



UNIVERSIDADE ESTADUAL DA PARAÍBA
CAMPUS I - CAMPINA GRANDE
PRÓ-REITORIA DE PÓS-GRADUAÇÃO E PESQUISA
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA
MESTRADO EM QUÍMICA

DARLEI GUTIERREZ DANTAS BERNARDO OLIVEIRA

**CLASSIFICAÇÃO E DETERMINAÇÃO DOS ATRIBUTOS DE QUALIDADE
DAS FIBRAS DE ALGODÃO UTILIZANDO UM ESPECTRÔMETRO
RAMAN PORTÁTIL E QUIMIOMETRIA**

CAMPINA GRANDE PB
JANEIRO 2022

DARLEI GUTIERREZ DANTAS BERNARDO OLIVEIRA

**CLASSIFICAÇÃO E DETERMINAÇÃO DOS ATRIBUTOS DE QUALIDADE
DAS FIBRAS DE ALGODÃO UTILIZANDO UM ESPECTRÔMETRO
RAMAN PORTÁTIL E QUIMIOMETRIA**

Trabalho de dissertação apresentado ao
Programa de Pós-Graduação em Química
da Universidade Estadual da Paraíba,
como requisito para obtenção do título de
Mestre em Química

Área de concentração: Metodologias
Analítica Avançada

Orientadora: Profa. Dra. Simone da Silva Simões

Coorientador: Dr. Paulo Everaldo de Medeiros

CAMPINA GRANDE PB
JANEIRO 2022

- O48c Oliveira, Darlei Gutierrez Dantas Bernardo.
Classificação e determinação dos atributos de qualidade das fibras de algodão utilizando um espectrômetro Raman portátil e quimiometria [manuscrito] / Darlei Gutierrez Dantas Bernardo Oliveira. - 2022.
102 p. : il. colorido.
- Digitado.
- Dissertação (Mestrado em Química - Mestrado) - Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2022.
- "Orientação : Prof. Dr. Simone da Silva Simões , Departamento de Química - CCT."
- "Coorientação: Prof. Dr. Paulo Everaldo de Medeiros , Embrapa Algodão"
1. Plumas de algodão. 2. Controle de qualidade. 3. Análise HVI. 4. Análises espectrais. 5. Análise Multivariada. I. Título
21. ed. CDD 633.51

DARLEI GUTIERREZ DANTAS BERNARDO OLIVEIRA

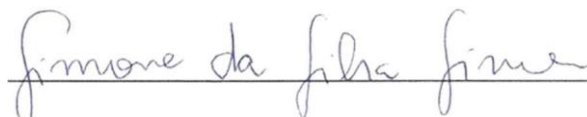
**CLASSIFICAÇÃO E DETERMINAÇÃO DOS ATRIBUTOS DE QUALIDADE
DAS FIBRAS DE ALGODÃO UTILIZANDO UM ESPECTRÔMETRO
RAMAN PORTÁTIL E QUIMIOMETRIA**

Trabalho de dissertação apresentado ao
Programa de Pós-Graduação em Química
da Universidade Estadual da Paraíba,
como requisito para obtenção do título de
Mestre em Química.

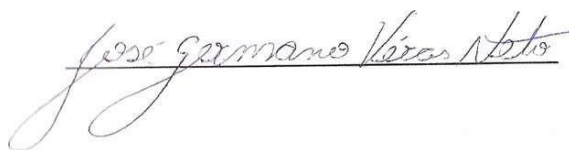
Área de concentração: Metodologias
Analíticas Avançadas

Aprovada em: 25/janeiro/2022

BANCA EXAMINADORA



Profa. Dra. Simone da Silva Simões (Orientadora)
Universidade Estadual da Paraíba (UEPB)



Prof. Dr. José Germano Vêras Neto - Examinador Interno
Universidade Estadual da Paraíba (UEPB)



Prof. Dr. Neirivaldo Cavalcante da Silva – Examinador Externo
Universidade Federal do Pará (UFPA)

AGRADECIMENTOS

Agradeço a DEUS por todas as coisas boas ocorridas na minha vida, em especial nessa etapa do mestrado. A ELE dedico todas as minhas conquistas.

Agradeço a minha família, em especial minha Mãe Janecleide que apesar da distância ajuda em todos os momentos possíveis. Também agradeço a Clenilda, Odair, minhas madrinhas Raimunda e Odília, meus avós Elvira e Francisco, meus irmãos Ágata e Andrei, meus irmãos (primos) Altair e Aldair e minha tia Janeleide.

Agradeço a minha companheira de vida que está comigo em mais uma conquista, Yara Duarte. Você tem sido muito importante em todos esses momentos, pois me dá forças quando preciso e também torna mais significativo os bons momentos, muito obrigado por toda paciência e acolhimento para comigo. Também agradeço a sua família que tem sido maravilhosa comigo a Branca, Walter e Kiara.

Agradeço a todos os amigos construídos ao longo da graduação e também no mestrado em especial a Everton Vieira, José Gorete, Itamar, Winício, Damião, Antônio, Mabson, Valéria, Flávio, Tamires, Willyan, Vitor, Elizandro, Renally, Elane, Mirelly, Tiago, Carol, Ronier, Guilherme e Jeferson.

Agradeço ao grupo de pesquisa GMAQ que foi fundamental para construção dessa pesquisa, pois os encontros e reuniões foram muito ricos em conhecimentos e discussões importantes.

Agradeço a todos os funcionários que fazem uma UEPB cada vez melhor. A todo o corpo docente e administração do PPGQ que contribuiu muito com ensinamentos, sendo importante para construção de toda trajetória durante o mestrado. Em especial agradeço a minha Orientadora Dra. Simone Simões, pois tem sido incrível durante esse percurso com muita paciência e dedicação para que esse trabalho pudesse ser desenvolvido, além de todo ensinamento para eu ser um bom profissional que levarei também para a vida.

Agradeço ao Dr. Everaldo que oportunizou a parceria com a Embrapa Algodão de Campina Grande para que a pesquisa pudesse ser desenvolvida. Também agradeço a FAPESQ que apesar das dificuldades tem procurado investir na pesquisa através de bolsas para alunos, sendo muito importante no processo de conclusão do mestrado.

Muito Obrigado!

RESUMO

A pluma de algodão é uma importante fibra natural com diversas aplicações. O seu valor de mercado está associado à qualidade, expressa por meio de variáveis e considerações de aspectos como localidade de produção, modificações genéticas e maturidade. Essas medidas são realizadas por instrumento de alto volume (HVI, do inglês *High Volume Instrument*). Essa tecnologia de análise demanda custos elevados, tempo de análise demorado, preparo de amostras trabalhoso, entre outras condições. Com a espectroscopia Raman tais medidas podem ser realizadas de modo simples, rápido e economicamente viável. Sendo assim, objetiva-se desenvolver uma metodologia analítica por meio da espectroscopia Raman e análise multivariada para determinação de atributos de qualidade e classificação de fibras de algodão de diferentes classes. Para isso, foram registrados espectros de 213 amostras de fibras de algodão na região de 400 a 2300 cm^{-1} com um espectrômetro Raman portátil. Com o HVI, realizou-se as medidas de referência dos atributos comprimento, índice de fibras curtas, índice de micronaire, resistência e reflectância. A análise exploratória foi realizada através das PCA's com os dados de atributos e espectros Raman. Igualmente, para classificação das amostras construíram-se modelos de PLS-DA com os dados de atributos de qualidade medidos e espectrais Raman. Já a determinação dos atributos deu-se através dos modelos de PLS. A partir do perfil espectral das amostras selecionou-se uma faixa de trabalho na região de 700 a 1700 cm^{-1} . Nos modelos PLS-DA Raman obteve-se para as amostras de diferentes maturidades um RMSE de 17% na etapa de teste. Com os dados de atributos de qualidade obteve-se para as amostras da localidade B 90,51% de acerto na etapa de teste. No PLS, o modelo para o parâmetro de índice de fibras curta apresentou um RMSEC de 0,590 e RMSEP de 0,525, inferiores ao coeficiente de variação 8,55. Deste modo, o método apresenta potencial para classificação de amostras através da espectroscopia Raman, tendo em vista que é uma técnica de simples aplicação e de baixo custo para análise, além da rapidez que o método pode apresentar para classificar as fibras quanto ao perfil produtivo. Também, é possível fazer a classificação de fibras de diferentes localidades a partir dos modelos construídos com os dados de HVI. Já para os modelos de quantificação são necessários alguns ajustes experimentais para melhorar sua capacidade preditiva.

Palavras-chaves: Plumas de algodão. Controle de qualidade. Análise HVI. Análises espectrais. Análise Multivariada.

ABSTRACT

Cotton down is an important natural fiber with many applications. Its market value is associated with quality, expressed through variables and considerations of aspects such as production location, genetic modifications and maturity. These measurements are performed by a High-Volume Instrument (HVI). This analysis technology demands high costs, lengthy analysis time, laborious sample preparation, among other conditions. With Raman spectroscopy such measurements can be performed in a simple, fast and economically viable way. Therefore, the objective is to develop an analytical methodology through Raman spectroscopy and multivariate analysis to determine quality attributes and classification of cotton fibers of different classes. For this, spectra of 213 cotton fiber samples were recorded in the region of 400 to 2300 cm^{-1} with a portable Raman spectrometer. With the HVI, reference measurements were made for the attributes length, short fiber index, micronaire index, strength and reflectance. The exploratory analysis was performed through the PCA's with the data of attributes and Raman spectra. Likewise, for the classification of samples, PLS-DA models were built with the data of measured quality attributes and Raman spectral. The determination of the attributes took place through the PLS models. From the spectral profile of the samples, a working range in the region of 700 to 1700 cm^{-1} was selected. In the PLS-DA Raman models, an RMSE of 17% was obtained for the samples of different maturities in the test stage. With the data of quality attributes, 90.51% accuracy was obtained for the samples from locality B in the test stage. In PLS, the model for the short fiber index parameter presented an RMSEC of 0.590 and RMSEP of 0.525, lower than the coefficient of variation 8,55. Thus, the method has potential for classifying samples through Raman spectroscopy, considering that it is a technique of simple application and low cost for analysis, in addition to the speed that the method can present to classify the fibers according to the productive profile. Also, it is possible to classify fibers from different locations from the models built with the HVI data. As for the quantification models, some experimental adjustments are needed to improve their predictive capacity.

Keywords: Cotton Plumes. Quality Control. HVI Analysis. Spectral Analysis. Multivariate Analysis.

LISTA DE FIGURA

Figura 01 - Representação da progressão de desenvolvimento da fibra de algodão	22
Figura 02 - Estrutura da fibra de algodão	23
Figura 03 - Teores de celulose e açúcares redutores de fibra de algodão durante a sua formação.	25
Figura 04 - Estrutura Química de quatro resíduos da cadeia de celulose.....	27
Figura 05 - Diagrama de nível de energia espectral para o Raman e Raman ressonante	37
Figura 07 - Representação gráfica bidimensional de dois grupos de amostra	47
Figura 08 - Esquematização da construção do modelo PLS-DA	49
Figura 09 - Esquematização da organização do conjunto de calibração	51
Figura 10 - Espectrômetro Raman Portátil.....	56
Figura 11 - Sistema HVI para medidas dos atributos de qualidade das fibras de algodão	57
Figura 12 - Espectros Raman de 213 amostras de fibras de algodão na região 400 a 2300 cm^{-1} destacando-se as bandas características.....	60
Figura 13 - Espectros Raman de 213 amostras de fibras de algodão na região de 700 cm^{-1} a 1700 cm^{-1} brutos (a) e com pré-processamento base NVU + WLS + SF31 (b)	62
Figura 14 - PCA's aplicados a dados de atributos de qualidade das 213 amostras de fibras de algodão de diferentes classes. Gráfico de T2 de Hotelling vs Q residual para (a) amostras de diferentes localidades, (c) com ou sem modificação genética e (e) relativa à maturidade. Escores de PC1 vs PC2 para (b) amostras de diferentes localidades, (d) com ou sem modificação genética e (f) relativa à maturidade.	64
Figura 15 - Gráficos de PCA's aplicados a dados espectrais Raman de 213 amostras de fibras de algodão de diferentes classes. Gráfico de T2 de Hotelling vs Q residual para (a)	

amostras de diferentes localidades, (c) com ou sem modificação genética e (e) relativa à maturidade. Escores de PC1 vs PC2 para (b) amostras de diferentes localidades, (d) com ou sem modificação genética e (f) relativa à maturidade. 66

Figura 16 - Gráficos de VIP escores dos dados de atributos de qualidade das fibras de algodão da localidade A (a); gráfico de VIP escores das amostras da localidade B (b); gráfico de VIP escores das amostras da localidade C (c) 70

Figura 17 - Gráficos de VIP escores dos dados de atributos de qualidade das fibras de algodão com modificações genéticas 71

Figura 18 - Gráficos de VIP escores dos dados de atributos de qualidade das fibras de algodão de diferentes maturidades 71

Figura 19 - Gráficos de predição do modelo PLS-DA-At das classes de fibras de algodão com diferentes localidades (a), modificações genéticas (b) e diferença na maturidade (c) 72

Figura 20 - Gráficos de VIP escores dos modelos PLS-DA-Raman das fibras de algodão da localidade A (a); gráfico de VIP escores das amostras da localidade B (b); gráfico de VIP escores das amostras da localidade C (c). 78

Figura 21 - Gráficos de VIP escores dos dados espectrais Raman das fibras de algodão de diferenças genéticas 79

Figura 22 - Gráficos de VIP escores dos dados espectrais Raman das fibras de algodão de diferenças de maturidade 80

Figura 23 - Gráficos de predição dos espectros Raman de fibras de algodão de diferentes localidades (a), diferença genética (b) e diferença na maturidade (c) 81

Figura 24 - Gráficos dos valores previstos versus os valores de referência HVI obtidos no modelo PLS para os parâmetros UHM, g.tex-1, MIC, SFI e %Rd 87

LISTA DE TABELA

Tabela 01 - Indicadores de produção e exportação mundial de algodão.....	18
Tabela 02 - Composição típica da fibra de algodão	28
Tabela 03 - Teor de metal no algodão	30
Tabela 04 - Exemplificação de uma tabela de confusão para duas classes de amostras	49
Tabela 05 - Número de amostras de treinamento e teste de cada classe de fibras de algodão	58
Tabela 06 - Principais frequências de bandas Raman característicos de fibras de algodão	61
Tabela 07 - Valores das figuras de mérito dos modelos PLS-DA-At construídos com dados de atributos de qualidade de diferentes classes de fibras de algodão.....	68
Tabela 08 - Valores das figuras de mérito dos modelos PLS-DA-Raman construídos com dados espectrais Raman de diferentes classes de fibras de algodão.....	75
Tabela 09 - Comparação das figuras de mérito na etapa de teste dos modelos PLS-DA-Raman e PLS-DA-At.....	83
Tabela 10 - Medidas dos atributos de qualidade das fibras de algodão pelo método HVI	84
Tabela 11 - Resultados obtidos no modelo PLS para os parâmetros de comprimento (UHM), índice de fibras curtas (SFI), micronaire (MIC), resistência (TR) e reflectância (Rd).....	85

LISTA DE ABREVIATURAS E SIGLAS

- ABRAPA – Associação Brasileira dos Produtores de Algodão
- CARS – Espalhamento Raman coerente Anti-Stolkes (do inglês, *coherent Anti-Stolkes Raman Scattering*)
- CCD - dispositivo de carga acoplada (do inglês, *charge coupled devices*)
- CMF - Microfibrilas de celulose (*Cellulose microfibrils*)
- CV – Coeficiente de variação (do inglês, *coefficient of variation*)
- DPA – Dia pós-antese (do inglês, *Days posthantesis*)
- FN – Falso negativo
- FP – Falso positivo
- HVI – Instrumento de alto volume (do inglês, *High Volume Instrument*)
- ICAC – Comitê Consultivo Internacional do Algodão (do inglês, *International Cotton Advisory committee*)
- IUPAC – União de Química pura e aplicada (do inglês, *International Union of Pure and Applied Chemistry*)
- GP - Grau de polimerização (GP)
- LDA – Análise discriminante linear (do inglês, *Linear Discriminant Analysis*)
- LV – Variável latente (do inglês, *latent variable*)
- MIC – Índice de micronaire (do inglês, *micronaire index*)
- MLR – Regressão linear múltipla (do inglês, *multiple linear regression*)
- MSC – Correção do espalhamento multiplicativo (do inglês, *multiplicative scatter correction*)
- NVU – Normalização pelo vetor unitário (do inglês, *normalization by unit vector*)
- PCR – Regressão em componentes principais (do inglês, *principal component regression*)
- MLR – Regressão linear múltipla (do inglês, *multiple linear regression*)
- PLS – Mínimos quadrados parciais (do inglês, *partial least square*)
- PCA – Análise de componentes principais (do inglês, *principal component analysis*)
- PCW - Biossíntese da parede celular primária (do inglês, *Primary cell wall*)
- PLS-DA - Análise discriminante de mínimos quadrados parciais (do inglês, *partial least square discriminant analysis*)
- Rd – Grau de reflectância (do inglês, *reflectance degree*)
- RMSE – Raiz quadrada do erro médio quadrático (do inglês, *Root mean square error*)

RMSEC – Raiz quadrada do erro médio quadrático de calibração (do inglês, *Root mean square error of calibration*)

RMSECV – Raiz quadrada do erro médio da validação cruzada (do inglês, *Root mean square error of cross validation*)

RMSEP – Raiz quadrada do erro médio quadrático de predição (do inglês, *Root mean Squares Error of Prediction*)

SCW - Biossíntese da parede celular secundária (do inglês, *Secondary cell wall*)

SEP – Erro padrão relativo (do inglês, *standard error prediction*)

SFI – Índice de fibras curtas (do inglês, *short fiber index*)

SIMCA – Modelagem independente suave de classe análoga (do inglês, *soft independent modeling of class analog*)

SNV – Variação normal padrão (*standard normal variate*)

SPXY – Partição da amostra baseado na distância X-Y (do inglês, *Samples set Partioning based on joint*)

SRS – Espalhamento Raman estimulado (do inglês, *stimulated Raman Scattering*)

SS - Sacarose sintase

SVM - Máquina de vetores de suporte (SVM, do inglês *support vector machine*)

UHM – Metade superior do comprimento médio (do inglês, *Upper Half Mean Lenght*)

UI – Índice de uniformidade (do inglês, *uniformity index*)

USDA – Departamento de agricultura dos Estados unidos (do inglês, *United States Department of Agriculture*)

VN – Verdadeiro negativo

VP – Verdadeiro positivo

WLS - Peso mínimos quadrados (do inglês, *Weight Least Squares*)

SUMÁRIO

1 INTRODUÇÃO	14
2 OBJETIVOS	16
2.1 Objetivo Geral	16
2.2 Objetivos Específicos.....	16
3 ESTADO DA ARTE.....	17
3.1 Fibras de algodão: contexto histórico, importância econômica e principais aplicações.....	17
3.2 Desenvolvimento biológico, estrutura e composição química da fibra de algodão.....	18
3.2.1 Estágio de Iniciação	19
3.2.2 Biossíntese de alongamento da parede celular primária	20
3.2.3 Biossíntese da celulose na parede celular secundária	20
3.2.4 Processo de Maturação.....	21
3.2.5 Estrutura da Fibra de algodão	23
3.2.6 Composição Química da Fibra de Algodão	25
3.2.6.1 Estrutura Química da Celulose	26
3.2.6.2 Outros constituintes da fibra de algodão	27
3.2.6.3 Teores de metais na fibra de algodão	29
3.3 Atributos de qualidade das fibras de algodão.....	31
3.3.1 Comprimento	31
3.3.2 Índice de Fibras Curtas	32
3.3.3 Resistência Específica	33
3.3.4 Índice de Micronaire	34
3.3.5 Grau de Reflectância	35
3.4 A espectroscopia raman aplicada nas análises de fibras de algodão	36
3.5 Métodos quimiométricos.....	39
3.5.1 Pré-Processamento dos dados	41
3.5.1.1 Centragem na média.....	42
3.5.1.2 Escalamento pela variância	42
3.5.1.3 Autoescalamento.....	43
3.5.1.4 Normalização	44
3.5.1.5 Correção de linha de base.....	45

3.5.1.6 Técnica de alisamento ou suavização	45
3.5.2 Análise de Componentes Principais (PCA)	46
3.5.3 Classificação Supervisionada.....	48
3.5.3.1 PLS-DA.....	48
3.5.4 Calibração Multivariada	50
3.5.4.1 Regressão por Mínimos Quadrados Parciais (PLS)	52
3.5.4.2 Parâmetros de qualidade do modelo	53
4 PERCURSO METODOLÓGICO	55
4.1 Localização da pesquisa	55
4.2 Amostras estudadas.....	55
4.3 Etapas da pesquisa	55
4.3.1 Primeira Etapa: Aquisição dos dados espectrais Raman.....	55
4.3.2 Segunda Etapa: Análises dos atributos de qualidade das fibras de algodão	56
4.3.3 Terceira Etapa: Construção dos modelos Quimiométricos	57
4.3.3.1 Softwares Quimiométrico	57
4.3.3.2 Análise exploratória dos dados	57
4.3.3.3 Construção do modelo PLS-DA	58
4.3.3.4 Construção do modelo PLS	59
5 RESULTADOS E DISCUSSÕES	60
5.1 análise dos espectros RAMAN.....	60
5.2 Análise exploratória dos dados.....	63
5.2.1 PCA dos dados de atributos de qualidade das fibras de algodão	63
5.2.2 PCA dos dados espectrais Raman das fibras de algodão	65
5.3 Modelo de classificação supervisionada	67
5.3.1 PLS-DA dos atributos de qualidade das fibras de algodão (PLS-DA-At)	67
5.3.2 PLS-DA dos dados espectrais Raman das fibras de algodão (PLS-DA-Raman).....	73
5.4 Calibração multivariada: modelos PLS's.....	84
6 CONSIDERAÇÕES FINAIS.....	88
7 PERSPECTIVAS FUTURAS.....	90
REFERÊNCIAS	91
APÊNDICE A – Gráficos de T² de hotelling e Q residual do modelo PLS-DA-At.....	100
APÊNDICE B - Gráficos de T² de hotelling e Q residual do modelo PLS-DA-Raman.....	101

APÊNDICE C – Gráficos de T^2 de <i>hotelling</i> versus Q residual dos modelos PLS para determinação dos atributos de qualidade das fibras de algodão	102
---	------------

1 INTRODUÇÃO

O algodão é uma importante fibra natural para economia de mais de 80 países do mundo (SIDDIQUI; WANG; MENON, 2020). Nos últimos anos a produção e o consumo das fibras de algodão têm crescido significativamente de modo que a estimativa de produção para safra de 2021/2022 é de 25,59 milhões de toneladas de algodão, um aumento de 5% em relação à safra anterior. Para respectiva safra, a estimativa de consumo é de 25,5 milhões de toneladas, ainda abaixo dos 26,0 milhões de toneladas comercializados sobre condições normais, isto é, sem a pandemia como relata o Comitê Consultivo Internacional do Algodão (ICAC, do inglês *International Cotton Advisory committee*) (ABRAPA, 2021).

A alta demanda por fibras de algodão é dada pela possibilidade de aplicações em diversos ramos industriais. Tendo na fiação com destino à indústria têxtil a sua principal aplicação, as fibras de algodão também são utilizadas para aplicações na medicina, desenvolvimento de móveis, indústria automobilística, indústria de celulose entre vários outros setores (ALVES et al., 2021). Essa valorização tem levado países como Brasil, Estados Unidos, Índia entre outros a produzirem e exportarem a pluma de algodão.

O destaque do Brasil no mercado internacional algodoeiro está associado a qualidade da pluma produzida (ALVEZ et al., 2021). Com isso, avaliação da qualidade das fibras de algodão mostra-se fundamental para determinação do seu valor de mercado e suas aplicabilidades. Para a determinação dos parâmetros de qualidade de fibra, e consequentemente seu valor comercial, utiliza-se o instrumento de alto volume (HVI, do inglês *High Volume Instruments*) a partir da qual é possível determinar variáveis como comprimento, grau de reflectância, resistência, índice de fibras curtas, índice de micronaire (GORDON; HSIEH, 2007; GONÇALVES et al., 2016; ZHU et al., 2021; MANIÇOBA, 2019). A qualidade da fibra do algodão também pode ser determinada por fatores como local de plantio, maturidade e presença de modificação genética. Estas características podem influenciar nos valores comerciais e segurança das fibras comercializadas (GORDON, 2006).

Destarte, é significativo o desenvolvimento de metodologias capazes de classificar as fibras de algodão quanto ao seu perfil produtivo e determinar os atributos de qualidade. Algumas tecnologias, como a espectroscopia Raman, vêm se destacando nas pesquisas científicas por serem rápidas na obtenção dos resultados, não necessitar do preparo de amostras e serem economicamente viáveis para análises de produtos agrícolas (XIE et al.,

2007; LIU et al., 2017, ROCHA, 2020). Além disso a miniaturização dos espectrômetros tem popularizado o uso de técnicas espectroscópicas pela facilidade de uso e robustez dos equipamentos.

Diante do exposto, o presente trabalho tem como objetivo verificar o potencial da espectrometria Raman com equipamento portátil aliada a técnicas quimiométricas para classificação de fibras de algodão quanto ao local de plantio, maturidade e modificação genética, como também determinar alguns dos seus atributos qualidade.

2 OBJETIVOS

2.1 Objetivo geral

- ✓ Desenvolver uma metodologia analítica utilizando a espectrometria Raman aliada a técnicas quimiométricas para a classificação e determinação de atributos de qualidade de fibras de algodão.

2.2 Objetivos específicos

- ✓ Registrar os espectros das fibras de algodão na região do visível e infravermelho próximo utilizando um espectrômetro Raman portátil;
- ✓ Realizar o tratamento dos dados utilizando os recursos das análises exploratória através de modelos de PCA;
- ✓ Desenvolver modelos PLS-DA para classificar fibras de algodão quanto a localidade, modificações genéticas e maturidade com dados de atributos de qualidade determinados por HVI;
- ✓ Desenvolver um modelo PLS-DA para classificar fibras de algodão quanto a localidade, modificações genéticas e maturidade com dados espectrais Raman;
- ✓ Comparar a capacidade preditiva dos modelos PLS-DA desenvolvidos utilizando HVI e Raman;
- ✓ Construir um modelo PLS para quantificar os atributos de qualidade das fibras de algodão como comprimento, índice de micronaire, resistência, índice de fibras curtas e grau de reflectância.

3 ESTADO DA ARTE

3.1 Fibras de algodão: contexto histórico, importância econômica e principais aplicações

O algodão é uma fibra natural que está presente na história da humanidade a milhares de anos. Povos do Sul da Arábia, Paquistão, Egito domesticaram e utilizam o algodão desde 3000 a 2000 anos a.C, com destaque para Índia, considerada a principal região de utilização da cultura do algodão (FISHER, 1981). As principais aplicabilidades das fibras de algodão na época consistiram na confecção de cordas para caça e pesca, além da produção de tecidos, como os têxteis, confortáveis contra o frio (KILINÇ; DURMUŞKAHYA; SEYDIBEYOĞLU, 2017).

Com a prática da agricultura tendo se espalhado por diversas regiões do mundo e consequentemente o cultivo e domesticação do algodão, diferentes espécies foram surgindo. Uma das primeiras espécies identificadas foi o *Gossypium arboreum*, cultivado na Índia e Paquistão chegando até as regiões africanas e asiáticas. Com uma genética diferente da natureza do algodão anteriormente citado, o *Gossypium herbaceo* é outra espécie que teve seu cultivo identificado na atual região da Arábia e Síria (KILINÇ; DURMUŞKAHYA; SEYDIBEYOĞLU, 2017).

Kilinç e colaboradores (2017) descrevem que foram identificados diferentes tipos de fibras de algodão cultivadas no continente americano, tendo o *Gossypium Hirsutum* identificado no México e o *Gossypium Barbadosense* nas regiões do Peru e Brasil. Os povos originários das américas foram uns dos principais utilizadores dessas espécies de algodão e empregavam a matéria prima na confecção de redes, para atividades de pescas, além de aplicações artesanais (PASSOS; CANÉCHIO FILHO; JOSÉ, 1973).

O potencial comercial da fibra de algodão é observado nos dias atuais, sendo importante para economia de mais de 80 países do mundo (SIDDIQUI; WANG; MENON, 2020). No Brasil, por exemplo, foram exportados cerca de 2,336 milhões de toneladas no período de agosto de 2020 a julho de 2021, resultando no segundo lugar do ranking de maiores exportadores do mundo. O Estados Unidos segue em primeiro com 3,100 milhões de toneladas de algodão exportados na safra 2020/2021, como é destacado pelo Comitê Consultivo Internacional do Algodão (ICAC, do inglês *International Cotton Advisory committee*) e evidenciado no relatório de safra da Associação Brasileira dos Produtores de Algodão (ABRAPA, 2021).

Apesar da desaceleração causada pela Pandemia da Covid 2019 no período 2019/2020, o comércio de algodão vem se recuperando e exportação mundial do algodão pode ter um aumento de 11,75% para 2021/2022. Na Tabela 01 estão os indicadores e a expectativa para o consumo e comércio global do algodão.

Tabela 01 - Indicadores de produção e exportação mundial de algodão

Indicador	2019/2020	2020/2021	2021/2022
Estoque Inicial	19,32	22,36	20,98
Produção Mundial	26,14	24,22	25,59
Oferta	45,46	46,58	46,58
Consumo	22,75	25,59	25,80
Exportação	9,03	10,09	10,01
Estoque Final	22,36	20,98	20,77

Fonte: ABRAPA (2021), relatório da ICAC. Dados em Milhões de Toneladas

Diante da Tabela 01, percebe-se que há uma perspectiva de aumento de 5% na produção do algodão, evidenciando a importância do algodão para economia de diversos países. Também pode ser destacado a previsão na diminuição dos estoques finais em 2021/2022, devido a menor produção e aumento no consumo, situação que não ocorre desde a safra 2015/2016 (ABRAPA, 2021).

