



**UNIVERSIDADE ESTADUAL DA PARAÍBA
CAMPUS I
CENTRO DE CIÊNCIAS E TECNOLOGIA
PROGRAMA DE PÓS GRADUAÇÃO EM QUÍMICA**



PRISCILLA DANTAS ROCHA

Metodologia para distinção entre genótipos convencionais e transgênicos de algodão utilizando espectroscopia NIR e Imagens Hiperespectrais

**CAMPINA GRANDE – PB
2020**

PRISCILLA DANTAS ROCHA

Metodologia para distinção entre genótipos convencionais e transgênicos de algodão utilizando espectroscopia NIR e Imagens Hiperespectrais

Trabalho de Dissertação apresentada ao Programa de Pós-Graduação em Química da Universidade Estadual da Paraíba, como requisito à obtenção do título de mestre em Química.

Área de concentração: Química Analítica.

Orientador: Profa. Dra. Simone da Silva Simões.

Coorientador: Dr. Everaldo Paulo de Medeiros.

**CAMPINA GRANDE – PB
2020**

É expressamente proibido a comercialização deste documento, tanto na forma impressa como eletrônica. Sua reprodução total ou parcial é permitida exclusivamente para fins acadêmicos e científicos, desde que na reprodução figure a identificação do autor, título, instituição e ano do trabalho.

R672m Rocha, Priscilla Dantas.

Metodologia para distinção entre genótipos convencionais e transgênicos de algodão utilizando espectroscopia NIR e Imagens Hiperespectrais [manuscrito] / Priscilla Dantas Rocha. - 2020.

79 p. : il. colorido.

Digitado.

Dissertação (Mestrado em Química - Mestrado) - Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2020.

"Orientação : Profa. Dra. Simone da Silva Simões , Departamento de Química - CCT."

"Coorientação: Prof. Dr. Everaldo Paulo de Medeiros , Embrapa Algodão"

1. Gossypium L.. 2. Metodologias alternativas. 3. Quimiometria. 4. Organismos Geneticamente Modificados. I. Título

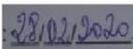
21. ed. CDD 633.51

PRISCILLA DANTAS ROCHA

**Metodologia para distinção entre genótipos convencionais e transgênicos
de algodão utilizando espectroscopia NIR e Imagens Hiperespectrais**

Trabalho de Dissertação
apresentada ao Programa de Pós-
Graduação em Química da
Universidade Estadual da Paraíba,
como requisito à obtenção do título
de mestre em Química.

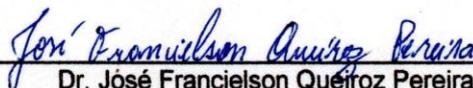
Área de concentração: Química
Analítica.

Aprovada em: 

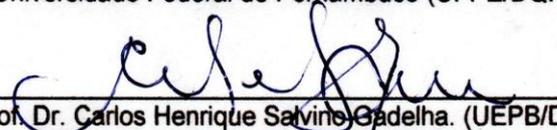
BANCA EXAMINADORA



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



Dr. José Francielson Queiroz Pereira
Universidade Federal de Pernambuco (UFPE/DQF)



Prof. Dr. Carlos Henrique Salvino Gadelha. (UEPB/DB)
Universidade Estadual da Paraíba (UEPB/DQ)



Prof. Dr. José Germano Veras Neto
Universidade Estadual da Paraíba (UEPB/DQ)

À minha família e amigos, pela dedicação,
companheirismo e amizade, DEDICO.

AGRADECIMENTOS

Agradeço a Deus em primeiro lugar pelas oportunidades, pelas bênçãos concedidas e por sempre abençoar o meu caminho, colocando pessoas maravilhosas para me ajudar nessa caminhada.

Agradeço aos meus pais Isauro Dantas Junior e Joseane Rocha Dantas pelo apoio, pelas palavras, pelos ensinamentos, por sempre me incentivarem a trilhar uma caminhada na fé, e por me incentivarem a ser uma pessoa independente e correr atrás dos meus sonhos, sem passar por cima de ninguém, tendo os pés no chão e respeitando o próximo.

Obrigada ao meu irmão Tássio, minha cunhada Iolanda por me presentear com a minha sobrinha Maria Cecilia, com ela eu conheci o amor mais puro que uma pessoa pode conhecer, e pela chegada em alguns meses do meu sobrinho Théo que eu não conheço ainda mais já amo tanto.

Um agradecimento mais que especial ao meu primo Luan Dantas que está seguindo comigo a tanto tempo, compartilhando o sonho de conquistar algo, e de construir um futuro melhor através dos estudos.

E a minha prima Bruna Rocha que é mais que prima é minha irmã de alma que sempre está ao meu lado me ajudando em todos os momentos difíceis e comemorando comigo nos momentos de conquistas. Espero que o seu caminho seja de sucesso que você merece o mundo.

Obrigada a Ingrid e a Juliana pela parceria para a coleta desses dados, onde por muitas vezes elas tiveram que coletar esses dados no meu lugar, e obrigada pelo profissionalismo.

Obrigada Dra. Carolina Santos Silva por ser tão gentil, delicada, tão generosa, e por todo suporte no tratamento das Imagens Hiperespectrais.

Obrigada a Dr. Everaldo que sempre está me apoiando na Embrapa Algodão e me guiando para fazer sempre um ótimo trabalho.

Agradeço também a todos os docentes que compõem a PPGQ- UEPB vocês foram nossos mestres, nos ensinando não só assuntos inerentes das disciplinas, mas também ensinando sobre a vida. E a todos os discentes que caminharam comigo nesta fase de tanto aprendizado.

Agradeço a todos os meus professores da PPGQ que se doam á esse programa que mesmo sendo tão novo é grandioso. Professores esses que nos inspiram e que nos fazem continuar na procura de um melhoramento profissional mais também humano.

Agradeço a UEPB por ser a instituição que me acolheu nos meus dois cursos de graduação me dando todo o suporte e base profissional como também a oportunidade de fazer mestrado na mesma.

Um agradecimento mais do que justo e necessário a minha orientadora Dra. Simone Simões que a princípio eu escutava muitos dos seus orientandos à chamando de Mãe, mais depois ela passou a ser mais do que isso pra mim, ela passou a ser um anjo que Deus enviou pra minha vida, um ser humano magnifico, uma profissional excelente e uma incentivadora pra vida.

Para finalizar vou citar um trecho de Emmanoel Rocha Carvalho “ Qualquer que seja a área de conhecimento humano, por mais simples ou valiosa que seja a obra produzida, não se deve nunca atribuir a um homem só a gratidão do feito. Certamente, outras pessoas, naturais ou jurídicas, do

passado ou do presente, contribuíram para que isso acontecesse. Mais que um simples obrigado, essas pessoas, coadjuvantes, merecem receber um reconhecimento profícuo e uma gratidão imorredoura”

RESUMO

Os organismos geneticamente modificados (OGM's) são quaisquer organismo em que o material genético não tenha sido alterado de uma maneira natural por acasalamento ou/e recombinação natural. As modificações genéticas são realizadas para garantir, por exemplo, que as plantas sejam tolerantes a herbicidas ou resistentes a fungos e insetos, permitem variedades mais robustas, de maior qualidade e rendimento. Como os OGM's se consolidaram como a tecnologia mais rápida adotada na história recente da agricultura moderna, algumas pessoas, empresas, órgãos e governos se mantiveram céticos quanto aos impactos que esses organismos podem causar aos seres humanos e ao meio ambiente. Diante disso algumas leis têm sido elaboradas em relação ao uso, identificação, quantificação e diferenciação de cultivares convencionais dos geneticamente modificados. Atualmente os métodos convencionais utilizados para esse fim são laboriosos, caros, demorados e não preservam as amostras. Diante do exposto o objetivo deste trabalho é desenvolver uma metodologia alternativa às convencionais para a distinção rápida e precisa entre cultivares convencionais e transgênicos de algodão utilizando a espectroscopia NIR (NIRS) e imagens hiperespectrais (HSI). A metodologia adotada envolve a aquisição das amostras foram 2390 sementes analisadas na espectroscopia NIR e 70 sementes analisadas por imagens espectrais. Para a construção do modelo PLS-DA baseado em NIRS, o algoritmo SP-Xy foi utilizado para selecionar amostras dos conjuntos de treinamento e de teste. Em seguida os espectros foram pré-processados para remoção de características espectrais não relacionadas ao sinal das amostras. O modelo PLS-DA-NIRS foi então validado e aplicado às amostras do conjunto de teste. Já para a construção do modelo PLS-DA-HSINIR, foram utilizadas no conjunto de treinamento as imagens de 20 sementes convencionais e 20 transgênicas, obtidas separadamente, e para o conjunto de teste foi utilizada a imagem de 30 sementes dispostas de forma aleatória no suporte de aquisição de imagem. O tratamento das imagens foi realizado e o modelo PLS-DA-HSINIR construído foi projetado nas amostras de treinamento. As técnicas utilizadas são não destrutivas, não utilizam reagentes químicos, rápidas e representam metodologias alternativas que podem auxiliar a classificação/identificação de sementes de algodão transgênico. A metodologia PLS-DA-NIRS apresentou uma excelente capacidade preditiva, obtendo uma taxa de classificação correta de 97,95% na etapa de predição. Já os modelos PLS-DA-HSINIR classificou de forma inequívoca 70% das sementes analisadas. 7 das amostras tiveram classificação inconclusiva (23.3%). 2 amostras convencionais foram classificadas de forma incorreta na classe das transgênicas representando um falso positivo. O modelo construído mostra a qualidade de classificar corretamente 100% das amostras transgênicas em sua respectiva classe, não sendo observado a presença de falsos negativos. Pode-se concluir que as metodologias desenvolvidas foram capazes de fazer a distinção das amostras transgênicas das convencionais, sendo então técnicas eficazes para este fim.

Palavras-chave: *Gossypium* L. Metodologias Alternativas. Quimiometria.

ABSTRACT

Genetically modified organisms (GMOs) are any organism in which the genetic material has not been altered in a natural way by mating or / and natural recombination. Genetic modifications are carried out to ensure, for example, that plants are tolerant to herbicides or resistant to fungi and insects, allowing for more robust varieties, of higher quality and yield. As GMOs have established themselves as the fastest technology adopted in the recent history of modern agriculture, some people, companies, agencies and governments have remained skeptical about the impacts that these organisms can cause on humans and the environment. In view of this, some laws have been developed in relation to the use, identification, quantification and differentiation of conventional cultivars from genetically modified ones. Currently, the conventional methods used for this purpose are laborious, expensive, time consuming and do not preserve the samples. Given the above, the objective of this work is to develop an alternative methodology to conventional ones for the quick and precise distinction between conventional cultivars and cotton GMOs using NIR spectroscopy (NIRS) and hyperspectral images (HSI). The adopted methodology involves the acquisition of the samples: 2390 seeds were analyzed in the NIR spectroscopy and 70 seeds analyzed by spectral images. For the construction of the PLS-DA model based on NIRS, the SP-Xy algorithm was used to select samples from the training and test sets. Then the spectra were pre-processed to remove spectral characteristics unrelated to the sample signal. The PLS-DA-NIRS model was then validated and applied to the samples in the test set. For the construction of the PLS-DA-HSINIR model, the images of 20 conventional and 20 transgenic seeds, obtained separately, were used in the training set, and for the test set, the image of 30 seeds randomly placed on the support was used. image acquisition. The image treatment was performed and the PLS-DA-HSINIR model built was designed on the training samples. The techniques used are non-destructive, do not use chemical reagents, fast and represent alternative methodologies that can help the classification / identification of transgenic cotton seeds. The PLS-DA-NIRS methodology presented an excellent predictive capacity, obtaining a correct classification rate of 97.95% in the prediction stage. The PLS-DA-HSINIR models uniquely classified 70% of the analyzed seeds. 7 of the samples had an inconclusive classification (23.3%). 2 conventional samples were classified incorrectly in the transgenic class representing a false positive. The constructed model shows the quality of correctly classifying 100% of the transgenic samples in their respective class, without the presence of false negatives. It can be concluded that the methodologies developed were able to distinguish transgenic samples from conventional ones, being therefore effective techniques for this purpose.

Keywords: Gossypium L. Alternative Methodologies. Chemometrics.

LISTA DE FIGURAS

Figura 1: Esquema do modo de ação da modificação genética em MON 88913.....	22
Figura 2: Diferentes modos de vibração em uma molécula triatômica heteronuclear	28
Figura 3: (a) Diagrama de energia potencial para os osciladores harmônicos e (b) anarmônico.....	29
Figura 4: Matriz de dados de imagens (a) em escalas de cinza, (b) em RGB e (c) Hiperspectrais.	34
Figura 5: Desdobramento das matrizes 3D de imagens (a) RGB e (b) hiperspectrais	35
Figura 6: Desdobramento da matriz tridimensional de dados em uma matriz bidimensional e sua decomposição em perfis de concentração relativa e espectros puros.	36
Figura 7: Análise de componentes principais (PCA) – primeira e segunda PCs no espaço tridimensional das variáveis.	38
Figura 8: Processo de Modelagem do SIMCA	39
Figura 9: Esquema da construção de um modelo PLS-DA.	42
Figura 10: Sementes de BRS 368 RF e BRS Aroeira.	43
Figura 11: Espectrofotômetro XDS near-infrared Rapid Content TM Analyser da FOSS Analytical	44
Figura 12: Espectrômetro sisuChema SWIR - Specim®, responsável pela obtenção das imagens hiperspectrais.....	45
Figura 13: Espectros brutos das sementes convencionais e transgênicas	48
Figura 14: PCA das amostras convencionais e transgênicas.....	49
Figura 15: Espectros com o pré-processamento de Suavização com polinômio Savitzky-Golay de 2ª ordem e janela de 15 pontos.	49
Figura 16: Q residual versus amostras de treinamento.....	51
Figura 17: T ² de Hotelling versus amostras de treinamento.....	51
Figura 18: Curva ROC e respostas estimadas da etapa de treinamento	52
Figura 19: Loadings nas Variáveis Latentes 1 (a) e 7 (b) versus Variáveis.....	52
Figura 20: Escores para as amostras de treinamento.....	53
Figura 21: Curva ROC e respostas estimadas na etapa de teste	54
Figura 22: Escores das amostras de teste	54
Figura 23: Imagens originais convencionais (a) e transgênicas (b) e imagens com redução de dimensão convencionais (c) e transgênicas (d)	55
Figura 24: Scores das primeiras 5 PCs (a) e histograma de frequência dos pixels (b)	56
Figura 25: Escores da PC1 e Histograma dos pixels	57
Figura 26: Máscaras morfológicas com 4 erosões.	57
Figura 27: PCA das amostras com o pré-processamento 1ª Derivada de Savitzky- Golay com polinômio de 2º grau e janela de 7 pontos	58
Figura 28: Espectros brutos (a) e pré-processados (b,c,d).	59
Figura 29: PCA das amostras transgênicas e convencionais da calibração.	61
Figura 30: Imagem média inicial (a) e borda da imagem média reduzida (b).....	62
Figura 31: Scores das primeiras 5 PCs e histograma de frequência dos pixels.....	63
Figura 32: Scores da PC1 e Histograma dos pixels	63
Figura 33: Número de Variáveis Latentes.	64
Figura 34: Curvas ROC e respostas estimadas da Calibração	64
Figura 35: Resultado da Imagem de Previsão	65

LISTA DE TABELA

Tabela 1: Classificação das amostras de teste	66
Tabela 2: Dados da sensibilidade e especificidade	66

LISTA DE ABREVIATURA

CTNBio - Comissão Técnica Nacional de Biossegurança
DNA - Ácido Desoxirribonucleico (Deoxyribonucleic Acid)
ELISA - Ensaio de Imunoabsorção Enzimática (Enzyme-Linked Immunosorbent Assay)
HSI-NIR - Imagem Hiperespectral no Infravermelho Próximo
ICAC - Comitê Consultivo Internacional para o Algodão (International Cotton Advisory Committee)
IFL - Imunoensaio de Fluxo Lateral
LV – Variáveis Latentes (<i>Latent Variables</i>)
MIR - Infravermelho Médio (Middle Infrared)
NIR – Infravermelho Próximo (Near Infrared)
OGM's - Organismos Geneticamente Modificados
PAT - Tecnologia Analítica de Processo
PC – Componentes Principais (Principal Component)
PCA – Análise de Componentes Principais (Principal Component Analysis)
PCR - Reação em Cadeia da Polimerase (Polymerase Chain Reaction)
PLS – Mínimos Quadrados Parciais (Partial Least Squares)
PLS-DA - Análise Discriminante por Mínimos Quadrados Parciais (Partial Least Squares – Discriminant Analysis)
RGB - Vermelho, Verde e Azul (Red, Green and Blue)
RNC - Registro Nacional de Cultivares
SIMCA - Modelagem Independente Flexível por Analogia de Classes (Soft Independent Modeling of Class Analogy)
SNV - Variação Normal Padrão (Standard Normal Variate)
WHO – Organização Mundial da Saúde (World Health Organization)

Sumário

1. INTRODUÇÃO	12
2. OBJETIVOS	14
2.1 Objetivo geral	14
2.2 Objetivos específicos	14
3. REVISÃO DA LITERATURA	15
3.1 Organismos Geneticamente Modificados	15
3.2 Uso de Transgenia nos Algodoeiros	17
3.3 Processo para Obtenção de uma Planta Transgênica	19
3.4 Vantagens e Desvantagens dos OGMs	23
3.5 Metodologias convencionais para confirmação dos OGMs	24
3.6 Espectroscopia na Região do Infravermelho	27
3.6.1 Utilização da Espectroscopia NIR nos transgênicos	31
3.7 Imagem hiperespectral no infravermelho próximo	32
3.8 Reconhecimento de Padrões	37
3.8.1 Análise de Componentes Principais (PCA)	37
3.8.2 SIMCA	38
3.8.3 PLS-DA	41
4 METODOLOGIA	43
4.1 Local do experimento	43
4.2 Amostras	43
4.3 Instrumentação	45
4.3.1 Espectroscopia NIR	45
4.3.2 Imagens Hiperespectrais no NIR- HSINIR	46
4.3.3 Softwares	47
4.4 Construção dos modelos de reconhecimento de padrões	46
4.4.1 Modelos PLS-DA baseados em espectros NIR	47
4.4.2 Modelos PLS-DA baseados em Imagens Hiperespectrais	47
5 RESULTADOS E DISCUSSÕES	48
5.1 Modelos baseados na espectroscopia NIR	48
5.2 Modelos PLS-DA baseados nas Imagens Hiperespectrais	56
6 CONCLUSÕES	68
REFERÊNCIAS	71

1. INTRODUÇÃO

De acordo com a Organização Mundial de Saúde (WHO, 2014), os Organismos Geneticamente Modificados (OGM's) podem ser definidos como “organismos nos quais o material genético (DNA) foi alterado de uma maneira que não ocorre naturalmente”. Podendo ser plantas, animais ou microorganismos, cujo material genético sofreu alterações por meio da engenharia genética, através da introdução de ácido desoxirribonucleico (DNA) exógeno (TERADA, 2002).

Os OGM's se consolidaram como a tecnologia mais rápida adotada na história recente da agricultura moderna (ISAAA, 2018), podem ser feitas algumas modificações para garantir, por exemplo, que as plantas sejam tolerantes a herbicidas ou resistentes a fungos e insetos, características que permitem variedades mais robustas, de maior qualidade e rendimento (KLUMPER, 2014).

Uma problemática envolvendo os algodoeiros são os danos causados por insetos e pragas, deste modo as mudanças feitas no algodão é de suma importância e essa melhoria está sendo feita utilizando principalmente tecnologias transgênicas (BAOHANG, 2019), onde pode-se obter uma nova variedade com maior teor de proteínas e óleos das sementes (WILKINS,2000), recuperar variedades mais férteis e resistentes a insetos (BOLEK et al, 2005).

Dados de 2018 mostraram que a área global de cultivos desenvolvidos por meio da biotecnologia cresceu 113 vezes quando se compara a área plantada de 1996 até 2018. No Brasil uma cultura que teve grande aumento de área para plantio foi o algodão, devido a alta demanda do mercado interno e externo, aumento da demanda doméstica e global de proteína para alimentação humana e animal (ISAAA, 2018). Esses e outros fatores só comprovam como essa cultura tem um papel importante na economia mundial, sendo produzida desde pequenos a grandes agricultores, com valor de produção de aproximadamente US\$40 bilhões (ABAPA, 2019).

Como toda tecnologia nova, os OGM trazem um pouco de incertezas e geram opiniões diversas, principalmente pela falta de informações quanto aos danos que pode vir a gerar na natureza e nos seres humanos. Por esse motivo, governos, órgãos e cientistas tem lutado para criar medidas para detectar os OGM's nos produtos e gerar uma maior confiança para o consumidor (BOCCIA, 2015; VARGAS et al. 2018).

Os métodos convencionais utilizados para detectar e identificar a presença de OGM's nos alimentos incluem métodos baseados em proteínas e em DNA. Esses métodos muitas vezes são laboriosos, caros, demorados, técnicas destrutivas que não preservam as amostras e estão longe dos grandes centros de produção se tornam inviáveis (HAO, et al., 2019).

Uma metodologia que pode ser uma alternativa útil às técnicas convencionais é a espectroscopia no Infravermelho Próximo (NIR), que traz como nenhuma ou mínima preparação da amostra, tempo e custo de análise reduzido, caracterização simultânea de múltiplos componentes influenciados pelo genótipo e viabilidade da análise online (HAI-FENG et al, 2012).

