



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



DISSERTAÇÃO DE MESTRADO

**Aplicações de Imagens Digitais e Análise Multivariada para
Classificação e Determinação de Parâmetros de Qualidade
em Plumas de Algodão**

MARIA IVANDA SILVA GONÇALVES

JOÃO PESSOA –PB – BRASIL
AGOSTO/2015



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



DISSERTAÇÃO DE MESTRADO

Aplicações de Imagens Digitais e Análise Multivariada para Classificação e Determinação de Parâmetros de Qualidade em Plumas de Algodão

Maria Ivanda Silva Gonçalves*

Dissertação de Mestrado apresentada ao
Programa de Pós-Graduação em Química
da Universidade Federal da Paraíba, em
cumprimento as exigências para obtenção
do Título de Mestre em Química.

Área de Concentração: Química Analítica.

Orientador: Prof. Dr. Márcio José Coelho de Pontes (UFPB)

2º Orientador: Dr. Everaldo Paulo de Medeiros (EMBRAPA)

***Bolsista**



JOÃO PESSOA – PB – BRASIL

AGOSTO/2015

G635a Gonçalves, Maria Ivanda Silva.

Aplicações de imagens digitais e análise multivariada para classificação e determinação de parâmetros de qualidade em plumas de algodão / Maria Ivanda Silva Gonçalves.- João Pessoa, 2015.

110f. : il.

Orientadores: Márcio José Coelho de Pontes e Everaldo Paulo de Medeiros

Dissertação (Mestrado) - UFPB/CCEN

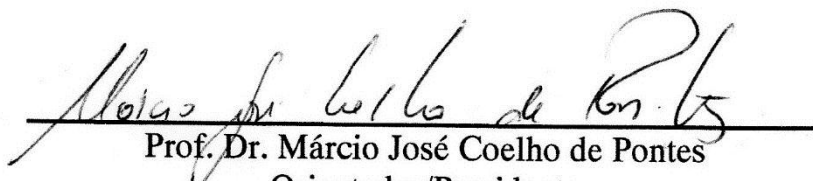
1. Química analítica. 2. Plumas de algodão. 3. Imagens digitais. 4. Seleção de variáveis. 5. Classificação multivariada. 6. Calibração multivariada.

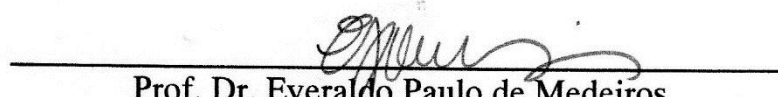
UFPB/BC

CDU: 543(043)

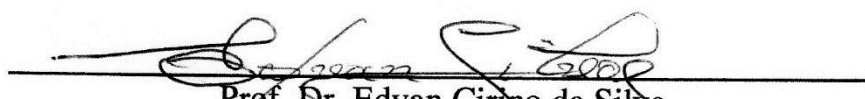
Aplicações de imagens digitais e análise multivariada para classificação e determinação de parâmetros de qualidade em plumas de algodão.

Dissertação de Mestrado apresentada pela aluna Maria Ivanda Silva Gonçalves e aprovada pela banca examinadora em 31 de agosto de 2015.


Prof. Dr. Márcio José Coelho de Pontes
Orientador/Presidente


Prof. Dr. Everaldo Paulo de Medeiros
2º. Orientador


Prof. Dr. Francisco Fernandes Gambarra Neto
Examinador


Prof. Dr. Edvan Cirino da Silva
Examinador

Dedicatória

A Deus “...Meu refúgio e minha fortaleza meu Deus eu confio em ti”!...(salmo 91: 2).

A minha mãe com a qual aprendi o que é amor, fé, verdade, a ser responsável, a dar valor as coisas, a conquistar meus sonhos, você que sempre me protege, perto ou longe, com suas forças e orações, é um exemplo em minha vida, onde sempre dizia “eu quero ser igual a minha mãe”, a que sempre tentou dar a melhor educação a suas filhas e enfrentou e enfrenta tantas tribulações, com uma força que só Deus sabe explicar. Obrigada por tudo, você é e sempre será: A melhor mãe do mundo! Eu te amo! E sou feliz porque você existe.

Com muito amor dedico.

AGRADECIMENTOS

A **Deus** e a **Nossa Senhora Aparecida**, por sempre estarem ao meu lado, ouvirem minhas preces, pela força, por me iluminarem, e permitirem a realização deste trabalho.

Aos meus pais **Ivã** e **Euza** em especial a minha mãe por ser um exemplo em minha vida e por todo seus ensinamentos, incentivo, amor e orações.

A minha irmã **Erica**, meu sobrinho **Kauê** e minha cunhada **Elza** pelos momentos de alegria proporcionados, quando estava preocupada e em dias estressantes.

Ao meu esposo **Eilzo Marcos** pelo amor, orações, força, conselhos e conversas ao longo desta jornada.

Aos meus **avós** e **avôs** paternos e maternos em especial ao meu avô **Miguel Jorge** pelo amor e incentivo e por tanto fazer pela educação de seus filhos e hoje por seus netos para que tenham um nível escolar superior.

Aos meus **familiares**, em especial ao meu tio **Francisco Diniz**. Por acreditarem e torcerem por mim.

A **Luzia** e **Welma** que me acolheram, pelas conversas e ensinamentos, que me ajudaram muito a continuar nesta caminhada e pela amizade verdadeira.

Ao Prof. Dr. **Márcio José Coelho de Pontes**, por sua valiosa orientação, dedicação, confiança, motivação e compreensão com as dificuldades encontradas ao longo do trabalho. Muito obrigada.

Ao Prof. Dr. **Everaldo Paulo de Medeiros**, pela acolhida, confiança, motivação e oportunidade cedida para realizar este trabalho.

Aos professores, **Edvan Cirino**, **Ricardo Alexandre Cavalcanti** e **Francisco Fernandes** pelas contribuições na pré - banca e banca final deste trabalho.

Ao Professor **Mário César Ugulino Araújo**, pela acolhida no LAQA.

A todos da **Família LAQA**, em especial, Adenilton, Adriano, Amanda, Anna Luiza, Flaviano, Karla, Manu, Mayara, Neto, Renato e Sofácles pelo apoio e ensinamentos valiosos para a minha formação profissional e companheirismo durante toda a trajetória

À **Embrapa Algodão**, pelas amostras cedidas e por disponibilizar suas dependências e equipamentos para realização deste trabalho, em especial ao LATECTQ, ao Laboratório de Fibras e recepção dos profissionais.

A **Edjane** do LATECTQ pela acolhida e ajuda sempre que necessária.

A Universidade Federal da Paraíba.

A **CAPES** pelo apoio financeiro.

A todos que de forma direta ou indiretamente contribuíram na realização deste Trabalho.

Obrigada!

“...Mil cairão ao teu lado, e dez mil à tua direita, mas tu não serás atingido....” (Salmo 91: 7)

... “Pois ele ordenou aos seus anjos que guardem você em seus caminhos. Eles os levarão nas mãos para que seu pé não tropece numa pedra....” (Salmo 91: 11 e 12)

“... Quando me invocar, eu o atenderei; na tribulação estarei com ele. Hei de livrá-lo e o cobrirei de glória....” (Salmo 91: 14 e 15)

“Entrega o teu caminho ao Senhor; confia nele, e Ele tudo fará”. (Salmos 37: 4 e 5)

SUMÁRIO

LISTA DE FIGURAS.....	iv
LISTA DE TABELAS.....	viii
LISTA DE ABREVIATURAS.....	x
RESUMO.....	xi
ABSTRACT.....	xii
1 INTRODUÇÃO	2
1.1 Caracterização do Problema.....	2
1.2 OBJETIVOS.....	4
1.2.1 <i>Objetivo Geral</i>	4
1.2.2 <i>Objetivos Específicos</i>	4
2 FUNDAMENTAÇÃO TEÓRICA	6
2.1 A Cultura do Algodão	6
2.1.1 <i>Algodão Colorido</i>	7
2.1.1.1 BRS 200 Marrom.....	8
2.1.1.2 BRS Verde.....	8
2.1.1.3 BRS Rubi.....	9
2.1.1.4 BRS Safira.....	9
2.1.1.5 BRS Topázio	9
2.1.2 <i>A Classificação do Algodão e os Sistemas de Análise</i>	9
2.2 Fatores que Influenciam a Percepção das Cores.....	13
2.3 Imagens Digitais.....	16
2.4 Modelos de Cores	19
2.4.1 <i>Modelo de Cores RGB</i>	19
2.4.2 <i>Modelo de Cores HSV</i>	22
2.5 Quimiometria	23
2.5.1 <i>Técnicas de Reconhecimento de Padrões</i>	23
2.5.1.1 <i>Análise por Componentes Principais (PCA)</i>	24
2.5.1.2 <i>Análise Discriminante pelos Mínimos Quadrados Parciais (PLS-DA)</i> ..	25
2.5.1.3 <i>Análise Discriminante Linear (LDA)</i>	25
2.5.2 <i>Calibração Multivariada</i>	26

2.5.2.1 Regressão Linear Múltipla (MLR)	27
2.5.2.2 Regressão em Mínimos Quadrados Parciais (PLS)	28
2.5.3 Seleção de Variáveis	28
2.5.3.1 Algoritmo das Projeções Sucessivas (SPA)	29
2.5.3.2 Stepwise (SW)	30
2.5.4 Seleção de Amostras	31
2.6 Revisão da Literatura.....	32
3 METODOLOGIA	36
3.1 Amostras	36
3.1.1 Estudo para Classificação	36
3.1.2 Estudo para Calibração	37
3.2 Aquisição das Imagens Digitais	38
3.2.1 Sistema de Aquisição de Imagens Digitais	38
3.2.1.1 Compartimento	38
3.2.1.2 Iluminação do Sistema	39
3.2.1.3 Webcam e Notebook	39
3.2.2 Obtenção das Imagens Digitais	39
3.3 Análise de Referência.....	41
3.3.1 Obtenção do Grau de amarelamento (+b) e Reflectância (Rd)	41
3.3.2 Obtenção do Teor de Cera (WAX)	42
3.4 Procedimentos Quimiométricos e Software	44
3.4.1 Reconhecimento de Padrões/Classificação	44
3.4.2 Calibração Multivariada	45
4 RESULTADOS E DISCUSSÃO	48
4.1 Estudo de Classificação.....	48
4.1.1 Histogramas.....	48
4.1.2 Análise por Componentes Principais (PCA)	51
4.1.3 Modelos de Classificação	55
4.1.3.1 Modelo LDA.....	57
4.1.3.2 Modelo PLS-DA.....	58
4.2. Estudo de Calibração	61
4.2.1 Histogramas.....	61
4.2.2 Modelos de Calibração Multivariada	62

5 CONCLUSÕES	71
5.1 Perspectivas	73
REFERÊNCIAS.....	75
ANEXO	89

LISTAS DE FIGURAS

Figura 2.1.	Diagrama de cor (USDA) HVI para algodões herbáceo (<i>Gossypium hirsutum</i> L.r. <i>latifolium</i> Hutch).....	11
Figura 2.2.	Espectro Eletromagnético.....	14
Figura 2.3.	Reflexão da luz.....	15
Figura 2.4.	Transmissão da luz.....	15
Figura 2.5.	Imagem monocromática e sua representação em imagem digital.....	17
Figura 2.6.	Os três planos monocromáticos de uma imagem, separadamente, e a imagem colorida resultante da composição desses três planos.....	18
Figura 2.7.	Histogramas referentes à distribuição de frequência de todos os possíveis valores de um pixel nos níveis de cinza, vermelho, verde e azul de uma imagem.....	19
Figura 2.8.	Cores primárias aditivas e as cores resultantes de suas combinações.....	20
Figura 2.9.	Modelo de cores RGB.....	21
Figura 2.10.	Modelo de cores HSV.....	22
Figura 3.1.	Plumas de Algodão Naturalmente Colorido (a) BRS 200 Marrom e BRS Topázio (Grupo I) e (b) BRS Rubi e BRS Safira (Grupo II).....	36
Figura 3.2.	Plumas das cultivares de algodão naturalmente branco e colorido utilizadas no estudo de calibração.....	37
Figura 3.3.	Sistema usado na aquisição das imagens de amostras de plumas de algodão naturalmente colorido (a) compartimento, (b) lâmpada fluorescente, (c) suporte de <i>Teflon</i> ®, (d) webcam e (e) notebook.....	38
Figura 3.4.	Procedimento para obtenção dos dados analíticos.....	40
Figura 3.5.	Representação da matriz de dados.....	41
Figura 3.6.	Sistema HVI (High Volume Instruments) modelo HVI 1000.....	41
Figura 3.7.	Processo de extração de cera de plumas de algodão com álcool em extratores de Randall.....	42

Figura 3.8.	Separação líquido-líquido para determinação de cera.....	43
Figura 4.1.	Histograma completo para o grupo (a) I e (b) II. R: Red, G: Green, B: Blue, H: Hue, S: Saturation e V: Value.....	49
Figura 4.2.	Histograma Médio para o grupo I. R: Red, G: Green, B: Blue, H: Hue, S: Saturação e V: Value.....	50
Figura 4.3.	Histograma Médio para o grupo II. R: Red, G: Green, B: Blue, H: Hue, S: Saturation e V: Value.....	51
Figura 4.4.	Gráfico dos escores obtidos pela PCA aplicada aos histogramas do Grupo I (●: BRS 200 Marrom, ■: BRS Topázio). (a) todos os canais; b) tons de cinza; c) R, (d) G, (e) B, (f) RGB, (g) H, (h) S, (i) V, (j) HSV, (l) Cinza-RGB, (m) Cinza-HSV, (n) RGB-HSV. A variância explicada para cada PC é encontrada entre parêntese.....	52
Figura 4.5	Gráfico dos pesos das variáveis em cada componente de cor para o grupo I.R: Red, G: Green, B: Blue, H: Hue, S: Saturação e V: Value.....	53
Figura 4.6.	Gráfico dos escores obtidos pela PCA aplicada aos histogramas do Grupo II (▲: BRS Rubi, ◆: BRS Safira). (a) todos os canais; b) tons de cinza; c) R, (d) G, (e) B, (f) RGB, (g) H, (h) S, (i) V, (j) HSV, (l) Cinza-RGB, (m) Cinza-HSV, (n) RGB-HSV. A variância explicada para cada PC é encontrada entre parêntese.....	54
Figura 4.7:	Gráfico dos pesos das variáveis em cada componente de cor para o grupo II.R: Red, G: Green, B: Blue, H: Hue, S: Saturação e V: Value.....	55
Figura 4.8.	Figura 4.8: Média dos histogramas para cada cultivar, com as Variáveis selecionadas (●,●) pelo LDA/SPA (a e b) e pelo LDA/SW (c e d) para o Grupo I (a e c) e Grupo II (b e d).....	57
Figura 4.9.	Gráfico do escores obtido por LDA usando algoritmos de seleção de variáveis SPA (a e b) e SW (c e d), (a e c) Grupo I (b e d) Grupo II. ●: BRS 200 Marrom, ■: BRS Topázio, ▲: BRS Rubi e ◆: BRS Safira.....	58

Figura 4.10.	Taxa de erro de validação cruzada <i>versus</i> o número de variáveis latentes (a) Grupo I; (b) Grupo II. As setas identificam o número de variáveis latentes selecionadas.....	59
Figura 4.11.	Histogramas completos, das plumas de algodão. R: Red, G: Green, B: Blue, H: Hue, S: Saturation e V: Valor.....	61
Figura 4.12.	Gráfico da variância em y <i>versus</i> o número de fatores dos modelos PLS, para (a) +b, (b) Rd e (c) WAX.....	66
Figura 4.13.	Gráficos das variáveis selecionadas <i>versus</i> RMSECV para MLR-SPA (a) +b, (c) Rd e (e) WAX. Histograma médio com as variáveis selecionadas pelo SPA (b) +b, (d) Rd e (f) WAX. As setas identificam o número de variáveis selecionadas.....	67
Figura 4.14.	Gráfico dos valores preditos <i>versus</i> os valores de referências obtidos pelos modelos PLS (a,c,e) e MLR-SPA (b,d,f) para +b (a-b), Rd (c-d) e WAX (e-f). Amostras de calibração (○) e amostra de predição (●).....	69

LISTAS DE TABELAS

Tabela 2.1	Produção Mundial do Algodão e os principais países produtores..	6
Tabela 2.2	Cores básicas da região visível e suas faixas no espectro.....	14
Tabela 3.1	Número de amostras de treinamento, validação e teste para cultivares do grupo I.....	44
Tabela 3.2	Número de amostras de treinamento, validação e teste para cultivares do grupo II.....	45
Tabela 3.3	Número de amostras de calibração e previsão usados na determinação do grau de amarelamento (+b), reflectância (Rd) e teor de cera (WAX).....	46
Tabela 4.1	Taxa de classificação correta (TCC) obtidos com os modelos LDA/SPA, LDA/SW e PLS-DA para o subconjunto de teste dos Grupo I e II.....	56
Tabela 4.2	Resultados da classificação para o subconjunto de teste obtidos com modelos LDA/SPA, LDA/SW e PLS-DA, para o grupo I e II. A taxa de classificação correta obtida por cada modelo é indicada em parênteses. N número de amostras preditas empregadas neste estudo.....	60
Tabela 4.3	Resultados obtidos para os modelos PLS e MLR-SPA para o grau de amarelamento em cada canal (e combinação entre eles). Os valores entre parênteses correspondem ao número de fatores e variáveis empregadas nos modelos PLS e MLR-SPA, respectivamente. A faixa de calibração do parâmetro é apresentada entre colchetes, cv. (validação cruzada) e prev. (previsão).....	63
Tabela 4.4	Resultados obtidos para os modelos PLS e MLR-SPA para a reflectância em cada canal (e combinação entre eles). Os valores entre parênteses correspondem ao número de fatores e variáveis empregadas nos modelos PLS e MLR-SPA, respectivamente. A faixa de calibração do parâmetro é apresentada entre colchetes, cv. (validação cruzada) e prev. (previsão).....	64

Tabela 4.5	Resultados obtidos para os modelos PLS e MLR-SPA para o teor de cera em cada canal (e combinação entre eles). Os valores entre parênteses correspondem ao número de fatores e variáveis empregadas nos modelos PLS e MLR-SPA, respectivamente. A faixa de calibração do parâmetro é apresentada entre colchetes, cv. (validação cruzada) e prev. (previsão).....	65
Tabela 4.6	Resultados obtidos pelos modelos PLS e MLR para o conjunto de previsão (em termos de RMSEP, valores de correlação e de determinação). N é o número de previsão.....	68
Tabela 4.7	Resultados para o teste F (nível de confiança de 95%) obtidos na comparação dos modelos (PLS e MLR-SPA).....	68
Tabela 4.8	Resultados para o teste t -pareado (nível de confiança de 95%) obtidos na comparação dos valores de referências e valores previstos pelos modelos PLS e MLR-SPA.....	69

LISTAS DE ABREVIATURAS

ASTM	<i>American Society for Testing and Materials</i> (Sociedade Americana de Testes e Materiais)
+b	Grau de Amarelamento
CCD	<i>Charge Coupled Device</i> (Dispositivo de Carga Acoplada)
CIE	<i>Commission Internationale de l'Eclairage</i> (Comissão Internacional de Iluminação)
CMY	<i>Cyan, Magenta, Yellow</i> (Ciano, Magenta, Amarelo)
CMYK	<i>Cyan, Magenta, Yellow, Black</i> (Ciano, Magenta, Amarelo, Preto)
Conab	Companhia Nacional de Abastecimento
FD	Função Discriminante
FT-NIR	<i>Fourier Transform Near Infrared</i> (Transformada de Fourier do infravermelho próximo)
GA	<i>Genetic Algorithm</i> (Algoritmo Genético)
HCA	<i>Hierarchical Cluster Analysis</i> (Análise de Agrupamento Hierárquico)
HSB	<i>Hue, Saturation, Brightness</i> (Matiz, Saturação, Brilho)
HSI	<i>Hue, Saturation, Intensity</i> (Matiz, Saturação, Intensidade)
HSV	<i>Hue, Saturation, Value</i> (Matiz, Saturação, Valor)
HLS	<i>Hue, Luminosity, Saturation</i> (Matiz, Luminosidade, Saturação)
HVI	<i>High Volume Instrument</i> (Instrumento de Alto Volume)
iPLS	<i>Interval Partial Least Square</i> (mínimos quadrados parciais por intervalo)
JK	Jack Knife
KS	<i>Kennard-Stone</i>
LDA	<i>Linear Discriminant Analysis</i> (Análise Discriminante Linear)
MLR	<i>Multiple Linear Regression</i> (Regressão Linear Múltipla)
NIR	<i>Near Infrared spectroscopy</i> (Espectroscopia do Infravermelho Próximo)
PCs	<i>Principal Component</i> (Componentes Principais)
PCA	<i>Principal Component Analysis</i> (Análise por Componentes Principais)

PLS	<i>Partial Least Square</i> (Mínimos Quadrados Parciais)
PLS-DA	<i>Partial Least Squares - Discriminant Analysis</i> (Análise Discriminante pelos Mínimos Quadrados Parciais)
PCR	<i>Principal Component Regression</i> (Regressão Por Componentes Principais)
r	Coeficiente de correlação
R ²	Coeficiente de determinação
Rd	Reflectância
RGB	<i>Red, Green, Blue</i> (Vermelho, Verde, Azul)
RMSECV	<i>Root Mean Square Error of Cross Validation</i> (Raiz quadrada do erro médio quadrático de validação cruzada)
RMSEP	<i>Root Mean Square Error of Prediction</i> (Raiz quadrada do erro médio quadrático de predição)
RMSEV	<i>Root Mean Square Error of Validation</i> (Raiz quadrada do erro médio quadrático de validação)
SCV	Sinal Composto de Vídeo
SIMCA	<i>Soft Independent Modeling of Class Analogy</i> (Modelagem Independente e Flexível por Analogia de Classe)
SLR	Single-Lens Reflex (Lente com reflexo único)
SPA	<i>Successive Projections Algorithm</i> (Algoritmo das Projeções Sucessivas)
SPXY	<i>Sample set Partitioning based on joint X-y distances</i> (Partição de amostra baseado na distância de X-y)
SW	<i>Stepwise</i>
UV	Ultravioleta
USDA	<i>United States Department of Agriculture</i> (Departamento de Agricultura dos Estados Unidos)
VIS	Visível
WAX	Teor de cera

RESUMO

GONÇALVES, M^a. I. S.. Dissertação de Mestrado (**Aplicações de Imagens Digitais e Análise Multivariada para Classificação e Determinação de Parâmetros de Qualidades em Plumas de Algodão** Mestrado em Química) – Laboratório de Automação e Instrumentação em Química Analítica, Universidade Federal da Paraíba, João Pessoa, 2015.

Orientadores: Prof. Dr. Márcio José Coelho de Pontes e Dr. Everaldo Paulo de Medeiros

Nos últimos anos, plumas de algodão comerciais têm-se desenvolvido com melhor qualidade, apresentando características diferentes, mas com coloração similar. Isto pode ser um problema porque a identificação destas amostras é, em larga escala, realizada por meio de uma inspeção visual, que é um método subjetivo e sujeito a erros. Outra forma disponível para classificação dessas amostras consiste no uso do sistema HVI (High Volume Instruments) na determinação de parâmetros físicos de qualidade. Contudo, tal equipamento apresenta um alto custo, se comparado a técnica de imagens digitais, além do mais tem-se a necessidade de uma infraestrutura adequada e de um analista treinado para o procedimento de análise. Este trabalho propõe o desenvolvimento de uma nova metodologia analítica baseada na utilização de imagens digitais e análise multivariada para (1) classificação de plumas de algodão naturalmente colorido de acordo com o tipo de cultivar e (2) determinação simultânea de grau de amarelamento (+b), reflectância (Rd) e teor de cera (WAX). A aquisição das imagens digitais das plumas de algodão foi realizada por meio de uma *webcam* e foram obtidos os histogramas contendo as distribuições nos níveis de cores no padrão RGB (vermelho-verde-azul), escala de cinza e o sistema HSV (*matiz-saturação-valor*). Na classificação das amostras, modelos baseados na análise discriminante pelos mínimos quadrados parciais (PLS-DA) e análise discriminante linear (LDA) com seleção de variáveis pelo algoritmo das projeções sucessivas (SPA) ou pelo *stepwise* (SW) foram avaliados. Para a determinação dos parâmetros de +b, Rd e WAX, modelos PLS e regressão linear múltipla (MLR) com seleção de variáveis pelo SPA foram desenvolvidos e comparados. Os melhores resultados de classificação foram obtidos com o modelo LDA/SW, com uma taxa de classificação correta (TCC) de 96% para o conjunto de teste utilizando a combinação HSV. Quanto aos métodos de calibração, resultados de previsão satisfatórios foram obtidos para ambos os modelos (PLS e MLR-SPA), com valores de RMSEP próximos à repetitividade do método de referência. Além disso, nenhum erro sistemático foi observado e não foram encontradas diferenças significativas entre os valores previstos e de referência, de acordo com um teste *t*-pareado ao nível de confiança de 95%. Como vantagens o método é simples, de baixo custo, não utiliza reagente, não destrói a amostra e realiza análise em curtos intervalos de tempo.

Palavras-Chaves: Plumas de algodão; Imagens Digitais; Seleção de Variáveis; Classificação multivariada e Calibração multivariada

ABSTRACT

GONÇALVES, M^a. I. S. **Aplicações de Imagens Digitais e Análise Multivariada para Classificação e Determinação de Parâmetros de Qualidades em Plumas de Algodão**. Dissertação de Mestrado (Mestrado em Química) – Laboratório de Automação e Instrumentação em Química Analítica, Universidade Federal da Paraíba, João Pessoa, 2015.

Orientadores: Prof. Dr. Márcio José Coelho de Pontes e Dr. Everaldo Paulo de Medeiros

In recent years, commercial cotton lint have been developed with better quality, presenting different characteristics, but with similar coloring. This can be a problem because these samples is identified, large-scale, performed by a visual inspection, which is a very subjective method and error prone. Another way available for classification of samples is the use of HVI system (High Volume Instruments) to determine physical quality parameters. However, this apparatus has a high cost when compared to digital imaging technique, furthermore has the need for adequate infrastructure and a trained analyst for analysis procedure. This work proposes the development of a novel analytical method based on the use of digital image and multivariate analysis to (1) naturally colored cotton plumes classification according to the type of cultivar and (2) simultaneous determination of degree of yellowness (+b), reflectance (Rd) and wax content (WAX). The acquisition of digital images of cotton lints was carried out through a webcam and histograms containing distributions in levels of colors in standard RGB (red-green-blue), grayscale and HSV system (hue-saturation-value) they were obtained. In the classification of samples, models based discriminant analysis by partial least squares (PLS-DA) and linear discriminant analysis (LDA) with variable selection by the successive projections algorithm (SPA) or stepwise (SW) were evaluated. For the determination of the parameters +b, Rd and WAX, PLS models and multiple linear regression (MLR) with variable selection by the SPA were developed and compared. The best classification results were obtained with LDA / SW model with a correct classification rate (TCC) of 96% for the test group using the HSV combination. As the calibration methods, satisfactory prediction results were obtained for both models (PLS and MLR-SPA) with values of RMSEP near repeatability of the reference method. Furthermore, no systematic error was observed and there were no significant differences between the predicted values and reference, according to a paired *t*-test at 95% confidence. As advantages of the method is simple, low cost, does not use reagent, does not destroy the sample and realizes analysis at short time intervals.

Key Words: Cotton lint; Digital Images; Variable Selection; Multivariate Classification and Multivariate Calibration.



CAPÍTULO I

INTRODUÇÃO



1 INTRODUÇÃO

1.1 Caracterização do Problema

O algodão é a mais importante fibra têxtil natural em todo o mundo e a sexta maior fonte mundial de óleo vegetal [1], com grande importância na economia de muitos países, onde 150 produzem ou consomem cerca de 20 milhões de toneladas anuais de algodão em pluma [2-3]. Estima-se que desde o plantio até a comercialização dos produtos feitos com a fibra, a cadeia mundial movimenta cerca de US\$ 12 bilhões por ano e empregue em média 350 milhões de pessoas [4].

Na maioria das espécies primitivas, o algodão possui fibras coloridas que são controladas por um gene dominante [2,5]. Entretanto, algumas cores podem ser influenciadas pela luz solar, tipo de solo, entre outros fatores [5-6]. Pesquisas demonstraram que o algodão branco e colorido são igualmente antigos, porém, o algodão colorido apresenta características inferiores (resistência, estabilidade, uniformidade, entre outros) ao algodão branco [6-7]. Mais recentemente novas cultivares de algodão colorido têm sido desenvolvidas com qualidade superior a fibras de algodão branco [8].

No ano de 1988, Sally Fox [9] conseguiu desenvolver, nos Estados Unidos, fibras de algodão naturalmente coloridas que resistem ao processamento de fiação têxtil, moderno e de alta velocidade. Com isso, o interesse pela fibra colorida aumentou e países como Israel, Brasil, Peru, Grécia, Turquia, entre outros [10] trabalham com melhoramento genético com intuito de desenvolver novas plumas coloridas e com melhor qualidade.