Contudo, até que a fibra de algodão esteja disponível para ser comercializada com os setores industriais, todas as etapas, isto é, do plantio à colheita, podem influenciar na quantidade e qualidade da fibra. Por isso, torna-se importante compreender o desenvolvimento biológico e as composições químicas presentes nas fibras de algodão que impactam na sua qualidade.

3.2 Desenvolvimento biológico, estrutura e composição química da fibra de algodão

O algodão apresenta característica de crescimento de uma planta perene, onde em regiões tropicais e subtropicais pode atingir estaturas de até dois metros. No entanto, na produção comercial são utilizados fitoreguladores para que a altura da planta de algodão não ultrapasse 1,5 m. Ainda sobre a produção comercial, a safra de colheita do algodão é anual, porém a planta não pereniza, pois devido a aplicação de produtos químicos e destruição mecânica o seu tempo de vida é reduzido (FANG, 2018).

O algodão apresenta uma das mais complexas morfologias que corrobora para um crescimento indeterminado. Os ciclos de crescimento do algodão podem variar entre precoce (ciclo produtivo fechando em 130 dias), ou médio tardio (ciclo produtivo

fechando em 170 dias). Assim, a utilização de escala fenológica do algodoeiro é de suma importância para as práticas de manejo necessárias (MARUR; RUANO, 2004).

Portanto, a escala fenológica pode ser dada por: Estádio Vegetativo (V), cuja a relação envolve o crescimento da folha que pode atingir um comprimento de 2,5 cm; Estádio floral (B), que inicia com o surgimento do primeiro botão floral até a primeira transformação do ramo frutífero em flor; Estádio flor (F), trata-se justamente da abertura da primeira flor que pode atingir um estágio de F_n (número de flor do ramo frutífero); Por fim, o estágio Capulho (C), dado pela abertura do primeiro capulho, na qual ocorre o rompimento das chamadas cápsulas (fruto do algodão) (MAURUR; RUANO, 2004; KILINÇ; DURMUŞKAHYA; SEYDIBEYOĞLU, 2017).

As fibras de algodão crescem e amadurecem nas sementes presentes nas cápsulas supramencionadas. A depender do genótipo, pode-se obter entre 10.000 à 20.000 fibras de algodão por sementes (ZHANG, et al 2011). O desenvolvimento das fibras de algodão nas sementes pode ocorrer em quatro estágios sobrepostos sendo a iniciação, biossíntese da parede celular primária (PCW, do inglês *Primary cell wall*) para alongamento, biossíntese da parede celular secundária (SCW, do inglês *Secondary cell wall*) para produção de celulose e maturação (NAITHANI; RAO; SINGH, 1982; FRENCH; KIN, 2018).

3.2.1 Estágio de Iniciação

As fibras de algodão são células não vivas produzidas nas epidermes de sementes localizadas dentro dos frutos. As primeiras evidências de desenvolvimento das fibras iniciais ocorrem entre 01 e 02 dias antes da antese (maturação das flores). Nesse processo, as células epidérmicas do óvulo são distintas das células de fibras e essa diferenciação é dada pela morfologia da célula epidérmicas. Isto é, quando a respectiva célula passa a apresentar uma conformação arredondada e sua superfície começa a projetar-se para fora. Chegando na antese, as fibras iniciais já começam um processo acelerado de alongação na célula epidérmicas do óvulo (ECHER; RESOLEM; RAPHEL, 2018; FANG, 2018).

Esse estágio inicial de desenvolvimento da fibra de algodão tornou-se muito importante para melhorar o rendimento da produção através de biotecnologias (LEE; WOODWARD; CHEN, 2007). Estudos indicam que o número de células epidérmicas do óvulo que se diferenciam das fibras iniciais pode ser um indicador de produtividade do

algodão. Foram observados que uma quantidade maior de fibras iniciais no estágio de iniciação, isto é, antes da antese, resultou em um maior rendimento na produção e também em fibras finas (ZHANG 2011; KIM 2018).

3.2.2 Biossíntese de alongamento da parede celular primária

A depender da espécie de fibra de algodão o alongamento pode iniciar no primeiro dia da antese e estender-se até 20 dias pós-antese (DPA, do inglês Days postanthesis). Outras fibras têm crescimento do comprimento estendido até 25 a 30 DPA. O alongamento ocorre em um crescimento diário de 01 mm até atingir o comprimento máximo, podendo esse variar entre 12 e 65 mm. Esse tamanho final é influenciado tanto pelas condições ambientais de produção, como também pelas características genéticas (ECHER; RESOLEM; RAPHEL, 2018; FANG, 2018; ZHANG et al., 2017).

É no estágio de biossíntese PCW para o alongamento que propriedades como comprimento da fibra de algodão é determinado. Nessa fase de desenvolvimento os principais constituintes da parede celular primária são polímeros de carboidratos (2 a 6,4% de hemicelulose, até 5% de lignina e entre 1 e 7% de pectina) e uma composição inferior a 15% de celulose (FANG, 2018; FRENCH; KIN, 2018). Estudos com espectroscopia Raman identificaram também a presença de componentes de alquil de cadeia longa na região de 2883 a 2933 cm^{-1} e hidrocarbonetos insaturados em 1614 e 1660 cm^{-1} no período de 10 DPA (KIM et al., 2017).

3.2.3 Biossíntese da celulose na parede celular secundária

Com aproximadamente 12 a 16 DPA a biossíntese de SCW inicia e pode ir até mais de 35 DPA (FANG, 2018). Nesse estágio o processo de alongamento passa a ser mais lento, enquanto que o crescimento da espessura da parede celular passa a predominar. Esse crescimento é ocasionado pela deposição de celulose em camadas internas concêntricas (ECHER; RESOLEM; RAPHEL, 2018). Com o cessamento da biossíntese de outros polímeros da parede celular durante a transição PCW para SCW a síntese de celulose é altamente amplificada podendo constituir entre 90 a 96% do peso da fibra de algodão (WAKELIN, 2006).

No estágio de biossíntese SCW, o componente de celulose sintetizado passa exercer um papel importante em algumas propriedades das fibras de algodão, como por exemplo, a resistência, a maturidade e a finura (KIM, 2018). A celulose, cuja a estrutura é β -Polímero linear de 1,4-d-glucopiranoose, formam um agregado chamado de microfibrilas de celulose (CMF, do inglês *Cellulose microfibrils*). Essas CMF são depositadas helicoidalmente formando camadas no eixo longitudinal da fibra. Periodicamente ocorrem reversões, ou seja, a direção de deslocamento helicoidal das CMF é alterada (WAKELIN, 2006; FANG, 2018).

Com isso, quando o ângulo de orientação da CMF em relação ao eixo longitudinal é relativo a fibra é mais resistente. Já com um ângulo de orientação maior a fibra é menos resistente. Assim como em outras células vegetais, a orientação da CMF é regida pelo citoesqueleto. Esse é constituído por estruturas filamentosas denominadas de microtúbulos e microfilamentos (WAKELIN, 2006; FANG, 2018).

Outro aspecto que afeta a resistência da fibra de algodão no estágio de biossíntese da SCW é o grau de polimerização da celulose, isto é, o número de moléculas de glicose por polímero. Deste modo, conforme esse grau de polimerização aumenta na parede celular secundária, também aumenta a resistência da fibra de algodão (TIMPA; RAMEY, 1994). Em um outro tópico será detalhada a estrutura química da celulose.

3.2.4 Processo de Maturação

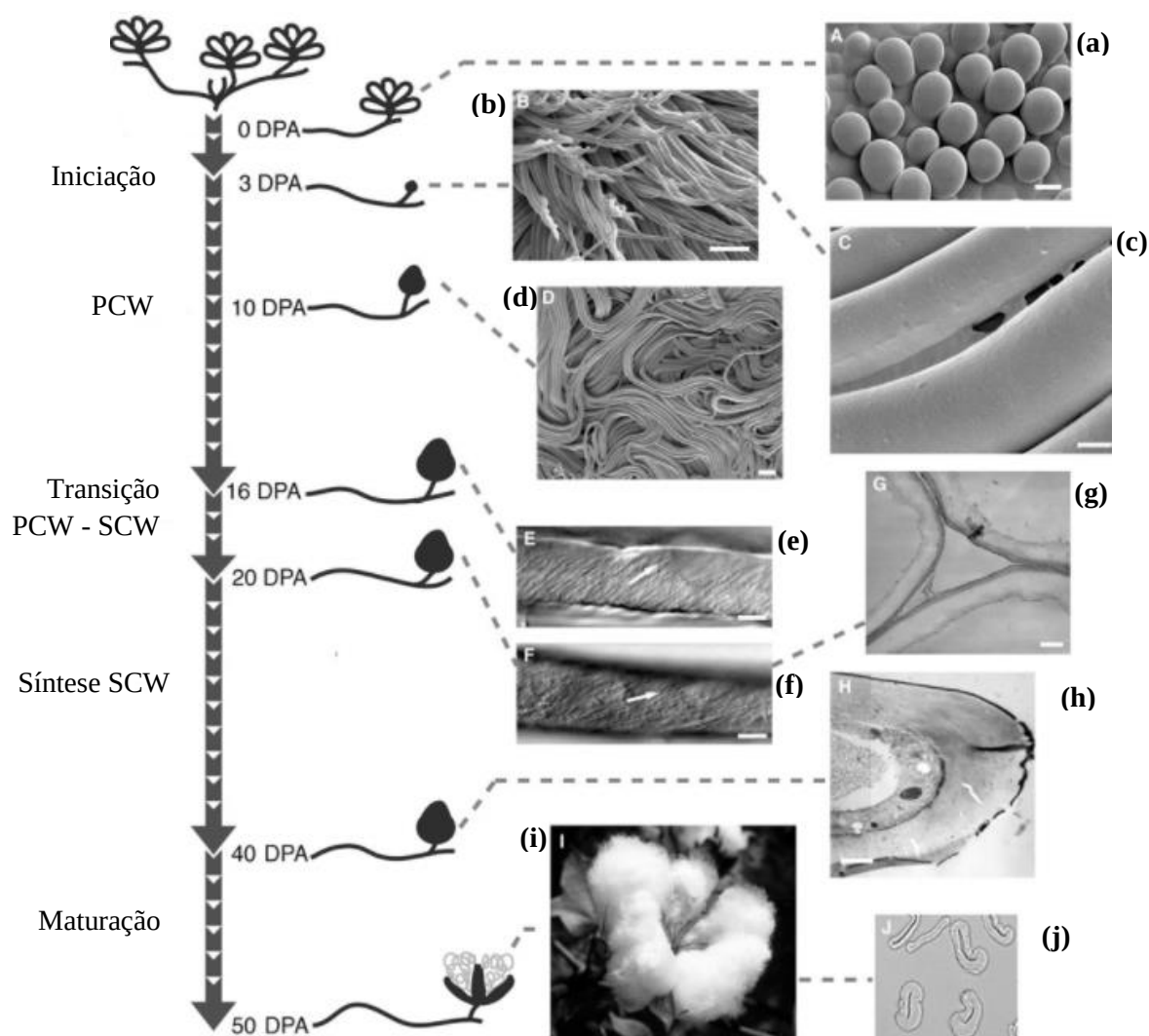
De modo geral, a maturação das fibras de algodão prolonga-se entre 45 e 75 dias após a fecundação. Esse período de maturação é influenciado pelas condições ambientais de produção, como também o genótipo do algodão (ECHER; RESOLEM; RAPHEL, 2018). Um exemplo disso é observado no algodão *herbáceo* colocado sobre estresse, na qual apresentou um desenvolvimento atrasado apresentando um período de maturação pronlogado (BRADOW; DAVIDONIS, 2000).

A maturidade da fibra de algodão é determinada pelo espessamento da parede celular. Deste modo, a identificação da maturidade passa ser importante para os setores industriais têxteis, pois fibras imaturas são facilmente quebráveis e isso corrobora para o surgimento de *neps* (espessamentos nodulares irregulares). Os *neps*, por sua vez, prejudicam etapas importantes como a fiação e tingimento. Em consequência, os produtos

finais acabam perdendo a qualidade e tornam-se desvalorizados (BACHELIER; GOURLOT, 2018; ECHER; RESOLEM; RAPHEL, 2018).

Diante disso, para fomentar a compreensão sobre o estágio de maturação, assim como os demais estágios de desenvolvimento da fibra de algodão pode ser observado na Figura 01 imagens microscópicas de cada estágio DPA das fibras de algodão.

Figura 01 - Representação da progressão de desenvolvimento da fibra de algodão



Fonte: Haigler et al (2012). Fibras iniciais (a); Torção e alongamento da fibra (b); CFM esticada entre duas fibras (c); Feixe de fibras ordenadas dentro da cápsula (d); Indicação do ângulo da microfibrila (e), (f); Estágio inicial de espessamento da parede secundária (g); Estágio avançado do espessamento da parede secundária (h); cápsula de algodão maduro (i); Seção transversal da fibra madura vista por um microscópio e luz (j).

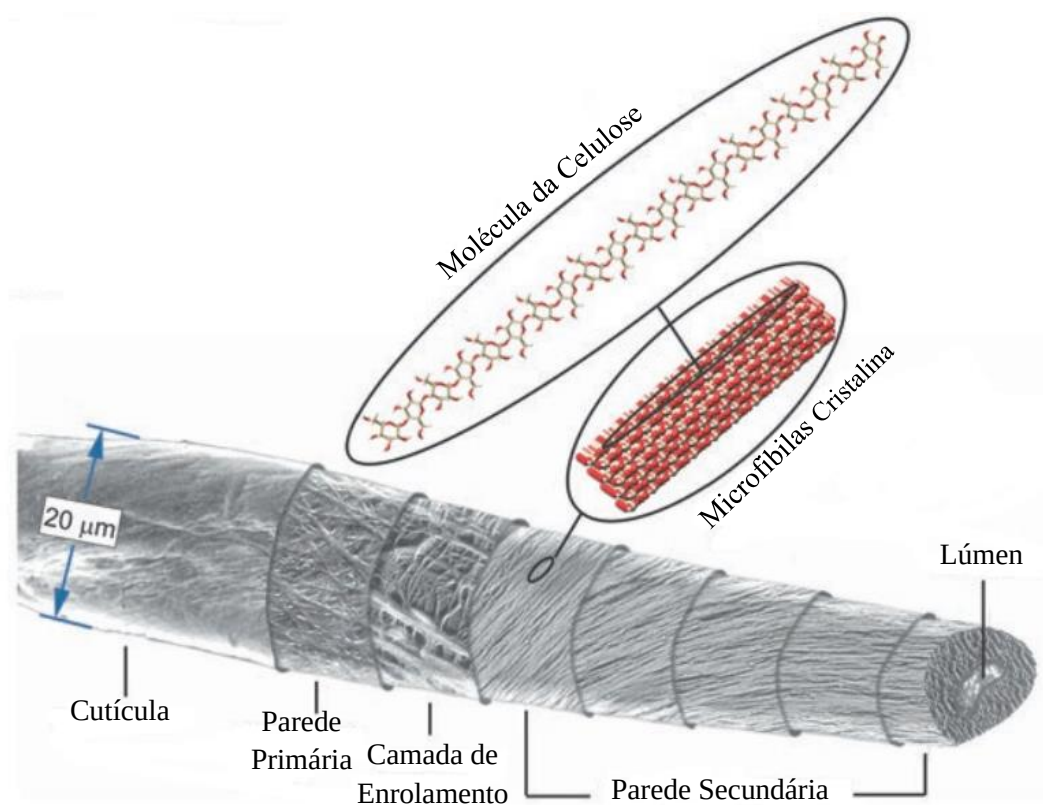
Como observado na Figura 01 e discutido anteriormente, a fibra de algodão passa por diversos estágios até atingir a maturação final para realização da colheita. Cada estágio corrobora para o desenvolvimento estrutural da fibra de algodão, como também a

composição química constituinte. Deste modo, compreender a estrutura da fibra de algodão, assim como a composição química é fundamental para o entendimento dos atributos que as fibras de algodão apresentam como: Comprimento, resistência, índice de micronaire, índice de fibras curtas, refletância entre outras.

3.2.5 Estrutura da Fibra de algodão

Os estágios de desenvolvimento da fibra de algodão são classificados de acordo com as propriedades físicas, conteúdo de celulose, espessura da parede celular, componentes químicos, cristalinidade e conjunto de microfibrilas. Com isso, a compreensão de cada estágio de desenvolvimento da fibra de algodão é dada pelo conhecimento da sua estrutura, podendo essa ser observada na Figura 02:

Figura 02 - Estrutura da fibra de algodão



Fonte: French e Kin (2018)

Na Figura 02, as dimensões observadas referenciam-se a fibra de algodão *G. hirsutum* que pode apresentar diferenças entre outras espécies a respeito de diâmetro e

comprimento. De modo geral, a fibra é tubular com uma base larga e achatada nas extremidades a formação é dada pela deposição de celuloses em camadas resultando em uma conformação final espiralada. Como pode ser observado, a fibra de algodão é constituída por diversas partes cuja formação, dada por diversos processos de biossíntese ocorridos ao longo do seu desenvolvimento (WAKELIN, 2006; ECHER; RESOLEM; RAPHEL, 2018).

Cada constituinte da estrutura da fibra de algodão apresenta funções bioquímicas e biológicas e composições em quantidades específicas. A cutícula, situada na parte mais externa, proporciona proteção à fibra após a abertura do capulho, no armazenamento e também nos processos industriais. Os constituintes que podem serem identificados na cutícula são cera, pectinas, óleos entre outros (SANTANA; WNADERLEY; BELTRÃO, 1998).

A parede primária é outra estrutura importante na fibra de algodão. Uma das primeiras a ser formada após a maturação, a parede primária possui uma espessura entre 01 a 02 μm . A parede primária fica delimitada externamente com a cutícula (cama externa que cobre a parede primaria) e internamente com a parede secundária. A principal função exercida é sobre o comprimento e finura da fibra de algodão e em algumas propriedades intrínsecas finais (BACHELIER; GOURLOT, 2018; ECHER; RESOLEM; RAPHEL, 2018).

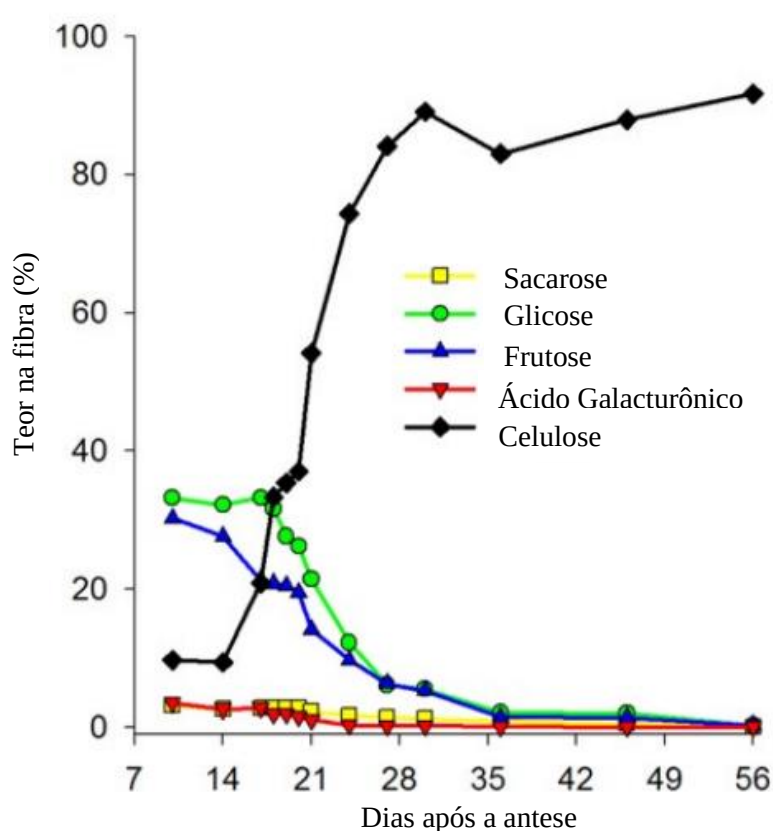
Uma das mais importantes partes estruturais da fibra de algodão é a parede secundária. Nela está presente o componente com maior quantidade, a celulose. A parede secundária acondiciona três camadas concêntricas sendo uma externa, uma interna e uma média formando uma espiral. Dentre as camadas concêntricas, a camada média apresenta uma espessura maior devido ser constituída essencialmente por celulose (BACHELIER; GOURLOT, 2018).

Por fim, outra parte importante que compõe a estrutura da fibra de algodão é o lúmen. Formado durante a maturação, o lúmen é um canal central constituído de resíduos proteicos da célula. A estrutura final do lúmen, isto é, seu achatamento, irá depender da quantidade de espessura da celulose depositada; em fibras madura essa deposição de celulose chega a 5 μm de espessura, conferindo a ela uma estrutura final circular, devido à redução no diâmetro e no achatamento do lúmen (BACHELIER; GOURLOT, 2018; ECHER; RESOLEM; RAPHEL, 2018).

3.2.6 Composição Química da Fibra de Algodão

A fibra de algodão é majoritariamente composta por celulose, podendo constituir até 96% na composição química total e final da fibra (WAKELYN, 2006). Como pode ser observado na Figura 03, a composição de celulose na fibra de algodão vai aumentando gradativamente durante a sua formação. Em contrapartida, alguns constituintes (açúcares redutores) que inicialmente estão em maior quantidade vão diminuindo, como é o caso da glicose e da frutose.

Figura 03 - Teores de celulose e açúcares redutores de fibra de algodão durante a sua formação.



Fonte: Echer, Rosolen e Raphael (2018), adaptado de Abidi, Hequet e Cabrales (2010)

Analisando a Figura 03, percebe-se que diferentemente dos demais constituintes a sacarose e o ácido galacturônico tem um teor durante todo desenvolvimento da fibra de algodão. Apesar disso, a sacarose tem impactos importantes na propriedade da fibra de algodão. Como enzima, a sacarose sintase (SS) pode atuar na síntese e hidrólise da sacarose, além de apresentar influência significativa nos processos iniciais de formação do fruto de uma planta (BASSINELLO et al., 1999). Nesse sentido, pode ser observado o envolvimento da SS nos processos de alongamento da fibra de algodão, em

consequência, influencias sobre a propriedade de comprimento da fibra de algodão (RUAN; LLEWELLYN; FURNANK, 2003).

Os teores de proteínas e açúcares redutores são maiores nas primeiras DPA, mas em decorrências de reações de polimerização de sacarídeos simples, como a glicose e frutose, o teor vai diminuindo (ABIDI; HEQUET; CABRALES, 2010). Devido a essas reações e o acúmulo de celulose na parede secundária, o aumento do teor da celulose é acelerado após a 16 DPA (Figura 03). Por outro lado, os outros constituintes presentes passam a diminuir, principalmente entre 14 e 16 DPA, justamente o período de formação da SCW (KIM, 2018). Após o processo de maturação, a fibra de algodão fica constituída basicamente de celulose, enquanto que os açúcares redutores apresentam-se com teores muito baixos.

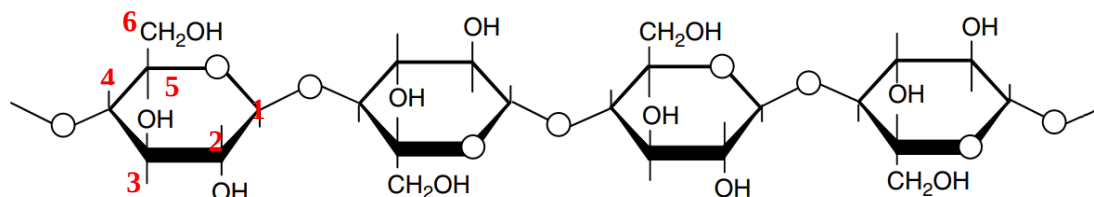
Com isso, tendo em vista os altos teores de celulose presentes nas fibras de algodão, é de suma importância uma discussão detalhada sobre sua estrutura química e compreender seu envolvimento sobre algumas propriedades referentes as fibras de algodão.

3.2.6.1 Estrutura Química da Celulose

A celulose é uma das matérias orgânicas mais abundantes na terra, sendo um dos constituintes mais importante da parede celular de um vegetal, como o algodão. Trata-se de um polissacarídeo constituído por unidades lineares de β -glutapirranose ligados covalentemente por 1,4 glicosídicas. A fórmula química da celulose apresenta-se como $(C_6H_{10}O_5)_n$, cuja o comprimento final irá depender do grau de polimerização (GP), isto é, do número de monômeros de glicose repetidas. Essa quantidade varia de acordo com a cultivar e o tratamento, podendo chegar a um GP de 8 mil a 15 mil (BACHELIER; GOURLOT, 2018; ABIDI, 2018).

Em outras palavras, a celulose é constituída unicamente de uma celobiose, ou seja, da união entre duas moléculas de glicose por uma ligação glicosídica formando uma macromolécula de celulose (COCINERO et al., 2009). Na macromolécula de celulose, cada unidade da glicose gira em um ângulo de 180° em torno do eixo molecular, na qual cada uma delas possuem três grupos hidroxilas nas posições C1, C2 e C3 (ABIDI, 2018). A estrutura química da macromolécula de celulose pode ser observada na Figura 04.

Figura 04 - Estrutura Química de quatro resíduos da cadeia de celulose



Fonte: Wakelyn (2006)

Como pode ser observado na Figura 04, a celulose apresenta uma estrutura homogênea com a presença de muitos grupos OH livres. As posições das hidroxilas em cada molécula da glicose são classificadas em primária (hidroxila do C-6) e secundária (hidroxila do C-2 e C-3). A presença das hidroxilas possibilita que as moléculas de celuloses se conectem através das ligações de hidrogênio intermolecular e intramolecular. A conexão dessas moléculas ocorre paralelamente na forma de folheto, conferindo a boa parte da celulose um arranjo cristalino ordenado (BACHELIER; GOURLOT, 2018; ABIDI, 2018).

Esses arranjos cristalinos formam agregados maiores denominados de fibrilas. As fibrilas, unidade cristalina básica, apresenta um diâmetro de aproximadamente 3,5 nm formado por cerca de 1,5 a 2 mil moléculas de celulose. Um agrupamento das fibras resulta em microfibrilas com diâmetros variando entre 10 a 30 nm, construindo assim a fibra de algodão (BACHELIER; GOURLOT, 2008).

Essa disposição cristalina ordenada, apresenta regiões com cadeias celulósicas distribuídas aleatoriamente, chamadas de regiões amorfas. Com isso, a celulose pode ser identificada tanto na forma cristalina quanto na forma amorfa. Na fibra de algodão, 70 a 80% da celulose apresenta-se na forma cristalina e 20% a 30% na forma amorfa. Com isso, propriedades como resistência da fibra estará relacionada com a forma cristalina da celulose, enquanto que a elasticidade, flexibilidade e capacidade de absorção de água, corantes e acabamentos químicos estão relacionadas a forma amorfa da celulose (ABIDI, 2018).

3.2.6.2 Outros constituintes da fibra de algodão

Além da celulose e dos açúcares redutores abordados até aqui, a fibra de algodão apresenta outras substâncias típicas. Os constituintes não celulósicos que podem ser

identificados na fibra de algodão são proteínas, aminoácidos, cera, ácidos orgânicos, compostos contendo nitrogênio, substâncias pécicas e alguns pigmentos em quantidades muito pequenas. Os teores desses constituintes podem ser observados na Tabela 02, que esquematiza a composição química em porcentagem de amostras secas.

Tabela 02 - Composição típica da fibra de algodão

Constituintes	Composição (% em amostra seca)	
	%Típica	%Alcançada
Celulose	95,0	88,0 a 96,0
Proteína (% N x 6,25)	1,3	1,1 a 1,9
Substâncias Pécicas	0,9	0,7 a 1,2
Cinza	1,2	0,7 a 1,6
Cera	0,6	0,4 a 1,0
Açucares Totais	0,3	0,1 a 1,0
Ácidos Orgânicos	0,8	0,5 a 1,0
Traço de Pigmento	Vestígios	-
Outras	1,4	-

Fonte: Wakelyn (2006).

Como observado na Tabela 02, a celulose pode corresponder até 96% da composição da fibra de algodão. Entre os constituintes não celulósicos, os compostos com a presença de Nitrogênio apresentam a maior porcentagem, quando esse é expresso em porcentagem de proteína (WARWICKER et al., 1996). A maior parte de nitrogênio é identificado no lúmen como protoplasmático, mas também pode ser identificada na parede primária da fibra. Os demais compostos irão aparecer em pequenas quantidades, mas podem influenciar alguns aspectos da fibra de algodão (WAKELYN, 2006). A cera por exemplo, é importante no processo de fiação atuando como uma espécie de lubrificante. Em contrapartida, a cera pode diminuir a resistência a tração da fibra, além de dificultar o tingimento e acabamento da fibra de algodão (GONÇALVES et al., 2016).

Com relação as substâncias pectinas, que se apresentam como um polímero linear livre, o ácido (1,4) -D-galacturônico e sais insolúveis de Cálcio, Magnésio e Ferro. Esses compostos apresentam entre 0,7% e 1,2% do peso da fibra seca, sendo identificados principalmente na camada mais externa da parede celular primária da fibra de algodão

(ABIDID; HEQUET; CABRALES, 2010). Apesar de existirem métodos para remoção dessas substâncias, os efeitos causados na resistência a trações e algumas propriedades do fio e do tecido não são significativos (WAKELYN, 2006).

Ácidos orgânicos presentes, em pequena quantidade, na fibra do algodão são: O ácido 1-málico com um teor de 0,5% e o ácido cítrico com até 0,07% do peso da fibra seca. Esses ácidos são identificados na região do lúmen da fibra de algodão como resíduos metabólicos. Por serem solúveis em água, a remoção desses ácidos pode ser feita apenas por meio de uma lavagem (WAKELYN, 2006).