Diante desses inconvenientes das técnicas convencionais uma abordagem mais prática pode consistir no uso de técnicas de imagens, como a imagem hiperespectral NIR (HSI-NIR) que combina as vantagens da imagem digital e espectroscopia para adquirir informações espaciais e espectrais, respectivamente. A HSI-NIR tem sido aplicado com sucesso em vários estudos envolvendo sementes, como classificação de tipos de sementes, identificação da origem geográfica (GAO et al., 2013), caracterização dos grãos e sementes individuais (RODRÍGUES-PULIDO et al., 2013), entre outros.

A técnica de imagens hiperespectrais no infravermelho próximo (HSI-NIR) tem o potencial de ser usada como uma abordagem não destrutiva para a análise de novas matrizes microbiológicas de interesse da agricultura. A mesma é também uma ferramenta analítica com ampla capacidade para a verificação e quantificação dos compostos de interesse em sistemas biológicos (AIRES et al, 2018).

Portanto, o presente trabalho visa desenvolver uma metodologia simples e precisa capaz de distinguir os genótipos convencionais e transgênicos de algodão utilizando espectroscopia NIR e de imagens hiperespectrais (HSI), aliada a métodos quimiométricos de reconhecimento de padrões como PLS-DA (Análise Discriminante de Mínimos Quadrados Parciais).

2. OBJETIVOS

2.1 Objetivo geral

Desenvolver uma metodologia baseada na espectroscopia NIR e Imagens Hiperespectrias e métodos quimiométricos de reconhecimento de padrões, capazes de distinguir entre genótipos convencionais e transgênicos de algodão.

2.2 Objetivos específicos

- Registrar os espectros e as imagens dos genótipos convencionais e transgênicos de algodões utilizando o NIR e o HSI-NIR;
- Avaliar qual o melhor pré-tratamento espectral;
- Avaliar a capacidade das técnicas espectroscopia no infravermelho próximo e imagem hiperespectral aliadas a técnica de reconhecimento de padrões supervisionada (PLS-DA) quanto a distinção entre as sementes convencionais e transgênicas;
- Avaliar a capacidade preditiva dos modelos PLS-DA construídos.

3. REVISÃO DA LITERATURA

3.1 Organismos Geneticamente Modificados

Desde os tempos remotos, o homem vem exercendo a prática do melhoramento vegetal, a partir da qual buscava selecionar os tipos de plantas que melhor se adaptavam às suas necessidades. Esta foi uma forma de domesticação das plantas e uma forma revolucionária para a obtenção dos alimentos. Segundo Borém e Milach (2010), “é difícil concluir se foi a agricultura que incentivou a prática do melhoramento de plantas pelos primeiros agricultores ou vice-versa”.

A genética se firmou como uma nova ciência a partir dos experimentos de Mendel, em 1865. Por volta de 1953 ocorreu o descobrimento da dupla hélice do DNA por James Watson e Francis Crick (CARRER, 2010), que contribuíram para os avanços da biotecnologia. O desenvolvimento mais acelerado neste campo de conhecimento só se deu a partir de 1973 por Stanley Cohen e Herbert Boyer, com o desenvolvimento da engenharia genética, cuja informação permitia a alteração direta do material genético do DNA/RNA (GIEHL, 2006; VIEIRA, 2006).

O termo biotecnologia em si foi originalmente utilizado pelo húngaro Karl Ereky, em 1919, para se referir a todas as linhas de trabalho cujos produtos eram produzidos a partir de matéria bruta com o auxílio de organismos vivos (STEINBERG, 2002) e dessa forma a biotecnologia engloba todas as tecnologias que lidam com organismos vivos (EuropaBio 2017). Nos dias de hoje muitos pesquisadores utilizam a biotecnologia para atender a determinadas demandas, onde o conhecimento desta área da ciência permite a aquisição de plantas geneticamente modificadas de maneira mais controlada, podendo por exemplo certificar que ao transferir um gene para o genoma de outra espécie, o gene vai determinar nessa espécie a mesma característica que determinava na espécie doadora (BORÉM, 2005).

Dessa forma a biotecnologia é considerada uma tecnologia essencial, podendo ter uma gama de aplicações como na agricultura, medicina, processamento de alimentos, proteção ambiental, mineração e até na nanoeletrônica (MAHGOUB, NOLLET, 2018).

A variedade de aplicações vai desde usos tradicionais, como a produção de pão, álcool e queijo, até aplicações mais modernas, como o uso na produção de transgênicos. Nesse sentido, a produção de transgênico cai sob a égide da biotecnologia, sendo uma das aplicações modernas da biotecnologia (MAHGOUB, NOLLET, 2018).

Como definição os Organismos Geneticamente Modificados (OGM's) são qualquer organismo em que o material genético tenha sido alterado de uma maneira que não ocorra naturalmente por acasalamento e/ou recombinação natural (MAHGOUB, NOLLET, 2018). O termo Organismos Geneticamente Modificados e transgênicos estão sendo utilizados de forma errônea como sinônimos, mas é importante diferenciá-los. Alguns autores consideram como transgênicos todo organismo cujo material genético sofreu alteração em sua estrutura a partir da técnica do DNA recombinante, com a introdução de fragmentos genéticos de um organismo de espécie diferente na espécie destinatária. Deste modo os transgênicos podem ser considerados OGM, mas os OGM's não são obrigatoriamente considerados transgênicos (BECHANE, 2012).

Nos últimos anos a produção de transgênicos se expandiu praticamente em todas as regiões agrícolas do planeta (JAMES, 2013) devido às inúmeras vantagens que se tem quando utilizados cultivares transgênicas, como aumento na produtividade, menos uso de herbicidas, culturas adaptadas às mais diversas regiões, preço mais baixo, maior durabilidade ou maior valor nutricional (MAHGOUB, NOLLET, 2018).

A área global de cultivos desenvolvidos por meio da biotecnologia cresceu aproximadamente 113 vezes, de 1.7 milhões de hectares em 1996 para 191,7 milhões de hectares em 2018, isso faz das culturas biotecnológicas a tecnologia de culturas adotada mais rapidamente nos últimos anos (ISAAA, 2018).

Os EUA lideraram o plantio de biotecnologia em 2018 em 75 milhões de hectares, seguidos pelo Brasil (51,3 milhões de hectares), Argentina (23,9 milhões de hectares), Canadá (12,7 milhões de hectares) e Índia (11,6 milhões de hectares), totalizando 174,5 milhões de hectares, representando 91% da área global. Assim, a biotecnologia beneficiou mais de 1,95 bilhão de pessoas nos cinco países ou 26% da população mundial atual de 7,7 bilhões (ISAAA, 2018).

Em relação ao Brasil a área destinada ao cultivo de algodão transgênicos aumentou significativamente em 2018 em comparação a 2017. Algumas das razões

para esse crescimento foram a lucratividade, preços mais altos, alta demanda do mercado nacional e internacional, e a tecnologias de sementes disponíveis. Foi realizado um estudo onde indicou uma imensa redução na aplicação de pesticidas por hectare e menos perdas causadas por pragas nos últimos 20 anos do cultivo de biotecnologia no Brasil. Conseqüentemente, a produtividade e o rendimento das culturas biotecnológicas têm sido, em média, superiores às culturas convencionais. (ISAAA, 2018).

Com todo o avanço dessa tecnologia é importante ter algumas medidas de controle para gerenciar e limitar o contato entre OGM, seres humanos e o ambiente ao redor. Isso é necessário para fornecer e garantir um alto nível de segurança: uma questão muito voltada para a produção de OGM. Já que qualquer atividade que envolva OGM's necessita de barreiras físicas, químicas ou biológicas que serão usadas como medidas de controle para limitar o contato com OGM's, seres humanos e meio ambiente e fornecer um alto nível de segurança. Isso está relacionado ao processo real dos transgênicos e ao uso, armazenamento, transporte e destruição de transgênicos (MAHGOUB, NOLLET, 2018).

3.2 Uso de Transgenia nos Algodoeiros

O papel do algodoeiro foi fundamental na história da economia mundial, permanecendo até a atualidade. O algodoeiro é cultivado em mais de 100 países, em todos os continentes, equivalendo a 2,8% das terras aráveis do globo. A cultura do algodoeiro é explorada por grandes e pequenos agricultores, exercendo grande importância em termos econômicos, sociais e produtivos (ABAPA, 2019).

O Comitê Consultivo Internacional para o Algodão, (do inglês *International Cotton Advisory Committee - ICAC*) aponta que o valor da produção mundial de algodão é de aproximadamente US\$40 bilhões. E cerca de 100 milhões de famílias, em torno de 350 milhões de pessoas no mundo, estão diretamente ligadas à sua produção, seja na agricultura ou nos diversos elos da cadeia produtiva, como transporte, beneficiamento, empacotamento e armazenamento (ABAPA, 2019).

A matéria prima obtida diretamente do algodoeiro são os fios e tecidos. No entanto o algodão é muito versátil e suas subpartes também possuem valor comercial. Esses subprodutos provenientes das sementes do algodoeiro servem para alimentação de gado, fertilizantes, óleo de cozinha e tantas outras aplicações

(CAMPBELL, HINZE, 2010; CAMPBELL et al, 2016). O óleo do caroço do algodão é o segundo óleo de semente mais importante a nível global, que é utilizado para fins culinários. O resíduo do bagaço da produção de óleo é uma ração rica em proteína, servindo para alimentar gado ruminante (SINGH, 2018).

Mesmo sabendo da versatilidade da semente do algodoeiro, os programas de melhoramento de algodão não deram grande ênfase à melhoria dos mesmos. Em um esforço para maximizar o retorno dos investimentos feitos pelos produtores em sistemas de produção de algodão, há interesse em desenvolver programas de melhoramento que também melhorem as características da semente de algodão. Em vez da semente de algodão ser considerada apenas como subproduto, poderia ser estabelecido um sistema conjunto de produção de fibras de algodão e sementes que maximizasse o valor dos componentes de fibra de algodão e sementes (KOTHARI et al., 2015; CAMPBELL et al, 2016).

Para que o algodoeiro possa desenvolver seu potencial produtivo em sua totalidade, tem que ter um cuidado especial desde o início do processo de plantio até a colheita e o seu processamento. Observar logo no início do processo o nível de competição que possa existir entre o algodão e outras espécies vegetais para obtenção de água, nutrientes e luz. A presença de plantas daninhas no final do ciclo, por exemplo acaba dificultando a colheita mecânica e aumenta o nível de impurezas na fibra, reduzindo assim a qualidade da fibra e a rentabilidade do produtor (EMBRAPA, 2007).

O Brasil é um dos poucos países que conseguem produzir algodão em condições tropicais úmidas, com excesso de chuvas durante o ciclo da planta. Essa é uma das razões por que o espectro de pragas e doenças que prejudicam o cultivo é amplo e diversificado, exigindo diferentes medidas de controle, tanto cultural como químicas ou biológicas (BÉLOT, BARROS, MIRANDA, 2016).

Muitos dos danos aos algodoeiros são causados por insetos e pragas, a principal praga é o bicudo (*Anthrenus grandis*), um coleóptero que induz a abscisão de estruturas reprodutivas ao se alimentar ou ovopositar em botões florais. Outra praga importante é o pulgão do algodoeiro (*Aphis gossypii*), sendo uma das primeiras pragas a aparecer na cultura do algodão, preferem se alimentar das folhas e brotos da planta provocando deformações (PAPA, 2006; SILVIE, BÉLOT, MICHEL, 2007; PESSOA et al., 2017) e vetor do Cotton Leaf Roll D, art Virus, que transmite um vírus capaz de causar grandes prejuízos em cultivares susceptíveis à doença

azul (RICHETTI,2006). Para combater essas pragas, é utilizado uma grande quantidade de inseticidas, Segundo Brasil, 2005 somente na cultura do algodoeiro no Brasil são consumidos mais de 10 toneladas de inseticidas e esse uso excessivo de inseticidas não específicos causam impactos ambientais negativos, como a grave redução da população de organismos benéficos e a potencialização do surgimento de pragas resistentes aos inseticidas convencionais

Nos dias atuais, o uso de herbicidas é uma realidade na cultura do algodoeiro. Mas com o avanço das técnicas transgênicas no meio agrícola, e com o surgimento de variedades de algodão transgênico tolerantes a herbicidas e a insetos e pragas, essas técnicas constituem ferramentas importantes para o manejo da cultura (SANTOS, TORRES, 2010). Elas podem ser introduzidas nos diferentes sistemas de produção, reduzindo os prejuízos causados pelas pragas. O benefício da inserção de cultivares transgênicas tolerantes a herbicidas ou resistentes a insetos ocorre por causa da redução no número de aplicação de inseticidas, diminuindo assim a depreciação dos equipamentos envolvidos nessa operação e aumento da população de inimigos naturais de insetos-praga, diminuindo as perdas na lavoura (BENNETT; MORSE; ISMAEL, 2006).

No caso do algodão quando se deseja implantar uma variedade transgênica em um determinado local tem a necessidade de fazer uma avaliação de risco ambiental cientificamente fundamentada, para ter certeza de que a variedade transgênica não contamine variedades nativas. Existe uma série de itens básicos que são solicitados pela CTNBio (Comissão Técnica Nacional de Biossegurança) que possui instruções normativas sobre este processo. Este procedimento é necessário para que seja preservada a identidade genética dos algodões nativos, como por exemplo a variedade *G. barbadense* que está presente em comunidades indígenas (EMBRAPA, 2007).

3.3 Processo para Obtenção de uma Planta Transgênica

Genoma é toda a informação hereditária de um organismo, que está codificada em seu DNA (ou, em alguns vírus, no RNA). Isto inclui tanto os genes como as sequências não-codificadoras que são muito importantes para a regulação gênica, dentre outras funções. Assim todo organismo vivo como bactéria, plantas, animais bem como em vírus, possuem um genoma, e este abriga as informações

genéticas de cada ser, estando presente em todas as células de cada organismo e definindo todas as suas características. Sabendo que a presença do DNA é universal nos organismos vivos, e que tanto a função quanto a atuação do DNA nesses organismos ocorrem de forma semelhante, pode-se identificar as características de interesse e transferi-las para outro organismo. (DELATORRE, 2005).

A transformação genética em vegetais só foi possível a partir do desenvolvimento das técnicas de cultura de tecido vegetal. Essas técnicas acabaram possibilitando a obtenção de uma planta a partir de uma única célula vegetal. O processo para a obtenção de uma planta transgênica é composto por alguns passos como o isolamento, identificação e introdução do DNA exógeno na célula, a seleção e crescimento das células transformadas e o estabelecimento de um sistema simples e eficiente de regeneração das células transformadas (BRASILEIRO & DUSI, 1999).

Segundo Powell (2015) O processo da produção de um OGM envolve as seguintes etapas básicas: identificar uma característica de interesse, isolar essa característica genética, inserir essa característica no genoma de um organismo desejado e, em seguida, cultivar o organismo manipulado. As etapas estão detalhadas a seguir:

- Etapa 1: Extração de ácido nucleico (DNA / RNA)

Em qualquer procedimento de isolamento, a etapa inicial é a ruptura da célula, a fim de quebrar a parede celular do organismo que contém a característica desejada, que pode ser viral, bacteriana ou vegetal. Isso ajuda a extração do ácido nucleico;

- Etapa 2: Clonagem Genética

O processo de clonagem envolve quatro etapas: geração de fragmentos de DNA, adesão a um vetor, propagação em uma célula hospedeira e seleção da sequência necessária;

- Etapa 3: Design e embalagem de genes

Uma vez que o gene desejado tenha sido clonado, ele deverá estar ligado a pedaços de DNA que controlarão sua expressão dentro da célula vegetal;

- Etapa 4: Transformação (inserção de gene);

- Etapa 5: criação de retrocessos.

Essas etapas são iniciadas no laboratório de biotecnologia, depois transferidas para a estufa para o crescimento inicial e finalmente levadas para o campo onde a cultura se desenvolve até a maturidade (MAHGOUB, NOLLET, 2018).

Por fim existem procedimentos de regulamentação envolvidos na autorização de um novo evento transgênico, até o lançamento das sementes transgênicas no mercado, que são complexos e passam por diversos setores do poder público, a aprovação comercial é concedida pela Comissão Técnica Nacional de Biossegurança (CTNBio) e o posterior registro das cultivares derivadas desse evento junto ao Registro Nacional de Cultivares (RNC) do Ministério da Agricultura, Pecuária e Abastecimento do Governo Federal do Brasil (MOTA, 2011).

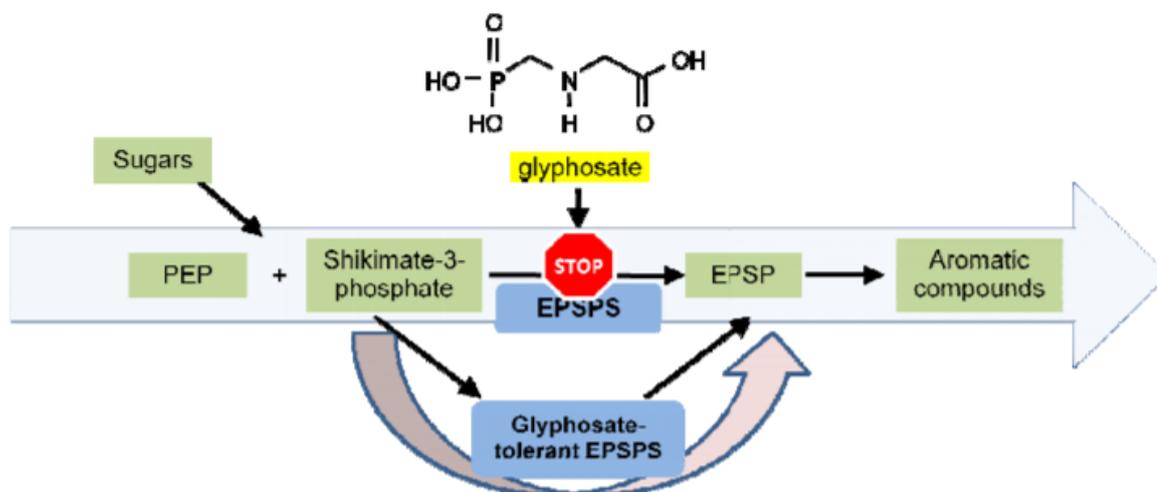
3.3.1 Processo de Obtenção da Cultivar BRS 368 RF

A cultivar transgênica, BRS 368 RF, foi originada de um cruzamento biparental, realizado pela Embrapa Algodão, obtido a partir de um programa de melhoramento de retrocruzamento para incorporação do evento comercial MON 88913, que é conhecido como Roundup Ready® Flex (RF), para conferir resistência ao herbicida glifosato em germoplasma de algodão. Esse cruzamento se deu com a cultivar convencional BRS 286, proveniente da Embrapa, que foi utilizada como pai recorrente e a cultivar de algodão de alto rendimento Sure Grow 125 utilizada como doadora de características, proveniente da Monsanto (BOWMAN et al, 2006; BARROSO et al, 2017).

A Monsanto Company desenvolveu um produto de algodão tolerante ao glifosato através da transformação mediada por *Agrobacterium tumefaciens*. O algodão MON 88913 expressa a proteína 5 enolpiruvilshiquimato-3 fosfato sintase (EPSPS) da *Agrobacterium tumefaciens* CP4 (CP4 EPSPS), e quando a CP4 EPSPS é tratada com glifosato, as plantas continuam se desenvolvendo normalmente. Esse produto de algodão tolerante ao glifosato é projetado para fornecer aos cultivadores um controle aprimorado de ervas daninhas através da tolerância a taxas mais altas de glifosato e maior flexibilidade para aplicação de herbicida glifosato (MONSANTO EMEA, 2015).

Os herbicidas contêm o ingrediente ativo glifosato, um herbicida de amplo espectro que atua através da inibição da proteína EPSPS. O EPSPS é encontrado naturalmente em todas as plantas, fungos e bactérias e é importante na produção de aminoácidos aromáticos essenciais. A inibição do EPSPS pelo glifosato bloqueia a produção desses aminoácidos, interferindo no crescimento e levando à morte das plantas (veja a Figura 1).

Figura 1: Esquema do modo de ação da modificação genética em MON 88913



PEP = Phosphoenolpyruvate
 EPSP = Enolpyruvyl-skikimate-phosphate
 EPSPS = 5-enolpyruvyl-skikimate-3-phosphate synthase

Fonte: MONSANTO EMEA, 2015.

As plantas de algodão MON 88913 contêm um EPSPS tolerante ao glifosato, isolado da cepa CP4 da bactéria comum do solo *A. tumefaciens*. A presença do EPSPS tolerante ao glifosato garante a função contínua da via de aminoácidos aromáticos, mesmo na presença do herbicida glifosato. Como resultado dessa modificação genética, o glifosato de algodão MON 88913, permite o controle seletivo de plantas daninhas na cultura (MONSANTO EMEA, 2015).

A cultivar BRS 286 tem como características: a boa qualidade de fibra, precocidade, adaptação ao cerrado nordestino, possui níveis adequados de resistência às principais doenças, sendo resistente à mancha angular, mosaico da nervura, mosaico comum, entre outros. Já o Sure Grow 125 RF é susceptível à doença do algodão azul, causada pelo vírus anão do algodoeiro Leafroll e não tem boa qualidade de fibra (BARROSO et al, 2017).

