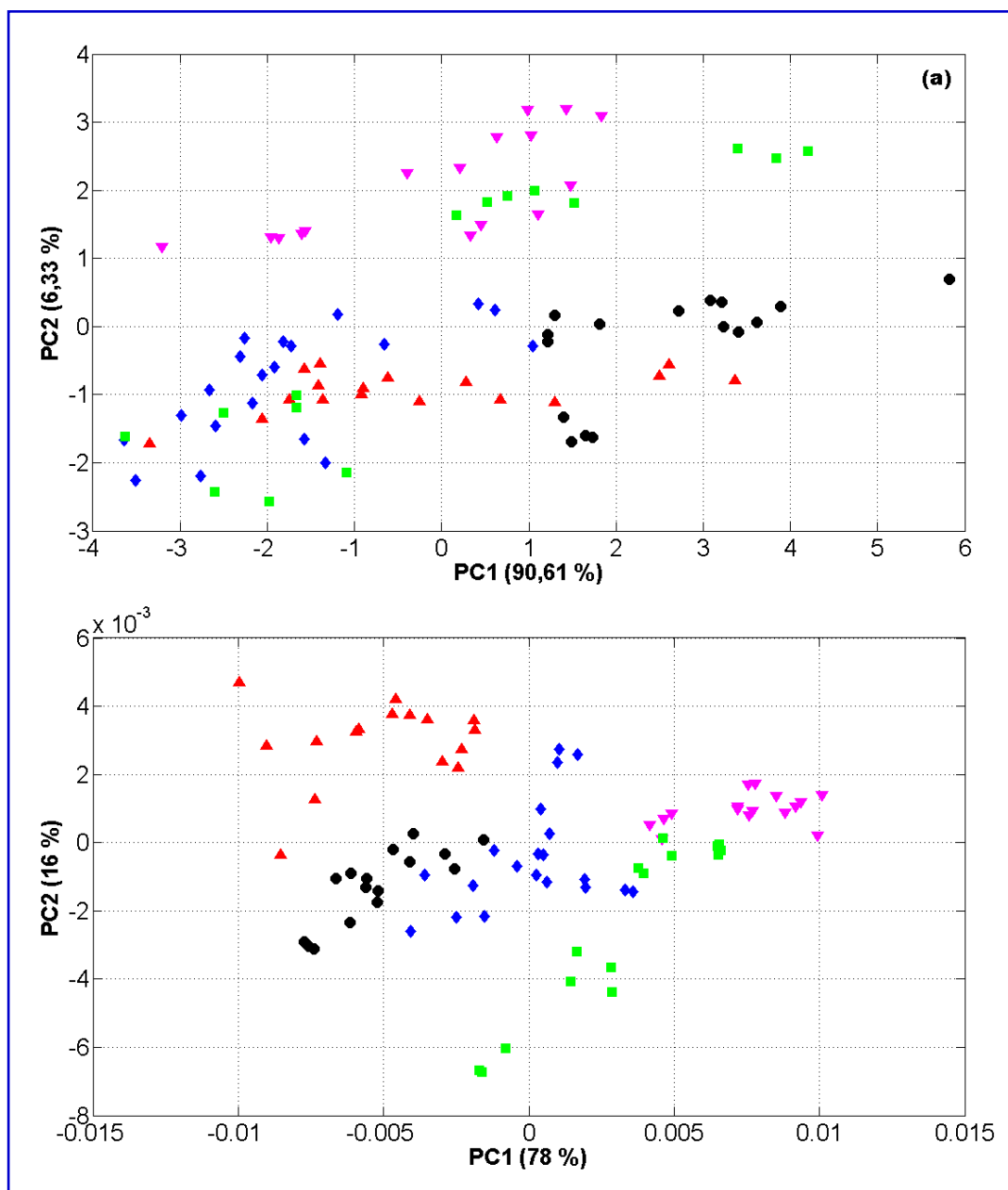


Figura 4.2. Gráfico de escores de PC1 × PC2 para as 84 amostras de chás estudadas usando os espectros (a) brutos e (b) derivados. (▼) Chá verde brasileiro, (▲) chá preto brasileiro, (■) chá verde argentino, (◆) chá preto argentino, (●) chá preto srilanquense. A variância explicada em cada PC é indicada em parênteses.

Na **Fig. 4.2a** observa-se que a utilização dos dados brutos apresenta uma tendência de separação entre as amostras apenas em relação à sua variedade: chás verdes ou pretos. Já na **Fig. 4.2b** verifica-se que, embora uma ligeira sobreposição entre os grupos seja evidente, o pré-processamento dos dados pode

conduzir a uma classificação mais apropriada das amostras de chás em relação ao seu tipo e origem geográfica simultaneamente. Assim, técnicas de classificação foram aplicadas aos espectros pré-processados das amostras e os resultados são apresentados na seção seguinte.

4.2.3. Classificação

4.2.3.1. SIMCA

Um modelo SIMCA foi construído para cada classe de chá estudada. A **Tab. 4.2** apresenta os resultados da classificação SIMCA para o conjunto de teste usando espectros NIR das amostras de chá.

Tabela 4.2. Resultados da classificação SIMCA em quatro níveis de significância para o teste-*F* (1%, 5%, 10% e 25%) para o conjunto de teste usando espectros NIR das amostras de chá.

Classe	ArP				BrP				SrP				ArV				BrV			
	1	5	10	25	1	5	10	25	1	5	10	25	1	5	10	25	1	5	10	25
ArP (4)	5	5	5	5	-	-	-	-	3	1	-	-	-	-	-	-	-	-	-	-
BrP (4)	2	2	2	1	5	5	5	3	3	2	1	-	-	-	-	-	-	-	-	-
SrP (5)	-	-	-	-	2	-	-	-	4	4	4	4	-	-	-	-	-	-	-	-
ArV (4)	2	1	1	-	-	-	-	-	-	-	-	-	5	5	5	5	4	4	2	5
BrV (5)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	5	5	5	5

ArP: chá preto argentino; BrP: chá preto brasileiro; SrP: chá preto srilanquense; ArV: chá verde argentino; BrV: chá verde brasileiro. O número de ótimo de PCs para cada classe é indicado em parênteses.

Com base na **Tab. 4.2** é possível observar os bons resultados obtidos com o SIMCA. Aumentando-se o nível de significância estatística α de 0,05 para 0,10, verificamos que há uma melhora de 40% para a classe de chás verdes brasileiros. Já quando o nível de significância estatística α aumenta para 0,25, nota-se uma redução significativa nos erros de classificação para todas as classes estudadas, exceto para os chás verdes argentinos, que obtiveram 40% de erro.

4.2.3.2. PLS-DA

Um único modelo PLS foi construído para as cinco classes de chá e o número de ótimo de variáveis latentes foi selecionado empregando validação cruzada completa. Um valor de corte de 0,5 foi adotado para inclusão de uma amostra em uma dada classe. Na **Tab. 4.3** são apresentados os resultados de classificação PLS-DA para o conjunto de teste usando espectros NIR das amostras de chá. O número de variáveis latentes é indicado em parênteses.

Tabela 4.3. Resultados de classificação PLS-DA para o conjunto de teste usando espectros NIR das amostras de chá.

Classe	ArP (4)	BrP (4)	SrP (5)	ArV (4)	BrV (5)
ArP	5	2	-	-	-
BrP	-	5	-	-	-
SrP	-	-	4	-	-
ArV	-	-	-	5	-
BrV	-	-	-	-	5

ArP: chá preto argentino; BrP: chá preto brasileiro; SrP: chá preto srilanquense; ArV: chá verde argentino; BrV: chá verde brasileiro. O número de variáveis latentes é indicado em parênteses.

Com base na **Tab. 4.3** verifica-se que todas as amostras de chás estudadas foram classificadas corretamente dentro de suas respectivas classes, isto é, não há a presença de erros do tipo I. Em relação aos do tipo II, apenas duas amostras de chá preto argentino foram classificadas incorretamente com sendo pertencentes à classe de chás pretos brasileiros. Considerando-se conjuntamente os erros tipo I e tipo II, uma média de 95,8% de classificação correta foi obtida para o conjunto de teste.

4.2.3.3. SPA-LDA

O número ideal de variáveis para o SPA-LDA foi determinado a partir do mínimo da função G custo exibido na **Fig. 4.3a**, que foi obtida para seis variáveis (4088,8; 4721,4; 4902,7; 5280,7; 5813,0 e 7062,8 cm^{-1}) indicadas na **Fig. 4.3b**.

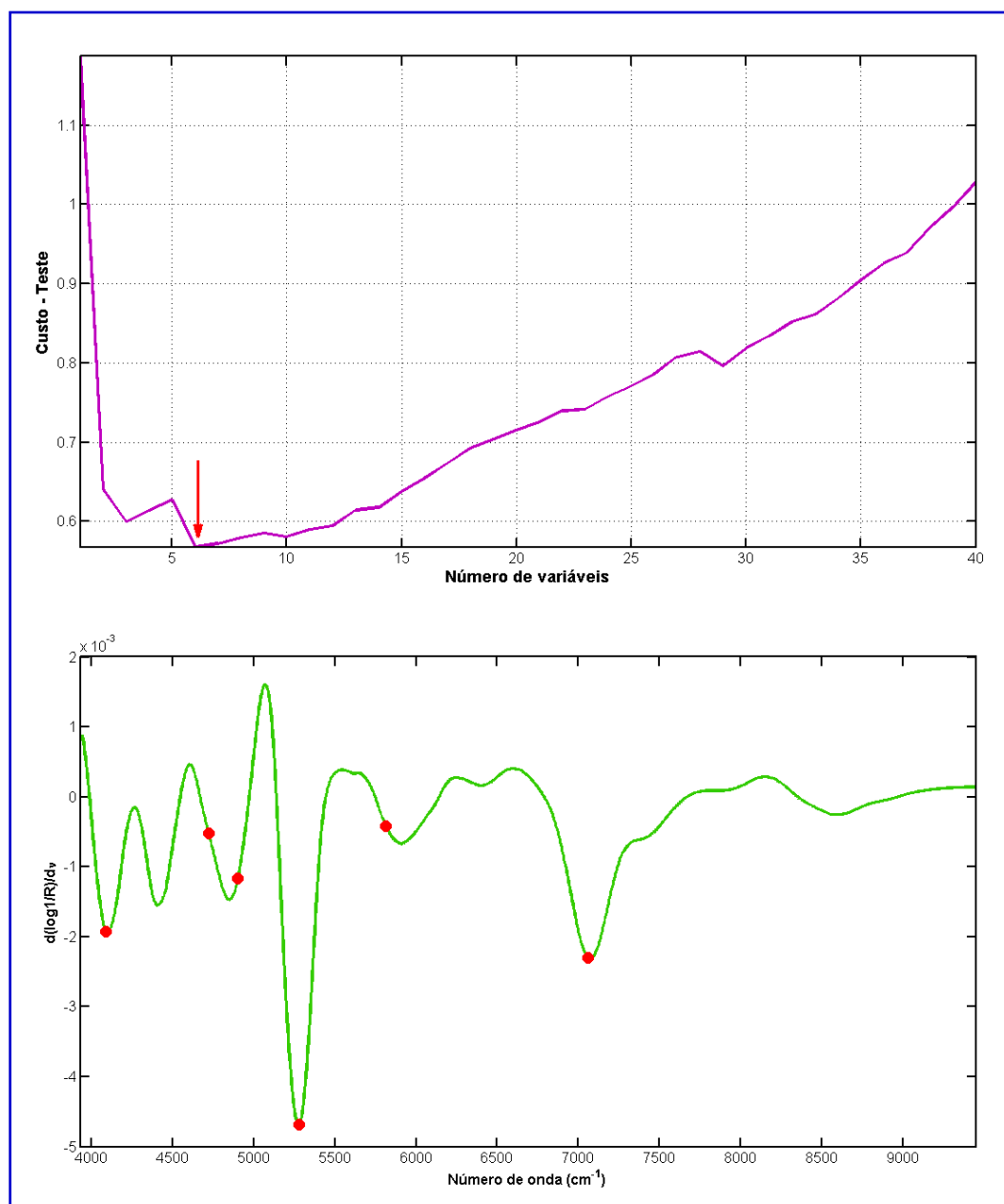


Figura 4.3. (a) Determinação do número ótimo de variáveis em SPA-LDA e (b) espectro médio derivado das 84 amostras de chá indicando os números de onda selecionados por SPA-LDA.

A **Tab. 4.4** apresenta os resultados de classificação para o conjunto de teste usando as 6 variáveis selecionadas pelo SPA-LDA nos espectros NIR dos 84 chás.

Tabela 4.4. Resultados de classificação para o conjunto de teste usando as 6 variáveis selecionadas pelo SPA-LDA nos espectros NIR das amostras de chá.

Classe	ArP	BrP	SrP	ArV	BrV
ArP	5	-	-	-	-
BrP	-	5	-	-	-
SrP	-	-	4	-	-
ArV	-	-	-	5	-
BrV	-	-	-	-	5

ArP: chá preto argentino; BrP: chá preto brasileiro; SrP: chá preto srilanquense; ArV: chá verde argentino; BrV: chá verde brasileiro.

O modelo LDA resultante classificou corretamente todas as amostras de chás estudadas, como pode ser visualizado na **Fig. 4.4**.

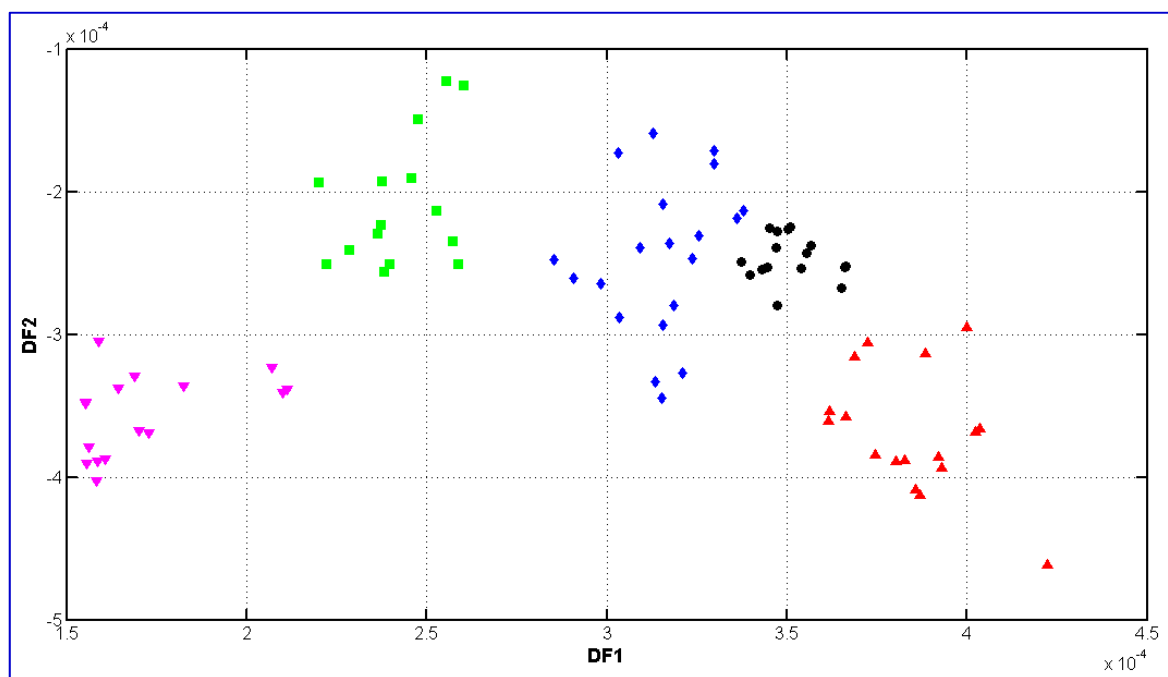


Figura 4.4. Funções discriminantes obtidas por SPA-LDA na classificação das 84 amostras de (▼) chá verde brasileiro, (▲) chá preto brasileiro, (■) chá verde argentino, (◆) chá preto argentino, (●) chá preto srilanquense.

Capítulo 5

Classificação de chás usando composição química e quimiometria



CLASSIFICAÇÃO DE CHÁS USANDO COMPOSIÇÃO QUÍMICA E QUIMIOMETRIA

5.1. EXPERIMENTAL

5.1.1. Amostras e reagentes

Foram adquiridas 60 amostras de chás em supermercados da cidade de Bahía Blanca (Argentina): 20 de chás verdes argentinos, 20 de chás pretos argentinos e 20 de chás pretos srilanquenses. Foram adquiridos cinco lotes diferentes de cada marca, de modo assegurar a representatividade das amostras. O sistema de embalagem e os procedimentos de quarteio das amostras foram análogos aos descritos na [Seção 3.3.1](#).

Todos os reagentes usados nas determinações cafeína, fluoreto, polifenóis totais e metais nas folhas e infusões foram de grau analítico. Água bidestilada desmineralizada foi empregada em todas as determinações. Todos os frascos e vidrarias foram lavados em ácido nítrico 10% (v/v) (Carlo Erba) durante 24 horas. A lavagem foi completada com rinsagens sucessivas com água bidestilada desmineralizada.

5.1.2. Procedimentos analíticos

5.1.2.1. Determinações nas folhas de chá

5.1.2.1.1. Umidade

Para a determinação de umidade, foram pesados 5 g de cada amostra de chá em pesa filtros com tampa previamente tarados. Em seguida, eles foram colocados para secar em estufa a 100 – 105 °C até atingir peso constante. Deixou-se esfriar em dessecador e se pesou. Informou-se a perda de peso como umidade.



A legislação argentina estabelece que o teor de umidade nas folhas de chá não deve exceder 7% (m/m) em base seca (CAA, 2013).

5.1.2.1.2. Cinzas totais

A determinação do teor de cinzas totais consiste da calcinação de 5 g de cada amostra em cadinhos de porcelana previamente tarados usando uma mufla, realizando uma rampa de temperatura com incrementos de 50 °C a cada 30 minutos até alcançar 550 °C. Esta temperatura foi mantida por 24 horas com a finalidade de garantir a completa calcinação e o peso constante das amostras. Este procedimento garante, ainda, a remoção de toda a matéria orgânica das amostras. A legislação argentina estabelece um teor máximo de 8% (m/m) de cinzas totais no chá em base seca (CAA, 2013).