3.2.6.3 Teores de metais na fibra de algodão

As fibras de algodão podem apresentar metais na sua composição. A incorporação de metais no algodão é normalmente proveniente de sais minerais, pois, assim como todo processo de crescimento de uma planta, Potássio e outros metais são absorvidos como forma de nutrientes (ECHER; RESOLEM; RAPHEL, 2018). A concentração desses metais varia de acordo com a maturidade da planta, as condições ambientais, a prática agrícola, em especial a colheita na qual pode ser depositado metais diretamente na fibra. A ausência desses metais pode prejudicar o desenvolvimento da fibra e consequentemente afetar a qualidade final (CONSTABLE; BANGE, 2007).

Por outro lado, o excesso desses metais também pode afetar a qualidade da fibra, principalmente no que se refere ao grau de amarelamento tanto do algodão bruto como do fio acabado (CARVALHO et al., 2000; VITERO et al., 2017). Outros problemas que podem ser identificados pela presença de metais na fibra de algodão é a fiação no processamento têxtil e o desgaste da agulha na tricotagem. A presença dos metais também causa preocupação ambiental, pois durante a produção comercial dos tecidos podem ser gerados resíduos sólidos e líquidos poluentes (BRUSHWOOD, 2000). Na Tabela 03 podem ser identificados teores de metais possivelmente presentes nas fibras de algodão.

Tabela 03 - Teor de metal no algodão

METAIS	ppm^a
Potássio (K)	2000 - 6500
Magnésio (Mg)	400 - 1200
Cálcio (Ca)	400 - 1200
Sódio (Na)	100 - 300
Ferro (Fe)	30 - 90
Manganês (Mn)	01 - 10
Cobre (Cu)	01 - 10
Zinco (Zn)	01 - 10

Fonte: Brushwood e Perkins (1994). (a) Partes por milhão

Como pode ser identificado na Tabela 03, o Potássio é o cátion com maior abundância na fibra de algodão, correspondendo 70% do conteúdo catiônico. A sua importância no desenvolvimento da fibra de algodão está relacionada com a capacidade fotossintética de síntese e transporte de carboidratos. Baixos teores de K podem afetar a deposição de microfibras de celulose na SCW, pois dificulta a aquisição de matéria prima para formação da fibra. Nesse sentido, pode comprometer, principalmente, propriedades como maturidade e micronaire, além do comprimento, resistência e a uniformidade (TEMIZ et al., 2009; YANG et al., 2016; KAPPES et al., 2016).

Outros metais com abundância na fibra de algodão é o Cálcio e Magnésio, correspondendo respectivamente 14% de peso entre os cátions presentes. Assim como o K, o Ca e Mg constituem o conteúdo de cinza como sais inorgânicos. Na fibra de algodão, o Cálcio pode ser identificado no lúmen e na fração de pectina, atuando como agente de reticulação. As influências que os metais Ca e Mg pode apresentar sobre a fibra de algodão são no processo de tingimento, o branqueamento com peróxido pode ser afetado pelos sais de Mg. De modo geral, esses metais não impactam diretamente na qualidade da fibra, mas a sua ausência pode afetar o desenvolvimento metabólico da planta de algodão, corroborando uma baixa produção (STAUT; KURIHARA, 1998; WAKELYN, 2006)

Em relação aos demais metais, Ferro, Sódio, Manganês, Cobre e Zinco são identificados na fibra de algodão em baixas quantidades. Eles podem causar efeitos nas propriedades de reflectância da fibra de algodão e problemas ambientais no processo de produção. O Fe e Cu podem contribuir para um alto grau de amarelamento, o Fe, em

especial, contribui para uma cor permanente entre marrom e rosa, afetando o processo de tingimento. No caso do Zn e Cu além de afetarem o tingimento têxtil, apresentam uma preocupação com os efluentes gerados com a presença desses metais (BRUSHWOOD, 2000; WAKELYN, 2006).

Diante disso, percebe-se que a composição química da fibra de algodão, assim como todo processo de desenvolvimento pode impactar na qualidade da final da fibra. Deste modo, torna-se importante compreender as atribuições que as propriedades físicas e químicas apresentam na determinação da qualidade da fibra de algodão, e consequentemente os valores comerciais empregado sobre as fibras de algodão.

3.3 Atributos de qualidade das fibras de algodão

Órgãos como o Departamento de Agricultura dos Estados Unidos (USDA, do inglês *United States Department of Agriculture*) estabelecem faixa de valores para os atributos de qualidade das fibras de algodões. Com isso, para que as fibras de algodão proporcionem a fabricação de produtos têxteis de qualidade e/ou qualquer outro produto que esteja relacionado à sua aplicabilidade, é indispensável que as medidas dos atributos estejam próximas ou iguais aos estabelecidos por instituições especializadas nas análises de qualidade da fibra de algodão (SIDDIQUI; WANG; MENON, 2020).

Assim, torna-se indispensável avaliar os parâmetros de qualidade das fibras de algodão, pois as fibras apresentam um significativo potencial econômico e o seu valor comercial está diretamente relacionado com suas características de qualidade (GONÇALVES et al., 2016). A USDA destaca algumas propriedades das fibras de algodão que podem ser avaliados para determinar sua qualidade: Comprimento; Índice de Micronaire; Resistência Específica; Grau de Refletância; Índice de Fibras Curtas.

3.3.1 Comprimento

As fibras de algodão passam por um processo de desenvolvimento que envolve quatro diferentes etapas que se sobrepõem: Iniciação; alongamento (comprimento); espessamento da parede celular secundária e maturação. Diferentes fatores durante o desenvolvimento da fibra de algodão podem afetar a qualidade total. Um dos fatores é o tipo de genótipo e o ambiente, pois estão relacionados com o tempo de cada etapa do

desenvolvimento da fibra de algodão. Assim, o comprimento final da fibra de algodão, será determinado pela taxa e tempo de alongamento do desenvolvimento das células de fibras de algodão (ZHU et al., 2021).

O alongamento das fibras algodão consiste em um “processo fisiológico e bioquímico complexo que envolve dilatação vacuolar, relaxamento da parede celular, biossíntese e transporte de lipídios da membrana e componentes da parede celular” (ZHU et al., 2021, p. 01). Logo, esses fatores podem afetar o comprimento das fibras de algodão e comprometer sua qualidade.

A classificação como fibra de comprimento médio da metade superior (UHM, do inglês *Upper Half Mean Length*), corresponde a valores de 1,3-1,6 de polegadas equivalentes a 29,8-30,5 mm de comprimento. Já as fibras de algodão classificadas como extralargas pode apresentar valores na faixa de 1,7 a 2,0 e isso é proporcional a 34,8 a 36,1 mm de comprimentos (COSTA et al., 2006).

Uma das técnicas utilizadas para realizar a medida do comprimento de fibras de algodão é o instrumento de alto volume. As medidas de comprimento da fibra de algodão por HVI leva em consideração o conceito de fibrograma, na qual relaciona-se os comprimentos das fibras em função da frequência da amostra (FONSECA; SANTANA, 2002). Assim, gera-se uma curva e medindo os valores em UHM utilização do HVI, que também fornecerá informações sobre o Índice de Uniformidade (UI, do inglês *uniformity index*) (SAYEED et al., 2021).

3.3.2 Índice de Fibras Curtas

O índice de fibras curtas (%SFI, do inglês *short fiber index*) é um importante atributo para a indústria têxtil na avaliação da qualidade das fibras de algodão. o parâmetro leva em consideração a porcentagem de fibras com comprimentos inferiores a 0,50 polegadas presentes em determinada amostra (BACHELIER; GOURLOT, 2018). De acordo com Maniçoba (2019), valores baixos desse parâmetro (%SFI) confere a fibra de algodão uma significativa capacidade de fabricação do fio, pois, a resistência à torção e estiramento no processo de fiação tornam-se maiores.

Assim como no comprimento, o genótipo do algodão também influenciará nos valores dos índices de fibras curtas. Também, poderá impactar no %SFI o processo de cultivo do algodão como, por exemplo, a disponibilidade hídrica para irrigação

(BRADOW; DAVIDONIS, 2000). Com isso, o mercado avalia como favorável a compra da fibra de algodão quando os valores do índice de fibra curta estão abaixo de 10%, (LANA; RIBAS; CHASELME, 2014; BRADOW; DAVIDONIS, 2000).

A medição do índice de fibras curtas pode ser feita pelo instrumento HVI ou utilizando o instrumento Uster AFIS. No primeiro, o processo envolve ensaios em feixes, enquanto que no Uster AFIS os ensaios são individualizados. O %SFI pode ser classificado em: Muito baixo (<6%); Baixo (6 a 9%); Médio (10 a 13%); Alto (14 a 17%); muito alto (>18%). Essas referências são de acordo com o algodão americano *Upland* (BACHELIER; GOURLOT, 2018).

É importante destacar que ao analisar o %SFI também é preciso considerar o índice de uniformidade. Considerando que determinada fibra apresente um baixo valor de uniformidade e encontra partida tenha um alto %SFI, isso poderia comprometer o desempenho do processo de fiação, pois favorece as fibrilas flutuarem, assim como as impurezas nos setores de estocagem proporcionando a acumulação de pó e micropó comprometendo a qualidade do produto têxtil fabricado (MARTINS, 2020).

3.3.3 Resistência Específica

Também denominada como tenacidade a rotura de fibra (gf.tex^{-1}), a resistência específica da fibra de algodão está relacionada com a força necessária para romper um feixe de fibras de 01 tex, que é definido como a massa expressa em grama de mil metros de material têxtil que está na forma de fibras, mantas, fitas, pavios e fios (BACHELIER; GOURLOT, 2018).

As medidas de resistência específica podem ser feitas através de ensaios nos instrumentos HVI, na qual as fibras de algodão são inseridas no fibrógrafo aleatoriamente, na sequência, elas são presas pelas garras e tracionadas até o rompimento, com isso, mede-se a força máxima de rotura (*Max Force*). De acordo com Bachelier e Gourlot (2018) essa medição consiste em determinar a força necessária para romper o feixe de fibras amostrado.

A tenacidade à rotura de uma fibra de algodão pode variar entre valores inferiores a 21 gf.tex^{-1} , sendo assim classificadas como resistência muito fraca até maiores do que 31 gf.tex^{-1} consideradas como muito forte (BACHELIER; GOURLOT 2018). Fibras de algodão com resistência forte são bem avaliadas pelas indústrias têxteis, pois a capacidade

de torção das fibras está relacionada com esse atributo. As fibras devem apresentar capacidade suficiente para suportar forças mecânicas exercidas em etapas como beneficiamento (MARTINS, 2020).

3.3.4 Índice de Micronaire

Para que a fibra de algodão possibilite o desenvolvimento de produtos qualificados e eficientes, os índices de alguns parâmetros de qualidade precisam estar em uma faixa de valor próximo ou igual aos padrões estabelecidos. A determinação do índice de micronaire (%MIC, do inglês *micronaire index*), por exemplo, é importante para compreensão sobre a qualidade da fibra de algodão, pois influenciam na resistência, na uniformidade de massa dos fios, no índice de pilosidade dos fios e tecidos e no tingimento de fibras, fios, linhas e tecidos (LIMA, 2018).

Lima (2018) ressalta os cuidados com o processamento da fibra de algodão que apresente um baixo índice de micronaire. Uma má execução desse processo pode levar a formação de *neps* (defeitos identificados nas amostras de fibras de algodão), reduzir a uniformidade do comprimento e aumentar o conteúdo de fibras curtas (BACHELIER; GOURLOT, 2018). Portanto, ainda que apresente baixo valor no índice de micronaire, mas com porcentagem significativa de maturidade, resistência, alongamento e sendo processada delicadamente, a produtividade ainda será significativa.

O índice de micronaire é adimensional e está relacionada com a maturidade e complexo de finura da fibra (MARTINS, 2020). Na determinação do índice de micronaire analisa-se apenas um corpo de prova, na qual é detalhado por Bachelier e Gurlot (2018, p. 67):

A massa de fibras do corpo de prova é inserida na câmara do aparelho micronaire, onde é submetida a uma pressão de ar comprimido. Os valores da massa do corpo de prova, do volume da câmara e da pressão do ar comprimido injetado são constantes. Em seguida, o aparelho determina a vazão do ar que passou pela massa de fibras, que é variável e que está em função do diâmetro externo e da espessura da parede secundária das fibras, formada por camadas de celulose, e que também estão relacionados diretamente com a maturidade das fibras.

De acordo com o foi supracitado, esta análise irá fornecer valores que determinam a espessura da fibra de algodão, de modo que estas podem ser classificadas, para esse parâmetro, como: Muito fina, em que os índices de microanire são abaixo de 3,0 %MIC; Fina, variando entre 3,0 a 3,6 %MIC; Média, quando o índice fica entre 3,7 e 4,7; Grossa,

apresentando índices na faixa de 4,8 a 5,4; Em índices superior ou igual a 5,5 %MIC as fibras são classificadas como muito grossa (USTER, 2002).

3.3.5 Grau de Reflectância

O grau de reflectância (%Rd, do inglês *reflectance degree*) é um dos atributos relacionados com a determinação da cor da fibra do algodão, avaliando o brilho ou a opacidade, isto é, o grau de cinza de uma determinada amostra (WANG; FAROOQ; MENON, 2020). Deste modo, o grau de reflectância passa a ser importante na classificação do algodão, tendo em vista que a cor afeta a capacidade de tingimento e acabamento da fibra de algodão e consequentemente influenciando no valor econômico (VIK et al., 2017).

Sendo assim, o %Rd é dado a partir dos valores de luminosidade e do reflexo da cor branca emitida pelas fibras de algodão. Uma das formas de obter esse percentual é através do calorímetro de Nickerson e Hunter. Esse método foi desenvolvido em 1940, para checagem de padrões da USDA, que apresenta um diagrama com uma escala de cores. Apesar de o método mencionado ter sido desenvolvido para algodão *Upland*, pode-se aplicá-lo para medições em plumas equivalentes (BACHELIER; GOURLOT 2018).

A análise é feita em um aparelho com capacidade de correlacionar sua leitura com a visual, possibilitando reproduzir padrões físicos oficiais para o tipo de cor do algodão. O procedimento para registrar o %Rd consiste em substituir a amostra de algodão por uma superfície lisa. posteriormente, a superfície é inserida e pressionada no visor do calorímetro e em seguida são obtidos os valores registrados (BACHELIER; GOURLOT 2018).

Com a determinação do grau de reflectância juntamente com índice de amarelamento é possível identificar o grau de cor das fibras de algodão. Isso possibilita um melhor monitoramento na preparação de fios e tecidos, pois conhecendo o grau de cor é possível neutralizar a variação de cor através do controle de uniformidade dos fardos que alimentam a linha de abertura, da limpeza, da mistura e cardagem das fibras durante a preparação para a fiação (LIMA, 2018).

Como observado na discussão ao longo desse tópico, para determinar a qualidade da fibra de algodão é preciso avaliar alguns parâmetros. Em muitos casos a análise desses parâmetros demanda tempo, gastos econômicos, impossibilidade de análises em campo

entre outros empecilhos. Deste modo, pesquisas científicas com tal proposta são realizadas com intuito de melhorar as técnicas analíticas no sentido de tornar o processo de análise mais rápido, eficaz, com menor custo.

3.4 A espectroscopia Raman aplicada nas análises de fibras de algodão

A espectroscopia Raman recebe esse nome em homenagem ao descobridor do fenômeno de difusão inelástica ou espalhamento Raman, o físico Chandrasekhara Vencata Raman em 1928. Dois anos mais tarde o professor Raman recebeu o prêmio Nobel de Física pela descoberta. Esse fenômeno, até então desconhecido, trata-se de um espalhamento disperso inelasticamente ocasionado pela interação entre os fótons incidentes em uma determinada amostra e as ligações químicas das moléculas presentes na amostra, resultando na medida dos níveis de energia vibracionais das espécies analisadas (SHIPP; SINJAB; NOTHINGER, 2017; KUHAR et al., 2018; DELFINO et al., 2019).

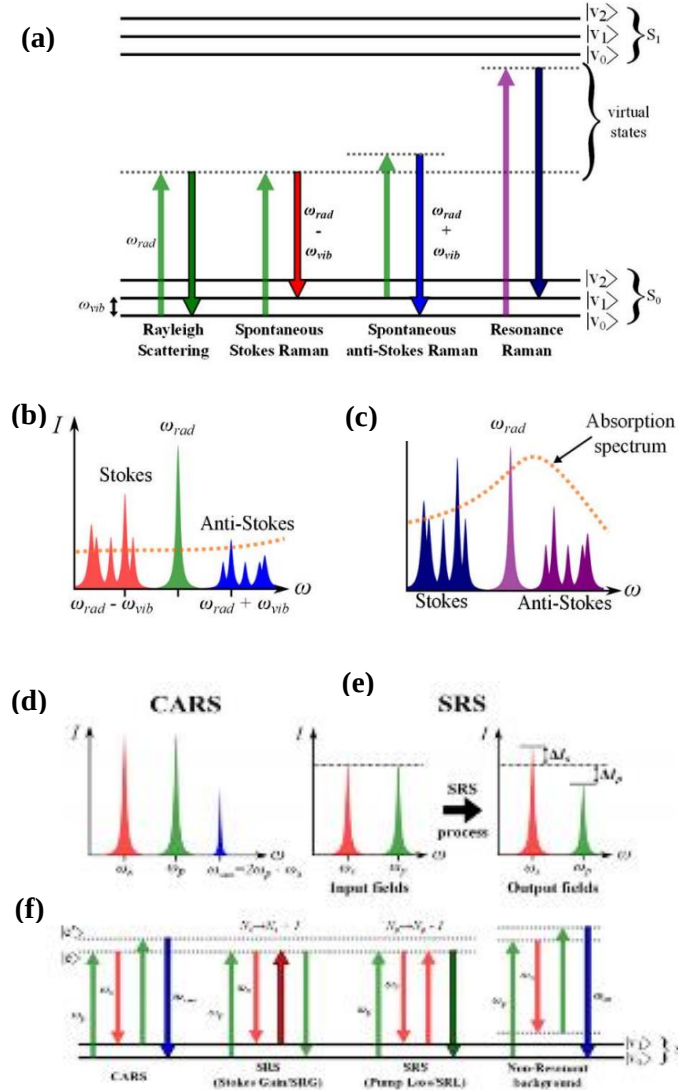
O espectro Raman é obtido após a excitação de uma amostra por um laser de alta intensidade. Deste modo, as informações sobre grupos funcionais presentes nas amostras estão relacionadas ao deslocamento inelástico que é medido por meio de um deslocamento em relação ao feixe de alta frequência incidido. Assim, obtém a impressão digital molecular, pois cada espécie apresenta um padrão de deslocamento energético entre fótons incidentes e refletidos (DELFINO et al., 2019).

Inicialmente, as aplicações da espectroscopia Raman tinham certa limitação devido ao sinal de espalhamento ser fraco, sendo necessário a excitação do sinal através de uma luz monocromática no visível ou infravermelho próxima (SHIPP; SINJAB; NOTHINGER, 2017). Esse problema envolvendo a análise Raman foi solucionada com o aumento da potência das fontes de radiação e detectores como os lasers de diodo e dispositivos de cargas acopladas (CCD's, do inglês *charge coupled devices*), ou seja, pequenos espectrômetros que foram utilizados para direcionar o espalhamento Raman e consequentemente tornar possível o melhoramento na detecção do sinal (KUHAR et al., 2018; FAKAYODE et al., 2020).

Com seu aprimoramento, a espectrometria Raman passou a ser utilizada em diversas áreas como a farmacêutica, ciências forenses, meio ambiente, produtos alimentícios, produtos agrícolas, podendo assim incluir a fibra de algodão (GAUTAM, et

al., 2015). Na Figura 05 observa-se esquematização de como ocorre o fenômeno do espalhamento Raman e o melhoramento para a detecção do sinal.

Figura 05 - Diagrama de nível de energia espectral para o Raman e Raman ressonante



Fonte: Shipp, Sinjab e Nothinger (2017, p. 12). Diagrama de nível de energia para espalhamento espontâneo (a); Espalhamento Raman fraco (b); Espalhamento Raman de Ressonância (c); Dispersão coerente Raman Anti-stokes (d); Espalhamento Raman estimulado (e); Diagrama de nível de energia CARS, SRS e não ressonante (f).

Yang e Ying (2011), assim como Herrero (2008) destacam que quando a radiação incidente e a radiação espalhada são diferentes ocorre o fenômeno Raman. Como observado na Figura 05(a), pode-se ter o espalhamento de *Rayleigh* ou espalhamento elástico que ocorre quando a radiação espalhada tem uma frequência igual à luz incidente. A dispersão de *Stokes* Raman, na qual a radiação espalhada perdeu energia para exercitar os modos vibracionais ou rotacionais da molécula, tornando-se menor do que a radiação incidente. Por fim, o espalhamento Raman *Anti-Stokes*, no qual a radiação espalhada

ganha energia igual aos modos vibracionais da molécula incidida (SHIPP; SINJAB; NOTHINGER, 2017).

Na espectroscopia Raman ocorre um espalhamento espontâneo, porém, com um sinal fraco [Figura 05(b)]. Esse sinal pode ser melhorado sintonizando a luz incidente com uma transição eletrônica (condição de ressonância) denominado de espalhamento Raman de ressonância (SALA, 1986; SHIPP; SINJAB; NOTHINGER, 2017). O aprimoramento pode ser observado ao comparar a intensidade do sinal da Figura 05b que se apresenta visivelmente menor em relação ao da Figura 05(c), na qual o sinal da radiação incidente é aproximado de uma transição eletrônica.

Outra forma de melhoramento pode ser realizada através do espalhamento coerente *Anti-Stokes* (CARS, do inglês *coherent Anti-Stokes Raman Scattering*) e/ou espalhamento Raman estimulado (SRS, do inglês *stimulated Raman Scattering*) observada nas Figuras 05(d), 05(e) e 05(f). Nos respectivos processos está envolvida a utilização de dois feixes de entrada de modo a conduzir coerentemente os modos vibracionais das moléculas de um dado volume da amostra (SHIPP, SINJAB; NOTHINGER, 2017).

Com a técnica Raman CARS a separação do sinal *Anti-Stokes* e a fluorescência são garantidas. Nesse processo, as trocas de energia só ocorrem nos campos de entrada e saída, resultando em estados vibracionais iniciais e finais idênticos [Figura 05(d)]. Esse processo é chamado de paramétrico. Já no processo de espalhamento Raman estimulado, os campos trocam energia com o meio em um processo dissipativo, isso resulta em um ganho de energia corroborando em maior eficiência em relação ao espalhamento Raman espontâneo (DUARTE, 2014; SHIPP, SINJAB; NOTHINGER, 2017).

À vista disso, os espectros gerados por meio da técnica Raman representam bandas característica de uma específica ligação química. Em outras palavras, irá fornecer a impressão digital de uma molécula possibilitando extrair informações pertinentes sobre uma determinada amostra (RIOS, 2018). Além da informação qualitativa, com a técnica Raman é possível realizar análise quantitativa, pois a intensidade do sinal ativo no espalhamento Raman é diretamente proporcional à concentração da molécula (FAKAYODE et al., 2020).

Nessas circunstâncias, pode-se empregar a técnica Raman para realizar análises tanto em amostras sólidas quanto em soluções aquosas, fornecendo informações sobre mudanças estruturais, classificação e determinação de conteúdo (YANG; YING, 2011). Portanto, tendo em vista que as fibras de algodão são produtos agrícolas sólidos, a

espectroscopia Raman poderia ser empregada para realização de análises como na avaliação dos atributos de qualidade das fibras de algodão.

Uma das principais vantagens da espectroscopia Raman é capacidade de realizar análises não destrutivas em um curto período de tempo. Como discutido anteriormente sobre as técnicas convencionais para avaliar os parâmetros de qualidade das fibras de algodão, a espectroscopia Raman apresenta características não destrutivas que divergem das técnicas convencionais, podendo ser utilizada para tais finalidades.

No entanto, os dados analíticos fornecidos nas análises por espectroscopia Raman são complexos, de difícil interpretação e em alguns casos multivariados, necessitando-se recorrer aos princípios e conceitos abordados na Quimiometria. Assim, através dessa combinação é possível obter informações importantes sobre a amostra analisada de acordo com os objetivos traçados para a pesquisa.

3.5 Métodos quimiométricos

A priori, os métodos empregados na quimiometria eram aplicados em outros ramos científicos (SIMÕES, 2008). A primeira evidencia dos métodos multivariados no tratamento de dados químicos (Quimiometria) ocorreu em 1969, onde um grupo de pesquisadores publicou uma série de artigos na revista *Analytical Chemistry*. Anos mais tarde, 1972, o termo “Quimiometria” foi designado em uma publicação realizada pelo químico Svante Wold no periódico sueco *Kemisk Tidskrift* (FERREIRA, 2015).

Com isso, a Quimiometria passou a ser considerada como disciplina e linha de pesquisa em 1970, justamente pela colaboração de Svante Wold e o matemático e químico analítico estadunidense Bruce Kowalski. Estes dois cientistas fundaram a Sociedade Internacional de Quimiometria (ICS do inglês, International Chemometrics Society) em 1974 (GELADI; ESBENSEN, 1990). Assim, com os avanços tecnológicos, equipamentos como espectrômetros e cromatógrafos passaram a fazer parte da instrumentação analítica, corroborando no fornecimento de uma imensa quantidade e variedade de dados químicos, necessitando dos conceitos quimiométricos para serem interpretados (FERREIRA, 2015).

Deste modo, a Quimiometria é definida pela União Internacional de Química Pura e Aplicada (IUPAC, do inglês International *Union of Pure and Applied Chemistry*) como uma ciência que possibilita a interpretação de dados multivariados de análises e processos químicos relacionando-os a métodos e modelos matemáticos e estatísticos (HIBBERT,

2016). Uma das proficiências da Quimiometria é permitir que métodos analíticos possam ser usados para análise multivariada como os dados espectrais Raman, sem a necessidade de resolução do sinal analítico, isto é, realizar uma separação matemática dos sinais ao invés de separar química ou fisicamente os interferentes (SENA; ALMEIDA, 2018).

O sinal analítico supramencionado trata-se de uma diversidade de informações implícitas sobre a amostra ou processo estudado. Eles podem ser adquiridos através das variáveis de técnicas analíticas como a espectroscopia (número de onda), cromatografia (tempo de retenção) entre outras. Com isso, utilizam-se os modelos quimiométricos para traduzir os dados em informações químicas, mas é necessário que estes estejam organizados em uma matriz de dados, conforme observado na Equação 01 (FERREIRA, 2015):

$$\mathbf{X} = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1J} \\ X_{21} & X_{22} & \dots & X_{2J} \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ X_{I1} & X_{I2} & \dots & x_{IJ} \end{bmatrix} \quad \text{Eq. 01}$$

Onde: A matriz $\mathbf{X}$ é sempre uma letra em maiúsculo em negrito. As linhas (I) irão corresponder às amostras, objetos, mistura ou composto. As colunas (J) irão corresponder as avariáveis, como exemplo os comprimentos de ondas, tempo de retenção, potencial de eletrodo. Portanto, a matriz de dados é organizada no sentido de $\mathbf{X} = [I \times J]$ em que X_{IJ} equivale ao resultado da variável em relação a amostra.

Atualmente, existem diversos métodos quimiométricos que podem ser aplicados na análise de um conjunto de dados analíticos (matriz $\mathbf{X}$). Com isso, a escolha do método é muito importante para as considerações acerca da pesquisa, sendo então determinado pela natureza dos dados, isto é, a forma que foi adquirida e pelos objetivos previamente definidos (SENA; ALMEIDA, 2018). Entre os métodos quimiométricos, têm-se: O Planejamento e Otimização de Experimentos; Reconhecimento de padrões (métodos supervisionados e não supervisionados); Calibração Multivariada e Resolução de curvas (SILVA et al, 2019; FILZMOSER et al., 2020).

Também conhecida como método não supervisionado, a análise exploratória tem como principal característica criar modelos sem qualquer tipo de informação sobre a natureza dos dados. A análise por componentes principais (PCA, do inglês *Principal component analysis*) é um dos principais métodos utilizados na análise exploratória, fornecendo informações através de diversas ferramentas (gráfico de *score* bidimensional

e/ou tridimensional; gráfico de *loadings*; T^2 de *hotelling*; Q residual). A PCA é capaz de verificar naturalmente a existência ou não de agrupamentos entre as amostras a partir de suas informações químicas. (HOTELLING, 1993; SZYMANSKA, 2015; SILVA et al, 2019).

Na análise supervisionada, informações a priori sobre as amostras precisam ser fornecidas ao modelo de classificação. Sistemáticamente, o método supervisionado atribui as amostras desconhecidas classes previamente definidas. Para esses procedimentos, pode-se utilizar uma análise discriminante de mínimos quadrados parciais (PLS-DA, do inglês *Partial least square discriminant analysis*), SIMCA (do inglês, *Soft independent modeling of class analog*), entre outros métodos supervisionados (FERREIRA, 2015; LEE; HAN; NAM, 2017).

Além dos modelos de classificação, outro conjunto de métodos quimiométricos frequentemente utilizado, principalmente na quantificação analítica, é a calibração multivariada. Na calibração multivariada o principal objetivo é estabelecer um modelo capaz de quantificar uma determinada propriedade a partir da medição de um sinal analítico. Assim como nas técnicas de reconhecimento de padrões, a calibração multivariada possibilita a interpretação de diversas propriedades referentes às amostras sem resolução do sinal analítico.

Entre os métodos de Calibração Multivariada pode-se citar: O método de mínimos quadrados parciais (PLS, do inglês *Partial least square*); a regressão em componentes principais (PCR, do inglês *principal Component regression*); o método de regressão linear múltipla (MLR, do inglês *Multiple linear regression*) (SENA; ALMEIDA, 2018), entre outros.