A cultivar BRS 368 RF foi lançado para atender à demanda por cultivares de algodão de alto desempenho e características transgênicas. Ela é recomendada para todas as áreas de cultivo de algodão do cerrado norte e nordeste brasileiro (BARROSO et al, 2017), tem como características ser de porte baixo, tem o ciclo de médio a precoce, com rendimento de fibra de 40% e potencial produtivo entre 4.200 e 4.500 kg/ha e é resistente ao herbicida glifosato. Por esse motivo, a mesma oferece maior flexibilidade no controle de plantas daninhas, permitindo a aplicação em qualquer fase do desenvolvimento da cultura sem necessidade de pulverizações com herbicidas não seletivos em jato dirigido (EMBRAPA, 2017).

Já a cultivar convencional (BRS Aroeira), produz em média 3841 kg/ha, dentre as cultivares plantadas no Brasil, esta é a que dispõe de maior teor de óleo na semente (25% a 27%) comparada com as outras que possuem em média 14% de óleo, possui fibras mais finas, com maior comprimento, representam uma alternativa para a produção de custo baixo e possui resistência múltipla a doenças como ramulose, viroses, mancha de stemphylium, além de tolerância a bacteriose, manchas de ramularia, entre outros (FREIRE, 2002).

3.4 Vantagens e Desvantagens dos Transgênicos

As principais controvérsias e questionamentos relacionadas aos produtos geneticamente modificados (OGM's) são os riscos que podem ser causados por estes alimentos, se eles devem ou não ser rotulados, qual deve ser o papel e as atitudes a serem tomadas dos governos frente ao tema, quais os efeitos que trazem ao meio ambiente e a saúde dos consumidores, qual o impacto que poderá ter na vida dos agricultores, qual o papel dessas culturas na alimentação da crescente população mundial e no sistema de agricultura industrial. As culturas transgênicas oferecem algumas alternativas para soluções de alguns problemas como adaptação de alguma cultivar em ambientes com climas extremos. Como toda tecnologia nova, a mesma pode apresentar certos riscos, principalmente quando se trata de OGMs que podem reunir combinações de genes que não são encontrados na natureza, dessa forma se torna um organismo questionável quanto ao seu uso (KUMAR et al, 2018).

Alguns defensores dos transgênicos afirmam que estes alimentos são tão seguros quanto quaisquer outros alimentos convencionais. Eles acreditam que os reguladores e o processo regulatório são suficientemente objetivos e rigorosos,

confiam que a tecnologia dos OGM's é a chave para alimentar uma crescente população mundial, e veem essa tecnologia como uma continuação da manipulação de plantas que os humanos conduziram por milênios. Sendo a diferença entre o melhoramento genético convencional e o transgênico, esta requer algumas ressalvas laboratórias e uma avaliação rigorosa das plantas transgênicas e do risco potencial que elas representam (KUMAR et al, 2018).

A maioria dos OGM's tem como características serem tolerantes a herbicidas ou resistentes a vírus, fungos ou insetos, características estas que permitem a obtenção de variedades mais robustas, de maior qualidade e rendimento (MARIOTTI et al., 2002; KLUMPER e QUAIM, 2014; RIBEIRO, 2017). Também trazem benefícios como à diminuição da quantidade de defensivos que é utilizado para controle de pragas, houve uma redução de 32% na dosagem desses defensivos aplicados para o algodão, isso reflete a quantidade de ingrediente ativo (defensivos químicos) aplicado nas lavouras que também diminuiu (CIB, 2018). A inclusão de novos traços acaba oferecendo maior produtividade agrícola, ou melhor qualidade e características nutricionais e de processamento, que podem contribuir diretamente para melhorar a saúde e o meio ambiente, pois pode haver benefícios indiretos, como a redução do uso de agrotóxicos, a melhoria da renda agrícola, entre outros (HASLBERGER, 2005).

3.5 Metodologias convencionais para confirmação dos Transgênicos

Os OGM's se consolidaram como a tecnologia mais rápida adotada na história recente da agricultura moderna (ISAAA, 2018), com o crescimento de aprovações para o plantio se dando na mesma velocidade, mas para que um organismo transgênico seja liberado para comercialização o mesmo deve passar por rigorosos testes (BRASIL, 2005). Para isso, mais de 120 instituições públicas e privadas foram credenciadas junto aos órgãos reguladores, para desenvolver pesquisas referentes aos OGM's, e nesses últimos 20 anos de uso de alimentos transgênicos em larga escala, não foram registrados nenhum impacto negativo no meio ambiente ou na saúde humana e animal (EMBRAPA, 2017). Apesar de todos esses fatos a opinião pública sobre o tema é ainda muito divergente, pelo simples fato da insuficiência de pesquisas e de informações concretas sobre os possíveis impactos que os OGM's possam ter nos seres humanos e na natureza, deste modo

o tema acabou chegando ao centro das atenções públicas e os mesmos estão sempre tentando implantar leis que limitem algumas ações frente aos OGM's. Sabe-se que a simples detecção desses organismos não garante a segurança dos alimentos produzidos com OGM's, mas essa detecção se faz necessária, pois permite verificar algum tipo de contaminação em diversos cultivos expostos a riscos biológicos e com implicações econômicas, sociais e ambientais. Também se faz necessários para que o consumidor tenha informações adequadas sobre os produtos que estão adquirindo, e para gerarem confiança nesses alimentos (CONCEIÇÃO, 2006).

Geralmente alguns modos de detecção dos OGM's são caracterizados pela presença de um ou mais segmentos de DNA exógenos, podendo ou não acomodar a expressão de novas proteínas. Sendo assim, a detecção destes organismos é baseada na sequência do DNA exógeno ou na proteína transgênica. Em se tratando de produtos alimentícios contendo OGM's, tem-se uma análise de rotina que é realizada em três etapas: a detecção; a identificação do OGM presente na amostra, para verificar se tem autorização para comercialização; e a quantificação do OGM no produto com a finalidade de rotulagem ou não, conforme a legislação (PETIT et al., 2003).

Alguns métodos de detecção são baseados na presença de proteínas como os bioensaios e os imunoenaios. Geralmente os OGM's utilizados na alimentação são tolerantes a herbicidas ou resistentes a insetos, fungos ou vírus (MARIOTTI, 2002), os quais podem ser analisados por bioensaios, nos quais se detecta a presença do herbicida. Em geral são métodos simples e práticos, em que as sementes são colocadas para germinar em um meio contendo uma solução diluída de herbicida. Se a semente for resistente ao herbicida glifosato, ocorrerá a germinação da plântula. Esse bioensaio tem algumas limitações como a necessidade de um longo tempo para a obtenção dos resultados e a utilização restrita à OGM's resistentes à herbicidas (TORRES et al., 2003).

Segundo Stave (1999), os imunoenaios são ideais para a detecção qualitativa e quantitativa de proteínas em misturas complexas. É baseado na concentração típica de proteínas transgênicas em tecidos vegetais ($> 10\mu\text{g. g}^{-1}$ de tecido). O limite de detecção de um imunensaio é de aproximadamente 1%, no entanto, nem sempre é possível a detecção de OGM's por este método. Como exemplo, na produção de alimentos, quando o nível de expressão da proteína

transgênica é muito baixo (AHMED, 1995), ou quando a concentração da proteína transgênica nos tecidos varia em função da idade, variedade e condições ambientais (STAVE, 2002). Outra limitação é quando ocorre o refinamento de alguns xaropes e óleos, ou a desnaturação dos mesmos, alterando a sua conformação e impedindo o seu reconhecimento pelo anticorpo (STAVE, 2002; LIU, 2004). Dentre os métodos de imunoenaios pode-se destacar o ensaio imunoenzimático (ELISA), o imunoensaio de fluxo lateral (IFL) e o Western blot, que são bastantes utilizados na detecção e quantificação de alimentos contendo OGM's (MAGIN, 2000).

Existem também métodos de detecção baseados na presença de DNA, já que geralmente pode-se estabelecer uma correlação linear entre a quantidade de OGM e DNA exógeno quando a modificação genética é nuclear. No entanto, em se tratando de modificação genética não nuclear, como é o caso de quando o DNA exógeno é introduzido em uma organela como o cloroplasto, essa correlação não é possível. Até o momento, todos os OGM's comercializados possuem modificação genética nuclear (CONCEIÇÃO et al., 2004).

Por ser mais estável que a maioria das proteínas, o DNA é passível de ser detectado em alimentos processados, e o principal método utilizado na detecção e quantificação de alimentos contendo OGM's é a Reação em Cadeia da Polimerase (PCR), que consiste na amplificação seletiva de sequências específicas da molécula de DNA. Este método é sensível, específico, seguro, capaz de detectar uma ampla série de eventos (BERTHEAU et al., 2002; GIOVANNINI & CONCILLO, 2002), e distinguir as variedades geneticamente modificadas que apresentam diferentes construções gênicas, mesmo que expressam a mesma proteína (YAMAGUCHI et al., 2003).

As principais limitações do PCR são as dificuldades na construção dos indicadores, pois suas informações sobre a sequência da modificação genética geralmente são confidenciais (HOLST-JENSEN et al., 2003), necessidade de pessoal treinado e de equipamentos especiais, custo elevado, já que cada teste é específico para uma única modificação genética, necessitando também de material de referência certificado, cuja disponibilidade geralmente é limitada (MIRAGLIA et al., 2004).

Outras formas de detecção e quantificação de OGM's vem sendo desenvolvidas, essas novas metodologias têm maior sensibilidade, confiabilidade, custos mais baixos, com certa rapidez quando comparadas aos métodos

convencionais mostrados anteriormente. Dentre esses novos métodos vem se destacando a cromatografia, a espectrometria de massa, a espectrometria no infravermelho próximo, a eletroforese capilar (HOLST-JENSEN et al., 2003; BORDONI et al., 2004).

Já a espectroscopia de infravermelho próximo (NIR), pode ser empregada na detecção e quantificação de OGM's. Como vantagens essa técnica é rápida, não requer a destruição da amostra, mesmo diante dessas vantagens, a quantificação ainda precisa ser validada (HURBURGH, et al., 2000; KOK et al., 2002). No entanto, os espectrômetros NIR não conseguem captar gradientes internos nos produtos alimentícios, podendo levar a uma discrepância entre as previsões da composição medida (Ariana, Lu, & Guyer, 2006).

3.6 Espectroscopia na Região do Infravermelho

Existem várias técnicas de espectroscopia que têm sido aplicadas para diversos processos analíticos, tais como a espectroscopia de infravermelho próximo, a espectroscopia de infravermelho médio (MIR) e a espectroscopia Raman. Nos últimos anos, a espectroscopia de infravermelho próximo (NIR) tem despertado grande interesse num contexto PAT (Tecnologia Analítica de Processo), devido à sua versatilidade analítica, largamente reconhecida como um método analítico poderoso para análises farmacêuticas e químicas. Devido principalmente à sua potencialidade em medições qualitativas, como por exemplo na identificação, e quantitativas, como no caso do doseamento, sendo rápidas e não destrutivas num processo (CEITIL, 2007).

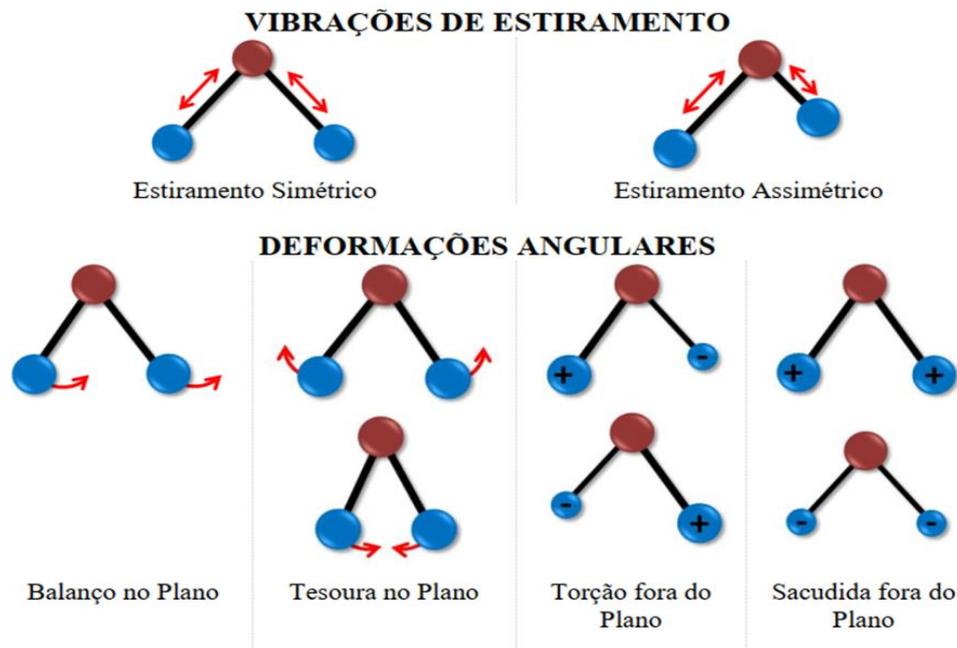
A região do infravermelho compreende a radiação eletromagnética com comprimentos de onda de 780 a 1.000.000 nm, sendo subdividida em três sub-regiões: infravermelho próximo (780 – 2.500 nm), infravermelho médio (2.500 – 50.000 nm) e infravermelho distante – FAR (50.000 – 1.000.000 nm) (SKOOG; HOLLER; NIEMAN, 2009). Assim, a espectroscopia de infravermelho próximo (NIR) é um tipo de espectroscopia de alta energia (PASQUINI, 2013; PASQUINI, 2018).

As frequências de radiação na região do infravermelho são baixas e, quando absorvidas por um sistema molecular são convertidas em energias rotacionais e vibracionais (SILVERSTEIN et al., 2005). É preciso então que a molécula possua uma variação no momento dipolo de uma ligação (sobretons) durante seu

movimento rotacional ou vibracional para que ela absorva a radiação infravermelha (SKOOG; HOLLER; NIEMAN, 2009).

As vibrações e rotações podem ocorrer de diferentes formas nas ligações da molécula, as posições relativas aos átomos oscilam continuamente, não sendo fixas, essas vibrações podem ser classificadas em estiramento e deformação (SKOOG; HOLLER; NIEMAN, 2009). Podemos ver esses comportamentos na Figura 2.

Figura 2: Diferentes modos de vibração em uma molécula triatômica heteronuclear



Fonte: SKOOG, 2006.

As bandas no infravermelho próximo são muito mais fracas que as vibrações fundamentais no infravermelho médio. Esta natureza fraca da absorção de NIR é uma vantagem analítica visto que permite uma análise direta de matrizes com absorções fortes com alta dispersão de luz, tais como lamas, suspensões, pastas e pós. As bandas de absorção de NIR são amplas e muito sobrepostas. Quando os comprimentos de onda com bandas longas, as bandas são fortes, nítidas e com melhores resoluções (ARNOLD, 2002).

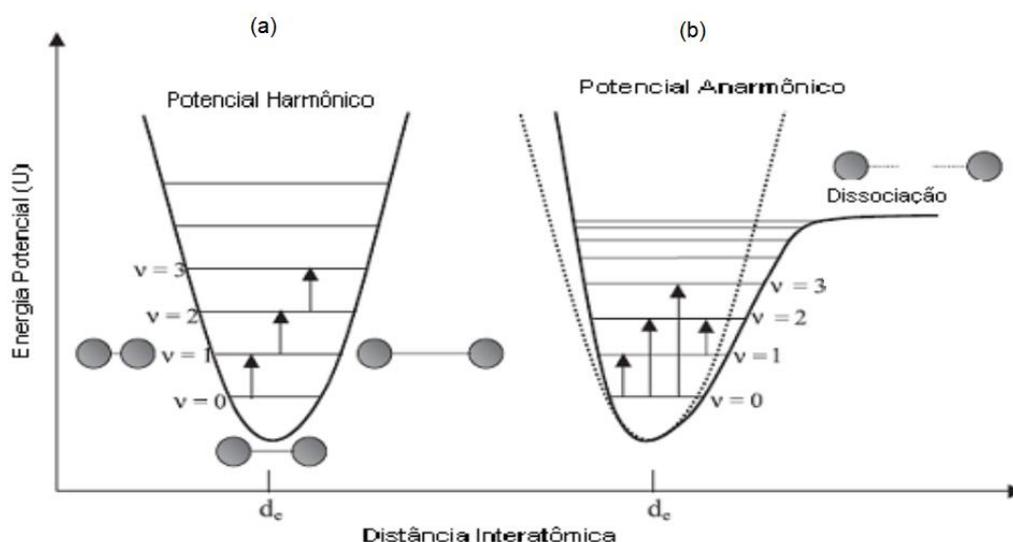
A interação da radiação NIR com a amostra pode ser promovida de várias maneiras, como na transmitância, refletância difusa, transmitância difusa, interatividade, entre outros. Esses termos são empregados para se referir a diferentes arranjos geométricos do feixe de radiação de sondagem, amostra e sistema de detecção usados para coletar informações espectrais analíticas sobre a

amostra (PASQUINI, 2003, OZAKI, 2012; PASQUINI, 2018). Dentre esses modos de medição, o que predomina para análise de sólidos são os baseados em refletância difusa. A absorção e a dispersão da radiação são os principais fenômenos que afetam o espectro de refletância, sendo frequentemente usada por que a NIR tem sido usada para a análise não destrutiva e direta de sólidos podem ser sondados de maneira não invasiva por refletância difusa para produzir informações analíticas úteis (PASQUINI, 2018).

Os espectros NIR, tanto contém informação sobre a composição química da amostra, como também relacionam a informação física da mesma. Assim se tiver o aumento do comprimento de onda, aumenta também a absorvidade e a eficiência da dispersão de luz, enquanto a sobreposição das bandas e a profundidade de penetração são inversamente proporcionais. Portanto, a região espectral ideal para ser usada para a análise de uma amostra é determinada pela correspondência entre as propriedades espectrais da região NIR com a performance analítica requerida, as requisições de amostragem e as propriedades físicas da amostra (METROHM, 2013).

Em se tratando do princípio teórico, a radiação ao incidir na molécula provoca a excitação dos modos vibracionais, correspondendo à diferença de dois níveis energéticos. Assim é possível que ocorra absorção por parte da molécula. Deste modo, as absorções ocorrem em um determinado comprimento de onda, correspondendo à energia que permite a transição (ALMEIDA, 2009).

Figura 3: (a) Diagrama de energia potencial para os osciladores harmônicos e (b) anarmônico.



Fonte: PASQUINI, 2003

A intensidade de absorção da radiação NIR está relacionada ao grau de anarmonicidade elétrica e mecânica do modo vibracional e à mudança no momento dipolar que elas produzem. Isso, em princípio, deve conferir uma sensibilidade relativamente menor aos métodos analíticos baseados no NIRS. No entanto, a existência de fontes de radiação mais brilhantes e bons detectores para essa região espectral compensa, pelo menos em parte, esse aspecto negativo. Devido a essas características, é possível concluir que a NIRS é capaz de produzir uma resposta analítica, considerando amostras de composição orgânica (PASQUINI, 2018).

As informações analíticas contidas nos espectros do NIR apresentam algumas desvantagens como serem muito complexas, alta correlação e difíceis de interpretar diretamente. Na maioria das vezes, as informações são dispersas por todo o espectro. Isso ocorre porque, geralmente, a maioria dos constituintes da matriz de amostra também fornecerá alguma contribuição para o espectro de absorção / reflectância do NIR, além das informações espectrais devidas ao próprio analito. Além disso, propriedades físicas, como a distribuição granulométrica, proporcionam efeitos significativos e geralmente indesejados sobre o espectro NIR (PASQUINI, 2018).

Dessa forma os espectros necessitam de um pré-processamento para que as diferenças espectrais sejam sobressaídas, assim à primeira vista pode não ser possível detectá-las. Além do mais, tem a exigência de se ter pessoas com experiência para aplicar as ferramentas matemáticas no desenvolvimento de modelos de calibração (MAGALHÃES, 2014). Algumas das desvantagens enumeradas são ultrapassadas, com a utilização de técnicas quimiométricas (PEREIRA, 2011).

3.6.1 Utilização da Espectroscopia NIR nos Transgênicos

Quando se trata da lógica por trás da análise transgênica utilizando o NIR primeiro se observa a característica da absorção espectral de ligações moleculares, tais como C-H, C-N, O-H e S-H, que está relacionada às mudanças fenotípicas causadas por mudanças genotípicas (nível de expressão). Assim a espectroscopia NIR se torna uma ferramenta alternativa útil para a identificação de transgênicos, pois possui como características ser um método que não necessita da preparação da amostra, o tempo e o custo de análise são reduzidos, pode ter a caracterização

simultânea de múltiplos componentes influenciados pelo genótipo, viabilidade de se fazer a análise online, é uma técnica não destrutiva e não invasiva (HAI-FENG et al, 2012).

Alishashi et al (2010) publicaram em um artigo vários trabalhos que tinham sido desenvolvidos até então envolvendo a espectroscopia NIR e transgênicos. Dentre esses, vale ressaltar a pesquisa de:

Hurburgh et al (2000) usaram espectroscopia NIR para distinguir grãos transgênicos. Eles mostraram que o método PCA teve um bom desempenho para separar grãos completamente transgênicos de não transgênicos. Concluindo então que a utilização da tecnologia NIRS como método não destrutivo tinha uma eficiência adequada para detectar grãos transgênicos.

No trabalho de Munck et al. (2004) os mesmos conseguiram classificar a cevada transgênica e não-transgênica por espectroscopia NIR e PCR. Os resultados mostraram que, embora a espectroscopia NIR não possa detectar as alterações na estrutura do DNA e nos genótipos, ela pode identificar as alterações fenotípicas como consequência das alterações genotípicas.