5.1.2.1.3. Cafeína

O teor de cafeína foi determinado pelo método de Cortés modificado (CAA, 1989). Para isso, 1 g de chá foi colocado em um erlenmeyer de 100 mL, no qual foram adicionados 4 mL de H₂SO₄ concentrado (Mallinckrodt) e aquecido em banho maria durante 15 minutos a fim de destruir toda a matéria orgânica e formar o sulfato de cafeína. Em seguida, 50 mL de água destilada fervente foram adicionados e mantidos em banho maria por mais 20 minutos. Então, a suspensão ainda quente foi filtrada e a solução que contém a cafeína foi coletada em uma ampola de 150 mL. A solução foi esfriada a temperatura ambiente durante 10 minutos e alcalinizada com NaOH 40% (Mallinckrodt) até um pH de aproximadamente 8,0.

Após estas etapas são realizadas três extrações utilizando 25 mL de clorofórmio (Mallinckrodt) cada uma. A fase orgânica de cada extração que contém a

cafeína foi coletada em um balão de destilação previamente tarado. O balão foi então colocado em um rotoevaporador para reduzir o volume da solução orgânica, que foi coletada novamente em clorofórmio para ser reutilizada nas extrações seguintes. Por fim, o balão foi levado a uma estufa a 100 – 105 °C para eliminar os resíduos do solvente até atingir um peso constante. Desta forma, os cristais brancos de cafeína são obtidos e quantificados por gravimetria. A legislação argentina estabelece um teor mínimo de cafeína no chá de 1,6% (m/m) em base seca (CAA, 2013).

5.1.2.1.4. Metais totais

Um espectrofotômetro de emissão ótica sequencial com plasma acoplado indutivamente, modelo Shimadzu ICPS-1000, foi usado para a determinação de metais nas folhas das amostras de chás verdes e pretos. A operação do RF foi de 1,2 kW. Foi selecionada uma altura de observação para as medidas de 12 mm, com as seguintes vazões de argônio: plasma 14 L/min, auxiliar 1,2 L/min e nebulizador 1,0 L/min. O tempo de integração foi de 5 segundos. Foram usados estoques de soluções-padrão (1000 mg/L) para todos os analitos (Na, Mg, Al, P, K, Ca, Cr, Mn, Fe, Co, Ni, Cu, Zn, Cd e Pb) a partir de Padrões Cica-Merck. As soluções de calibração dos 15 elementos foram preparadas em ácido nítrico 0,1 mol/L a partir de diluições sucessivas de seus padrões estoque.

A determinação de metais totais (Na, Mg, Al, P, K, Ca, Cr, Mn, Fe, Co, Ni, Cu, Zn, Cd e Pb) foi feita calcinando 5 g de cada amostra de chá em uma mufla usando uma rampa de temperatura com incrementos de 50°C a cada 30 minutos até alcançar 450°C, mantendo esta temperatura por 24 horas. O produto obtido foi lentamente dissolvido em 10 mL de ácido clorídrico 6 mol/L (Carlo Erba) e aquecido

em uma placa de aquecimento até quase secura. O resíduo foi redissolvido em ácido nítrico 0,1 mol/L (Carlo Erba) e levado a 25 mL. A solução foi acondicionada em garrafas de plástico Nalgene e mantida refrigerada a 4°C. Uma filtração usando filtros Millipore de 0,45 µm foi realizada a fim de evitar entupimento do nebulizador do ICP. Os brancos foram preparados de maneira idêntica, mas omitindo-se a amostra. A **Tab. 5.1** apresenta os comprimentos de onda escolhidos para as determinações dos metais e seus respectivos limites de determinação.

Tabela 5.1. Comprimentos de onda escolhidos para as determinações dos metais e seus respectivos limites de determinação.

Elemento	Comprimento de onda (nm)	Limites de determinação (µg/L)
Al	396,153	20
Ca	422,673	0,5
Cd	226,502	5
Co	228,616	5
Cr	267,716	5
Cu	327,396	5
Fe	259,940	5
K	766,491	100
Mg	279,553	0,5
Mn	257,610	1
Na	588,995	50
Ni	231,604	5
P	178,287	50
Pb	220,351	50
Zn	213,856	5

Amostras fortificadas com concentrações de metais de 5,0 e 50 mg/kg foram utilizadas para avaliar possíveis efeitos de matriz e/ou interferências interelementos.



Não foram observados efeitos significativos para os níveis de concentração utilizados.

5.1.2.2. Determinações nas infusões de chás

5.1.2.2.1. Preparação das infusões

As infusões foram preparadas utilizando 2 g de cada amostra de chá em 100 ml de água bidestilada a 90°C, e deixou-se repousar durante 5 minutos. As infusões foram filtradas usando papel de filtro de 8 µm e levadas a 100 mL com água bidestilada. Em seguida, as soluções foram armazenadas em fracos de Nalgene na geladeira a 4°C.

5.1.2.2.2. Fluoreto

Para a determinação de fluoreto foram pesados 2 g de amostra, sobre o qual foram adicionados 100 mL de água bidestilada a 90 °C e deixou-se repousar por 5 minutos. A solução foi então filtrada com papel de filtro de porosidade média em balão de 100 mL; deixou-se esfriar e aferiu-se o volume. Em um balão de 10 mL foram adicionados 8 mL da solução contendo fluoreto e, em seguida, 1 mL de NaNO_3 0,1 mol/L e 0,1 mL de solução TISAB (tampão de ajuste de força iônica total); aferiu-se o volume com água bidestilada. A adição de NaNO_3 previne a formação de complexos de fluoreto com alumínio e ferro, além de ajudar a manter a força iônica e o pH constantes. A solução TISAB foi preparada por dissolução de 22,05 g de citrato de sódio desidratado e 0,80 g de hidróxido de sódio com 100 mL de água bidestilada; o pH foi ajustado para 5,3 com HClO_4 0,1 mol/L. As determinações de fluoreto foram realizadas usando-se um eletrodo seletivo para

fluoreto Cole Parmer 27502-19 e um potenciômetro Orion 710A. Uma solução padrão de fluoreto de 1000 mg/L foi utilizada.

5.1.2.2.3. Polifenóis totais

O teor de polifenóis totais foi estimado pelo ensaio fotométrico de Folin-Ciocalteu (SINGLETON et al., 1999). O método foi proposto por e caracteriza-se por uma mistura de ácidos fosfotungístico e fosfomolibídico (coloração amarelada) em um meio básico. Os fenóis contidos nas amostras são energeticamente oxidados em meio básico, resultando na formação do O^{2-} , o qual reage com os ácidos formando um complexo de coloração azul intensa com absorvância máxima na faixa de 760 nm. Os compostos fenólicos determinados são frequentemente expressos em ácido gálico equivalente, que é usado como padrão de calibração. As medidas foram realizadas em um espectrofotômetro UV-vis GBC Cintra 20.

5.1.2.2.4. Metais

A determinação de metais (Na, Mg, Al, P, K, Ca, Cr, Mn, Fe, Co, Ni, Cu, Zn, Cd e Pb) nas infusões recém-preparadas foram realizadas por espectrometria de emissão óptica com plasma indutivamente acoplado (ICP-OES) utilizando as mesmas condições descritas na [Seção 5.1.2.1.4](#).

5.1.3. Procedimento quimiométrico

As amostras de chá foram divididas conjuntos de calibração (45) e teste (15) aplicando-se o algoritmo (KENNARD e STONE, 1969). Em seguida, os dados foram autoescalados e uma análise exploratória usando PCA foi realizada. Foram avaliados diferentes métodos de reconhecimento de padrões: (a) SIMCA, (b) PLS-

DA e (c) SPA-LDA. Os modelos foram validados com validação cruzada completa. As variáveis escolhidas pelo SPA foram aquelas que correspondem ao menor valor de custo da função G. Os procedimentos de modelagem foram realizados usando as amostras de calibração e validação (incluindo a seleção de variáveis para a modelagem SPA-LDA e a determinação de componentes principais para cada modelagem SIMCA). As amostras de teste foram utilizadas apenas para a capacidade preditiva dos modelos e, conseqüentemente, comparação do desempenho dos modelos de classificação.

Os cálculos para PCA, SIMCA e PLS-DA foram realizados usando o programa The Unscrambler[®] 9.7 (CAMO S/A), já os algoritmos KS e SPA-LDA utilizaram o programa Matlab[®] 2009b (Mathworks Inc.). Todos os cálculos foram realizados para um nível de confiança de 95%.

2.1. RESULTADOS E DISCUSSÃO

5.1.4. Considerações iniciais

O Código Alimentar Argentino (CAA, 2013) recomenda o controle do teor de talos e pecíolos, umidade, cinzas totais, cinzas insolúveis em HCl, cinzas solúveis em água, extrato aquoso, cafeína, tanino e fibra crua. No entanto, o grande número de procedimentos para a determinação destes parâmetros torna as análises muito laboriosas, demoradas e requerem muitos passos para o tratamento de amostras. Além disso, os resultados das determinações destes parâmetros não fornecem informação diretamente relacionada à procedência do chá. Portanto, seria importante propor uma estratégia que permita a classificação de chás usando um número reduzido de parâmetros químicos, tornando as futuras análises mais simples, rápidas e baratas. Isso pode ser feito usando um algoritmo para encontrar

as informações analíticas mais relevantes para o problema em consideração, combinando as variáveis para melhorar o desempenho relacionado ao resultado analítico.

Portanto, a fim de encontrar a abordagem analítica mais simples para a classificação simultânea de chás de acordo com o tipo (chá preto ou verde) e a origem geográfica (Argentina ou na Ásia), foram testados três diferentes técnicas de classificação multivariada (SIMCA, PLS-DA e SPA-LDA) usando quatro diferentes conjuntos de dados:

- (1) O primeiro conjunto de dados (“*Todos*”) foi constituído pelos resultados de todos os parâmetros químicos determinados: umidade, cinzas totais, cafeína, fluoreto, polifenóis totais e metais nas folhas de chá e infusões.
- (2) O segundo conjunto de dados (“*CAA*”) foi formado pelos resultados das determinações de umidade, cinzas totais, cafeína e polifenóis totais, exigido pela legislação argentina. Os resultados para fluoreto também foram inseridos neste conjunto, porque o teor de fluoreto é um importante parâmetro químico que pode ser usado como indicador na avaliação da qualidade de chás (LU, GUO e YANG, 2004).
- (3) O terceiro conjunto de dados (“*Metais totais*”) foi constituído apenas dos resultados das determinações do teor de Na, Mg, Al, P, K, Ca, Cr, Mn, Fe, Co, Ni, Cu, Zn, Cd e Pb nas folhas de chá.
- (4) O último conjunto de dados (“*Metais na infusão*”) foi constituído apenas pelos resultados das determinações dos metais mencionados acima nas infusões de chá. Uma vez que o teor de cádmio estava abaixo do limite de detecção, este parâmetro não foi incluído neste conjunto.

É importante ressaltar que as determinações de fluoreto, polifenóis totais e metais nas folhas e infusões de chá foram escolhidas alternativamente ao Código Alimentar Argentino, visto que são de fundamental importância para a qualidade dos chás. As justificativas para tais escolhas são descritas a seguir.

Como a determinação de taninos identifica apenas um grupo de compostos fenólicos, atualmente é aconselhável abandonar o uso do termo "tanino" e substituir a sua determinação pela de polifenóis totais devido a sua importante contribuição no aroma, sabor e cor da infusão de chá (BRIGNARDELLO E PARODI, 2011).

O chá é uma importante fonte de flúor, visto que suas folhas acumulam mais flúor por contaminação do solo e ar que qualquer outro tipo de planta. Estima-se que o teor de flúor nas folhas da *Camellia sinensis* pode ser de 1000 e de 2–7 vezes maior que o teor do flúor dissolvido na água e no solo, respectivamente. Além disso, o flúor absorvido do solo forma complexos F–Al, o qual transporta o Al para as folhas e, em seguida, às catequinas. O consumo excessivo de flúor pode causar fluorose e de alumínio pode estar associado com mal de Alzheimer (LU, GUO e YANG, 2004).

O chá pode ser uma boa fonte alimentar de metais traços essenciais para os seres humanos. Investigar a composição elementar do chá é extremamente importante do ponto de vista médico, toxicológico e ambiental, já que depende de vários fatores, incluindo as características do solo (composição mineral, pH, tamanho das partículas, etc.), as condições ambientais (temperatura, precipitação, umidade, qualidade do ar, etc.) e a planta do chá (cultivar e estado biológico, por exemplo). Além disso, o *fingerprint* elementar já provou ser útil para a classificação de chás de acordo com a origem geográfica (MOREDA-PIÑEIRO et al., 2003; FERNÁNDEZ-CÁCERES et al., 2001; HERRADOR e GONZÁLEZ, 2001).

5.1.5. Apresentação e avaliação preliminar dos resultados

Os resultados das determinações de umidade, cinzas totais, cafeína, fluoreto, polifenóis totais e metais nas folhas e infusões de chá para as três classes estudadas são apresentados no **Anexo 1**. Dentre estes parâmetros, o Código Alimentar Argentino (CAA, 2013) estabelece limites apenas para umidade, cinzas totais e cafeína.

Em relação à umidade, observa-se que as amostras de chás pretos argentinos apresentam maior umidade que os chás pretos srilanquenses e verdes argentinos. Essa diferença pode ser devida ao fato de os chás pretos argentinos serem comercializados sem invólucro, enquanto que os demais chás disponibilizados em sachês de papel ou alumínio. Por outro lado, é sabido que a transferência de água dentro destes sistemas é incerta, uma vez que cada um dos elementos que compõem a embalagem (cartão, papel e alumínio) têm suas próprias isotermas de absorção que se combinam, em um pseudo-equilíbrio, com a própria isoterma do chá. Neste sentido, um fator agravante para a retenção de umidade nos chás são as más condições de armazenamento. Por isso, é importante destacar que praticamente todas as amostras apresentaram teor de umidade acima de 7% (m/m) em base seca, recomendado pela legislação argentina.

Para as determinações de cinzas totais, verificou-se que todas as amostras analisadas encontraram-se abaixo do valor máximo estabelecido pelo Código Alimentar Argentino, que é de 8% (m/m) em base seca. Já para as determinações de cafeína, verificou-se que 12 das 20 amostras de chás verdes apresentaram teores inferiores ao mínimo estabelecido pela legislação argentina, que é de 1,6% (m/m) em base seca. As outras 6 amostras de chás verdes e demais 40 de chás pretos estavam de acordo com o valor permitido. Uma possível explicação para isto é que o

processo de fermentação muda a composição química do chá, fazendo com que a cafeína geralmente se encontre em quantidades ligeiramente maiores no chá preto que no chá verde, que não é fermentado.

Apesar do exposto acima, a objetivo desta 3ª aplicação era avaliar quimiometricamente a utilização de dados de composição química para fins de classificação de amostras comerciais de chás de acordo com o tipo e a origem geográfica, conforme será apresentado nas próximas seções.

5.1.6. Análise de componentes principais

A **Fig. 5.1** apresenta o gráfico de escores e pesos para as duas primeiras PCs que empregam os resultados das determinações para os quatro conjuntos de dados estudados: (a) “*Todos*”, (b) “*CAA*”, (c) “*Metais totais*” e (d) “*Metais na infusão*”.

Como se pode observar, as **Figs. 5.1a** e **c** são muito semelhantes. De fato, analisando-se os gráficos de pesos destes modelos, podemos verificar que o teor de metais (principalmente K, Ca, Al, Mg, Mn e P (**Figs. 5.1a'** e **c'**)) nas folhas de chá foram responsáveis por explicar a variância descrita nas duas primeiras PCs. As **Figs. 5.1b** e **d** mostram uma pequena tendência de separação entre as três classes estudadas. O gráfico de pesos para “*Metais na infusão*” (**Fig. 5.1d**) apresenta os mesmos metais (Al, Mn, Mg, Ca, P e K) como sendo os mais importantes para as duas primeiras PCs, o que corrobora com o fato de que estes metais são facilmente extraíveis das folhas para as infusões. Já o gráfico de pesos para “*CAA*” (**Fig. 5.1b**) apresentam os teores de polifenóis totais e fluoreto como sendo os mais relevantes para as duas primeiras PCs. No entanto, sobreposições entre as classes estudadas são observadas em todos os casos, o que motiva o uso de técnicas de classificação supervisionadas.