Uma etapa importante que precede a construção de modelos de Calibração Multivariada, de análise exploratória ou de classificação supervisionada é a avaliação da necessidade de realização de pré-tratamento dos dados. Esse processo tem como objetivo melhorar o conjunto de dados diminuindo as variações aleatórias, ruídos entre outros problemas que prejudicam o sinal de interesse, corroborando para uma melhor capacidade preditiva dos modelos desenvolvidos (ENGEL et al., 2013; FERREIRA, 2015).

3.5.1 Pré-Processamento dos dados

Normalmente, antes de construir os modelos propriamente dito realiza-se um pré-tratamento nos dados a serem utilizados. De acordo com Ferreira (2015) há duas possibilidades de aplicação de um pré-tratamento: transformações nas amostras (as linhas da matriz); pré-processamento nas colunas (variáveis). A escolha do pré-processamento a ser utilizado é muito importante, pois a escolha, combinação e/ou sequência inadequados pode reduzir o desempenho do modelo quimiométrico (ENGEL et al., 2013; MISHRA et al, 2021).

As aplicações de pré-tratamento nas amostras são distintas das aplicações nas variáveis. Para os pré-processamento, que no caso envolve o tratamento de todas as variáveis de uma amostra por vez, têm-se a centralização na média, a ponderação e o auto-escalamento. (FERREIRA, 2015).

Com relação ao tratamento de dados realizados nas amostras será destacado: A Normalização; Correção de linha de base; Técnicas de alisamento (*Savitzky-Golay*); Correção do espalhamento multiplicativo (MSC, do inglês *multiplicative scatter correction*); Variação normal padrão (SNV, do inglês *standard normal variate*).

3.5.1.1 Centragem na média

A centralização na média é um dos pré-processamentos mais utilizados. Uma das principais vantagens da utilização desse método é a capacidade de remover todas as expressões de interceptação do modelo (OLIVIERI, 2018). A matemática envolvida na centragem dos dados na média consiste em: calcular a média aritmética de uma determinada variável e subtrair do resultado de cada amostra na respectiva variável, como expresso na Equação 02 (FERREIRA, 2015):

$$X_{ij(cm)} = X_{ij} - \bar{X}_j \quad \text{Eq.02}$$

Onde: $X_{ij(cm)}$ referem-se aos dados centrados na média, na qual j corresponde a uma variável da amostra i; X_{ij} , trata-se do valor correspondente a variável j na amostra i, por fim, $\bar{X}_j$ é a média de uma determina variável, isto é, a média de uma coluna da matriz de dados.

3.5.1.2 Escalamento pela variância

No escalamento pela variância os valores dos dados tornam-se invariantes, isto é, os dados ficam adimensionais. O efeito causado pela aplicação desse pré-processamento torna as variáveis com pesos iguais mesmo com a presença de picos, espalhamentos ou até ruído na linha de base dos comprimentos de onda (FERREIRA, 2015). Assim, cada elemento da coluna na matriz é dividido pelo desvio padrão da respectiva coluna conforme a Equação 03:

$$x_{ij(EV)} = \frac{x_{ij}}{S_j} \quad \text{Eq.03}$$

Onde: $x_{ij(EV)}$ refere-se ao escalamento pela variância da variável j na amostra i ; já x_{ij} é valor correspondente a variável j na amostra i e S_j trata-se do desvio padrão à variável j .

3.5.1.3 Autoescalamento

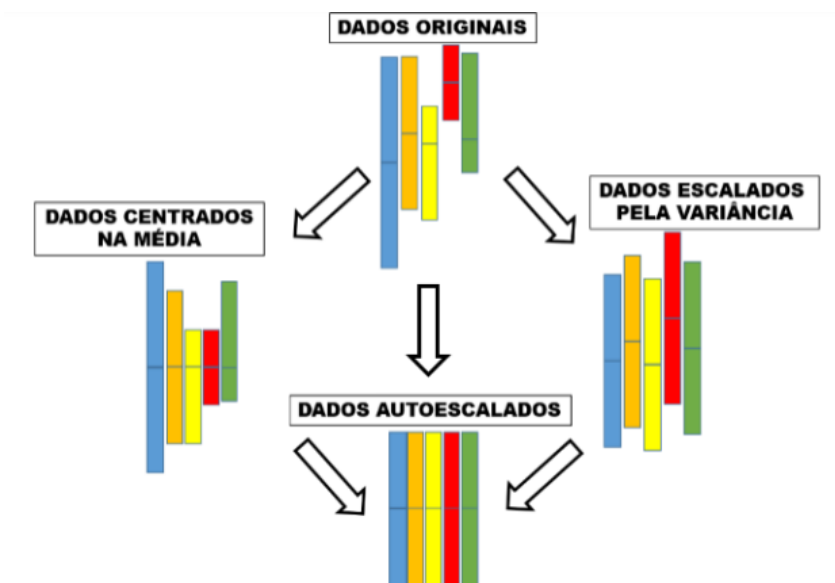
O autoescalamento normalmente é aplicado com intuito de diminuir os efeitos de variáveis indesejáveis que prejudicam o modelo, normalmente aplicado em dados discretos que apresentam unidades e escalas distintas (OLIVEIRA et al., 2020). Trata-se de um pré-processamento resultante da combinação entre a centralização na média e o escalonamento, podendo ser expresso matematicamente na Equação 04 (FERREIRA, 2015):

$$X_{ij(a)} = \frac{x_{ij} - \bar{X}_j}{S_j} \quad \text{Eq.04}$$

Onde: $X_{ij(a)}$ é auto-escalamento, cuja o resultado obtém-se pela subtração de cada espécie da matriz de dados X_{ij} de uma determinada coluna pela média da respectiva coluna $\bar{X}_j$ e dividir pelo desvio padrão S_j referente a variável da coluna. Resumidamente, cada variável da matriz será centrada na média e escalonada através do desvio padrão.

Na Figura 06 observa-se uma esquematização do comportamento dos dados quando aplicado os pré-processamento mencionado.

Figura 06 – Esquema de pré-processamento aplicado nas variáveis



Fonte: Rios (2018)

Ao observar a Figura 06, percebe-se as diferenças nos efeitos causados pela aplicação de cada pré-processamento nas variáveis. Com isso, a escolha adequada do pré-processamento vai estar relacionada com a natureza dos dados, pois cada um irá causar um efeito específico sobre o conjunto de dados.

3.5.1.4 Normalização

A normalização é um tipo de pré-tratamento aplicado nas amostras com o propósito de reduzir ao máximo o número de variações que não contribuem para construção do modelo. Existem diferentes métodos de normalizar um conjunto de dados, como por exemplo, a normalização por vetor unitário, normalização pela área total, normalização infinita ou “sup”, entre outros. Na Equação 05 observa-se matematicamente como a normalização é aplicada a cada elemento contido em uma linha da matriz de dados (FERREIRA, 2015):

$$X_{ij(norm)} = \frac{X_{ij}}{\|X_i\|}, \quad j = 1, 2, \dots, J. \quad \text{Eq.05}$$

Onde: $(X_{ij(norm)})$ expressa a normalização pelo vetor unitário, cuja o valor é dado pelas variáveis de uma determinada amostra (X_{ij}) dividido pela norma da respectiva amostra $\|X_i\|$.

3.5.1.5 Correção de linha de base

Os sinais registrados por um instrumento espectroscópico pode apresentar contribuições estocásticas, também denominada como ruído, fontes de variação com baixa frequência que muitas vezes não estão relacionadas com a investigação realizada. Uma variação nos espectros indesejada é o deslocamento da linha de base que pode ser contatante, quando ocorre no eixo das ordenadas; deslocamento variável quando há uma inclinação na linha de base; pode ocorrer no eixo das abscissas, correspondendo as variáveis (FERREIRA, 2015).

Na espectroscopia Raman a linha de base pode ser prejudicada por efeitos multiplicativos, aditivos e curvaturas oriundas da fluorescência. A partir de uma visualização dos dados espectrais o analista irá verificar a necessidade da correção da linha de base, e qual método será empregado para tal finalidade. Diversos métodos matemáticos podem ser empregados como etapa de pré-tratamento quimiométrico (DUCKWORTH, 2004).

Para correção de linha de base espectros Raman é frequentemente empregado uma correção por mínimos quadrados ponderados (WLS, do inglês *Weighted least squares*). O WLS trata-se de um algoritmo capaz de ajustar a linha de base de cada espectro. Para isso, utiliza uma linha de base de referência e ajusta os espectros que estão acima ou abaixo em relação a referência (CHEN et al., 2006).

O ajuste de linha de base por regressão linear, é um exemplo de método matemático que pode ser empregado para correção da linha de base. Porém, algumas ressalvas são necessárias na escolha desse método, pois a aplicação é indicada quando a linha de base está muito inadequada e também com número baixo de amostras disponível. Outro método utilizado para correção da linha de base refere-se a linha linear de dois pontos. Nesse método é selecionado duas linhas de base no espectro e posteriormente conectando-as com uma linha e subtraindo das respostas espectrais (DUCKWORTH, 2004).

3.5.1.6 Técnica de alisamento ou suavização

A suavização é um método de alisamento com aplicações em dados espectrais e cromatográficos, cujo propósito é eliminar as variações que ocorrem nos processos de

obtenção dos dados que podem afetar a qualidade do modelo. Uma das formas de realizar esse procedimento é através de um polinômio móvel usando o algoritmo *Savitzky-Golay*.

O *Savitzky-Golay* calcula uma média ponderada dos pesos ajustada por um polinômio. Para uma utilização adequada precisa-se considerar o grau do polinômio e o número de pontos da janela (RINNAN; VAN DEN BERG; ENGELSEN, 2009). A escolha do número de pontos precisa ser cautelosa, pois um valor alto pode retirar informações importantes do conjunto de dados e prejudicar a análise. Uma expressão geral é observada na Equação 06 para aplicação do *Savitzky-Golay*:

$$X_k = a_0 + a_1k + a_2k^2 + \dots + a_nk^n \quad k = -m, \dots, 0, \dots, +m \quad \text{Eq.06}$$

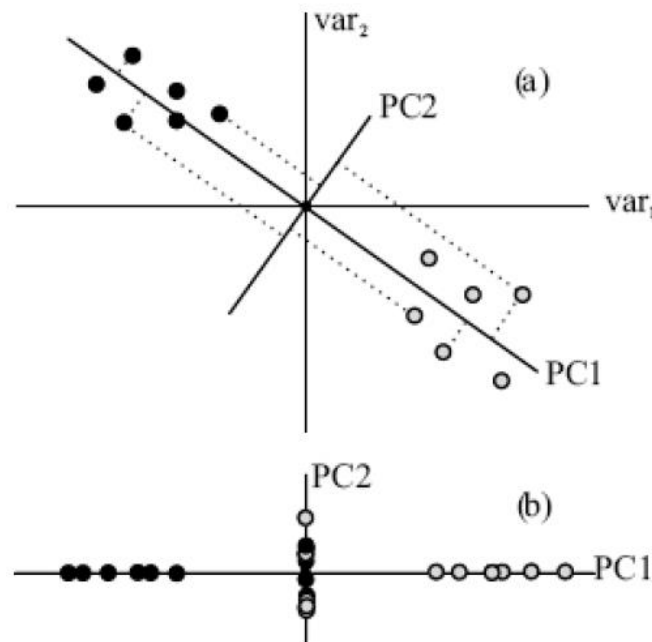
Onde: n é o grau do polinômio escolhido e k o número de pontos da janela.

3.5.2 Análise de Componentes Principais (PCA)

A PCA é um método de análise exploratória multivariada, sendo comumente empregado em tratamento de dados Químicos com imensa diversidade de variáveis. Uma das principais características desse método é capacidade de analisar uma matriz de dados sem informações precedentes sobre as amostras (BRO; SMILDE, 2014; SCHNEIDER et al., 2018).

De acordo com Marini (2013), ao realizar uma PCA ocorre uma decomposição/projeção bilinear no conjunto de dados, na qual são modificados em uma pequena quantidade de parâmetros denominados de componentes principais (PC's). As PC's identificam os níveis de semelhanças e diferenças contidos nas amostras e variáveis da matriz de dados. Sendo identificadas, as PC's são representadas em um gráfico de score em um espaço bidimensional observado na Figura 07.

Figura 06 - Representação gráfica bidimensional de dois grupos de amostra



Fonte: Ferreira (2015)

Como observado na Figura 07, verifica-se que em uma componente principal é possível identificar as semelhanças e diferenças entre as amostras e as variáveis, diferentemente de PC2. Pois a PC1 é uma combinação linear de máxima variância dos dados originais, enquanto a PC2 sendo ortogonal a PC1 explica a variância dos dados nessa dimensão. Deste modo, percebe-se que todos os PC's tem uma dimensionalidade de explicação própria, mas envolvendo-as na análise fomenta a capacidade de explorar e interpretar o conjunto de dados (BRO; SMILDE, 2014; SCHNEIDER et al., 2018).

Em termos matemáticos, Marcelo, Pozebon e Ferrão (2015) descrevem que a PCA consiste na transformação de uma matriz $\mathbf{X}$ em duas novas matrizes, sendo referente aos escores $\mathbf{T}$ e uma matriz transposta de pesos (*loadings*) $\mathbf{P}^T$, além de dados não explicado pelo modelo $\mathbf{E}$ (erro da matriz). Essa transformação é expressa na Equação 07 descrita por Marini (2013):

$$\mathbf{X}_{(I,J)} = \mathbf{T}_{IA} \mathbf{P}_{AJ}^T + \mathbf{E}_{(I,J)} \quad \text{Eq.07}$$

Onde: A refere-se ao número de componentes ou estruturas subjacentes. $\mathbf{T}_{IA}$ expressa as relações entre as amostras, isto é, os escores, enquanto $\mathbf{P}_{AJ}^T$ expressa os pesos referente as variáveis que são organizadas nas colunas da matriz transposta. Já $\mathbf{E}_{(I,J)}$ trata-se da matriz de resíduos.

3.5.3 Classificação Supervisionada

As aplicações da química analítica qualitativa vão além das análises de identificação ou classificação de compostos químicos individuais. Com a Quimiometria, torna-se possível o desenvolvimento de um amplo espectro de métodos de análises de dados. Com isso, é possível aplicar diversas técnicas instrumentais para qualificar amostras, compostos e até grupos de compostos (SZYMANSKA et al., 2015).

Diante disso, na Quimiometria o método de classificação supervisionada é desenvolvido para reconhecer padrões através de um modelo matemático. Assim, torna-se possível classificar amostras desconhecidas a classes que foram previamente determinadas. Para isso, utiliza-se um conjunto de variáveis descritivas para classificação e um conjunto de variável qualitativa, justamente as pertencentes as classes definidas (SZYUMANSKA et al., 2015; SENA; ALMEIDA, 2018).

Existem diversos métodos de classificação supervisionada, podendo-se dividi-los em lineares ou não lineares, paramétricos ou não paramétricos, de forma mais geral em discriminantes ou de modelagem de classe. Dentre esses métodos tem-se: O SIMCA um método de modelagem de classe; o método discriminante linear LDA (do inglês, *Linear Discriminant Analysis*); a análise discriminante por mínimos quadrados parciais (PLS-DA) (BARKER; RAYENS, 2003; BRERETON; LLOYD, 2014; SENA; ALMEIDA, 2018).

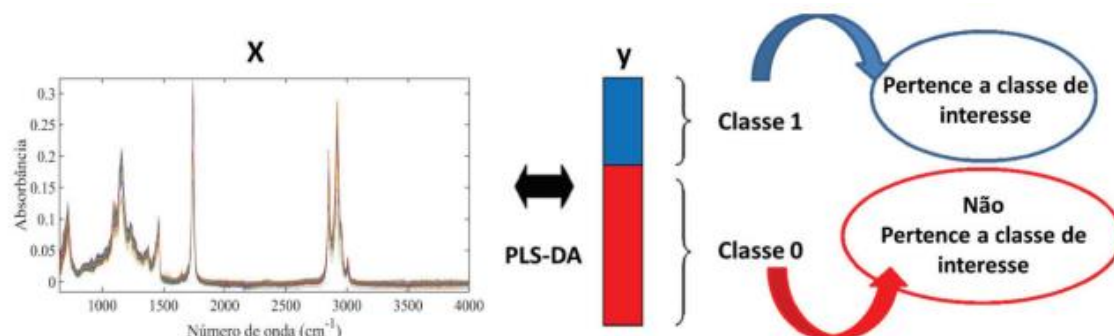
3.5.3.1 PLS-DA

O método PLS-DA realiza uma relação linear entre a medida analítica (Ex: Espectroscopia Raman, medidas HVI), denominadas de variáveis independentes “**X**” com as propriedades de interesse, cuja as variáveis são dependentes “**Y**”. Essas variáveis dependentes assumem valores binários codificados de 1 ou 0, na qual irá classifica-la como pertencente ou não a classe previamente definida (BALLABIO; TODESCHINI, 2009).

Sendo assim, a matriz **Y** possuirá uma dimensão ($l \times L$), sendo L o número de classes envolvidas ($l = 1, 2...L$). Com isso, na primeira classe $l = 1$, será atribuído o valor 1 para as amostras pertencentes aquela respectiva classe e 0 para as demais amostras não pertencentes a classe. Na coluna seguinte da matriz **Y**, quando $l = 2$, as amostras dessa

classe passam a receber o valor de um (1), enquanto as demais amostras são codificadas em 0. Essa esquematização segue-se até que todas as classes sejam codificadas (FERREIRA, 2015). Na Figura 08 observa-se o processo de construção do modelo PLS-DA.

Figura 07 - Esquematização da construção do modelo PLS-DA



Fonte: Santana (et al., 2020)

Como pode ser observado na Figura 08, tem-se uma matriz **X** com as variáveis independentes, dadas pelos valores de absorbância de cada comprimento de onda. E as variáveis dependentes dispostos em uma matriz **Y**, codificadas em 1 para as amostras pertencentes a classe e 0 para as amostras não pertencentes a classe. Assim, o número de colunas na matriz **Y** dependerá do número de classes das amostras.

Para que o modelo seja considerado eficiente para aplicações posteriores é preciso avaliá-lo sobre alguns aspectos. Na classificação supervisionada uma das formas de verificar o desempenho dos modelos é através da tabela de confusão, que mostra o número de amostras o número de amostras previstas em cada classe. Na Tabela 01 nota-se um exemplo de tabela confusão para duas classes.

Tabela 04 - Exemplificação de uma tabela de confusão para duas classes de amostras

Tabela de confusão 2x2	Previsto pelo Modelo		
		Classe 1	Classe 0
Referência	Classe 1	VP	FN
	Classe 0	FP	VN

Fonte: Santana et al., (2020)

A tabela de confusão é construída com as amostras de teste que foram corretamente indicada para sua classe, onde o modelo pode classificar em: Verdadeiro positivo (VP – uma amostra positiva classificada verdadeiramente como positiva); falso positivo (FP – uma amostra negativa é classificada como positiva); verdadeiro negativo

(VN – uma amostra realmente negativa classificada como negativa); falso negativo (FN – nessa situação a amostra positiva é classificada erroneamente como negativa) (VALENTINI; SOMMER; MATIOLI, 2007; MORAIS, 2014).

A partir dos resultados obtidos na tabela de confusão é possível obter figuras de méritos com taxas utilizadas como parâmetros para avaliação da qualidade do modelo. Algumas figuras de méritos têm a nomenclatura similares entre um método qualitativo e quantitativo como sensibilidade e especificidade. Para um modelo de classificação, a sensibilidade refere-se na capacidade do modelo classificar amostras como VP (verdadeiro positivo), cuja o valor numérico pode ser obtido realizando o cálculo da Equação 07 (SANTANA et al., 2020):

$$\text{Sensibilidade} = \frac{VP}{VP+FN} \quad \text{Eq. 08}$$

Já com relação a especificidade, o modelo é avaliado quanto a sua capacidade de identificar corretamente as amostras negativas (VN) como realmente negativas, isto é, ao invés de avaliar as amostras positivas, como a sensibilidade, a especificidade do modelo leva em consideração as amostras negativas. O cálculo para obtenção da taxa de especificidade do modelo pode ser realizado a partir da Equação 08 (SANTANA et al., 2020):

$$\text{Especificidade} = \frac{VN}{VN+FP} \quad \text{Eq. 09}$$

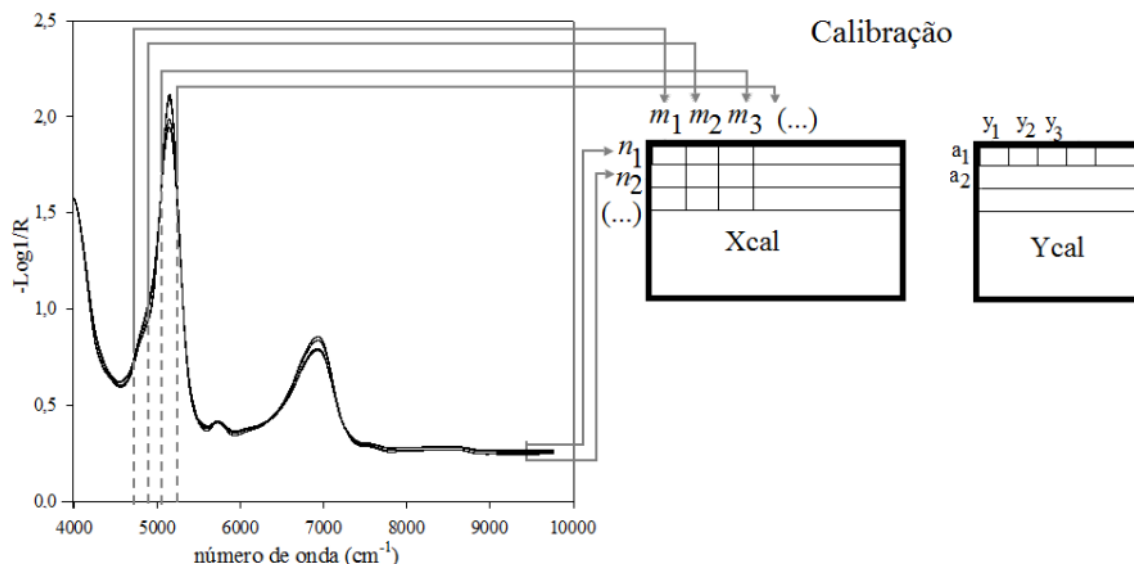
3.5.4 Calibração Multivariada

Com os avanços instrumentais, tornou-se possível a aquisição de um grande número de variáveis sobre um objeto analisado. A análise por espectroscopia Raman é um exemplo técnica capaz de extrair múltiplas informações de uma amostra. Porém, para realizar a leitura dessas informações com clarezas é precisa-se da Quimiometria e normalmente emprega-se a calibração multivariada para construção dos modelos quimiométricos (GHASEMI; NIAZI; LEARDI, 2003).

Uma das vantagens da calibração multivariada é a capacidade de diminuir a quantidade de ruído e simultaneamente determinar as propriedades referentes à amostra analisada com sinais multivariados (KUMAR et al., 2014). Complementando essa ideia, Simões (2008) destaca que a calibração multivariada relaciona a construção dos modelos matemáticos com as medidas instrumentais analíticas de referência, na qual contém informações sobre as amostras. A construção do conjunto de calibração é realizada com uma ou mais variáveis Y baseado na seletividade de um conjunto X de variáveis

relevantes (MARINI, 2013). Uma explicação sobre essa modelagem é observada na Figura 09.

Figura 08 - Esquematização da organização do conjunto de calibração



Fonte: Oliveira (2015)

A matriz $\mathbf{X}_{cal}$ (Figura 09) é estruturada por n-linhas que representam o número total das amostras analisadas e m-colunas, referindo-se aos comprimentos obtidos na análise espectral como por exemplo a espectroscopia Raman. Os valores internos, isto é, que ficam dentro da matriz $\mathbf{X}_{cal}$ são as absorvâncias das amostras em determinados comprimentos de onda. Assim, a matriz $\mathbf{X}_{cal}$ construída é chamada de variável independente.

Para a construção de um modelo de calibração multivariada, é necessário realizar uma análise química, cujo propósito é defini-la como método de referência. Com isso, as variáveis analisadas no método de referência são utilizadas para construção do vetor $\mathbf{y}_{cal}$ (Figura 09). Assim como na matriz $\mathbf{X}$, o vetor $\mathbf{y}_{cal}$ contém n-linhas e uma coluna, porém, o valor de cada linha da coluna é relacionado com a propriedade física ou química da amostra analisada pelo método de referência. Caso seja avaliado duas ou mais propriedades das amostras constrói-se uma matriz $\mathbf{Y}_{cal}$ que terá mais de uma coluna.

Além da construção de um conjunto de calibração é necessário construir um conjunto de validação e previsão. Na validação, o propósito é validar os modelos construídos na calibração e valida-los através de ferramentas estatística. Enquanto que na previsão, os modelos validados são aplicados para prever as propriedades das amostras

(SIMÕES, 2008). Sendo assim, uma alternativa para construir um modelo de calibração é através das técnicas de regressão e o PLS é uma das mais utilizadas.

3.5.4.1 Regressão por Mínimos Quadrados Parciais (PLS)

O PLS é um dos métodos mais empregado na quimiometria para análise de dados multivariados. No entanto, a denominação como PLS ocorreu inicialmente em dados econométrico por Herman Wold na década de 1970, anos mais tarde Svante Wold introduziu-a em abordagens quimiométricas (WOLD; SJÖSTRÖM; ERIKSSON, 2001; FERREIRA, 2014). A relevância do método PLS é justificada pela capacidade de tratar matrizes com grande variabilidade de dados, de modo a obter as informações mais relevantes das amostras (GHASEMI; NIAZI; LEARDI, 2003).

Sendo assim, desde que a PLS tenha informações parciais sobre o conjunto de dados a sua utilização pode ser de uma significativa utilidade. Tanto, que dependendo do conjunto de dados os modelos PLS podem serem muito robustos. Outra característica dos modelos PLS, trata-se de considerar os erros que ocorrem nas estimativas de concentrações, como também os erros nos espectros (BRERETON, 2000).

A PLS utiliza variáveis latentes (LV, do inglês *latente variables*) capazes de realizar a diminuição dimensional e correlacionar a matriz espectral $\mathbf{X}$ com as propriedades físicas e/ou química de interesse organizada na matriz $\mathbf{Y}$ (SOUZA et al., 2013). Deste modo, as matrizes $\mathbf{X}$ e $\mathbf{Y}$ são decompostas em VL, podendo ser explicada matematicamente na Equação 09 e 10:

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E} \quad \text{Eq.10}$$

$$\mathbf{Y} = \mathbf{UQ}^T + \mathbf{F} \quad \text{Eq.11}$$

Em que: A matriz $\mathbf{X}$ é dada pelo produto de pesos $\mathbf{P}^T$ com escores somados com a matriz residual $\mathbf{E}$. Semelhantemente, na matriz $\mathbf{Y}$ o produto ocorre entre os pesos expressos por $\mathbf{Q}^T$ e os escores $\mathbf{U}$ somados a matriz de resíduos $\mathbf{F}$.

Assim, os novos eixos formados, as VL's, equivalem as PCA's. em que os scores também funcionam como coordenadas das amostras. A diferença em relação a PCA é que no PLS as variáveis latentes não são ortogonais como os componentes principais, mas irão atuar em um modelo linear até atingir a máxima covariância entre $\mathbf{X}$ e $\mathbf{Y}$ (TEÓFILO, 2007; SOUZA et., 2013; FERREIRA, 2015). Essa correlação linear trata-se do modelo

final de PLS, podendo ser representado matematicamente na Equação 11 e 12 (SIMÕES, 2008):

$$\mathbf{U} = \mathbf{BT} + \mathbf{G} \quad \text{Eq.12}$$

$$\mathbf{Y} = \mathbf{BTQ}^T + \mathbf{H} \quad \text{Eq.13}$$

Na qual: os coeficientes de regressão estão representados pela matriz **B**; Já **G** refere-se a matriz de resíduos dos escores e por fim, **H** expressa a matriz de resíduos de concentração.

3.5.4.2 Parâmetros de qualidade do modelo

Para que a modelagem do modelo seja realizada com qualidade precisa-se escolher um número ideal de variáveis latentes (nLV). Pois, um nLV alto pode modelar informações que não apresentam contribuições tornando o modelo sobreajustado, enquanto que um nLV pequeno pode não incluir informações que seriam relevantes favorecendo para um modelo subajustado (FABER; RAJKÓ, 2007).

Um dos métodos utilizados para a escolha do número adequado de variáveis latentes para um modelo de calibração é o *Leave-one-out* (LOO). Nesse método é excluído uma amostra do conjunto de calibração, de modo que a capacidade preditiva do modelo seja avaliada a partir da amostra deixada de fora do conjunto de calibração (FERREIRA, 2015).

Nesse contexto, o LOO é aplicado em todas as amostras do conjunto de calibração e o número de VL escolhido depende do menor erro de validação cruzada (RMSECV, do inglês *root mean square error of cross validation*) expressa na Equação 13 (SENA; ALMEIDA, 2018):

$$RMSECV = \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n_c}} \quad \text{Eq.14}$$

Onde: a soma quadrática (expressão na raiz quadrada) é dada por n_c que representa o número de amostra do conjunto de calibração; $\hat{Y}_i$ refere-se ao valor previsto pelo modelo e Y_i corresponde ao valor de referência da i-ésima amostra.

Para avaliação da qualidade do modelo é indispensável avaliar a sua capacidade preditiva, pois é importante que ele seja capaz de prever novas amostra, de modo que se apresente com uma robustez significativa. Assim, uma maneira de avaliar esse aspecto destacado por Valderrama, Braga e Poppi (2009) é o RMSEP (do inglês, *Root Mean*

Squares Error of Prediction), ou seja, o somatório da raiz média quadrática de predição (Equação 14):

$$RMSEP = \sqrt{\frac{\sum_{p=1}^P (Y_p - \hat{Y}_p)^2}{P}} \quad \text{Eq.15}$$

Em que: p refere-se ao número de amostra do conjunto de predição; Y_p o número previsto pelo modelo e $\hat{Y}_p$ corresponde ao valor de referência.