O algodão colorido contribui para a sustentabilidade ecológica, dispensando o processo de tingimento têxtil com o uso de substâncias químicas [2,6-7,10]. Com isso, tem-se um maior valor de mercado em comparação ao algodão de fibra branca [11]. Desse modo, torna-se crescente seu cultivo no Brasil, em especial, na região nordeste pela agricultura familiar tanto em manejo convencional, orgânico e agroecológico [7].

Entre plumas coloridas, a cor marrom é, sem sombra de dúvidas, a mais comum. Podem ser encontrados em diferentes tons, que varia de marrom-claro até marrom-avermelhado intenso [12]. Nos últimos anos, com o surgimento de novas cultivares, para fibras especiais e de qualidade superior tem-se requisitado novas

maneiras de identifica-las de forma rápida e em larga escala, empregando tecnologias com maior poder de discriminação, que apenas a inspeção visual.

A determinação de parâmetros físicos e químicos no algodão também é de grande importância para avaliação do controle de qualidade da fibra, uma vez que implica em seu valor no mercado, sendo imprescindível tanto para o melhoramento genético como gerenciamento de fiações [13-14].

A classificação do algodão em relação aos parâmetros físicos começou a ser determinada manualmente e por percepção visual [13]. Todavia, são constituídos de processos subjetivos e sujeitos a erros. Esse método foi aprimorado com o instrumento de alto volume (HVI: *High Volume Instruments*), que surgiu entre as décadas de 1960 e 1970 e foi introduzido pelo Departamento de Agricultura dos Estados Unidos (USDA: *United States Department of Agriculture*) [13]. Dentre as medições realizadas por este equipamento, destacam-se: índice micronaire, comprimento, uniformidade de comprimento, índice de fibras curtas, índice de fiabilidade, resistência, alongamento a ruptura, impurezas, reflectância e grau de amarelamento [13-15]. O HVI é um instrumento rápido e preciso para medir características físicas de qualidades da fibra. Porém, trata-se de um equipamento de alto custo, além de requerer infraestrutura adequada e pessoal treinado para o procedimento da análise, não permitindo a mobilidade para medidas *in situ*. Ambas as determinações (classificação manual/visual ou classificação por HVI) utilizam padrões físicos de referência determinados pelo USDA e outras organizações [14]. Quanto aos parâmetros químicos, por exemplo, o teor de cera é determinado por metodologia clássica, que envolve um procedimento gravimétrico. Tal método é laborioso, usa reagentes químicos tóxicos e demanda de um longo tempo de análise.

Em face do exposto, torna-se evidente a necessidade de desenvolvimento de métodos analíticos alternativos que sejam simples, econômicos e confiáveis para auxiliar no controle de qualidade de amostras de plumas de algodão (brancas e naturalmente coloridas).

O uso de dados de imagens digitais acoplados à análise multivariada estabelece uma alternativa atrativa para o desenvolvimento de métodos analíticos simples, rápidos, precisos, de baixo custo, capazes de realizar análises não destrutivas em amostras de algodão.

1.2 OBJETIVOS

1.2.1 Objetivo Geral

Desenvolver uma nova metodologia analítica baseada no uso de imagens digitais e análise multivariada para a classificação de plumas de algodão com relação à cultivar e para determinação de três importantes parâmetros de qualidade: grau de amarelamento (+b), reflectância (Rd) e teor de cera (WAX).

1.2.2 Objetivos Específicos

- ✓ Aperfeiçoar um sistema para aquisição de imagens digitais de plumas de algodão, em termos de iluminação e configuração da câmera;
- ✓ Desenvolver e comparar modelos baseados na LDA com seleção de variáveis (SPA e SW) e PLS-DA para a classificação de plumas de algodão naturalmente coloridas com relação à cultivar (BRS 200 Marrom, BRS Topázio, BRS Rubi e BRS Safira);
- ✓ Desenvolver e comparar modelos de calibração multivariada (PLS e MLR-SPA) para a determinação simultânea do grau de amarelamento (+b), reflectância (Rd) e teor de cera (WAX) em plumas de algodão branca e naturalmente coloridas.



CAPÍTULO II

FUNDAMENTAÇÃO

TEÓRICA



2 FUNDAMENTAÇÃO TEÓRICA

2.1 A Cultura do Algodão

O algodão é uma das dez principais culturas cultivadas em mais de 60 países, distribuído nos cinco continentes [3-4]. Seus primeiros registros são datados a 7 mil anos atrás, onde atualmente encontra-se Paquistão e Índia [4].

A **Tabela 2.1** apresenta a produção mundial do algodão em pluma na safra de 2009/10 até a perspectiva para safra de 2014/15, onde pode ser visto uma queda na produção em 2012/13 e 2013/14 decorrente, da quantidade excedente de estoque produzido suficiente para um consumo mundial para um período de dez meses [16]. Com isso, fica clara a queda de produção, porém os produtores buscam alternativas para se ajustar às expectativas do cenário futuro.

Entre os principais países produtores da fibra, encontram-se China, Índia, Estados Unidos e Paquistão. O Brasil, em especial, ocupa a quinta posição na produção mundial da pluma de algodão [17] e também é um dos principais exportadores.

Tabela 2.1: Produção Mundial do Algodão e os principais países produtores.

Milhões de toneladas						
	SAFRA					
	2009/10	2010/11	2011/12	2012/13	2013/14	2014/15
Total Mundial	22,334	25,425	27,820	26,660	26,238	26,272
PRINCIPAIS PAÍSES						
China	6,925	6,400	7,400	7,300	6,929	6,463
Índia	5,185	5,865	6,239	6,205	6,770	6,770
USA	2,654	3,942	3,391	3,770	2,811	3,570
Paquistão	2,158	1,948	2,311	2,002	2,076	2,103
Brasil	1,194	1,960	1,877	1,310	1,705	1,507
Uzbequistão	0,850	0,910	0,880	1,000	0,940	0,940
Outros	3,369	4,401	5,722	5,073	5,008	4,918

FONTE: ABRAPA. Associação Brasileira dos produtores de Algodão. Estatística da Produção Mundial do algodão, 2014. Disponível em:

<<http://www.abrapa.com.br/estatisticas/Paginas/producaomundial.aspx>>.

Acesso em: 29 janeiro de 2015.



No Brasil, a região centro-oeste se destaca por apresentar a maior produção de algodão, tendo o Estado do Mato Grosso como maior produtor [17-18].

Na produção do algodoeiro, são explorados comercialmente (1) a fibra como principal produto com várias aplicações industriais, como confecções de fios para vários tipos de tecidos, algodão hidrófilo utilizado em área-médica hospitalar, obtenção de celulose entre outros; (2) o óleo da semente (18%-25% m/m), muito utilizado na alimentação humana, na fabricação de margarina, sabões e para produção de energia (biodiesel); (3) a torta obtida na extração do óleo como coproduto, rica em proteína (40% a 45%), é frequentemente empregada na alimentação animal [3,19-20].

A pluma de algodão comercial é quase totalmente constituída por algodão de fibra branca. Consequentemente, corantes são necessários durante o processamento têxtil para colorir o tecido [21]. O alto consumo de corantes naturais ou sintéticos resulta em uma atenção ambiental especial [21]. Uma alternativa seria a utilização da fibra de algodão colorido que poderá ser obtida em tons verde e marrom, sendo ambientalmente inócuos [21-22].

2.1.1 Algodão Colorido

Os primeiros registros do algodão colorido são por volta de 2700 a.C. no Paquistão, Egito e Peru [22-23]. Durante anos foram descartados pela indústria têxtil de vários países, onde o produtor considerava-os contaminantes indesejáveis dos algodões de fibra branca [11]. Além do mais possuíam fibras curtas, grossas e fracas, favoráveis apenas para fiação manual [6], ou seja, sem nenhuma importância econômica.

No Brasil, o algodão colorido, em especial, à fibra marrom foi utilizada como planta ornamental e em confecções de artesanatos, até sua exploração em plantios comerciais [24].

Com desenvolvimento de programas de melhoramento genético, tendo a finalidade de aperfeiçoar as características da fibra colorida, hoje é crescente o interesse pelo algodão colorido em muitos países. De fato, o cultivo de algodão colorido traz uma série de benefícios na diminuição dos custos de obtenção do tecido durante o processo de industrialização, como: redução do uso de produtos químicos

e da quantidade de água utilizada no tingimento têxtil, uma substancial diminuição da energia e da quantidade de efluentes a serem tratados ^[11].

Apesar das grandes vantagens em se utilizar o algodão colorido, seu cultivo é, ainda, muito limitado. De uma forma geral, isso pode ser atribuído, entre outros fatores, à falta de regras e critérios para classificação de sua qualidade, tornando seu processamento e comércio global mais difícil ^[10]. Assim, o algodão colorido é, em sua grande maioria, confeccionado por pequenas empresas, produzindo roupas, acessórios, artigos de decoração e outros.

A Embrapa (Empresa Brasileira de Pesquisa Agropecuária) já lançou comercialmente entre 2000 e 2010 cinco cultivares de fibras coloridas: BRS 200, BRS Verde, BRS Rubi, BRS Safira, BRS Topázio e continua trabalhando com melhoramento genético, com a finalidade de produzir mais cultivares com boas características como: comprimento, resistência, uniformidade e estabilidade ^[7].

Devido aos benefícios gerados pelo algodão colorido, os produtos dessas plumas podem ser alvos de falsificação.

2.1.1.1 BRS 200 Marrom

Lançada no segundo semestre no ano de 2000, a BRS 200 Marrom foi a primeira cultivar colorida oriunda de algodoeiros arbóreos coletados na região Nordeste ^[25]. Como características marcantes, podem ser citadas: cor marrom-clara, ciclo semi-perene, ou seja, produz economicamente até 3 anos, elevada resistência a seca, podendo ser plantada nas regiões do Seridó e Sertão. Seu valor de mercado é em torno de 30-50% superior às fibras de algodão branco ^[23,25]. Sua produtividade média é de 1.300 kg/ha e, em condições de plantio irrigado, pode chegar até 3.000kg/ha ^[23].

2.1.1.2 BRS Verde

Cultivar lançada no ano de 2003 ^[7], de algodoeiro herbáceo ou anual, a BRS verde é derivada do cruzamento entre um material introduzido dos Estados Unidos, o Arkansas Green (de fibra verde), com a cultivar CNPA 7H (de fibra branca), de ampla adaptação à região Nordeste ^[23,26]. É uma cultivar que tem uma instabilidade em relação à sua coloração, pois os pigmentos que causam a cor verde são sensíveis à

luz solar. Assim, recomenda-se que a fibra seja colhida em duas etapas, evitando que a mesma fique muito exposta ao sol no campo [7, 23].

2.1.1.3 BRS Rubi

Cultivar lançada em 2005, a BRS Rubi é o resultado do cruzamento de um material introduzido dos Estados Unidos de fibra marrom-escuro com a CNPA 7H [7,27]. É de algodoeiro herbáceo e constantemente explorada na região Nordeste em regime de sequeiro e irrigado. [23,27]. Sua fibra possui uma cor marrom-escuro ou marrom-avermelhada, sendo a primeira cultivar do Brasil com esta cor da pluma, e seu ciclo de colheita é de 120-140 dias [7,23].

2.1.1.4 BRS Safira

Cultivar também lançada em 2005, a BRS Safira originou-se do cruzamento entre um material introduzido dos Estados Unidos (de fibra marrom-escuro) com a CNPA 87-33 (de fibra branca), que assim como CNPA 7H, possui boa qualidade e ampla adaptação à região Nordeste [7,28]. Como a BRS Rubi, a cor de sua fibra é marrom-escuro ou marrom-avermelhada. A BRS Safira é uma cultivar de algodoeiro herbáceo ou anual e ciclo de colheita 120-140 dias [7,23].

2.1.1.5 BRS Topázio

Lançada em 2010, é uma cultivar de algodoeiro herbáceo, de fibra marrom-clara, derivada do cruzamento entre as cultivares Suregrow 31 e Delta Opal [23,8]. Reúne as melhores características da fibra entre as demais cultivares coloridas e supera as características de algumas fibras brancas, com alto rendimento de fibra, uniformidade e resistência [7,23,8].

2.1.2 A Classificação do Algodão e os Sistemas de Análise

A classificação do algodão é feita com base em suas características extrínsecas e intrínsecas [29], sendo o meio de controle de qualidade da fibra que determina seu valor no mercado.



O controle da qualidade do algodão começou a ser realizado depois de 1907, por várias dificuldades encontradas em sua comercialização, incluindo as grandes diferenças de preços. Estabeleceram-se, então, normas (padrões de referências físicos), desenvolvidas pelo Departamento de Agricultura dos Estados Unidos (USDA: *United States Department of Agriculture*) [30-31] que são anualmente atualizadas. A partir dessas normas, o algodão pode ser classificado, beneficiando os agricultores, comerciantes, fiandeiros, e institutos de pesquisa, garantindo um algodão de boa qualidade.

A classificação do algodão por muito tempo foi realizada por especialistas treinados e certificados, de forma manual/visual, porém são poucas as características de qualidade do algodão determinadas, apenas o comprimento da fibra, grau de folha (conteúdo de trash) e a cor [31].

Com o crescimento do comércio do algodão no mundo, houve a necessidade de desenvolver técnicas mais rápidas e precisas para determinar as características de qualidade do algodão, bem como outras características, como índice de fibras curtas, alongamento a ruptura, resistência, maturidade, dentre outros, não determinadas pelos classificadores. Com isso em 1960 o USDA junto à indústria americana de eletrônicos desenvolveu o primeiro “*High Volume Instrument*” (HVI) pela Motion Control (Dallas, Texas) e outros fabricantes. Porém, desde 1990 é produzido pela *Uster Technologies*. O sistema HVI é capaz de realizar medidas em larga escala e em um espaço mínimo de tempo [31]. Assim, a classificação do algodão passou a ser instrumental sem destruir a amostra. Mas, vale ressaltar que alguns países, incluindo o Brasil, utilizam ainda tanto o método instrumental, como o manual/visual.

Entre os parâmetros físicos determinados pelo sistema HVI, destacam-se o grau de amarelamento (+b) e a reflectância (Rd), que são características do algodão de grande relevância e estão diretamente associadas à sua cor.

O grau de amarelamento é obtido com ajuda de um filtro amarelo e verifica o quanto amarela é a amostra. Já a reflectância verifica o quanto cinza ou clara é a amostra, ou seja, a quantidade de luz que é refletida na amostra [14-15,29].

A classificação de plumas de algodão com respeito à cor apresenta-se como uma análise de grande importância e de interesse comercial, uma vez que tal parâmetro pode contribuir na indicação da qualidade dos produtos têxteis [32]. Para essa classificação, o sistema HVI fornece os valores de +b e Rd que se interseccionam



em um diagrama de Nickerson-Hunter (**Figura 2.1**), gerando um código referente ao índice de cor do algodão [14,30-31].

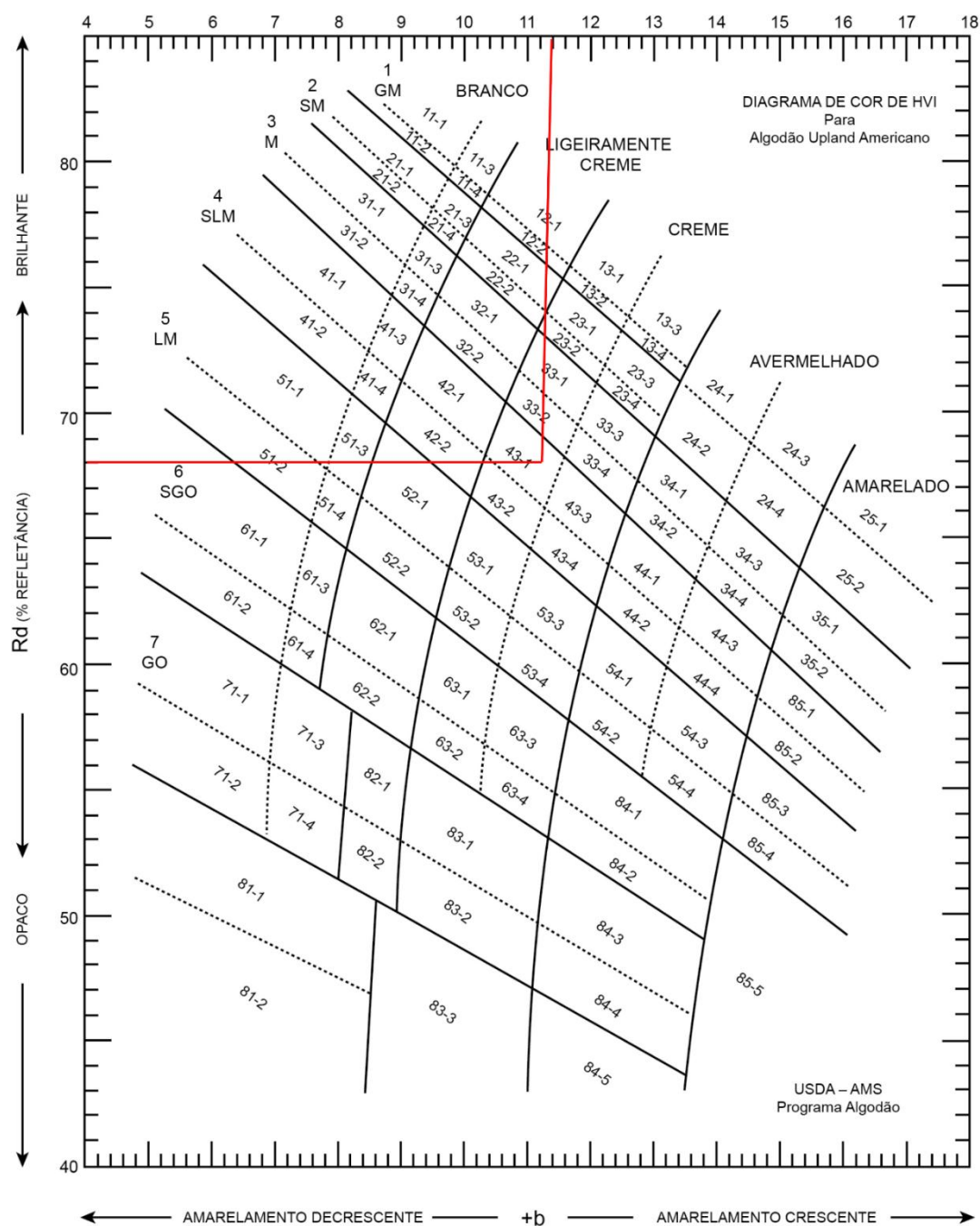


Figura 2.1: Diagrama de cor (USDA) HVI para algodões herbáceo (*Gossypium hirsutum* L.r. *latifolium* Hutch). Adaptado de: <http://www.abrapa.com.br/noticias/Documentos/00ProjetoDeRTdoAlgodao-ARQUIVOPARACOSULTAPUBLICA.pdf> Acessado em: 15 de fevereiro de 2015.



Devido ao alto custo e alguns inconvenientes das análises realizadas com o sistema HVI, vários métodos têm sido propostos [32-36] na literatura para a determinação do +b e Rd com confiança, rapidez e economia.

Duckett et.al. [33], Matusiak e Walawska [34], Rodgers et.al. [35] utilizaram métodos espectrofotométricos e os espaços de cor da Comissão Internacional de Iluminação (CIE: *Commission Internationale de l'Eclairage*) para determinar os parâmetros de cor do algodão. Os resultados obtidos foram comparados com aqueles determinados pelo sistema HVI. Além disso, Duckett et. al. [33] ainda utilizaram imagens digitais obtidas por um *scanner* para avaliar a uniformidade da cor, teor de resíduo e manchas amarelas nas amostras de algodão. Matusiak e Walawska [34] apresentaram, em outros estudos, uma avaliação comparativa entre a classificação da cor do algodão determinadas por classificadores com o sistema HVI.

Kang e Kim [36] utilizaram imagens digitais para medir os parâmetros da cor do algodão e os resíduos presentes na fibra e aplicaram um sistema de redes neurais artificiais para classificar a cor do algodão.

Recentemente, Cui et. al. [32] utilizaram imagens obtidas por um *scanner* para avaliar a distribuição e variação da cor dentro da amostra do algodão. Os resultados obtidos pelo sistema HVI foram empregados como estratégia de comparação.

Os parâmetros químicos também são de grande importância para análise da qualidade da fibra. Por exemplo, o teor de cera é essencial para a eficiência do processo de fiação, uma vez que fornece uma camada lubrificante capaz de reduzir o atrito entre a fibra e o metal do maquinário têxtil. Contudo, um teor de cera elevado pode ser uma desvantagem, visto que esta camada pode atuar, também, como uma barreira impermeável à entrada de moléculas de água e corantes na fibra. Tal barreira deverá ser removida por lavagem e/ou branqueamento [13] para que o processo de tingimento seja bem sucedido. Em termos de qualidade da fibra sua presença não é desejável, uma vez que correlaciona-se negativamente com propriedades importantes, como o teor de celulose e rendimento da fibra [37].

O método mais comum e de referência para a determinação do teor de cera no algodão foi desenvolvido por Conrad [38] em 1944, no qual envolve o uso de solventes orgânicos em um procedimento gravimétrico. Tal procedimento é demorado, laborioso, destrutivo e envolve o uso de reagentes químicos tóxicos. Recentemente, Silva [39] propôs um novo método analítico baseado no uso da espectroscopia visível e na região do infravermelho próximo e calibração multivariada para a determinação



do teor de cera em plumas de algodão. O método proposto é não destrutivo e rápido, mas o equipamento empregado na análise ainda apresenta um custo elevado, quando comparado à técnicas empregando imagens digitais obtidas por *webcam*.

Para classificar plumas de algodão de acordo com sua classe poucos métodos têm sido propostos [40-42].

Liu et.al.[40] utilizaram espectroscopia com transformada de Fourier do infravermelho próximo (FT-NIR) e PCA para comparar as diferentes características dos algodões naturalmente brancos e coloridos.

Liu et.al.[41] utilizaram a espectroscopia UV/visível/NIR associada ao método PLS, para prever os atributos de cor e outras propriedades de qualidades físicas da fibra de algodão. Além desse estudo, os autores também avaliaram métodos de reconhecimento de padrão com intuito de classificar três classes de amostras de algodão com respeito ao valor de micronaire.

Sousa [42] desenvolveu um método com o uso da espectroscopia na região do visível e infravermelho próximo aplicando a técnica de Modelagem independente e flexível por analogia de classe (SIMCA); para classificação de plumas e tecidos de algodão colorido, resultando em uma taxa de classificação correta de 100%.

O desenvolvimento de métodos que empregam a análise por imagens digitais vem crescendo substancialmente nos últimos anos [43], incluindo em trabalhos voltados à área de Química Analítica [43]. De fato, a disponibilidade e o baixo valor comercial das câmeras digitais, bem como os aparelhos de digitalização de uma forma geral, têm contribuído para o aumento do número de pesquisas nesta área.

Nas próximas seções, será descrita uma abordagem teórica referente às imagens digitais com diferentes modelos de cores. Contudo, torna-se importante apresentar os fatores que influenciam a percepção das cores.

2.2 Fatores que Influenciam a Percepção das Cores

A percepção das cores é resultante da interação de três elementos fundamentais: a fonte de luz, o objeto e o observador (sistema visual humano) [44-45].

A luz consiste de uma radiação eletromagnética e apresenta um comportamento ondulatório caracterizado por sua frequência (f) e comprimento de onda (λ) [45-47].



Na **Figura 2.2**, apresenta-se uma síntese do espectro eletromagnético com os tipos de radiações em ordem crescente de comprimento de onda, destacando-se a faixa de luz visível.

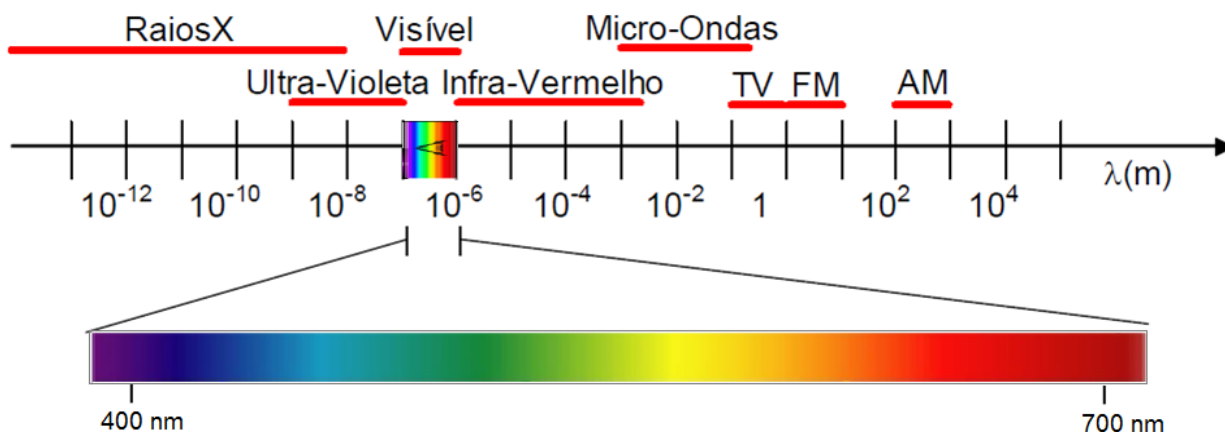


Figura 2.2: Espectro Eletromagnético

A luz visível é a faixa do espectro eletromagnético à qual o sistema visual humano é sensível e está compreendida na faixa de 380 a 740 nm. Nesta região, cada comprimento de onda corresponde quantitativamente a uma cor [47]. Quando as fontes de radiação emitem um único comprimento de onda, tem-se uma radiação monocromática, logo uma cor espectral pura [46]. Porém, são mais comuns as fontes de radiação emitirem uma mistura de comprimentos de ondas diferentes [47].

Na **Tabela 2.2** são apresentadas as sete cores básicas com os comprimentos de ondas característicos que varrem toda a região visível do espectro eletromagnético.

Tabela 2.2: Cores básicas da região visível e suas faixas no espectro

Cor	Faixa Correspondente no espectro (nm)
Violeta	380-440
Azul	440-485
Ciano	485-500
Verde	500-565
Amarelo	565-590
Laranja	590-625
Vermelho	625-740





As cores que os seres humanos percebem em um objeto são determinadas pela natureza da luz refletida a partir do objeto [48]. Quando uma fonte de luz incide sobre um objeto, três diferentes fenômenos podem ocorrer: reflexão, absorção e transmissão, que de acordo com o material podem absorver ou refletir determinados comprimentos de onda [45].

Na reflexão, quando a luz incide sobre um objeto, atinge os átomos e moléculas internos desse objeto e retorna ao seu meio de origem. Durante a interação da luz com os átomos e moléculas, o objeto absorve alguns comprimentos de ondas e reflete outros, processo denominado de reflectância espectral. Nesse processo a luz incidente é diferente da luz refletida [44,49] (**Figura 2.3**).

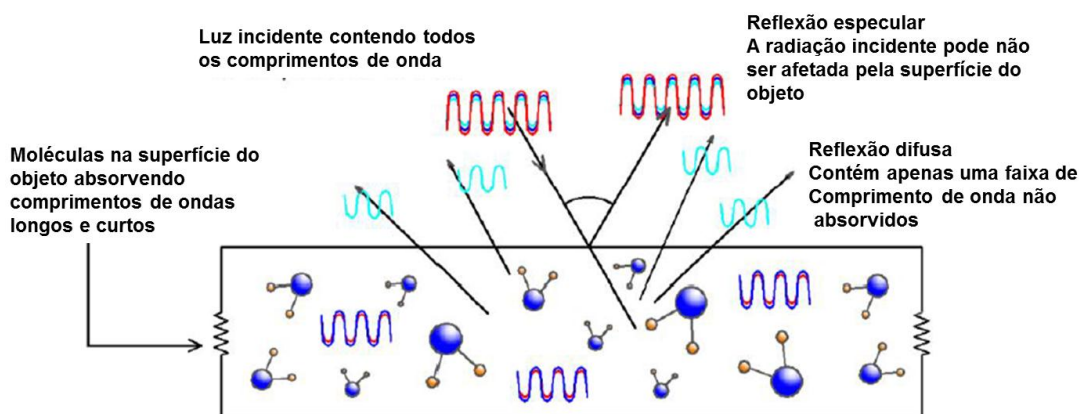


Figura 2.3: Reflexão da luz

Fonte: Vilar (2014) [50]

Acesso: 24 de fevereiro de 2015

Na transmissão, um objeto transmissivo afeta os comprimentos de onda da mesma forma como o objeto reflexivo [49], desde que os objetos transmissivos não possuam um elevado grau de transparência, pois toda luz incidente passará por ele [44] (**Figura 2.4**).

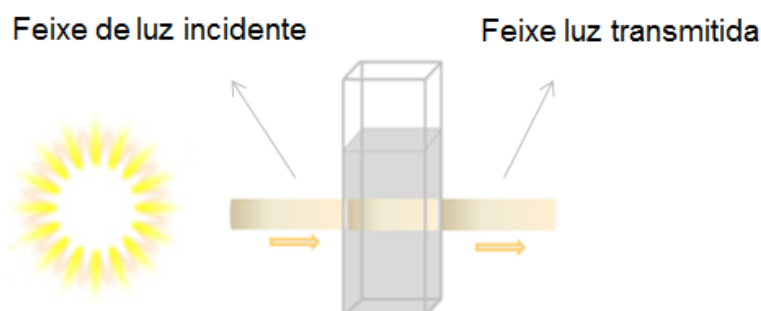


Figura 2.4: Transmissão da luz

Fonte: Milanez (2013) [49] adaptada

Acesso: 24 de fevereiro de 2015



Em se tratando do sistema visual humano (observador), como já mencionado, o observador percebe a cor por meio de uma energia luminosa. O olho humano possui uma rede de fotorreceptores sensível à luz (cones e bastonetes) distribuída na retina, a luz recebida é convertida em sinais elétricos que são transmitidos e interpretados pelo cérebro, tendo como resultado a sensação da visão com características como a cor. Três cones receptores do olho humano possui sensibilidade aos comprimentos de onda vermelho, verde e azul e a partir da combinação dos mesmos consegue-se visualizar todas as cores [45,51-52].