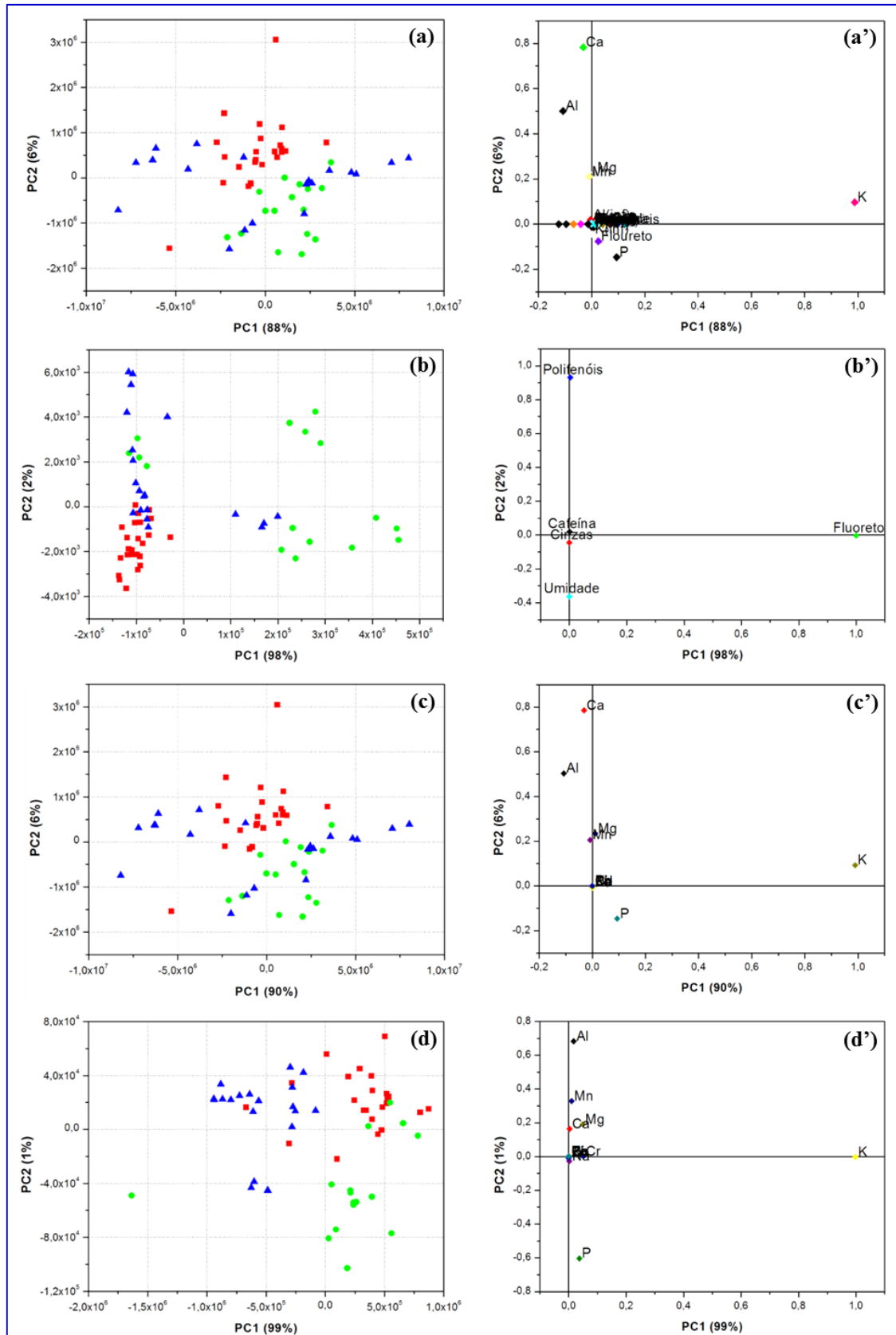


Figure 5.1. Gráficos de escores (a–d) e seus respectivos pesos (a’–d’) para as duas primeiras PCs para os quatro conjuntos de dados estudados: (a) “Todos”, (b) “CAA”, (c) “Metais totais” e (d) “Metais na infusão”. Chá preto argentino (■), chá preto srilanquense (●) e chá verde argentino (▲).

5.1.7. Classificação

A **Tab. 5.2** mostra de erros de classificação para o conjunto de teste usando SIMCA, PLS-DA e SPA-LDA. O número ótimo de PCs para cada classe para SIMCA, o número de variáveis latentes para PLS-DA e o número de variáveis selecionadas pelo SPA-LDA também são indicados.

Tabela 5.2. Erros de classificação para o conjunto de teste usando SIMCA, PLS-DA e SPA-LDA para os quatro conjuntos de dados estudados: “*Todos*”, “*CAA*”, “*Metais totais*” e “*Metais na infusão*”.

	Todos			CAA			Metais totais			Metais na infusão		
	ArP	SrP	ArV	ArP	SrP	ArV	ArP	SrP	ArV	ArP	SrP	ArV
SIMCA^a												
Número de PCs	6	5	2	1	1	1	5	4	2	1	2	2
ArP	5	2	-	5	-	4	5	3	2	5	-	-
SrP	3	5	-	5	5	-	5	5	0	2	5	-
ArV	-	-	5	-	0	5	-	-	5	-	-	5
PLS-DA												
Var. latentes	9			4			8			12		
ArP	5	-	-	5	-	-	5	-	-	5	-	-
SrP	-	5	-	-	5	-	-	5	-	-	5	-
ArV	-	-	5	-	-	3	-	-	3	-	-	5
SPA-LDA												
Var. selec.	9			3			9			3		
ArP	5	-	-	4	-	1	5	-	-	5	-	-
SrP	-	5	-	1	5	-	-	5	-	-	5	-
ArV	-	-	5	-	-	4	-	-	5	-	-	5

^anível de significância de 95%; ArP: chá preto argentino; SrP: chá preto srilanquense; ArV: chá verde argentino.

5.1.7.1. SIMCA

Com base na **Tab. 5.2**, podemos observar que todas as amostras foram classificadas corretamente em suas respectivas classes para todos os conjuntos de



dados utilizados, ou seja, não há erros tipo I. Levando-se em consideração os erros do tipo I e do tipo II conjuntamente, o melhor resultado foi obtido para “*Metais na infusão*”, que atingiu uma média de 94,4% de classificação correta.

5.1.7.2. PLS-DA

Em todos os casos, PLS-DA apresentou menos erros de classificação que seus respectivos resultados obtidos para SIMCA. Apenas os conjuntos “CAA” e “*Metais totais*” apresentaram 2 erros do tipo I para a classe de chás verdes argentinos. Como pode ainda ser observado na [Tab. 5.2](#), nenhum erro foi encontrado para os conjuntos “*Todos*” e “*Metais na infusão*”, isto é, obtiveram 100 % de classificação correta, enquanto que os outros dois conjuntos obtiveram uma média de 86,7%.

5.1.7.3. SPA-LDA

O estudo envolvendo SPA-LDA visava investigar se bons resultados de classificação poderiam ser alcançados por meio da seleção de um conjunto mínimo de parâmetros químicos, em vez das todas as variáveis. Os parâmetros selecionados pelo SPA para cada modelo são apresentados a seguir: (a) “*Todos*”: umidade e oito metais nas folhas (K, Ca, P, Mn, Al, Mg, Fe e Na); (b) “CAA”: cafeína, fluoreto e polifenóis; (c) “*Metais totais*”: K, Ca, P, Mn, Al, Mg, Fe, Na, e Co; e (d) “*Metais na infusão*”: K, Al e Mg.

A [Fig. 5.2](#) mostra o gráfico de escores das funções discriminantes de Fisher obtidas usando apenas os parâmetros químicos selecionados pelo SPA-LDA para “*Metais na infusão*”. Como se pode observar, a separação entre as classes é mais aparente quando comparadas com os gráficos de escores de PCA apresentados na

Fig. 5.1d. Além disso, três dos parâmetros indicados nos gráficos de pesos da **Fig. 5.1d'** foi selecionada pelo SPA, o que demonstra a sua excelente viabilidade para a seleção de informação analítica útil.

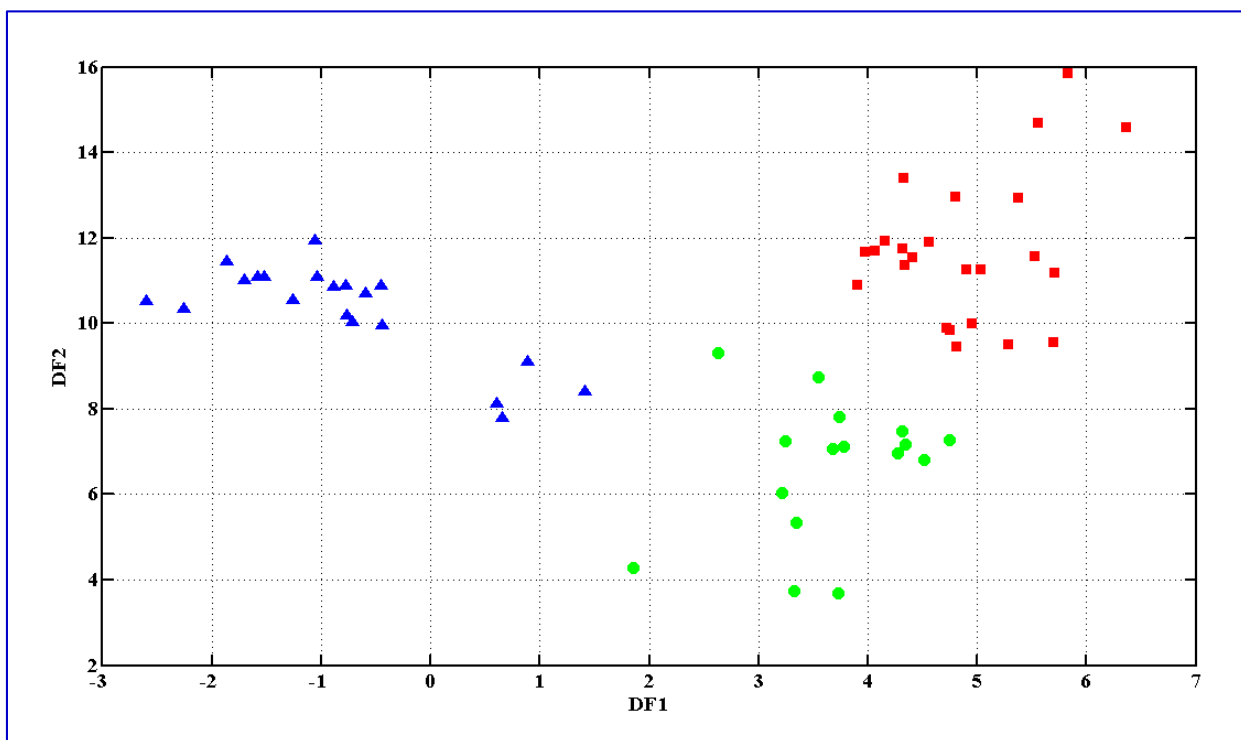


Figura 5.2. Gráficos de escores das funções discriminantes de Fisher para as amostras de chás estudadas usando as três variáveis (K, Al e Mg) selecionadas por SPA-LDA para “*Metais na infusão*”. Chá preto argentino (■), chá preto srilançense (●) e chá verde argentino (▲).

Em relação ao número de amostras classificadas incorretamente, os resultados de SPA-LDA obtiveram um melhor desempenho do que SIMCA e resultados semelhantes em comparação a PLS-DA. Apenas o conjunto “CAA” apresentou uma taxa de classificação correta inferior a 100% utilizando SPA-LDA; neste caso, apenas 87,8%. Os outros modelos classificaram corretamente todas as amostras no conjunto de teste, o que indica que os parâmetros selecionados transmitem informação analítica suficiente para discriminar as amostras de chás estudadas.

Apesar do bom desempenho de todas as técnicas de classificação supervisionadas utilizando o conjunto “*Metais na infusão*”, SPA-LDA fornece o resultado mais importante, uma vez que seleciona apenas três parâmetros químicos. Este resultado é extremamente positivo do ponto de vista da análise química, porque a quantificação de apenas três elementos majoritários (K, Al e Mg) em uma simples infusão de chá pode proporcionar uma metodologia mais simples, rápida e barata. De fato, a assinatura química (impressão digital) das folhas e infusões de chá em relação a seu perfil elementar é apropriada para se obter uma classificação confiável, já que ela é específica para uma determinada área (origem geográfica) onde as plantas de chá crescem e não apresentam mudanças para um período de tempo relativamente longo (SZYMCZYCHA-MADEJA et al., 2012). Além disso, do ponto de vista biológico, pode-se observar que: o potássio é um elemento essencial para o crescimento das plantas e o magnésio é encontrado na clorofila das plantas, que é um complexo de porfirina-magnésio envolvido na fotossíntese. Por outro lado, a *Camellia sinensis* é uma das poucas plantas que toleram o acúmulo de alumínio, fato este que deve ser mais bem investigado visto que a ingestão humana de alumínio pode conduzir a problemas de saúde devido a suas neuro e osteotoxicidades.

Uma vez que esta metodologia não foi empregada para amostras de chás brasileiras, a futura quantificação dos três parâmetros selecionados por SPA-LDA nestas amostras pode ser realizada e, em seguida, uma nova classificação contendo as amostras de chás brasileiras, argentinas e srilanqueses pode ser realizada. Desta forma, outra estratégia para a avaliação da qualidade de chás pode ser obtida, o que pode ser útil para a elaboração de normas para o controle de qualidade de chás no Brasil, que ainda é inexistente.

Capítulo 6

Conclusão



6. CONCLUSÃO

Levando-se em consideração que o Brasil não possui uma legislação específica para avaliar e controlar a qualidade de chás, três diferentes estratégias para avaliação da qualidade de chás produzidos no Brasil, Argentina e Sri Lanka foram propostas nesta Tese de Doutorado.

Na 1ª aplicação foi desenvolvida uma estratégia baseada na utilização de imagens capturadas com uma câmera digital comercial simples para classificação simultânea de chás de acordo com o tipo e a origem geográfica. Através da utilização de histogramas de cor, obtidos a partir das imagens digitais, combinada com SPA-LDA foi possível classificar corretamente todas as amostras de chá estudadas. O posicionamento da amostra, a uniformidade da iluminação e a distância focal também são fatores importantes, já que, uma vez resolvidos, evitam uma posterior manipulação das imagens. A metodologia proposta apresenta vantagens intrínsecas: não requer nenhum tratamento prévio das amostras, não usa reagentes e não gera resíduos. Além disso, ela pode fornecer uma ferramenta adicional, rápida e barata aos produtores primários de chá de modo a garantir um produto de qualidade diferenciada.

Em uma 2ª aplicação, foi desenvolvida uma metodologia baseada em espectroscopia NIR e técnicas quimiométricas de reconhecimento de padrão também para a classificação simultânea de chás segundo o tipo e a origem geográfica. Mais uma vez, todas as amostras estudadas foram classificadas corretamente utilizando-se SPA-LDA, demonstrando, assim, a sua boa capacidade de classificação entre as classes estudadas. A metodologia proposta destaca-se por ser simples, rápida, não destrói a amostra, não gera resíduos, além de ser de fácil

implementação na indústria de modo a promover uma análise exploratória rápida da qualidade do chá.

Na 3ª aplicação foi desenvolvido um estudo que explora diferentes abordagens para classificação de chás baseada na seleção de atributos de composição química e técnicas de classificação multivariada. Dados quantitativos de umidade, cinzas totais, cafeína, fluoreto, polifenóis totais e quinze metais nas folhas e infusões de chá foram usados como informação analítica. A partir daí, buscou-se uma estratégia que permitisse a classificação de chás usando um número reduzido de parâmetros químicos. Deste modo, a utilização de SPA-LDA proporcionou o melhor resultado de classificação a partir da seleção de apenas três parâmetros químicos (K, Al e Mg) determinados nas infusões de chá. Este resultado confirma que a impressão digital elementar (assinatura química) pode ser usada para identificar simultaneamente o tipo e a origem geográfica de chás. O resultado é extremamente positivo do ponto de vista de análises químicas, porque quantificações feitas com menos elementos naturalmente fornecem metodologias mais simples, mais rápidas e menos dispendiosas.

Os bons resultados obtidos pelas metodologias ora descritas são bastante promissores no sentido em que contribuem significativamente para o desenvolvimento de estratégias mais rápidas e confiáveis para uma avaliação da qualidade de chás, principalmente brasileiros, para serem internacionalmente aceitável. No entanto, a fim de garantir uma melhor generalização das metodologias propostas, seria necessário o uso de mais amostras de chás de diferentes origens geográficas e tipos (branco, amarelo, oolong e pu-ehr, por exemplo).

2.1. Propostas futuras

Além dos estudos ora descritos neste trabalho para classificação de chás usando diferentes técnicas, pretende-se explorar ainda o uso da espectroscopia UV-vis para esta mesma finalidade, tendo-se em vista que ela é uma técnica simples e facilmente acessível em laboratórios de análises de rotina. Outros estudos que envolvem o uso de determinação de parâmetros químicos e físico-químicos também serão abordados.

Do ponto de vista toxicológico, uma vez que as amostras de chás possuem uma ampla variedade de metais, torna-se necessário também o desenvolvimento de Cartas de Controle Multivariado de modo definir os teores limites destes metais aceitáveis para o consumo humano. Além disso, seria importante desenvolver também uma avaliação toxicológica das amostras com base nos teores de alumínio e fluoreto em infusões de chás.

Assim, pretende-se, de forma objetiva, como proposta de continuidade deste trabalho:

- Classificar simultaneamente o tipo e a origem geográfica de chás empregando espectroscopia UV-vis e SPA-LDA;
- Determinar umidade, cinzas, cafeína e polifenóis totais empregando espectrometria NIR e calibração multivariada;
- Desenvolver cartas de controle multivariado para o perfil metálico das folhas e infusões de chás;
- Avaliação das relações existentes entre os teores de alumínio e fluoreto em infusões de chás.

Referências



REFERÊNCIAS

- ABRAMOFF, M. D., MAGALHAES, P. J., RAM, S. J. ImageJ. *Biophotonics International*, 11 (2004) 36–42.
- ANDERSEN, C. M., BRO, R. Variable selection in regression – a tutorial. *Journal of Chemometrics*, 24 (2010) 728–737.
- ANDRADE, S. I. E., LIMA, M. B., BARRETO, I. S., LYRA, W. S., ALMEIDA, L. F., ARAÚJO, M. C. U., SILVA, E. C. A digital image-based flow-batch analyzer for determining Al(III) and Cr(VI) in water. *Microchemical Journal*, 109 (2013) 106–111.
- ANESINI, C., FERRARO, G. E., FILIP, R. Total Polyphenol content and antioxidant capacity of commercially available tea (*Camellia sinensis*) in Argentina. *Journal of Agricultural and Food Chemistry*, 56 (2008) 9225–9229.
- ARAÚJO, M. C. U., SALDANHA, T. C. B., GALVÃO, R. K. H., YONEYAMA, T., CHAME, H. C., VISANI, V. The successive projections algorithm for variable selection in spectroscopic multicomponent analysis. *Chemometrics and Intelligent Laboratory Systems*, 57 (2001) 65–73.
- BERRUETA, L. A., ALONSO-SALCES, R. M., HÉBERGER, K. Supervised pattern recognition in food analysis. *Journal of Chromatography A*, 1158 (2007) 196–214.