Outro parâmetro capaz de avaliar a qualidade do modelo é o erro-padrão relativo (SEP, do inglês *standard error prediction*) que foi expresso na Equação 15 por Ferreira (2015):

$$SEP = \sqrt{\frac{\sum_{p=1}^P (y_p - \hat{y}_p - \text{viés})^2}{P-1}} \quad \text{Eq.16}$$

Onde: $\text{Viés} = \frac{1}{I} \sum_{i=1}^I (y_i - \hat{y}_i)$, isto identificar a presença de tendências do modelo; P o número de amostras no conjunto de predição; $\hat{y}_p$ resultado previsto pelo modelo; por fim, y_p corresponde ao valor de referência.

4 PERCURSO METODOLÓGICO

4.1 Localização da pesquisa

Este trabalho foi realizado no Laboratório Avançado de Tecnologia Química- (LATECQ) localizado na Embrapa Algodão (Empresa Brasileira de Pesquisa Agropecuária), unidade de Campina Grande PB. E nas dependências do Laboratório de Síntese Inorgânica e Quimiometria (LabSIQ) localizado no Campus I da Universidade Estadual da Paraíba, Campina Grande, Paraíba, Brasil.

4.2 Amostras estudadas

Foram utilizadas um total de 213 amostras de fibras de algodão, fornecidas pela Embrapa Algodão, unidade de Campina Grande PB. O conjunto total das fibras são proveniente de 3 localidades distintas identificadas como localidade A (69 amostras), localidade B (72 amostras) e localidade C (72 amostras). Destas, as amostras provenientes da localidade A são convencionais (69 amostras) e as provenientes das outras duas localidades (134 amostras) são geneticamente modificadas. Além disso as amostras provenientes da localidade A podem ser classificadas de acordo com seu desempenho produtivo como: precoce (22 amostras) ou médio tardio (47 amostras).

4.3 Etapas da pesquisa

4.3.1 Primeira Etapa: Aquisição dos dados espectrais Raman

Para obtenção dos dados espectrais foi utilizado um equipamento Raman Portátil (Figura 10), Mira Cal DS, Metrohm®, com faixa espectral de 400 a 2300 cm^{-1} , resolução espectral de 8-10 cm^{-1} , que possui um único laser de comprimento de onda de 785 nm $\pm$ 0.5nm e potência de aproximadamente 100 mW e interface USB 2.0 com fonte de alimentação de transmissão de dados com cabo USB.

Figura 9 - Espectrômetro Raman Portátil



Fonte: Metrohm (2021)

Foram realizados os registros de todas as amostras a partir de três repetições para cada amostra, ou seja, registrou os espectros de três porções de fibras com massa aproximadamente iguais e esse processo seguiu-se para todas as amostras de fibras de algodão. Em seguida fez-se a média aritmética dos registros espectrais, com isso, resultou-se em uma matriz constituída de 404.700 dados espectrais, organizados em 1900 colunas (comprimentos de ondas) e 213 linhas (amostras).

4.3.2 Segunda Etapa: Análises dos atributos de qualidade das fibras de algodão

Foram realizadas medidas dos atributos de qualidade das fibras de algodão. Para realização dessas análises foi utilizado a instrumentação HVI (Figura 11), na qual foram determinados: Comprimento da Fibra; o índice de fibras curtas; a resistência da fibra; o índice de micronaire e a reflectância. Todas as medidas foram realizadas em triplicata e aplicado uma média para o tratamento dos dados.

Figura 10 - Sistema HVI para medidas dos atributos de qualidade das fibras de algodão



Fonte: USTER® HVI 1000 (2022)

4.3.3 Terceira Etapa: Construção dos modelos Quimiométricos

4.3.3.1 Softwares Quimiométrico

O tratamento dos dados espectrais e a construção dos modelos de calibração multivariada foi realizada utilizando o pacote PLS-toolbox®, versão 6.0.1 da marca Eigenvector em um ambiente Matlab®.

4.3.3.2 Análise exploratória dos dados

Após uma análise inicial do perfil espectral, selecionou-se uma região de trabalho (700 a 1700 cm^{-1}) e construiu-se uma PCA utilizando todas as amostras. Nessa análise, também foi avaliada possíveis amostras com comportamentos anômalos através de gráficos como T^2 de *hotelling* e Q residual. Assim, foram construído uma PCA para cada classe de fibra de algodão, isto é, referente a localidade, diferenças genéticas e diferenças na maturidade.

Além das PCA's com os dados espectrais, também foi realizado uma análise exploratória com os dados de atributos de qualidade. Nesses modelos, foram avaliados os possíveis agrupamentos e realizados estudos para identificação de possíveis *outliers* através dos gráficos de T^2 de *hotelling* e Q residual. Assim como para os dados espectrais, como os atributos foram construídos PCA's referente as fibras de localidades distintas, modificações genéticas e diferentes maturidade.

4.3.3.3 Construção do modelo PLS-DA

Foram desenvolvidos três modelos de PLS-DA para cada conjunto de dados, espectrais Raman e atributos de qualidade das fibras de algodão. Esses modelos foram aplicados para classificar amostras quanto a sua localidade, modificações genéticas e maturidade. Com isso, organizou-se uma matriz **X**, constituída com as variáveis independentes e uma matriz **Y** com as variáveis dependentes codificadas em 1 (pertencente a classe) e 0 (não pertencente a classe).

Posteriormente, os dados foram divididos em conjunto de treinamento (dois terços das amostras) e conjunto de teste (um terço das amostras). Para selecionar as amostras nos subconjuntos foi aplicado o algoritmo KS em cada classe separadamente (KENNARD, 1969; PONTES et al., 2005). Na Tabela 05 pode ser observar a organização dos subconjuntos para cada classe de fibra.

Tabela 05 - Número de amostras de treinamento e teste de cada classe de fibras de algodão

Classes de fibras de algodão analisadas		Subconjuntos	
		Treinamento	Teste
Diferentes localidades	Localidade A	46	23
	Localidade B	48	24
	Localidade C	48	24
Modificações genética	Convencional	46	23
	Transgênica	96	48
Diferentes maturidades	Precoce	13	09
	Tardio	30	17

Fonte: Arquivo Próprio (2022)

Em todos os modelos, o número adequado de variáveis latentes foi baseado na porcentagem de amostras classificadas corretamente em cada classe na validação cruzada. Nessa determinação foram avaliados os gráficos de variância explicada versus o número de VL. Essa avaliação consistiu em selecionar o número de VL que não tornasse o modelo sobreajustado ou subajustado, ou seja, que incluísse toda informação útil para construção do modelo, como também não relacionasse aspectos como ruídos espectrais (SANTANA et al., 2020).

Por fim, nos dados espectrais foram testadas diferentes combinações de pré-processamento, afim de identificar a combinação que resultasse na melhor capacidade preditiva no modelo. Para avaliação da qualidade modelo, foram considerado os valores

de especificidade e sensibilidade e porcentagem de erro nas etapas de treinamento e teste. As avaliações foram consideradas tanto para os dados espectrais Raman quanto para os dados de atributos de qualidade das fibras de algodão.

4.3.3.4 Construção do modelo PLS

O número de amostras selecionadas para compor o conjunto de calibração foi de dois terços do total das amostras, com isso utilizou-se para o conjunto de calibração 142 amostras e para o conjunto de predição foi selecionado 71 amostras. Vale destacar que para seleção das amostras de ambos os conjuntos foi utilizado o algoritmo **SPXY** (do inglês, *Samples set Partitioning basead on joint X-Y distances*) (GALVÃO et al., 2005).

Nesse contexto, construiu-se um modelo de calibração para cada parâmetro avaliado pela análise de referência, totalizando 05 modelos PLS's. Para cada um dos modelos verificou-se qual pré-processamento e/ou sequência de pré-processamentos adequava-se mais, com base na capacidade preditiva dos modelos gerados.

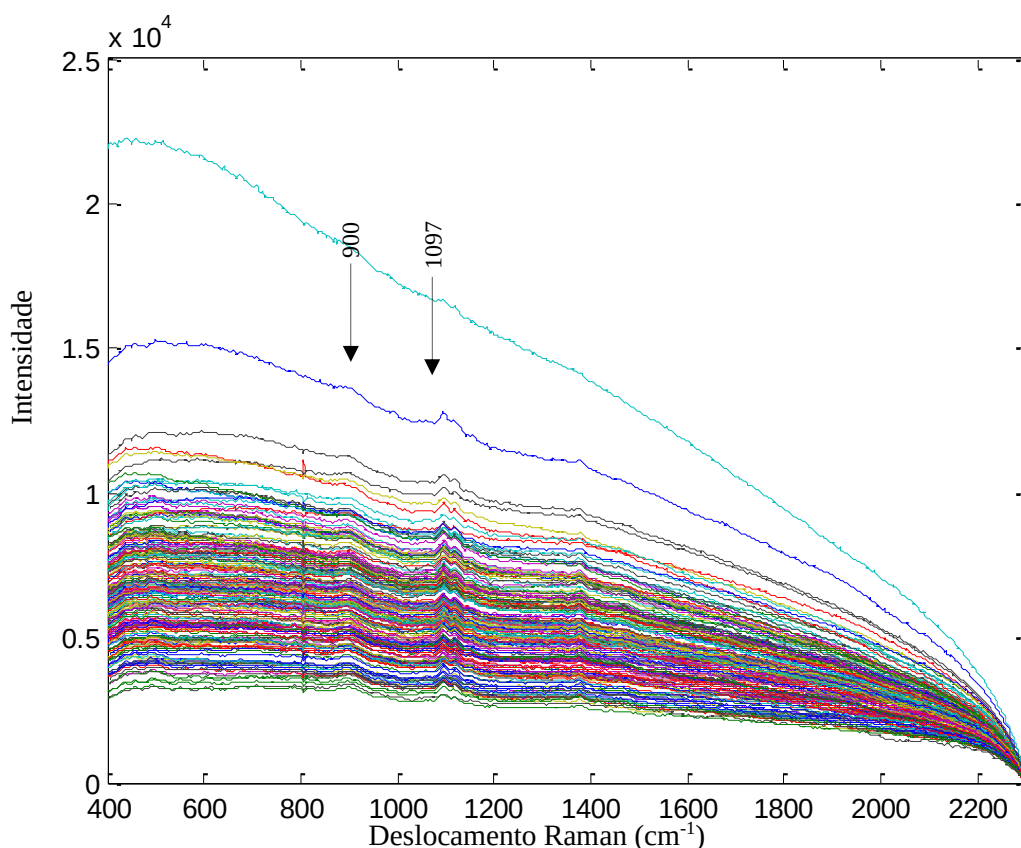
Por fim, para avaliação da qualidade do modelo foram avaliadas figuras de mérito como raiz quadrada do erro médio quadrático (RMSE, do inglês *Root mean square error*) das etapas de calibração, validação e predição; R^2 (coeficiente de determinação) e Bias, além da análise dos gráficos de valores preditos pelo modelo *versus* o de referência.

5 RESULTADOS E DISCUSSÕES

5.1 Análise dos espectros Raman

Antes da construção dos modelos de reconhecimento de padrões (PLS-DA) e de calibração multivariada (PLS) foi realizada uma análise do comportamento dos espectros das amostras em toda faixa espectral registrada, que compreendeu a região de 400 a 2300 cm^{-1} (Figura 12). Esse estudo foi importante para o entendimento sobre quais bandas poderiam contribuir significativamente para construção dos modelos. Nos espectros brutos verifica-se sinais nas regiões de 900 e 1097 cm^{-1} . Já a literatura reporta que as fibras de algodão apresentam sinais característicos de espectroscopia Raman em bandas como 1483, 1381, 1340, 1097, e 900 cm^{-1} (BIANCHI et al., 2016).

Figura 11 - Espectros Raman de 213 amostras de fibras de algodão na região 400 a 2300 cm^{-1} destacando-se as bandas características



Fonte: Próprio Autor (2022)

Ao analisar o gráfico da Figura 12, percebeu-se uma curvatura nos extremos do espectro, devido a influência do sinal de fluorescência, além de um deslocamento de linha

de base e regiões com baixa relação sinal ruído (BURGER e GOWENE, 2011). Deste modo, excluiu-se as extremidades dos espectros na faixa espectral que compreende as regiões de 400 a 699 cm^{-1} e 1701 a 2300 cm^{-1} . O que motiva esta exclusão é que as regiões mencionadas, por sua inclinação poderiam prejudicar a aplicação de alguns pré-processamentos espectrais e consequentemente o desempenho preditivo dos modelos de classificação e calibração.

Outra justificativa para retirada das extremidades dos espectros foi pautada nas informações presentes da Tabela 06, na qual observa-se as regiões características mais informativas e atribuições de bandas referentes aos modos vibracionais das ligações químicas de espécies presentes nas fibras de algodão.

Tabela 06 - Principais frequências de bandas Raman característicos de fibras de algodão

Comprimento de onda (cm^{-1})	Indicação na figura 08b	Atribuição de banda	Fonte pesquisada
1614 e 1660	I	Hidrocarbonetos insaturados/Proteína	Kim (et al., 2017); Lee (et al., 2014)
1160, 1122 e 1096	II	Alongamento simétrico das ligações de anel 1,4-glicosídica	Liu, Kokot e Sambi (1998)
1000 – 1150, 1300 – 1500	III	Picos que podem ser atribuídos a alongamento assimétrico da celulose	Kim (et al., 2017); Lee (et al., 2014); Bianchi (et al., 2016)
972, 1380, 1410, 1462	IV	δ (CH_2)	Lee (et al., 2014); Liu, Kokot e Sambi (1998)
1471	V	Carboidratos; Celulose δ (CH_2)	Lee (et al., 2014)
900	VI	γ (COC)	Liu, Kokot e Sambi (1998)
896	VII	H_2O ; Celulose	Lee (et al., 2014)

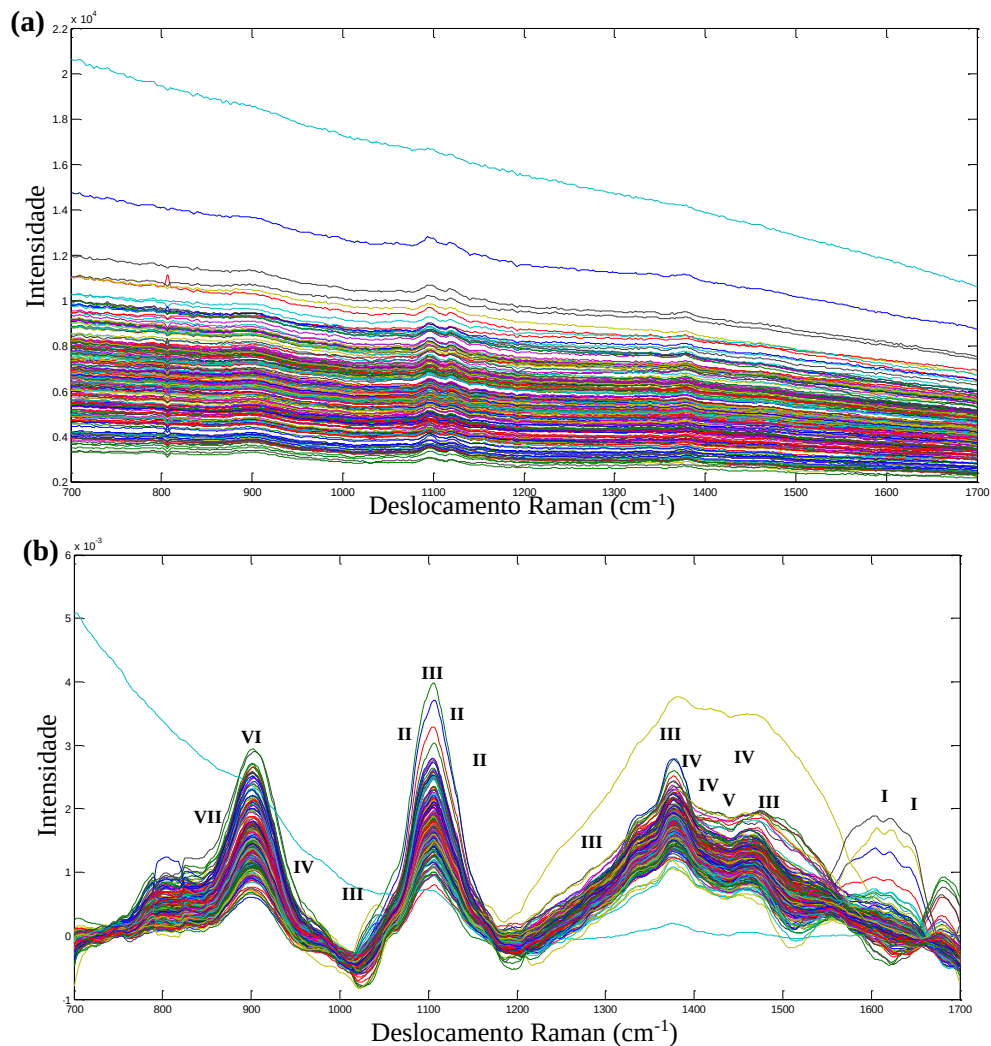
Fonte: Próprio Arquivo (2022)

Como esquematizado na Tabela 06, algumas regiões dos espectros Raman das fibras de algodão apresentam sinais característicos de espécies químicas que constituem as fibras de algodão. Dentre essas espécies, a celulose destaca-se por apresentar sinais Raman em diversos comprimentos ondas, isto é, modos vibracionais de ligações químicas da celulose são ativados em diversas bandas espectrais. As regiões com sinais mais

intensos situam-se no comprimento de onda de 900 cm^{-1} e na região de 1000 a 1150 cm^{-1} (LEE et al., 2014; BIANCHI et al., 2016).

Diante disso, selecionou-se uma região de trabalho entre 700 cm^{-1} a 1700 cm^{-1} . Nessa região foram observados problemas como espalhamento de sinal, presença de ruído espectral e deslocamento da linha de base (Figura 13a). Deste modo, foram testados pré-processamentos espectrais para corrigir os problemas mencionados, que consistiram em: Normalização pelo Vetor Unitário (NVU, do inglês *Normalization by unit vector*); correção de linha de base como descrita por Duckworth (2004), com o algoritmo WLS e suavização Savitzky-Golay com polinômio de 1ª ordem e janela de 31 pontos (SF31) (Figura 13b).

Figura 12 - Espectros Raman de 213 amostras de fibras de algodão na região de 700 cm^{-1} a 1700 cm^{-1} brutos (a) e com pré-processamento base NVU + WLS + SF31 (b)



Fonte: Próprio Arquivo (2022). Regiões com sinais espectrais Raman: I (1614 e 1660 cm^{-1}); II (1160 , 1122 e 1096 cm^{-1}); III ($1000 - 1150$, $1300 - 1500\text{ cm}^{-1}$); IV (972 , 1380 , 1410 , 1462 cm^{-1}); V (1471 cm^{-1}); VI (900 cm^{-1}); VII (896 cm^{-1}) (b).

Analisando os espectros Raman da Figura 13, nota-se uma significativa melhora no perfil espectral das amostras. Com isso, constata-se a eficiência dos pré-processamento aplicados. Pode-se observar algumas amostras com perfil espectral distinto, as quais poderão ser excluídas do conjunto de dados após um estudo mais detalhado para identificação de amostras anômalas. Deste modo, após a avaliação do perfil espectral, a pesquisa seguiu com uma análise exploratória (PCA) dos dados espectrais. A PCA também foi realizada com os dados de atributos de qualidade das fibras de algodão obtidas por HVI.

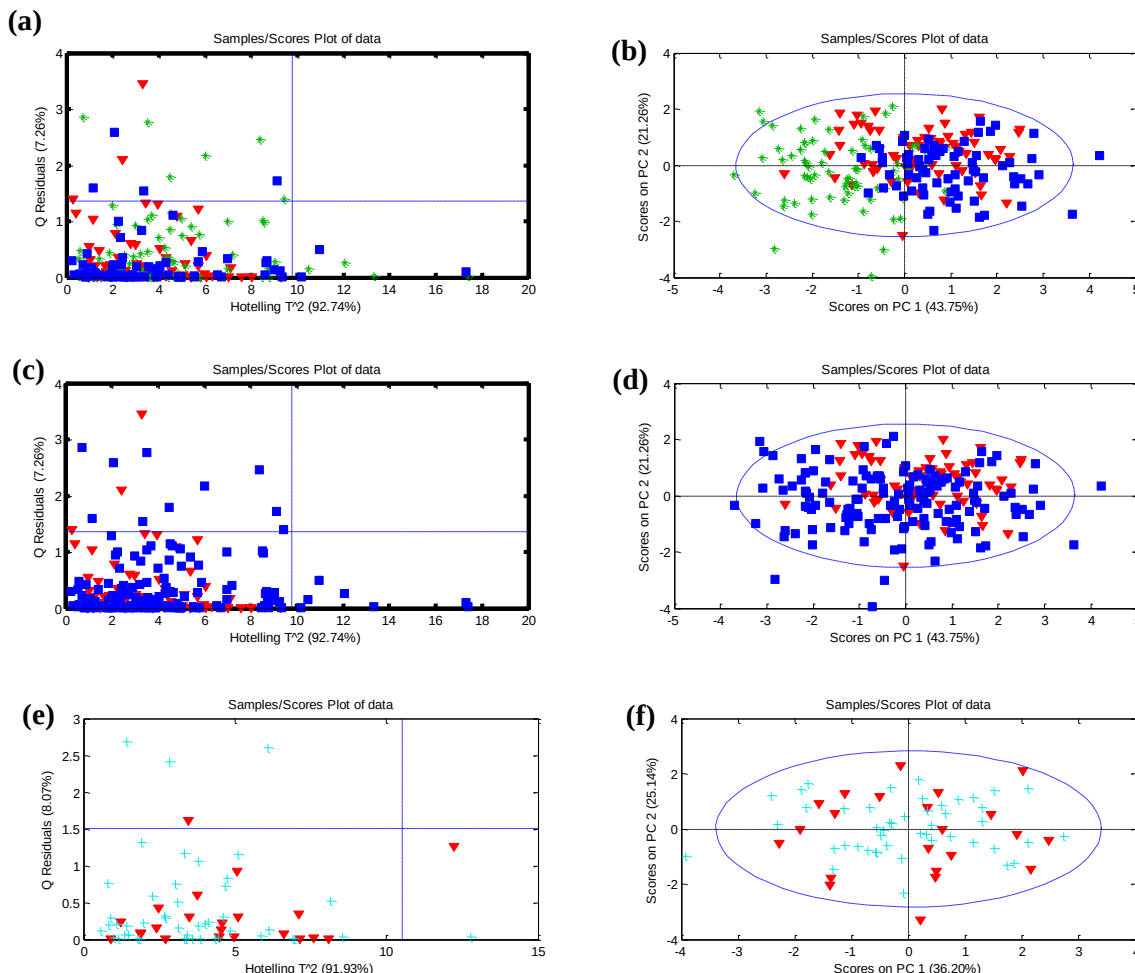
5.2 Análise exploratória dos dados

5.2.1 PCA dos dados de atributos de qualidade das fibras de algodão

Foi realizada uma análise exploratória com os dados de atributos de qualidades das fibras de algodão. Nesse estudo foram construídas PCA's referentes a cada classe das amostras. Os modelos foram construídos utilizando 05 variáveis, isto é, 05 atributos de qualidade das fibras de algodão. Esses atributos foram medidos através de um instrumento HVI, e são comprimento da fibra, índice de fibras curtas, micronaire reflectância e resistência.

Deste modo, organizou-se os dados em uma matriz constituída de 05 colunas (variáveis) e 213 linhas (amostras de fibras de algodão). A partir disso, modelou-se uma PCA aplicando o autoescalamento como pré-processamento. Os resultados dos modelos construídos com os dados de atributos de qualidade da fibra podem ser observados na Figura 14.

Figura 13 - PCA's aplicados a dados de atributos de qualidade das 213 amostras de fibras de algodão de diferentes classes. Gráfico de T2 de Hotelling vs Q residual para (a) amostras de diferentes localidades, (c) com ou sem modificação genética e (e) relativa à maturidade. Escores de PC1 vs PC2 para (b) amostras de diferentes localidades, (d) com ou sem modificação genética e (f) relativa à maturidade.



Fonte: Arquivo Próprio (2022). **(a) e (b)** Amostras de diferentes localidades: Localidade A (vermelhas), Localidade B (Verdes) e Localidade C (Azul); **(c) e (d)** Amostras convencionais (Vermelha) e transgênicas (Azul); **(e) e (f)** Amostras: Precoce (Vermelha) e tardio (Ciano).

Analisando os gráficos da Figura 14b, evidencia-se uma tendência de separação nos escores correspondentes as classes de localidade das fibras de algodão. Dentre as classes de localidade, nota-se que as amostras de B agrupam-se principalmente nos escores negativos de PC1, enquanto que as amostras da localidade C tendem a ficar nos escores positivos de PC1. Já em relação as amostras da localidade A ficam distribuídas com uma grande dispersão das amostras, não sendo observada tendência de agrupamento. Com relação a PCA para as amostras convencionais e transgênicas, nota-se uma grande dispersão das amostras e não é possível observar agrupamentos. Os valores correspondentes a variância explicada em PC1 é de 43,75 e PC2 de 21,26%. Assim como nos gráficos de escores da localidade algumas amostras ficaram fora da elipse de

confiança, mas sem apresentar altos valores de T^2 Hotelling e Q residual. Vale destacar que em ambas situações os modelos foram construídos utilizando 04 PC's, porém não foram observadas tendências de agrupamentos claros.

Na PCA construída com as amostras de diferentes maturidades não se observa uma tendência de separação. Os valores de variância explicada de PC1 foi de 36,20% e de PC2 25,14%.

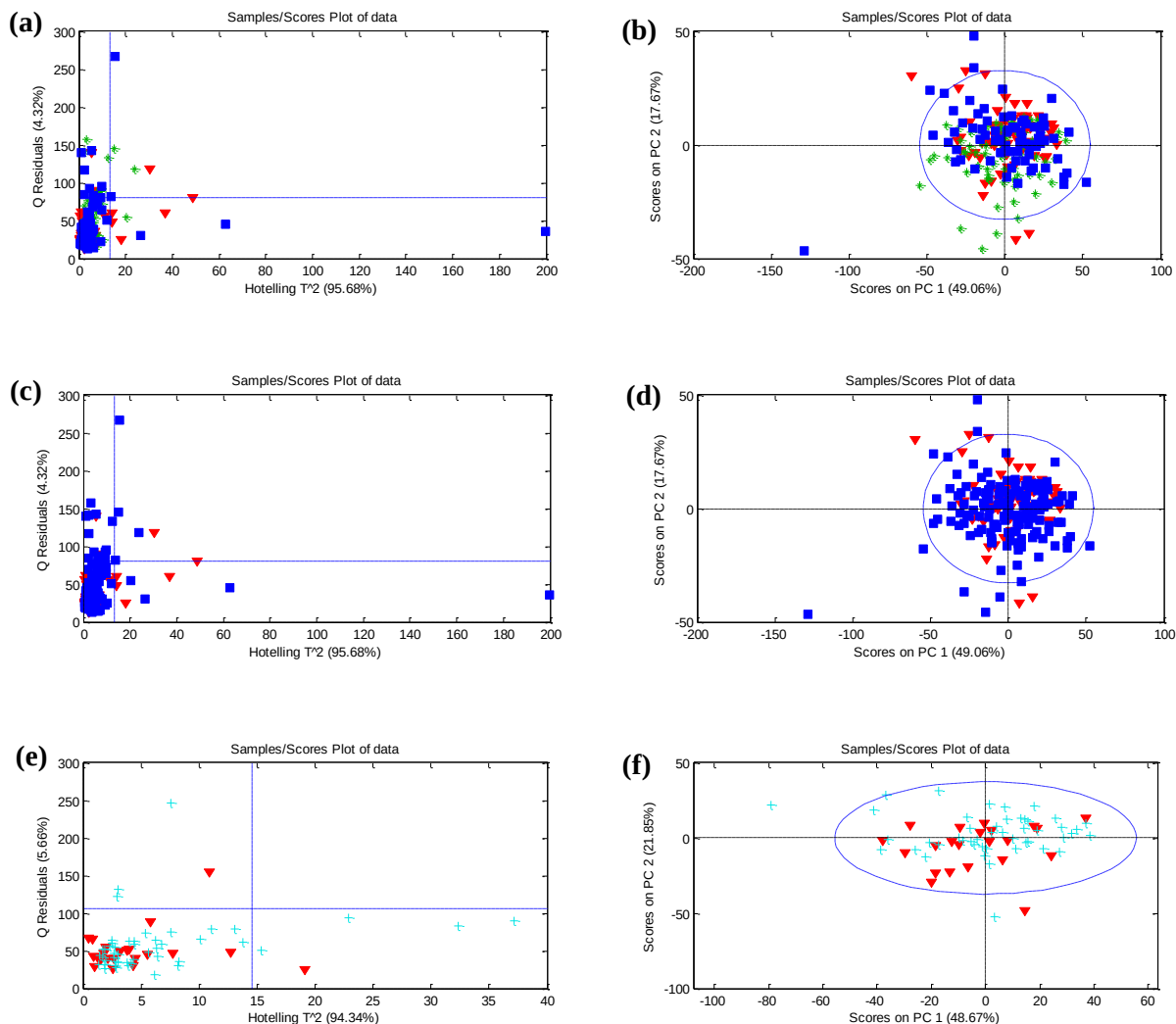
De forma geral podemos observar uma grande dispersão entre as amostras e com uso da PCA não foi possível ver alguma distribuição uniforme de amostras semelhantes. Apesar disso, foi construído modelos PLS-DAs para verificar a viabilidade de classificação das amostras em grupos.