Já Xie et al.(2007) consideraram o potencial da tecnologia NIR para classificar os tomates transgênicos. Eles usaram vários métodos para avaliar os tomates transgênicos e apresentaram as vantagens e desvantagens. Nesta pesquisa as discrepâncias entre as diversas variedades de tomate transgênico e não transgênico foram reconhecidas adequadamente pelo NIRS, e chegaram à conclusão que a utilização desta técnica em comparação com outros métodos (ou seja, PCR, ELISA) foi muito melhor.

Outros trabalhos também foram publicados como o de Cui (2012), voltado para a discriminação automática e rápida de genótipos de algodão por Espectroscopia no Infravermelho Próximo e Quimiometria, onde foi investigado a influência de diferentes métodos de pré-processamento de dados no desempenho da previsão do modelo, obtendo resultados que demonstraram que o rPCA pode efetivamente detectar discrepâncias e manter a eficiência da análise discriminante. A precisão da classificação foi aprimorada quando utilizou a derivada de segunda ordem e a variável normal padrão (SNV). Os melhores modelos de análise discriminante de mínimos quadrados parciais (PLS-DA) obtiveram precisão total de classificação de 100% e 97,6% para sementes e folhas, respectivamente.

Luna et al (2013) conseguiu distinguir óleos de soja transgênicos dos não-transgênicos usando espectroscopia NIR e métodos quimiométricos, onde foi utilizada a classificação multivariada. Os melhores resultados de classificação foram obtidos com a Análise de Discriminação de Vetores de Suporte (SVM-DA) com um acerto de 100% e Análise Discriminante por Mínimos de Quadrados Parciais (PLS-DA). Após centralização média mais correção de dispersão multiplicativa a porcentagem de classificação bem-sucedida foi de 95% e 100% no grupo de treinamento para amostras de óleo de soja não transgênicos e transgênicos, respectivamente, no grupo de validação obteve 100% para não-transgênicos e 80% para transgênicos. Como conclusão a espectroscopia NIR pode fornecer um método rápido, não destrutivo e confiável para distinguir óleos de soja não transgênicos e transgênicos.

O trabalho de Hao et al, (2019) foi voltado para a identificação de variedades de arroz e características transgênicas em espectroscopia de reflexão difusa no infravermelho próximo e quimiometria. Os resultados mostraram que a espectroscopia NIR combinada com métodos quimiométricos poderia ser usada para identificar variedades de arroz e características transgênicas (Bt63).

3.7 Imagem hiperespectral no infravermelho próximo

As imagens são ferramentas úteis e versáteis para se obter informações analíticas, tanto qualitativas quanto quantitativas, por representarem grandes conjuntos de dados visualmente. A aplicação das imagens como ferramenta analítica se deu graças ao desenvolvimento de novos dispositivos para aquisição de imagens químicas, além de computadores modernos e softwares capazes de tratar grandes conjuntos de dados, afim de extrair as informações relevantes (GOWEN et al, 2007; GRIFFITHS, 2011; VIDAL, AMIGO, 2012).

Um aspecto importante ligado ao caráter local das imagens é correlacionar as propriedades físicas e químicas nas vizinhanças dos pixels locais. Dessa forma, estudar imagens como se fossem um conjunto de pixels individuais isolados acaba não explorando todo o potencial da medição, mas utilizar o conceito de vizinhança de pixel ajuda a definir melhor certas características de uma imagem de amostra, como heterogeneidade, ou a mapear a complexidade da composição local (DE JUAN, 2017).

A classificação das imagens pode ser feita de acordo com algumas propriedades como funcionalidade, tipo de radiação espectroscópica, estrutura espacial. Esta última é provavelmente a classificação mais interessante quando se trata de análise de dados. De uma forma geral, uma imagem é uma representação espacial de um objeto, a resolução espacial que a imagem contém varia dependendo da proporção da imagem e o tamanho do pixel (VIDAL, AMIGO, 2012).

As imagens de canal único ou em escala de cinza são, provavelmente, as imagens mais simples, tendo apenas dois valores possíveis para cada pixel, sendo conhecidas como imagens binárias como por exemplo, preto e branco. As imagens binárias geralmente surgem no processamento de imagens digitais como máscaras ou como resultado de certas operações, como segmentação ou limiar. As imagens de canal único também podem ser compostas por tons de cinza, denotando diferentes intensidades para cada pixel (VIDAL, AMIGO, 2012).

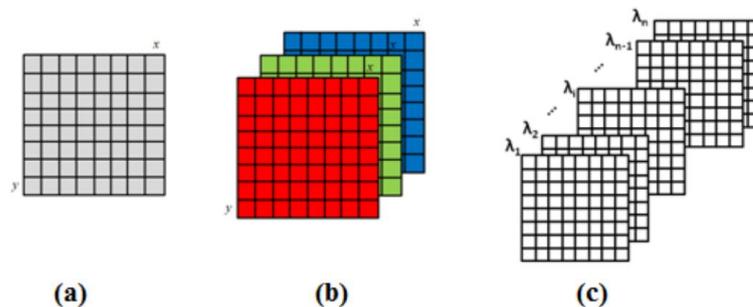
Originalmente na análise de imagens em escala de cinza, como não existia informações sobre as cores, a textura, que pode ser definida como uma função da variação espacial das intensidades de pixel, se tornou uma característica muito importante para reconhecer e descrever objetos em uma imagem (PRATS-MONTALBÁN, DE JUAN, FERRER, 2011).

Já as imagens de cores são aquelas em que a cor de cada pixel é definida pela combinação de cores diferentes em uma determinada dimensão, o mais usado é em escala de Vermelho, Verde e Azul do inglês (RGB – Red, Green, Blue). O modelo de cores RGB segue aproximadamente os receptores de cores no olho humano. Nesse caso, cada pixel contém três canais de informações pertinentes as três cores citadas acima (VIDAL, AMIGO, 2012). As informações relacionadas a cada pixel cresceram de maneira muito mais significativa com as imagens multiespectrais, que possuem poucos canais espectrais, e hiperespectrais, que possuem um espectro completo por pixel. O aumento da direção espectral acabou oferecendo muitas possibilidades, já que agora as diferenças na superfície varrida não dizem apenas sobre a textura da superfície, como também sobre a composição química (PRATS-MONTALBÁN, DE JUAN, FERRER, 2011).

Para imagens digitais, o menor elemento formador é o *pixel* (no caso de imagens de superfície, ou seja, em duas dimensões - 2D) ou o *voxel* (para imagens tridimensionais – 3D). Essas imagens são formadas por coordenadas espaciais e valores numéricos para cada pixel, representando a escala de cores utilizada. Na

escala cinza por exemplo, tem-se uma matriz de dados bidimensionais e cada pixel condiz a um valor de intensidade na escala cinza (Figura 4.a). Para imagens RGB acontece da mesma forma, a diferença é que para estas existem três canais de cores, tendo assim uma matriz de dados tridimensional, onde cada pixel terá três valores de intensidade para cada canal de cor (Figura 4.b), (PRATS-MONTALBÁN, DE JUAN, FERRER, 2011).

Figura 4: Matriz de dados de imagens (a) em escalas de cinza, (b) em RGB e (c) Hiperespectrais.



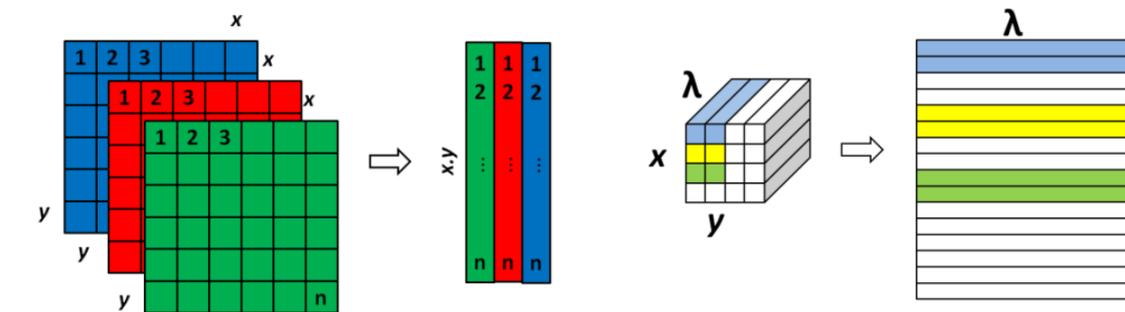
Fonte: DE JUAN, 2009.

Já as Imagens Hiperespectrais podem ser definidas como estruturas tridimensionais (hipercubos) formadas por três dimensões, sendo duas espaciais ($x - y$) e uma espectral, referente ao comprimento de onda (λ) (BURGER e GOWEN, 2011; GOWEN et al., 2011). Deste modo além de fornecer as informações espaciais sobre a amostra, também fornece informações a respeito da composição química desta. As imagens hiperespectrais possuem um espectro de alta resolução ou alguns comprimentos de onda associados a cada pixel da imagem (Figura 5.c) (DE JUAN et al, 2009). Essa dualidade de informações espectrais e espaciais (distributivas) deve ser levada em consideração para extrair o máximo de informações desse tipo de medição (PRATS-MONTALBÁN, DE JUAN, FERRER, 2011).

Imagens hiperespectrais são um tipo específico de imagem para o qual um espectro completo é registrado por cada pixel. No entanto, o fato de ser obtido um espectro completo por pixel indica que o uso dessas imagens se concentra mais nas informações relacionadas à composição química da amostra, bem refletidas na forma e intensidade dos espectros obtidos, em vez de propriedades físicas, como por exemplo, textura, superfície ou volume digitalizado (GRAHN, GELADI, 2007; SALZER, SIESLER, 2009; PRATS-MONTALBÁN, DE JUAN, FERRER, 2011).

Como os hipercubos possuem um grande volume de dados, estes necessitam de um tratamento matemático para poderem fornecer informações sobre as amostras (WU e SUN, 2013). Para desdobrar o hipercubo realiza-se uma aplicação de métodos matemáticos, para que o mesmo passe das três dimensões (x, y, z) para duas dimensões $\{(X \times Y), \lambda\}$, (FERRARI et al., 2013). Cada pixel pode ser tratado como uma amostra e os valores da intensidade estarão dispostos nas colunas respeitando a ordem apresentada na Figura 5 que mostra o desdobramento das imagens em RGB (5a) e para as Imagens Hipererspectrais (5b), respectivamente.

Figura 5: Desdobramento das matrizes 3D de imagens (a) RGB e (b) hiperespectrais



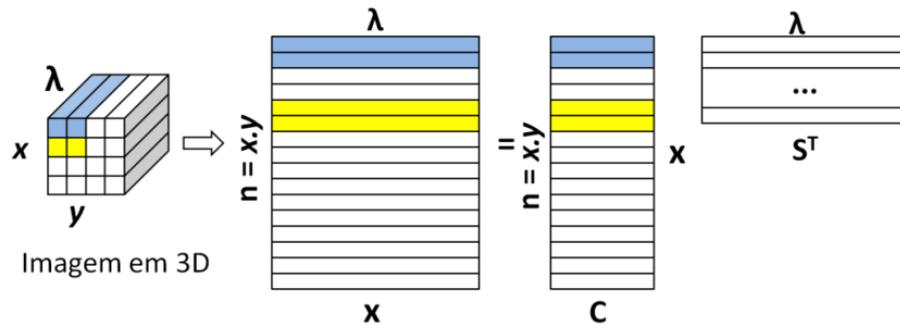
Fonte: PRATS-MONTALBÁN, DE JUAN, FERRER, 2011

Para avaliar uma imagem hiperespectral é necessário entender que tanto os dados espectroscópicos clássicos quanto os espectros das imagens hiperespectrais obedecem à linearidade da lei de Lambert-Beer, assim seguem a equação 1.

$$\mathbf{X} = \mathbf{CS}^t + \mathbf{E} \quad (1)$$

Em que \mathbf{X} é a matriz dos dados, \mathbf{C} é a matriz de intensidade relativa de cada componente presente, \mathbf{S} é a matriz que contém os espectros dos constituintes puros e \mathbf{E} é a matriz de erros que podem estar associados as informações que o modelo não consegue explicar. (PRATS-MONTALBÁN, DE JUAN, FERRER, 2011). O desdobramento de uma imagem hiperespectral está mostrado na Figura 6:

Figura 6: Desdobramento da matriz tridimensional de dados em uma matriz bidimensional e sua decomposição em perfis de concentração relativa e espectros puros.



Fonte: PRATS-MONTALBÁN et al., 2011

A princípio as imagens hiperespectrais foram desenvolvidas para aplicação no sensoriamento remoto (GOETZ,1985), mas devido suas vantagens podem ser aplicadas na ciência de alimentos (BELLON-MAUREL, 2009), na área farmacêutica (AMIGO,2010), no controle de qualidade (VIDAL, AMIGO, 2012) e até mesmo na medicina (NATTKEMPER,2004).

O tratamento de imagens hiperespectrais se faz pela Análise Multivariada de Dados ou de Imagens, a partir das quais as imagens são analisadas utilizando ferramentas matemáticas e estatísticas, permitindo que o desenvolvimento de estudos mais complexos seja desenvolvido. (PRATS-MONTALBÁN, DE JUAN, FERRER, 2011).

A técnica de imagens hiperespectrais vem sendo bastante utilizada na área de reconhecimento de padrões. Se utiliza o reconhecimento de padrões quando se quer obter modelos matemáticos que sejam capazes de reconhecer ou classificar amostras (IQBAL, 2014). Para se aumentar a qualidade dos resultados ao final da análise, se faz necessário nas imagens hiperespectrais, realizar correções nos dados. Para este fim existem algumas técnicas de pré-processamento que são bastantes utilizadas para contornar alguns problemas típicos que estão relacionados não só com as variações de medidas de imagens como também com as variações comuns observadas em um conjunto clássico de espectros. Alguns pré-processamentos como normalização, suavização, derivadas, correções de linhas de base, dentre outras (DE JUAN et al, 2009).

3.8 Reconhecimento de Padrões

Como já mencionado o reconhecimento de padrões é utilizado visando a obtenção de modelos matemáticos que sejam capazes de reconhecer ou classificar amostras (IQBAL et al.,2014). A escolha de qual deve ser utilizada irá depender das informações disponíveis e do objetivo pretendido (GREDILLA et al., 2013). Nas técnicas supervisionadas deve existir um conhecimento prévio sobre a identidade das amostras, para que seja possível a classificação. Já nas técnicas não supervisionadas não é exigido o conhecimento prévio sobre a identidade das amostras (YI et al., 2013).

Quando existe informação a priori sobre as classes ou objetos em uma imagem, as técnicas estatísticas de reconhecimento de padrões podem ser diretamente aplicadas à imagem. Na ausência dessas informações, a imagem pode ser sujeita primeiro a um estágio de segmentação e as classes detectadas podem ser submetidas à ferramenta de reconhecimento de padrões (PRATS-MONTALBÁN, DE JUAN, FERRER, 2011).

3.8.1 *Análise de Componentes Principais (PCA)*

A PCA faz parte das técnicas de reconhecimento de padrões não-supervisionados, cujo princípio é manipular matematicamente um conjunto de dados de natureza multivariada. A mesma é capaz de identificar similaridades e diferenças entre amostras do conjunto de dados, resultando na redução da dimensionalidade das variáveis e representação das variações existentes na forma de componentes principais (PCs), (HUANG, 2015). PCA é provavelmente a técnica que possui uma maior aceitação pela comunidade científica e é largamente utilizada em diversas aplicações (BRERETON, 2003).

No campo de análise de imagens, a PCA se tornou o método mais comum para reduzir a dimensão dos conjuntos de dados, mantendo as informações relevantes. Os componentes principais são novas variáveis não correlacionadas e aproximadamente normalmente distribuídas que fornecem representações fiéis da imagem. A compressão usando componentes principais mantém todas as informações relevantes da imagem e, ao mesmo tempo, permite entender o relacionamento entre as variáveis usadas para construir o modelo, analisando as

estruturas de correlação internas fornecidas pelas cargas (PRATS-MONTALBÁN, DE JUAN, FERRER, 2011).

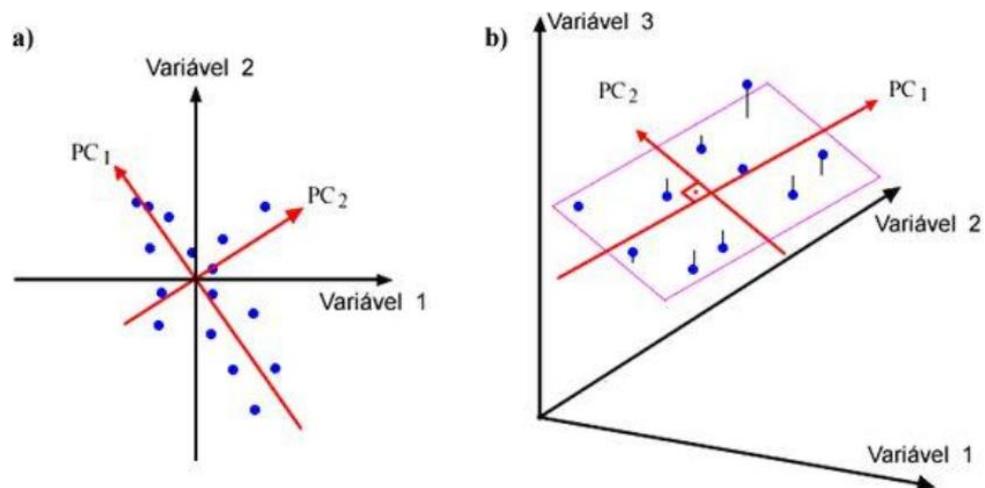
O modelo que representa a transformação da matriz de dados em um modelo bilinear obedece a equação exposta abaixo:

$$\mathbf{X} = \mathbf{T}\mathbf{L}^T + \mathbf{E} \quad (2)$$

Onde na PCA ocorre à decomposição da matriz \mathbf{X} em um produto de duas matrizes, \mathbf{T} é a matriz de escores (são as coordenadas das amostras no novo sistema de eixo) e \mathbf{L} a matriz dos pesos (Loadings, que são os cossenos dos ângulos entre os eixos originais e as componentes principais) e \mathbf{E} é a matriz dos erros.

O novo modelo que é formado a partir de uma combinação linear das variáveis originais e nelas contidas estão as informações mais relevantes dos dados. Assim as representações dos novos eixos formados são denominadas de componentes principais, onde serão sempre ortogonais (perpendiculares) umas às outras e representarão, de forma decrescente, a maior variância dos dados (BORRÁS et al., 2014).

Figura 7: Análise de componentes principais (PCA) – primeira e segunda PCs no espaço tridimensional das variáveis.



Fonte: HUANG et al., 2015

Nos resultados mostrados na PCA tem as imagens dos escores que podem fornecer informações importantes sobre o conjunto de dados, principalmente quando combinados aos gráficos de pesos. Também é possível obter um gráfico de dispersão com a matriz desdobrada de uma imagem de escores de uma

componente versus outra. Este tipo de gráfico pode fornecer informações sobre *outliers* e diferentes classes no conjunto de dados analisados, (GELADI et al., 1989).

PCA é uma técnica bastante utilizada como base para outras técnicas quimiométricas multivariadas como, por exemplo, o SIMCA, que é uma técnica de reconhecimento de padrão supervisionado.

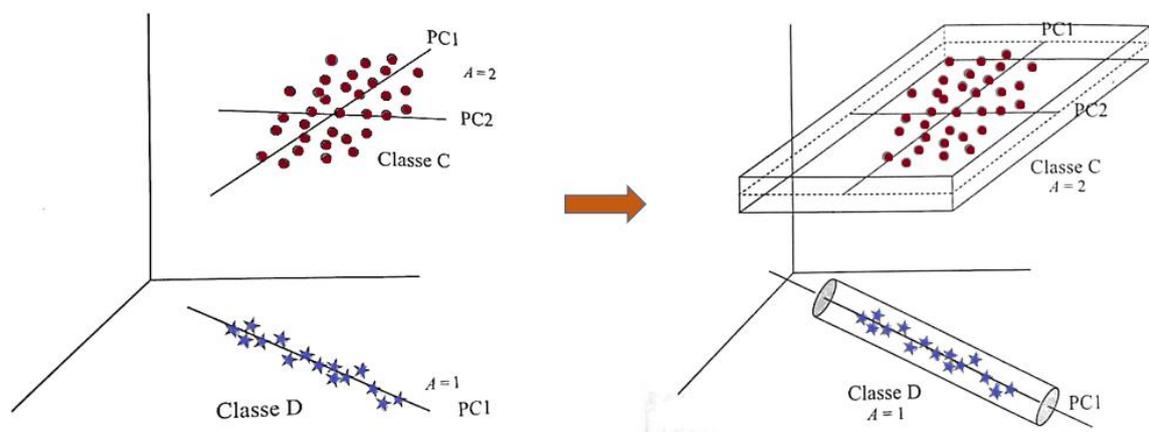
3.8.2 SIMCA

Esse método foi introduzido por Svante Wold nos anos 1970. Como a Quimiometria teve seu início na área de reconhecimento de padrões, SIMCA foi o nome do primeiro pacote computacional, por esse motivo a palavra SIMCA era considerada sinônimo de Quimiometria (FERREIRA, 2015).

A modelagem independente flexível por analogia de classes (SIMCA) é a principal técnica para o desenvolvimento de modelos supervisionados que modela a localização das classes no espaço multimensional pelo cálculo das componentes principais separadamente para cada categoria (BRERETON, 2003).