- BOKOBZA, L. Origin of near-infrared absorption bands. In: SIESLER, H. W., OZAKI, Y., KAWATA, S., HEISE H. M. (Eds.), **Near infrared spectroscopy: principles, instruments, applications**. Weinheim: Wiley-VCH, 2002. pp. 11–41.
- BORSE, B. B., RAO, L. J. M., NAGALAKSHMI, S., KRISHNAMURTHY, N. Fingerprint of black teas from India: identification of the regio-specific characteristics. *Food Chemistry*, 79 (2002) 419–424.
- BRERETON, R. G. **Chemometrics: data analysis for the laboratory and chemical plant**. Chichester: Wiley, 2003.
- BRIGNARDELLO, A. E., PARODI, N. B. Calidades químicas de té negro de Misiones-Argentina y de un producto derivado (bt) destinados a exportación. *Revista de Ciencia y Tecnología*, 15 (2011) 30–37.
- CHATURVEDULA, V. S. P., PRAKASH, I. The aroma, taste, color and bioactive constituents of tea. *Journal of Medicinal Plants Research*, 5 (2011) 2110–2124.
- CHEN, Q., ZHAO, J., ZHANG, H., MUHUA, L., FANG, M. Qualitative identification of tea by near infrared spectroscopy based on soft independent modeling of class analogy pattern recognition. *Journal of Near Infrared Spectroscopy*, 13 (2005) 327–322.
- CHEN, Q., ZHAO, J., ZHANG, H., WANG, X. Feasibility study on qualitative and quantitative analysis in tea by near infrared spectroscopy with multivariate calibration. *Analytica Chimica Acta*, 572 (2006) 77–84.



- CHEN, Q., ZHAO, J., FANG, C. H., WANG, D. Feasibility study on identification of green, black and Oolong teas using near-infrared reflectance spectroscopy based on support vector machine (SVM). *Spectrochimica Acta Part A*, 66 (2007) 568–574.
- CHEN, Q., ZHAO, J., LIU, M., CAI, J. Nondestructive identification of tea (*Camellia sinensis L.*) varieties using FT-NIR spectroscopy and pattern recognition. *Czech Journal of Food Sciences*, 26 (2008) 360–367.
- CHEN, Q., ZHAO, J., LIU, M. Study on discrimination of Roast green tea (*Camellia sinensis L.*) according to geographical origin by FT-NIR spectroscopy and supervised pattern recognition. *Spectrochimica Acta Part A*, 72 (2009) 845–850.
- CHOW, H.-H. S., HAKIM, I. A. Pharmacokinetic and chemoprevention studies on tea in humans. *Pharmacological Research*, 64 (2011) 105–112.
- CHU, D. C., JUNEJA, L. R. General composition of green tea and its infusion. In: YAMAMOTO, T., JUNEJA, L. R., CHU, D., KIM, M. (Eds.). **Chemistry and applications of green tea**. pp. 1–12. Boca Raton: CRC Press, 1997.
- CIURCZAK, E. W. Principles of near-infrared spectroscopy. In: BURNS, D. A., CIURCZAK E. W. (Eds.), **Handbook of near-infrared analysis**, 2nd ed. New York/Basel: Marcel Dekker Inc., 2001. pp. 7–18.
- Código Alimentario Argentino (CAA). Res 101/89 MSAS. Boletín Oficial 26.590. 14.6– Determinación de la Cafeína. A. Método de Cortés Modificado. Argentina, 1989.

- DANIEL, G. B. Digital imaging. *Veterinary Clinics of North America: Small Animal Practice*, 39 (2009) 667–676.
- DINIZ, P. H. G. D., DANTAS, H. V., MELO, K. D. T., BARBOSA, M. F., HARDING, D. P., NASCIMENTO, E. C. L., PISTONESI, M. F., BAND, B. S. F., ARAÚJO, M. C. U. Using a simple digital camera and SPA-LDA modeling to screen teas. *Analytical Methods*, 4 (2012) 2648–2652.
- Eurachem Guide: The fitness for purpose of analytical methods: a laboratory guide to method validation and related topics. Teddington: Eurachem, 1998.
- FISHER, R. A. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7 (1936) 179–188.
- FOLEY, J. D., VAN DAM, A., FEINER, S. K., HUGHES, J. F. Computer graphics, principles and practice. Reading: Addison-Wesley, 1990.
- FREITAS, S. K. B., NASCIMENTO, E. C. L., DIONÍZIO, A. G. G., GOMES, A. A., ARAÚJO, M. C. U., GALVÃO, R. K. H. A flow-batch analyzer using a low cost aquarium pump for classification of citrus juice with respect to brand. *Talanta*, 107 (2013) 45–48.
- GAIÃO, E. N., MARTINS, V. L., LYRA, W. S., ALMEIDA, L. F., SILVA, E. C., ARAÚJO, M. C. U. Digital image-based titrations. *Analytica Chimica Acta*, 570 (2006) 283–290.
- GAMBARRA-NETO, F. F., MARINO, G., ARAÚJO, M. C. U., GALVÃO, R. K. H., PONTES, M. J. C., MEDEIROS, E. P., LIMA, R. S. Classification of edible

- vegetable oils using square wave voltammetry with multivariate data analysis. *Talanta*, 77 (2009) 1660–1666.
- GHASEMI-VARNAMKHAZI, M., MOHTASEBI, S. S., RODRIGUEZ-MENDEZ, M. L., GOMES, A. A., ARAÚJO, M. C. U., GALVÃO, R. K. H. Screening analysis of beer ageing using near infrared spectroscopy and the successive projections algorithm for variable selection. *Talanta*, 89 (2012) 286–291.
- GILL, G.S., KUMAR, A., AGARWAL, R. Monitoring and grading of tea by computer vision – a review. *Journal of Food Engineering*, 106 (2011) 13–19.
- GONZALES, R. C., WOODS, R. E. **Digital image processing**. Reading: Addison-Wesley, 1992.
- GUTMAN, R. L., RYU, B.-H. Rediscovering tea. An exploration of the scientific literature, *HerbalGram*, 37 (1996) 33–48.
- HAGEMAN, J. A., STREPPPEL, M., WEHRENS, R., BUYDENS, L. M. C. Wavelength selection with tabu search. *Journal of Chemometrics*, 17 (2003) 427–437.
- HALL, M.N., ROBERTSON, A., SCOTTER, C. N. G. Near-infrared reflectance prediction of quality, theaflavin content and moisture content of black tea. *Food Chemistry*, 27 (1988) 61–75.
- HE, W., HU, X., ZHAO, L., LIAO, X., ZHANG, Y., ZHANG, M., WU, J. Evaluation of Chinese tea by the electronic tongue: correlation with sensory properties and classification according to geographical origin and grade level. *Food Research International*, 42 (2009) 1462–1467.

- HE, Y., LI, X. L., DENG, X. F. Discrimination of varieties of tea using near-infrared spectroscopy by principal component analysis and BP model. *Journal of Food Engineering*, 79 (2007) 1238–1242.
- HOLLAND, J. H. **Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence**. Ann Arbor: University of Michigan Press, 1975.
- HORCHNER, U., KALIVAS, J. H. Simulated-annealing-based optimization algorithms: Fundamentals and wavelength selection applications. *Journal of Chemometrics*, 9 (1995) 283–308.
- HUANG, H., YU, H., XU, H., YING, Y. Near infrared spectroscopy for on/in-line monitoring of quality in foods and beverages: a review. *Journal of Food Engineering*, 87 (2008) 303–313.
- JÄHNE, B. (Ed.). **Digital image processing**. 6th edition. Heidelberg: Springer, 2005.
- JAIN, A., MANGHANI, C., KOHLI, S., NIGAM, D., RANI, V. Tea and human health: the dark shadows. *Toxicology Letters*, 220 (2013) 82–87.
- KEMSLEY, E. K. Discriminant analysis of high-dimensional data: a comparison of principal components analysis and partial least squares data reduction methods. *Chemometrics and Intelligent Laboratory Systems*, 33 (1996) 47–61.
- KENNARD, R. W., STONE, L. A. Computer aided design of experiments. *Technometrics*, 11 (1969) 137–148.

- LADDI, A., PRAKASH, N. R., SHARMA, S., MONDAL, H. S., KUMAR, A., KAPUR, P. Significant physical attributes affecting quality of Indian black (CTC) tea. *Journal of Food Engineering*, 113 (2012) 69–78.
- LADDI, A., SHARMA, S., KUMAR, A., KAPUR, P. Classification of tea grains based upon image texture feature analysis under different illumination conditions. *Journal of Food Engineering*, 115 (2013) 226–231.
- LAVINE, B. K. **Encyclopedia of analytical chemistry**. Chichester: Wiley, 2000.
- LAVINE, B.K. Pattern recognition. *Critical Reviews in Analytical Chemistry*, 36 (2006) 153–161.
- LI, X., HE, Y. Discriminating varieties of tea plant based on Vis/NIR spectral characteristics and using artificial neural networks. *Biosystems Engineering*, 99 (2008) 313–321.
- LI, X., NIE, P., QIU, Z.-J., HE, Y. Using wavelet transform and multi-class least square support vector machine in multi-spectral imaging classification of Chinese famous tea. *Expert Systems with Applications*, 38 (2011) 11149–11159.
- LIMA, J. D., MAZZAFERA, P., MORAES, W. S., SILVA, R. B. Chá: aspectos relacionados à qualidade e perspectivas. *Ciência Rural*, 39 (2009) 1270–1278.
- LIU, S. L., TSAI, Y. S., OU, A. S.-M. Classifying the variety, production area and season of Taiwan partially fermented tea by near infrared spectroscopy. *Journal of Food and Drug Analysis*, 18 (2010) 34–43.

- LOPEZ-MOLINERO, A., LIÑAN, D., SIPIERA, D., FALCON, R. Chemometric interpretation of digital image colorimetry. Application for titanium determination in plastics. *Microchemical Journal*, 96 (2010) 380–385.
- LYRA, W. S., SANTOS, V. B., DIONÍZIO, A. G. G., MARTINS, V. L., ALMEIDA, L. F., GAIÃO, E. N., DINIZ, P. H. G. D., SILVA, E. C., ARAÚJO, M. C. U. Digital image-based flame emission spectrometry. *Talanta*, 77 (2009) 1584–1589.
- LYRA, W. S., SANCHES, F. A. C., CUNHA, F. A. S., DINIZ, P. H. G. D., LEMOS, S. G., SILVA, E. C., ARAÚJO, M. C. U. Indirect determination of sodium diclofenac, sodium dipyron and calcium gluconate in injection drugs using digital image-based (webcam) flame emission spectrometric method. *Analytical Methods*, 3 (2011) 1975–1980.
- MAESSCHALCK, R., JOUAN-RIMBAUD, D., MASSART, D. L. Tutorial - The Mahalanobis distance. *Chemometrics and Intelligent Laboratory Systems*, 50 (2000) 1–18.
- MASSART, D. L., VANDEGINSTE, B. G. M., BUYDENS, L. M. C., DE JONG, S., LEWI, P. J., SMEYERS-VERBEKE, J. **Handbook of Chemometrics and Qualimetrics, Part A**. Amsterdam: Elsevier, 1997.
- MEI, L., LUNDIN, P., BRYDEGAARD, M., GONG, S., TANG, D., SOMESFALEAN, G., HE, S., SVANBERG, S. Tea classification and quality assessment using laser-induced fluorescence and chemometric evaluation. *Applied Optics*, 51 (2012) 803–811.

- MØLLER, S. F., VON FRESE, J., BRO, R. Robust methods for multivariate data analysis. *Journal of Chemometrics*, 19 (2005) 549–563.
- MOREIRA, E. D. T., PONTES, M. J. C., GALVÃO, R. K. H., ARAÚJO, M. C. U. Near infrared reflectance spectrometry classification of cigarettes using the successive projections algorithm for variable selection. *Talanta*, 79 (2009) 1260–1264.
- NAGALAKSHMI, S. Tea: An Appraisal of Processing Methods and Products. In: CHARKRAVERTY, A., MUJUMDAR, A. S., RAGHAVAN, G. S. V., RAMASWAMY, H. S. (Eds.) **Handbook of Postharvesting Technology: Cereals, Fruits, Vegetables, Tea, and Spices**. pp 741–778. New York: Marcel Dekker, 2003.
- O'DRISCOLL, S., MACCRAITH, B. D., BURKE, C. S. A novel camera phone-based platform for quantitative fluorescence sensing. *Analytical Methods*, 5 (2013) 1904–1908.
- OLINGER, J. M., GRIFFITHS, P. R., BURGER, T. Theory of diffuse reflection in the NIR region. BURNS, D. A., CIURCZAK E. W. (Eds.), **Handbook of near-infrared analysis**, 2nd ed. New York/Basel: Marcel Dekker Inc., 2001. pp. 19–51.
- Organização das Nações Unidas para Alimentação e Agricultura (ONUAA) (2012). Committee on commodity problems, intergovernmental group on tea. Current situation and medium term outlook for tea. Disponível em: <http://www.fao.org/fileadmin/templates/est/COMM_MARKETS_MONITORING/Tea/Documents/IGG_20/12-CRS7-CurrentSit_01.pdf>. Acesso em 19 de junho de 2013.

- OTTO, M. **Chemometrics statistics and computer application in analytical chemistry**. Weinheim: Wiley-VCH, 1999.
- PASQUINI, C. Near infrared spectroscopy: fundamentals, practical aspects and analytical applications. *Journal of the Brazilian Chemical Society*, 14 (2003) 198–219.
- PINTO, M. S. Tea: a new perspective on health benefits. *Food Research International*, (2013). doi: [10.1016/j.foodres.2013.01.038](https://doi.org/10.1016/j.foodres.2013.01.038)
- PONTES, M. J. C., CORTEZ, J., GALVÃO, R. K. H., PASQUINI, C., ARAÚJO, M. C. U., COELHO, R. M., CHIBA, M. K., ABREU, M. F., MADARI, B. E. Classification of Brazilian soils by using LIBS and variable selection in the wavelet domain. *Analytica Chimica Acta*, 642 (2009) 12–18.
- PONTES, M. J. C., GALVÃO, R. K. H., ARAÚJO, M. C. U., MOREIRA, P. N. T., PESSOA NETO, O. D., JOSÉ, G. E., SALDANHA, T. C. B. The successive projections algorithm for spectral variable selection in classification problems. *Chemometrics and Intelligent Laboratory Systems*, 78 (2005) 11–18.
- PONTES, M. J. C., PEREIRA, C. F., PIMENTEL, M. F., VASCONCELOS, F. V. C., SILVA, A. G. B. Screening analysis to detect adulteration in diesel/biodiesel blends using near infrared spectrometry and multivariate classification. *Talanta*, 85 (2011) 2159–2165.
- PREEDY, V. **Tea in health and disease prevention**. 1st edition. San Diego: Academic Press, 2012.

- REICH, G. Near-infrared spectroscopy and imaging: basic principles and pharmaceutical applications. *Advanced Drug Delivery Reviews*, 57 (2005) 1109–1143.
- REN, G., WANG, S., NING, J., XU, R., WANG, Y., XING, Z., WAN, X., ZHANG, Z. Quantitative analysis and geographical traceability of black tea using Fourier transform near-infrared spectroscopy (FT-NIRS), *Food Research International*, (2013). doi: [10.1016/j.foodres.2012.10.032](https://doi.org/10.1016/j.foodres.2012.10.032)
- SANG, S., LAMBERT, J. D., HO, C.-T., YANG, C. S. The chemistry and biotransformation of tea constituents. *Pharmacological Research*, 64 (2011) 87–99.
- SAVITZKY, A., GOLAY, M. J. E. Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36 (1964) 1627-1639.
- SCHONBERG, A., MOUBACHER, R. The Strecker degradation of α -amino acids. *Chemical Reviews*, 50 (1952) 261–277.
- SEETOHUL, L. N., ISLAM, M., O'HARE, W. T., ALI, Z. Discrimination of teas based on total luminescence spectroscopy and pattern recognition. *Journal of the Science of Food and Agriculture*, 86 (2006) 2092–2098.
- SHAMSIPUR, M., ZARE-SHAHABADI, V., HEMMATEENEJAD, B., AKHOND, M. Combination of ant colony optimization with various local search strategies. a novel method for variable selection in multivariate calibration and QSPR study. *QSAR Combinatorial Sciences*, 28 (2009) 1263–1275.

- SHARANGI, A. B., Medicinal and therapeutic potentialities of tea (*Camellia sinensis* L.) – a review. *Food Research International*, 42 (2009) 529–535.
- SIEBERT, K. J. Chemometrics in brewing: a review. *Journal of the American Society of Brewing Chemists*, 59 (2001) 147–156.
- SILVA, A. C., PONTES, L. F. B. L., PIMENTEL, M. F., PONTES, M. J. C. Detection of adulteration in hydrated ethyl alcohol fuel using infrared spectroscopy and supervised pattern recognition methods. *Talanta*, 93 (2012b) 129–134.
- SILVA, C. S., BORBA, F. S. L., PIMENTEL, M. F., PONTES, M. J. C., HONORATO, R. S., PASQUINI, C. Classification of blue pen ink using infrared spectroscopy and linear discriminant analysis. *Microchemical Journal*, 109 (2013) 122–127.
- SINGLETON, V. L., ORTHOFER, R., LAMUELA-RAVENTOS, R. M. Analysis of total phenols and other oxidation substrates and antioxidants by means of Folin-Ciocalteu reagent. *Methods in Enzymology*, 299 (1999) 152–178.
- SOARES, A. S., GALVÃO, R. K. H., ARAÚJO, M. C. U., SOARES, S. F. C., PINTO, L. A. Multi-core computation in chemometrics: case studies of voltammetric and NIR spectrometric analyses. *Journal of the Brazilian Chemical Society*, 21 (2010) 1626–1634.
- SOARES, S. F. C., GOMES, A. A., GALVÃO FILHO, A. R., ARAÚJO, M. C. U., GALVÃO, R. K. H. The successive projections algorithm. *Trends in Analytical Chemistry*, 42 (2013) 84–98.