5.2.2 PCA dos dados espectrais Raman das fibras de algodão

O NVU e o WLS foram utilizados como pré-processamentos padrão para construção das PCA's com os dados espectrais Raman das fibras de algodão. A combinação proporciona uma diminuição na relação sinal ruído e o WLS corrige o deslocamento da linha de base. Para cada classe de fibra foi construído uma PCA aplicando o pré-processamento padrão (NVU+WLS) e testado outras combinações de pré-processamentos, tais como suavização com polinômio Savitzky-Golay de 1ª ordem e janela de 21 pontos (SF21) e 31 pontos (SF31), SNV e MSC.

Após esses testes verificou-se que a combinação NVU+WLS+SNV+SF31 resultou nos melhores resultados, pois houve uma menor quantidade de amostras fora da elipse de confiança em todas as PCA's construídas; não foi constatado amostras com altos valores de T^2 de hotelling em concomitante com o Q residual. Assim, na Figura 15 é possível observar os modelos de PCA's para cada classe de fibra de algodão, as quais foram construídas utilizando 6 PC's.

Figura 14 - Gráficos de PCA's aplicados a dados espectrais Raman de 213 amostras de fibras de algodão de diferentes classes. Gráfico de T² de *Hotelling* vs Q residual para (a) amostras de diferentes localidades, (c) com ou sem modificação genética e (e) relativa à maturidade. Escores de PC1 vs PC2 para (b) amostras de diferentes localidades, (d) com ou sem modificação genética e (f) relativa à maturidade.



Fonte: Arquivo Próprio (2022). **(a) e (b)** Amostras de diferentes localidades: Localidade A (vermelhas), localidade B (Verdes) e localidade C (Azul); **(c) e (d)** Amostras convencionais (Vermelha) e transgênicas (Azul); **(e) e (f)** Amostras: Precoce (Vermelha) e tardio (Ciano).

Em todas as PCA's da Figura 15 foi verificado algumas amostras com altos valores de T² *hotelling* ou Q residual, além de se encontrarem fora da elipse de confiança, o que pode caracterizar um comportamento anômalo (BALLABIO, CONSONNI, 2013). De forma geral não foi observada a formação de agrupamentos bem definidos relativos as classes estudadas. Assim como para os dados de HVI, os dados obtidos por espectrometria Raman apresentaram uma grande dispersão, não apresentando agrupamentos claros entre as classes estudadas, sendo construído modelos PLS-DA

5.3 Modelo de classificação supervisionada

5.3.1 *PLS-DA dos atributos de qualidade das fibras de algodão (PLS-DA-At)*

Inicialmente para a distinção das fibras quanto a localidade do plantio, presença de modificação genética e maturidade os modelos PLS-DA foram desenvolvidos utilizando os dados dos parâmetros de qualidade obtidos por HVI. O modelo relativo à localidade possuía 3 classes (localidade A; localidade B; localidade C), presença de modificação genética (CONV: Convencionais e TRANG: Transgênicas) e maturidade (PRE: Precoce e TAR: Tardio).

Nos modelos PLS-DA construídos com os atributos de qualidade (PLS-DA-At) utilizou-se como pré-processamento o autoescalamento na matriz de dados. Para o modelo relativo à localidade foram utilizadas 04 VL que corresponderam 89,42% da variância dos dados, o mesmo número de VL foi adequado para os demais modelos, na qual a variância capturada foi de 85,35% para o modelo referente as modificações genéticas e 88,48% para amostras relativos à maturidade.

Isto posto, a qualidade do modelo de classificação supervisionada (PLS-DA-At) foi avaliado com base nos valores das figuras de méritos (Tabela 07) (SUBRAMANIAN; SILVA; COUTINHO, 2005)

Tabela 07 - Valores das figuras de mérito dos modelos PLS-DA-At construídos com dados de atributos de qualidade de diferentes classes de fibras de algodão

Etapas e parâmetros de qualidade do modelo			Classes dos Modelos PLS-DA-At						
			Diferentes Localidades		Modificações Genéticas		Diferenças na maturidade		
			A	B	C	CONV	TRAN	PRE	TARD
Modelo com todas as amostras	Treinamento	Sensibilidade	0,804	0,833	0,833	0,804	0,635	0,615	0,700
		Especificidade	0,583	0,798	0,745	0,635	0,804	0,615	0,700
		% Erro	30,6	18,4	21,0	28,0	28,0	34,2	34,2
	Teste	Sensibilidade	0,913	0,917	0,833	0,913	0,604	0,778	0,647
		Especificidade	0,708	0,894	0,723	0,604	0,913	0,747	0,778
		% Erro	19,0	9,49	22,2	24,1	24,1	28,8	28,8
Modelo sem outliers	Treinamento	Sensibilidade	0,783	0,830	0,813	0,804	0,632	0,750	0,633
		Especificidade	0,579	0,787	0,753	0,632	0,804	0,633	0,750
		% Erro	32,0	19,1	21,7	28,2	28,2	30,8	30,8
	Teste	Sensibilidade	0,913	0,917	0,792	0,870	0,604	0,889	0,529
		Especificidade	0,729	0,894	0,723	0,604	0,870	0,529	0,889
		% Erro	17,9	9,49	21,7	26,3	26,3	29,1	29,1

Fonte: Arquivo Próprio (2022). *dentificação das classes de fibras de algodão: CONV (Convencionais); TRAN (Transgênicas); PRE (Precoce); TARD (Tardio).

Dentre as medidas utilizadas para avaliar a capacidade preditiva do modelo, pode-se citar a sensibilidade, a especificidade e a porcentagem de erro na etapa de treinamento e teste. Nesse sentido, avaliando os dados dispostos na Tabela 07, os erros nas etapas de treinamento e teste são elevados em alguns modelos (modificações genéticas e maturidade). No modelo de localidade os resultados obtidos (sensibilidade, especificidade e %erro) podem ser úteis para ensaios de rotina, isto é, otimizar o tempo e custo de análise para distinguir fibras de diferentes localidades, tendo em vista a competitividade que a cotonicultura apresenta.

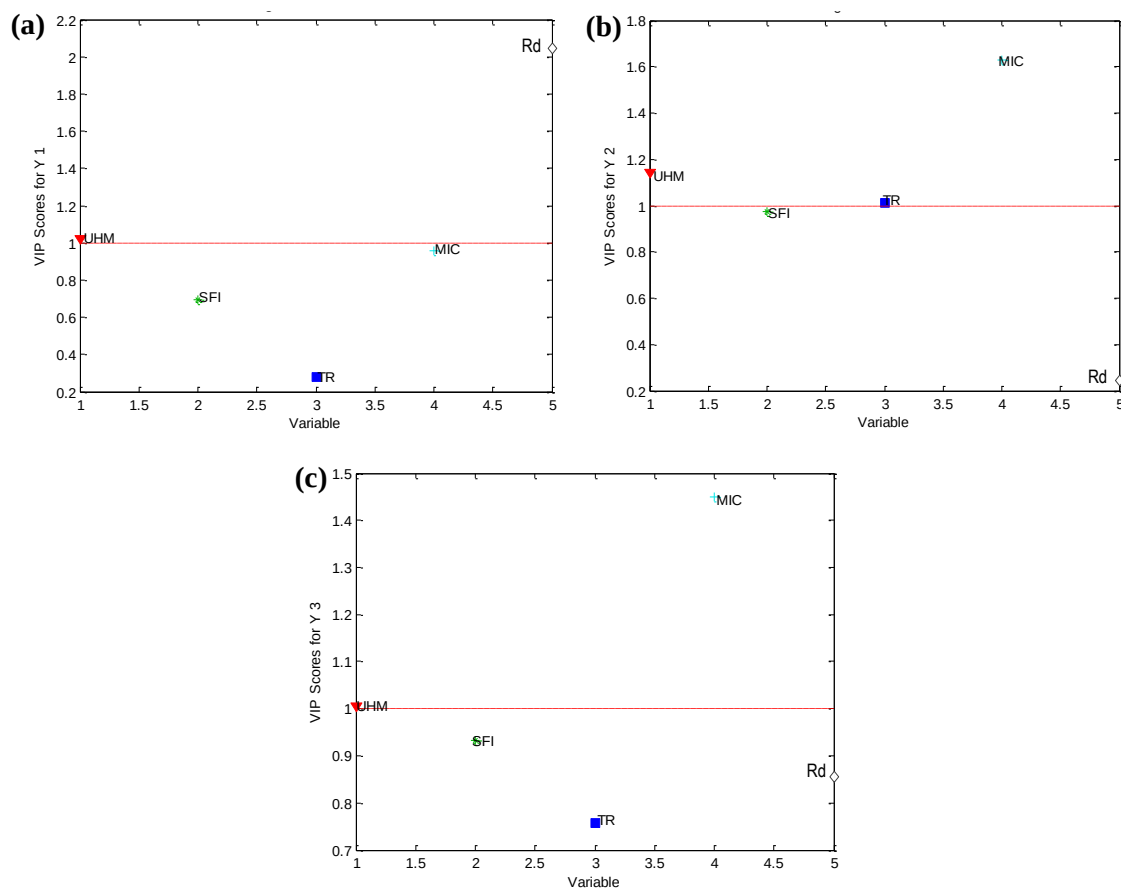
Com isso, constatou-se valores para a sensibilidade e especificidade para os modelos relativos as amostras B de 0,917 e 0,894 e para modelo da localidade C sensibilidade de 0,833 e 0,723 de sensibilidade. Esses valores resultaram em erros na etapa de teste de 9,49% (B) e 22,2 % (C). Por outro lado, a classe A apresentou uma especificidade muito baixa na etapa de treinamento, levando-a classificar os elementos da classe B e C como pertencentes a classe A. Apesar da retirada de *outlier*, o modelo não apresentou melhoras nos parâmetros de desempenho, tanto para as classes da localidade A, como para B e C.

Em relação ao modelo destinado a distinguir as amostras convencionais (CONV) das que sofreram modificações genéticas (TRAN), os modelos apresentaram na etapa de teste uma % de erro de aproximadamente 30% o que é reflexo dos valores de especificidade e sensibilidade alcançados. Apesar da retirada de amostra anômalas não se verificou melhoras significativas na capacidade preditiva dos modelos.

Similarmente, em relação a ocorrência de erros do tipo I (falso positivo) e II (falso negativo), foi observada para o modelo relativo à maturidade das fibras. Neste caso a retirada de amostras anômalas, baseadas nos gráficos de T^2 de *hotelings* e Q residual (Apêndice A), elevou a % de erro de classificação.

Na Figura 16 nota-se o gráfico de VIP's escores para os modelos construídos com as amostras de diferentes localidades. Nessa análise foi possível compreender a influência dos atributos de qualidade das fibras de algodão para a separação das classes.

Figura 15 - Gráficos de VIP escores dos dados de atributos de qualidade das fibras de algodão da localidade A (a); gráfico de VIP escores das amostras da localidade B (b); gráfico de VIP escores das amostras da localidade C (c)

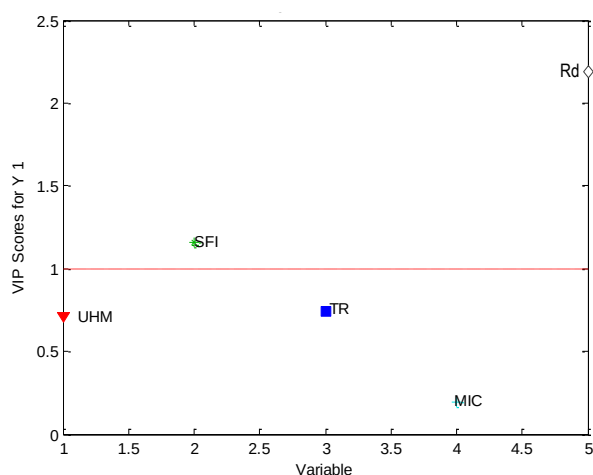


Fonte: Arquivo Próprio (2022). Parâmetros de comprimento (UHM); índice fibras curtas (SFI); Resistência (TR); Índice de micronaire (MIC); Reflectância (Rd).

De acordo com o gráfico da Figura 16a o parâmetro de qualidade reflectância (Rd) é o que apresenta maior relevância para a classe de amostras cultivadas na localidade A, enquanto o parâmetro micronaire (MIC) foi o mais relevante para as classes de amostras da localidade B e C.

Em relação às amostras das classes CONV e TRAN os parâmetros mais significativos foram reflectância (Rd) e índice de fibras curtas (SFI) Figura 17.

Figura 16 - Gráficos de VIP escores dos dados de atributos de qualidade das fibras de algodão com modificações genéticas

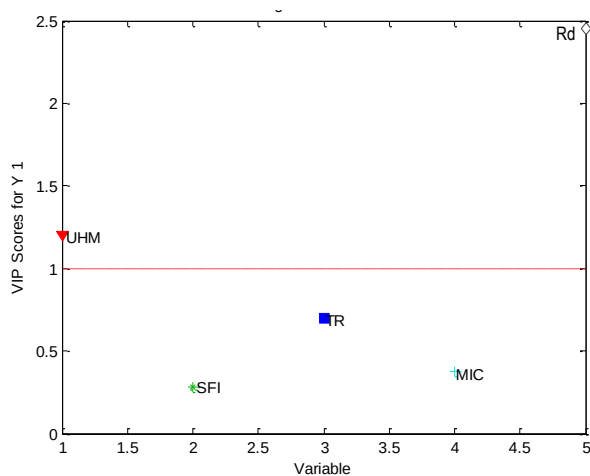


Fonte: Arquivo Próprio (2022). Parâmetros de comprimento (UHM); índice fibras curtas (SFI); Resistência (TR); Índice de micronaire (MIC); Reflectância (Rd).

A %SFI é um importante atributo para a indústria têxtil na avaliação da qualidade das fibras de algodão, que pode prever a capacidade de fabricação do fio. Este parâmetro pode ser influenciado pelo tipo de genótipo do algodão e por isto pode ser importante para distinguir fibras provenientes de genótipos convencionais e modificados geneticamente.

Já para o modelo relativo à maturidade os parâmetros mais relevantes foram reflectância (Rd) e comprimento (UHM), Figura 18.

Figura 17 - Gráficos de VIP escores dos dados de atributos de qualidade das fibras de algodão de diferentes maturidades



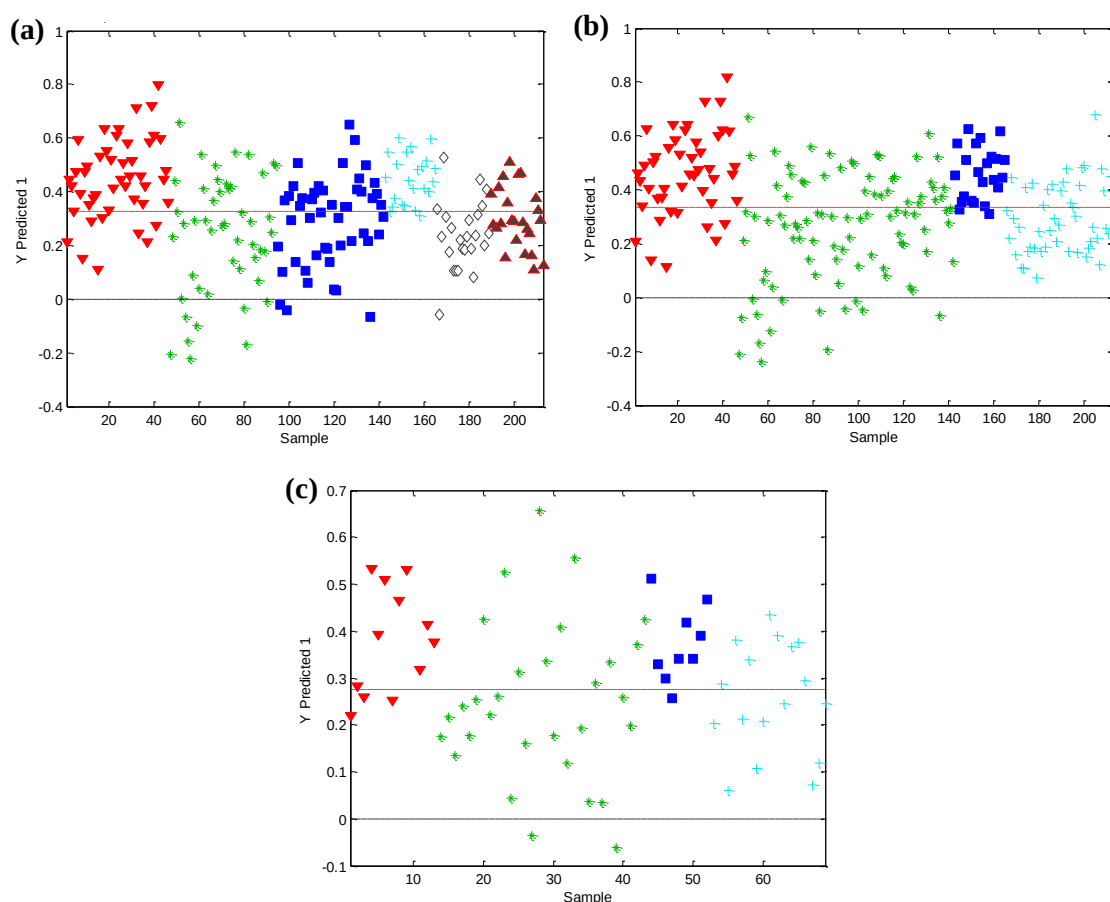
Fonte: Arquivo Próprio (2022). Parâmetros de comprimento (UHM); índice fibras curtas (SFI); Resistência (TR); Índice de micronaire (MIC); Reflectância (Rd).

O alongamento (comprimento) é uma das etapas de desenvolvimento fisiológico da fibra de algodão que pode variar de acordo com o tipo de genótipo e o ambiente e por este motivo pode ser um importante parâmetro de discriminação entre fibras de algodão de diferentes maturidades.

O grau de reflectância com grande relevância para os modelos relativos à modificação genética e maturidade é um dos atributos relacionados com a determinação da cor da fibra do algodão e é bastante importante na classificação do algodão.

A avaliação da capacidade preditiva dos modelos PLS-DA-At deu-se a partir da aplicação modelos construídos na etapa de treinamento para prever as classes para as amostras teste. Para isso, foram considerados os resultados com todas as amostras cujas os gráficos de predição podem ser observados na Figura 19a, 19b e 19c.

Figura 18 - Gráficos de predição do modelo PLS-DA-At das classes de fibras de algodão com diferentes localidades (a), modificações genéticas (b) e diferença na maturidade (c)



Fonte: Arquivo Pessoal (2022). Em **(a)** Conjunto de treinamento: Vermelhas (A); Verdes (B); Azuis (C). Conjunto de teste: Ciano (A); Brancas (B); Vermelho-preto (C). Em **(b)** Conjunto de Treinamento: Vermelhas (CONV); Verdes (TRAN). Conjunto de Teste: Azuis (CONV); Ciano (TRAN). Em **(c)** Conjunto de Treinamento: Vermelhas (PRE); Verdes (TARD). Conjunto de Teste: Azuis (PRE); Ciano (TARD).

Na Figura 19a, pode-se observar que as amostras das três localidades estão bastante dispersas e que a um limiar de 0,4 é possível separar a classe das amostras A das B e C. Mesmo na etapa de treinamento essa separação não é inequívoca. Esta dispersão das amostras ocorre para os três modelos o que dificulta a classificação correta. Levando a erros do tipo I e tipo II. Comparando com a PCA (Figura 14), havia-se uma tendência de separação entre as amostras B e A, porém no PLS-DA-At essa discriminação não foi claramente observada em ambas as etapas.

A respeito da Figura 19b, as amostras convencionais apresentaram uma boa classificação na etapa de teste, com a maioria das amostras classificadas corretamente. Por outro lado, as amostras transgênicas não apresentaram comportamento semelhante, tendo diversas amostras classificadas erroneamente, ou seja, amostras transgênicas classificadas como convencionais. Isso pode justificar os valores baixos de sensibilidade e especificidade.

Na Figura 19c, obteve-se uma boa classificação das amostras precoce. Nessa classe a melhor discriminação ocorre na etapa de teste. Com relação a classe tardias observa-se uma dispersão das amostras, na qual o modelo não distingue claramente as amostras TARD no limiar estabelecido nas etapas de treinamento e teste.

A fim de obter-se modelos com melhores capacidades preditivas, a espectrometria Raman portátil foi testada.

5.3.2 PLS-DA dos dados espectrais Raman das fibras de algodão (PLS-DA-Raman)

A construção do modelo de PLS-DA utilizando os dados espectrais deu-se pela necessidade de identificar possíveis agrupamentos, tendo em vista que no modelo PLS-DA-At e na análise exploratória isso não foi evidenciado claramente, inviabilizando a aplicação de modelos baseados em PCA como é o caso do SIMCA. Com isso, foram construídos modelos para diferenciar as classes relativa a localidade das amostras (localidade A; localidade B; localidade C), presença de modificação genética (Convencionais e Transgênicas) e maturidade (Precoce e Tardio).

Para construção desses modelos foram testadas diferentes combinações de pré-processamentos afim de fornecer a melhor capacidade preditiva aos modelos. Após os testes verificou-se que a combinação dos pré-processamentos NVU+WLS+SNV+SF31 resultou nos modelos com as melhores capacidades preditivas. Além disso, também foi

avaliado um número de VL's que melhor explicasse a variância dos dados (FABER; RAJKÓ, 2007).

Foram utilizadas 07 VL para construção dos respectivos modelos. No modelo referente a classe de localidade das amostras a variância explicada foi de 96,61%, esse valor também foi identificado para o modelo relativo à presença de modificação genética. Diferentemente, o modelo de maturidade das amostras foi construído com 05 VL e apresentou uma variância explicada de 88,08%.

Com os modelos construídos, a sua capacidade preditiva foi avaliada considerando os valores das figuras mérito tanto da etapa de treinamento quanto na etapa de teste (Tabela 08). Além disso, também foi construído os modelos com a retirada de alguns possíveis *outliers*.

Tabela 08 - Valores das figuras de mérito dos modelos PLS-DA-Raman construídos com dados espectrais Raman de diferentes classes de fibras de algodão

Etapas e parâmetros de qualidade do modelo			Classes dos Modelos PLS-DA-Raman						
			Diferentes Localidades			Diferenças Genéticas		Diferenças na maturidade	
			A	B	C	CONV	TRAN	PRE	TARD
Modelo com todas as amostras	Treinamento	Sensibilidade	0,913	0,917	0,667	0,913	0,823	0,923	0,883
		Especificidade	0,771	0,862	0,787	0,823	0,913	0,883	0,923
		% Erro	15,8	11,1	27,3	13,2	13,2	12,2	12,2
	Teste	Sensibilidade	0,826	0,927	0,458	0,826	0,750	0,889	0,765
		Especificidade	0,813	0,638	0,872	0,750	0,826	0,765	0,889
		% Erro	18,1	22,3	33,5	21,2	21,2	17,3	17,3
Modelo sem outliers	Treinamento	Sensibilidade	0,913	0,875	0,702	0,935	0,842	0,923	0,862
		Especificidade	0,789	0,882	0,745	0,842	0,935	0,862	0,923
		% Erro	14,9	12,2	27,7	11,1	11,1	10,7	10,7
	Teste	Sensibilidade	0,826	0,917	0,667	0,783	0,708	0,778	0,765
		Especificidade	0,750	0,723	0,809	0,708	0,783	0,765	0,778
		% Erro	21,1	18,0	26,2	25,5	25,5	22,8	22,8

Fonte: Arquivo Próprio (2022). Identificação das classes de fibras de algodão: CONV (Convencionais); TRAN (Transgênicas); PRE (Precoce); TARD (Tardio)

Avaliando os dados da Tabela 08, verifica-se que em alguns modelos com dados espectrais Raman apresentaram valores de sensibilidade e especificidade melhores em relação aos modelos com os dados de atributos. Assim, levando em consideração a rapidez que o método tem em fornecer resposta sobre a análise do perfil produtivo do algodão, juntamente com o baixo custo de análise e possibilidade de análises em campo, os modelos podem ser úteis para aplicações e contribuir para controle de qualidade no setor produtivo, como por exemplo: na distinção de amostras convencionais e transgênicas; na classificação de amostras com diferentes maturidades e a localidade de produção.

Sendo assim, os principais resultados obtidos no modelo relativo à localidade A foi a sensibilidade de 0,826 e especificidade 0,813 culminando em uma porcentagem de erro de 18%, sendo o erro mais baixo em relação aos modelos das outras localidades. Nesses, verificou-se que os valores de sensibilidade e especificidade do modelo da localidade B podem levar a erros do tipo II, ou seja, classificar como pertencente a classe B os elementos da classe A e C. Para esta classe foi obtido uma porcentagem de erro de 22,3% na etapa de teste. Já para a classe C foi alcançado um baixo valor de sensibilidade, levando a erros tipo I, podendo levar o modelo a classificar amostras pertencentes a classe C como não pertencentes a esta classe. Para esta classe foi obtido o maior valor de porcentagem de erro, 33,5%.

Para o modelo referente à modificação genética foram obtidos na etapa de teste valores de sensibilidade e especificidade que resultaram em uma porcentagem de erro de 21,2%. Assim, em uma situação com demanda alta de análise e agilidade na obtenção das respostas sobre fibras convencionais e transgênicas o método pode ser adequado e a porcentagem de erro considerada.

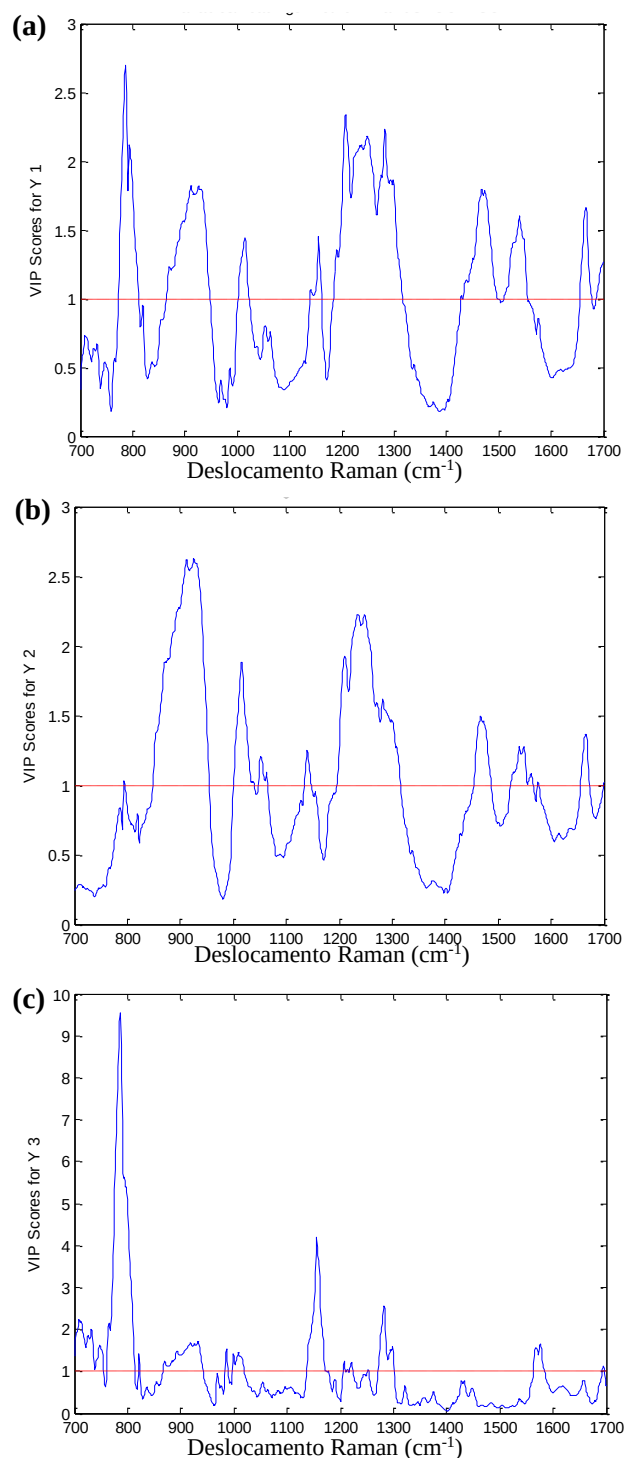
Para o modelo referente a maturidade foram obtidos valores de sensibilidade e especificidade de 0,8 e 0,7 nas etapas de treinamento e teste, respectivamente, e porcentagens de erro classificação de 12% e 17% foram obtidos nas duas etapas. Em comparação com os demais modelos com dados espectrais Raman, a classe de maturidade apresentou a menor porcentagem de erro. Podendo então ser uma alternativa economicamente viável para classificação de fibras de diferentes maturidades.

A presença de amostras anômalas foi avaliada de acordo com os gráficos de T^2 de *hotelling* e Q residual (Apêndice **B**). Amostras com altos valores de T^2 de *Hotelling* e Q residual foram excluídas do conjunto de dados já que valores de Q residual elevado

implica em resíduo que não pode ser explicado pelo modelo e altos valores de T^2 de *Hotelling* indica que a amostra possui grande influência sobre o modelo (BALLAIO; CONSONNI, 2013). Após a exclusão das amostras anômalas o modelo PLS-DA-Raman foi construído e os valores para as figuras de méritos descritas na Tabela 08. Pode-se observar que em sempre a exclusão de amostras levou a melhores resultados para as figuras de mérito e consequentemente não levaram a uma melhor capacidade preditiva para os modelos.

A fim de compreender quais variáveis espectrais contribuíram para a distinção das classes estudadas nos modelos, analisou-se os gráficos dos VIP's escores. Na Figura 20 verifica-se os VIP's escores referentes as amostras das três localidades.

Figura 19 - Gráficos de VIP escores dos modelos PLS-DA-Raman das fibras de algodão da localidade A (a); gráfico de VIP escores das amostras da localidade B (b); gráfico de VIP escores das amostras da localidade C (c).