Esse método assume que os valores medidos para um grupo de amostras parecidas tenderão para uma distribuição uniforme e modelável, assim uma amostra desconhecida pode ser classificada como pertencente a uma, a mais de uma ou a nenhuma categoria predefinidas, dessa forma o SIMCA faz um tipo de embalagem para cada classe.

Figura 8: Processo de Modelagem do SIMCA



Fonte: FERREIRA, 2015

No SIMCA um modelo de componentes principais é ajustado a cada classe do conjunto de treinamento, dando origem a um classificador para cada classe de forma independente, o que evita a influência de uma sobre as outras no processo de definição das classes, essa é uma das grandes vantagens desse método. A outra vantagem é que o SIMCA é capaz de detectar *outliers*, o que significa que essa técnica é capaz de identificar amostras que não pertencem a quaisquer das classes determinadas.

Tem que ser tomados os seguintes passos no SIMCA, (BERRUETA et al., 2007):

- (i) Precisa fazer uma seleção das amostras para pertencerem ao conjunto de treinamento, de validação e de teste, que consistem de objetos de associação de classes conhecidas para as quais as variáveis são medidas;
- (ii) Uma seleção das variáveis é feita para eliminar aquelas que correspondem a ruído ou que não tenham poder discriminante;
- (iii) Constrói um modelo utilizando o conjunto de treinamento, gerando algumas categorias que se agrupam de acordo com as suas semelhanças;
- (iv) Faz uma validação do modelo utilizando um conjunto independentes de amostras para avaliar a consistência da classificação.

Existe o fator do número de PCs, onde não deve ser muito alto, pois existe um risco de fazer um modelo baseado em ruído ou em dados aleatórios, e nem muito baixo, já que pode deixar de incluir dados importantes, diminuindo o poder exploratório do conjunto de dados. (BERRUETA et al., 2007).

O ideal para a obtenção de um modelo satisfatório é o fato de as amostras serem independentes tanto para o grupo de treinamento quanto para a validação. Se isso não for satisfeito e não se tenham amostras suficientes para tal coisa, faz uma validação cruzada interna, onde o próprio conjunto de dados é dividido em dois subconjuntos: treinamento e validação (ALMEIDA, 2013).

Essa validação cruzada interna tem a característica de deixar de fora um segmento e fazer o modelo com o restante desses segmentos, constituindo o conjunto de treinamento, para validar só pegar o segmento que foi deixado de fora. Esse procedimento é repetido várias vezes. Assim ao final da validação cruzada, haverá não apenas um modelo, mas um grupo de modelos que descrevem o conjunto de dados (ENKILDE, et al., 2007). O modelo escolhido será então o que tiver o menor erro.

3.8.3 Análise Discriminante de Mínimos Quadrados Parciais - PLS-DA

A Análise Discriminante de Mínimos Quadrados Parciais (PLS-DA) é um método de classificação linear que combina as propriedades da regressão de mínimos quadrados parciais (*Partial Least Squares* - PLS) com o poder discriminante de uma técnica de classificação. O PLS-DA é baseado no algoritmo de regressão PLS, que procura por variáveis latentes com covariância máxima com as variáveis da matriz Y . (BALLABIO, CONSONNI, 2013).

Para entender a construção de um modelo PLS é preciso realizar uma regressão entre os scores das matrizes X e Y . Já que essa técnica consiste na construção de um modelo que considera as relações interna dos dados de cada matriz (X e Y , separadamente) estabelecendo um vínculo entre as duas (GELADI, 1986).

Fazendo uma ligação entre a PCA e o PLS, temos que a primeira calcula os componentes com a mínima correlação entre as variáveis originais e que maximizam a variância em X . Já a segunda procura as variáveis latentes que expliquem a máxima correlação entre as variâncias das matrizes X e Y . Assim, pode interpretar geometricamente as variáveis latentes como uma leve rotação das PCs de forma a aumentar a correlação entre os dois conjuntos de dados (WOLD, 2001).

Pode-se observar matematicamente que o PLS consiste na decomposição das matrizes X e Y , como mostrado nas equações, respectivamente:

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E} \quad (3)$$

$$\mathbf{Y} = \mathbf{UQ}^T + \mathbf{F} \quad (4)$$

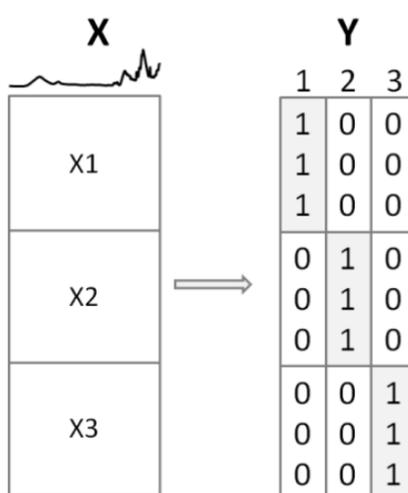
Essas equações são análogas às equações existentes na PCA, onde \mathbf{T} e \mathbf{U} são análogas às matrizes dos escores e \mathbf{P} e \mathbf{Q} , análogas às matrizes dos pesos (ver equação 2). Assim, o PLS ajusta as equações 3 e 4 de forma que \mathbf{E} e \mathbf{F} , que são os desvios de da representação dos dados originais, no novo espaço de projeção, não entram na correlação e constrói uma relação entre \mathbf{U} e \mathbf{T} (equação 5) (KALIVAS, 2006):

$$\mathbf{U} = \mathbf{TW} \quad (5)$$

Na prática quando se faz uma classificação com o método da Análise Discriminante por Mínimos Quadrados Parciais (PLS-DA), e considerado uma matriz

Y que contenha apenas 0 e 1 em seus elementos. Vamos supor agora que um sistema contenha 3 classes que pertence a matriz X como mostrado na figura 9. Essas matrizes estão concatenadas contendo os espectros referentes às amostras das classes 1, 2 e 3. A matriz Y como já dito antes terá em seus elementos 0 e 1 em que cada coluna estará relacionada com a classe do componente (1, 2 ou 3). Assim quando se fizer o modelo que for classificar as amostras o valor que pertencer a determinada classe terá o valor de 1 na matriz Y e quando não pertencer obterá o resultado 0. Normalmente utiliza-se um limiar para a classificação das amostras, onde se a amostra apresentar um valor maior ou igual que 0,5 significa que ela pertence à classe em questão.

Figura 9: Esquema da construção de um modelo PLS-DA.



Fonte: Silva, 2013.

A principal vantagem do PLS-DA é que as fontes relevantes de variabilidade de dados são modeladas pelas chamadas Variáveis Latentes (VLs), que são combinações lineares das variáveis originais, e, conseqüentemente, permite a visualização gráfica e compreensão dos diferentes padrões de dados e relações por escores e loadings dessas variáveis originais. Os loadings são os coeficientes das variáveis nas combinações lineares que determinam as variáveis latentes e, portanto, podem ser interpretadas como a influência de cada variável em cada VL, enquanto os escores representam as coordenadas das amostras no hiperespaço da projeção da VL (BALLABIO, CONSONNI, 2013).

4 METODOLOGIA

4.1 Local do experimento

Este trabalho foi realizado no Laboratório Avançado de Tecnologia Química (LATECQ) localizada na Embrapa Algodão, em Campina Grande, PB.

4.2 Amostras

As amostras analisadas neste trabalho foram sementes de cultivares convencionais e transgênicas de algodoeiro (*Gossypium* L.), cedidas pela Embrapa Algodão, localizada na cidade de Campina Grande-PB. A espécie transgênica foi denominada por BRS 368 RF e a espécie convencional foi denominada BRS Aroeira.

Após a aquisição das amostras foi realizada a limpeza das mesmas, a fim de separar a pluma das sementes e retirar toda as sujidades, como folhas secas ou areia. Para que não restasse nenhum resíduo de pluma na sua superfície das sementes, foi feito o deslinteramento que caracteriza-se na eliminação total ou parcial do línter, que são os restos de pluma presente na superfície da semente, pode ser feito através do emprego dos processos: mecânicos, químicos e por flambagem (VIEIRA, BELTRÃO, 1999). O que foi utilizado neste trabalho foi o deslinteramento por um processo químico.

As amostras foram condicionadas a uma temperatura de 20°C e umidade de 64% aproximadamente, por pelo menos uma hora antes da aquisição dos espectros e das imagens.

O aspecto das duas cultivares, convencional e transgênica, são semelhantes, sendo impossível distingui-las visualmente, como mostra a Figura 10.



Fonte: Própria, 2019.

4.3 Instrumentação

4.3.1 Espectroscopia NIR

As medidas espectrais de reflectância difusa na região do visível e infravermelho próximo foram adquiridas por meio do espectrômetro VIS/NIR modelo XDS Analyser (Foss Analytical, Hogans, Sweden) equipado com uma célula de quartzo de 3cm de diâmetro. Cada espectro foi adquirido como a média de 32 varreduras na faixa de 400 a 2500 nm, com intervalos de 0,5 nm, como ilustrado na Figura 11 a seguir.

Figura 11: Espectrofotômetro XDS near-infrared Rapid Content TM Analyser da FOSS Analytical



Fonte: Própria, 2019.

Foram utilizadas 1195 sementes, divididas em 40 sub-amostras, cada sub-amostra continha aproximadamente 30 sementes que foram lidas individualmente. Dessa forma a matriz completa de dados das amostras convencionais e transgênicas ficou com as dimensões 2390 x 4200.

4.3.2 Imagens Hiperespectrais no NIR - HSINIR

A obtenção das imagens hiperespectrais na região do NIR, faixa espectral de 900 a 2500 nm, das sementes de algodão convencionais e transgênicos foi realizada utilizando-se uma câmera espectral sisuChema SWIR -Specim® (Spectral Imaging Ltd., Oulu, Finland), lente de 50 mm, resolução espectral de 6,25 nm, resolução espacial de 150 μ m, como está mostrando a Figura 12.

Figura 12: Espectrômetro sisuChema SWIR - Specim®, responsável pela obtenção das imagens hiperespectrais



Fonte: Própria, 2019.

Para a aquisição das HSI-NIR, as amostras foram depositadas sobre a bandeja Figura 13 b e transportadas para dentro do equipamento. A velocidade do transportador foi fixada em 15.60 mm/s para assegurar a mesma forma espacial do isolado na imagem. A aquisição das imagens hiperespectrais inclui a digitalização da matriz linear pelo detector ao longo do eixo Y, o movimento da bandeja sobre o eixo X, e os espectros NIR no eixo Z da amostra. A HSI-NIR capturada constitui um hipercubo 3D contendo informações espaciais e espectrais sobre a amostra que serão utilizadas na identificação das sementes das cultivares de algodões.

4.3.3 Softwares

O tratamento das imagens foi realizado utilizando o programa Hypertools® e a construção dos modelos PLS-DA foi desenvolvida utilizando o PLS toolbox®, ambos em ambiente Matlab®, sendo Matlab® R2010a e Matlab® 2015, respectivamente.

4.4 Construção dos modelos de reconhecimento de padrões

4.4.1 Modelo PLS-DA baseado em espectros NIR

Foram analisados inicialmente a partir de uma PCA os 2390 espectros brutos de todas as amostras, a fim de verificar o seu comportamento em relação a formação de agrupamentos distintos relativos as duas classes (convencional e transgênica). Em seguida utilizou-se o algoritmo SP-Xy (do inglês: Sample set Partitioning based on joint X-y distances) (GALVÃO et al, 2005) para selecionar as amostras de treinamento (1590 amostras, sendo 871 convencionais e 719 transgênicas) e teste (800 amostras, sendo 324 convencionais e 476 transgênicas) para a construção e validação do modelo de reconhecimento de padrão supervisionado, respectivamente. Em seguida, a fim de remover ou minimizar características espectrais indesejadas, alguns pré-processamentos espectrais foram aplicados. Dentre eles, a correção de linha de base, correção do espalhamento multiplicativo MSC, variação normal padrão SNV e 1° derivada, utilizando suavização com polinômio Savitzky Golay de segunda ordem e janela de 15 pontos.

4.4.2 Modelo PLS-DA baseado em imagens hiperespectrais

As imagens hiperespectrais das sementes convencionais (a) e transgênicas (b) foram adquiridas em bandejas contendo 20 sementes de cada classe (Figuras 23 a e b). Com o objetivo de facilitar o tratamento de dados, as bordas da imagem média (Figura 23 a e b) foram reduzidas (Figura 23 c e d) por representarem informação relacionada apenas a bandeja. Em seguida, foi selecionada a região de interesse (do inglês Region of Interest: ROI), que é aquela relacionada aos pixels referentes a amostra, com base nos escores obtidos da análise por componentes principais e no gráfico dos histogramas de frequência dos pixels.

Após este tratamento prévio, as imagens hiperespectrais que estavam na forma de um hiper cubo (X,Y,λ) , foram desdobradas em uma matriz de duas dimensões $\{(X \times Y),\lambda\}$, como descrito no item 3.7 deste trabalho, para que fosse construído o modelo de reconhecimento de padrões supervisionados. Os dados espectrais foram pré-processados antes da construção do modelo PLS-DA afim de

eliminar informações não relacionadas a amostra como deslocamento de linha de base, espalhamento de sinal, ruído espectral, entre outros.

Para a construção do modelo PLS-DA, foram utilizados como dados de entrada a matriz X, contendo os dados das imagens das sementes convencionais e transgênicas, um vetor de atributo Y contendo informações sobre as classes as quais pertenciam as sementes. A validação do modelo foi feita pelo método de validação cruzada com blocos contínuos, utilizando 20 o número máximo de variáveis latentes e 19 divisões de dados. Nesta etapa foram avaliados a sensibilidade, seletividade, r^2 e poder de distinção entre as classes.

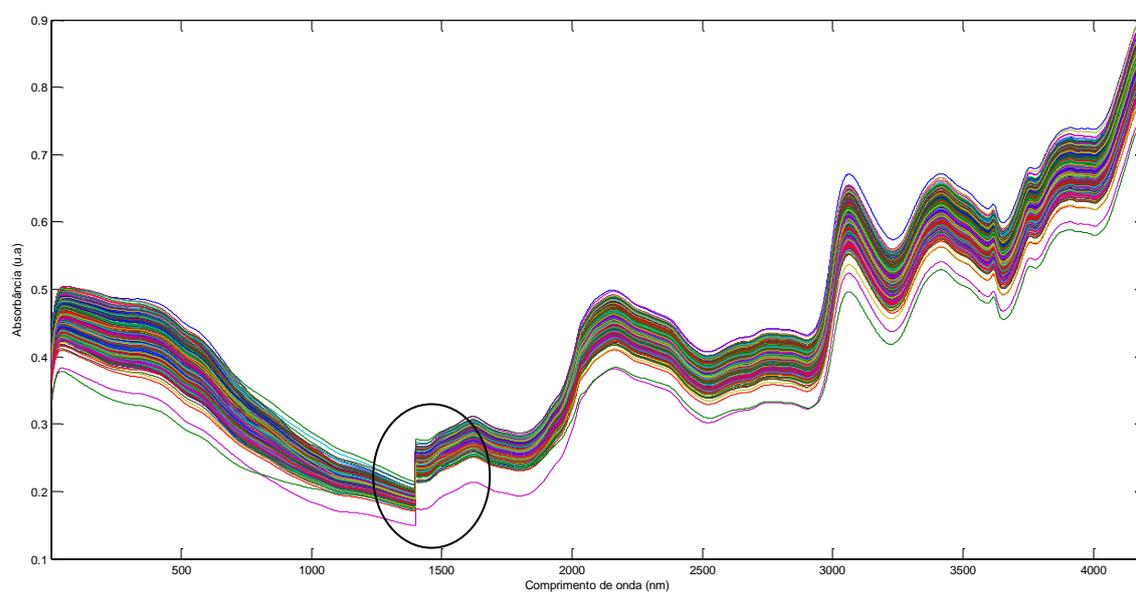
Após a validação, o modelo PLS-DA construído foi utilizado para prever a classe de 30 sementes distribuídas de forma aleatória na bandeja, as quais tiveram suas imagens HSI-NIR registradas. As bordas da imagem média foram reduzidas por representarem informação relacionada apenas a bandeja. As imagens das amostras de predição passaram pelas mesmas etapas que as imagens de treinamento, exceto o desdobramento da imagem.

5 RESULTADOS E DISCUSSÕES

5.1 Modelos baseados na espectroscopia NIR

A Figura 13 mostra os espectros brutos, na região de 400 a 4200 nm, de todas as 2390 sementes analisadas. Pode-se notar que na região de 1400 nm (destacado na figura 13) existe uma discrepância que pode ser atribuída a mudança de detector no espectrômetro utilizado. De acordo com os estudos realizados, esta discrepância não afeta os modelos construídos.

Figura 13: Espectros brutos das sementes convencionais e transgênicas

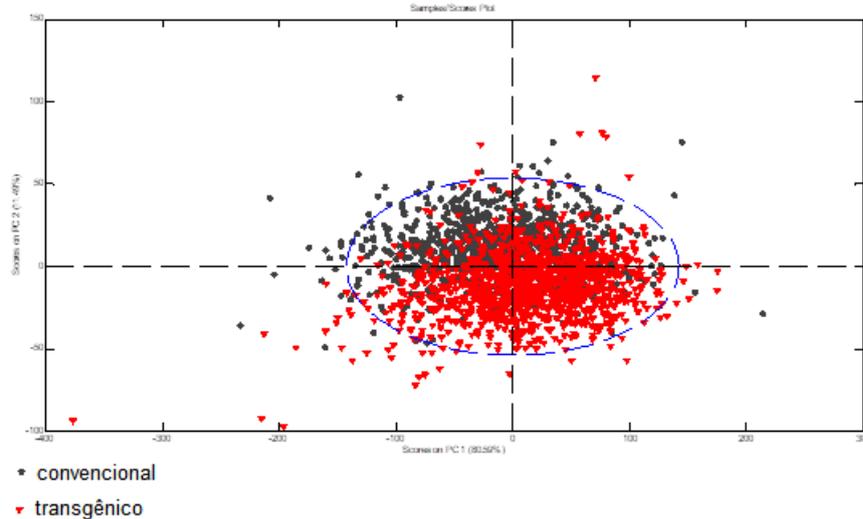


Fonte: algoritmo PLS Toolbox

Com o objetivo verificar se havia tendência de separação entre as classes convencional (pretas) e transgênica (vermelhas) foi realizada uma análise exploratória dos dados a partir da análise em componentes principais utilizando a matriz contendo todos os espectros brutos, apresentada na Figura 14. Nessa análise inicial não foram excluídas amostras com alto T^2 de Hotelling ou alto Q residual, devido a seu caráter investigativo. Não é possível notar a formação clara dos dois agrupamentos esperados, mas sim uma leve tendência de separação onde a maioria das amostras convencionais (pretas) ficaram localizadas nos escores positivos de PC2 e a maioria das amostras transgênicas (vermelhas) tenderam a se

localizar nos escores negativos de PC2, de acordo com os gráfico dos escores apresentado na figura 14. O mesmo comportamento foi apresentado para os dados pré-processados.

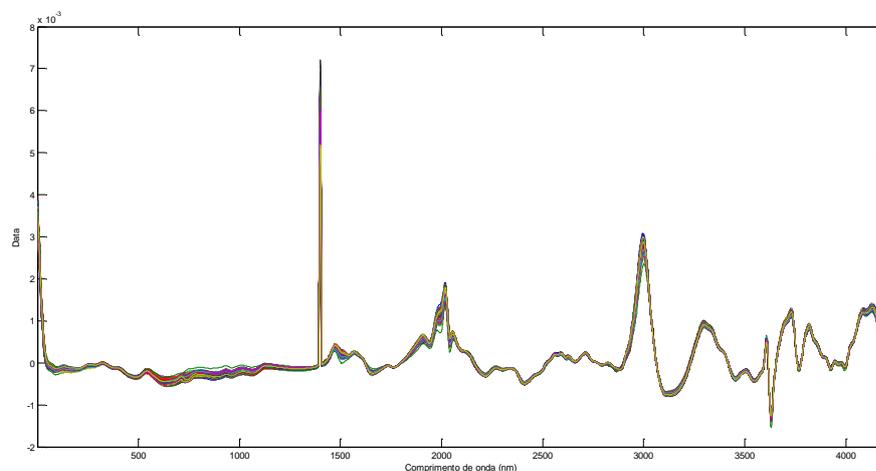
Figura 14: PCA das amostras convencionais e transgênicas



Fonte: Programa PLS Toolbox

Na tentativa de melhorar a separação entre as classes alguns pré-processamentos foram testados e o que apresentou melhor desempenho na construção do modelo PLS-DA foi a aplicação da 1ª derivada com suavização utilizando o polinômio Savitzky-Golay de 2ª ordem e janela de 15 pontos (Figura 15). Na construção de todos os modelos foi aplicado o auto escalonamento dos dados após a aplicação de cada pré-processamento.

Figura 15: Espectros com o pré-processamento de Suavização com polinômio Savitzky-Golay de 2ª ordem e janela de 15 pontos.