- SOHN, M., HIMMELSBACH, D. S., KAYS, S. E., ARCHIBALD, D. D., BARTON, F. E. NIR-FT/Raman spectroscopy for nutritional classification of cereal foods. *Cereal Chemistry*, 82 (2005) 660–665.
- SOUSA, A. C., SOARES, A. S., COELHO, C. J., GAIÃO, E. N., PEREIRA, A. F. C., GALVÃO, R. K. H., ARAÚJO, M. C. U. Screening analysis of seston from a domestic wastewater treatment plant using digital images. *Analytical Methods*, 4 (2012) 2375–2380.
- SOUTO, U. T. C. P., PONTES, M. J. C., SILVA, E. C., GALVÃO, R. K. H., ARAÚJO, M. C. U., SANCHES, F. A. C., CUNHA, F. A. S., OLIVEIRA, M. S. R. UV–Vis spectrometric classification of coffees by SPA–LDA. *Food Chemistry*, 119 (2010) 368–371.
- SUN, D.-W. **Computer vision technology for food quality evaluation**. New York: Academic Press, 2007.
- SUN, D.-W. **Infrared spectroscopy for food quality, analysis and control**. New York: Academic Press, 2009.
- SUN, D.-W. **Hyperspectral imaging for food quality analysis and control**. New York: Academic Press, 2010.
- SZYMCZYCHA-MADEJA, A., WELNA, M., POHL, P. Elemental analysis of teas and their infusions by spectrometric methods. *Trends in Analytical Chemistry*, 35 (2012) 165–181.

- TAN, S.-M., LUO, R.-M., ZHOU, Y.-P., GONG, H., TAN, Z. Rapid and non-destructive discrimination of tea varieties by near infrared diffuse reflection spectroscopy coupled with classification and regression trees. *African Journal of Biotechnology*, 11 (2012) 2303–2312.
- TÔRRES, A. R., LYRA, W. S., ANDRADE, S. I. E., ANDRADE, R. A. N., SILVA, E. C., ARAÚJO, M. C. U., GAIÃO, E. N. A digital image-based method for determining of total acidity in red wines using acid–base titration without indicator. *Talanta*, 84 (2011) 601–606.
- WANG, L.-F., LEE, J.-Y., CHUNG, J.-O., BAIK, J.-H., SO, S., PARK, S.-K. Discrimination of teas with different degrees of fermentation by SPME-GC analysis of the characteristic volatile flavour compounds. *Food Chemistry*, 109 (2008) 196–206.
- WEHLING, R. L. Infrared spectroscopy. In: NIELSEN, S. S. (Ed.) **Food analysis**. 4th edition. pp. 407–420. New York: Springer, 2009.
- WOLD, S., ANTTI, H., LINDGREN, F., ÖHMAN, J. Orthogonal signal correction of near-infrared spectra. *Chemometrics and Intelligent Laboratory Systems*, 44 (1998) 175–185.
- WU, D., YANG, H., CHEN, X., HE, Y., LI, X. Application of image texture for the sorting of tea categories using multi-spectral imaging technique and support vector machine. *Journal of Food Engineering*, 88 (2008) 474–483.



- WU, D., SUN, D.-W. Colour measurements by computer vision for food quality control – a review. *Trends in Food Science e Technology*, 29 (2013) 5–20.
- WU, W., MALLET, Y., WALCZAK, B., PENNINGCKX, W., MASSART, D. L., HEUERDING, S., ERNI, F. Comparison of regularized discriminant analysis, linear discriminant analysis and quadratic discriminant analysis applied to NIR data. *Analytica Chimica Acta*, 329 (1996) 257–265.
- YE, N. S. A mini review of analytical methods for the geographical origin analysis of teas (*Camellia sinensis*). *Critical Reviews in Food Science and Nutrition*, 52 (2012) 775–780.
- ZHAO, J., CHEN, Q., HUANG, X., FANG, C. H. Qualitative identification of tea categories by near infrared spectroscopy and support vector machine. *Journal of Pharmaceutical and Biomedical Analysis*, 41 (2006) 1198–1204.
- ZHAO, J., CHEN, Q., CAI, J., QUYANG, Q. Automated tea quality classification by hyperspectral imaging. *Applied Optics*, 19 (2009) 3557–3564.

Anexos





ANEXO 1: Tabela de resultados da 3ª aplicação

Anexo 1. Resultados das determinações de umidade, cinzas, cafeína, fluoreto, polifenóis e metais nas folhas e infusões de chá para as três classes estudadas: chá preto argentino, chá preto srilanquense e chá verde argentino. Todos os valores das determinações são expressos em base seca.

	Chá preto argentino			Chá preto srilanquense			Chá verde argentino		
	Faixa	Média	Desvio-padrão	Faixa	Média	Desvio-padrão	Faixa	Média	Desvio-padrão
Determinações nas folhas de chás									
Umidade (% m/m)	7,50 – 11,90	9,07	0,83	7,00 – 8,20	7,61	0,31	5,80 – 8,30	6,95	0,62
Cinzas (% m/m)	6,00 – 7,40	6,68	0,30	5,00 – 6,40	6,00	0,22	5,40 – 7,80	6,12	0,48
Cafeína (% m/m)	1,20 – 2,60	1,80	0,29	1,40 – 3,00	2,14	0,28	0,80 – 2,30	1,44	0,32
Al (mg/kg)	1110,80 – 3165,63	1641,65	343,63	291,17 – 1231,33	692,88	210,62	440,25 – 2169,14	1178,00	524,50
Ca (mg/kg)	2732,48 – 6387,19	4449,75	452,65	2576,31 – 4710,32	3589,95	512,64	2784,92 – 5178,88	4081,78	538,59
Cd (mg/kg)	0,04 – 0,39	0,22	0,06	0,10 – 0,29	0,18	0,04	0,10 – 0,40	0,23	0,08
Co (mg/kg)	0,05 – 0,24	0,16	0,03	0,03 – 0,24	0,12	0,07	0,09 – 0,24	0,19	0,04
Cr (mg/kg)	0,66 – 3,13	1,83	0,52	0,25 – 3,65	1,78	1,05	0,77 – 2,58	1,18	0,28
Cu (mg/kg)	0,58 – 26,73	12,84	2,46	11,31 – 23,29	17,64	3,82	6,15 – 27,56	13,54	4,42
Fe (mg/kg)	40,92 – 182,01	94,02	25,36	57,89 – 289,49	132,18	57,05	58,90 – 210,24	105,93	38,37

K (mg/kg)	8279,10 – 17202,37	13282,10	1262,03	11491,53 – 17334,59	14881,89	1299,46	5540,37 – 21676,20	13247,60	4188,58
Mg (mg/kg)	1094,08 – 1633,15	1396,67	88,66	734,99 – 1492,53	1046,79	219,20	85,28 – 1397,03	855,87	299,14
Mn (mg/kg)	537,27 – 1191,68	927,31	141,05	64,81 – 1164,29	383,31	321,17	195,93 – 1346,45	1017,33	192,84
Na (mg/kg)	21,87 – 46,22	33,84	6,12	32,26 – 75,82	55,87	12,10	35,74 – 71,89	53,62	12,13
Ni (mg/kg)	1,65 – 10,97	3,99	0,90	2,71 – 5,57	4,04	0,56	2,13 – 8,43	3,80	0,85
P (mg/kg)	865,77 – 2354,70	1952,14	213,91	1679,20 – 3348,80	2382,65	357,38	1383,76 – 2964,22	2008,69	413,04
Pb (mg/kg)	0,86 – 6,48	2,83	0,73	1,11 – 3,28	1,67	0,51	0,17 – 8,16	2,44	1,75
Zn (mg/kg)	4,57 – 26,62	18,82	3,07	14,45 – 34,91	23,97	5,82	12,34 – 30,19	18,97	4,87
Determinações nas infusões de chá									
Fluoreto (mg/kg)	73,90 – 182,80	109,56	18,53	95,20 – 666,40	415,98	150,74	90,40 – 410,40	168,19	82,38
Polifenóis (mg/kg)	6,56 – 9,60	7,77	0,68	7,78 – 15,11	10,95	1,93	7,94 – 14,87	10,49	1,83
Al (mg/kg)	50,92 – 133,62	97,22	12,67	13,93 – 82,47	45,52	15,90	19,44 – 86,99	62,78	17,05
Ca (mg/kg)	24,41 – 49,32	41,08	5,49	15,82 – 37,60	29,25	6,29	23,48 – 56,53	36,24	5,05
Cd* (mg/kg)	-	-	-	-	-	-	-	-	-
Co (mg/kg)	0,01 – 0,20	0,05	0,03	0,01 – 0,10	0,06	0,02	0,01 – 0,09	0,04	0,03

Cr (mg/kg)	0,01 – 0,21	0,11	0,04	0,01 – 0,09	0,03	0,03	0,01 – 0,06	0,02	0,02
Cu (mg/kg)	0,35 – 0,60	0,52	0,04	0,36 – 4,48	0,79	0,48	0,09 – 2,84	0,37	0,26
Fe (mg/kg)	0,16 – 1,32	0,40	0,24	0,19 – 0,38	0,26	0,05	0,04 – 0,62	0,42	0,12
K (mg/kg)	994,15 – 2531,88	1978,79	246,21	19,85 – 2439,53	1858,31	284,08	720,93 – 1577,86	1120,88	223,18
Mg (mg/kg)	34,34 – 176,84	118,01	24,41	12,33 – 182,32	127,08	21,27	45,24 – 153,47	83,75	33,16
Mn (mg/kg)	15,38 – 66,61	39,19	10,85	4,28 – 71,74	21,62	16,00	19,31 – 69,01	36,76	12,24
Na (mg/kg)	14,18 – 20,32	17,86	1,54	15,05 – 21,02	18,00	1,42	2,58 – 23,43	14,62	3,83
Ni (mg/kg)	0,25 – 0,54	0,43	0,06	0,21 – 0,65	0,41	0,13	0,18 – 0,58	0,38	0,06
P (mg/kg)	39,27 – 125,95	94,74	17,01	110,28 – 166,22	135,61	14,26	39,46 – 93,01	64,53	16,18
Pb (mg/kg)	0,01 – 0,67	0,35	0,18	0,01 – 0,66	0,22	0,22	0,01 – 0,46	0,17	0,17
Zn (mg/kg)	0,31 – 0,82	0,61	0,09	0,69 – 1,99	1,29	0,30	0,45 – 1,07	0,73	0,16

* Abaixo do limite de detecção.



ANEXO 2: Artigo publicado no periódico *Analytical Methods* referente à 1ª aplicação desenvolvida nesta Tese de Doutorado

Using a simple digital camera and SPA-LDA modeling to screen teas

Paulo H. G. D. Diniz,^a Hebertty V. Dantas,^a Karla D. T. Melo,^a Mayara F. Barbosa,^a David P. Harding,^a Elaine C. L. Nascimento,^a Marcelo F. Pistonesi,^b Beatriz S. F. Band^b and Mário C. U. Araújo†^{*a}

Received 11th May 2012, Accepted 6th July 2012

DOI: 10.1039/c2ay25481f

Classification or screening analysis of natural unprocessed teas using simple digital images and a variable selection algorithm is described. The proposed methodology uses color histograms generated on free downloadable software *ImageJ 1.44p* as a source of analytical information. Two chemometric methods were compared for classification of the resulting images, namely Soft Independent Modeling of Class Analogy (SIMCA), and Linear Discriminant Analysis (LDA) with variable selection by the Successive Projections Algorithm (SPA). The results were evaluated in terms of errors found in a sample set separate from the modeling process. The choice of more informative photometric color attributes (red-green-blue (RGB), hue (H), saturation (S), brightness (B), and grayscale) for screening the tea samples was made during the color modeling because SIMCA failed to give good results. Therefore the data treatment used SPA-LDA, which correctly classified all samples according to their geographical regions, whether from Brazilian, Argentinian or foreign soils.

1. Introduction

After water, teas from the leaves of *Camellia Sinensis* are the most widely consumed beverages in the world.¹ Various factors such as altitude, climate, soil, *etc.* affect the plant which leads to differing leaf textures, physical appearance, and chemical properties. Depending on the manufacturing process, teas are classified into three major types: non-fermented green tea, semi-fermented oolong tea, and fermented black tea.^{2,3} Tea has a wide variety of compounds such as flavonoids, phenolic acids, amino acids, chlorophyll, pigments, carbohydrates, organic acid, caffeine and other alkaloids, minerals, vitamins and enzymes.^{1,4} Unlike black and oolong teas, green tea is not fermented, and more of the active constituents remain in the leaves. In terms of consumption, black teas are the most consumed around the world, and their quality is judged mainly on the basis of their components and color.

India and China are first in world tea production, being favoured by both climatic and topographic characteristics.^{5,6} Argentinian production is around 70 000 tons per annum, 80% of which is sold in foreign markets. The province of *Misiones* in Argentina accounts for 90% of the country's output. Only 10% of the cultivated area is sown with seeds of national origin, the rest is planted with seeds from China and India⁷ which is a source of considerable variation in the quality of the crops. Though they lack flavour, Brazilian teas are ideal for blending, and the majority of Brazilian tea is produced for this purpose. About 70% of the total tea production is sold to the United States. During the 1970s, the annual tea production was about 11 000 tons, yet there has been a steady decline ever since. In recent years, tea producers have been focusing their efforts on increasing the quality of Brazilian teas and have seen an increase in market share.

Quality control of tea depends mainly on its appearance, taste, and aroma. Traditionally, human sensory assessment by skillful tasters is employed for evaluating tea quality. However, the results of this type of assessment are not precise and can be markedly partial.⁸ The "Digital Image", on the other hand, is an analytical tool with great potential for use in qualitative and quantitative analysis.^{9–13} This is one of the emerging frontiers of advanced research, and deals with the process of capturing, conditioning, and measurement of the Digital Camera based *digital image* by using advanced soft-computing algorithms that extract important information and features from the acquired images.^{14–16} The technique is non-invasive, facilitates the acquisition of data, and decreases costs. It also can reproduce a human evaluation in a standardized, impartial, and robust way.¹⁷

Many methodologies have been proposed to determine the physical/chemical properties of foods, using chemometrics, pattern recognition, and image analysis techniques. Applications for digital image processing techniques are expanding rapidly in the food processing industries.^{18,19} Multi- and hyperspectral imaging, artificial olfactory devices, and spectrometric techniques that use wavelet transform, neural networks, and support vector machines have been developed but are expensive, and overly complex.^{18,20–31}

Non-destructive classification is one of the most important applications for image analysis. Many studies have used Principal Component Analysis (PCA) to compress the information, and the derived latent variables are applied later with a classifier (*e.g.* Fisher's Linear Discriminant Analysis).^{32–34} From a database made up of useful classification variables, and applying Soft Independent Modelling of Class Analogy (SIMCA), and Partial Least Squares-Discriminant Analysis (PLS-DA), the dimension of the data structure

^aUniversidade Federal da Paraíba, CCEN, Departamento de Química, João Pessoa, PB, Brazil. E-mail: laqa@quimica.ufpb.br; Fax: +55 83 3216 7437; Tel: +55 83 3216 7438

^bUniversidad Nacional del Sur, Departamento de Química, Bahía Blanca, Argentina

† Present address: Department of Chemistry, CCEN, Federal University of Paraíba, Caixa Postal 5093, CEP 58051-970, Joao Pessoa, Brazil.

is reduced and the classification accuracy rate achieved by simpler classification techniques is improved upon.^{35,36} Linear Discriminant Analysis (LDA) as a classifier, with variable selection by the Successive Projections Algorithm (SPA)³⁷ has been successfully used for this purpose in different classification problems, including the analysis of edible vegetable oils,^{38,39} soils,⁴⁰ cigarettes,⁴¹ coffees,⁴² diesel/biodiesels,⁴³ alcohol fuels,⁴⁴ beers⁴⁵ and pen inks.⁴⁶

SIMCA is a well-known multivariate pattern recognition method which discriminates between different classes of samples,^{47–51} given various classes, classification rules are defined by a set of samples from known classes (the training set), together with distinct measurement values from each sample, and a sample validation set is included. The rules are then used to classify the new unknown samples (test set) on the basis of similar measurements.⁵² LDA classification methods employ linear decision boundaries (hyperplanes), which are defined in order to maximize the ratio of inter-class to intra-class dispersion.⁵³ In order to have a well-posed problem, the number of calibration (training) objects must be larger than the number of variables included in the LDA model. The use of LDA for classification of spectral data usually requires an appropriate variable selection procedure.^{37,41,54,55} The Successive Projections Algorithm (SPA) was adapted³⁷ for use in classification problems. In the original formulation,⁵⁶ the candidate variable subsets are formed as a result of projection operations intended to minimize multi-collinearity effects, which are known to cause poor performance in LDA.^{39,57}

This study aimed to classify or screen non-processed green and black teas purchased in Brazil and Argentina using simple images captured on a digital camera. The methodology used chemometrically evaluated color histograms generated on free software *ImageJ 1.44p*.⁵⁸ The chemometric evaluation utilized SPA in combination with LDA. The study determined a relationship between the digital image and geographical origin of the tea, this being useful for investigating possible sample adulterations. It could provide an additional tool for primary tea producers to secure both quality and a differentiated product.

2. Experimental

2.1. Samples and image acquisition

One hundred samples of green and black teas were purchased from local supermarkets; 40 Brazilian samples (20 each green and black), and 40 Argentinian samples (20 each green and black), and 20 imported black samples. For all of the above, five brands of each type whether green or black were purchased to assure variability or uniform distribution of the model.