2.3 Imagens Digitais

A imagem digital, dita como a representação de uma cena, em forma bidimensional, requer duas etapas indispensáveis: aquisição e digitalização [53]. A aquisição da imagem ocorre por meio de dispositivos eletrônicos, como câmeras fotográficas, *webcam* dentre outros, que captam uma cena real tridimensional e a convertem em uma imagem analógica (imagem eletrônica). Este processo é denominado transdução optoeletrônica e envolve a redução da dimensionalidade da cena por meio de um sensor, onde o mais utilizado em dispositivos eletrônicos para aquisição de imagens é o CCD (*Charge Coupled Device* - Dispositivo de Carga Acoplada), que tem como função converter a energia luminosa incidente sobre o objeto em sinal elétrico. Tal sinal sai na forma de sinal composto de vídeo (SCV) em imagem analógica (monocromática ou colorida) [53].

Contudo para que a imagem analógica possa ser processada no computador, a mesma deve ser digitalizada, ou seja, convertida em uma matriz do tipo $M \times N$ pontos (dados na horizontal e vertical) (**Figura 2.5**), onde cada ponto é denominado *pixel* ou *elemento de imagem* (etapa de digitalização) [46,53]. Cada localização do *pixel* na imagem contém apenas um único valor numérico que representa o nível de sinal naquele ponto, podendo ter 256 níveis de cores que varia de (0-255) [46,54].

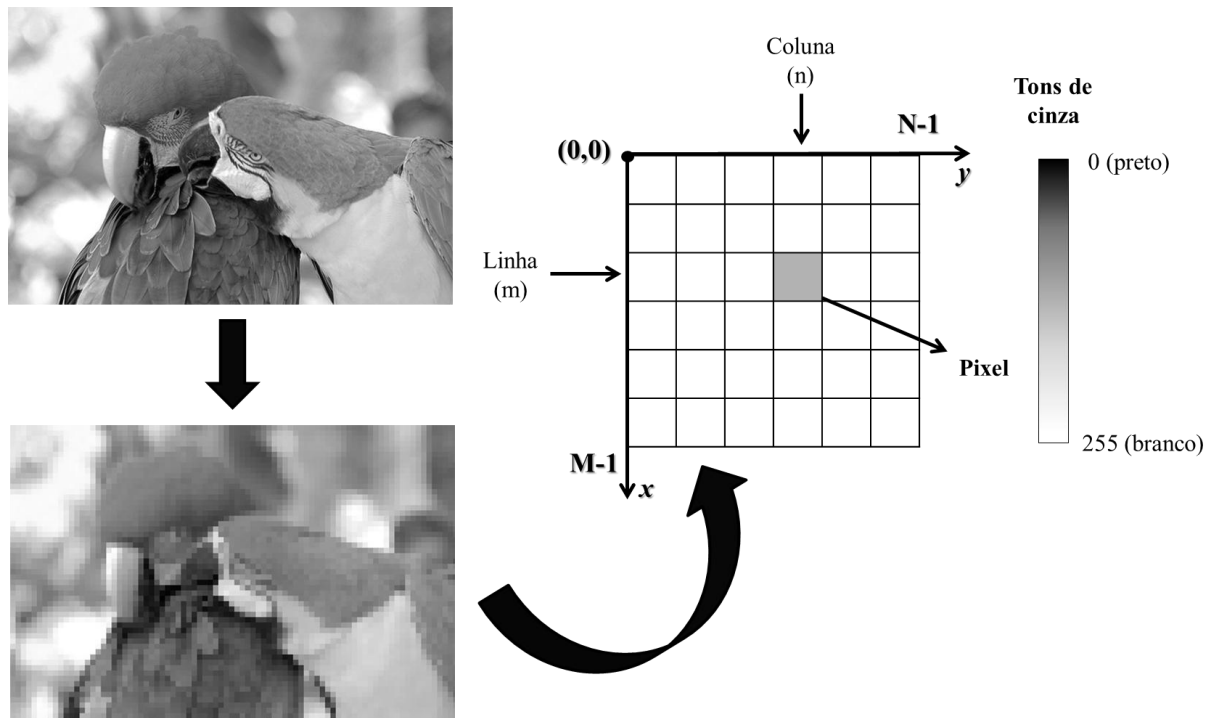


Figura 2.5: Imagem monocromática e sua representação em imagem digital.

Fonte: Queiroz e Gomes (2001) [46] adaptada.

Em uma imagem digital monocromática, cada pixel corresponde ao nível de cinza variando do preto (0) ao branco (255). Em um sistema RGB (imagem digital colorida), um *pixel* pode ser considerado um vetor, composto por três imagens monocromáticas, em que os componentes representam as intensidades de vermelho (R), verde (G) e azul (B) [46,52]. Matematicamente esse conceito pode ser representado pela **equação 2.1**:

$$f(x, y) = f_R(x, y) + f_G(x, y) + f_B(x, y) \quad (2.1)$$

Na **Figura 2.6**, são representados os planos monocromáticos nos canais vermelho (R), verde (G) e azul (B) de uma imagem e a combinação ou composição dos três planos (RGB).



Figura 2.6: Os três planos monocromáticos de uma imagem, separadamente, e a imagem colorida resultante da composição desses três planos.

O número de pixels presente em uma imagem pode ser representado por um gráfico denominado *histograma*. Nele, tem-se a distribuição de frequência dos valores que um pixel pode assumir, ou seja, quantas vezes um valor de cor variando (0-255) pode aparecer na imagem. Por exemplo, em uma imagem monocromática com tons de cinza, o histograma fornece o percentual de pixels correspondentes na imagem para cada nível de cinza [53].

A **Figura 2.7** apresenta uma imagem digital colorida, e os histogramas resultantes de sua decomposição nos canais de cores tons de cinza, vermelho (R), verde (G) e azul (B).

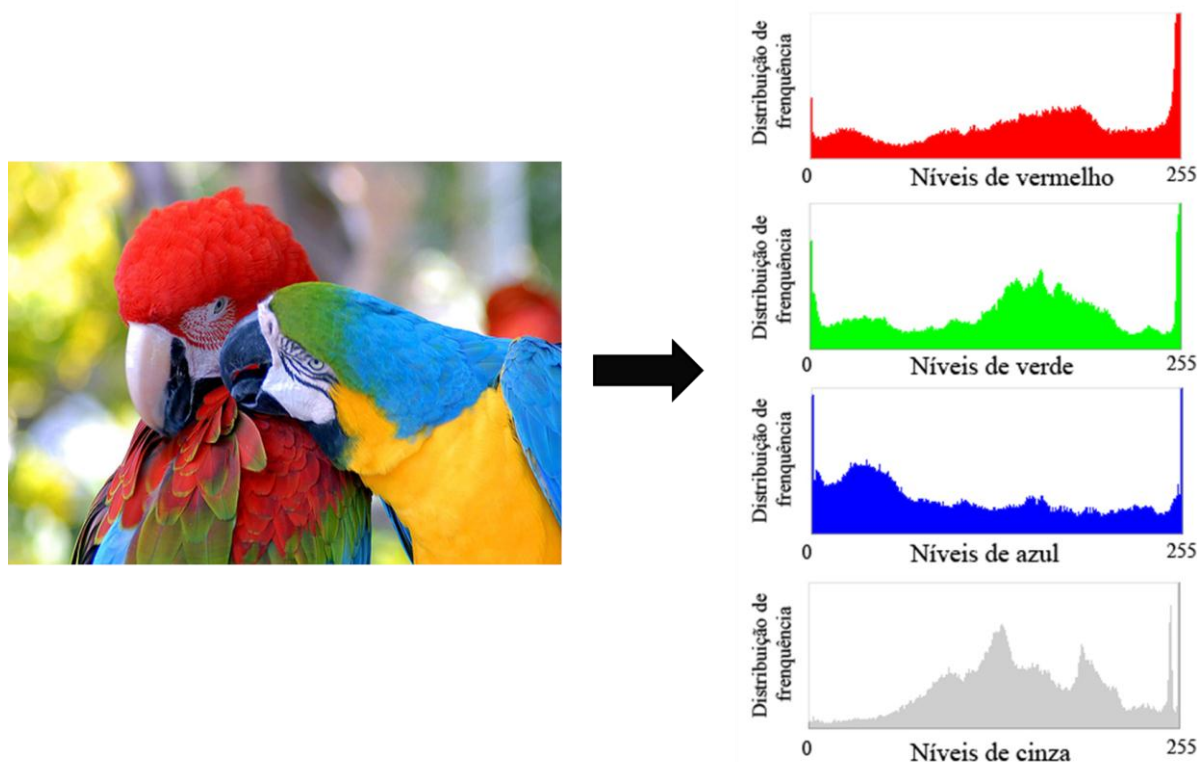


Figura 2.7: Histogramas referentes à distribuição de frequência de todos os possíveis valores de um pixel nos níveis de cinza, vermelho, verde e azul de uma imagem.

2.4 Modelos de Cores

Segundo Gonzalez e Woods [48], um modelo de cor, denominado também de espaço de cor ou sistema de cor, é uma especificação de um sistema de coordenadas tridimensionais, em que cada ponto representa uma cor, ou seja, de forma padronizada permite a especificação das cores.

Os modelos de cores mais utilizados são *RGB* (Red, Green, Blue), *CMY* (cyan, magenta, yellow), *CMYK* (cyan, magenta, yellow, black), *HSI* (hue, saturation, intensity), e *HSV* (hue, saturation, value) [48, 55]. Nas seções seguintes, os modelos RGB e HSV, empregados neste trabalho serão detalhados.

2.4.1 Modelo de Cores RGB

O RGB é o espaço de cor mais comum utilizado para a representação da imagem digital, uma vez que corresponde às três cores primárias (vermelho, verde e azul) [55]. Baseia-se no mecanismo de formação de cores do olho humano, na qual

relaciona à combinação de radiações monocromáticas dessas três cores primárias, tornando-se, então a formação de cores como um processo aditivo. [52] (**Figura 2.8**).

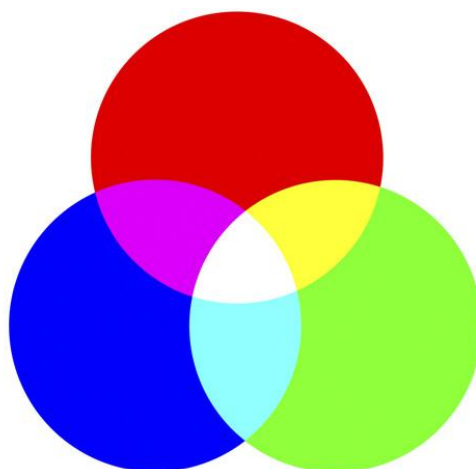


Figura 2.8: Cores primárias aditivas e as cores resultantes de suas combinações

Fonte: Batista (2005) [52]

Acesso: 04 de março de 2015

Tal mecanismo de formação de cores é formado por absorção e reflexão da luz, já que a percepção das cores ao olho humano possui uma relação direta com a luz que um objeto é capaz de refletir [52].

Os objetos podem refletir ou absorver total ou parcialmente as radiações com comprimentos de onda visíveis. Na reflexão total, a cor branca é percebida pelo observador e na absorção total, tem-se ausência de cor e o preto é percebido. Quando os objetos são parcialmente refletidos e/ou absorvidos, trata-se de um objeto colorido, que favorece a reflectância em variações limitadas do espectro visível, exibindo apenas alguns tons de cores [48]. Por exemplo, um objeto de cor ciano, reflete as radiações azul e verde e absorve a vermelha denominada de cor oposta ao ciano [53].

O modelo RGB é representado por um cubo definido sobre os eixos R, G e B (**Figura 2.9**), que assumem valores de (0-255), ou seja, 256 níveis de cores. Assim, cada canal de cor é formado por um conjunto de 8 bits resultando em uma imagem de 24 bits ou 16,7 milhões de cores, (esses valores são obtidos da seguinte forma: $2^8 = 256$ níveis de cores para cada canal, logo tem-se $3 \times 8 = 24$ bits para imagem RGB, assim $2^{24} = 16.777.216$ cores diferentes) [48,52,54].

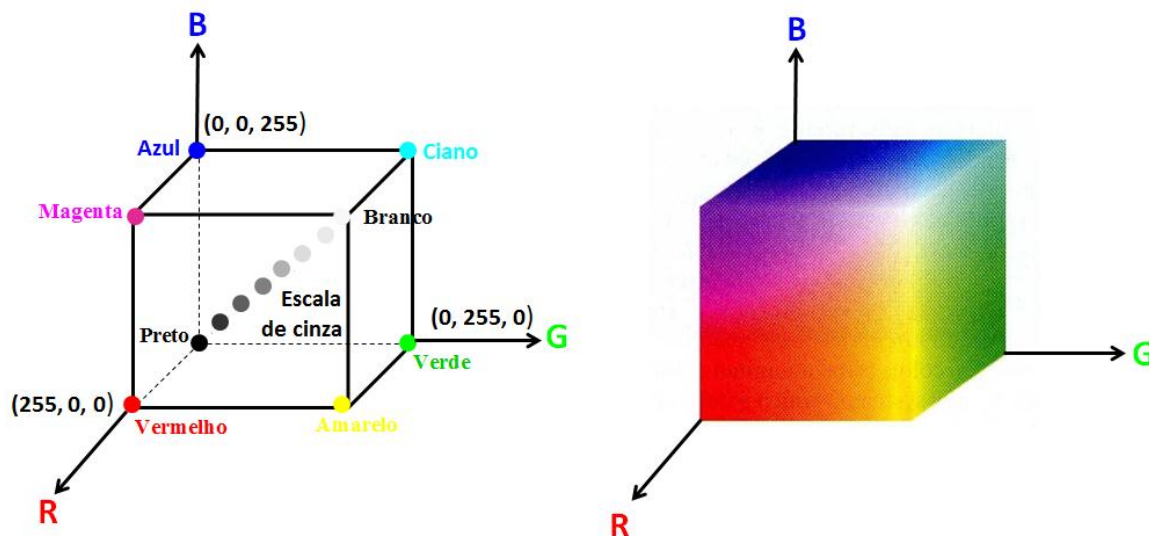


Figura 2.9: Modelo de cores RGB

FONTE: <http://www.teoriadacor.alojamentogratico.com/modelos.htm?ckattempt=1>.

Adaptada. Acesso: 04 de março de 2015.

Observando o cubo, nas arestas em R ou G ou B têm-se as cores primárias e as faces nos planos GB, BR, RG tem-se as secundárias (ciano, magenta e amarelo), formadas da combinação de duas cores primárias. Na origem do cubo, tem-se o preto e o vértice mais afastado da origem corresponde à cor branca. Na diagonal entre esses dois pontos, tem-se a escala de cinza [48].

De acordo com Solomon e Breckon [54] uma imagem no espaço de cor RGB pode ser convertida em uma imagem com tons de cinza por meio da **equação 2.2**, que é dada pela soma ponderada dos canais de cor vermelho, verde e azul.

$$I_{\text{escala de cinza}}(n, m) = \alpha I_{\text{cor}}(n, m, r) + \beta I_{\text{cor}}(n, m, g) + \gamma I_{\text{cor}}(n, m, b) \quad (2.2)$$

Em que:

- ✓ (n, m) é um pixel individual em uma imagem na escala de cinza;
- ✓ (n, m, c) é a localização de um pixel (n, m) , na imagem colorida em relação a cada componente de cor c (red, green e blue) presentes na imagem;
- ✓ $(\alpha, \beta$ e $\gamma)$ são coeficientes de ponderação com peso $\alpha=0,2989$, $\beta= 0,5870$ e $\gamma= 0,1140$, atribuídos em relação à resposta de percepção do olho humano a cada um dos canais de cores vermelho, verde e azul.



2.4.2 Modelo de Cores HSV

O modelo *HSV* é um dos sistemas de cores utilizados em aplicações de análise de imagens, mas o seu uso frequência é na computação gráfica. Ele tenta representar as imagens na cor real (verdadeira) de uma maneira que é mais natural para a percepção humana do que a representação RGB [54,56].

Este modelo descreve as cores em três atributos fundamentais: matiz, saturação e valor (ou brilho) do inglês (*hue, saturation, value*-HSV), em que o *H* (matiz) é o comprimento de onda dominante da cor, ou seja, define a tonalidade de uma área, por exemplo, vermelho, azul, verde, *S* (saturação) é a "pureza" da cor, no sentido da quantidade de luz branca misturada com matiz e *V* (valor) é o brilho da cor (também conhecido como luminância) [54].

O sistema de cor HSV é baseado em coordenadas polares, representado em forma hexagonal, uma pirâmide invertida de seis lados (**Figura 2.10**). Em cada vértice do hexágono estão presentes as cores primárias e secundárias. Neste modelo, a *matiz* é medida em ângulos, dispostos em torno do eixo central, variando de 0° a 360° , em que 0° corresponde à cor vermelha, 120° a cor verde e 240° à cor azul. As cores complementares (secundárias) estão em posições opostas (180°) no círculo graduado. A *saturação* é dada pela distância do eixo central às bordas e pode variar de 0 a 1, em que no zero tem-se a cor totalmente branca, e em 1 tem-se a cor pura. E o *valor* corresponde à altura da pirâmide, e pode variar de 0 (preto) a 1 (branco) ao longo do eixo *V* da pirâmide, onde se encontra a escala de cinza [57-58].

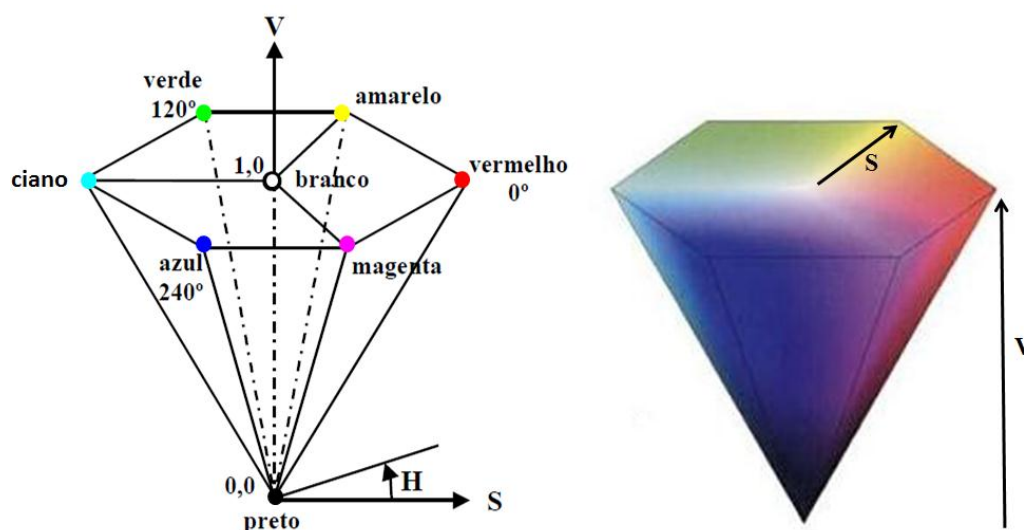


Figura 2.10: Modelo de cores HSV
FONTE: Alves (2010) [57] adaptada
Acesso em 08 de março de 2015



2.5 Quimiometria

Quimiometria pode ser entendida como a aplicação de modelos matemáticos e estáticos a dados químicos [59-60]. Dentro dessa perspectiva, a mesma pode ser adotada para planejar e delinear experimentos, assim como obter um maior número de informações a partir de dados oriundos das análises químicas, principalmente, os de natureza multivariada [59-61].

A Quimiometria é inserida a diferentes áreas da Química, sendo mais empregada na área de Química Analítica. Diversos setores, desde os acadêmicos como os industriais têm sido beneficiados com o uso da Quimiometria. [62].

As ferramentas quimiométricas para o tratamento de dados são desenvolvidas de acordo com o interesse de estudo, seja para análise qualitativa ou quantitativa. Com esta perspectiva, duas ferramentas ganharam destaque no contexto das aplicações analíticas [63-66]: (a) Reconhecimento de Padrões/Classificação e (b) Calibração Multivariada.

2.5.1 Técnicas de Reconhecimento de Padrões

As técnicas de reconhecimento de padrões, têm como finalidade avaliar a existência de similaridades e diferenças entre diferentes grupos de amostras que foram submetidas a algum tipo de análise [59,67]. Os métodos de reconhecimento de padrões podem ser classificados em dois tipos:

- ✓ Não supervisionados: aplicados para fazer uma análise exploratória dos dados. Muito utilizado no estágio inicial de uma investigação, tendo como objetivo encontrar ou identificar uma tendência, ou mesmo, uma formação de agrupamentos naturais sem usar qualquer tipo de informação prévia das classes [59,68]. Os principais métodos deste tipo são: análise de agrupamento hierárquico (HCA: *Hierarchical Cluster Analysis*) e análise de componentes principais (PCA: *Principal Componentes Analysis*) [59,68].
- ✓ Supervisionados: Tem como objetivo construir modelos a partir de um conjunto de treinamento com grupos de amostras de natureza conhecida. Então, torna-se



necessária uma informação prévia das classes para classificar futuras amostras de origem desconhecida, que poderão pertencer a uma classe conhecida, a várias classes ou a nenhuma classe [59]. Três métodos se destacam com grande aplicabilidade em problemas analíticos de classificação: análise discriminante linear (LDA: *Linear Discriminant Analysis*), análise discriminante pelos mínimos quadrados parciais (PLS-DA: *Partial Least Squares-Discriminant Analysis*) e a modelagem independente e flexível por analogia de classes (SIMCA: *Soft Independent Modeling of Class Analogy*). Porém, neste trabalho apenas os dois primeiros foram empregados.

2.5.1.1 Análise por Componentes Principais (PCA)

É uma técnica de reconhecimento de padrão não supervisionada, desenvolvida para extrair e visualizar as principais informações em dados multivariados, como a formação de agrupamento entre as amostras; identificar as variáveis que possuem maior contribuição para diferenciá-las; detectar amostras anômalas [68], entre outras funcionalidades.

A PCA permite reduzir a dimensionalidade do conjunto de dados destacando as informações importantes e eliminando as redundantes, o que facilita a interpretação dos dados [69]. Ela usa uma manipulação matemática (**Equações 2.3 e 2.4**), que transforma as variáveis originais em novas variáveis chamadas componentes principais (PCs), fatores ou variáveis latentes, que são combinações lineares das variáveis originais. Ortogonais entre si, as PCs são determinadas em ordem decrescente da variância, na qual a maior variância está contida na PC1, a PC2 possui maior informação que a PC3, e assim sucessivamente [60,70].

$$\mathbf{X} = \mathbf{t}_1 \mathbf{l}_1^T + \mathbf{t}_2 \mathbf{l}_2^T + \mathbf{t}_n \mathbf{l}_n^T + \mathbf{E} \quad (2.3)$$

$$\mathbf{X} = \mathbf{TL}^T + \mathbf{E} \quad (2.4)$$

Em que:

- ✓ **X**: matriz dos dados originais com n linhas (amostras) e m colunas (variáveis);
- ✓ **T**: matriz dos escores (número de PCs escolhido);
- ✓ **L^T**: transposta da matriz de pesos;



- ✓ E : matriz dos resíduos (variância não explicada);
- ✓ t_n : vetores dos escores;
- ✓ I_n : vetor de pesos.

2.5.1.2 Análise Discriminante pelos Mínimos Quadrados Parciais (PLS-DA)

A técnica PLS-DA, baseia-se no método de regressão PLS (*Mínimos Quadrados Parciais*) [71], cuja abordagem teórica é apresentada na seção 2.5.2.2. O método busca uma relação linear da resposta instrumental “ X ” (*variáveis independentes*), com propriedades de interesse “ Y ” (*variáveis dependentes*). No contexto da classificação de amostras, a matriz Y representa valores qualitativos, codificada com números binários 0 ou 1 [72].

Quando a classificação é entre duas classes, utiliza-se o método PLS1 e a presença de três ou mais classe utiliza-se o PLS2. Os dados são modelados por meio da análise de componentes principais, ou seja, pela obtenção das variáveis latentes (ou fatores), que tenha a covariância máxima entre a matriz de dados instrumentais e uma matriz Y [71].

Na discriminação das classes um limiar é selecionado no ponto em que o número de falsos positivos e falsos negativos seja minimizado [71].

2.5.1.3 Análise Discriminante Linear (LDA)

A análise discriminante linear tem como objetivo determinar funções discriminantes (FD) (*variáveis latentes*), as quais maximiza a distância entre as classes e minimize a distância entre os objetos ou amostras dentro de cada classe [73-74]. Essas novas variáveis são obtidas por meio de combinações lineares das variáveis originais, que melhor discrimine as classes de amostras [74].

A LDA possui algumas limitações, necessitando de dados de pequena dimensão, uma vez que na construção do modelo de classificação pode apresentar problemas de colinearidade. Desse modo, o número de variáveis deve ser menor que o número de amostras [74]. Devem-se, então, empregar métodos que reduzam a dimensionalidade dos dados (como PCA ou Transformada *Wavelet*) ou seleção de variáveis (como algoritmo das projeções sucessivas (SPA), *stepwise* (SW) ou



algoritmo genético (GA)) [74]. Apenas o SPA e SW foram empregados neste trabalho e serão apresentados na seção **2.5.3**.

2.5.2 Calibração Multivariada

A calibração multivariada tem como função construir um modelo matemático a partir de dados multivariados, para prever de forma rápida, propriedades de interesse das amostras como, por exemplo, parâmetros físicos ou químicos específicos [59,61,68]

O processo de calibração multivariada consiste basicamente de três fases: *calibração*, *validação* e *previsão*. Na calibração são construídos os modelos matemáticos, relacionando a matriz da resposta instrumental ($\mathbf{X}_{cal}$) com um vetor que possui os valores de referência ($\mathbf{y}_{cal}$) (propriedade de interesse).

Na validação é realizada uma avaliação dos modelos de modo a verificar se são satisfatórios para determinação das propriedades de interesse. Para isso, duas técnicas de validação são bastante empregadas: a validação por série de teste que usa subconjuntos separados (calibração e validação) e a validação cruzada (*cross-validation*) que usa apenas o conjunto de calibração. O processo de validação é avaliado, entre outros parâmetros, pela raiz quadrada do erro médio quadrático de validação (RMSEV) ou validação cruzada (RMSECV) [59,75]. Neste trabalho, apenas validação cruzada foi empregada e a equação que descreve o RMSECV é apresentada a seguir:

$$\text{RMSECV} = \sqrt{\frac{1}{K_c} \sum_{i=1}^{K_c} (y_c^i - \hat{y}_c^i)^2} \quad (2.5)$$

Em que K_c é o número de amostras do conjunto de calibração, y_c^k e $\hat{y}_c^k$ são respectivamente valores medidos e previsto para o parâmetro de interesse na i -ésima amostra de calibração.

Na etapa posterior, a habilidade dos modelos construídos e validados é avaliada com amostras externas (que não participaram do processo de construção e validação dos modelos). O valor de RMSEP (raiz quadrada do erro médio quadrático de previsão) (**Equação 2.6**) [75] é empregada como métrica de desempenho.



$$\text{RMSEP} = \sqrt{\frac{1}{K_p} \sum_{i=1}^{K_p} (y_p^i - \hat{y}_p^i)^2} \quad (2.6)$$

Os métodos de calibração multivariada mais aplicados em Química Analítica baseiam-se na regressão linear múltipla (MLR: *Multiple Linear Regression*) e na regressão por mínimos quadrados parciais (PLS: *Partial Least Square*) [76].

2.5.2.1 Regressão Linear Múltipla (MLR)

Entre os métodos de calibração multivariada, o MLR é o mais simples. Nele, os parâmetros (**y**) são empregados como uma função das respostas instrumentais (**X**), ou seja, assume-se uma relação linear entre cada variável dependente do vetor **y** com as variáveis independentes da matriz **X** (Equação 2.7) [68,75-76].

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e} \quad (2.7)$$

Em que **y** refere-se ao parâmetro que se pretende determinar; **X** é a matriz contendo a resposta instrumental; **b** o vetor dos coeficientes lineares de regressão e “**e**” o vetor dos resíduos (não é modelado em **y**) [75].

O coeficiente **b** é determinado de acordo com o critério dos “mínimos quadrados” e na etapa de calibração, pela (Equação 2.8):

$$\mathbf{b} = (\mathbf{X}^T\mathbf{X})^{-1} \mathbf{X}^T\mathbf{y} \quad (2.8)$$

A resolução desta equação requer a inversa da matriz $\mathbf{X}^T\mathbf{X}$. Porém, para que isso seja possível o número de variáveis deve ser igual ou menor que o número de amostras, evitando que se tenha um possível sistema indeterminado [68,75].