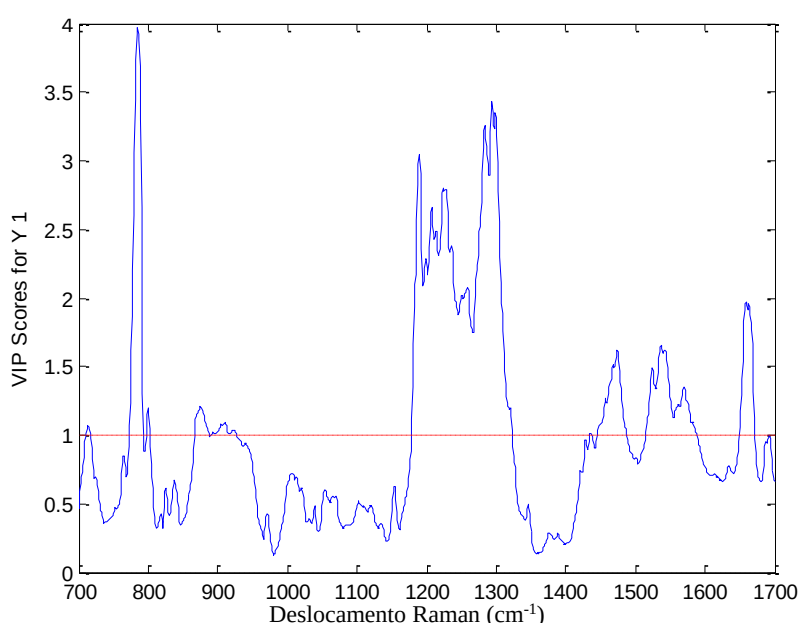
Fonte: Arquivo Próprio (2022)

Analisando os gráficos da Figura 20 nota-se que em todas as localidades das fibras de algodão há informações espectrais que contribuem para discriminação do modelo. Os três gráficos têm perfil bastante semelhantes, podendo ser observada apenas mudança nas

intensidades dos picos. As regiões espectrais mais significativas apresentam valores superiores a 1. Com isso, verifica-se que a região de 800 cm^{-1} (H_2O) a 900 cm^{-1} (celulose) tem influência espectral tanto nas fibras da localidade A quanto na localidade C. Já na localidade B, as principais influências espectrais estão relacionadas as regiões correspondentes a celulose (900 cm^{-1} , 1200 a 1300 cm^{-1}) e ligações γ COC (900 cm^{-1}).

Em relação aos gráficos de VIP escores para o modelo destinado a distinguir entre as amostras convencionais e geneticamente modificadas, as regiões mais informativas podem ser observadas na Figura 21.

Figura 20 - Gráficos de VIP escores dos dados espectrais Raman das fibras de algodão de diferenças genéticas

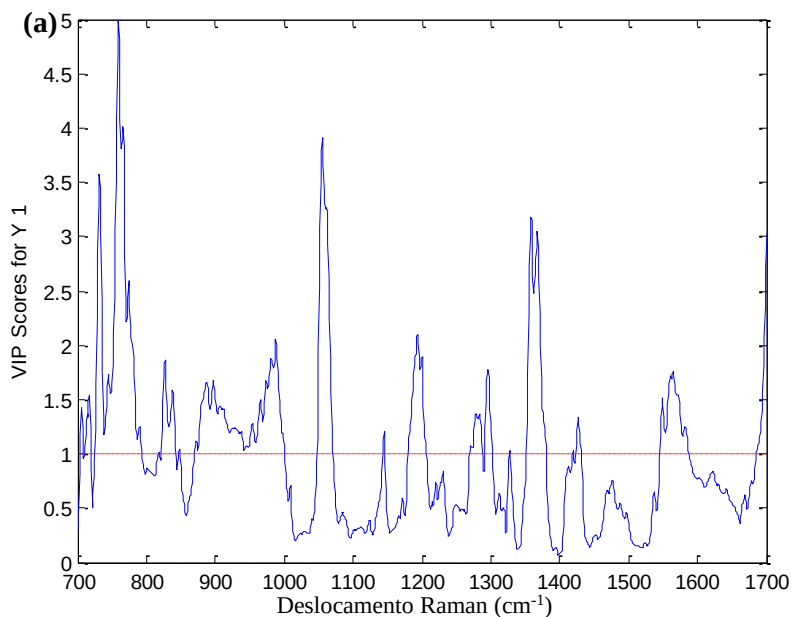


Fonte: Arquivo Próprio (2022).

Observa-se que as variáveis que possuem maior contribuição para a separação das classes convencionais e transgênicas estão nas regiões de 750 a 800 cm^{-1} , porém não foi verificado espécies químicas com sinais nessa região; nas regiões de 1200 a 1300 cm^{-1} cuja os picos podem estar relacionados a celulose; na região de 1660 cm^{-1} que está associada a proteínas e hidrocarbonetos.

Para o modelo relativo à maturidade, o gráfico de VIP escores pode ser observado na Figura 22.

Figura 21 - Gráficos de VIP escores dos dados espectrais Raman das fibras de algodão de diferenças de maturidade

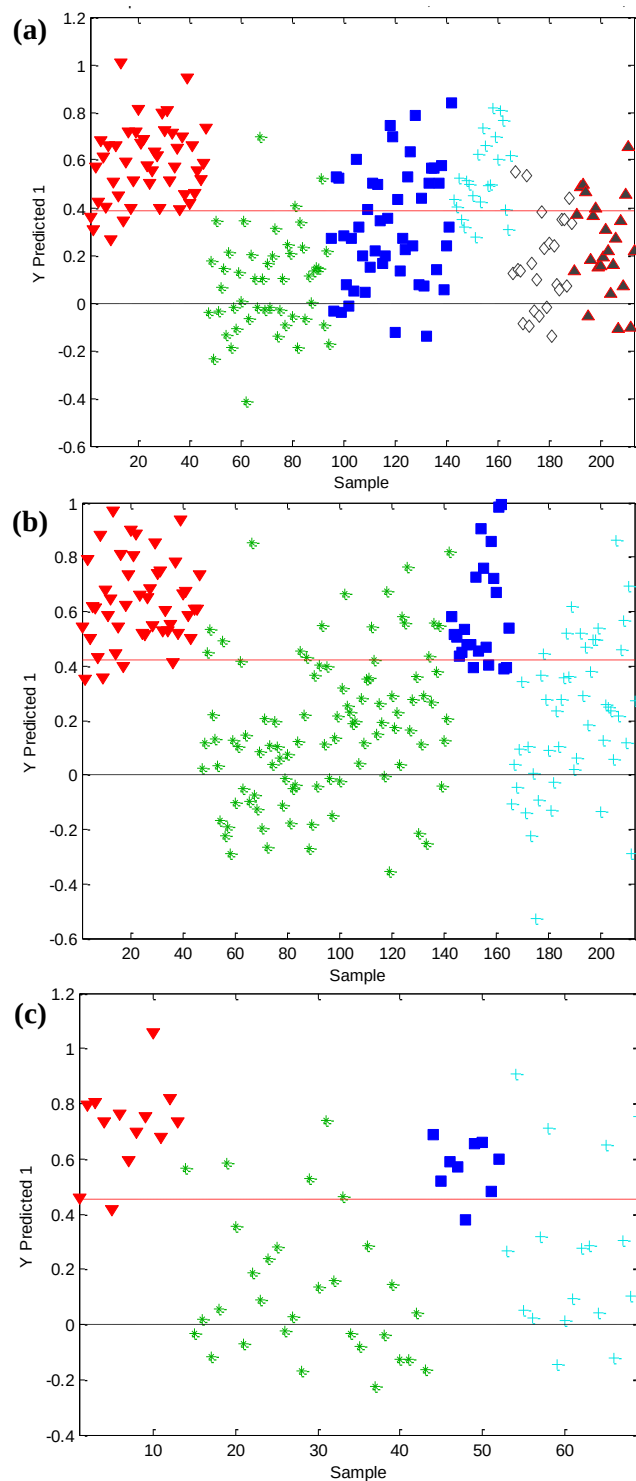


Fonte: Arquivo Próprio (2022).

As variáveis espectrais com maiores contribuições para a discriminação entre as classes precoce e tardio foram as bandas entre 700 e 800 cm^{-1} , 1050 cm^{-1} , 1200 cm^{-1} , 1350 cm^{-1} e 1550 cm^{-1} . Algumas dessas regiões são referentes as moléculas de celulose constituídas nas fibras de algodão.

A avaliação da capacidade preditiva dos modelos foi realizada aplicando-se os modelos construídos as amostras de treinamento para prever a classe de amostras de testes. Os resultados das predições podem ser observados na Figura 23a, 23b e 23c.

Figura 22 - Gráficos de predição dos espectros Raman de fibras de algodão de diferentes localidades (a), diferença genética (b) e diferença na maturidade (c)



Fonte: Arquivo Pessoal (2022). Em **(a)** Conjunto de treinamento: Vermelhas (A); Verdes (B); Azuis (B). Conjunto de teste: Ciano (A); Brancas (B); Vermelho-preto (C). Em **(b)** Conjunto de Treinamento: Vermelhas (CONV); Verdes (TRAN). Conjunto de Teste: Azuis (CONV); Ciano (TRAN). Em **(c)** Conjunto de Treinamento: Vermelhas (PRE); Verdes (TARD). Conjunto de Teste: Azuis (PRE); Ciano (TARD).

Mais uma vez pode-se afirmar que há uma grande dispersão entre as amostras. A Figura 23a que corresponde a predição das classes relativas à localidade, percebe-se que ocorre uma discriminação clara entre das amostras A e as de B e C. Esta distinção é observada tanto na etapa de treinamento quanto de teste. Já as classes B e C não apresentam distinção clara no limiar estabelecido pelo modelo.

Para o modelo destinado à distinção entre as amostras convencionais e modificadas geneticamente, Figura 23b, foi possível observar uma menor dispersão entre as amostras convencionais, enquanto as transgênicas se mostraram mais dispersas. Este comportamento é observado tanto na etapa de treinamento quanto de teste. Isso também é identificado na Figura 23c, na qual nas etapas de teste e treinamento há uma baixa dispersão entre as amostras precoce, enquanto as amostras tardias ficaram mais dispersas.

Na Tabela 09 verifica-se uma comparação entre a capacidade preditiva (etapa de teste) dos modelos construídos com os dados obtidos por HVI e os dados obtidos por espectroscopia Raman.

Tabela 09 - Comparação das figuras de mérito na etapa de teste dos modelos PLS-DA-Raman e PLS-DA-At

Parâmetros de qualidade do modelo		Modelos PLS-DA-At						
		Diferentes Localidades			Diferenças Genéticas		Diferenças na maturidade	
		A	B	C	CONV	TRAN	PRE	TARD
Todas as amostras	Sensibilidade	0,913	0,917	0,833	0,913	0,604	0,778	0,647
	Especificidade	0,708	0,894	0,723	0,604	0,913	0,747	0,778
	% Erro	19,0	9,49	22,2	24,1	24,1	28,8	28,8
Sem <i>outliers</i>	Sensibilidade	0,913	0,917	0,792	0,870	0,604	0,889	0,529
	Especificidade	0,729	0,894	0,723	0,604	0,870	0,529	0,889
	% Erro	17,9	9,49	21,7	26,3	26,3	29,1	29,1
Parâmetros de qualidade do modelo		Modelos PLS-DA-Raman						
		Diferentes Localidades			Diferenças Genéticas		Diferenças na maturidade	
		A	B	C	CONV	TRAN	PRE	TARD
Todas as amostras	Sensibilidade	0,826	0,927	0,458	0,826	0,750	0,889	0,765
	Especificidade	0,813	0,638	0,872	0,750	0,826	0,765	0,889
	% Erro	18,1	22,3	33,5	21,2	21,2	17,3	17,3
Sem <i>outliers</i>	Sensibilidade	0,826	0,917	0,667	0,783	0,708	0,778	0,765
	Especificidade	0,750	0,723	0,809	0,708	0,783	0,765	0,778
	% Erro	21,1	18,0	26,2	25,5	25,5	22,8	22,8

Fonte: Arquivo Próprio (2022). Identificação das classes de fibras de algodão: CONV (Convencionais); TRAN (Transgênicas); PRE (Precoce); TARD (Tardio)

De forma geral os modelos destinados a distinguir as amostras em relação a localidade com base nos dados de HVI mostraram um desempenho melhor do que os baseados em Raman apresentando erros de predição em torno de 20%. Já os modelos destinados à distinção das amostras em relação maturidade e modificação genética com base na espectrometria Raman mostraram menores erros de predição e consequentemente melhor capacidade preditiva em comparação com os modelos construídos com os dados de HVI.

5.4 Calibração multivariada: modelos PLS's

A construção dos modelos de calibração PLS deu-se a partir da relação dos dados obtidos nas análises de espectroscopia Raman com os atributos de qualidade: Comprimento (UHM), resistência (g.tex^{-1}), micronaire (%MIC), índice de fibras curtas (%SFI) e refletância (%Rd) das fibras de algodão, sendo medidos através do método de referência HVI. Com isso, na Tabela 10 estão esquematizados os valores dos cinco atributos de qualidade referente: a faixa de trabalho; a média aritmética; desvio padrão (DP); coeficiente de variação (%CV).

Tabela 10 - Medidas dos atributos de qualidade das fibras de algodão pelo método HVI

Atributos de qualidade das fibras de algodão	V_{MIN} e V_{MAX}	Média ± DP	CV (%)
Comprimento (UHM)	28,6 – 34,3	31,25 ± 1,35	2,9
Índice de Fibras Curtas (%SFI)	5,6 – 9,9	7,12 ± 0,62	8,55
Índice de Micronaire (%MIC)	3,3 – 5,0	4,25 ± 0,155	6,45
Resistência (g.tex^{-1})	26,9 – 37,9	31,85 ± 2,15	6,7
Refletância (%Rd)	72,7 – 83,9	60,32 ± 1,58	2,00

Fonte: Arquivo Próprio (2022). V_{MIN} e V_{MAX}, Valores mínimos e máximos; DP, desvio padrão; CV coeficiente de variação.

Para construção dos modelos PLS foram testadas diversas combinações de pré-processamento com os dados espectrais Raman. Sendo então selecionado a combinação com resultante na melhor predição de cada atributo de qualidade. Também é importante a escolha correta do número de VL's, pois um número menor do que o ideal pode sobajustar o modelo, enquanto que um número maior do que o ideal pode sobreajustar (FABER; RAJKÓ, 2007).

Para construção dos modelos PLS's foi selecionado o número de VL's que resultasse nas melhores respostas para RMSEC, RMSEP, R^2 e Bias. Além disso, é importante que a porcentagem de variância explicada seja superior a 90% (SENA; ALMEIDA, 2018). Portanto, para o modelo: PLS-UHM 07 VL explica 97,7%; PLS-SFI 09 VL a variância foi de 97,1%; PLS-MIC 08 VL explica 96,97%; PLS-TR 06 VL explica 91,33%; PLS-Rd 07 VL a variância explicada foi de 96,45%.

Diante do exposto, na Tabela 11 nota-se os valores das figuras de mérito para os modelos PLS's construídos para cada atributo de qualidade das fibras de algodão, como também as combinações de pré-processamento que proporcionaram os melhores resultados para cada modelo.

Tabela 11 - Resultados obtidos no modelo PLS para os parâmetros de comprimento (UHM), índice de fibras curtas (SFI), micronaire (MIC), resistência (TR) e reflectância (Rd)

Modelos PLS	VL	Parâmetros avaliativos dos modelos					Pré-processamento Aplicado
		RMSEC	RMSEP	Bias	R^2		
PLS-UHM	07	1,094	0,784	0,066	0,092		NVU+WLS+SF31
PLS-SFI	09	0,590	0,525	0,022	0,263		NVU+WLS+SF15
PLS-MIC	08	0,280	0,665	0,032	0,107		NVU+WLS+SNV+SF31
PLS-TR	06	2,088	2,032	-0,234	0,087		NVU+WLS+SNV+SF15
PLS-Rd	07	1,694	1,434	-0,095	0,093		NVU+WLS+SF15

Fonte: Arquivo Próprio (2022). NVU: Normalização pelo vetor unitário; WLS: Correção de linha de base; SF15: Suavização com polinômio Savitzky-Golay de 1ª ordem e janela de 15 pontos e SF31: Suavização com polinômio Savitzky-Golay de 1ª ordem e janela de 31; SNV.

Analisando os dados da Tabela 11, nota-se um baixo valor no coeficiente de determinação (R^2) em todos os modelos PLS construídos. Os valores de R^2 mostraram que as informações químicas presentes na matriz de dados Raman utilizada não estão relacionadas com as variações nos parâmetros do HVI. Pois como destaca Subramanian, Silva e Coutinho (2005) é importante que os valores R^2 estejam próximo de 1, pois trata-se uma medida sobre a proximidade dos dados com linha de regressão, estatisticamente, a partir do R^2 compreende-se o quanto o modelo é capaz de explicar a correlação entre os valores do método de referência com os modelos multivariados desenvolvidos. Nesse sentido, o modelo PLS-SFI e PLS-MIC apresentam os valores mais altos para R^2 de 0,263 (%SFI) e 0,107 (%MIC), porém ainda distante do que é considerado ideal para correlação dos dados.

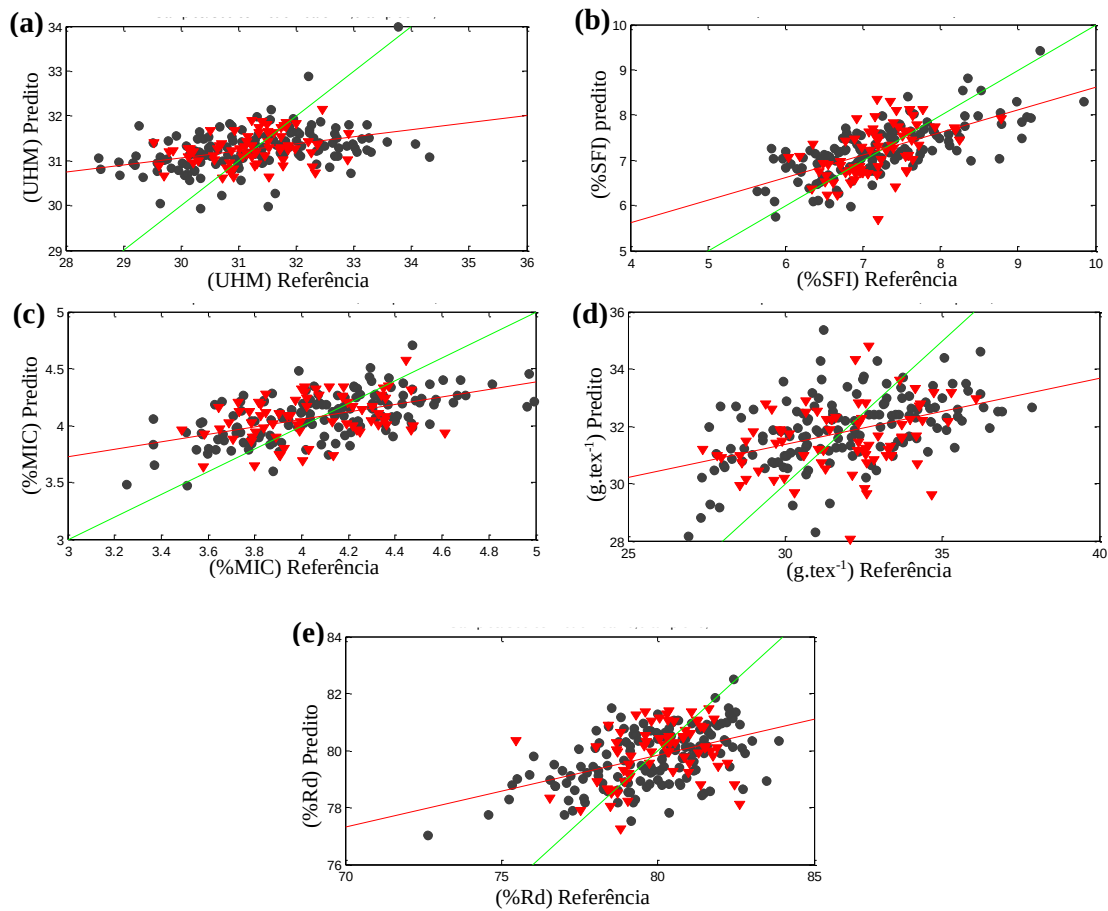
Ainda na Tabela 11, verifica-se que o menor RMSEP obtido foi no modelo para índice de fibras curtas atingindo 0,525 (%SFI). Comparando com o RMSEC verifica-se que não há discrepância entre os valores. Isso também pode ser observado para os demais modelos construídos. Com relação aos Bias, novamente o modelo PLS-SFI destaca-se com o menor valor obtido 0,022 (%SFI). A avaliação desse parâmetro possibilita a identificação de erros sistemáticos no modelo através do cálculo da diferença média e o valor verdadeiro (CURRIE, 1999).

Também pode-se avaliar a qualidade do modelo comparando-o com a CV% (VALDERRAMA; BRAGA; POPPI, 2009). Deste modo, nota-se que todos os modelos apresentaram os valores de RMSEC e RMSEP inferiores ao seu respectivo CV% (destacado na Tabela 09).

No estudo dos modelos PLS's construídos foi avaliado a presença de possíveis *outliers*. A presença de amostras com comportamentos anômalos na calibração multivariada prejudica ainda mais o modelo afetando a sua capacidade preditiva (MARTENS; NAES, 1996; VALDERRAMA; BRAGA; POPPI, 2007). Para este estudo foi avaliado os gráficos de T^2 *hotelling* e Q residual (Apêndice C) de todos os modelos PLS construído. Não foi identificado amostras com altos valores em concomitante dos parâmetros mencionados, mas algumas amostras com valores acima dos limites de confiança T^2 *hotelling*. Assim, construiu-se os modelos sem as referidas as amostras, porém não apresentou melhoras significativas na capacidade preditiva do modelo.

Além desses resultados, um comparativo sobre a qualidade dos modelos foi realizado através das observações na Figura 24. Nela, verifica-se a relação entre os valores previsto pelo modelo com os resultados medidos pelo método de referência HVI, cuja os atributos de qualidade avaliados são: UHM, g.tex⁻¹, MIC, SFI, %Rd.

Figura 23 - Gráficos dos valores previstos versus os valores de referência HVI obtidos no modelo PLS para os parâmetros UHM, g.tex-1, MIC, SFI e %Rd



Fonte: Arquivo Próprio (2022). PLS-UHM (a); PLS-SFI (b); PLS-MIC (c); PLS-TR (d); PLS-Rd (e). Amostras da Calibração (▲); Amostras de Predição (●).

Analisando os gráficos Figura 24, verifica-se em todos os modelos há amostras espalhadas na reta, porém essa distribuição ocorre com grandes erros, justificando os baixos valores R^2 identificado nos modelos. Esta baixa correlação entre a matriz dos dados espectrais e o vetor y, do parâmetro de qualidade pode ser atribuída a dispersão entre os espectros que não foi possível de ser corrigida através de métodos matemáticos.

Diante dessa análise, percebe-se que os modelos necessitam de melhorias para serem capazes de prever satisfatoriamente os parâmetros de qualidade da fibra de algodão utilizando um espectrômetro Raman portátil.

6 CONSIDERAÇÕES FINAIS

De acordo com a avaliação exploratória dos dados por meio de PCA, verifica-se que não há uma tendência de agrupamento entre as amostras. Essa análise foi realizada tanto com dados espectrais Raman (medidos por um Raman portátil), como também dados das variáveis de qualidade de fibras de algodão (medidos por um HVI). A partir da PCA foi observada uma grande dispersão entre as amostras e ausência de agrupamentos bem definidos em relação as classes estudadas. Por conta disso, foram construídos modelos de classificação supervisionada discriminante PLS-DA-Raman e PLS-DA-At.

Nesses modelos, os valores de sensibilidade e especificidade obtidos com o PLS-DA-Raman são baixos intrinsecamente, mas em uma perspectiva de aplicação para o setor produtivo da fibra de algodão os modelos construídos podem ser utilizáveis devidos os benefícios que o método apresenta: rapidez na análise, baixo custo, não precisar de preparo de amostras, possibilidades de análises em campo e facilidade do uso. Nessa perspectiva, os modelos são capazes de discriminar principalmente amostras convencionais e transgênicas com uma porcentagem de 79% de acerto; com 82,7 % de acerto em discriminar amostras precoce e tardio; classificar amostras da localidade B com 82% de acerto.

Já o modelo PLS-DA-At consegue ter uma boa discriminação para as amostras de diferentes localidades, pois apresenta valores de sensibilidade e especificidade que confere a esses modelos a capacidade de 90,51% em classificar as amostras corretamente as amostras da localidade B. E 78,3% de classificar corretamente as amostras da localidade C como sendo da respectiva localidade.

A respeito dos modelos PLS's dos atributos de qualidade, os valores obtidos de RMSEC e RMSEP são relativamente baixos quando comparado ao %CV de cada modelo. O Bias é outro parâmetro que apresentou um baixo valor. Porém, não se verifica correlação entre os valores preditos pelos modelos e as medidas de referência. Dentre todos, a apenas o PLS-SFI mostrou potencial para determinar o índice de fibras curtas com a utilização de um espectrômetro Raman Portátil. Para isso é preciso corrigir o procedimento da obtenção dos dados, pois há muita dispersão entre os espectros e os modelos construídos não foram capazes de corrigir essa problemática.

De modo geral, a pesquisa realizada pode contribuir significativamente para o setor produtivo de algodão, como o estágio de prova de teste da tecnologia estudada. Podendo levar a melhorias e futuramente a modelos mais robustos e precisos.

7 PERSPECTIVAS FUTURAS

- ✓ Desenvolver um artigo científico para publicação dos principais resultados obtidos, compartilhando com a comunidade científica discussões a respeito de análises envolvendo espectros Raman de fibras e algodão;
- ✓ Realizar novas medidas espectrais para construção de novos modelos de calibração e classificação, tendo em vista que os resultados obtidos são bem promissores para determinar atributos de qualidade e classificar fibras de algodão de diferentes classes;
- ✓ Realizar novas estratégias de análises, como por exemplo a fusão de dados espectrais NIR e espectrais Raman para determinar atributos de qualidade das fibras de algodão;
- ✓ Empregar outros métodos de calibração multivariada como o iPLS, máquina de vetores de suporte (SVM, do inglês *support vector machine*);
- ✓ Determinar os atributos de qualidade da fibra de algodão, como por exemplo: o índice de cera e grau de amarelamento.

REFERÊNCIAS

ABIDI, N; HEQUET, E; CABRALES, L. Changes in sugar composition and cellulose content during the secondary cell wall biogenesis in cotton fibers. **Cellulose**, v. 17, n. 1, p. 153-160, 2010.

_____, N. Chemical Properties of Cotton Fiber and Chemical Modification. In: **Cotton Fiber: Physics, Chemistry and Biology**. Springer, Cham. p. 95-115, 2018.

ABRAPA. **Associação Brasileira dos Prudutores de Algodão**. Relatório da Safra: Principais Indicadores da temporada de exportação de algodão 2020 /2021, Brasília DR, 2021. Disponível em: <https://www.abrapa.com.br/SiteAssets/Lists/Noticias/EditForm>. Acesso em: 01 de dezembro de 2021

ALVES, L. R. A et al. Cadeia agroindustrial e transmissão de preços do algodão ao consumidor brasileiro. **Revista de Economia e Sociologia Rural**, v. 59, n. 3, 2021.

BACHELIER, B; GOURLOT J. P. **A fibra de algodão: origem, estrutura, composição e caracterização**. In: Manual de qualidade da fibra da AMPA. Bélot Jean-Louis (ed.). Cuiabá: IMAmt-Ampa, pp. 28-57, 2018.

BALLABIO, D.; CONSONNI, V. Classification tools in Chemistry. Part 1: Linear models. PLS-DA. *Analytical Methods*, v.5, 2013.

_____, D; TODESCHINI, R. Multivariate Classification for Qualitative Analysis in: **Da-wen-Sun. Infraed Spectroscopy for food quality analysis an control**. 1ed. Elsevier inc. p. 97, 2009.

BARKER, M.; RAYENS, W. Partial least squares for discrimination. *Journal of Chemometrics*, v. 17, n. 3, p. 166-173, 2003.

BASSINELLO, P. Z. et al. Distribuição da sacarose-fosfato sintase e sacarose sintase em bananas durante o amadurecimento. **Food Science and Technology**, v. 19, p. 102-106, 1999.

BIANCHI, F., et al. Differentiation of aged fibers by Raman spectroscopy and multivariate data analysis. **Talanta**, v. 154, p. 467-473, 2016.

BRADOW, J. M.; DAVIDONIS, G. H. Quantization of fiber quality and the cotton productionprocessing interface: a physiologist's perspective. **The Journal of Cotton Science**, v.4, p.34-64, 2000.

BRERETON, Richard G. Introduction to multivariate calibration in analytical chemistryElectronic Supplementary Information available. **Analyst**, v. 125, n. 11, p. 2125-2154, 2000.

_____, R. G.; LLOYD, G. R. Partial least squares discriminant analysis: taking the magic way. *Journal of Chemometrics*, v. 28, n. 4, p. 213-225, 2014.

BRO, R; SMILDE, A. K. Principal component analysis. **Analytical methods**, v. 6, n. 9, p. 2812-2831, 2014.

BRUSHWOOD, D. E.; PERKINS, H. H. Determining the metal content of cotton. **Textile Chemist And Colorist**, v.26, p. 32-35, 1994.

_____, D. E. Survey of metals found in non-domestic raw cottons. In: **Proceedings of the Beltwide Cotton Conference**. 2000. p. 1546-1549.

CARVALHO, V. A. M., et al. **Pesquisa da influência do ph, temperatura e tempo no alvejamento com hipoclorito de sódio**. In: Simpósio Internacional de Engenharia Têxtil, 2000, Natal.