The method of imaging takes into consideration the overall visual feature of the sample surface. All samples were homogenized, and photographed without treatment using a digital camera, Olympus® model SP-510 UZ, with 7.1 Megapixels. To achieve standardization, manufacturers resolve questions of color balance red-green-blue (RGB), hue (H), saturation (S), and brightness (B), as well as the grayscale intensity in ways particular for each camera model sold. These manufacturer's considerations do not affect the overall method, but rather the individual scores for each value; RGB, HSB, grayscale. A total of 300 images (3 for each sample) were obtained. To ensure reproducibility, a mechanical digital camera support and a circular fluorescent lamp were mounted to maintain fixed positioning, luminosity, sample-to-camera distance, and focus. Fig. 1

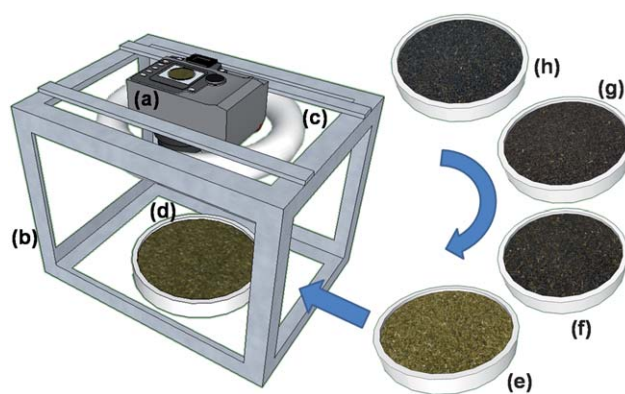


Fig. 1 Apparatus built for tea image capturing and examples of green and black tea samples. (a) Digital camera, (b) metallic support, (c) circular lamp, (d) Brazilian green, (e) Argentinian green, (f) Brazilian black, (g) Argentinian black, and (h) Imported black tea.

shows the apparatus built for image capturing, and actual examples of the five groups (Argentinian green and black, Brazilian green and black, and imported black teas) in line to be photographed. The digital camera was set above the tea vertically. The distance between the lens and the tea was 12 cm, and between the illumination and the tea was 10 cm. The digital camera was placed in the center of a circular fluorescent lamp. The sample holder used in this study was composed of polytetrafluoroethylene, which minimizes light scattering and fluorescence effects, and their effects on image color histograms. Therewith, the proposed methodology dispensed with the need for further image manipulation.

2.2. Histograms and data analysis

Color histograms describe the statistical distribution of the pixels as a function of the recorded color component, and not a physical-chemical behavior directly. The digital images were treated using free downloadable software *ImageJ 1.44p*, which created a histogram for each; red, green, blue, hue, saturation, brightness, and grayscale intensity. Fig. 2 shows an example of each histogram type (and its corresponding image), for a single green tea sample.

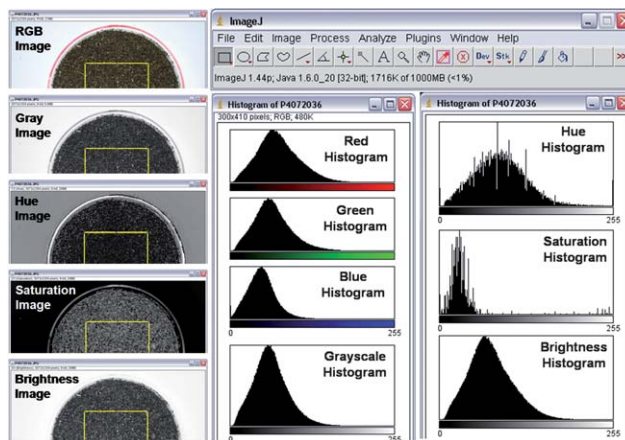


Fig. 2 The *ImageJ* software and the corresponding histograms and images for a green tea sample.

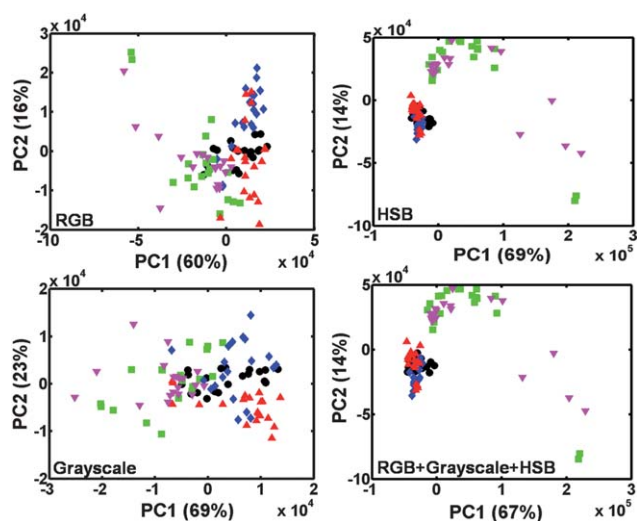


Fig. 3 PCA scores obtained from RGB, HSB, grayscale and RGB + grayscale + HSB histograms for all tea samples. Brazilian green tea (∇), Brazilian black tea (\blacktriangle), Argentinian green tea (\blacksquare), Argentinian black tea (\blacklozenge) and Imported black tea (\bullet). The variance for each principal component is indicated in parenthesis.

For data treatment we defined a 1000×1000 pixel square region at the center of each image, which represented about 14% of the total image area. Using only the selected region of the images, histograms employing red, green, blue, hue, saturation, brightness, and grayscale were constructed. Each color component of the models is composed of 256 tones, which are used as analytical information. In order to check for each color's relative influence, we selected four different color models employing (a) RGB, (b) HSB, (c) grayscale, (d) RGB + grayscale + HSB. The models were composed of (a) 3×256 , (b) 3×256 , (c) 1×256 and (d) $768 + 256 + 768$ variables, respectively. From three photos for each sample a mean histogram was calculated.

The analytical information extracted from the histograms above was employed to construct the chemometric classification models using SIMCA and SPA-LDA. The data obtained from each histogram were separated into: training (60%), validation (20%) and prediction (20%) sets using the Kennard–Stone algorithm.⁵⁹ The Kennard–Stone (KS) algorithm was applied separately to each class. It is a classic method to extract a representative set of objects from a given dataset by maximizing the minimal Euclidean distance between already selected objects and the remaining objects. The calibration,

validation and prediction samples were initially selected in accordance with the Kennard–Stone results.^{37,59–61} Chemometric data treatment was implemented with The Unscrambler® 9.7 (CAMO S/A), and Matlab® 2009b (Mathworks Inc.) software.

3. Results and discussion

3.1. Exploratory analysis

In Fig. 1, we can observe that it is difficult to distinguish between Fig. 1(d) and (e), in the case of green teas, and between Fig. 1(f)–(h) for black teas. Since there is variation in the colors of individual tea particles, the images are characterized by a natural stochastic image texture, *i.e.* they allow us to extract information about the spatial arrangement of color and intensities. Color histograms describe the statistical distribution of the pixels as a function of the recorded color component, and not physical–chemical properties.

An exploratory data analysis was made using principal component analysis employing the three color models RGB, HSB, and grayscale, and the one model using all the histograms together (RGB + grayscale + HSB), as described before. Fig. 3 shows the plotted scores of the two first principal components using these four colors models.

The (RGB) and (grayscale) plots present overlap between the samples, and very little separation. Plots (HSB) and (RGB + grayscale + HSB) present similar results which separate the samples into two major clusters of black and green tea, yet with some remaining overlap. This was expected since green and black teas are very different in color, surface texture, particle size, *etc.* The three black teas overlap more, while the two green teas are better resolved. This occurs because there is repetition of the image information, which requires variable selection to discover which tones of the histogram lead to a better classification model. A variable selection algorithm (SPA-LDA) was used to resolve this and consequentially to classify the tea samples. The SPA-LDA results were subsequently compared with the SIMCA classification results, as discussed in the next section.

3.2. Classification

Table 1 shows the SPA-LDA and SIMCA results with cost function values, for both green and black tea samples according to the geographical region for the four histogram classification models. Green teas were classified into Brazilian and Argentinian clusters; and black teas into Brazilian, Argentinian, and Imported clusters. Each class was composed of 20 tea histograms, which were separated into

Table 1 SIMCA and SPA-LDA classification results for green and black tea samples according to their geographical origin using color histograms in classification models^a

Histograms used in the classification models	SPA-LDA						SIMCA ^b				
	Selected variables/cost function value	No. of errors ^c (%)					No. of errors ^c (%)				
		Green	Black		Black		Green	Black		Black	
		Br	Ar	Br	Ar	Im	Br	Ar	Br	Ar	Im
RGB	20/0.74728	0	37.5	50.0	37.5	12.5	100	87.5	87.5	100	87.5
HSB	16/0.54008	0	0	12.5	37.5	0	100	100	100	87.5	87.5
Grayscale	12/0.81383	50.0	50.0	12.5	12.5	12.5	100	75.0	100	87.5	87.5
RGB + grayscale + HSB	17/0.54907	0	0	0	0	0	100	75.0	100	87.5	87.5

^a Br = Brazilian tea; Ar = Argentinean tea; Im = Imported tea. ^b 95% confidence level. ^c Validation + test errors.

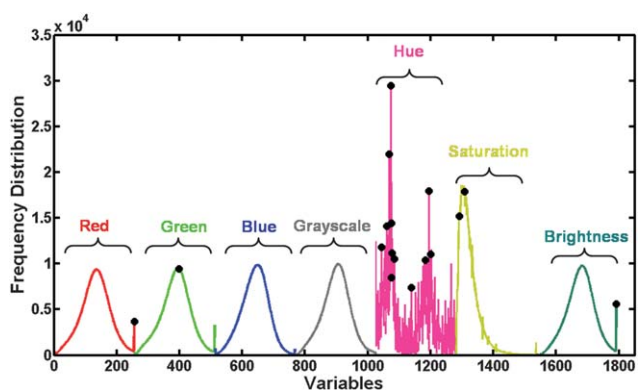


Fig. 4 The 17 variables selected by SPA-LDA (●) from the red, green, blue, grayscale, hue, saturation, and brightness histograms.

training (12), validation (4) and prediction (4) sets. The number of classification errors is presented as validation + prediction errors; their respective percentages are also shown. The SIMCA models were calculated with a 95% confidence level.

For the SIMCA models, all training, validation, and prediction samples separated by the Kennard–Stone algorithm were correctly classified into their respective class, but they are to a large degree classified into other classes as well. This can be ascribed to the high overlap between the classes, reaching errors of above 75%. SPA-LDA improved meaningfully the geographical discrimination ability of the classification models. Fig. 4 shows the 17 variables selected by SPA-LDA from the red, green, blue, grayscale, hue, saturation, and brightness histograms, which correspond to less than one percent of all the variables. These 17 variables (color tones, hue, saturation, and brightness) serve as the basis for a more intelligent discrimination between class characteristics. Although a similar separation pattern into two major classes (black and green teas) had been obtained using principal component analysis (see Fig. 3), the SIMCA model uses all of the information contained in the databases and its predictive ability becomes limited by superfluous information.

Fig. 5 shows plots of the four resulting SPA-LDA discriminant functions (DF), corresponding to the linear combination of the selected variables. We observed good discrimination between the green teas, into Brazilian and Argentinian clusters, and with respect to the black teas, the *Imported* tea class was separated from the Brazilian and Argentinian black tea clusters by the superior DF's, 3–4.

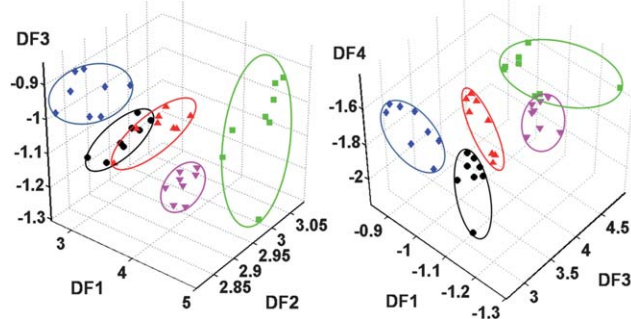


Fig. 5 Discriminant functions plot (DF1–4) obtained employing the SPA-LDA selected variables for Brazilian green tea (▼), Brazilian black tea (▲), Argentinian green tea (■), Argentinian black tea (◆), and Imported black tea (●).

4. Conclusions

This paper demonstrates the use of analytical information extracted from digital image generated color histograms. An appropriate variable selection algorithm (SPA), in combination with LDA, served as an analytical tool for discriminating natural green and black teas according to their geographical regions, whether from Brazilian and Argentinian, or foreign soils. The value of this method is that it approximates the expert/master's eye for discerning the origin of the tea being examined, yet remains impartial. The importance of sample positioning, lighting uniformity, and focal distance is worth mentioning, with these resolved, further image manipulation was unnecessary. The choice of an appropriate combination of histograms, and subsequent treatment using SPA-LDA reduced by about 99% the number of variables. The proposed method presents inherent advantages and requires no prior handling of the samples, no reagents, and generates no waste.

Acknowledgements

The authors gratefully acknowledge the Brazilian agencies CNPq and CAPES, and the Argentinian agencies CONICET and CIC for scholarships.

References

- 1 C. J. Dufresne and E. R. Farnworth, *J. Nutr. Biochem.*, 2001, **12**, 404–421.
- 2 M. Bancirova, *Food Res. Int.*, 2010, **43**, 1379–1382.
- 3 L. F. Wang, D. M. Kim and C. Y. Lee, *J. Agric. Food Chem.*, 2000, **48**, 4227–4232.
- 4 Y. Lu, W. F. Guo and Q. X. Yang, *J. Agric. Food Chem.*, 2004, **52**, 4472–4476.
- 5 M. A. Delavar, L. M. Sann, K. G. Lin, S. Tajuddin, B. S. Hassan and P. Hanachi, *Res. J. Int. Stud.*, 2008, **8**, 117–124.
- 6 D. L. McKay and J. B. Blumberg, *J. Am. Coll. Nutr.*, 2002, **21**, 1–13.
- 7 C. Anesini, G. E. Ferraro and R. Filip, *J. Agric. Food Chem.*, 2008, **56**, 9225–9229.
- 8 J. Zhao, Q. Chen, J. Cai and Q. Quyang, *Appl. Opt.*, 2009, **19**, 3557–3564.
- 9 A. Lopez-Molinero, D. Liñan, D. Sipiara and R. Falcon, *Microchem. J.*, 2010, **96**, 380–385.
- 10 W. S. Lyra, F. A. C. Sanches, F. A. S. Cunha, P. H. G. D. Diniz, S. G. Lemos, E. C. Silva and M. C. U. Araújo, *Anal. Methods*, 2011, **3**, 1975.
- 11 S. I. E. Andrade, M. B. Lima, I. S. Barreto, W. S. Lyra, L. F. Almeida, M. C. U. Araújo and E. C. Silva, *Microchem. J.*, 2012, DOI: 10.1016/j.microc.2012.03.029.
- 12 E. N. Gaião, V. L. Martins, W. S. Lyra, L. F. Almeida, E. C. Silva and M. C. U. Araújo, *Anal. Chim. Acta*, 2006, **570**, 283–290.
- 13 A. C. Sousa, A. S. Soares, C. J. Coelho, E. N. Gaião, A. F. C. Pereira, R. K. H. Galvão and M. C. U. Araújo, *Anal. Methods*, 2012, DOI: 10.1039/C2AY055377F.
- 14 T. Oikawa, M. Tomizawa and S. Degawa, *Control Eng. Pract.*, 1997, **5**, 75–78.
- 15 R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Addison-Wesley, Boston, MA, 3rd edn, 1992.
- 16 R. Berstein, *IBM J. Res. Dev.*, 1976, **20**, 40–56.
- 17 G. S. Gill, A. Kumar and R. Agarwal, *J. Food Eng.*, 2011, **106**, 13–19.
- 18 M. S. Zenoozian and S. Devahastin, *J. Food Eng.*, 2009, **2**, 219–227.
- 19 A. R. Tórres, W. S. Lyra, S. I. E. Andrade, R. A. N. Andrade, E. C. Silva, M. C. U. Araújo and E. N. Gaião, *Talanta*, 2011, **84**, 601–606.
- 20 N. Bhattacharyya, S. Seth, B. Tudu, P. Tamuly, A. Jana, D. Ggosh, R. Bandyopadhyay, M. Bhuyan and S. Sabhapandit, *Sens. Actuators, B*, 2007, **122**, 627–634.
- 21 S. Borah and M. Bhuyan, *Insight J.*, 2003, **35**, 55–58.
- 22 P. K. Mahanta and M. Hazarika, *J. Sci. Food Agric.*, 1985, **36**, 1133–1139.