Entretanto, em dados multivariados de origem analítica (espectral, imagens, dentre outros) é comum o número de variáveis ser muito maior que o número de amostras, ou ainda em muitos casos, a existência de um número elevado de variáveis altamente correlacionadas. Para solucionar esses problemas e utilizar o MLR sem restrições, métodos de seleção de variáveis são úteis e devem, portanto, serem aplicados nessa ocasião [59].



2.5.2.2 Regressão em Mínimos Quadrados Parciais (PLS)

O PLS também utiliza o processo de calibração inversa. Neste método, as variáveis originais são substituídas por combinações lineares dos dados originais, denominadas fatores ou variáveis latentes [59].

Na construção do modelo, utilizam-se tanto as informações da matriz **X** (variáveis independentes) como as informações da matriz **Y** (variáveis dependentes) [76], isso para os conjuntos de calibração e validação, as quais são decompostas em suas matrizes de escores e de pesos de acordo com as **equações 2.9 e 2.10** [75].

$$\mathbf{X} = \mathbf{T}\mathbf{P}^t + \mathbf{E} \quad (2.9)$$

$$\mathbf{Y} = \mathbf{U}\mathbf{Q}^t + \mathbf{F} \quad (2.10)$$

Sendo **T**, **P** e **E** as matrizes dos escores, pesos e resíduos respectivamente para a matriz **X** e **U**, **Q** e **F** para a matriz **Y**.

O modelo resultante relaciona linearmente os escores da matriz **X** (**T**) com os escores da matriz **Y** (**U**), dados pelas **equações 2.11 e 2.12** [76].

$$\mathbf{U} = \mathbf{B}\mathbf{T}^t + \mathbf{G} \quad (2.11)$$

$$\mathbf{Y} = \mathbf{B}\mathbf{T}\mathbf{Q}^t + \mathbf{H} \quad (2.12)$$

Em que **B**, **G** e **H** são as matrizes dos coeficientes de regressão, resíduos dos escores e resíduos de concentração, respectivamente.

O número ótimo de variáveis latentes que é utilizado no modelo PLS é geralmente determinado por validação cruzada [59].

2.5.3 Seleção de Variáveis

Devido às limitações de alguns métodos quimiométricos aplicados a dados multivariados, tem-se empregado técnicas de seleções de variáveis com intuito de remover as variáveis não informativas e/ou redundantes e reter aquelas que resultem numa maior exatidão. A literatura detalha várias técnicas de seleção de variáveis [68,77-82] que podem ser empregadas tanto no contexto da classificação, como na calibração



multivariada. Destacam-se as seguintes técnicas: algoritmo genético (GA) [77], *stepwise* (SW) [78] algoritmo das projeções sucessivas (SPA) [79-81], regressão por mínimos quadrados parciais por intervalo (iPLS) [82], algoritmo Jack Knife (JK) [68], entre outras. Neste trabalho, apenas o SPA e o SW foram empregados.

2.5.3.1 Algoritmo das Projeções Sucessivas (SPA)

Foi proposto em 2001 por Araújo e colaboradores [79] no contexto da calibração multivariada. Em especial, como os modelos baseados em MLR.

O SPA tem por finalidade buscar subconjuntos representativos de variáveis com pequena colinearidade [79]. O melhor subconjunto é escolhido, por meio da modelagem do MLR, de acordo com o menor RMSEV ou RMSECV [80]. **(Equação 2.5)**. Maiores detalhes sobre o funcionamento do SPA no contexto da calibração multivariada podem ser encontrados em Araújo et al. [79] e Galvão et al. [80].

Pontes et al. [81] também adaptaram o algoritmo SPA no contexto de classificação, especialmente, para modelos baseados na LDA. Neste caso, a função de custo empregada para guiar a escolha das variáveis foi o risco médio G de uma classificação incorreta pela LDA, calculada com base no conjunto de validação. **(Equação 2.13)** [81].

$$G = \frac{1}{k_v} \sum_{i=1}^{k_v} g_k \quad (2.13)$$

Onde g_k (risco de uma classificação incorreta do objeto x_k da k -ésima amostra de validação) definido como:

$$g_k = \frac{r^2(x_k, \mu_{IK})}{\min_{Ij \neq IK} r^2(x_k, \mu_{IK})} \quad (2.14)$$

Em que:

- ✓ $r^2(x_k, \mu_{IK})$ corresponde ao quadrado da distância de Mahalanobis entre o objeto x_k (com índice de classe Ik) e a média de sua classe (μ_{IK}).
- ✓ $\min_{Ij \neq IK} r^2(x_k, \mu_{IK})$ corresponde ao quadrado da distância de Mahalanobis entre o objeto x_k e o centro da classe errada mais próxima.

Idealmente, $\mathbf{g}_k$ deverá ser tão pequeno quanto possível, ou seja, o objeto $\mathbf{x}_k$ deverá estar perto do centro da sua verdadeira classe e distante dos centros das demais classes.

É importante ressaltar que o SPA para classificação utilizado neste trabalho foi executado com a validação por série de teste. Todavia, é possível encontrar trabalhos na literatura [83] envolvendo seu uso com validação interna (validação cruzada) e, para isso, utiliza-se apenas o conjunto de treinamento para o cálculo do custo.

2.5.3.2 Stepwise (SW)

O *stepwise* é um método de seleção de variáveis que vem sendo comumente utilizado em problemas de classificação aplicado em LDA [84-87].

Com este algoritmo, verifica-se a importância das variáveis, adicionando-as ou excluindo-as do modelo, com base no fator de discriminação (D_i) individual de cada uma, que leva em conta as medidas de dispersão dentro da classe e entre as classes (Equação 2.15) [74,78].

$$D_i = \frac{SB_i}{SW_i} \quad (2.15)$$

Em que SB_i a dispersão entre as classes e o SW_i a dispersão dentro da classe.

A variável n_i que apresenta o maior fator de discriminação é selecionada e um procedimento de validação cruzada *leave-one-out* é efetuado, de modo a obter o número de erros. Para evitar problemas de colinearidade as variáveis que são altamente correlacionadas com a variável selecionada são descartadas. Esse procedimento é repetido com todas as variáveis até que o grupo de variáveis que apresentar o menor número de erros seja escolhido [74,78].

Neste método, pode-se testar diferentes valores de limiar (*threshold*) variando 0,10 a 0,99, comparados de acordo com o número de erros do conjunto de validação [74].

O algoritmo empregado neste trabalho foi conduzido apenas no contexto de classificação.



2.5.4 Seleção de Amostras

Na seleção de amostras em estudos de quimiometria, o algoritmo *Kernard-Stone* (KS) ^[88] é o mais utilizado. De acordo com Galvão et al. ^[89], o KS visa selecionar um subconjunto representativo de um conjunto de N amostras, maximizando as distâncias euclidianas entre os vetores das respostas instrumentais ($\mathbf{X}$) das amostras selecionadas. Essa distância é denotada como $d_x(p,q)$ entre os x -vetores de cada par (p,q) das amostras selecionadas, calculada pela **equação 2.16**.

$$d_x(p, q) = \sqrt{\sum_{j=1}^J [x_p(j) - x_q(j)]^2}; \quad p, q \in [1, N] \quad (2.16)$$

Sendo $x_p(j)$ e $x_q(j)$ as respostas instrumentais na j -ésima variável, para as amostras “ p ” e “ q ”, respectivamente. Já “ j ” indica o número de variáveis no conjunto de dados.

Galvão et al. ^[89] adaptaram o algoritmo KS para o contexto de calibração multivariada. Para esse propósito, criou-se o SPXY (*Sample set Partitioning based on joint X-y distances*- Partição de amostra baseado na distância de $\mathbf{X}$ - $\mathbf{y}$) que abrange as diferenças tanto de $\mathbf{X}$ como de $\mathbf{y}$ no cálculo das distâncias inter-amostras. Em que a distância $d_y(p,q)$ pode ser calculada para cada par de amostras “ p ” e “ q ”, como descrito na **equação 2.17**.

$$d_y(p, q) = \sqrt{(y_p - y_q)^2} = |y_p - y_q|; \quad p, q \in [1, N] \quad (2.17)$$

Com a finalidade de atribuir igual importância para a distribuição das amostras em $\mathbf{X}$ e no espaço $\mathbf{y}$, as distâncias $d_x(p,q)$ e $d_y(p,q)$ são divididas pelos seus valores máximos no conjunto de dados. Desta forma, as distâncias Xy são normalizadas de acordo com a **equação 2.18**.

$$d_{xy}(p, q) = \frac{d_x(p,q)}{\max_{p,q \in [1, N]} d_x(p,q)} + \frac{d_y(p,q)}{\max_{p,q \in [1, N]} d_y(p,q)}; \quad p, q \in [1, N] \quad (2.18)$$

2.6 Revisão da Literatura

Nos últimos anos, uma grande variedade de trabalhos científicos foram publicados, com o uso de imagens digitais junto à quimiometria, no desenvolvimento de modelos de classificação e calibração multivariada, aplicados para a análise de diferentes matrizes [43,84, 90-98].

Diniz et al. [90] utilizaram imagens digitais, obtidas por uma câmera digital com os modelos de cores RGB, HSB (matiz, saturação, brilho) e escala de cinza, na construção de modelos SIMCA e SPA-LDA para classificar chás naturais não processados. Os modelos de classificação foram comparados e o SPA-LDA forneceu resultados mais promissores, classificando corretamente todas as amostras de acordo com as suas regiões geográficas: solos brasileiros, argentinos ou estrangeiros.

Oliveira et al. [43] utilizaram imagens digitais obtidas por uma impressora multifuncional e PCA para a identificação de amostras de solos com concentrações de Cr (VI) superiores aos valores estabelecidos pelas legislações internacionais. As imagens foram analisadas nos níveis de cores (R, G, B, H, S, V, r, g, b e L) e, como resultado, houve boa separação entre os valores de concentração de Cr (VI), principalmente quando os valores não estão próximos às faixas estabelecidas pela legislação do Canadá, Suécia e Itália.

Milanez e Pontes [84] utilizaram imagens digitais obtidas por uma *webcam*, as quais foram decompostas em histogramas com três modelos de cores RGB, HSI (matiz, saturação, intensidade) e tons de cinza. Os dados foram tratados com técnicas de reconhecimento de padrões (LDA-SPA e LDA-SW), para classificar óleos vegetais comestíveis em relação ao tipo de oleaginosa e ao estado de conservação. De modo geral os melhores resultados foram obtidos para o modelo SPA-LDA.

Almeida et al. [91] utilizaram imagens digitais, também obtidas por uma simples *webcam*, tendo histogramas com os canais de cores RGB, HSV e tons de cinza. De acordo com os autores, técnicas de reconhecimento de padrão supervisionada (SIMCA, PCA-LDA, PLS-DA e SPA/LDA) foram aplicadas para classificar cinco diferentes tipos de bactérias (*Escherichia coli*, *Enterococcus faecalis*, *Streptococcus salivarius*, *Streptococcus oralis*, e *Staphylococcus aureus*). O melhor resultado foi obtido usando o SPA/LDA com a combinação RGB, resultando em uma taxa de classificação correta de 94% e 100% para os conjuntos de treinamento e teste, respectivamente.

Vilar et al. [92] utilizaram imagens digitais obtendo a distribuição de frequência de índices de cor em RGB, HSI e os canais de tons de cinza associada às técnicas LDA-SPA e PLS-DA, para a classificação de sementes de mamona com relação a quatro tipos de cultivares: BRS Nordestina, BRS Paraguaçu, BRS Energia e CNPA 2009-7. Os melhores resultados foram obtidos usando o modelo PLS-DA.

Souto et al. [93] utilizaram imagens digitais obtidas em um scanner, com a distribuição de frequência HLS (matiz-luminosidade-saturação), associada às técnicas de reconhecimento padrão supervisionadas: SPA/LDA, SIMCA e PLS-DA, para identificar adulterações em café torrado. Os modelos foram comparados e, de modo geral, o SPA/LDA forneceu um desempenho significativamente melhor em relação aos outros modelos, chegando a uma precisão média de 92,5% para ambos os conjuntos de treinamento e teste.

Santos et al. [94] utilizaram imagens digitais obtidas por um scanner e modelos de calibração MLR, PLS e PCR (Regressão por componentes principais), bem como PCA, para detecção e quantificação de adulterações em amostras de leite de vaca líquido. As adulterações foram feitas com água e NaOH e dez parâmetros de cor foram utilizados para avaliar a informação a partir das imagens: vermelho, verde, azul, matiz, saturação, valor, cores relativas (r, g e b), e intensidade. Na quantificação de água, os modelos PLS e PCR apresentaram um resultado mais satisfatório com erro médio quadrático de previsão menor do que 7%. Quanto à adulteração do leite com NaOH, os níveis de cores B, S, g, e b apresentaram as maiores diferenças entre as amostras.

Botelho et al. [95] desenvolveram uma nova metodologia para a quantificação de um corante artificial, o amarelo sunset (SY), em bebidas suaves (isotônicos e refrigerantes), utilizando imagens digitais obtidas por um *scanner* (histogramas RGB) e Regressão PLS. De acordo com os autores, erros de previsão relativos inferiores a 10% foram alcançados.

Dominguez e Centurión [96] aplicaram o PLS a dados provenientes de imagens digitais capturadas por uma *webcam* para determinar a cor em amostras de mel. Três modelos de cores (RGB, HSB e em tons de cinza) foram avaliados e os melhores resultados foram encontrados com o HSB.

Técnicas baseadas em imagens digitais vêm sendo aplicadas em trabalhos relacionados com o algodão, avaliando-se a cor, bem como sua distribuição, forma, textura e tamanho dos objetos [97]. Mais especificamente, determinando-se +b e Rd,

com o intuito de estabelecer um índice de cor do algodão para propósito de classificação [32,36] e identificação dos resíduos presentes em sua fibra [98-99].

Mustafic e Li [98] utilizaram imagens fluorescentes em amostras de algodão com uma câmera digital de lente com reflexo único (SLR: single-lens reflex) para identificar e classificar treze tipos de resíduos (botânicos e não-botânicos) presentes no algodão.

Liu et al. [99] propuseram um método para distinguir contaminantes brancos (papel, plásticos e pena) presentes no algodão com tonalidade branca usando imagens a lasers de linha vermelha de 50 MW de potência e comprimento de onda de 650 nm.

Poucos trabalhos [100-101] têm sido propostos utilizando imagens digitais com quimiometria para análise da qualidade em pluma de algodão.

Ji et al. [100] propuseram um classificador de contaminantes de fibras de algodão com base em uma máquina de vetor de suporte, utilizando imagens digitais adquiridas por uma câmera de cor de varredura de linha 3CCD, e um algoritmo para obter as características dos contaminantes.

Yang et al. [101] identificaram e classificaram materiais não vegetais presentes no algodão (fibras tingidas, filme plástico, cordas de polipropileno, dentre outros), por meio de imagens capturadas por uma câmera CCD de varredura de linha. Os autores utilizaram os canais R, G e B, extraíndo as características: cor, forma e textura para a classificação empregando um gráfico acíclico direto de máquina de vetor de suporte de multi-classe construído.

É importante ressaltar que trabalhos envolvendo o uso de imagens digitais associado à quimiometria, no desenvolvimento de modelos de classificação e calibração multivariada, aplicados à fibras de algodão não foram encontrados na literatura consultada até a data de redação desta dissertação.



CAPÍTULO III

METODOLOGIA

3 METODOLOGIA

3.1 Amostras

3.1.1 Estudo para Classificação

Neste trabalho foram utilizadas amostras comerciais de quatro cultivares de fibras naturalmente coloridas de algodão (BRS 200 Marrom, BRS Topázio, BRS Rubi e BRS Safira). Todas as amostras foram fornecidas pela Embrapa Algodão (Empresa Brasileira de Pesquisa Agropecuária), localizada na cidade de Campina Grande-PB, Brasil,

As amostras foram adquiridas em diferentes períodos de colheita do ano de 2013 e oriundas de lotes independentes. Um total de 400 amostras (100 para cada cultivar) foram obtidas dos lotes disponíveis.

A **Figura 3.1** apresenta as imagens das quatro cultivares de algodão empregadas neste trabalho. As cultivares foram divididas em dois grupos de acordo com sua similaridade. O primeiro grupo (**Figura 3.1a**) foi constituído pelas cultivares BRS 200 Marrom e BRS Topázio, de coloração marrom-clara. Já o grupo II (**Figura 3.1b**) formado pelas cultivares com coloração marrom-escura predominante: BRS Rubi e BRS Safira.

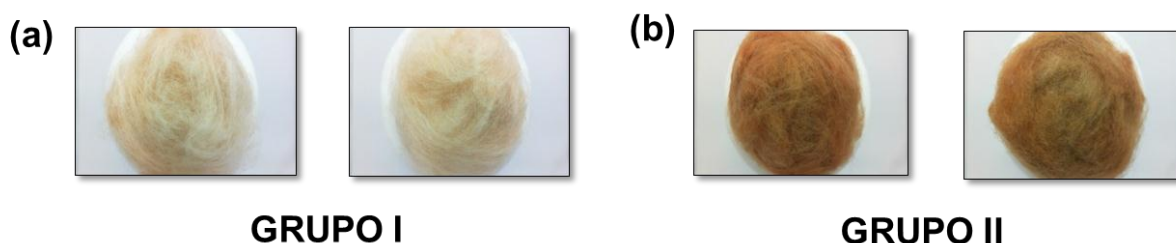


Figura 3.1: Plumas de Algodão Naturalmente Colorido (a) BRS 200 Marrom e BRS Topázio (Grupo I) e (b) BRS Rubi e BRS Safira (Grupo II).

Como pode ser visto na **Figura 3.1**, a grande semelhança entre as amostras dentro de cada grupo é evidenciada. Assim, torna-se difícil identifica-las por uma inspeção visual. Apesar da similaridade visual das plumas em cada grupo, tais cultivares têm características diferentes, o que impactam no seu valor comercial. O estudo de classificação envolvido neste trabalho foi realizado separadamente para cada grupo de cultivares.

3.1.2 Estudo para Calibração

O estudo de calibração multivariada envolveu amostras de plumas brancas (cultivares BRS 8H e BRS Aroeira) e naturalmente coloridas (BRS 200 Marrom, BRS Topázio, BRS Rubi, BRS Safira e BRS Verde). A **Figura 3.2** mostra as imagens das diferentes cultivares estudadas.



Figura 3.2: Plumas das cultivares de algodão branco e naturalmente colorido utilizadas no estudo de calibração.

Assim como no estudo de classificação, todas as amostras foram provenientes da Embrapa Algodão, sendo obtidas a partir de diferentes lotes e períodos do ano 2013. Um total de 76 amostras foi empregado para a determinação do grau de amarelamento e reflectância, enquanto que para o teor de cera, foram empregadas 63 amostras. Uma massa de 10,0 g foi utilizada para a realização de todas as análises, incluindo as imagens digitais e determinação dos parâmetros de qualidade.

Antes de iniciar as análises, tanto para o estudo de classificação como calibração, as amostras foram acondicionadas a uma temperatura e umidade relativa do ar de $20 \pm 2^\circ\text{C}$ e $65 \pm 2\%$, respectivamente, durante 24 horas, seguindo a Prática Padrão ASTM D 1776 -*Standard Practice for Conditioning and Testing Textiles*) ^[102].

3.2 Aquisição das Imagens Digitais

3.2.1 Sistema de Aquisição de Imagens Digitais

A **Figura 3.3** ilustra o sistema utilizado para a aquisição das imagens digitais. É um sistema simples e com poucos componentes: (a) compartimento de madeira com dimensões de 30 cm × 22 cm × 23 cm; (b) uma lâmpada fluorescente em forma de espiral; c) uma célula de *Teflon*®, com diâmetro e comprimento de 4,7 cm e 12 cm, respectivamente, empregada como suporte para as amostras de algodão, (d) uma *webcam* Microsoft com resolução HD (1280 x 720), modelo *LifeCam Cinema* e (e) notebook com processador core i3, com memória de 4 GB e HD de 500 GB.

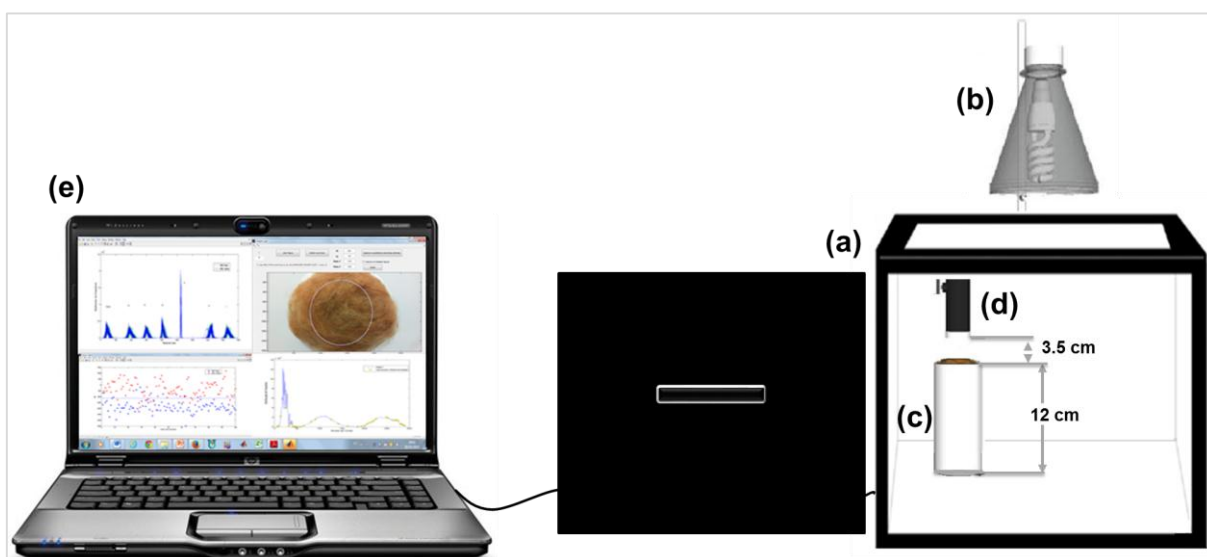


Figura 3.3: Sistema usado na aquisição das imagens de amostras de plumas de algodão naturalmente colorido (a) compartimento, (b) lâmpada fluorescente, (c) suporte de *Teflon*®, (d) *webcam* e (e) notebook

3.2.1.1 Compartimento

Como observado na **Figura 3.3**, o compartimento foi completamente fechado com intuito de isolar a amostra da luz externa, bem como para garantir uma melhor qualidade na obtenção das imagens. A parte inferior foi revestida com folha de papel branco (tamanho A4) para que todas as cores presentes na imagem das plumas de algodão fossem detectadas pela *webcam*.

3.2.1.2 Iluminação do Sistema

A lâmpada fluorescente utilizada para iluminação do sistema foi de 6 W de potência e temperatura de 4000 K (cor branca). A mesma foi alocada na parte externa do compartimento a uma distância de 12 cm da amostra. Assim, uma iluminação do sistema mais uniforme foi alcançada. Para auxiliar na dispersão da luz no interior da caixa, um difusor constituído por uma folha de papel branco A4 foi colocado na parte superior do compartimento, servindo também como uma tampa para o mesmo.

3.2.1.3 Webcam e Notebook

Uma *webcam*, devidamente conectada ao notebook, foi posicionada a uma distância de 3,5 cm da amostra para o registro das imagens. Assim, foi possível capturar e estocar as imagens, bem como realizar o tratamento quimiométrico.

3.2.2 Obtenção das Imagens Digitais

O registro das imagens foi realizado de forma aleatória e individual para cada amostra com o software da própria *webcam*. A configuração da câmera foi mantida constante durante todo o processo de medida.

Na aquisição das imagens de cada amostra, uma massa de 130,0 mg foi empregada e cinco imagens sequenciais foram registradas e armazenadas em formato JPEG, com 24 bits (16,7 milhões de cores) e uma resolução espacial de 2880 x 1620 pixels.

Ao término da captura das cinco imagens sequencias, em cada imagem selecionou-se uma região central em forma de círculo com dimensões $X_i = 702$ e $Y_i = 218$ que indicam a posição da área selecionada sobre a figura e são dados em pixels e Raio 1= 1244 e Raio 2= 1200 que correspondem ao tamanho da área sobre a figura, também em pixels. Essas dimensões foram mantidas constante para todas as amostras (algodões de fibra branca e coloridas). Usando apenas a região selecionada das imagens, as distribuições de frequência (histograma) dos índices de cor foram obtidas para cada uma das cinco imagens. Em seguida, a média dos cinco histogramas foi calculada, para ser utilizada em todo o tratamento quimiométrico.

Esse procedimento empregado para obtenção dos histogramas, foi realizado por uma interface implementada no Matlab denominada de Image_Gui. A **Figura 3.4** resume o procedimento de aquisição do conjunto de dados.

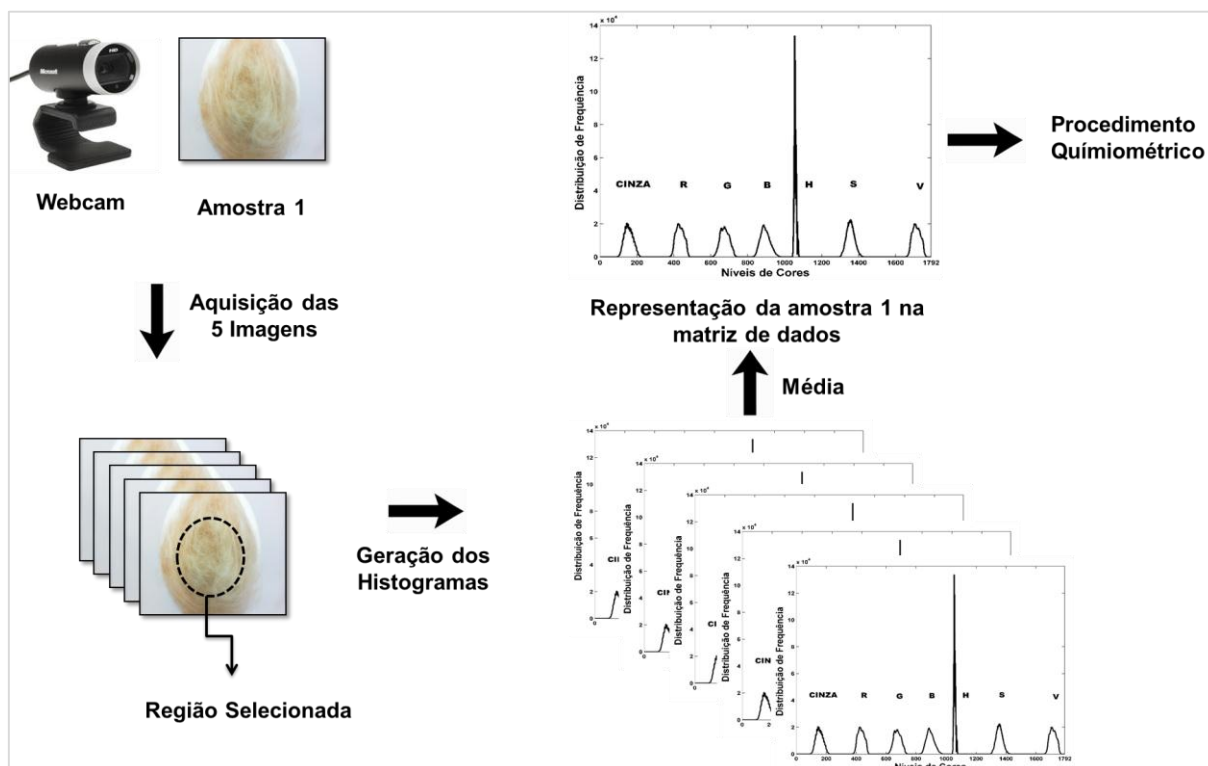


Figura 3.4: Procedimento para obtenção dos dados analíticos.

Para obter os histogramas, três modelos para a cor de um pixel foram utilizados neste estudo, tais como: o padrão RGB (R – vermelho, G - verde e B- azul) escala de cinza e o sistema HSV (H - matiz, S - saturação e V - Valor).

Cada componente de cor dos modelos é composto de 256 níveis de cores que variam de 0 a 255. Então os setes canais de cores envolvem um total de 1792 variáveis.

O conjunto de dados utilizado para o tratamento quimiométrico é inicialmente formado por uma matriz, em que as amostras estão dispostas nas linhas, enquanto as variáveis, correspondentes aos níveis de cores de cada canal, encontram-se nas colunas. A **Figura 3.5** ilustra a representação do conjunto de dados formado.

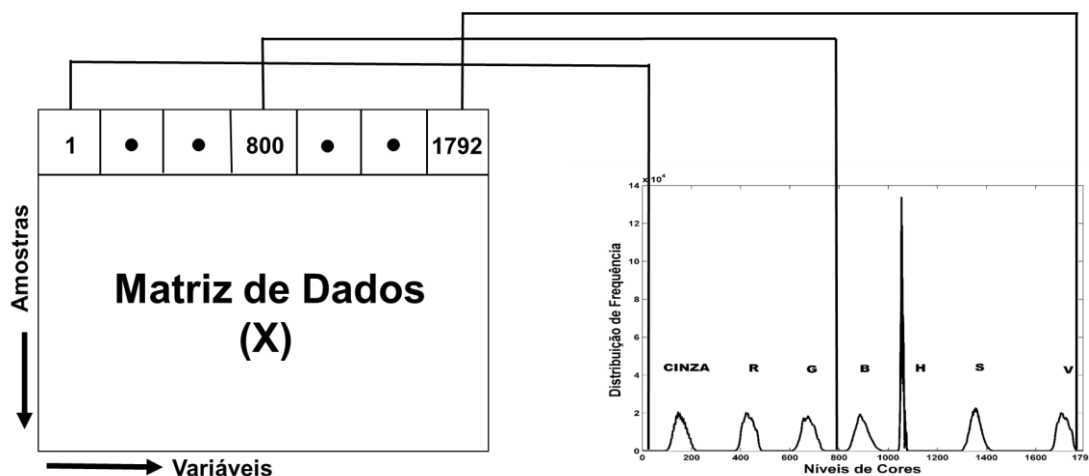


Figura 3.5: Representação da matriz de dados.