Fonte: algoritmo PLS Toolbox

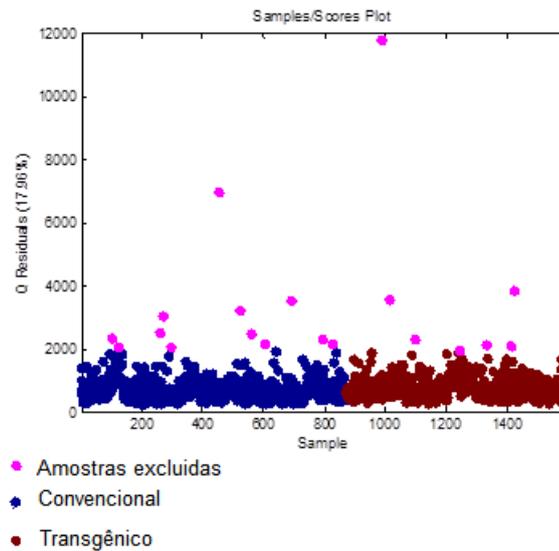
O comprimento de onda de maior intensidade, destacado na figura 16, é um spike, mesmo assim não mostrou influência no modelo, já que a performance do mesmo se mantêm, independente da exclusão do mesmo.

Previamente a construção dos modelos PLS-DA, foi aplicado o algoritmo SPXy (GALVÃO et al, 2005) para que fossem selecionadas as amostras do conjunto de treinamento de predição como mencionado na metodologia (item 4.4.1). Para isso os espectros de todas as amostras foram utilizados como matriz X (2390 linhas por 4200 colunas) e um vetor de classes 0 para as amostras convencionais e 1 para as transgênicas foram utilizadas como vetor y (2390 linhas por 1 coluna).

Em seguida, foi realizado um estudo para de detectar e excluir amostras com comportamento anômalo no conjunto de treinamento. A presença desses outliers (amostras anômalas) devem ser analisados para entender o motivo pelo qual diferem do restante dos dados. Após essa avaliação, é importante decidir se a necessidade de os excluir ou deixá-los (BALLABIO, CONSONNI, 2013). Esses outliers podem ser avaliados por meio de vários índices como Hotelling T^2 e Q Residual.

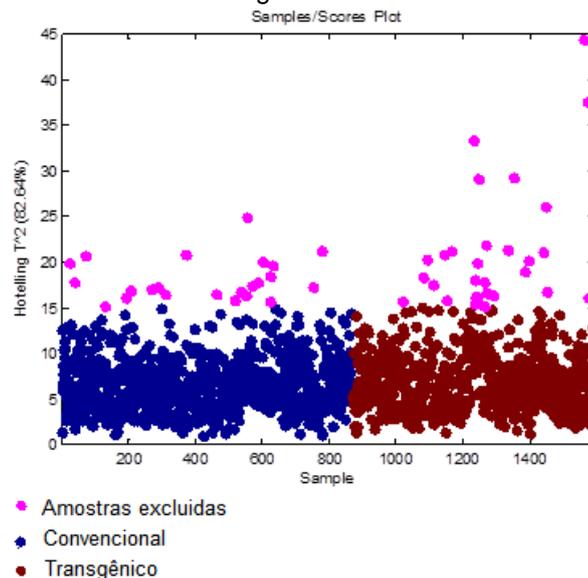
O primeiro parâmetro avaliado foi o Q residual das amostras de treinamento mostrado Figura 16, onde as amostras convencionais estão destacadas em azul escuro e as transgênicas em marrom. A estatística Q é usada para avaliar quão bem cada amostra está em conformidade com o modelo. É uma medida da diferença entre os dados originais e os dados reconstruídos com base no modelo calibrado. Os valores de Q estão associados a cada amostra e os grandes valores de Q indicam amostras que têm grandes resíduos fora do modelo (BALLABIO, CONSONNI, 2013).

De acordo com a Figura 16, foi escolhido um limiar de 2000 para aceitação das amostras. Deste modo, as amostras com Q residual superior a este limiar (marcadas em rosa) foram excluídas da etapa de treinamento. Foram excluídas 19 amostras nesta etapa.

Figura 16: Q residual versus amostras de treinamento

Fonte: algoritmo PLS Toolbox

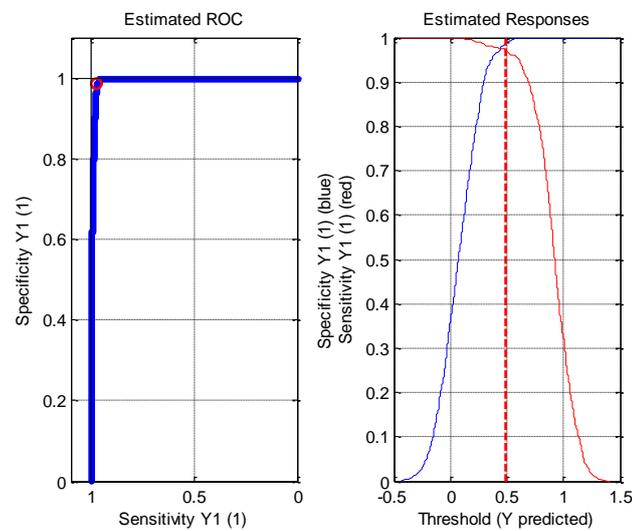
Em seguida avaliou-se o T^2 de Hotelling, que é baseada na soma dos escores quadrados normalizados, é uma medida da variação em cada amostra dentro do modelo. Os altos valores de T^2 de Hotelling indicam uma amostra com forte influência no modelo PLS-DA. É importante então evitar a presença de amostras com influência muito grande, ou seja, grandes valores significativos de T^2 de Hotelling (BALLABIO, CONSONNI, 2013). Neste caso foi escolhido um limiar de 15. Nesta etapa foram excluídas 51 amostras, como mostra a Figura 17.

Figura 17: T^2 de Hottelling versus amostras de treinamento

Fonte: algoritmo PLS Toolbox

O modelo que apresentou melhor desempenho na distinção das duas classes foi aquele construído com os dados pré-processados com 1ª derivada e suavização utilizando o polinômio Savitzky Golay de 2ª ordem e janela de 15 pontos. O modelo PLS-DA foi construído com 7 variáveis latentes, as quais explica 82,02% da variância dos dados. A etapa de treinamento apresentou uma sensibilidade de 0,972 e uma especificidade de 0,987, mostrando um alto poder de distinção entre as duas classes modeladas a um limiar de aproximadamente 0.5, como mostra a Figura 18, a partir da curva ROC e das respostas estimadas, respectivamente.

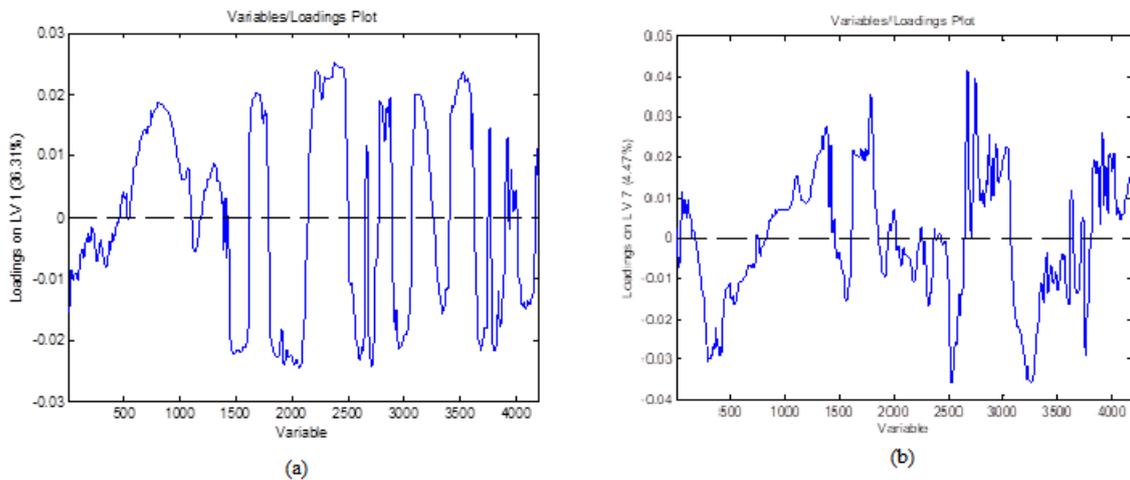
Figura 18: Curva ROC e respostas estimadas da etapa de treinamento



Fonte: algoritmo PLS Toolbox

De acordo com a Figura 19 dos Loadings nas variáveis latentes 1 e 7, a e b respectivamente, pode-se verificar que a maioria das variáveis espectrais tem contribuição para a separação das duas classes.

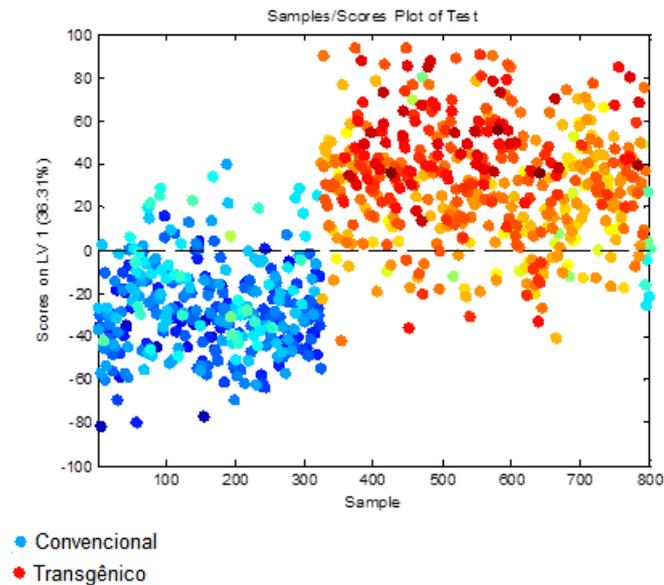
Figura 19: Loadings nas Variáveis Latentes 1 (a) e 7 (b) versus Variáveis



Fonte: algoritmo PLS Toolbox

A figura 20 mostra os escores para as amostras de treinamento, as amostras com valores de escores negativos (tons de azul) apresentam comportamento característico da classe convencional, enquanto as amostras com valores de escores positivos (tons de vermelho) apresentam comportamento característico da classe transgênica.

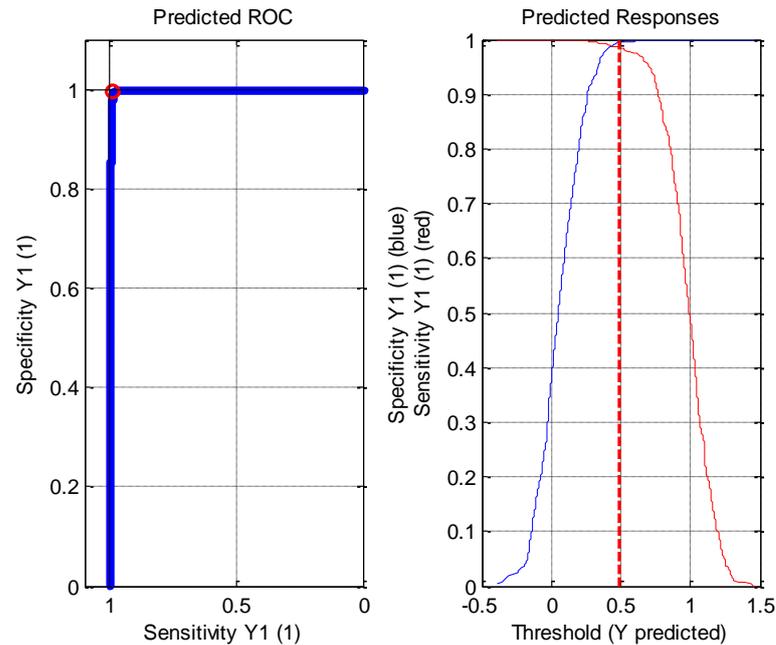
Figura 20: Escores para as amostras de treinamento



Fonte: algoritmo PLS Toolbox

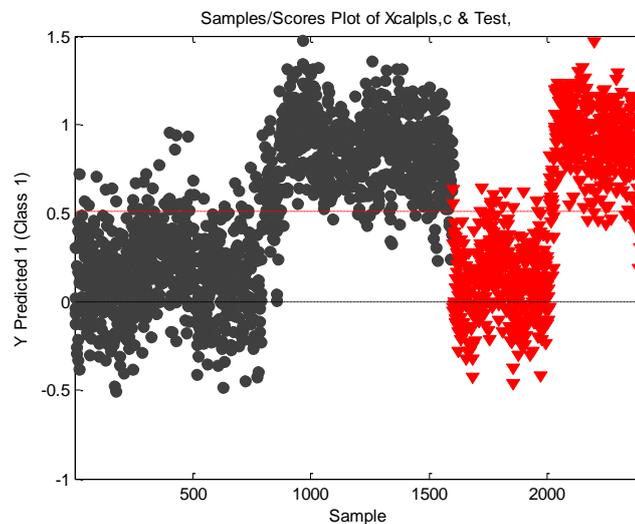
Nele pode-se notar uma boa separação entre as classes. No entanto, algumas amostras nomeadas como classe transgênica apresentaram comportamento similar aos das amostras convencionais (valores de escores negativo). A porcentagem de classificação incorreta na etapa de treinamento foi de 2,05%.

Já a Figura 21 mostra a curva ROC e as respostas previstas com o limiar de separação entre as classes, a e b, respectivamente. Nele pode-se observar que foram alcançados valores de sensibilidade e seletividade de 0.987 e 0.997, respectivamente, a um limiar de 0.5.

Figura 21: Curva ROC e respostas estimadas na etapa de teste

Fonte: algoritmo PLS Toolbox

O gráfico 22 mostra os escores para as amostras de teste. Nela pode-se observar uma clara distinção entre as classes convencional e transgênica a um limiar de 0.5. A porcentagem de erro na etapa de predição foi de 0,78%. As amostras que estão com a coloração preta são as amostras de treinamento e as amostras que estão em vermelho são as amostras do grupo de teste. Abaixo da limiar 0,5 são as amostras convencionais e acima de 0,5 são as amostras transgênicas.

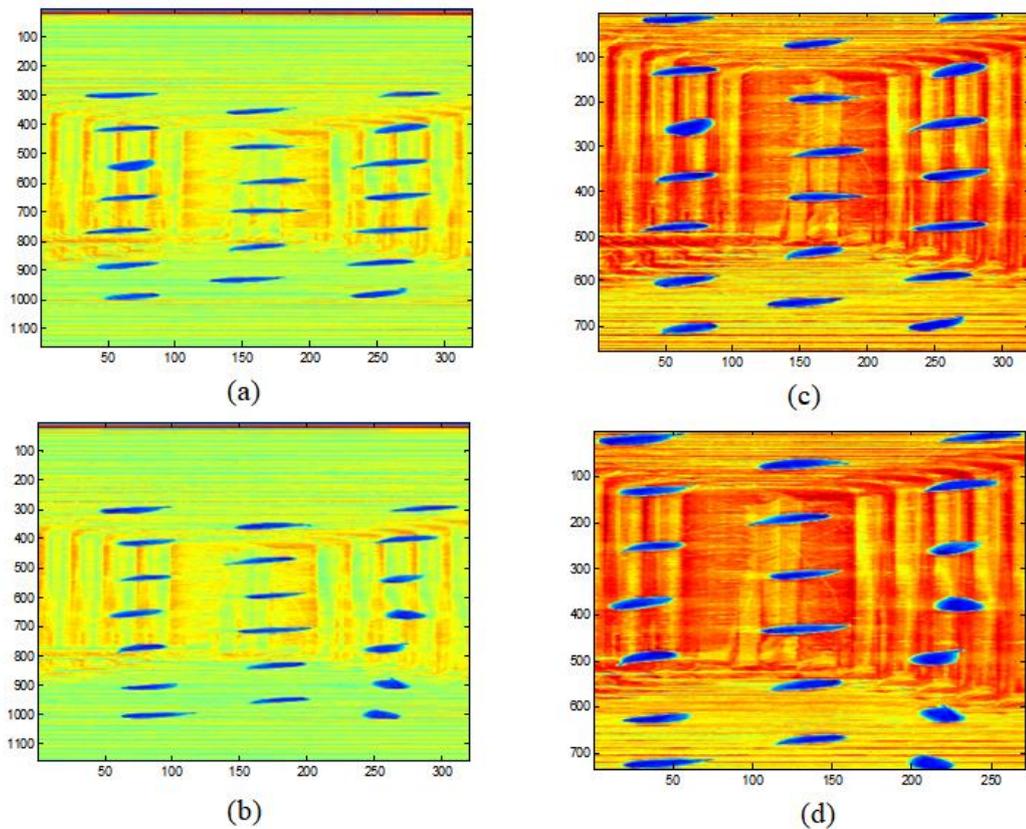
Figura 22: Escores das amostras de teste

Fonte: algoritmo PLS Toolbox

5.2 Modelo PLS-DA baseado nas imagens hiperespectrais

Com o objetivo de facilitar o tratamento das imagens e melhor a capacidade preditiva dos modelos de reconhecimento de padrões, um tratamento prévio das imagens foi realizado. A figura 23 mostra a imagem média original (a) e a com as bordas reduzidas (b). As bordas da imagem foram reduzidas por representarem informação relacionada apenas a bandeja.

Figura 23: Imagens médias convencionais (a) e transgênicas (b) e imagens com redução de dimensão convencionais (c) e transgênicas (d)

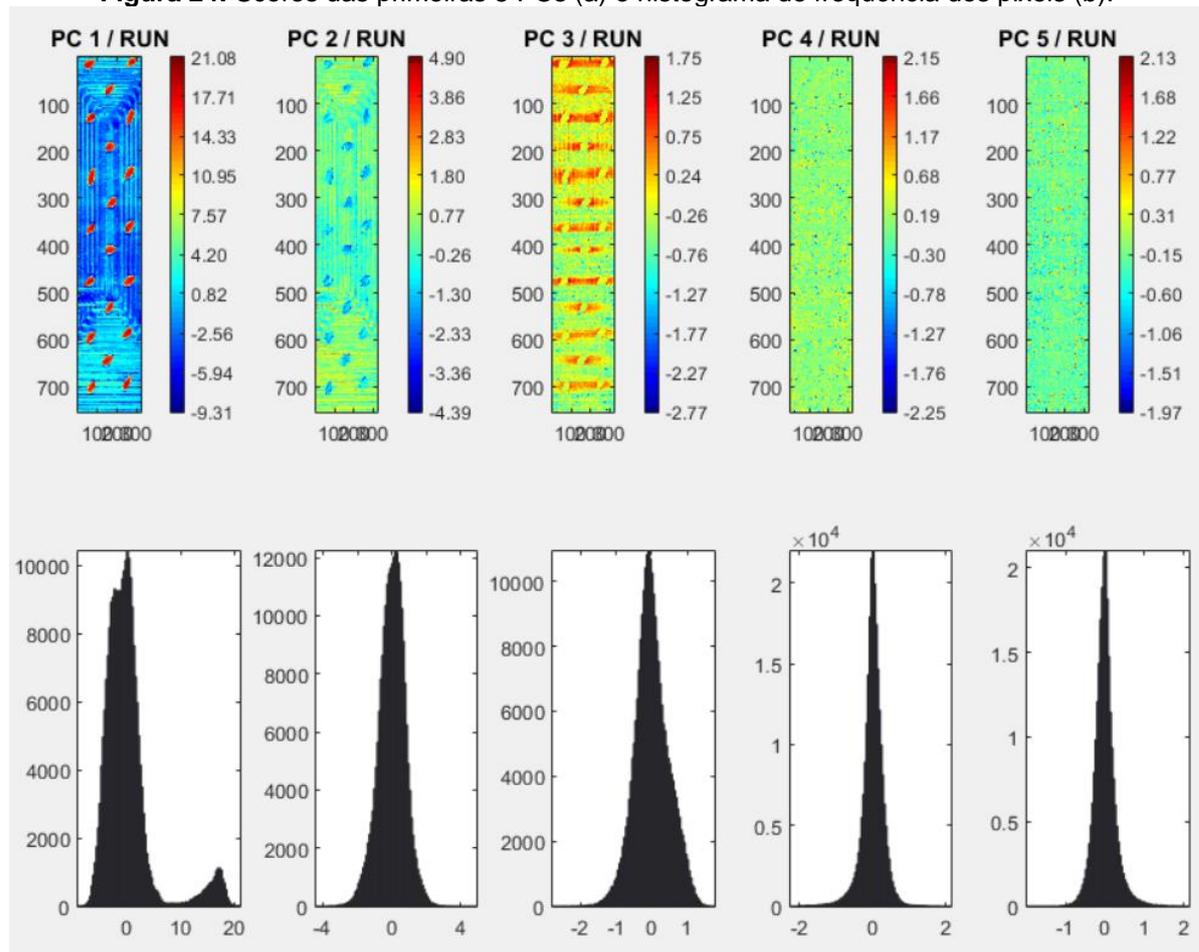


Fonte: Algoritmo do Hypertools®

Em seguida, a partir das imagens representadas na figura 23 c e d, foi selecionada a região de interesse (do inglês Region of Interest: ROI).