COCINERO, E. J. et al. The building blocks of cellulose: the intrinsic conformational structures of cellobiose, its epimer, lactose, and their singly hydrated complexes. **Journal of the American Chemical Society**, v. 131, n. 31, p. 11117-11123, 2009.

CONSTABLE, G. A.; BANGE, M.P. Producing and preserving fiber quality: from the seed to the bale. In: World Cotton Research Conference, 4, Abstracts. Lubbock, Texas, 2007.

COSTA, J. N., et al. Padrões universais para classificação do algodão. **Embrapa Algodão-Documentos (INFOTECA-E)**, 2006.

CURRIE, L. A.; Anal. Chim. **Acta**. 391, 105, 1999.

DELFINO, I., et al. Multivariate analysis of difference raman spectra of the irradiated nucleus and cytoplasm region of SH-SY5Y human neuroblastoma cells. **Sensors**, v. 19, n. 18, p. 3971, 2019.

DUARTE, A. S. **Aplicações de microscopia CARS multiplexada em nanotubos de carbono e tecido cerebral**. 106f. Tese (Doutorado em Física) – Universidade federal do Rio grande do Sul, Porto Alegre – RS, 2014.

DUCKWORTH, James. Mathematical data preprocessing. **Near-infrared spectroscopy in agriculture**, v. 44, p. 113-132, 2004.

ECHER, F. R.; ROSOLEM, C. A.; RAPHAEL, J. P. A. Desenvolvimento da planta e qualidade da fibra. **Manual de qualidade da fibra da AMPA**. Cuiabá: Instituto Mato-Grossense do Algodão–IMAmt, p. 206-237, 2018.

ENGEL, J., et al. Breaking with trends in pre-processing? **TrAC Trends in Analytical Chemistry**, v. 50, p. 96-106, 2013.

FABER, N. M.; RAJKÓ, R. How to avoid over-fitting in multivariate calibration – the conventional validation approach and an alternative. **Analytica Chimica Acta**, v. 595, n. 1/2, p. 98-106, 2007.

- FAKAYODE, S. O., et al. Molecular (Raman, NIR, and FTIR) spectroscopy and multivariate analysis in consumable products analysis1. **Applied Spectroscopy Reviews**, v. 55, n. 8, p. 647-723, 2020.
- FANG, D. D. General Description of Cotton. In: **Cotton Fiber: Physics, Chemistry and Biology**. Springer, Cham, p. 1-11, 2018.
- FERREIRA, M. M. C. **Quimiometria: conceitos, métodos e aplicações**. Editora da Unicamp, 2015.
- FILZMOSER, P., et al. Robust multivariate methods in chemometrics. **arXiv preprint arXiv:2006.01617**, 2020.
- FISHER, Charles H. History of natural fiber. **Journal of Macromolecular Science — Chemistry**, v. 15, n. 7, pág. 1345-1375, 1981.
- FONSECA, R. G; SANTANA, J. C. F. Resultados de ensaio HVI e suas Interpretações (ASTM D-4605). **Embrapa Algodão-Circular Técnica (INFOTECA-E)**, 2002.
- FRENCH, A. D.; KIM, H. J. Cotton fiber structure. In: **Cotton fiber: physics, chemistry and biology**. Springer, Cham, 2018. p. 13-39.
- GALVÃO, R. K. H., et al. A method for calibration and validation subset partitioning. **Talanta**, v. 67, n. 4, p. 736-740, 2005.
- GAUTAM, Rekha et al. Review of multidimensional data processing approaches for Raman and infrared spectroscopy. **EPJ Techniques and Instrumentation**, v. 2, n. 1, p. 1-38, 2015.
- GELADI, P.; ESBENSEN, K. The start and early history of chemometrics: selected interviews. Part 1. **Journal of Chemometrics**, v. 4, n. 5, p. 337-354, 1990.
- GHASEMI, J; NIAZI, A; LEARDI, R. Genetic-algorithm-based wavelength selection in multicomponent spectrophotometric determination by PLS: application on copper and zinc mixture. **Talanta**, v. 59, n. 2, p. 311-317, 2003.
- GONÇALVES, M. I. S., et al. An analytical method for determination of quality parameters in cotton plumes by digital image and chemometrics. **Computers and Electronics in Agriculture**, v. 123, p. 89-94, 2016.
- GORDON, S.; HSIEH, Y. L. **Cotton: Science and Technology**. Cambridge, Woodhead Publishing Limited, 2007.
- HAIGLER, C. H., et al. Cotton fiber: a powerful single-cell model for cell wall and cellulose research. **Frontiers in plant science**, v. 3, p. 104, 2012.
- HIBBERT, D. B. Vocabulary of concepts and terms in chemometrics (IUPAC Recommendations 2016). **Pure and Applied Chemistry**, v. 88, n. 4, p. 407-443, 2016.

HERRERO, A. M. Raman spectroscopy for monitoring protein structure in muscle food systems. **Critical reviews in food science and nutrition**, v. 48, n. 6, p. 512-523, 2008.

HOTELLING, H. Analysis of a complex of statistical variables into principal components. **Journal of Educational Psychology**, vol. 24, p. 417, 1993.

KAPPES, C.; ZANCANARO L.; FRANCISCO E.A.B. Nitrogen and potassium in narrow-row cotton. **Revista Brasileira de Ciência do Solo**, v.40, 2016.

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

KILINÇ, A. Ç; DURMUŞKAHYA, C; SEYDIBEYOĞLU, M. Ö. Natural fibers. In: **Fiber Technology for Fiber-Reinforced Composites**. Woodhead Publishing, 2017. p. 209-235.

KIM, Hee Jin. Cotton fiber biosynthesis. **Cotton fiber: physics, chemistry and biology**, p. 133-150, 2018.

_____, H. J et al. Comparative physical and chemical analyses of cotton fibers from two near isogenic upland lines differing in fiber wall thickness. **Cellulose**, v. 24, n. 6, p. 2385-2401, 2017.

KUHAR, N., et al. Challenges in application of Raman spectroscopy to biology and materials. **RSC advances**, v. 8, n. 46, p. 25888-25908, 2018.

KUMAR, N., et al. Chemometrics tools used in analytical chemistry: An overview. **Talanta**, v. 123, p. 186-199, 2014.

LANA V; RIBAS, P. V; CHANSELME, J. L. **Classificação do algodão**. In: Chanselme, J-L. Manual de beneficiamento. Cuiabá: AMPA; IMamt, p. 301-313, 2014.

LEE, C. M. et al. Probing crystal structure and mesoscale assembly of cellulose microfibrils in plant cell walls, tunicate tests, and bacterial films using vibrational sum frequency generation (SFG) spectroscopy. **Physical Chemistry Chemical Physics**, v. 16, n. 22, p. 10844-10853, 2014.

LEE, J. J.; WOODWARD, A. W.; CHEN, Z. J. Gene expression changes and early events in cotton fibre development. **Annals of botany**, v. 100, n. 7, p. 1391-1401, 2007.

LEE, Y; HAN, S. H; NAM, S. H. Soft independent modeling of class analogy (SIMCA) modeling of laser-induced plasma emission spectra of edible salts for accurate classification. **Applied spectroscopy**, v. 71, n. 9, p. 2199-2210, 2017.

LIMA, J.J. Classificação de algodão em pluma. In: **safra 2018 – Manual de Qualidade da Fibra AMPA**. Cuiabá, MT: IMamt – Instituto Matogrossense do Algodão, p. 58-115, 2018.

LIU, J., et al. Using terahertz spectroscopy to identify transgenic cottonseed oil according to physicochemical quality parameters. **Optik**, v. 142, p. 483–488, 2017

LIU, Y. KOKOT, S. SAMBI, T. J. Vibrational spectroscopic investigation of Australian cotton cellulose fibres Part 1. A Fourier transform Raman study. **Analyst**, v. 123, n. 4, p. 633-636, 1998.

LU, H. Introdução ao pré-tratamento enzimático do algodão. **Revista Química Têxtil**, São Paulo, n.89, p.06-10, 2007.

MANIÇOBA, R. M. **Manejo da irrigação em cultivares de algodão herbáceo no semiárido brasileiro**. 140f. Tese (Doutorado em Manejo de Solo e Água) – Universidade Federal Rural do Semiárido, Mossoró RN, 2019

MARCELO, C. A.; POZEBON, D; FERRÃO, M. F. Análise Multivariada do Perfil das Concentrações de Elementos Presentes na Erva-mate (*Ilex paraguariensis*) comercializada no MERCOSUL. **Revista Virtual de Química**, v. 7, n. 6, p. 2400-2421, 2015.

MARINI, F. **Chemometrics in food chemistry**. Newnes, 2013.

MARTENS, H.; NAES, T. Multivariate calibration. **Chichester: Wiley**. p. 419, 1989.

MARTINS, I. T. A. **Qualidade da fibra de diferentes cultivares brasileiras de algodão e sua relação com as condições meteorológicas**. 162f. Dissertação (Mestrado em Ciências) – Universidade de São Paulo, Piracicaba - SP, 2020.

MARUR, C. J; RUANO, O. Escala do algodão: um método para determinação de estádios de desenvolvimento do algodoeiro herbáceo. **Informações Agronômicas**, Piracicaba, v. 105, n. 1, p.3-4, mar. 2004.

METROHM. **Analísadores Raman portáteis Mira, Misa e SERS**. Disponível em: <https://www.metrohm.com/pt-br/produtos-geral/espectroscopia/espectrometros-portateis-mira/>. Acesso em: 25 dezembro de 2021.

MISHRA, P et al. Chemometric pre-processing can negatively affect the performance of near-infrared spectroscopy models for fruit quality prediction. **Talanta**, v. 229, p. 122303, 2021.

MORAIS, S. Cristiane. **Aplicação de Calibração Multivariada em dados de Espectroscopia UV-Visível para previsão da Acidez total em vinhos**. 2014. Trabalho de conclusão de curso – Tecnologia em Alimentos, Universidade Tecnológica Federal do Paraná (UTFPR). Campo Mourão, 2014.

NAITHANI, S.C.; RAO, N.R.; SINGH, Y.D. Physiological and biochemical changes associated with cotton - bre development. I. Growth kinetics and auxin content. **Physiologia Plantarum**, v.54, p.225-229, 1982.

OLIVEIRA, D. S. et al. Perfil metabólico dos extratos polares das folhas, galhos e cascas de *copaifera multijuga hayne* (copaíba mari-mari da amazônia) por esi-em. **Química Nova**, v. 43, n. 1, p. 72-77, 2020.

- OLIVIERI, Alejandro C. Mathematical Pre-processing. In: **Introduction to Multivariate Calibration**. Springer, Cham, p. 139-158, 2018
- PASSOS, M. G.; CANÉCHIO FILHO, V.; JOSÉ, A. **Principais culturas**. Instituto Campineiro de Ensino Agrícola, 1973. 403 p., 1973.
- PONTES, M. J. C., et al. The successive projections algorithm for spectral variable selection in classification problems. **Chemeometrics and inteligente Laboratory Systems**, 78, 11, 2005.
- RINNAN, A.; VAN DEN BERG, F.; ENGELSEN, S. B. Review of the most common pre-processing techniques for near-infrared spectra. **TrAC - Trends Analytical Chemistry**, v. 28, p. 1201–22, 2009
- RIOS, C. A. **Estudo de manchas de sangue: uma abordagem forense empregando espectroscopia raman, imagens digitais e ferramentas quimiométricas**. 145f. Dissertação (Mestrado em Química Analítica) Universidade Federal de Minas Gerais, Belo Horizonte, 2018.
- ROCHA, P. D. **Metodologia para distinção entre genótipos convencionais e transgênico de algodão utilizando espectroscopia NIR e imagens Hiperspectrais**. 2020. Dissertação (Mestrado em Química). Universidade Estadual da Paraíba- UEPB, Campina Grande-PB.
- RUAN, Y. L.; LLEWELLYN, D. J.; FURBANK, R. T. Suppression of sucrose synthase gene expression represses cotton fiber cell initiation, elongation, and seed development. **The Plant Cell**, v. 15, n. 4, p. 952-964, 2003.
- SALA, O. **Intensificação de espalhamento Raman por ressonância e por superfície: princípios e aplicações**. 53f. Tese (Doutorado em Química) - Instituto de Química, Universidade de São Paulo. São Paulo, 1986.
- SANTANA, F. B., et al. Experimento didático de quimiometria para classificação de óleos vegetais comestíveis por espectroscopia no infravermelho médio combinado com análise discriminante por mínimos quadrados parciais: um tutorial, parte V. **Química Nova**, v. 43, p. 371-381, 2020.
- SANTANA, J. C. F.; WANDERLEY, MJR; BELTRAO, N. M. Tecnologia da fibra e do fio do algodão, análises e interpretações dos resultados. **Embrapa Agropecuária Oeste-Capítulo em livro técnico (INFOTECA-E)**, 1998.
- SAYEED, M. A., et al. Characterizing the total within-sample variation in cotton fiber length using the High-Volume Instrument fibrogram. **Textile Research Journal**, v. 91, n. 1-2, p. 175-187, 2021.
- SCHNEIDER, M. et al. Métodos de Análise Exploratória Aplicadas ao Ensino de Espectroscopia na Região do Infravermelho. **Revista Virtual de Química**, v.10, n.2, p. 229-243, 2018.

SENA, M. M; ALMEIDA, M. R. Quimiometria Aplicada aos Dados Espectrais no Infravermelho Próximo. In: Casiane Salete Tibola. (Org.). **Espectroscopia no Infravermelho Próximo para Avaliar Indicadores de Qualidade Tecnológica e Contaminantes em Grãos**. 1aed. Passo Fundo: EMBRAPA Trigo, v 1., p. 32-52, 2018

SHIPP, D. W; SINJAB, F; NOTINGHER, I. Raman spectroscopy: techniques and applications in the life sciences. **Advances in Optics and Photonics**, v. 9, n. 2, p. 315-428, 2017.

SIDDIQUI, M. Q; WANG, H; MEMON, H. Cotton Fiber Testing. In: **Cotton Science and Processing Technology**. Springer, Singapore, p. 99-119, 2020.

SILVA, J. T., et al. Rapadura de Caldo de Cana Pura e com Adição de Açúcar Refinado: Análise Físico-Química e Classificação por Espectroscopia NIR e Quimiometria. **Revista Virtual de Química**, v. 11, n. 4, 2019.

SIMÕES, S. S. **Desenvolvimentos de métodos avaliados para determinação de captopril usando espectroscopia NIR e calibração multivariada**. 2008. 97 f. Tese (Doutorado em Química) – Departamento de Química, Universidade Federal da Paraíba, João Pessoa, 2008.

SOUZA, A. M., et al. Experimento didático de quimiometria para calibração multivariada na determinação de paracetamol em comprimidos comerciais utilizando espectroscopia no infravermelho próximo: um tutorial, parte II. **Química Nova**, v. 36, n. 7, p. 1057-1065, 2013.

SUBRAMANIAN, A; SILVA, L. B; COUTINHO, A. S. Aplicação de método e técnica multivariados para previsão de variáveis termo ambientais e perceptivas. **Production**, v. 17, n. 1, p. 52-70, 2007.

STAUT, L. A.; KURIHARA, C. H. Calagem, nutrição e adubação. **Embrapa Agropecuária Oeste-Capítulo em livro técnico (INFOTECA-E)**, 1998.

SZYMANSKA, E., et al. Chemometrics and qualitative analysis have a vibrant relationship. **TrAC – Trends in Analytical Chemistry**, v. 69, p. 34-51, June 2015.

TEMIZ, M., et al. Effect of foliar potassium and micronutrient additions on yield and fiber quality of cotton (*Gossypium hirsutum* L.). **Journal of Food, Agriculture and Environment**, v. 7, n. 1, p. 118-122, 2009.

TEÓFILO, R. F. **Métodos quimiométricos em estudos eletroquímicos de fenóis sobre filmes de diamante dopado com boro**. 329p. (Doutorado em Química). Universidade Estadual de Campinas, 2007.

TIMPA, J. D; RAMEY J. R, H. H. Relationship between cotton fiber strength and cellulose molecular weight distribution: HVI calibration standards. **Textile research journal**, v. 64, n. 10, p. 557-562, 1994.

USDA. Unidted States Departament of Agriculture. **Cotton Fiber Testing and Standardization**, 2014.

_____. Unidited States Departamentt of Agriculture. **Foreign Agricultural Service**, 2021.

USTER. USTER HVI 1000 – Application Handbook – Fiber bundle testing of cotton. 2002.

VALDERRAMA, P; BRAGA, J. W; POPPI, R. J. Estado da arte de figuras de mérito em calibração multivariada. **Química Nova**, v. 32, n. 5, p. 1278-1287, 2009.

VALDERRAMA, P.; BRAGA, J. W. B.; POPPI, R. J. Variable selection, outlier detection, and figures of merit estimation in a partial least-squares regression multivariate calibration model. A case study for the determination of quality parameters in the alcohol industry by near-infrared spectroscopy. **Journal of Agricultural and Food Chemistry**, v. 55, n. 21, p. 8331-8338, 2007.

VALENTINI, S.R.; SOMMER W.A.; MATIOLI, G. **Validação de métodos analíticos**. Arquivos do Mudi, v.11, n. 2, p. 26-31, 2007.

VIK, M., et al. Non-contact method for measurement of colour variation in a cotton sample. **Fibres & Textiles in Eastern Europe**, 2017.

VITERO, F., et al. Electrobleaching of cotton fabrics in sodium chloride solution. **Vlakna a Textil**, v. 24, n. 2, p. 32–36, 2017.

XIE, L., et al. Discrimination of transgenic tomatoes based on visible/near-infrared spectra. **Analytica Chimica Acta**, v. 584, p. 379–384, 2007.

ZHU, S., et al. GhAlaRP, a cotton alanine rich protein gene, involves in fiber elongation process. **The Crop Journal**, v. 9, n. 2, p. 313-324, 2021.

ZHANG, M., et al. Spatiotemporal manipulation of auxin biosynthesis in cotton ovule epidermal cells enhances fiber yield and quality. **Nature biotechnology**, v. 29, n. 5, p. 453-458, 2011.

ZHANG, Z., et al. Suppressing a putative sterol carrier gene reduces plasmodesmal permeability and activates sucrose transporter genes during cotton fiber elongation. **The Plant Cell**, v. 29, n. 8, p. 2027-2046, 2017.

WAKELYN, Phillip J. **Cotton fiber chemistry and technology**. CRC Press, 2006.

WANG, H; FAROOQ, A; MEMON, H. Influence of cotton fiber properties on the microstructural characteristics of mercerized fibers by regression analysis. **Wood Fiber Sci**, v. 52, n. 1, p. 13-27, 2020.

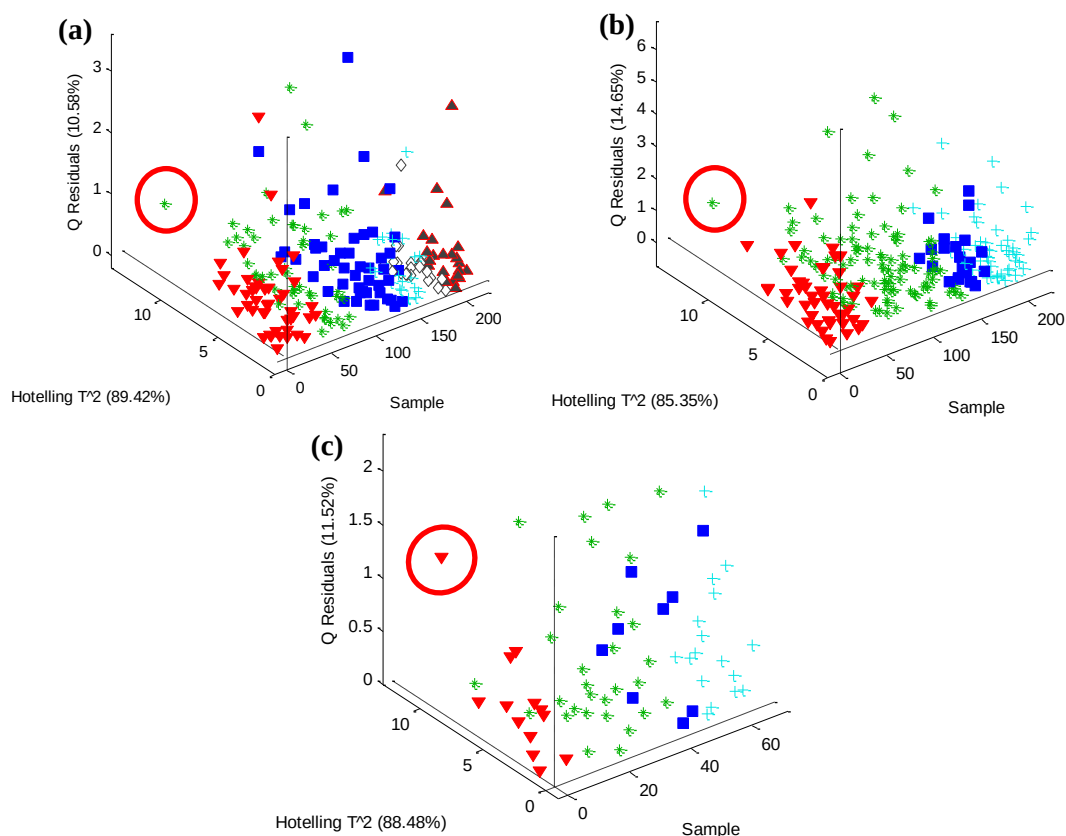
WARWICKER, J. O. et al. A review of the literature on the effect of caustic soda and other swelling agents on the fine structure of cotton. **A review of the literature on the effect of caustic soda and other swelling agents on the fine structure of cotton.**, 1966.

WOLD, S; SJÖSTRÖM, M; ERIKSSON, L. PLS-regression: a basic tool of chemometrics. **Chemometrics and intelligent laboratory systems**, v. 58, n. 2, p. 109-130, 2001.

YANG, D; YING, Y. Applications of Raman spectroscopy in agricultural products and food analysis: A review. **Applied Spectroscopy Reviews**, v. 46, n. 7, p. 539-560, 2011.

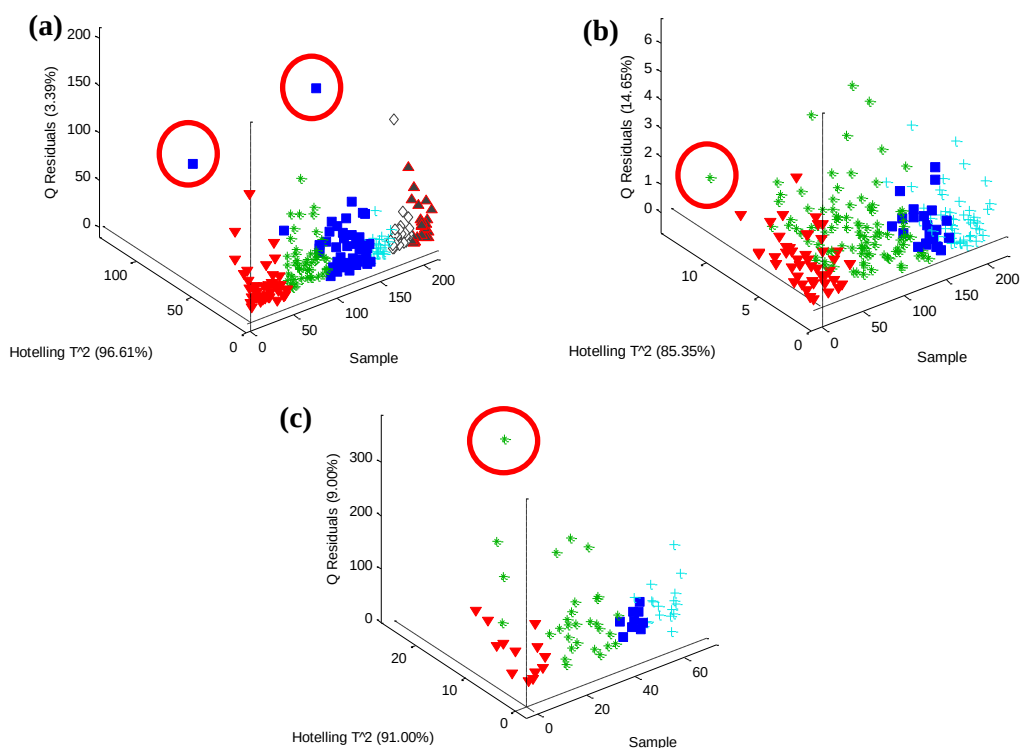
YANG, J., HU, W., ZHAO, W., MENG, Y., CHEN, B., WANG, Y., ZHOU, Z. Soil potassium deficiency reduces cotton fiber strength by accelerating and shortening fiber development. **Scientific Reports**, v.6, 28856, 2016.

APÊNDICE A – Gráficos de T^2 de *hotelling* e Q residual do modelo PLS-DA-At



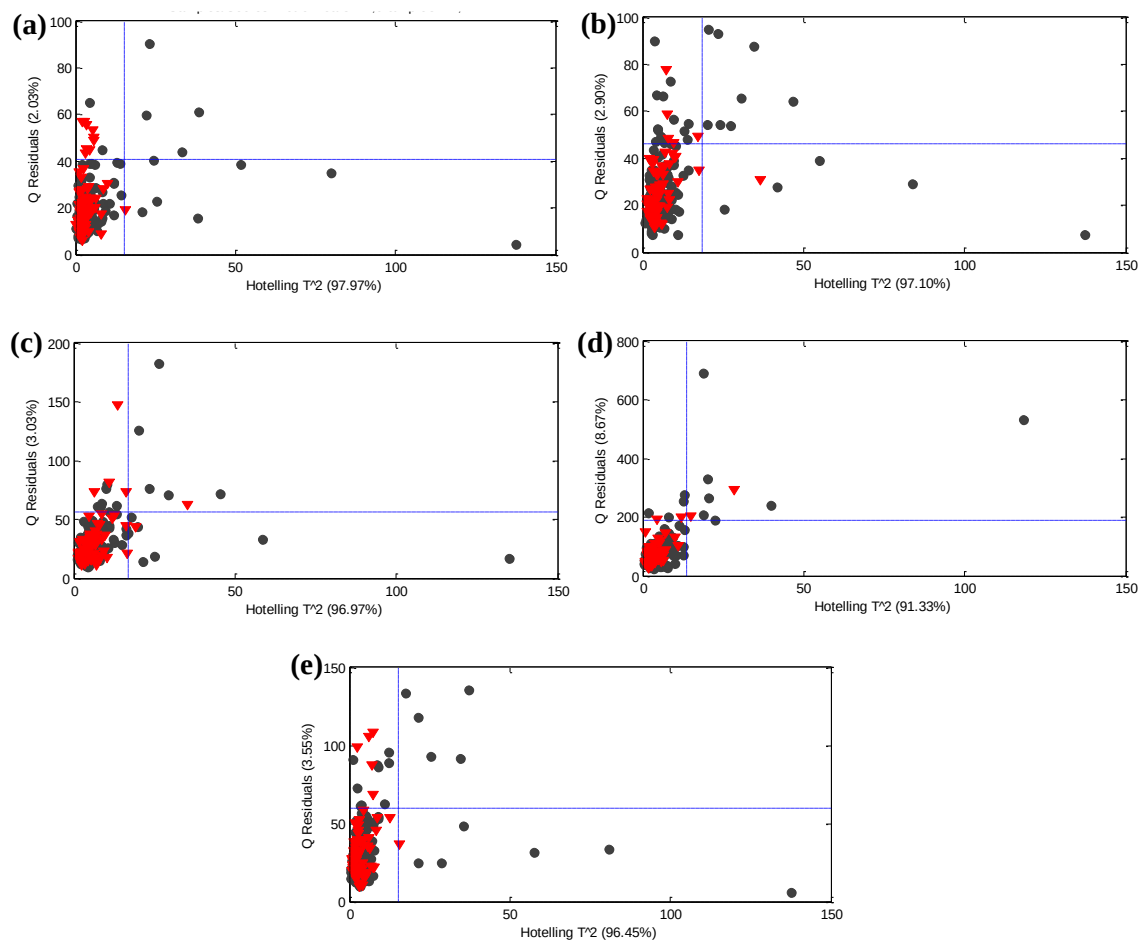
Fonte: Arquivo Próprio (2022). Em **(a)** Conjunto de treinamento: Vermelhas (A); Verdes (B); Azuis (C). Conjunto de teste: Ciano (A); Brancas (B); Vermelho-preto (C). Em **(b)** Conjunto de Treinamento: Vermelhas (CONV); Verdes (TRAN). Conjunto de Teste: Azuis (CONV); Ciano (TRAN). Em **(c)** Conjunto de Treinamento: Vermelhas (PRE); Verdes (TARD). Conjunto de Teste: Azuis (PRE); Ciano (TARD). Amostras com comportamentos Anômalos (círculo vermelho)

APÊNDICE B - Gráficos de T^2 de *hotelling* e Q residual do modelo PLS-DA-Raman



Fonte: Arquivo Próprio (2022). Em (a) Conjunto de treinamento: Vermelhas (A); Verdes (B); Azuis (C). Conjunto de teste: Ciano (A); Brancas (B); Vermelho-preto (C). Em (b) Conjunto de Treinamento: Vermelhas (CONV); Verdes (TRAN). Conjunto de Teste: Azuis (CONV); Ciano (TRAN). Em (c) Conjunto de Treinamento: Vermelhas (PRE); Verdes (TARD). Conjunto de Teste: Azuis (PRE); Ciano (TARD). Amostras com comportamentos Anômalos (círculo vermelho)

APÊNDICE C – Gráficos de T^2 de *hotelling* versus Q residual dos modelos PLS para determinação dos atributos de qualidade das fibras de algodão



Fonte: Arquivo Próprio (2022). PLS-UHM (a); PLS-SFI (b); PLS-MIC (c); PLS-TR (d); PLS-Rd (e). Amostras da Calibração (▲); Amostras de Predição (●).