3.3 Análise de Referência

3.3.1 Obtenção do Grau de amarelamento (+b) e Reflectância (Rd)

Esta análise foi realizada no Laboratório de Fibras e fios da Embrapa Algodão, Campina Grande-Paraíba, Brasil.

O equipamento HVI (High Volume Instrument), modelo Uster® HVI 1000 **Figura 3.6** foi empregado para a determinação dos parâmetros físicos de qualidade (+b e Rd) das amostras de plumas de algodão. Cada propriedade foi medida três vezes e a média foi aplicada para o tratamento dos dados. Nessa análise, utilizou-se uma massa de aproximadamente de 10,0g de cada amostra.



Figura 3.6: Sistema HVI (High Volume Instruments) modelo HVI 1000.

FONTE: USTER®. HVI 1000. **Application Report. Quality Characteristics used for Cotton Classification.** USDA/Uster Technologies, 2006. SE 598 [31]
Acesso em 07 de abril de 2015.

3.3.2 Obtenção do Teor de Cera (WAX)

Para a determinação do teor de cera, a metodologia empregada foi de acordo com Conrad (1944) [38]. Essa análise foi realizada no Laboratório Avançado de Tecnologia Química (LATECQ) da Embrapa Algodão.

O procedimento envolve o uso de extratores do tipo Randall devidamente identificados. Para cada amostra, cerca de 2,5 g (**m_a**) de pluma de algodão foi colocada em um cartucho confeccionado a partir de papel de filtro qualitativo dobrado e 50,0 mL de álcool etílico P.A. (99,5%) foi empregado como solvente. A temperatura de extração foi de 100°C durante 6 horas ininterruptas com reposição do solvente quando necessário (**Figura 3.7**).



Figura 3.7: Processo de extração de cera de plumas de algodão com álcool em extratores de Randall.

Ao término das 6 horas, a segunda etapa iniciou-se transferindo as amostras contidas nos balões de Randall para funis de separação com capacidade de 500 mL.

Os balões foram lavados três vezes com 5,0 mL de álcool etílico P.A. (99,5%) a fim de garantir que todo o extrativo obtido da pluma fosse transferido. Posteriormente, adicionaram-se 50,0 mL de clorofórmio P.A. (99%) e 40 mL de água destilada. A solução formada foi agitada por 5s e, em seguida, mantida em repouso por 24 h. Ao concluir este procedimento, observou-se a formação de duas fases distintas (**Figura 3.8**) nos funis.

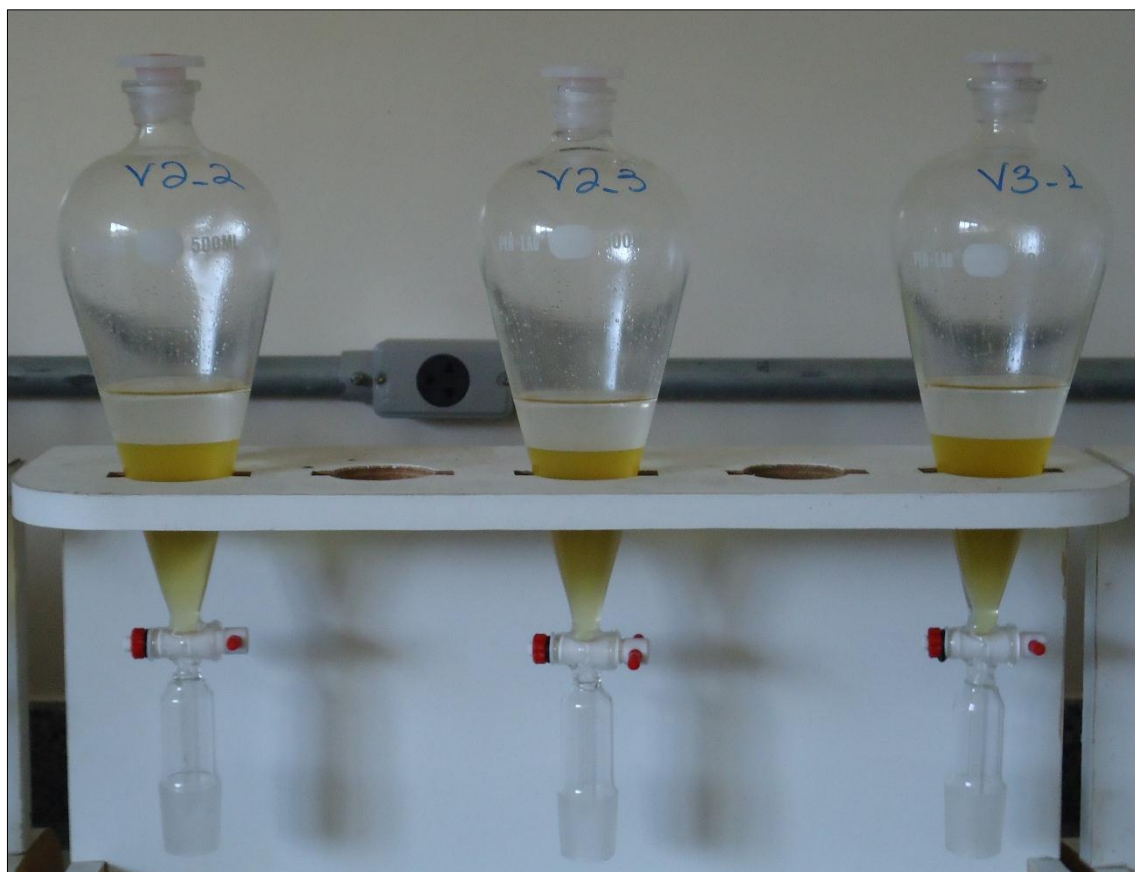


Figura 3.8: Separação líquido-líquido para determinação de cera.

Das duas fases distintas observadas na **Figura 3.8**, transferiu-se cuidadosamente a fase clorofórmica para balões de Randall previamente lavados, secos (em estufa a 120 °C por 30 minutos), identificados e pesados (m_o).

Finalmente os balões de Randall com a solução de cera e solventes foram colocados em banho-maria à temperatura de 100 °C até que todo o solvente evaporasse. Com isso, restou-se apenas a cera no balão de Randall. Deixou-se em média 20 min em dessecador e logo após sua massa foi medida (m_f)

O cálculo do teor de cera em fibras de algodão foi realizado aplicando a **equação 3.1**:

$$\text{WAX}(\%) = \frac{m_f - m_o}{m_a} \times 100 \quad (3.1)$$

Em que m_o e m_f são a massa inicial e final do balão de Randall respectivamente; m_a é a massa da pluma de algodão.

Para cada amostra de pluma algodão, a medida foi realizada em triplicata. Por conseguinte a média do teor de cera foi utilizada para o procedimento quimiométrico.

3.4 Procedimentos Quimiométricos e Software

3.4.1 Reconhecimento de Padrões/Classificação

No grupo I, o procedimento quimiométrico foi realizado para classificar as cultivares BRS 200 Marrom e BRS Topázio. Para o grupo II, a discriminação foi feita entre as cultivares BRS Rubi e BRS Safira.

Antes de construir os modelos de classificação, uma PCA foi realizada separadamente para cada grupo de amostras com intuito de realizar uma análise exploratória dos dados. Com isso, foi possível identificar semelhanças e diferenças nos diferentes tipos de amostras.

Na construção dos modelos de classificação os conjuntos de dados foram divididos em subconjuntos de treinamento (50% das amostras), validação (25% das amostras) e teste (25% das amostras) utilizando o algoritmo KS [88]. Este algoritmo foi aplicado separadamente para cada classe, como reportado por Pontes et al [81]. São apresentados nas **Tabelas 3.1 e 3.2** o número de amostras de treinamento, validação e teste para cada grupo de fibras de algodão colorido.

Tabela 3.1: Número de amostras de treinamento, validação e teste para cultivares do grupo I.

Cultivares	Subconjuntos		
	Treinamento	Validação	Teste
BRS 200 Marrom	50	25	25
BRS Topázio	50	25	25
Total	100	50	50

Tabela 3.2: Número de amostras de treinamento, validação e teste para cultivares do grupo II.

Cultivares	Subconjuntos		
	Treinamento	Validação	Teste
BRS Rubi	50	25	25
BRS Safira	50	25	25
Total	100	50	50

Os subconjuntos de treinamento e validação foram utilizados no procedimento de modelagem, enquanto que o subconjunto de teste foi empregado para avaliação do desempenho dos modelos resultantes. A métrica de comparação dos modelos foi realizada em função da taxa de classificação correta (TCC) (**Equação 3.2**).

$$\text{TCC (\%)} = \frac{(\text{Número de acertos de classificação})}{(\text{Número total de amostras do subconjunto de teste})} \times 100 \quad (3.2)$$

Os modelos de classificação desenvolvidos e avaliados foram: LDA/SPA e LDA/SW. Para fins de comparação, os modelos baseados PLS-DA também foram empregados usando o histograma completo.

Este trabalho adota a formulação SPA apresentado por Pontes et al. ^[81]. Para o algoritmo SW, seis valores de coeficientes de correlação múltipla (0,1; 0,3; 0,7; 0,9; 0,95 e 0,99) foram testados, a fim de decidir o descarte de variáveis.

Nos modelos PLS-DA, utilizou-se validação cruzada *leave-one-out*, com a escolha do número de fatores baseada no menor número de erros de classificação. O limiar para definir as classes no ponto em que o número de falsos positivos e falsos negativos seja minimizado, foi calculado com base no teorema de Bayes ^[103].

3.4.2 Calibração Multivariada

A fim de determinar os três parâmetros de qualidade das plumas de algodão (+b, Rd e WAX), foram construídos modelos de calibração multivariada utilizando os histogramas obtidos das imagens e os valores obtidos pelos métodos de referência. Para tanto, foram avaliados os modelos PLS (histograma completo) e o MLR com



seleção de variáveis pelo SPA. Antes da construção e validação dos modelos, os conjuntos de dados foram divididos em subconjuntos de calibração (60%) e previsão (40 %) aplicando tanto o algoritmo KS [88] como o algoritmo SPXY [89], os quais, foram comparados em função dos modelos desenvolvidos.

A **Tabela 3.3** detalha a distribuição das amostras para determinação dos parâmetros: grau de amarelamento (+b), reflectância (Rd) e teor de Cera (WAX).

Tabela 3.3: Número de amostras de calibração e previsão usadas na determinação do grau de amarelamento (+b), reflectância (Rd) e teor de cera (WAX).

Parâmetros	Número de amostras	Subconjuntos	
		Calibração	Previsão
+b e Rd	76	46	30
WAX	63	38	25

O subconjunto de calibração foi utilizado na construção e validação cruzada dos modelos, incluindo seleção de variáveis pelo SPA e fatores pelo PLS, enquanto que o subconjunto de previsão foi empregado para avaliação da habilidade preditiva dos modelos.

O desempenho dos modelos PLS e MLR-SPA foi avaliado de acordo com os valores de RMSEP, coeficientes de correlação (r) e de determinação (r^2). Um teste F , ao nível de confiança 95%, foi realizado com intuito de avaliar a existência de diferenças estatisticamente significativas entre os valores de RMSEP obtidos pelos modelos PLS (histograma-completo) e MLR-SPA. O teste t -pareado, ao nível de confiança de 95%, foi empregado para avaliar a existência de diferenças significativas entre o método proposto e o método de referência. Os valores de t -calculado foram comparados com os valores de t -crítico (0,95; n) tabelado.

A PCA e os modelos PLS foram obtidos utilizando Unscrambler® X.1 ®. (CAMO/SA). Os algoritmos KS, SPXY, LDA/SPA, LDA/SW, PLS-DA e MLR-SPA foram executados em Matlab® 7.10.



CAPÍTULO IV

RESULTADOS E

DISCUSSÃO

4 RESULTADOS E DISCUSSÃO

4.1 Estudo de Classificação

4.1.1 Histogramas

Os histogramas (com sete canais de cores: tons de cinza, vermelho, verde, azul, matiz, saturação, e valor) obtidos por meio das imagens digitais registradas das 400 amostras de pluma de algodão para os grupos I e II são apresentados nas **Figuras 4.1**. Os histogramas representam as distribuições de frequência dos níveis de cores presentes na imagem e a quantidade de pixels relacionados a cada nível de cor. Considerando os sete canais, 1792 níveis de cores os quais correspondem a 256 variáveis (0 a 255) por cada canal foram coletados.

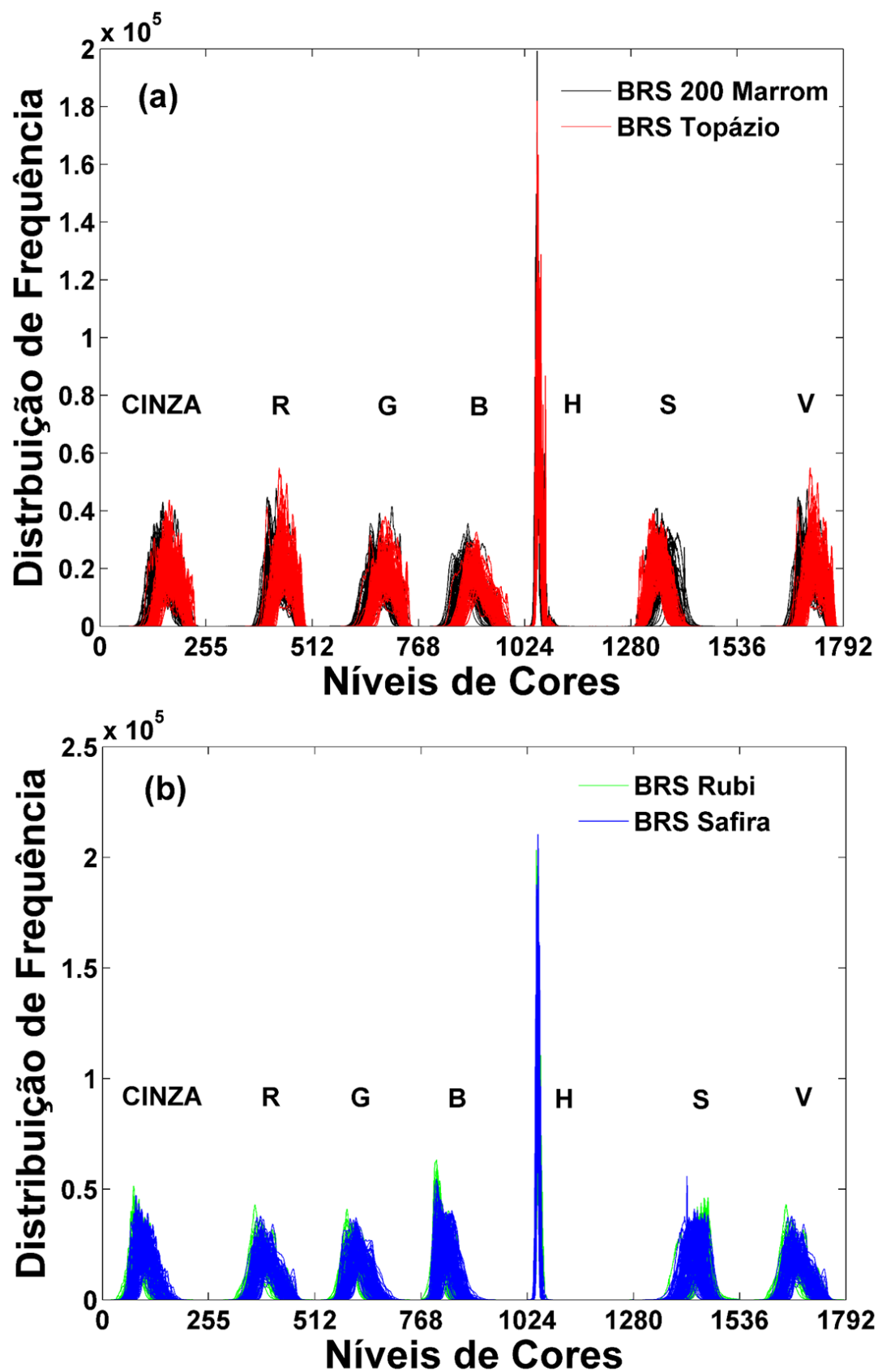


Figura 4.1: Histograma completo para o grupo (a) I e (b) II. R: Red, G: Green, B: Blue, H: Hue, S: Saturation e V: Value.

Observando as **Figuras 4.1a-b**, percebe-se que há variáveis com nível de cor igual a zero. Estas foram eliminadas, resultando em histogramas com 803 e 845 variáveis para o grupo I e II, respectivamente. Os histogramas resultantes foram utilizados no tratamento quimiométrico.

As **Figuras 4.2 e 4.3** apresentam os histogramas médios de cada cultivar após a remoção de variáveis não informativas para os grupos I e II, respectivamente. É possível perceber uma sutil diferença entre os perfis médios das duas cultivares, tanto para o grupo I, como para o grupo II. Tal comportamento não é tão evidente para o canal H (matiz). De fato, este canal corresponde ao comprimento de onda dominante em relação à cor e como as plumas de cada grupo individual são de coloração semelhante, é possível que as diferentes cultivares do grupo I absorvam no mesmo comprimento de onda visível. Comportamento parecido ocorre com as cultivares do grupo II.

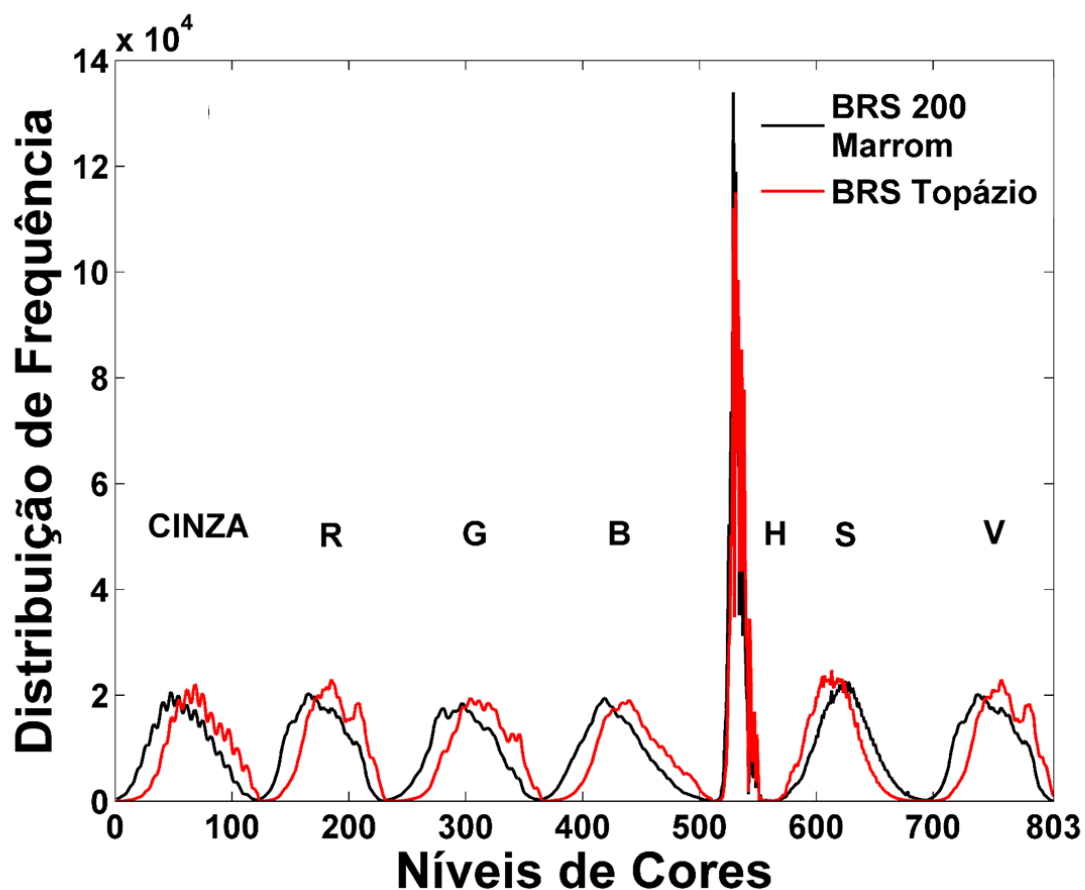


Figura 4.2: Histograma médio para o grupo I. R: Red, G: Green, B: Blue, H: Hue, S: Saturation e V: Value.

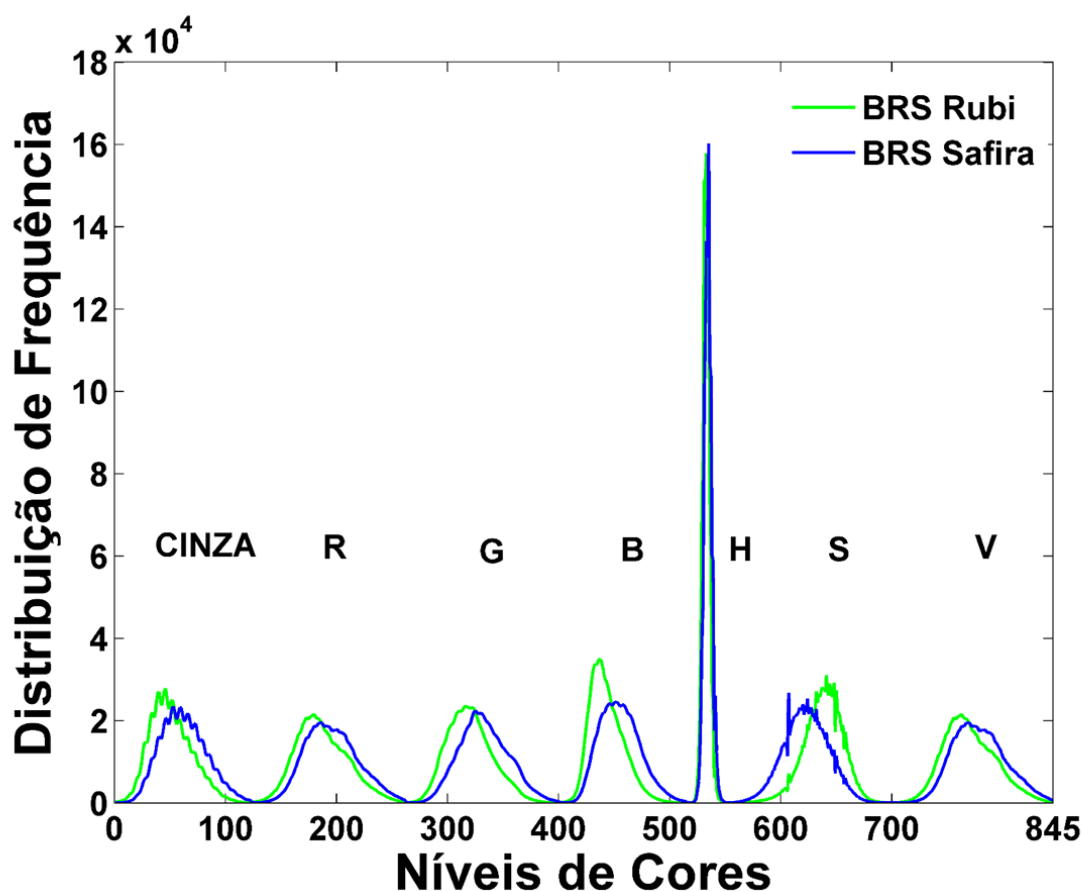


Figura 4.3: Histograma Médio para o grupo II. R: Red, G: Green, B: Blue, H: Hue, S: Saturation e V: Value.

4.1.2 Análise por Componentes Principais (PCA)

Um estudo exploratório dos dados foi realizado utilizando a PCA aplicada separadamente aos histogramas de cada grupo de pluma. Para esse propósito, a avaliação foi realizada no domínio dos histogramas completos (todos os canais), nos canais individuais, bem como na combinação entre eles. Como resultado, diferentes gráficos de escores e de pesos para a combinação de todos os canais foram obtidos.

A **Figura 4.4** e **4.5** correspondem aos gráficos de escores e pesos para o grupo I, enquanto que as **Figuras 4.6** e **4.7** aos gráficos de escores e pesos, respectivamente, para o grupo II.

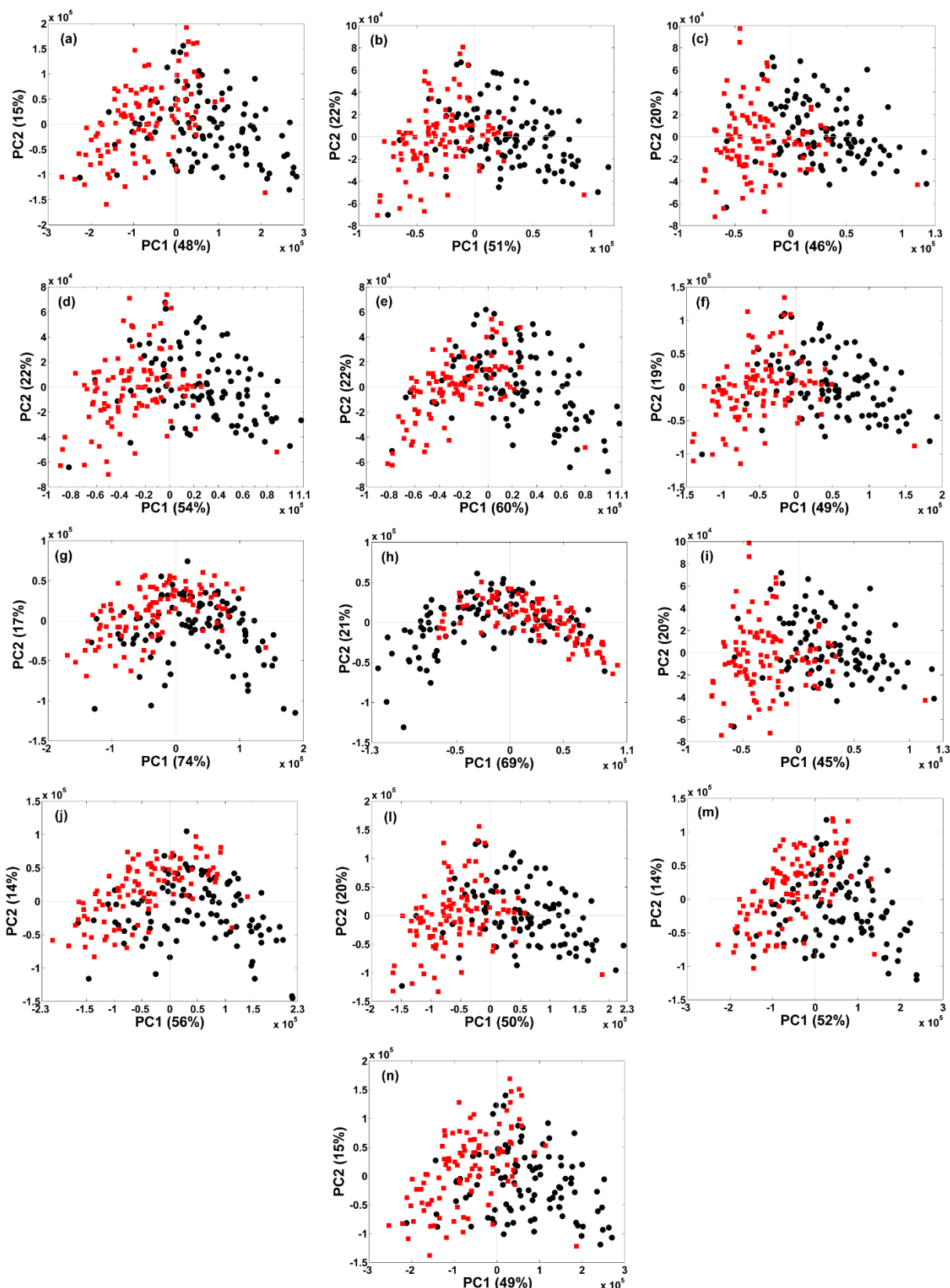


Figura 4.4: Gráfico dos escores obtidos pela PCA aplicada aos histogramas do grupo I (●: BRS 200 Marrom, ■: BRS Topázio). (a) todos os canais; b) tons de cinza; c) R, (d) G, (e) B, (f) RGB, (g) H, (h) S, (i) V, (j) HSV, (l) Cinza-RGB, (m) Cinza-HSV, (n) RGB-HSV. A variância explicada para cada PC é encontrada entre parêntese.