- 23 M. Pietikäinen, T. Ojala and Z. Xu, *Pattern Recognit.*, 2000, **33**, 43–52.
- 24 X. Li, P. Nie, Z. J. Qiu and Y. He, *Expert Syst. Appl.*, 2011, **38**, 11149–11159.
- 25 Q. S. Chen, J. W. Zhao and J. R. Cai, *Trans. Am. Soc. Agric. Biol. Eng.*, 2008, **51**, 615–621.
- 26 Q. S. Chen, J. W. Zhao, C. H. Fang and D. M. Wang, *Spectrochim. Acta, Part A*, 2007, **66**, 568–574.
- 27 S. Borah, E. L. Hines and M. Bhuyan, *J. Food Eng.*, 2007, **79**, 629–639.
- 28 Q. S. Chen, J. W. Zhao, H. D. Zhang, M. H. Liu and M. Fang, *J. Near Infrared Spectrosc.*, 2005, **13**, 327–332.
- 29 D. Wu, X. J. Chen and Y. He, *Spectrosc. Spectral Anal.*, 2009, **29**, 1382–1385.
- 30 X. J. Chen, D. Wu, Y. He, X. L. Li and S. Liu, *Spectrosc. Spectral Anal.*, 2008, **28**, 2527–2530.
- 31 D. Wu, H. Q. Yang, X. J. Chen, Y. He and X. L. Li, *J. Food Eng.*, 2008, **88**, 474–483.
- 32 H. Yu and J. Yang, *Pattern Recognit.*, 2001, **34**, 2067–2070.
- 33 A. Moreda-Piñero, A. Fisher and S. J. Hill, *J. Food Compos. Anal.*, 2003, **16**, 195–211.
- 34 J. E. Lee, B. J. Lee, J. A. Hwang, K. S. Ko, J. O. Chung, E. H. Kim, S. J. Lee and Y. S. Hong, *J. Agric. Food Chem.*, 2011, **59**, 10579–10585.
- 35 J. M. Prats-Montalbán, A. Juan and A. Ferrer, *Chemom. Intell. Lab. Syst.*, 2011, **107**, 1–23.
- 36 F. López, J. M. Valiente, J. M. Prats and A. Ferrer, *Pattern Recognit.*, 2008, **41**, 1161–1172.
- 37 M. C. U. Araújo and R. K. H. Galvão, Linear regression modeling: variable selection, in *Comprehensive Chemometrics: Chemical and Biochemical Data Analysis (Four-Volume Set)*, ed. B. Walczak, R. T. Ferré and S. Brown, Elsevier, Amsterdam, The Netherlands, 1st edn, 2009, vol. 03, pp. 233–283.
- 38 M. J. C. Pontes, R. K. H. Galvão, M. C. U. Araújo, P. N. T. Moreira, O. D. Pessoa Neto, G. E. José and T. C. B. Saldanha, *Chemom. Intell. Lab. Syst.*, 2005, **78**, 11–18.
- 39 F. F. Gambarra-Neto, G. Marino, M. C. U. Araújo, R. K. H. Galvão, M. J. C. Pontes, E. P. Medeiros and R. S. Lima, *Talanta*, 2009, **77**, 1660–1666.
- 40 M. J. C. Pontes, J. Cortez, R. K. H. Galvão, C. Pasquini, M. C. U. Araújo, R. M. Coelho, M. K. Chiba, M. F. Abreu and B. E. Madari, *Anal. Chim. Acta*, 2009, **642**, 12–18.
- 41 E. D. Moreira, M. J. C. Pontes, R. K. H. Galvão and M. C. U. Araújo, *Talanta*, 2009, **79**, 1260–1264.
- 42 U. T. C. P. Souto, M. J. C. Pontes, E. C. Silva, R. K. H. Galvão, M. C. U. Araújo, F. A. C. Sanches, F. A. S. Cunha and M. S. R. Oliveira, *Food Chem.*, 2010, **119**, 368–371.
- 43 M. J. C. Pontes, C. F. Pereira, M. F. Pimentel, F. V. C. Vasconcelos and A. G. B. Silva, *Talanta*, 2011, **85**, 2159–2165.
- 44 A. C. Silva, L. F. B. L. Pontes, M. F. Pimentel and M. J. C. Pontes, *Talanta*, 2012, **93**, 129–134.
- 45 M. Ghasemi-Varnamkhashtia, S. S. Mohtasebic, M. L. Rodriguez-Mendez, A. A. Gomes, M. C. U. Araújo and R. K. H. Galvão, *Talanta*, 2012, **89**, 286–291.
- 46 C. S. Silva, F. S. L. Borba, M. F. Pimentel, M. J. C. Pontes, R. S. Honorato and C. Pasquini, *Microchem. J.*, 2012, DOI: 10.1016/j.microc.2012.03.025.
- 47 S. Wold and M. Sjostrom, *Chemometrics: Theory and Application*, ed. B. R. Kowalski, American Chemical Society, Washington, 1973, p. 243.
- 48 M. D. Derde and D. L. Massart, *Anal. Chim. Acta*, 1986, **191**, 181–192.
- 49 M. Sjostrom and B. R. Kowalski, *Anal. Chim. Acta*, 1979, **112**, 11–30.
- 50 R. De Maesschalck, A. Candolfi, D. L. Massart and S. Heuerding, *Chemom. Intell. Lab. Syst.*, 1999, **47**, 65–77.
- 51 S. Wold, K. Esbensen and P. Geladi, *Chemom. Intell. Lab. Syst.*, 1987, **2**, 37–52.
- 52 F. S. Oliveira, L. S. G. Teixeira, M. C. U. Araujo and M. Korn, *Fuel*, 2004, **83**, 917–923.
- 53 R. O. Duda, P. E. Hart and D. G. Stork, *Pattern Classification*, John Wiley, New York, 2nd edn, 2001.
- 54 A. R. Caneca, M. F. Pimentel, R. K. H. Galvão, C. E. Matta, F. R. Carvalho, I. M. Raimundo Jr, C. Pasquini and J. J. R. Rohwedder, *Talanta*, 2006, **70**, 344–352.
- 55 Y. Mallet, D. Coomans and O. de Vel, *Chemom. Intell. Lab. Syst.*, 1996, **35**, 157–173.
- 56 M. C. U. Araújo, T. C. B. Saldanha, R. K. H. Galvão, T. Yoneyama, H. C. Chame and V. Visani, *Chemom. Intell. Lab. Syst.*, 2001, **57**, 65–73.
- 57 T. Naes and B. H. Mevik, *J. Chemom.*, 2001, **15**, 413–426.
- 58 M. D. Abramoff, P. J. Magalhaes and S. J. Ram, *Biophotonics Int.*, 2004, **11**, 36–42.
- 59 R. W. Kennard and L. A. Stone, *Technometrics*, 1969, **11**, 137–148.
- 60 E. Bouveresse, C. Hartmann, D. L. Massart, I. R. Last and K. A. Prebble, *Anal. Chem.*, 1996, **68**, 982–990.
- 61 K. R. Kanduc, J. Zupan and N. Majcen, *Chemom. Intell. Lab. Syst.*, 2003, **65**, 221–229.



ANEXO 3: Artigo publicado no periódico *Talanta* referente às atividades desenvolvidas no Doutorado-Sanduíche



ELSEVIER

Contents lists available at SciVerse ScienceDirect

Talanta

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

Eco-friendly sonoluminescent determination of free glycerol in biodiesel samples

Paulo Henrique Gonçalves Dias Diniz^a, Marcelo Fabián Pistonesi^b, Mário César Ugulino de Araújo^{a,*}, Beatriz Susana Fernández Band^b

^a Universidade Federal da Paraíba, Departamento de Química, Laboratório de Automação e Instrumentação em Química Analítica/Quimiometria (LAQA), Caixa Postal 5093, 58051-970 João Pessoa, PB, Brazil

^b Laboratorio FIA, Departamento de Química, INQUISUR, Universidad Nacional del Sur, Bahía Blanca, Buenos Aires, Argentina

ARTICLE INFO

Article history:

Received 9 February 2013

Received in revised form

2 April 2013

Accepted 5 April 2013

Available online 12 April 2013

Keywords:

Flow-batch

Glycerol

Biodiesel

Online extraction

Sonoluminescence

Water cavitation.

ABSTRACT

This paper proposes a flow-batch methodology for the determination of free glycerol in biodiesel that is notably eco-friendly, since non-chemical reagents are used. Deionized water (the solvent) was used alone for glycerol (sample) extractions from the biodiesel. The same water was used to generate water-cavitation sonoluminescence signals, which were modulated by the quenching effect associated with the amount of extracted glycerol. The necessarily reproducible signal generation was achieved by using a simple and inexpensive piezoelectric device. A linear response was observed for glycerol within the 0.001–100 mg/L range, equivalent to 0.004–400 mg/kg free glycerol in biodiesel. The lowest measurable concentration of free glycerol was estimated at 1.0 µg/L. The selectivity of the proposed method was confirmed by comparing the shape and retention of both real and calibration samples to standard solution chromatograms, presenting no peaks other than glycerol. All samples (after extraction) are greatly diluted; this minimizes (toward non-detectability) potential interference effects. The methodology was successfully applied to biodiesel analysis at a high sampling rate, with neither reagent nor solvent (other than water), and with minimum waste generation. The results agreed with the reference method (ASTM D6584-07), at a 95% confidence level.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

The commercialization of bioenergy is beginning providing an effective way to overcome the problems of both petroleum scarcity (through fossil fuel depletion), and environmental degradation. Biodiesel fuels have chemical properties that are very similar to conventional diesel; as such, there is no need for motor adaptations or new equipment to enable their use when blending (diesel/biodiesel), or substituting diesel fuel [1–3].

Glycerol is a natural product of biodiesel refinery: for every 9 kg of biodiesel, 1 kg of crude glycerol is produced as a co-product [4]. Glycerol is soluble in water (> 500 g/L), slightly soluble in esters (biodiesel), almost colorless, odorless, viscous, hygroscopic, and has a high boiling point. Glycerol is widely used in the pharmaceutical, chemical, and food industries [5–7], being neither toxic, nor volatile, and nor flammable at low temperatures [8,9].

In general, a result of inadequate processing through insufficient separation of the glycerin phase or incomplete water-washing after phase separation, free glycerol may be present in the biodiesel.

Glycerol is the major impurity found in commercial biodiesel; it can occlude fuel filters, impairing engine performance, cause fuel tank damage, and it can also release acrolein into the environment [10–13].

International regulations specify a limit for free glycerol in biodiesel of 0.020% (w/w) [14,15]. Considering the increasing use of biofuels, the need for fast, precise, reliable, and inexpensive analytical methods for the determination of free glycerol in biodiesel is clear. Several methods for determining free glycerol content in biodiesel have been developed based on chromatographic, amperometric, and spectrophotometric techniques [16–21]. However, these procedures are generally laborious, expensive, and typically use various chemicals, generating large amounts of waste. The development of alternative analytical methods, which require neither solvents nor reagents, and use both automation and miniaturization, has minimized substantially the adverse impacts once sustained by the environmental [22].

Flow-batch systems [23] explore programed (in-flow) multi-commutation to achieve very high sampling rates, and discrete (or batch) analytical precision. It can accommodate markedly different analytical procedures without significantly changing the physical configuration of the system, and modifying operational parameters becomes a function of programed valve-switch timing control.

During the last few years, many new analytical procedures using the sonochemical effect for a wide range of applications have

* Corresponding author. Tel./fax: +54 29 14595100x3557.

E-mail addresses: laqa@quimica.ufpb.br, mariougulino@gmail.com (M.C.U. de Araújo).

emerged, especially those involving degradation, nanoparticle formation, emulsification, and polymerization processes [24–28]. There are three classes of sonochemical reactions: homogeneous liquid sonochemistry; heterogeneous liquid–liquid, or liquid–solid sonochemistry; and sonocatalysis (which overlaps both) [28]. The underlying mechanism of sonochemistry is cavitation. The sonoluminescence (SL) occurs when brief light emissions occur due to the collapse of cavitation bubbles in a liquid irradiated with ultrasound. During a typical bubble collapse, high temperature and pressure are reached, which can generate highly reactive species (radicals). In the present case, ultrasonic waves cause water decomposition, and produce OH^\bullet radicals [28–31].



Light emission occurs from a transition



which is characteristic of an excited hydroxyl deactivation. Generally, OH^* is produced by exciting water to a level of $^1\text{A}_2$ [30].

Depending on the type of surface-active solutes, sonoluminescence intensity can either be enhanced (e.g., with low concentrations of ionic surfactants), or quenched (e.g., with certain aliphatic alcohols). SL quenching in water with aliphatic alcohols, including glycerol, may be related to the air/water interface concentration of solute molecules on the inside surface of the collapsing bubble. The most commonly accepted mechanism of SL quenching describes bubble collapse thermal energy, being consumed by endothermic bond breaking reactions of the solute molecules. A lower temperature inside the bubble occurs, producing a lower number of excited-state molecules, and consequently leading to a lower sonoluminescent intensity [28,32,33].

This paper proposes a notably eco-friendly flow-batch (FB) method for sonoluminescent determination of free glycerol in biodiesel. The proposed flow-batch sonoluminescent (FBS) system uses an online glycerol extraction method (from biodiesel) with deionized water as the single extraction solvent. The same water was also employed to generate analytical signals through water-cavitation sonochemiluminescence, which are modulated by quenching, a result proportional to the amount of extracted glycerol in the water. A reproducible water-cavitation signal generation was achieved using a simple and inexpensive piezoelectric device. This approach is an environmentally safe methodology, and is the first to completely avoid the use of any chemical reagents whatsoever during the entire analytical process of free glycerol determination in biodiesel samples.

2. Experimental

2.1. Solution and samples

Stock solution was prepared by dissolving 0.10002 g glycerol (99.5%, Mallinckrodt[®]) in 1000 ml of water. The glycerol-containing homogenized soybean biodiesel samples from different lots were purchased from the Petrobras[®] refinery located in Bahía Blanca City, in Argentina. Distilled–deionized water (18 M Ω /cm) was used exclusively throughout the entire study.

2.2. Flow-batch assembly

A schematic diagram of the flow-batch system used for sonoluminescent determination of free glycerol is shown in Fig. 1.

The FBS system consists of three three-way solenoid valves ($V_{\text{G/B}}$, $V_{\text{H}_2\text{O}}$, and V_{W}) (model 161T031, NResearch[®]) polyethylene tubing connectors with 0.8 mm i.d., and a peristaltic pump (model M312,

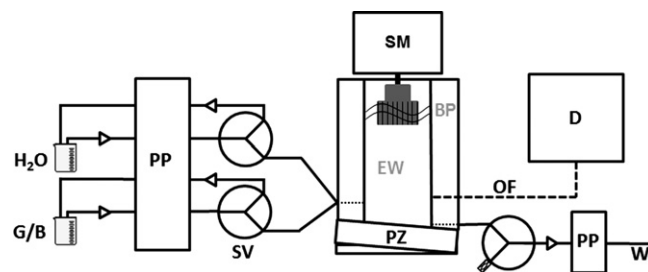


Fig. 1. Flow diagram of the flow-batch system for free glycerol sonoluminescent determination. SV—solenoid valve (three-way); PP—peristaltic pump; SM—stepper motor; BP—biodiesel phase, EW—extractive water, PZ—piezoelectric device; OF—optical fiber; D—lab-built luminometer; W—waste; H₂O: water; G/B—glycerol standard/biodiesel sample.

Gilson[®]) equipped with six pumping channels, and Tygon[®] pumping tubes (1.29 and 2.06 mm i.d.).

The lab-constructed flow-batch chamber was built in PTFE with ca. 2.0 mL of total volume. The chamber has a fixed stirrer on the top, which is turned by a stepper motor (MDN3, 200 rpm, 9 V DC), and has a piezoelectric device (taken from a commercial ultrasonic nebulizer ASPEN[®] NU400 [34]) at the bottom of the chamber. Fig. 2 displays the circuit for controlling the piezoelectric frequency. The chamber was connected by a fiber optic (P400-025-UV/SR, Ocean Optics[®]) probe to a lab-constructed luminometer, with a photomultiplier tube (PMT), (CE-A59/RX, 33-29-71 PMT, Bausch & Lomb[®]). The PMT was operated at 700 V, and the emission of photons was registered at 425 nm. The analytical signals were recorded by using a Data Chart 1250 Paperless Recorder (Monarch Instrument[®]).

A microcomputer connected with a USB interface (USB6009, National Instruments[®]) was used to control the FBS system, running software developed in LabView[®] 8.0 graphics language (National Instruments[®], Austin, TX, USA).

2.3. Analytical procedure

The operating steps carried out by the FBS system are shown in Table 1.

Before starting the analysis, the channels are filled with the appropriate fluids. The valves $V_{\text{G/B}}$, $V_{\text{H}_2\text{O}}$ and V_{W} are initially switched OFF, so that the solutions are continuously pumped to load the channels and are continuously returned to their respective containers. Valves $V_{\text{G/B}}$ and $V_{\text{H}_2\text{O}}$ are then simultaneously switched ON for 2 s, and the solutions are pumped toward the flow-batch (FB) chamber to fill the channels between the valves and the chamber. Then, V_{W} is switched ON, and the excess solution in the chamber is aspirated as waste for 5 s.

Deionized water is then inserted into the chamber, and the piezoelectric device is activated for 5 s in order to obtain the sonoluminescence signals through water cavitation. An optical fiber (P400-025-UV/SR Ocean Optics[®]) is used to transmit the emitted photons from the FB chamber to the detector.

The online (in chamber) preparation of calibration solutions (0.001–100 mg/L glycerol, being equivalent to 0.004–400 mg/kg free glycerol in biodiesel) are performed using the stock solution. In these preparations, only valve-switch-timing controls ($V_{\text{G/B}}$ and $V_{\text{H}_2\text{O}}$) are employed to insert differing amounts of stock solution, and water into the chamber. While the solutions are pumped into the chamber, the stepper motor is switched ON. Afterward, the stepper motor is stopped and the piezoelectric device is activated to sonicate the standard working solutions for 5 s. The SL signals are then recorded, with decreases in intensity being proportional to the increases in the concentration of glycerol. A wash cycle

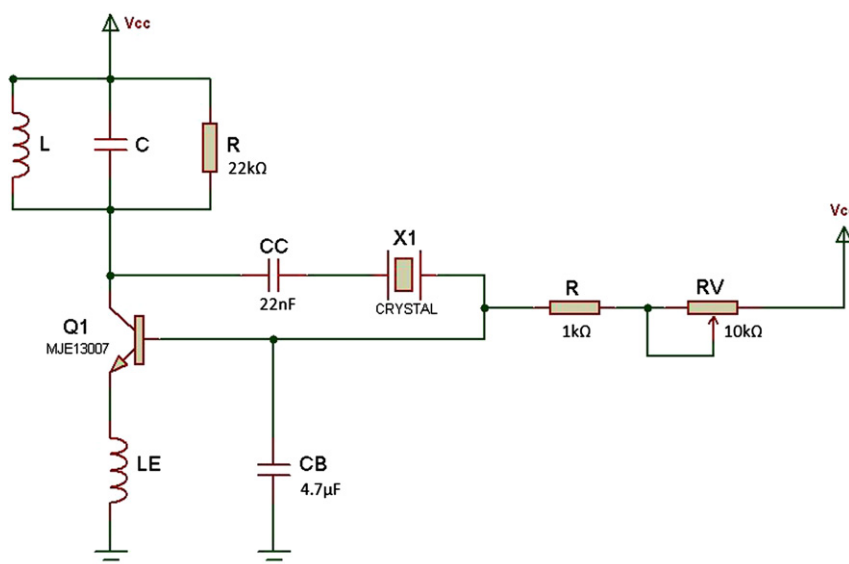


Fig. 2. Electronic circuit for piezoelectric frequency control.

Table 1
Operation of the FBS system for extraction and determination of free glycerol in biodiesel.

Step	Event	Time (s)	Volume (μL)	Pump rotation speed (rpm)	Pump tube (mm i.d.)
1	Biodiesel ($V_{G/B}$)	1.0	14.8	10.0	1.29
	Water (V_{H_2O})	3.9	985.2	48.0	2.06
2	Stirrer time ^a	240	–	0.0	–
3	Detection	5.0	–	0.0	–
4	Waste (V_W)	4.2	–	48.0	2.06
5 ^b	Water (V_{H_2O})	2.9	690.5	48.0	2.06
6 ^b	Waste (V_W)	4.2	–	48.0	2.06

^a Extraction time of the free glycerol in biodiesel.

^b Chamber cleaning.