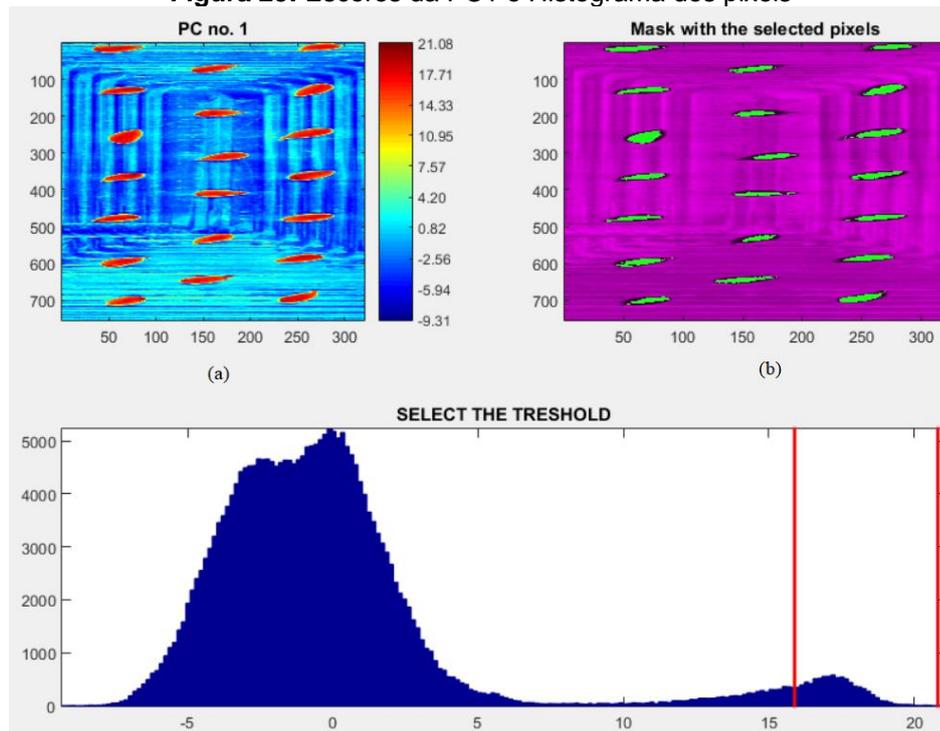
A figura 24a mostra os escores das primeiras 5 componentes principais (PC), nela pode-se observar que a imagem dos escores de PC1 é a que apresenta uma maior distinção entre os pixels das sementes (vermelho) e os da bandeja (azul ciano). Com base no gráfico dos histogramas de frequência dos pixels para PC1 (Figura 24b) pode-se escolher o limiar de separação entre os pixels referentes as sementes e aqueles referentes a bandeja.

Figura 24: Scores das primeiras 5 PCs (a) e histograma de frequência dos pixels (b).



Fonte: Algoritmo do Hypertools®

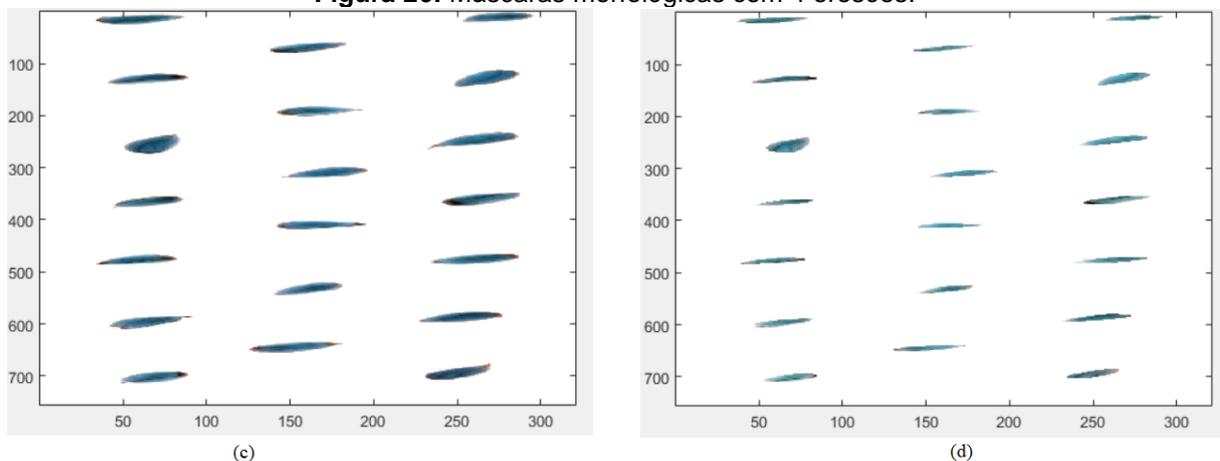
A Figura 25 mostra a imagem dos escores de PC1 e o gráfico de histogramas dos pixels onde foi escolhido o limiar de separação (16 até aproximadamente 21) usado para a criação da máscara. Neste gráfico pode-se notar duas curvas sendo a que possui valores mais altos relativa aos pixels referente a bandeja e a de menor intensidade relativa aos pixels da semente. A fim diminuir/eliminar os pixels referentes a bandeja foi desenvolvida uma máscara, deste modo selecionou-se a região do histograma de frequência dos pixels referente apenas aos pixels da semente. Diversas regiões na faixa de 10 a 21 foram testadas como limiar, porém a região de 16 a 21 foi selecionada a fim de evitar o efeito dos pixels que se encontravam na borda da semente.

Figura 25: Escores da PC1 e Histograma dos pixels

Fonte: Algoritmo do Hypertools®

Mesmo após a escolha da ROI, ainda restaram pixels relativos à bandeja, para eliminá-los, aplicou-se uma máscara morfológica onde foi feito um processo de erosão a fim de eliminar pixels espúrios.

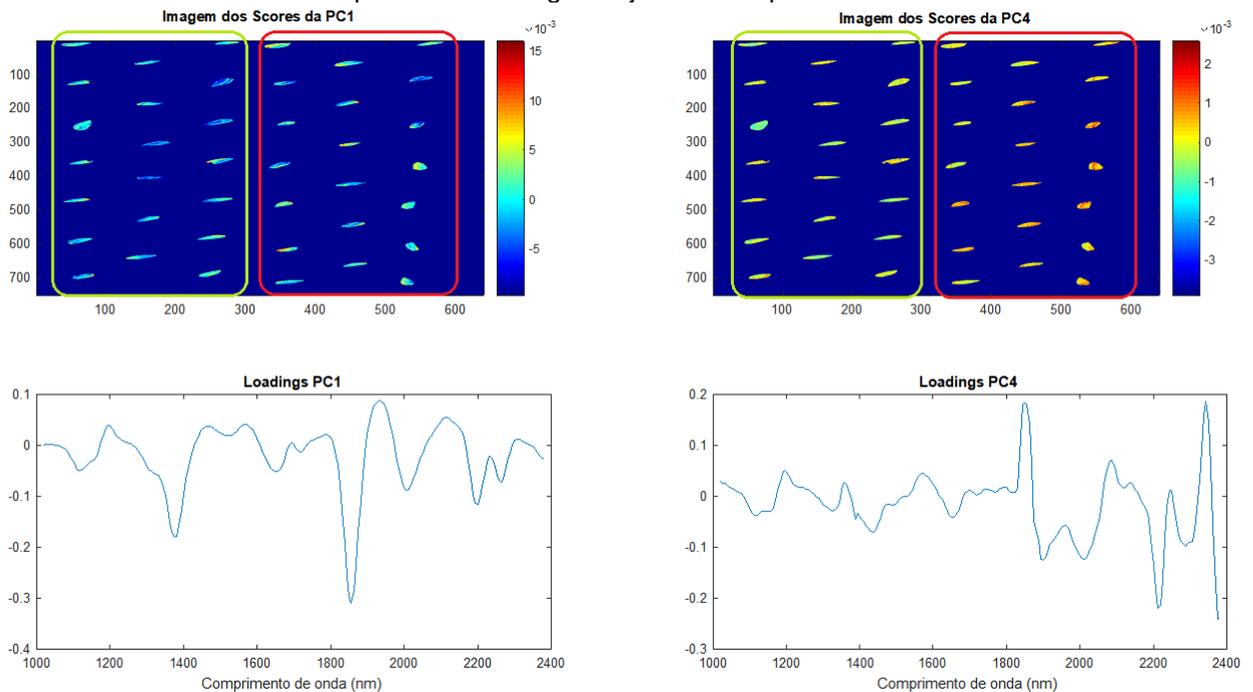
A figura 26 mostra a imagem antes da erosão (c) e após a erosão (d). Esse procedimento foi realizado para ambas as amostras, transgênicas e convencionais.

Figura 26: Máscaras morfológicas com 4 erosões.

Fonte: Algoritmo do Hypertools®

Previamente a construção do modelo de reconhecimento de padrões supervisionado foi realizada uma PCA utilizando diversos pré-processamentos dos dados espectrais a fim de verificar qual destes promoveria a melhor separação entre as classes. Foram feitas PCAs tanto com os dados brutos quanto com os pré-processamentos como MSC + Suavização com polinômio de 2º grau e janela de 5 pontos, Primeira Derivada de Savitzky-Golay com polinômio de 2º grau e janela de 7 pontos e Segunda Derivada de Savitzky-Golay com polinômio de 2º grau e janela de 7 pontos. O que obteve uma melhor separação foi os dados pré-processados com Primeira Derivada de Savitzky-Golay com polinômio de 2º grau e janela de 7 pontos, os resultados podem ser conferidos na Figura 27, que traz os resultados da PC1 e PC4 com seus respectivos loadings, às PC2 e PC3 não foram inseridas, pois davam resultados parecidos com a PC1.

Figura 27: PCA das amostras com o pré-processamento 1ª Derivada de Savitzky-Golay com polinômio de 2º grau e janela de 7 pontos

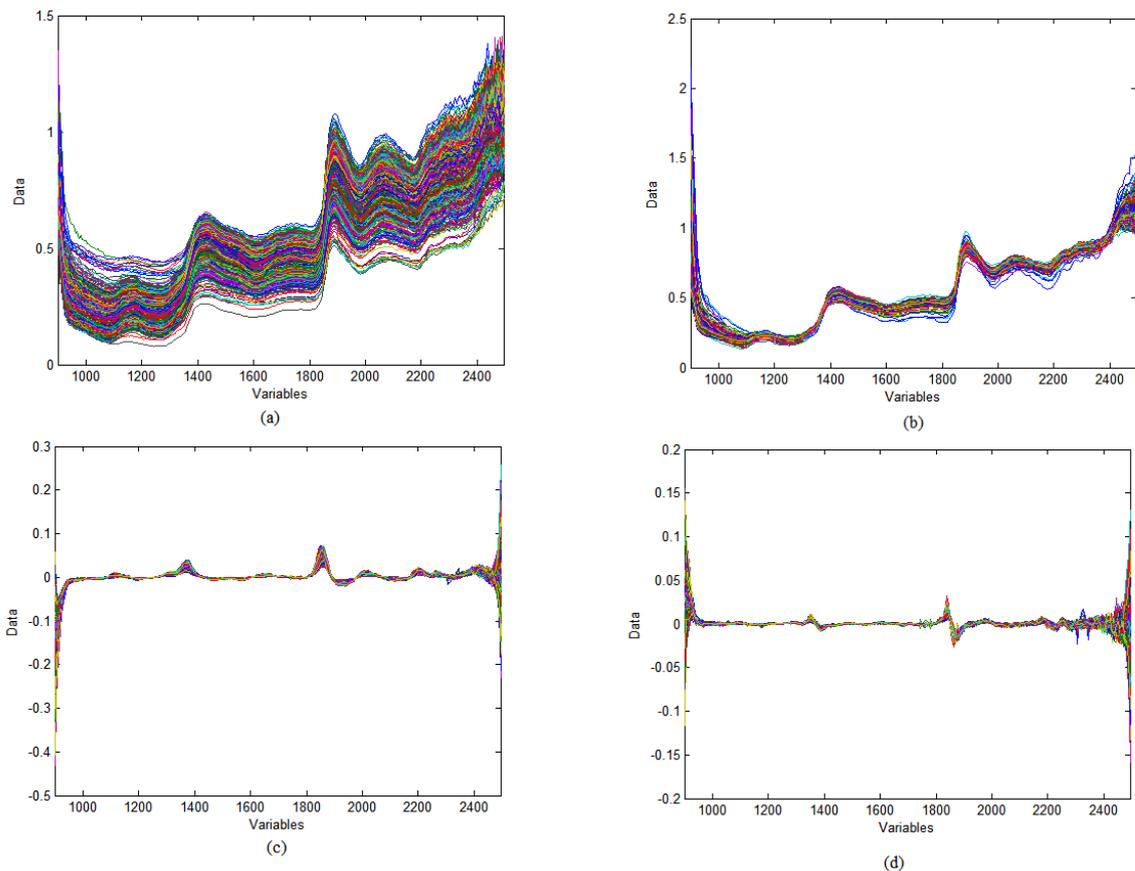


Fonte: Algoritmo do Hypertools®

Aparentemente a PC1 não nos mostra uma tendência de separação já que as amostras estão todas em tons de azul. Já a PC4 mostra tonalidades distintas para as amostras convencionais (tonalidades de ciano) e transgênica (tonalidades de vermelho). A região que está evidenciada em verde são as amostras convencionais e a região em vermelho são as amostras transgênicas (Figura 27).

Ainda na análise exploratória foi analisado também os espectros das amostras, na Figura 28 (a) são apresentados os espectros brutos, neles não é possível perceber diferença entre os espectros das amostras transgênicas e convencionais. Também é possível observar um espalhamento do sinal espectral devido à natureza das amostras. Dessa forma foi necessário utilizar técnicas de pré-processamento, visando reduzir a porção do sinal analítico que não está relacionada as amostras. Os pré-processamentos utilizados foram MSC + Suavização com polinômio de 2° grau e janela de 5 pontos (b), Primeira Derivada de Savitzky-Golay com polinômio de 2° grau e janela de 7 pontos (c) e Segunda Derivada de Savitzky-Golay com polinômio de 2° grau e janela de 7 pontos (d), como mostrado na Figura 28.

Figura 28: Espectros brutos (a) e pré-processados (b,c,d).



Fonte: Algoritmo do PLS Toolbox

Os principais objetivos dessas abordagens são melhorar as análises exploratórias subsequentes, melhorar os modelos de classificação ou previsão entre outros (MARTENS, NIELSEN, ENGELSEN, 2003; FEMENIAS et al, 2020).

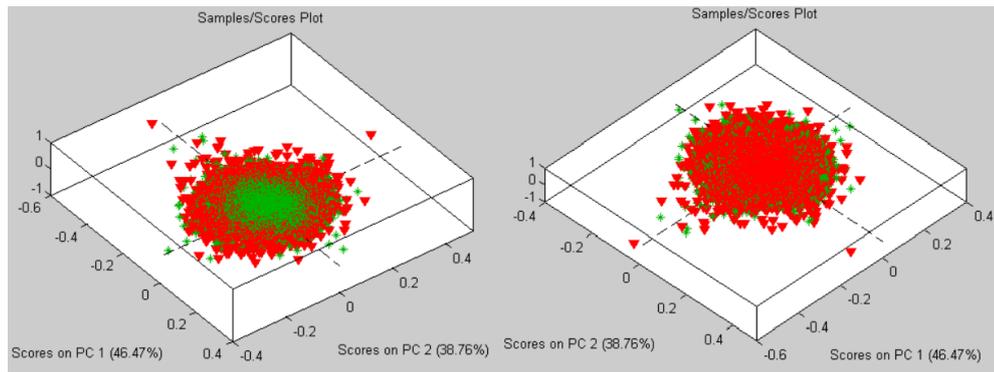
Também foi necessário retirar a região espectral abaixo de 1000 nm e acima de 2400 nm por se apresentarem bastante ruidosa.

Como o pré-processamento 1ª Derivada de Savitzky-Golay com polinômio de 2º grau e janela de 7 pontos foi o que apresentou a melhor tendência de separação entre as classes, o mesmo foi utilizado para desenvolver os modelos de reconhecimento de padrões.

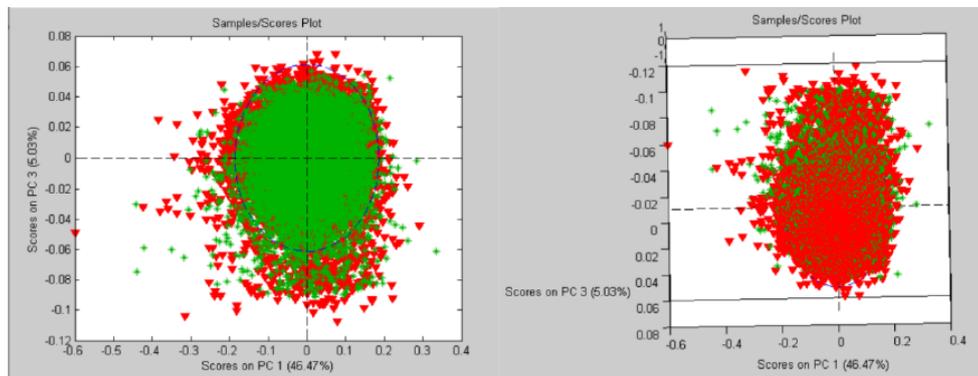
A Figura 29 (a, b, c e d), mostram os escores de PC1xPC2, PC1xPC3, PC1xPC4 e PC3xPC4, respectivamente. Nos escores de PC1xPC2 e PC3xPC4 não é possível notar distinção entre as amostras transgênicas e convencionais. Já nos escores de PC1xPC4 e PC1xPC3 pode-se observar uma tendência de separação entre as classes. Nelas, forma geral, há uma tendência de agrupamento das amostras transgênicas em cima (em verde) e convencionais abaixo (em vermelho). Também pode-se notar que as amostras transgênicas possuem menor dispersão que as convencionais. Esse comportamento pode estar associado a pleiotropia, que pode ser definido como um mecanismo genético controlador de várias características a partir de um único gene, que pode ocorrer nas espécies convencionais tornando-as mais diversas (maior variabilidade) do que as sementes transgênicas. Ou ainda aos efeitos epigenéticos, que são modificações do genoma herdadas pelas gerações descendentes, mas que não alteram a sequência do DNA. As variações epigenéticas (não-genéticas) adquiridas por um organismo pode ser frequentemente repassada às próximas gerações e depende de pequenas mudanças químicas no DNA e na proteína que o envolve, o que pode ser uma justificativa para a dispersão das amostras convencionais. (FANTAPPIE, 2013)

Neste sentido, os escores de PC1xPC4 é onde é observado a melhor separação entre as classes.

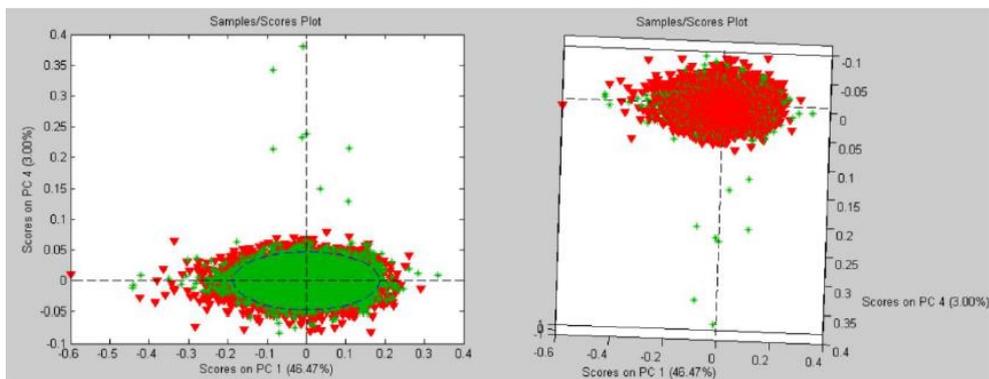
Figura 29: PCA das amostras transgênicas e convencionais da calibração.



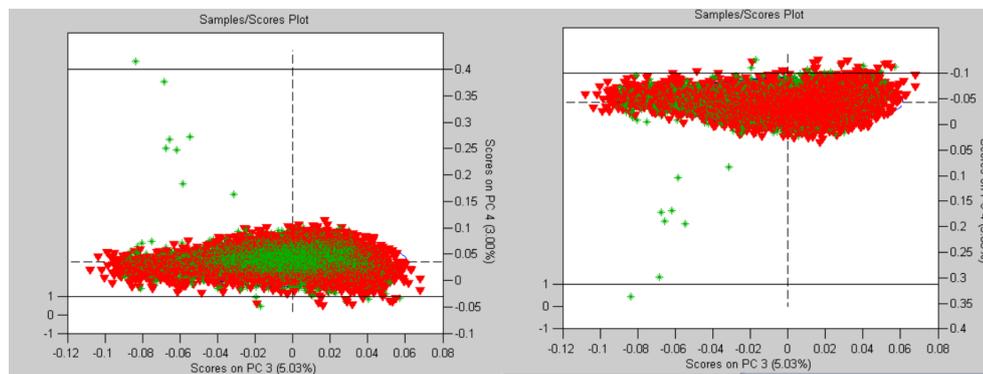
(a)



(b)



(c)



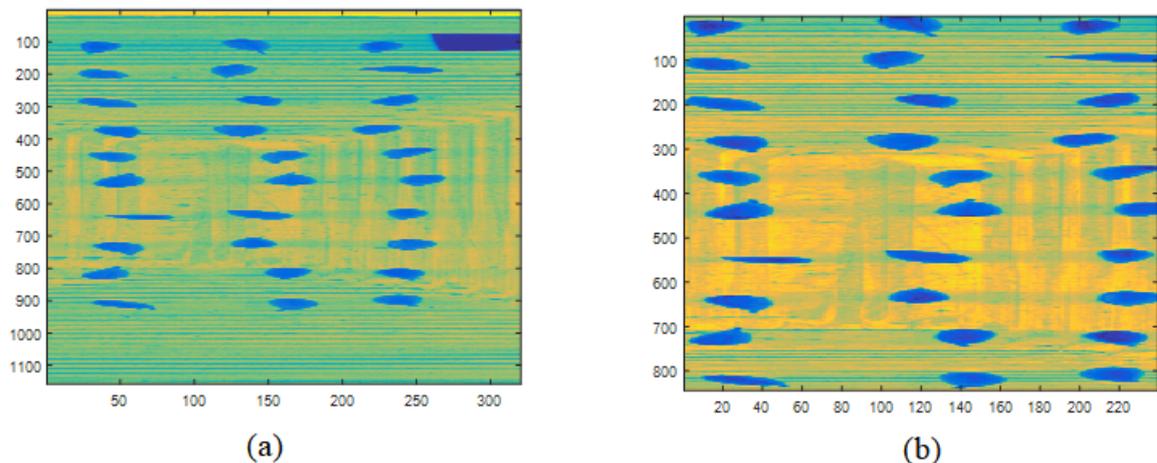
(d)

Fonte: Algoritmo do PLS Toolbox

Para a construção do modelo PLS-DA, foram utilizados como dados de entrada a matriz X, contendo os dados das imagens das sementes convencionais e transgênicas, um vetor de atributo Y contendo informações sobre as classes as quais pertenciam as sementes. A validação do modelo foi feita pelo método de validação cruzada com blocos contínuos, utilizando 20 como o número máximo de variáveis latentes e 19 divisões de dados. Nesta etapa foram avaliados a sensibilidade, seletividade, r^2 e poder de distinção entre as classes.