Para o grupo I, é possível perceber uma tendência de separação das cultivares analisadas ao longo de PC1 (**Figura 4.4 a-n**). De fato, a maioria das amostras pertencentes à cultivar BRS 200 Marrom encontra-se localizada com valores de escores mais positivos em PC1, quando comparadas com aquelas pertencentes à cultivar BRS Topázio. Contudo, uma sobreposição de algumas amostras é, ainda, aparente nos diferentes casos. Isto é particularmente verdade quando se avalia os escores de PC1 no canal H (**Figura 4.4 g**). Esta evidência corrobora com a análise realizada no domínio dos histogramas apresentados na **Figura 4.2**.

Para avaliar a contribuição de cada variável, o gráfico dos pesos para o grupo I com a combinação de todos os canais (**Figura 4.5**), revela que praticamente todos os componentes de cor contribuem de forma semelhante para a distribuição de escores ao longo de PC1 e PC2.

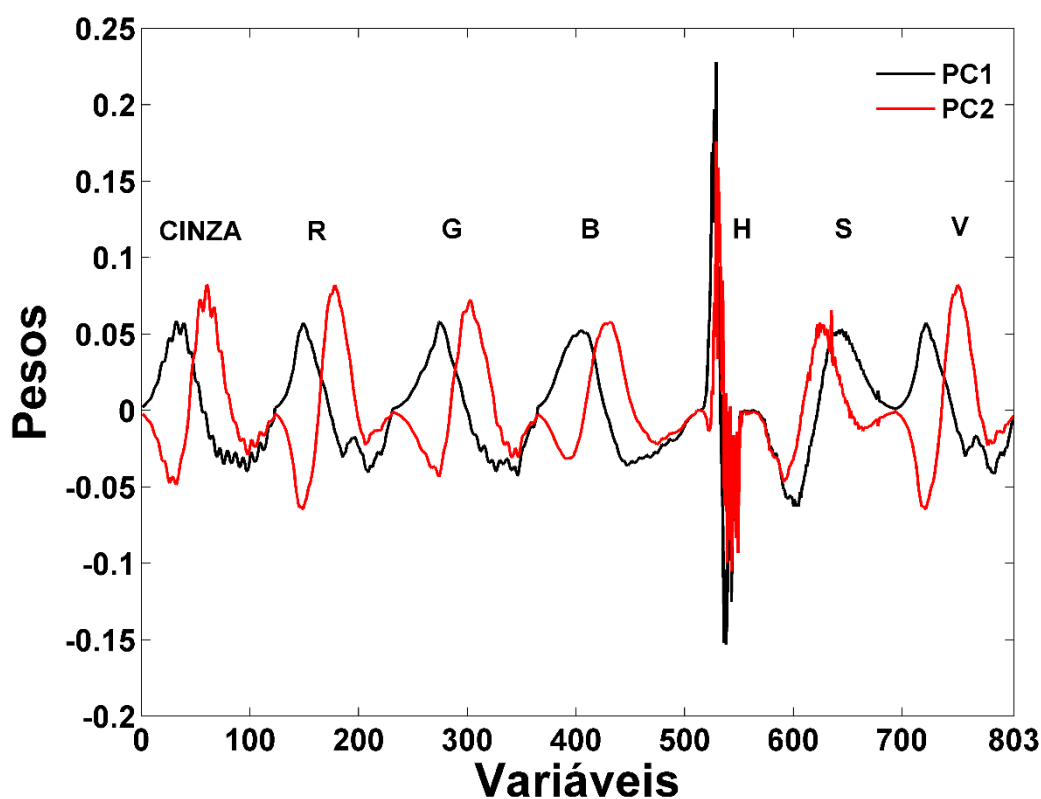


Figura 4.5: Gráfico dos pesos das variáveis em cada componente de cor para o grupo I. R: Red, G: Green, B: Blue, H: Hue, S: Saturação e V: Value.

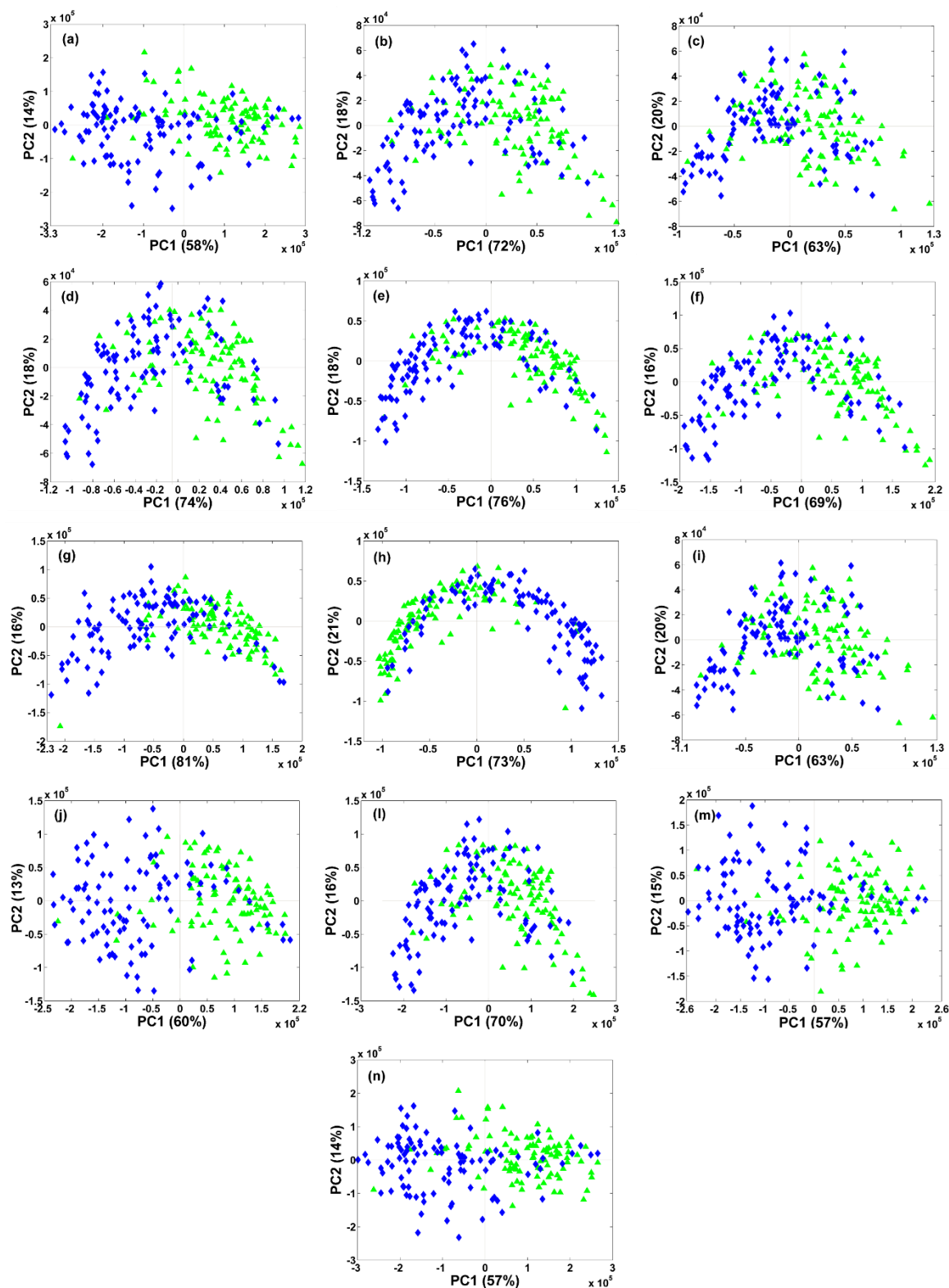


Figura 4.6: Gráfico dos escores obtidos pela PCA aplicada aos histogramas do grupo II ($\blacktriangle$: BRS Rubi, $\blacklozenge$: BRS Safira). (a) todos os canais; b) tons de cinza; c) R, (d) G, (e) B, (f) RGB, (g) H, (h) S, (i) V, (j) HSV, (l) Cinza-RGB, (m) Cinza-HSV, (n) RGB-HSV. A variância explicada para cada PC é encontrada entre parêntese.

Na **Figura 4.6**, é possível observar um comportamento similar aos gráficos dos escores do grupo I e uma sobreposição das amostras menos acentuada quando são empregados todos os canais, HSV, Cinza-HSV ou RGB-HSV. Assim como o gráfico dos pesos, para o grupo II (**Figura 4.7**), com a combinações todos os canais, também evidencia resultados similares ao grupo I.

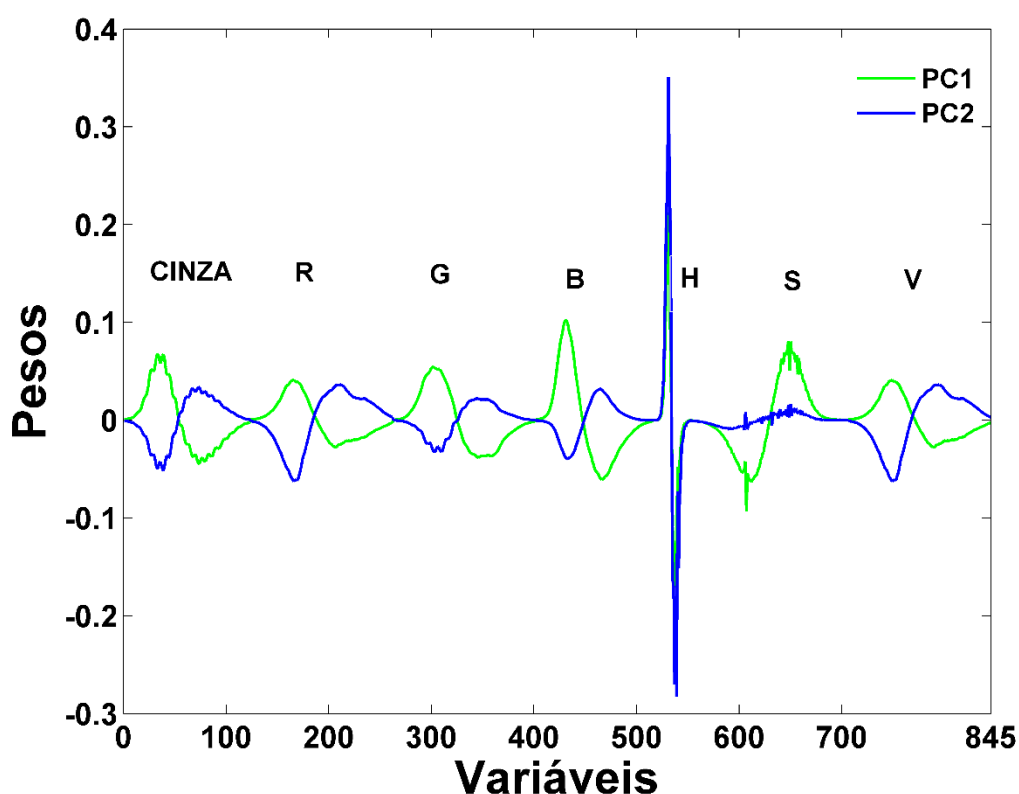


Figura 4.7: Gráfico dos pesos das variáveis em cada componente de cor para o grupo II. R: Red, G: Green, B: Blue, H: Hue, S: Saturação e V: Value

Apesar da PCA evidenciar um desempenho razoável no que diz respeito à discriminação das cultivares estudadas (no grupo I e II), técnicas de reconhecimento de padrões supervisionadas com seleção de variáveis são necessárias, tanto para construir modelos que sejam capazes de classificar amostras desconhecidas, como para melhorar a discriminação dessas amostras.

4.1.3 Modelos de Classificação

Modelos de classificação baseados na LDA/SPA, LDA/SW e PLS-DA foram desenvolvidos para cada canal de cor, bem como na combinação entre eles. A **Tabela 4.1** fornece os resultados de classificação obtidos pelos diferentes modelos para os

grupos I e II. Os resultados são apresentados em termos da TCC para o subconjunto de teste.

Os melhores resultados foram obtidos pelo modelo LDA/SW utilizando a combinação entre os canais HSV (para ambos os grupos). Um desempenho superior dos outros métodos (LDA/SPA e PLS/DA) também foi alcançado empregando o sistema HSV, quando o grupo II é avaliado. Para o grupo I, contudo, o LDA/SPA alcançou o melhor resultado utilizando todos os canais estudados, enquanto que o PLS-DA alcançou uma TCC superior quando o canal verde ou valor (brilho) é empregado.

Tabela 4.1: Taxa de classificação correta (TCC) obtidos com os modelos LDA/SPA, LDA/SW e PLS-DA para o subconjunto de teste dos Grupo I e II.

Taxa de Classificação Correta (%)						
Canais	Grupo I			Grupo II		
	LDA/SPA	LDA/SW	PLS-DA	LDA/SPA	LDA/SW	PLS-DA
Cinza-RGB-HSV	92	90	86	88	88	82
Cinza	84	92	90	80	82	88
Red	82	88	88	76	72	74
Green	88	94	92	84	84	86
Blue	88	94	88	74	80	86
RGB	86	82	88	88	92	86
Hue	82	86	74	82	88	86
Saturação	78	84	84	84	88	88
Valor	70	86	92	76	72	70
HSV	82	96	78	90	96	88
Cinza-RGB	86	88	86	82	92	86
Cinza-HSV	74	88	82	82	94	80
RGB-HSV	84	92	83,5	90	88	84

Os resultados detalhados dos canais escolhidos para cada modelo (LDA ou PLS-DA) serão apresentados nas próximas seções.

4.1.3.1 Modelo LDA

As **Figuras 4.8** apresentam os gráficos das variáveis selecionadas utilizadas na construção dos modelos LDA/SPA e LDA/SW para os canais de cores com melhores resultados destacados na **Tabela 4.1**.

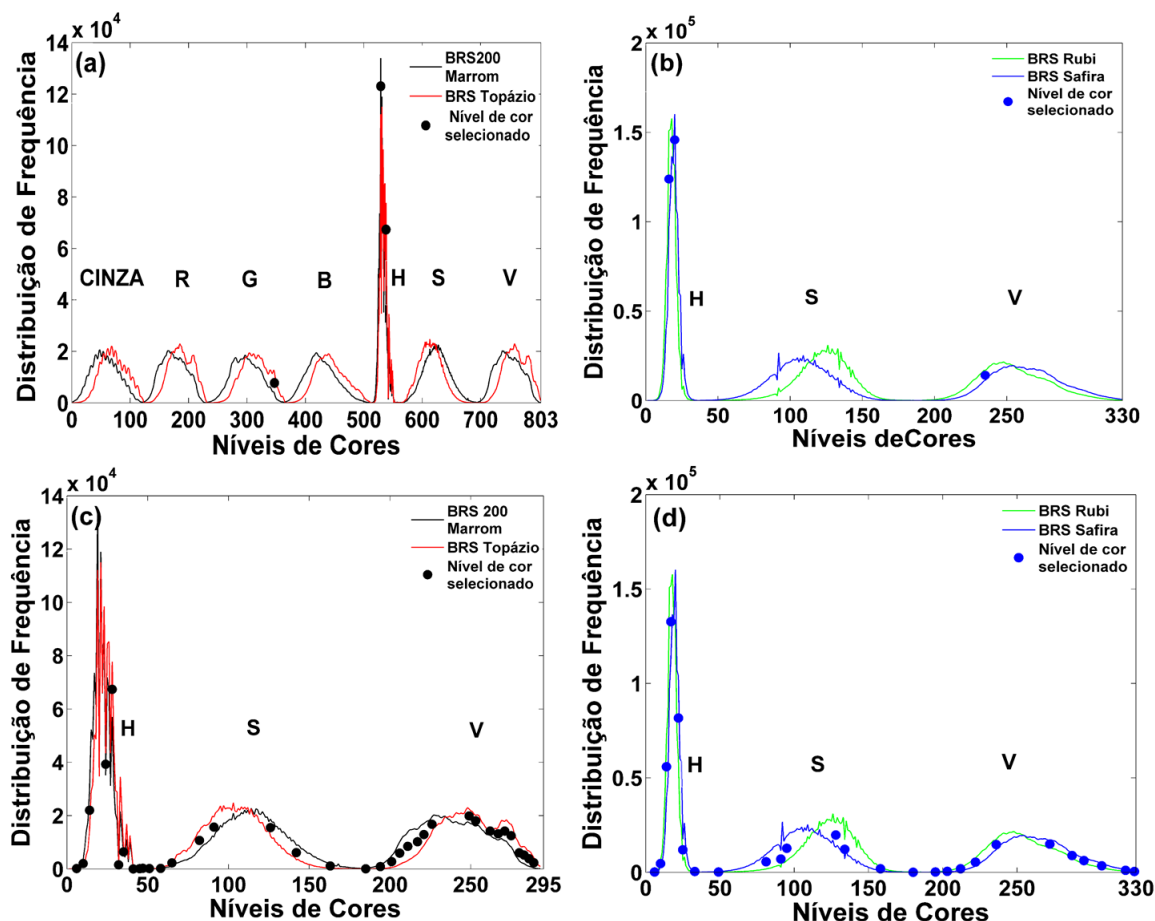


Figura 4.8: Média dos histogramas para cada cultivar, com as Variáveis selecionadas (●, ●) pelo LDA/SPA (a e b) e pelo LDA/SW (c e d) para o Grupo I (a e c) e Grupo II (b e d).

É possível perceber que, tanto para o grupo I como para o grupo II, os modelos LDA/SPA selecionaram apenas três variáveis, resultando em 9 e 8 erros de validação, respectivamente. Quando aplicados aos conjuntos de teste (validação externa), os modelos desenvolvidos com essas variáveis alcançaram uma taxa de classificação correta (TCC) de 92% e 90% para os grupos I e II, respectivamente.

Já para os modelos LDA/SW, um número bem maior de variáveis foi selecionado. Mais especificamente, 36 (grupo I) e 26 (grupo II), varrendo toda a faixa do histograma (**Figura. 4.8c-d**). Essas variáveis foram selecionadas com base num valor de coeficiente de correlação múltipla (*threshold*) de 0,95, resultando em 13 e 7

erros de validação para o grupo I e II, respectivamente. Os modelos desenvolvidos com essas variáveis alcançaram uma TCC de 96% para ambos os grupos.

O bom desempenho dos modelos de classificação é também evidenciado com o gráfico dos escores obtidos pela LDA (com prévia seleção de variáveis pelo SPA ou SW) aplicada ao conjunto completo de amostra (**Figura 4.9**). Os perfis dos gráficos de escores obtidos tanto pelo LDA/SPA (**Figuras 4.9a-b**) como pelo LDA/SW (**Figuras 4.9c-d**) são muito parecidos. Contudo, quando comparados com os resultados obtidos pela PCA (**Figuras 4.4 e 4.6**), a discriminação das duas cultivares em cada grupo é muito mais aparente.

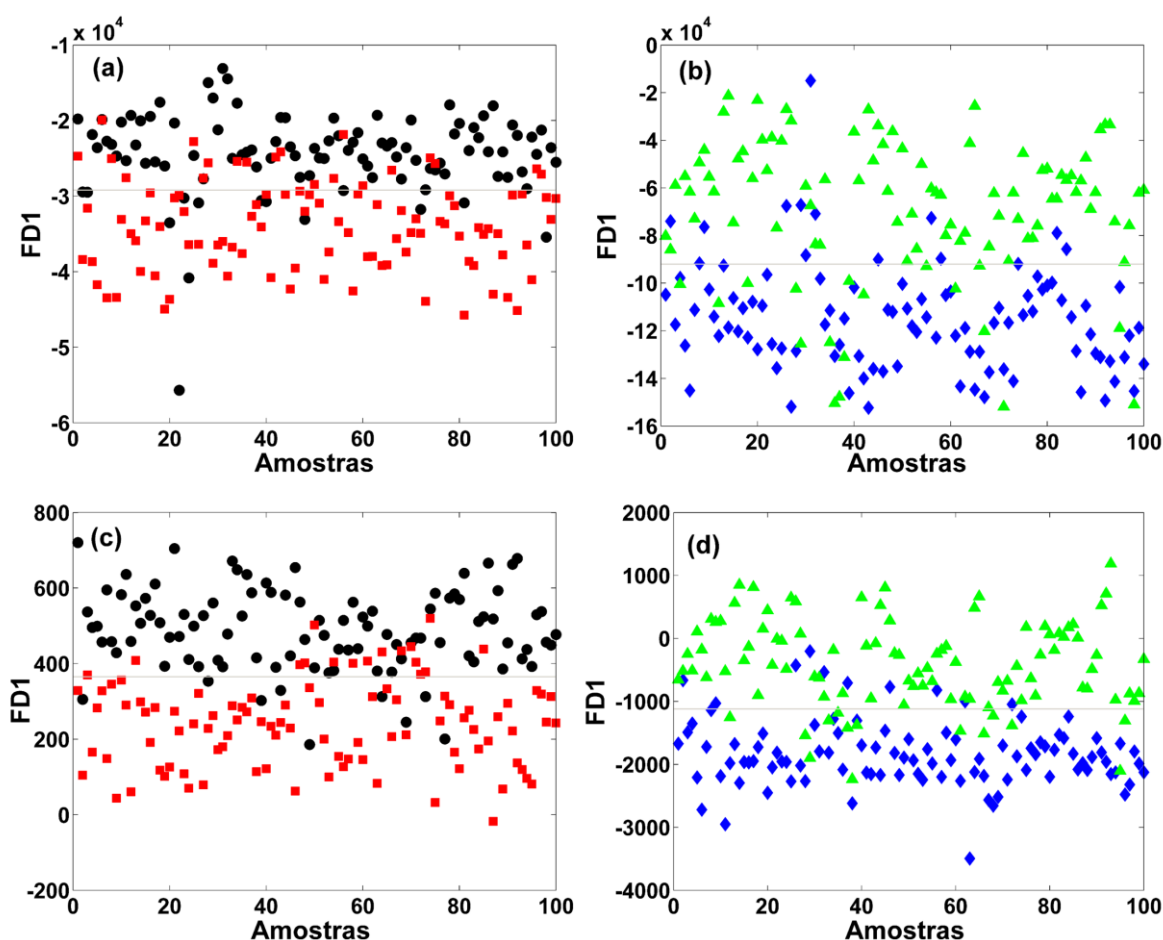


Figura 4.9: Gráfico do escores obtido por LDA usando algoritmos de seleção de variáveis SPA (a e b) e SW (c e d), (a e c) Grupo I (b e d) Grupo II. ●: BRS 200 Marrom, ■: BRS Topázio, ▲: BRS Rubi e ◆: BRS Safira.

4.1.3.2 Modelo PLS-DA

O número ótimo de fatores usado na construção do modelo PLS-DA foi obtido de acordo com o menor erro de validação cruzada. A **Figuras 4.10** mostra o número de fatores escolhidos para os resultados mais promissores apresentado na **Tabela**

4.1 com o canal de cor valor (V) e a combinação HSV para o grupo I e II respectivamente.

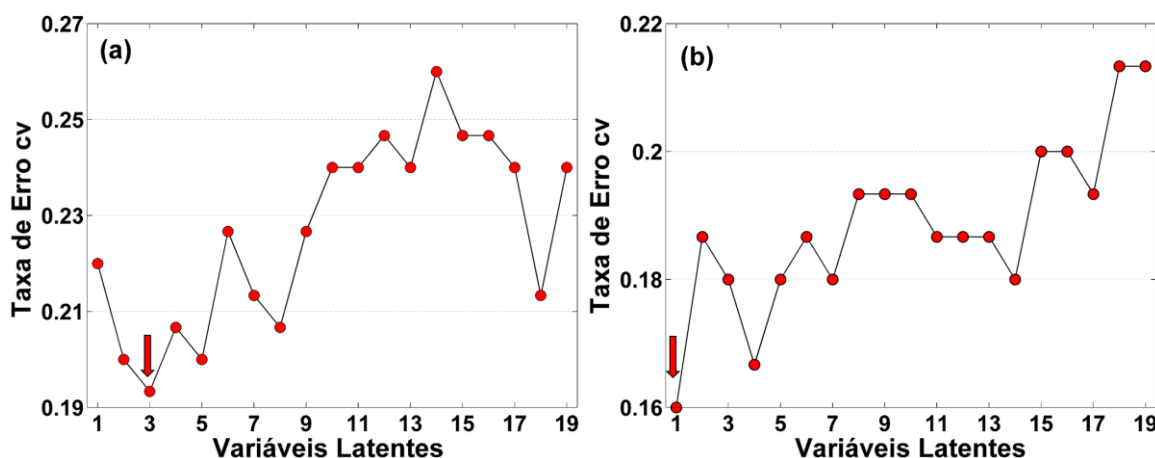


Figura 4.10: Taxa de erro de validação cruzada *versus* o número de variáveis latentes (a) Grupo I; (b) Grupo II. As setas identificam para o número de variáveis latentes selecionadas.

Como observado nas **Figuras 4.10a-b**, três variáveis latentes foram selecionadas com uma taxa de erro de validação cruzada de aproximadamente 19%, quando o grupo I é avaliado. Já para o grupo II, apenas uma variável latente foi escolhida com base numa taxa de 16%.

Quando aplicado ao conjunto de teste (externo), o modelo PLS-DA apresenta uma taxa de classificação correta igual ao modelo LDA/SPA (92%) e um pouco abaixo do modelo LDA/SW quando o grupo I é avaliado. Para o grupo II, o desempenho do PLS-DA é inferior aos outros métodos, resultando em uma TCC de apenas 88%.

A **Tabela 4.2** apresenta os resultados de classificação detalhados obtidos pelos modelos LDA/SPA, LDA/SW e PLS-DA aplicados aos conjuntos de teste.

Os valores apresentados na **Tabela 4.2** expressam tanto a classificação correta (classe predita igual à classe verdadeira), como a classificação incorreta (classe predita diferente da classe verdadeira). Podemos entendê-lo da seguinte forma: das vinte e cinco amostras de BRS 200 Marrom pertencentes ao subconjunto de teste, vinte e três foram corretamente classificadas como pertencente à classe BRS 200 Marrom e duas amostras foram incorretamente classificadas como BRS Topázio quando o modelo LDA/SPA é empregado.

Tabela 4.2: Resultados da classificação para o subconjunto de teste obtidos com modelos LDA/SPA, LDA/SW e PLS-DA, para o grupo I e II. A taxa de classificação correta obtida por cada modelo é indicada em parênteses. N é o número de amostras preditas empregadas neste estudo.

Grupo I		LDA-SPA (92%)		LDA-SW ^a (96%)		PLS-DA (92%)	
		Classe Predita		Classe Predita		Classe Predita	
Classe Verdadeira	N	BRS 200 Marrom	BRS Topázio	BRS 200 Marrom	BRS Topázio	BRS 200 Marrom	BRS Topázio
BRS 200 Marrom	25	23	02	24	01	24	01
BRS Topázio	25	02	23	01	24	03	22

Grupo II		LDA-SPA (90%)		LDA-SW ^a (96%)		PLS-DA (88%)	
		Classe Predita		Classe Predita		Classe Predita	
Classe Verdadeira	N	BRS Rubi	BRS Safira	BRS Rubi	BRS Safira	BRS Rubi	BRS Safira
BRS Rubi	25	23	02	25	–	24	01
BRS Safira	25	03	22	02	23	05	20

^a 0,95 (Grupo I e Grupo II).



Como apresentado na **Tabela 4.2**, os melhores resultados foram alcançados com o LDA/SW, resultando em uma classificação incorreta de apenas uma amostra para cada tipo de cultivar. Para o grupo II, os erros mais frequentes ocorrem com as amostras pertencentes a cultivar BRS Safira, com três, duas e cinco amostras do total de 25 amostras de teste sendo incorretamente classificadas como BRS Rubi para os modelos LDA/SPA, LDA/SW e PLS/DA, respectivamente.

4.2. Estudo de Calibração

4.2.1 Histogramas

A **Figura 4.11** apresenta os histogramas com sete canais de cores: tons cinza, vermelho, verde, azul, matiz, saturação e valor, obtidos das imagens digitais para todas as amostras de plumas de algodão empregadas neste estudo. Assim como no estudo anterior, os histogramas apresentam um total de 1792 níveis de cores.

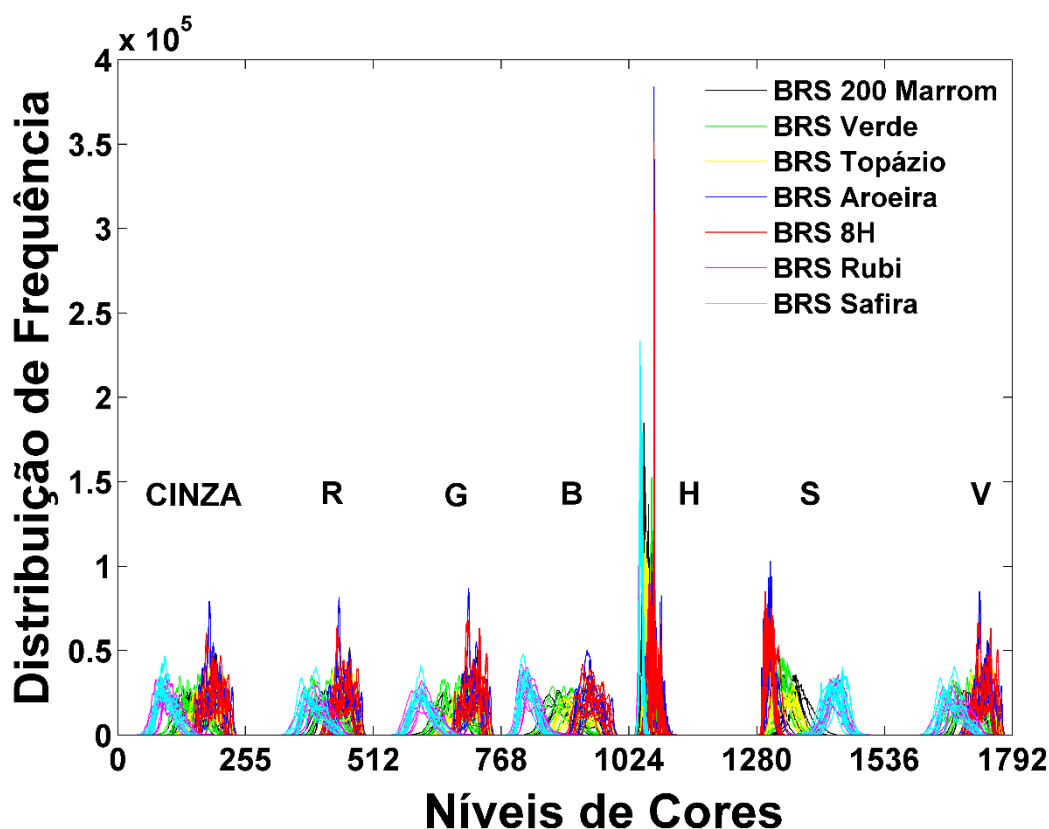


Figura. 4.11: Histogramas completos, das plumas de algodão. R: Red, G: Green, B: Blue, H: Hue, S: Saturation e V: Valor.