(with water) is performed after each measurement. The piezoelectric device is turned ON during the wash step.

The free glycerol extraction from the biodiesel samples was carried out (using water) inside the FB chamber (Fig. 1) as described by Lima et al. [21]. After extraction, the stepper motor is stopped and the piezoelectric device is activated to sonicate the solutions for 5 s. The analytical signals are then recorded.

2.4. Reference method

The chromatographic procedure was carried out according to the reference method (D6584-07 “Standard Test Method for the Determination of Free and Total Glycerin in B-100 Biodiesel Methyl Esters by Gas Chromatography”) [35]. A gas chromatograph, (Agilent Technologies[®] 6890 GC) with a flame ionization detector (FID), a programmed temperature vaporizing (PTV) inlet (CIS 4, GERSTEL[®]), Dual Rail MPS 2 robotic sampler with 10 μL on-column syringe, and an 80 μL sideport syringe with a diluter module (GERSTEL[®]) were used. An Rtx-Biodiesel TG Restek column (10 m \times 0.32 mm i.d.) was also used. Each sample solution was injected in triplicate, with a free glycerol retention time of 4.1 min, and the concentrations were calculated from a calibration curve.

3. Results and discussion

3.1. Study of the piezoelectric frequency

In order to generate reproducible SL signals, we investigated three operating frequencies for the piezoelectric device: (a) 320 kHz, (b) 358 kHz, and (c) 385 kHz. For this purpose, three glycerol solutions 1.0, 10, and 30 mg/L were used. No SL production at 320 kHz was observed, whereas at 385 kHz the analytical signals were no reproducible. The best frequency for the study was found while operating in a continuous mode of 358 kHz, which corroborates with the work of Sunartio et al. [29].

3.2. Variation of the temperature caused by piezoelectric

A study of possible piezoelectric-caused temperature variations in the standard solutions and samples was carried out (before and after) for the sonication time of 5 s. As the cavitation time is short, and the piezoelectric frequency is low, no significant temperature variation was observed. Each experiment was made in triplicate at room temperature (20 ± 1 °C). The thermocouple (MAX6675[®] with sensitivity of 41 $\mu\text{V}/^\circ\text{C}$) was removed from the FB chamber during sonication to avoid damage.

3.3. Location of the optical fiber

As indicated in Fig. 1, the optical fiber is located at the inferior lateral of the mixing chamber, in front of the aqueous phase (extractive water). Since the biodiesel phase (immiscible fraction) is located on top of the mixing chamber, there is no light scattering. In addition, the system no needs no time for phase separation; emulsion is avoided by a slow rotation speed (200 rpm), as was done by Lima et al. [21].

3.4. Analytical curve

A calibration curve was obtained for $I=0.302 (\pm 0.011) - 0.101 \log C (\pm 0.003)$, $r=0.996$, in which I is the relative SL intensity, and C is the glycerol concentration in g/L. A linear response was observed within the range of 0.001–100 mg/L glycerol, which is equivalent to 0.004–400 mg/kg free glycerol in biodiesel. Compared to the previously published works [12,18–21],

the proposed method provides a lower measurable concentration of glycerol (1.0 $\mu\text{g/L}$).

3.5. Determination of the free glycerol content in the real samples

Table 2 shows the results of the determinations of the free glycerol content in commercial biodiesel samples. The precision of the proposed method was estimated determining the relative standard deviation (RSD) for the repeated measurement of authentic triplicates of different real samples, acquired and analyzed on different days, by both the reference and proposed methods at 1.5% [36]. Accuracy was evaluated comparing results from the proposed method with those of a reference method. For this purpose, an ordinary least squares (OLS) fitting was obtained, and the estimated intercept (a) and slope (b) were compared (with their ideal values of 0 and 1) using the elliptical joint confidence region (EJCR) test [21,37]. The intercept $-(-0.02 \pm 0.13)$, and the slope (1.00 ± 0.03) values demonstrate that the joint confidence region certainly contains the theoretical (0, 1) point ($\alpha=0.05$, 2, $n-2=4$). Fig. 3 shows the EJCR plot for the proposed method. The concentrations of free glycerol obtained by our proposed method for all the biodiesel samples were in a good agreement with those obtained by gas chromatography.

Table 2

Determinations of free glycerol content (mg/L) in commercial biodiesel samples by using the proposed method and the reference method. The values of uncertainty were estimated by using the expression $\pm t_{N-1} s/\sqrt{N}$, where N is the number of replicate measurements, t_{N-1} is the statistic parameter often called Student's t (with $N=3$, at 95% level of confidence) and s is the standard deviation.

Samples	Proposed method (mg/L)	Reference method (mg/L)
(1)	3.29 ± 0.06	3.10 ± 0.08
(2)	8.74 ± 0.03	8.80 ± 0.05
(3)	0.12 ± 0.01	0.13 ± 0.01
(4)	0.19 ± 0.04	0.18 ± 0.00
(5)	0.21 ± 0.03	0.22 ± 0.00
(6)	0.20 ± 0.04	0.21 ± 0.01

3.6. Analytical features

Table 3 compares the analytical features of the proposed method with other procedures described in the literature. In relation to the others, the proposed method has a great advantage: no chemical reagents are employed in any phase of analysis. Deionized water alone is used as the solvent for online extraction of glycerol from the biodiesel, as was done by Lima et al. [21]. The same water is then employed as the medium, generating the analytical signal, and the extracted glycerol quenches the sonoluminescence produced during water cavitation. Reproducible signals were generated under robust conditions due to the use of a simple and inexpensive piezoelectric device.

In the process of the biodiesel production, methanol (or ethanol) reacts with the triglyceride (vegetable oil) using a strong alkaline catalyst (generally NaOH or KOH) to give the corresponding mono-alkyl methyl esters (biodiesel), and glycerol in a 9:1 ratio [4]. The reagents (alcohol and catalyst) used in the transesterification reaction remain present in the biodiesel as residuals. After production, the biodiesel is washed exhaustively with water in order to minimize undesirable residues of both reagents and co-products before the commercialization. However, glycerol remains as the major impurity. The selectivity of the proposed method was confirmed by comparing the shape and the retention of real and calibration samples with chromatograms of standard solutions, and no peaks other than glycerol were observed. Extensive (in chamber) dilution of the real samples is always performed after extraction, which greatly minimizes the probability of finding alcohol or catalyst within the limits of detection. Therefore, the effect of other possible interferences in the sonoluminescence needed not be evaluated.

4. Conclusion

The processes of both extraction and determination of glycerol can be controlled by implementing the flow-batch methodology, and the strategy represents an excellent alternative to conventional flow-based procedures. Automatic preparation of the calibration curve, along with the online extraction and detection (using only water) make the proposed system extremely attractive for routine analysis. Waste generation is reduced significantly,

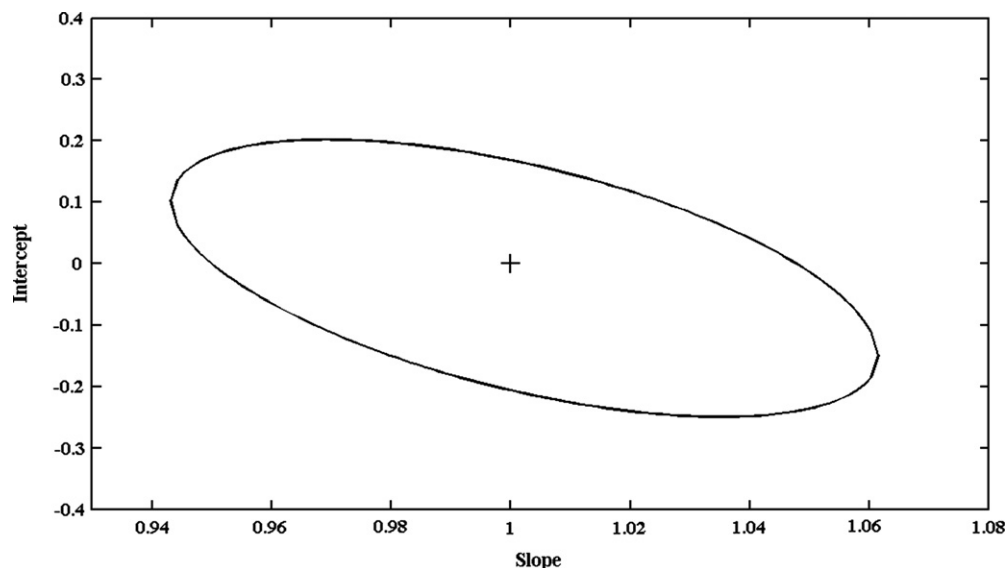


Fig. 3. Elliptical joint confidence region for the slope (b) and intercept (a) corresponding to the free glycerol content determinations of the proposed method. The cross marks the theoretical point ($a=0$, $b=1$).

Table 3
Comparative features of different methods for determination of free glycerol in biodiesel.

	MC [12]	FIA [18]	MC [19]	Batch [20]	FBF [21]	Proposed FBS
Detection limit [mg kg ⁻¹]	4.0	5.0	2.0	2.0	0.144	4.0 × 10 ⁻⁶
Sampling rate (h ⁻¹)	34	90	35	4	14	14
Reagents	Water, Acetic acid, Ammonium acetate, Acetylaceton Sodium periodate	Water, Sodium hydroxide	Water, Hydrochloric acid, Ammonium acetate, Sodium hydroxide, Acetylaceton Sodium periodate	Water, Ethanol, Acetic acid, Ammonium acetate, Potassium periodate, Acetylaceton	Water, Acetic acid, Ammonium acetate, Potassium periodate, Acetylaceton, Isooctane	Water
Extraction	Offline	Offline	Offline	–	Online	Online
Sample (mg)	1000	250	1000	1000	15.4	15.4
Instrumentation	Spectrophotometer	Potentiometer	Spectrofluorometer	Spectrophotometer	Spectrofluorometer	Lab made luminometer

MC—Multicommution in Flow Analysis; FIA—Flow-Injection Analysis; FBF—Flow-Batch with Fluorescence Detection; FBS—Flow-Batch with Sonoluminescence Detection.

a primary objective of green chemistry. Joining the principles of sonochemistry to the flow-batch methodology provided high sample rates, yet without the use of either chemical reagents or solvents. The new method was successfully applied to biodiesel analysis, and the results agreed with those of the reference method (ASTM D6584-07) at a 95% confidence level.

Acknowledgments

The authors gratefully acknowledge the Universidad Nacional del Sur, and CONICET (Consejo Nacional de Investigaciones Científicas y Técnicas) for their financial support. Researcher M.F. Pistonesi is also grateful to CIC (Comisión de Investigaciones Científicas de la Provincia de Buenos Aires). The authors acknowledge the support of CAPES/SPU (Brazil/Argentina international cooperation grant PPCP 013/2011), and CAPES (studentships). The authors also thank David Harding for reviewing the English in this article.

References

- [1] A.E. Atabani, A.S. Silitonga, H.C. Ong, T.M.I. Mahlia, H.H. Masjuki, I.A. Badruddin, H. Fayaz, *Renew. Sust. Energy Rev.* 18 (2013) 211–245.
- [2] H. Daming, Z. Haining, Lin Lin, *Energy Procedia* 16 (2012) 1874–1885.
- [3] A. Demirbas, *Energy Convers. Manage.* 50 (2009) 14–34.
- [4] N. Pachauri, B. He, Value added utilization of crude glycerol from biodiesel production: a survey of current research activities, in: *Proceedings of the ASABE Annual International Meeting*, Portland, USA, 2006.
- [5] A. Košmider, K. Leja, K. Czaczyk, *Biodiesel—Quality, Emissions and By-Products*, in: G. Montero, M. Stoytcheva (Eds.), *InTech*, Rijeka, Croatia, 2011, pp. 341–364.
- [6] M. Pagliaro, M. Rossi, *The Future of Glycerol*, 2nd ed., RSC Publishing, Cambridge, UK, 2010.
- [7] S. Abad, X. Turon, *Biotech. Adv.* 30 (2012) 733–741.
- [8] J.F. Izquierdo, M. Montiel, I. Palés, P.R. Outón, M. Galán, L. Jutglar, M. Villarrubia, M. Izquierdo, M.P. Hermo, X. Ariza, *Renew. Sust. Energy Rev.* 16 (2012) 6717–6724.
- [9] A.B. Leoneti, V.A. Leoneti, S.V.W.B. Oliveira, *Renew. Energy* 45 (2012) 138–145.
- [10] D.R. Lide, *CRC Handbook of Chemistry and Physics*, 85th ed., CRC, California, 2005.
- [11] I.P. Lôbo, S.L.C. Ferreira, R.S. Cruz, *Quim. Nova* 32 (2009) 1596–1608.
- [12] S.G. Silva, F.R.P. Rocha, *Talanta* 83 (2010) 559–564.
- [13] M. Lapuerta, O. Armas, R. Garcia-Contreras, *Energy Fuels* 23 (2009) 4343–4354.
- [14] United States Patent Application Publication US 2004/0137546A1.
- [15] ASTM D 6751, Test method for determination of free and total glycerin in B 100 biodiesel methyl esters by gas chromatography, ASTM International. Available from: (www.astm.org).
- [16] C.P. Prados, D.R. Rezende, L.R. Batista, M.I.R. Alves, N.R. Antoniosi Filho, *Fuel* 96 (2012) 476–481.
- [17] A.N. Dias, M.B.R. Cerqueira, R.R. Moura, M.H.S. Kurz, R.M. Clementin, M.G.M. D'Oca, E.G. Primel, *Fuel* 94 (2012) 178–183.
- [18] A.H. Maruta, T.R.L.C. Paixão, *Fuel* 91 (2012) 187–191.
- [19] S.G. Silva, A. Morales-Rubio, M. La Guardia, F.R.P. Rocha, *Anal. Bioanal. Chem.* 401 (2011) 365–371.
- [20] M.S. Ribeiro, F.R.P. Rocha, *Microchem. J.* 106 (2013) 23–26.
- [21] M.B. Lima, M. Insausti, C.E. Domini, M.F. Pistonesi, M.C.U. Araújo, B.S.F. Band, *Talanta* 89 (2012) 21–26.
- [22] S. Armenta, S. Garrigues, M. de la Guardia, *Trends Anal. Chem.* 27 (2008) 497–511.
- [23] P.H.G.D. Diniz, L.F. Almeida, D.P. Harding, M.C.U. Araújo, *Trends Anal. Chem.* 35 (2012) 39–49.
- [24] F. Priego-Capote, M.D. Luque de Castro, *Trends Anal. Chem.* 23 (2004) 644–653.
- [25] J.P. Lorimer, T.J. Mason, *Applied Sonochemistry: Uses of Power Ultrasound in Chemistry and Processing*, Wiley-VCH, Darmstadt, Germany, 2002.
- [26] D.M. Stasiak, Z.J. Dolatowski, *Pol. J. Food Nutr. Sci.* 57 (2007) 133–136.
- [27] D.L. Miller, N.B. Smith, M.R. Bailey, G.J. Czarnota, K. Hynynen, I.R.S. Makin, *J. Ultrasound Med.* 31 (2012) 623–634.
- [28] Y. Moliner-Martínez, R.A. González-Fuenzalida, R. Herráez-Hernández, P. Campíns-Falcó, J. Verdú-Andrés, *J. Chromatogr. A* 1263 (2012) 43–50.
- [29] D. Sunartio, F. Grieser, M. Ashokkumar, *Ultrason. Sonochem.* 16 (2009) 23–27.
- [30] F. Caupin, E. Herbert, C. R. Phys. 7 (2006) 1000–1017.
- [31] M.A. Margulis, *Ultrasonics* 23 (1985) 157–169.
- [32] G.J. Price, M. Ashokkumar, F. Grieser, *J. Am. Chem. Soc.* 126 (2004) 2755–2762.
- [33] A. Brotchie, J. Schneider, R. Pflieger, D. Shchukin, H. Möhwald, *Chem. Eur. J.* 18 (2012) 11201–11204.
- [34] *Ultrasonic Nebulizer ASPEN[®] NU400. Operation Manual.* (<http://www.aspen-salud.com.ar/manuales/Manual%20NU400.pdf>).
- [35] ASTM D6584-07, Test method for the determination of free and total glycerin in B 100 biodiesel methyl esters by gas chromatography, ASTM International, Available from: (www.gerstel.com).
- [36] P. Araujo, *J. Chromatogr. B* 877 (2009) 2224–2234.
- [37] D.J. Massart, B.M.G. Vandeginste, L.M.C. Buydens, S. De Jong, P.J. Lewi, J. Smeyers-Verbeke, *Handbook of Chemometrics and Qualimetrics: Part A*, 1st ed., Elsevier, Amsterdam, 1997.



ANEXO: Lista de referências das imagens utilizadas

Imagem de fundo da numeração das páginas:

Green tea pot. Disponível em: <http://www.mindbodygreen.com/0-5314/9-Reasons-Why-You-Should-Drink-Green-Tea.html>

FIGURA 1.1:

World tea production, FAO 2008. Disponível em:

<http://upload.wikimedia.org/wikipedia/commons/0/0b/WorldTeaProductionFAO2008.svg>

FIGURA 1.2:

Pren. *Tiếng Việt: Cây chè*. Disponível em:

https://upload.wikimedia.org/wikipedia/commons/9/93/Camellia_sinensis_%282%29.jpg

Dave Dahl. *Organic mountain-grown tea leaf*. Disponível em:

http://upload.wikimedia.org/wikipedia/commons/d/d7/Organic_mountain_grown_tea_leaf.jpg

FIGURA 1.3:

Disponível em: http://www.allposters.com.br/-sp/Tea-Tea-Plantations-Munnar-Kerala-India-posters_i8942855_.htm

Hermann Hammer. *Tea of different fermentation: From left to right: Green tea*

(Bancha from Japan), Yellow tea (Kekecha from China), Oolong tea (Kwai flower from China) and Black tea (Assam Sonipur Bio FOP from India). Disponível em:

http://upload.wikimedia.org/wikipedia/commons/0/04/Tea_in_different_grade_of_fermentation.jpg

FIGURA 1.4:

Disponível em: <http://www.teasun.com.au/rare-unique-tea/>