Devido à semelhança entre as diferentes cultivares analisadas neste estudo, é possível perceber, na **Figura 4.11**, que as cultivares brancas (BRS Aroeira e BRS 8H) estão se sobrepondo, assim como as cultivares BRS 200 Marrom, BRS Topázio e BRS Verde. Um perfil parecido entre os níveis de cores é encontrado entre as cultivares BRS Rubi e BRS Safira. As similaridades e diferenças entre os histogramas das diferentes classes estão em consonância com as cores das diferentes cultivares apresentadas na **Figura 3.2**. É importante ressaltar que, na matriz original, existiam algumas variáveis não informativas (com níveis de cor com frequência igual a zero) que foram removidas do conjunto de dados.

4.2.2 Modelos de Calibração Multivariada

Os modelos de calibração (PLS e MLR-SPA) foram construídos relacionando os histogramas com cada um dos seguintes parâmetros: grau de amarelamento (+b), reflectância (Rd) e teor de cera (WAX) das fibras de algodão, obtidos pelos métodos de referência. Os modelos foram construídos para cada canal, bem como para a combinação entre eles.

As **Tabelas 4.3-4.5** apresentam os resultados obtidos pelos modelos PLS e MLR-SPA para o conjunto de calibração e predição na determinação de + b, Rd e WAX. Um melhor desempenho dos modelos foi alcançado para grau de amarelamento e teor de cera, quando o algoritmo SPXY é empregado para selecionar subconjuntos de amostras de calibração e previsão. Já para reflectância, a aplicação do algoritmo KS levou a um resultado mais satisfatório.

Tabela 4.3: Resultados obtidos para os modelos PLS e MLR-SPA para o grau de amarelamento em cada canal (e combinação entre eles). Os valores entre parênteses correspondem ao número de fatores ou variáveis empregados nos modelos PLS e MLR-SPA, respectivamente. A faixa de calibração do parâmetro é apresentada entre colchetes, cv. (validação cruzada) e prev. (previsão).

Grau de Amarelamento (+b) [8,58-20,70]									
PLS					MLR-SPA				
Canais	r_{cv}	RMSECV	r_{prev}	RMSEP	r_{cv}	RMSECV	r_{prev}	RMSEP	
Cinza-RGB-HSV	0,96	1,36 (5)	0,97	0,88	0,98	0,94 (6)	0,97	0,95	
Cinza	0,94	1,49 (6)	0,97	0,94	0,95	1,35 (10)	0,95	1,21	
R	0,74	3,22 (13)	0,90	1,95	0,65	3,52 (4)	0,82	2,46	
G	0,83	2,58 (8)	0,88	1,75	0,79	2,76 (4)	0,68	2,84	
B	0,94	1,62 (14)	0,94	1,46	0,91	1,83 (4)	0,94	1,61	
RGB	0,90	1,97 (3)	0,95	1,27	0,89	2,07 (4)	0,92	1,51	
H	0,96	1,18 (5)	0,98	0,86	0,96	1,23 (4)	0,99	0,76	
S	0,97	1,16 (4)	0,97	0,83	0,97	1,09 (7)	0,98	0,81	
V	0,93	1,61 (7)	0,95	1,28	0,91	1,85 (8)	0,91	1,58	
HSV	0,96	1,31 (4)	0,98	0,81	0,97	1,10 (11)	0,97	1,02	
Cinza-RGB	0,91	1,85 (4)	0,96	1,20	0,92	1,74 (5)	0,94	1,40	
Cinza-HSV	0,96	1,31 (5)	0,98	0,83	0,96	1,23 (7)	0,96	1,12	
RGB-HSV	0,94	1,51 (9)	0,98	0,84	0,97	1,05 (12)	0,97	1,06	

Com base nos resultados apresentados na **Tabela 4.3**, é possível perceber que, quando o PLS é aplicado aos canais saturação (S), a combinação HSV ou Cinza-HSV, desempenhos satisfatórios e parecidos são encontrados, com baixos valores de RMSEP. Contudo, o canal S foi escolhido por apresentar tanto um número menor de fatores, como um valor mais baixo RMSECV. O modelo MLR-SPA apresentou um

melhor desempenho quando aplicado ao canal H. Os resultados detalhados do PLS e MLR-SPA aplicados ao canal S e H, respectivamente, para a determinação do grau de amarelamento serão apresentados posteriormente.

Tabela 4.4: Resultados obtidos para os modelos PLS e MLR-SPA para reflectância em cada canal (e combinação entre eles). Os valores entre parênteses correspondem ao número de fatores ou variáveis empregados nos modelos PLS e MLR-SPA, respectivamente. A faixa de calibração do parâmetro é apresentada entre colchetes, cv. (validação cruzada) e prev. (previsão).

Reflectância (Rd) [37,5-80,6]								
PLS					MLR-SPA			
Canais	r_{cv}	RMSECV	r_{prev}	RMSEP	r_{cv}	RMSECV	r_{prev}	RMSEP
Cinza-RGB-HSV	0,97	3,89 (2)	0,97	3,32	0,98	2,91 (6)	0,96	3,62
Cinza	0,93	5,17 (7)	0,96	4,29	0,94	4,90 (11)	0,85	8,08
R	0,81	9,68 (16)	0,87	7,65	0,60	12,08 (2)	0,55	13,51
G	0,88	7,14 (11)	0,87	7,08	0,85	7,62 (1)	0,74	12,04
B	0,93	5,38 (3)	0,91	5,71	0,94	5,07 (6)	0,93	5,56
RGB	0,93	5,49 (3)	0,97	3,57	0,92	5,98 (4)	0,93	5,32
H	0,98	3,36 (3)	0,99	2,49	0,99	2,58 (6)	0,99	2,53
S	0,98	3,12 (2)	0,97	3,44	0,98	2,99 (5)	0,97	3,38
V	0,90	6,49 (6)	0,95	6,00	0,87	7,33 (8)	0,92	6,52
HSV	0,98	2,87 (3)	0,99	2,81	0,98	2,92 (4)	0,97	3,65
Cinza-RGB	0,93	5,19 (4)	0,97	3,47	0,91	6,09 (7)	0,91	5,62
Cinza-HSV	0,98	3,21 (3)	0,98	2,81	0,98	2,78 (7)	0,97	3,50
RGB-HSV	0,97	3,32 (3)	0,98	2,92	0,99	2,48 (6)	0,95	4,17

De acordo com a **Tabela 4.4**, um valor mais baixo de RMSEP foi alcançado utilizando o canal H para ambos os modelos. Os mesmos serão detalhados mais adiante.

Tabela 4.5: Resultados obtidos para os modelos PLS e MLR-SPA para o teor de cera em cada canal (e combinação entre eles). Os valores entre parênteses correspondem ao número de fatores ou variáveis empregados nos modelos PLS e MLR-SPA, respectivamente. A faixa de calibração do parâmetro é apresentada entre colchetes, cv. (validação cruzada) e prev. (previsão).

Teor de cera (WAX) [0,45-7,18]									
PLS					MLR-SPA				
Canais	r_{cv}	RMSECV	r_{prev}	RMSEP	r_{cv}	RMSECV	r_{prev}	RMSEP	
Cinza-RGB-HSV	0,96	0,61 (3)	0,96	0,33	0,97	0,56 (13)	0,97	0,35	
Cinza	0,50	1,96 (1)	0,13	2,27	0,58	1,84 (1)	0,14	1,54	
R	0,31	2,28(1)	0,46	2,10	0,36	2,22 (1)	0,16	1,93	
G	0,72	1,65 (9)	0,85	0,78	0,65	1,71 (1)	0,39	1,22	
B	0,46	2,02 (1)	0,34	1,63	0,56	1,88 (1)	0,49	1,19	
RGB	0,69	1,68 (5)	0,57	1,01	0,76	1,48 (6)	0,68	0,88	
H	0,94	0,78 (6)	0,96	0,33	0,95	0,71 (8)	0,89	0,50	
S	0,84	1,24 (8)	0,91	0,48	0,88	1,09 (2)	0,87	0,55	
V	0,75	1,69 (15)	0,09	1,36	0,62	1,87 (2)	0,21	0,88	
HSV	0,97	0,53 (4)	0,98	0,26	0,96	0,66 (10)	0,95	0,38	
Cinza-RGB	0,67	1,71 (4)	0,55	0,96	0,76	1,49 (8)	0,31	1,14	
Cinza-HSV	0,97	0,52 (4)	0,97	0,30	0,97	0,60 (10)	0,96	0,33	
RGB-HSV	0,97	0,59 (3)	0,96	0,33	0,97	0,57 (12)	0,97	0,41	

Para determinação do teor de cera, resultados semelhantes foram encontrados pelos modelos PLS quando a combinação HSV e Cinza-HSV são avaliadas. Quanto ao desempenho do modelo MLR-SPA, a combinação Cinza-HSV parece ser mais adequada.

A **Figura 4.12** mostra os gráficos com os números de fatores utilizados na construção dos modelos PLS (histograma-completo) *versus* a variância explicada em *y* para os melhores resultados dos três parâmetros. Podemos observar que para o grau de amarelamento e teor de cera, quatro fatores foram necessários (**Figura 4.12a,c**), enquanto que para reflectância, três variáveis latentes foram adotadas (**Figura 4.12b**).

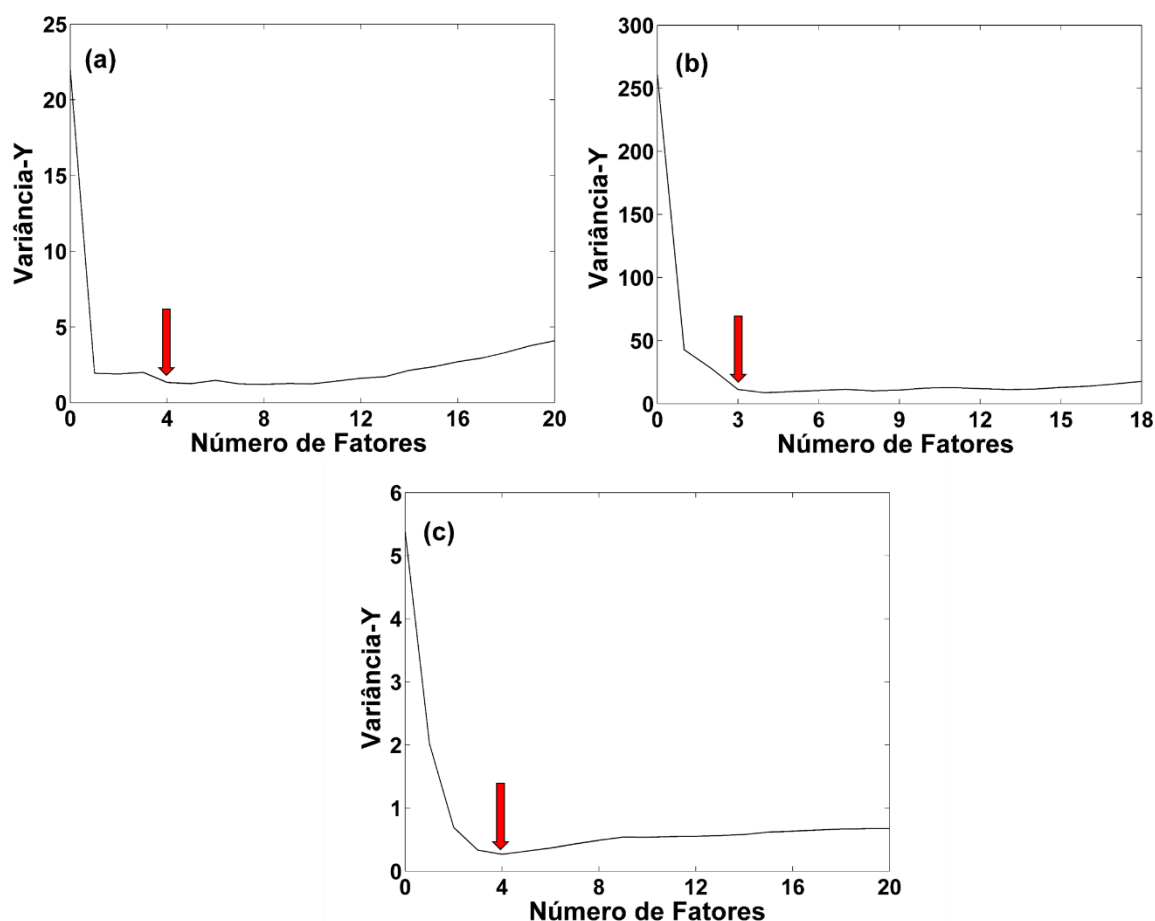


Figura 4.12: Gráfico da variância em *y* *versus* o número de fatores dos modelos PLS, para (a) +b, (b) Rd e (c) WAX.

As variáveis selecionadas pelos modelos MLR-SPA para a determinação dos três parâmetros são apresentadas na **Figura 4.13**. O modelo selecionou 4, 6 e 10 variáveis para os parâmetros +b, Rd e WAX, respectivamente. Tais variáveis são apresentadas nos histogramas médios (**Figura 4.13 b,d,f**).

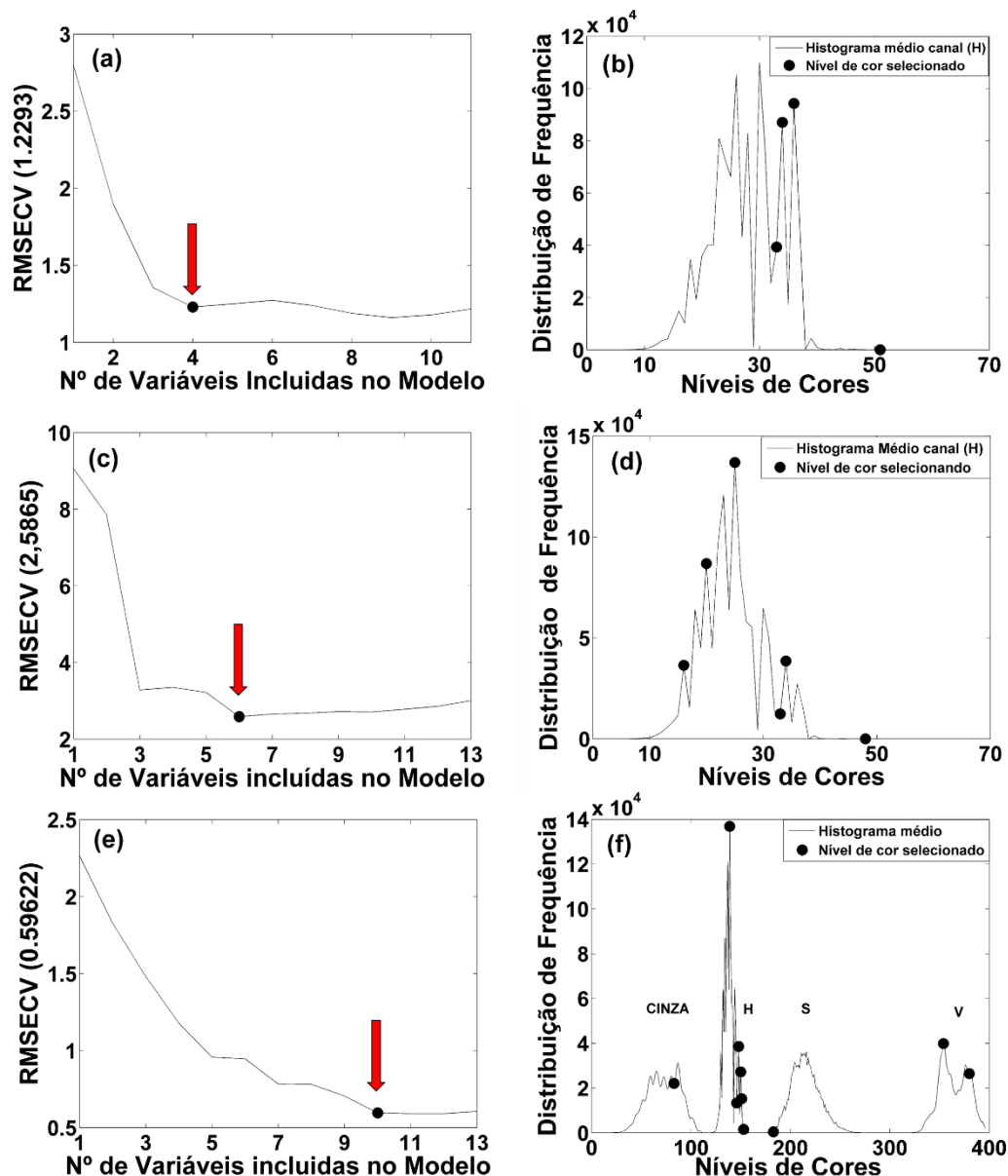


Figura 4.13: Gráficos das variáveis selecionadas *versus* RMSECV para MLR-SPA (a) +b, (c) Rd e (e) WAX. Histograma médio com as variáveis selecionadas pelo SPA (b) +b, (d) Rd e (f) WAX. As setas identificam o número de variáveis selecionadas.

Para avaliação e comparação dos modelos, a **Tabela 4.6** resume os resultados com as componentes de cores escolhidas no estudo inicial obtidos pelos modelos PLS (histograma-completo) e MLR-SPA, para o conjunto de previsão. As faixas de variação observada nos conjuntos de calibração foram 8,58-20,70 (+b); 37,5–80,6 % (Rd) e 0,45–7,18 % (WAX).

Tabela. 4.6: Resultados obtidos pelos modelos PLS e MLR para o conjunto de previsão (em termos de RMSEP, valores de correlação e de determinação). N é o número de amostras de previsão.

Parâmetro	PLS (hist.- completo)					MLR-SPA		
	N	Repetitividade	RMSEP	r_{pred}	r^2	RMSEP	r_{pred}	r^2
+b	30	0,41	0,83	0,97	0.95	0,76	0,99	0,97
Rd	30	2,19	2,49	0,99	0.97	2,53	0,99	0,97
WAX	25	0,30	0,30	0,97	0.93	0,33	0,96	0,91

Para todos os parâmetros avaliados, resultados de previsão satisfatórios foram obtidos, com valores de RMSEP baixos e próximos da repetitividade do método de referência. A repetitividade foi determinada de acordo com a referência [104].

Em termos de desempenho de previsão entre os dois métodos avaliados, o uso de seleção de variáveis pelo SPA não teve impacto significativo entre os resultados. De fato, um *Teste F* (**Tabela 4.7**), com um nível de confiança de 95% não indicou diferenças estatisticamente significativas entre os valores de RMSEP obtidos com o modelo PLS (histograma-completo) e MLR-SPA.

Tabela 4.7: Resultados para o teste *F* (nível de confiança de 95%) obtidos na comparação dos modelos (PLS e MLR-SPA).

Parâmetro	$F_{calculado}$	$F_{crítico}$
Grau de amarelamento (+b)	1,193	1,861
Reflectância (Rd)	1,032	1,861
Teor de cera (WAX)	1,210	1,984

A concordância entre valores previstos e de referência de + b, Rd e WAX são apresentados na **Figura 4.14**. Como pode ser observado, não há erro sistemático nas previsões, uma vez que os pontos estão distribuídos aleatoriamente ao longo da linha bissetriz em todo intervalo dos valores de y. Além disso, um teste t-pareado (**Tabela 4.8**) para o conjunto de previsão não indicam diferenças estatisticamente significativas entre os valores previstos e de referência a nível de confiança de 95%.

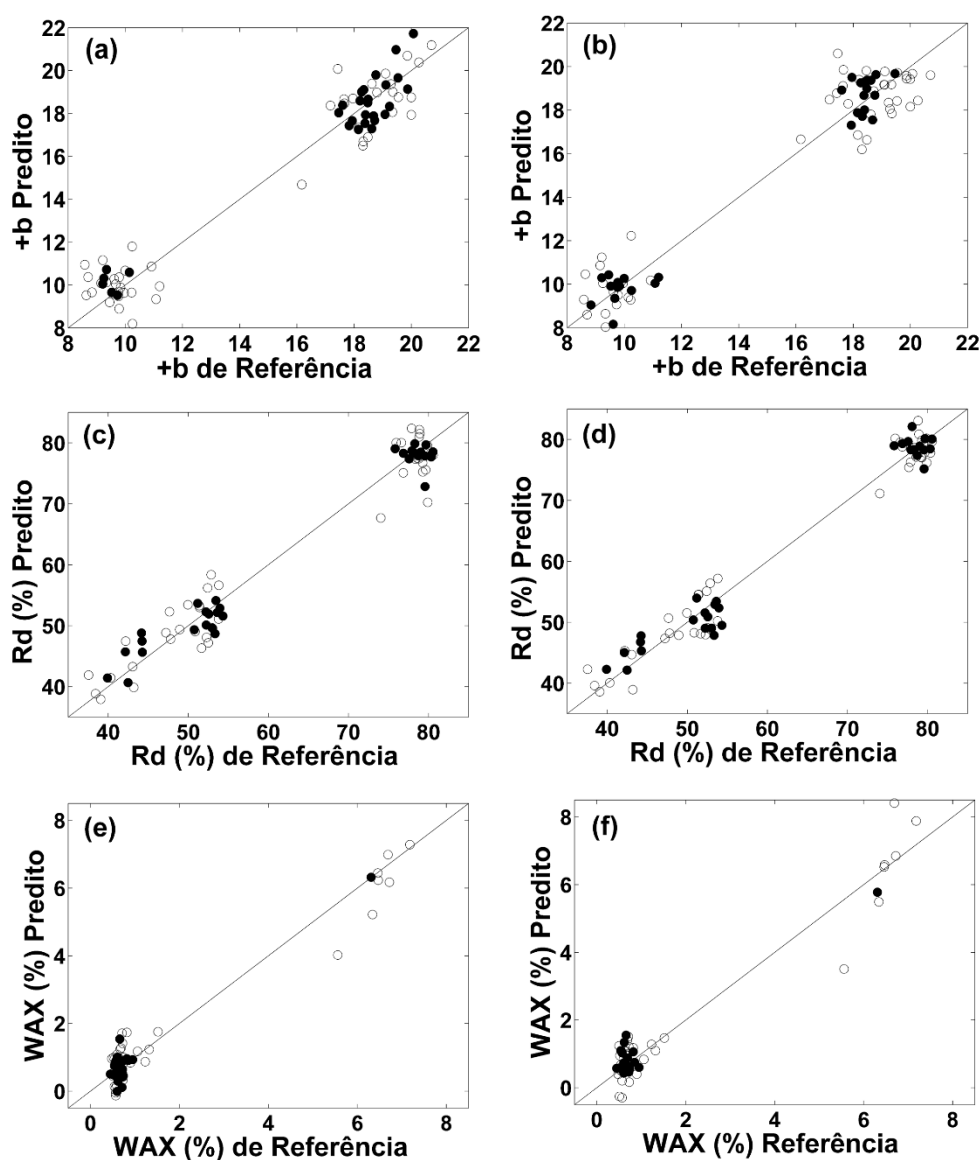


Figura 4.14: Gráfico dos valores preditos *versus* os valores de referências obtidos pelos modelos PLS (a,c,e) e MLR-SPA (b,d,f) para +b (a-b), Rd (c-d) e WAX (e-f). Amostras de calibração (○) e amostra de predição (●).

Tabela 4.8: Resultados para o teste t-pareado (nível de confiança de 95%) obtidos na comparação dos valores de referências e valores previstos pelos modelos PLS e MLR-SPA.

Parâmetro	t _{calculado} (PLS)	t _{calculado} (MLR)	t _{crítico}
Grau de amarelamento (+b)	0,649	1,097	1,699
Reflectância (Rd)	0,662	0,326	1,699
Teor de cera (WAX)	0,279	0,456	1,711



CAPÍTULO V

CONCLUSÃO

5 CONCLUSÕES

Este trabalho apresentou duas metodologias analíticas baseadas em imagens digitais e análise multivariada para (1) discriminação/classificação de duas cultivares de cada grupo de algodão naturalmente colorido (BRS 200 Marrom, BRS Topázio, BRS Rubi e BRS Safira); (2) Determinação de três importantes parâmetros de qualidade em amostras de fibras de algodão branco e naturalmente colorido: grau de amarelamento (+b), reflectância (Rd) e teor de cera (WAX).

Uma avaliação exploratória dos dados realizada pela PCA revelou a presença de padrões de amostra para cada uma das cultivares estudadas. Contudo, uma sobreposição de algumas amostras foi aparente, motivando o uso de técnicas supervisionadas com seleção de variáveis.

O estudo comparativo entre os métodos de classificação LDA-SPA, LDA-SW e PLS-DA permitiu o LDA-SW como mais eficiente na classificação de amostras de algodão naturalmente colorido, fornecendo uma TCC de 96% para o conjunto de previsão, quando o sistema HSV é empregado.

No estudo envolvendo a determinação simultânea de +b, Rd e WAX, modelos de calibração multivariada (PLS e MLR-SPA) foram desenvolvidos, validados e comparados. Resultados de previsão apropriados foram obtidos independentemente do modelo, com baixos valores de RMSEP e próximos à repetitividade do método de referência. Nenhum erro sistemático foi observado e não foram encontradas diferenças significativas entre os valores previstos e de referência, de acordo com um teste t-pareado ao nível de confiança de 95%.

Como vantagens, as metodologias desenvolvidas neste trabalho são simples, mantém a integridade das amostras, não utilizam métodos laboriosos e possuem um custo de instrumentação relativamente baixo. Além disso, as análises são realizadas em curtos intervalos de tempo, com resultados comparáveis aqueles obtidos pelos métodos de referência.

Finalmente, os estudos desenvolvidos neste trabalho podem auxiliar:

- ✓ No combate à falsificação de amostras tingidas;
- ✓ Na avaliação direta da cor da pluma por meio da determinação simultânea de +b e Rd. De fato, uma alteração nesses parâmetros pode interferir na cor natural do algodão, desvalorizando a fibra;

- ✓ Na indústria têxtil e melhoramento genético, com ferramentas, para selecionar fibras com baixo teor de cera, bem como no desenvolvimento de novas cultivares de algodão;
- ✓ Na etapa do tingimento. Como um alto teor de cera pode acarretar uma barreira à absorção dos corantes, a identificação prévia do teor de cera presente em cada tipo de pluma poderá ser útil nesse processo;
- ✓ Na classificação de outras cultivares, com a incorporação de um novo banco de dados e construção de novos modelos.

5.1 Perspectivas

Aplicar os modelos de classificação previamente construídos a amostras de plumas de algodão falsificadas (artificialmente coloridas);

Expandir os estudos para determinação do teor de celulose e pigmentação (flavonoides) determinados por métodos clássicos, empregando calibração multivariada, correlacionando com os dados adquiridos por meio das imagens digitais;

Avaliar estratégias de transferência de classificação e calibração para a obtenção de modelos desenvolvidos com outros dispositivos, como *smartphones* ou *scanners*).

Utilizar os algoritmos COW e ICOSHIFT, entre outros, no alinhamento dos histogramas no eixo x, para fazer uma análise dos algoritmos de classificação e de calibração multivariada



CAPÍTULO VI

REFERÊNCIAS

REFERÊNCIAS

- [1] ULLOA, M.; BRUBAKER, C.; CHEE, P. Cotton, em: KOLE, C. (Ed.), **Technical Crops: Genome Mapping and Molecular Breeding in Plants**. Berlin Heidelberg: Springer-Verlag, 2010, pp. 1-50.
- [2] DUTT, Y.; WANG, X. D.; ZHU, Y. G.; LI, Y. Y. Breeding for high yield and fibre quality in coloured cotton. *Plant Breeding*, **123:145**, 2004.
- [3] BELTRÃO, N. E. M.; AZEVEDO, D. M. P., Eds. **O agronegócio do algodão no Brasil**. Brasília: Embrapa Informação Tecnológica, v.1, 2 ed. 2008.
- [4] ABRAPA. Associação Brasileira dos produtores de Algodão. **Relatório de Gestão**, Biênio 2011/2012. Disponível em:
<http://www.abrapa.com.br/biblioteca/Paginas/biblioteca-institucional.aspx> . Acessado em 10 de janeiro de 2015.
- [5] FREIRE, E. C. Algodão colorido. *Biotecnologia, Ciência & desenvolvimento*, **2: 36**, 1999.
- [6] MURTHY MSS. Never Say Dye: the story of Coloured Cotton. *Resonance*, **6: 29**, 2001.
- [7] CARVALHO, L. P. de; ANDRADE F. P.; FILHO J. L. da S. Cultivares de Algodão Colorido no Brasil. *Revista Brasileira de Oleaginosa e Fibrosa*, **15: 37**, 2011.
- [8] VIDAL NETO, F. C.; ANDRADE, F. P.; SILVA, J. L.; CARVALHO, L. P. **BRS Topázio: fibra marrom claro uniforme resistente e macia**. Campina Grande: Embrapa Algodão, 2010. (Folder). Disponível em:
<<http://www.infoteca.cnptia.embrapa.br/handle/doc/859647>>. Acessado em 29 de janeiro, 2015.
- [9] FOX, S. **Naturally colored cottons**, Spin off, 1987.

- [10] MATUSIAK, M.; FRYDRYCH, I. Investigation of Naturally Coloured Cotton of Different Origin – Analysis of Fibre Properties. *Fibres & Textiles in Eastern Europe*, **107: 34**, 2014.
- [11] BARROSO, P. A. V.; ALMEIDA, V. C.; ARAÚJO, R. L.; NETO, F. A.; PINTO, F. S. L.; HOFFMANN, L.V.; SANTOS, J. W.; SANTOS, T. S.; FONSECA, R. G. **Intensidade da Cor da Fibra de Algodões Coloridos**. Campina Grande: Embrapa-CNPA, 2005. (Comunicado Técnico, 249).
- [12] SINGH, P.; SINGH, V. V.; WAGHMARE, V. N. **Naturally Coloured Cotton**. CICR Technical Bulletin, Central Institute for Cotton Research Nagpur. v. 4, p. 1-10. Disponível em: <<http://www.cicr.org.in/CICRPublications.html>>. Acessado em 05 de janeiro de 2015.
- [13] GORDON, S.; HSIEH, Y. L., Eds. **Cotton: Science and Technology**. Cambridge, Woodhead Publishing Limited, 2007.
- [14] USDA. **Cotton Standardization**. 2014. Disponível em: <<http://www.ams.usda.gov/AMSv1.0/cotton>>. Acessado em 13 de dezembro de 2014.
- [15] FONSECA, R. G. da; SANTANA, J. C. F. de. **Circular Técnica: Resultados de Ensaio HVI e Suas Interpretações** (ASTM D-4605). Campina Grande: Embrapa Algodão, 2002.
- [16] Conab (Companhia Nacional de Abastecimento). **Perspectiva para a agropecuária**. Safra 2014/2015. Algodão. Brasília, v.2, p. 14, 2014.
- [17] Conab (Companhia Nacional de Abastecimento). **Acompanhamento da safra brasileira. Grãos**, Safra 2013/14, v.1 n.12 - Décimo Segundo Levantamento, Brasília, p. 65, 2014.
- [18] INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA- IBGE. **Levantamento sistemático da produção agrícola: Estatística de produção**

agrícola- Dezembro de 2014- Pesquisa mensal de previsão e acompanhamento das safras agrícolas no ano civil. Disponível em:

[http://www.ibge.gov.br/Producao_Agricola/Levantamento_Sistematico_da_Producao_Agricola_\[mensal\]/Fasciculo/lspa_201403.pdf](http://www.ibge.gov.br/Producao_Agricola/Levantamento_Sistematico_da_Producao_Agricola_[mensal]/Fasciculo/lspa_201403.pdf). Acessado em 20 de janeiro de 2015.

[19] SANTOS, R. F. dos; BARROS, M. A. L. **Algodão: O produtor pergunta, a Embrapa responde**. Brasília: Embrapa de Informação Tecnológica, 2004.

[20] BUANAIN, A. M.; BATALHA, M. **Serie de agronegócio: Cadeia produtiva do Algodão**. Brasília: IICA: MAPA/SPA, v.4, 2007.13p e 15p.

[21] HUA, S.; WANG, X.; YUAN, S.; SHAO, M.; ZHAO X.; ZHU, S.; JIANG L. Characterization of Pigmentation and Cellulose Synthesis in Colored Cotton Fibers. *Crop Science*, **47:1540, 2007**.

[22] KHAN, A. I.; AWAN, F. S.; SADIA, B.; RANA, R. M. and KHAN, I. A. Genetic Diversity Studies among Coloured Cotton Genotypes by Using Rapd Markers. *Pakistan Journal of Botany*, **42: 71-77, 2010**.

[23] VALE, D. G.; GUIMARAES, F. M.; OLIVEIRA, G. S.; CARDOSO, G. D.; ALVES, I.; SILVA, J. C. A.; CARVALHO, L. P.; SILVA, O. R. R. F.; CARTAXO, W. V. **"Algodão colorido: "Tecnologia Embrapa para a geração de emprego e renda na agricultura familiar do Brasil"**. Campina Grande: Embrapa Algodão, 2011. (Folder). Disponível em:<<http://www.infoteca.cnptia.embrapa.br/handle/doc/906449>>. Acesso em: 20 de janeiro de 2015.

[24] CARVALHO, L. P. **Algodão de fibra colorida no Brasil**. In: 5º Congresso Brasileiro do Algodão, 2005, Salvador. **Anais...** Salvador: Embrapa Informação Tecnológica, 2005.

[25] FREIRE, E. C.; ANDRADE, F. P. de; SANTANA, J. C. F. de; BELTRÃO, N. E. de M.; PEDROSA, M. B.; GUEDES, A. R.; WANDERLEY, M. J. R.; ASSUNÇÃO, J. H. de; DANTAS, E. S. B.; SILVA, S. C. da. **BRS 200 Marrom. Cultivar de algodão de fibra colorida**, Campina Grande: Embrapa Algodão, 2010. (Folder). Disponível

<<http://ainfo.cnptia.embrapa.br/digital/bitstream/item/19210/1/BRS-200-Marrom.pdf>>.

Acessado em: 29 de janeiro. 2015.

[26] CARVALHO, L. P.; BELTRAO, N. E. M.; COSTA, J. N.; ANDRADE, F. P.; SILVA, O. R. R. F.; ARAUJO, G. P.; ALVES, I. **BRS Verde**. Campina Grande: Embrapa Algodão, Campina Grande, 2009. (Folder). Disponível em: <<http://www.infoteca.cnptia.embrapa.br/handle/doc/579005>>. Acessado em: 29 de janeiro, 2015.

[27] CARVALHO, L. P.; ARAUJO, G. P.; VIEIRA, R. M.; BELTRAO, N. E. M.; COSTA, J. N. **BRS Rubi**. Campina Grande: Embrapa Algodão, 2009. (Folder). Disponível em: <<http://www.infoteca.cnptia.embrapa.br/handle/doc/579001>>. Acessado em: 29 de janeiro, 2015.

[28] CARVALHO, L. P.; ARAUJO, G. P.; VIEIRA, R. M.; BELTRAO, N. E. M.; COSTA, J. N. **BRS Safira**. Campina grande: Empresa Brasileira de Pesquisa Agropecuária, 2009. (Folder). Disponível em: <<http://www.infoteca.cnptia.embrapa.br/handle/doc/579011>>. Acessado em: 29 de janeiro, 2015.

[29] LANA, V.; RIBAS, P.V. ; CHANSELME, J-L. **Manual de Beneficiamento**. AMPA-IMAmt, 2014.

[30] USDA, Agricultural Marketing Service, **The Classification of Cotton**, Agricultural Handbook 566, USDA, Washington, D.C., 2001.

[31] USTER®. HVI 1000. **Application Report. Quality Characteristics used for Cotton Classification**. USDA/Uster Technologies, 2006. SE 598. Disponível em: http://www.uster.com/fileadmin/customer/Knowledge/Textile_Know_How/Fibre_testing/U_HVI_1000_Quality_characteristics.pdf. Acessado em 05 de janeiro de 2015.

[32] CUI, X.; CAI, Y.; RODGERS, J.; MARTIN, V.; WATSON, M. An investigation into the intra-sample variation in the color of cotton using image analysis. *Textile Research Journal*, **84**: 214, 2014.

[33] DUCKETT, K.; ZAPLETALOVA, T.; CHENG, L.; GHORASHI, H.; WATSON M. Color Grading of Cotton Part I: Spectral and Color Image Analysis. Textile Research Journal, **69: 876, 1999**.

[34]. MATUSIAK M., WALAWSKA A.; Important Aspects of Cotton Colour Measurement. Fibres & Textiles in Eastern Europe, **18: 17, 2010**.

[35] RODGERS, J; SCHILD G.; CUI X.; DELHOM C. TEXTILE TECHNOLOGY: Portable Color Spectrophotometer Measurements of Cotton Color in Remote Locations, Journal of Cotton Science, **17: 202, 2013**.

[36] KANG, T. J.; KIM, C. S. Objective Evaluation of the Trash and Color of Raw Cotton by Image Processing and Neural Network. Textile Research Journal, **72: 776, 2002**.

[37] PAN, Z.; SUN, D.; SUN, J.; ZHOU, Z.; JIA, Y.; PANG, B.; MA, Z.; DU, X. Effects of fiber wax and cellulose content on colored cotton fiber quality. Euphytica, **173: 141, 2010**.

[38] CONRAD, C. M. Determination of Wax in Cotton Fiber A New Alcohol Extraction Method. Industrial and Engineering Chemistry, **16: 745, 1944**.

[39] SILVA, D. J. da. **Predição não destrutiva de propriedades físicas e química em fibras de algodão naturalmente colorido**. Campina Grande-PB, Programa de Pós-Graduação em Ciências Agrárias, UEPB, 2014, Dissertação de Mestrado, 39p.

[40] LIU, Y.; KOKOT, S.; SAMBI, T. J. Vibrational spectroscopy investigation of Australian cotton cellulose fibres. Part 2. A Fourier transform near-infrared preliminary study. Analyst, **123:1725, 1998**.

[41] LIU, Y.; GAMBLE, G.; THIBOBEAUX, D. UV/visible/near-Infrared Reflectance Models for the Rapid and Non-Destructive Prediction and Classification of Cotton Color and Physical Indices. Trans. ASABE, **53:1341, 2010**.

- [42] SOUSA, G. R. M. de. **Aplicação da espectroscopia NIR para análise exploratória de plumas e tecidos de algodão colorido**. Campina Grande-PB, Programa de Pós-Graduação em Ciências Agrárias, UEPB, 2013, Dissertação de Mestrado, 41p.
- [43] OLIVEIRA, L. F.; CANEVARI, N. T.; JESUS, A. M. D.; PEREIRA FILHO, E. R. I. Uso de Imagens Digitais e Análise de Componentes Principais na Identificação dos Níveis de Cr (VI) em Amostras de Solos Revista Virtual de Química, **5: 169, 2013**.
- [44] FRASER, B.; MURPHY, C.; BUNTING, F. **Real World Color Management**. 2nd Edition. United States of America: Peachpit Press, 2005.
- [45] MELCHIADES, F.G.; BOSCHI, A.O. Cores e Tonalidades em Revestimentos Cerâmicos, Cerâmica Industrial, **4: 11, 1999**.
- [46] QUEIROZ, J. E. R.; GOMES, H. M. Introdução ao Processamento Digital de Imagens, Revista RITA: instruções para preparação de documentos em Word, **8: 1, 2001**.
- [47] BROWN, T. L.; LEMAY Jr, H. E.; BURSTEN, B. E.; BURDGE, J. R. **Química: a ciência central**. Tradução de Robson Mendes Matos. 9ª Ed. São Paulo: Pearson Prentice Hall, 2005. Título original: Chemistry – The central Science.
- [48] GONZALEZ, R. C.; WOODS, R. E. **Digital Image Processing**. 2ª Ed. United States of America: Pretence Hall, 2002.
- [49] MILANEZ, K. D. T. M. **Classificação de óleos vegetais comestíveis usando imagens digitais e técnicas de reconhecimento de padrões**. João Pessoa – PB, Programa de Pós Graduação em Química, UFPB, 2013, Dissertação de Mestrado, 102p.
- [50] VILAR, W. T. **Classificação individual de sementes de mamona usando espectroscopia de reflectância no visível, imagens digitais e análises**

multivariadas. João Pessoa – PB, Programa de Pós-Graduação em Química, UFPB, 2014, Dissertação de Mestrado, 73p.

[51] LEÃO, A. C. **Gerenciamento de cores para imagens digitais.** Belo Horizonte, Programa de Pós-Graduação em Artes Visuais, UFMG, 2005. Dissertação de mestrado, 135p.

[52] BATISTA, L.V. Notas de aula da disciplina - **Introdução ao Processamento Digital de Imagens**, 2005, Disponível em:

<<http://www.di.ufpb.br/leonardo/pdi/PDI2005.pdf>> Acessado em 25 de fevereiro de 2015.

[53] MARQUES FILHO, O.; VIEIRA NETO, H. **Processamento Digital de Imagens**, Rio de Janeiro: Brasport, 1999.

[54] SOLOMON, C.; BRECKON, T. **Fundamentals of Digital Image Processing - A Practical Approach with Examples in Matlab.** 1st Edition. United States of America: John Wiley & Sons Ltd, 2011.

[55] GONZALEZ, R. C.; WOODS, R. E. **Processamento de imagens digitais.** Tradução Roberto Marcondes Cesar Júnior, Luciano da Fontoura Costa. 1ª Ed. São Paulo: Editora Blucher, 2000.

[56] PLATANOTIS, K. N.; VENETSANOPOULOS A. N. **Color Image Processing and Applications.** Berlin; Heidelberg; NewYork; Barcelona; Hong Kong; London; Milano; Paris; Singapore; Tokyo: Springer 2000.

[57] ALVES D. R. **Avaliação dos Modelos de Cores RGB e HSV na segmentação de Curvas de Nível em Cartas Topográficas Coloridas.** Belo Horizonte, Programa de Pós-Graduação em Engenharia Elétrica Belo Horizonte, Pontifícia Universidade Católica de Minas Gerais, 2010. Dissertação de mestrado 65p.

[58] LOPES, J. M. B. **Cor e Luz**. Instituto superior técnico, Universidade Técnica de Lisboa, Departamento de Engenharia em Informática, material didático disciplina computação gráfica, 2013.

[59] BEEBE, K. R.; PELL, R. J.; SEASHOLTZ, M. B. **Chemometrics: a practical guide**, New York: John Wiley and Sons, Inc., 1998.

[60] OTTO, M. **Chemometrics: statistics and computer application in analytical chemistry**, New York: Wiley-VCH, 1999.

[61] FERREIRA, M. M. C.; ANTUNES, A. M.; MELGO, M. S.; VOLPE, L. O. Quimiometria I: calibração multivariada, um tutorial. Química Nova, **22: 724, 1999**.

[62] CHAU, F-T.; LIANG, Y-Z.; GAO, J.; SHAO, X-G. **Chemometrics From Basics to Wavelet Transform**, v. 164, New Jersey: John Wiley & Sons, Inc., Publication, 2004.

[63] DINIZ, P. H. G. D.; GOMES, A. A. PISTONESI, M. F.; BAND, B. S. F.; ARAÚJO, M. C. U. Simultaneous Classification of Teas According to Their Varieties and Geographical Origins by Using NIR Spectroscopy and SPA-LDA. Food Analytical. Methods, **7: 1712, 2014**.

[64] BRITO, A. L. B.; ARAÚJO, D. A.; PONTES, M. J. C.; PONTES, L. F. B. L. Near infrared reflectance spectrometry classification of lettuce using linear discriminant analysis. Analytical Methods, **7: 1890, 2015**.

[65] VASCONCELOS, F. V. C.; SOUZA Jr., P. F. B.; PIMENTEL, M. F.; PONTES, M. J. C.; PEREIRA, C. F. Using near-infrared overtone regions to determine biodiesel content and adulteration of diesel/biodiesel blends with vegetable oils. Analytica Chimica Acta, **716: 101, 2012**.

[66] ESTEKI, M.; NOUROOZI, S.; SHAHSAVARI, Z. **A fast and direct spectrophotometric method for the simultaneous determination of methyl paraben and hydroquinone in cosmetic products using successive projections algorithm**. International Journal of Cosmetic Science, 2015. [DOI: 10.1111/ics.12241](https://doi.org/10.1111/ics.12241).

- [67] GONZÁLEZ, A. G. Use and misuse of supervised pattern recognition methods for interpreting compositional data. *Journal of Chromatography A*, **1158**: 215, 2007.
- [68] NAES, T.; ISAKSSON, T.; FEARN, T.; DAVIES, T. **A User-Friendly Guide to Multivariate Calibration and Classification**. Chichester: NIR Publications, 2002.
- [69] CORREIA, P. R. M.; FERREIRA, M. M. C. Reconhecimento de padrões por métodos não supervisionados: explorando procedimentos quimiométricos para tratamento de dados analíticos. *Química Nova*, **30**: 481, 2007.
- [70] SILVA, A. C. da. **Metodologias analíticas para a identificação de não conformidades em amostras de álcool combustível**. - João Pessoa – PB, Programa de Pós-Graduação em Química, UFPB, 2013, Dissertação de Mestrado, 82p.
- [71] BALLABIO, D.; CONSONNI, V. Classification tools in chemistry. Part 1: linear models PLS-DA. *Analytical Methods*, **5**: 3790, 2013.
- [72] BALLABIO D.; TODESCHINI R. Multivariate Classification for Qualitative Analysis in: Da-Wen-Sun. **Infrared Spectroscopy for food quality analysis and control**. 1ed. Elsevier Inc. 2009, p 97.
- [73] KUMAR, N.; BANSAL, A.; SARMA, G. S.; RAWAL, R. K. Chemometrics tools used in analytical chemistry: An overview. *Talanta*, **123**: 186, 2014.
- [74] PONTES, M. J. C. **Algoritmo das Projeções Sucessivas para seleção de variáveis espectrais em problemas de classificação**. João Pessoa, Programa de Pós-Graduação em Química, UFPB, 2009, Tese de doutorado, 123p.
- [75] LIRA, L. de F. B. **Desenvolvimento de Métodos Analíticos para Monitoramento da Qualidade do Biodiesel e suas Misturas**. Recife, Programa de Pós-Graduação em Química, UFPE. 2010. Tese de doutorado, 124p.

- [76] SIMÕES, S. da S. **Desenvolvimento de métodos validados para a determinação de captopril usando espectrometria NIRR e calibração multivariada**. João Pessoa, Programa de Pós-Graduação em Química, UFPB. 2008. Tese de doutorado, 83p
- [77] LEARDI, R. Genetic algorithms in chemometrics and chemistry: a review. *Journal of Chemometrics*, **15**: 559, 2001.
- [78] CANECA, A. R.; PIMENTEL, M. F.; GALVÃO, R. K. H.; MATTA, C. E.; CARVALHO, F. R.; RAIMUNDO JR., I. M.; PASQUINI, C.; ROHWEDDER, J. J. R., Assessment of infrared spectroscopy and multivariate techniques for monitoring the service condition of diesel-engine lubricating oils, *Talanta*, **70**: 344, 2006.
- [79] 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**: 65, 2001.
- [80] GALVÃO, R. K. H.; ARAÚJO, M. C. U.; SILVA, E. C.; JOSÉ, G. E.; SOARES, S. F. C.; PAIVA, H. M. Cross-validation for the Selection of Spectral Variables Using the Successive Projections Algorithm. *Journal of Brazilian Chemical Society*, **18**: 1580, 2007.
- [81] PONTES, M.J.C.; GALVÃO, R.K.H.; ARAÚJO, M.C.U.; MOREIRA, P.N.T.; NETO, O. D. P.; JOSÉ, G.E.; SALDANHA, T.C.B. The successive projections algorithm for spectral variable selection in classification problems. *Chemometrics and Intelligent Laboratory Systems*, **78**, 11: 2005.
- [82] NORGAARD, L.; SAUDLAND, A.; WAGNER, J.; NIELSEN, J. P.; MUNCK, L.; ENGELSEN, S. B. Interval Partial Least-Squares Regression (iPLS): A Comparative Chemometric Study with an Example from Near Infrared Spectroscopy. *Applied Spectroscopy*, **54**: 413, 2000.

- [83] SOARES, S. F. C.; GALVÃO, R. K. H.; PONTES, M. J. C.; ARAÚJO, M. C. U. A New Validation Criterion for Guiding the Selection of Variables by the Successive Projections Algorithm in Classification Problems. *Journal Brazilian Chemistry Society*, **25: 176, 2014.**
- [84] MILANEZ, K. D. T. M.; PONTES, M. J. C. Classification of edible vegetable oil using digital image and pattern recognition techniques. *Microchemical Journal*, **113: 10, 2014.**
- [85] 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: 129, 2012.**
- [86] SILVA, C. S.; BORBA, F. de 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: 122, 2013.**
- [87] BRITO, A. L. B.; BRITO, L. R.; HONORATO, F. A.; PONTES, M. J. C.; PONTES, L. F. B. L. Classification of cereal bars using near infrared spectroscopy and linear discriminant analysis. *Food Research International*, **51: 924, 2013.**
- [88] KENNARD, R. W.; STONE, L. A. Computer Aided Design of Experiments. *Technometrics*, **11: 137, 1969.**
- [89] GALVÃO, R. K. H.; ARAUJO, M. C. U.; JOSE, G. E.; PONTES, M. J. C.; SILVA, E. C.; SALDANHA, T. C. B. A method for calibration and validation subset partitioning. *Talanta*, **67: 736, 2005.**
- [90] 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: 2648, 2012.**

- [91] ALMEIDA, V. E. de; COSTA, G. B. da; FERNANDES, D. D. de S.; DINIZ, P. H. G. D.; BRANDÃO, D.; MEDEIROS, A. C. D. de; VÉRAS, G. Using color histograms and SPA-LDA to classify bactéria. *Analytical and Bioanalytical Chemistry*, **406**: 5989, 2014.
- [92] VILAR, W. T. S.; ARANHA, R. M; MEDEIROS, E.P.; PONTES M. J. C. Classification of Individual Castor Seeds Using Digital Imaging and Multivariate. *Analysis Journal of the Brazilian Chemical Society*, **26**: 102, 2015.
- [93] SOUTO, U. T. de C. P.; BARBOSA, M. F.; DANTAS, V. H.; PONTES, V. T.; LYRA, W. da S.; DINIZ, P. H. G. D.; ARAÚJO, M. C. U.de and SILVA, E. C. da. Screening for Coffee Adulteration Using Digital Images and SPA-LDA. *Food Analytical Methods*, **8**: 1515, 2015.
- [94] SANTOS, P. M.; WENTZELL, P. D.; PEREIRA FILHO, E. R. Scanner Digital Images Combined with Color Parameters: A Case Study to Detect Adulterations in Liquid Cow's Milk. *Food Analytical Methods* **5**: 89, 2012.
- [95] BOTELHO, B. G.; ASSIS, L. P. de; SENA, M. M. Development and analytical validation of a simple multivariate calibration method using digital scanner images for sunset yellow determination in soft beverages. *Food Chemistry*, **159**: 175, 2014.
- [96] DOMINGUEZ, M. A.; CENTURIÓN, M. E. Application of digital images to determine color in honey samples from Argentina. *Microchemical Journal*, **118**: 110, 2015.
- [97] DU, C. J; SUN, D. W. Recent developments in the applications of image processing techniques for food quality evaluation. *Trends in Food Science & Technology*, **15**: 230, 2004.
- [98] MUSTAFIC, A; LI, C. Classification of cotton foreign matter using color features extracted from fluorescent images. *Textile Research Journal*, **85**: 1209, 2015.

[99] LIU, F, SU, Z.; HE, X.; ZHANG, C.; CHEN, M.; QIAO, L. A laser imaging method for machine vision detection of white contaminants in cotton. *Textile Research Journal*, **84**: 1987, 2014.

[100] JI, R.; LI, D.; CHENG, L.; YANG, W. Classification and identification of foreign fibers in cotton on the basis of a support vector machine. *Mathematical and Computer Modelling*, **51**: 1433, 2010.

[101] YANG, W.; LU, S.; WANG S.; LI, D. Fast recognition of foreign fibers in cotton lint using machine vision. *Mathematical and Computer Modelling*, **54**: 877, 2011.

[102] ASTM D1776, **Standard Practice for Conditioning and Testing Textiles**, American Society for Testing and Materials, West Conshohocken, Pennsylvania, USA, 2004.

[103] PÉRES, N. F.; FERRÉ, J.; BOQUÉ, R. Calculation of the reliability of classification in discriminant partial least-squares binary classification. *Chemometrics and Intelligent Laboratory Systems*, **95**: 122, 2009.

[104] MAGNUSSON B.; ORNEMARK U. (eds.) Eurachem Guide: **The Fitness for Purpose of Analytical Methods – A Laboratory Guide to Method Validation and Related Topics**, (2nd ed. 2014). ISBN 978-91-87461-59-0. Disponível em: www.eurachem.org. Acessado em: 06 de junho de 2015.



CAPÍTULO VII

ANEXO

ANEXO

Artigo Publicado

Cite this: *Anal. Methods*, 2015, 7, 5869

A novel strategy for the classification of naturally colored cotton fibers based on digital imaging and pattern recognition techniques

Maria Ivanda S. Gonçalves,^a Welma T. S. Vilar,^a Everaldo Paulo Medeiros^b
and Márcio José Coelho Pontes^{*a}

This study proposes the use of digital images and multivariate analysis as an alternative methodology for the classification of naturally colored cotton fiber, according to cultivar type. For this purpose, two groups were evaluated: the first comprised of BRS 200 Marrom and BRS Topázio cultivars, while the second contained BRS Rubi and BRS Safira cultivars. Cotton fiber sample images were obtained using a webcam and the frequency distribution of color indexes on the grayscale, red–green–blue (RGB), hue (H), saturation (S), value (V), and grayscale channels was obtained. Classification models based on linear discriminant analysis (LDA) with prior variable selection by successive projection algorithm (SPA) and stepwise (SW) were used. For the purpose of comparison, partial least squares discriminant analysis (PLS-DA) applied to the full-histogram was also used. For both groups, the best results were achieved using the LDA/SW model, with a correct classification rate (CCR) of 96% for the prediction set using the HSV combination. The proposed method is simple, low cost, does not use a reagent, does not destroy the sample and provides results in a short period of time.

Received 1st May 2015
Accepted 4th June 2015
DOI: 10.1039/c5ay01116g
www.rsc.org/methods

1. Introduction

The cotton plant is a fibrous and oil plant. Worldwide, there are more than 50 species, but *Gossypium hirsutum* L., latifolium Hutch and *G. barbadense* L. are the most cultivated.¹ Cotton is the most important natural textile fiber and the sixth largest worldwide source of vegetable oil.² In the majority of primitive species, cotton has colored fibers that are controlled by a dominant gene.^{3,4} However, sunlight and soil type among other factors can also influence some colors.^{4,5} Research has shown that white cotton and colored cotton evolved at the same time. Colored cotton, however, has inferior characteristics to white cotton.^{4–6}

In 1988, working in the United States, Sally Fox⁷ managed to develop naturally colored cotton fibers tolerant to machine processing. With this work, interest in colored fibers increased. Soon, other countries such as Israel, Brazil, Peru, Greece and Turkey⁸ also worked with genetic modifications in order to develop new colored fibers with good quality.

Due to their non-toxic characteristics, naturally colored cottons have been given increasing importance by researchers in recent years. In fact, the process of bleaching and dyeing fiber

with the use of toxic chemicals, including environment polluting heavy metals,^{3,6,9} has been dispensed with due to the increased use of naturally colored fiber. Since many of the chemical reagents used in the dyeing process can cause serious health problems such as cancer, allergies and rashes,^{3,5,9} there is interest in clothing produced from naturally colored cotton.

Among the colored cottons, brown is most commonly found in different shades, which can range from a light brown to an intense reddish brown colour.⁹ Genetic modification programs have been used to develop better quality fibers with different characteristics, but with similar coloration. Commercial cotton plumes in Brazil, such as BRS 200 Marrom, feature a great similarity to BRS Topázio, while BRS Rubi resembles BRS Safira. This can be a problem because the identification of these samples is commonly accomplished through a visual inspection, which is a highly subjective method. According to the United States Department of Agriculture (USDA), the classification can also be performed by the determination of various physical properties of the samples using the high volume instrument system (HVI). This is a high cost method and requires adequate infrastructure and trained personnel for the analysis procedure. Additionally, this type of analysis does not allow mobility for performing measurements *in situ*.

Several studies have been reported in the literature using atomic absorption spectroscopy (AAS),¹⁰ inductively coupled plasma atomic emission spectroscopy (ICP-AES),¹⁰ Fourier transform infrared,¹¹ and laser-induced breakdown spectroscopy (LIBS),¹² among others^{13–15} for quality control of cotton.

^aDepartamento de Química, Universidade Federal da Paraíba, Laboratório de Automação e Instrumentação em Química Analítica/Quimiometria (LAQA), CEP 580051-970, João Pessoa, PB, Brazil. E-mail: marciocoelho@quimica.ufpb.br; Fax: +55 83 3216 7438; Tel: +55 83 3216 7438

^bEmbrapa, Centro Nacional de Pesquisa de Algodão, Campina Grande, PB, Brazil