Após a validação, o modelo PLS-DA construído foi utilizado para prever a classe de 30 sementes distribuídas de forma aleatória na bandeja, as quais tiveram suas imagens HSI-NIR registradas. As bordas da imagem média (Figura 30 a) foram reduzidas (Figura 30b) por representarem informação relacionada apenas a bandeja.

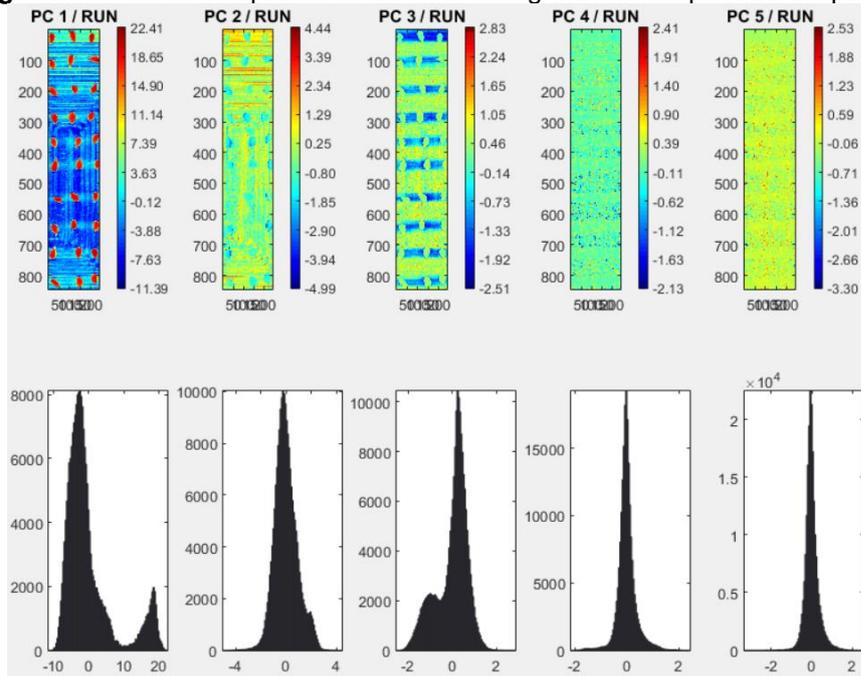
Figura 30: Imagem média inicial (a) e borda da imagem média reduzida (b).



Fonte: Algoritmo do Hypertools®

A imagem das amostras de predição passou pelas mesmas etapas que as imagens de treinamento, exceto a etapa de desdobramento da imagem (Figuras 31 e 32).

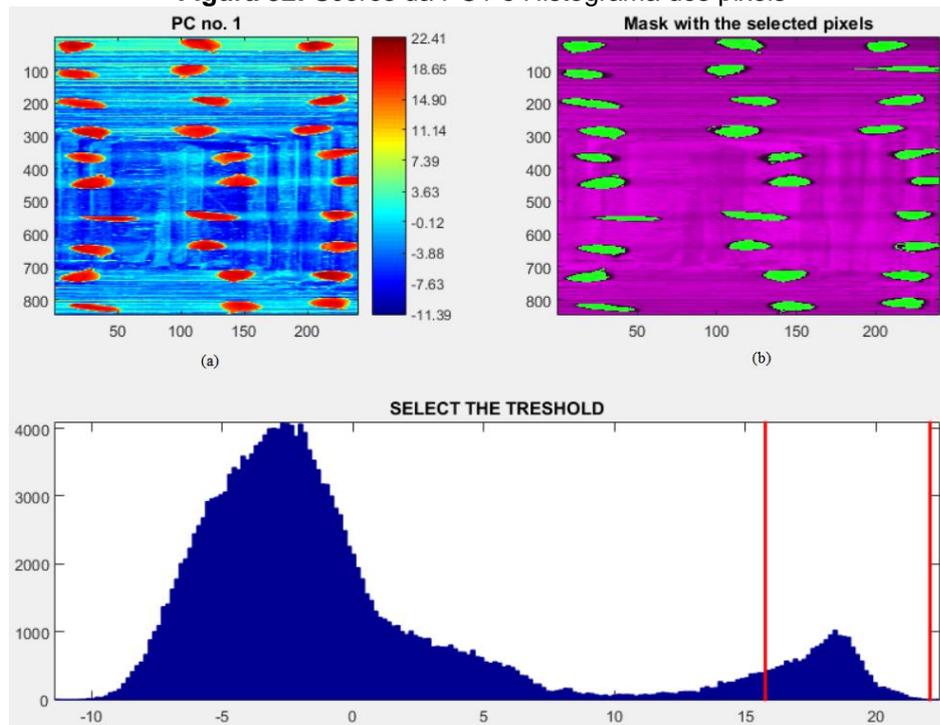
Figura 31: Scores das primeiras 5 PCs e histograma de frequência dos pixels



Fonte: Algoritmo do Hypertools®

A componente principal que apresentou a melhor separação entre os pixels das sementes e os da bandeja foi a PC1. Onde os pixels relacionados as sementes apresentam maiores valores de escores (vermelho) (Figura 32).

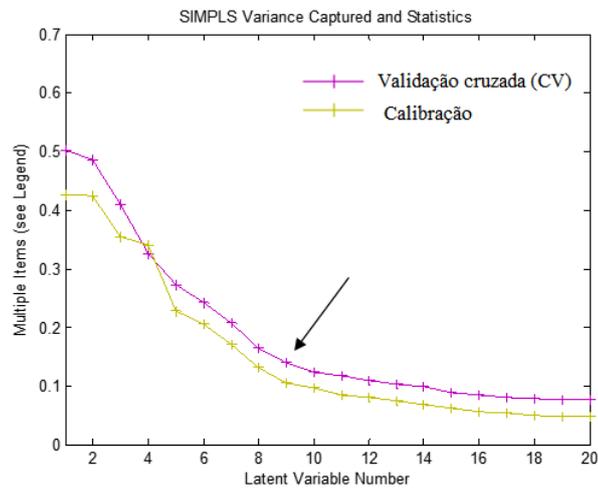
Figura 32: Scores da PC1 e Histograma dos pixels



Fonte: Algoritmo do Hypertools®

O modelo PLS-DA foi construído com 9 variáveis latentes (Figura 33) as quais explicam 99,57% da variância dos dados.

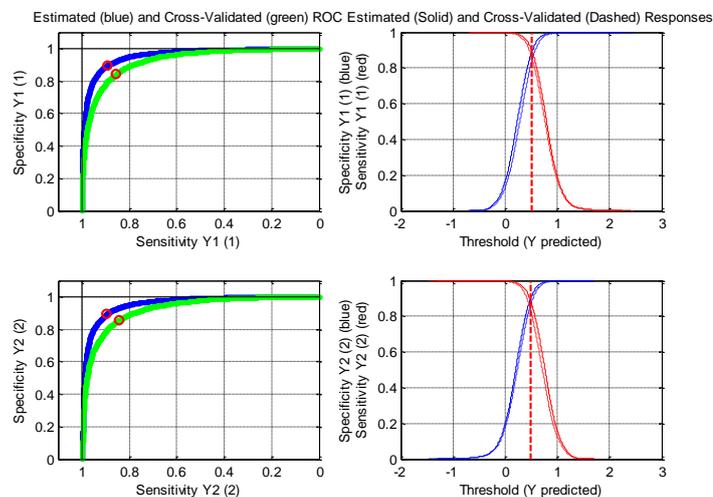
Figura 33: Número de Variáveis Latentes.



Fonte: Algoritmo do PLS Toolbox

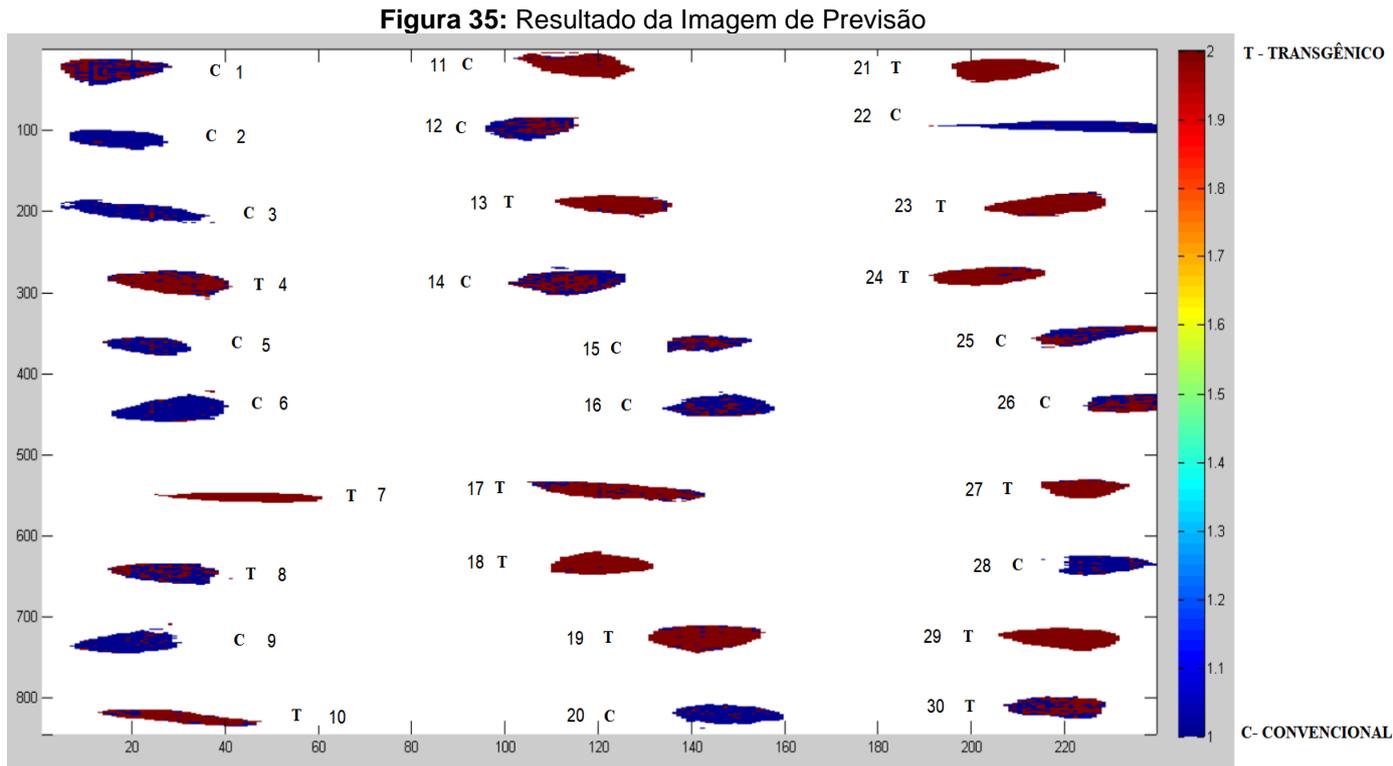
Na Figura 34 são apresentadas as curvas ROC (à direita) e as respostas estimadas (à esquerda). Nela pode-se observar que na etapa de treinamento foram alcançados valores de sensibilidade e seletividade de 0,8548 e 0,8938 a um limiar de 0,5196, respectivamente. E na etapa de validação cruzada (CV) de foram obtidos valores de sensibilidade e seletividade de 0,8653 e 0,844 a um limiar de 0,4831, respectivamente, o que demonstra uma boa performance do modelo. O limiar utilizado foi de 0,4831, já que o software utiliza sempre o limiar da validação cruzada.

Figura 34: Curvas ROC e respostas estimadas da Calibração



Fonte: Algoritmo do PLS Toolbox

Após a validação do modelo, o mesmo foi aplicado a imagem das amostras de predição. Os resultados obtidos podem ser observados na Figura 35, onde as manchas azuis se referem as amostras previstas pelo modelo como convencional e as manchas em vermelho são aquelas previstas como transgênicas pelo modelo. Ao lado de cada mancha foi colocado o rótulo C (convencional) ou T (transgênica) que se refere a classificação real das sementes.



Fonte: Matlab R.2010a

Como pode ser observado na Figura 35 e tabela 1 das 30 sementes analisadas, sendo 16 convencionais e 14 transgênicas, 9 amostras convencionais e 12 amostras transgênicas podem ser classificadas de forma correta e inequívoca em suas respectivas classes, já que a maioria dos pixels possuem altos escores (vermelho), no caso das amostras transgênicas, e baixos escores (azul), no caso das amostras convencionais. Desta forma 21 amostras foram classificadas na classe correta apresentando uma taxa de acertos de 70%. 7 amostras possuem classificação dúbia ou subjetiva (representadas por ? na tabela 1), já que não é possível concluir visualmente se possui maioria de pixels vermelhos ou azuis (representando 23,3% dos dados). 2 amostras convencionais foram classificadas incorretamente na classe transgênica (representadas por * na tabela 1), representando um falso positivo.

Tabela 1: Classificação das amostras de teste

Amostra	Classe real	Predição	Amostra	Classe real	Predição
1	C	?	16	C	C
2	C	C	17	T	T
3	C	C	18	T	T
4	T	T	19	T	T
5	C	C	20	C	C
6	C	C	21	T	T
7	T	T	22	C	C
8	T	?	23	T	T
9	C	C	24	T	T
10	T	T	25	C	?
11	C	T*	26	C	?
12	C	?	27	T	T
13	T	T	28	C	C
14	C	T*	29	T	T
15	C	?	30	T	?

A Tabela 2 mostrando os valores da sensibilidade, especificidade para a treinamento, validação cruzada e teste.

Tabela 2: Dados da sensibilidade e especificidade

	Convencional	transgênico
Sn (Cal)	0,896	0,907
Sp(Cal)	0,907	0,896
Sn(CV)	0,871	0,878
Sp(CV)	0,878	0,871
Sn(Pred)	0,420	0,975
Sp(Pred)	0,975	0,420

Sn = Sensibilidade
Sp = Especificidade
Cal = Calibração
CV = Validação Cruzada
Pred= Previsão

Fonte: Própria, 2020.

A sensibilidade é a capacidade do modelo em classificar uma amostra corretamente. Pode-se observar que na etapa de previsão foi obtido um maior valor

de sensibilidade para a classe transgênica, o que significa que a maioria dos pixels referentes as amostras transgênicas foram classificadas corretamente.

Em contra partida, os valores de especificidade para as amostras transgênicas na etapa de predição foi baixo, o que indica que o modelo consegue identificar a maioria dos pixels referentes as amostras transgênicas, mas não é tão eficiente em diferenciar as duas classes de pixels, o que pode levar a classificação incorreta das amostras.

Não foi observada a presença de falsos negativos na classificação o que é um bom indicativo, já que o modelo não apresenta a tendência de classificar uma amostra transgênica como convencional, o que poderia representar um risco de contaminação de áreas nativas e um grande prejuízo ao meio ambiente.

6 CONCLUSÕES

Os espectros na região do infravermelho das sementes de algodão convencional e transgênicas foram registrados e foi verificado que não era possível diferenciar as duas espécies com base em diferenças espectrais.

A aplicação da análise de componentes principais (PCA) dos espectros brutos, na etapa exploratória, não mostrou tendência de separação entre as classes, deste modo alguns pré-processamentos, o que apresentou a melhor tendência de separação foi a 1ª derivada com suavização utilizando o polinômio Savitzky-Golay de 2ª ordem e janela de 15 pontos.

O modelo PLS-DA-NIRS foi construído com 7 variáveis latentes, as quais explicam 82,02% da variância dos dados. A etapa de treinamento apresentou uma sensibilidade de 0,972 e uma especificidade de 0,987, mostrando um alto poder de distinção entre as duas classes modeladas a um limiar de aproximadamente 0,5. A porcentagem de erro na etapa de predição foi de 0,78% e de classificação correta na etapa de treinamento foi de 97,95%.

Em relação aos modelos PLS-DA-HSNIR, o melhor pré-processamento foi a 1ª Derivada com polinômio de 2ª ordem e janela de 7 pontos, o modelo foi construído com 9 variáveis latentes e explicou 88,57%. Na etapa de treinamento foram alcançados valores de sensibilidade e seletividade de 0,8548 e 0,8938 a um limiar de 0,5196, respectivamente. E na etapa de validação cruzada (CV) de foram obtidos valores de sensibilidade e seletividade de 0,8653 e 0,844 a um limiar de 0,4831, respectivamente. Na etapa de teste foram classificadas corretamente 70% das amostras.

No que diz respeito a comparação entre as duas metodologias podemos afirmar que a espectroscopia NIR é uma técnica menos honerosa, mais simples em relação ao tratamento de dados e que apresentou resultados bastante animadores. Já nas imagens hiperespectrais notamos que a mesma é um pouco mais complexa, mais honerosa, em contra partida fornece mais informações das amostras e de mais fácil interpretação e adequada para uma análise objetiva em relação a uma única semente.

De forma geral os modelos PLS-DA-NIR e PLS-DA-HSNIR foram capazes de fazer a distinção das amostras transgênicas das convencionais de maneira eficaz

mostrando-se uma boa ferramenta auxiliar aos métodos convencionais e adequados para uma análise prévia das amostras.

REFERÊNCIAS

ABAPA. **RELATÓRIO DE GESTÃO BIÊNIO, 2017-2018**, 2019. Disponível em: <http://abapa.com.br/wp-content/uploads/2019/02/Relat%C3%B3rio-de-Gest%C3%A3o-da-Abapa-Bi%C3%AAnio-2017-2018.pdf> Acesso em: 20 abr 2019.

ALMEIDA, F. M. **Espectroscopia de Infravermelho Próximo com Transformada de Fourier (FT-NIR) na Caracterização de Farinhas para Alimentação Pueril**. Instituto Superior Técnico - Universidade Técnica de Lisboa, 2009.

ALMEIDA, M. R.; CORREA, D. N.; ROCHA, W. F. C.; SCAFI, F. J. O.; POPPI, R. J. **Discrimination between authentic and counterfeit banknotes using Raman spectroscopy and PLS-DA with uncertainty estimation**. *Microchemical Journal*, v.109, p.170–177, 2013.

AIRES, P.S.R.; GAMBARRA-NET, F.F.; COUTINHA, W.M.; ARAÚJO, A.E.; DA SILVA, G.F.; GOUVEIA, J.P.G.; MEDEIROS, E.P. **Near infrared hyperspectral images and pattern recognition techniques used to identify etiological agents of cotton anthracnose and ramulosis**. *Spectral Imaging* 7, a8, 2018.

AMIGO, J.M.; **Practical issues of hyperspectral imaging analysis of solid dosage forms**, *Analytical and Bioanalytical Chemistry* 398, p.93-109, 2010.

ARAUJO, A. M. **Monitoramento de processos em tempo real via espectrometria no infravermelho próximo**. 2007. 92 p. Dissertação (Mestrado em Engenharia de Processos) – Universidade Tiradentes, Aracajú, 2007.

ARIANA, D.; LU, R.; GUYER, D. E. **Hyperspectral reflectance imaging for detection of bruises on pickling cucumbers**. *Computers and Electronics in Agriculture*, v. 53(1), p.60-70, 2006.

Arnold, S.A; Harvey, L.M. **Employing Near-Infrared Spectroscopic Methods of Analysis for Fermentation Monitoring and Control - Part 1, Method Development**. BioPharm International, 2002.

BARBOSA, L.C de A. **Espectroscopia no infravermelho na caracterização de compostos orgânicos**. Viçosa: UFV, 2008.

BARROSO, P.A.V; SUASSUNA, N.D; PEDROSA, M.B; MORELLO, C.L; FILHO, J.L.daS; LAMAS,F.M;BOGIANI, J.C; **BRS 368RF: A glyphosate tolerant, midseason upland cotton cultivar for Northeast and North Brazilian cerrado**. *Crop Breeding and Applied Biotechnology*, v. 17, p. 399-402, 2017

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

BECHANE, V. **Organismos geneticamente modificados**. *Jornada Científica*, v. 1, 2012.

BELLON-MAUREL, V.; DUBOIS, J.; **Near-infrared hyperspectral imaging in food and agricultural sciences**, in: R. Salzer, H.W. Siesler (Eds.), *Infrared and Raman Spectroscopic Imaging*, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, p. 259–294, 2009.

BÉLOT, J.L, BARROS, E.M, MIRANDA, J.E. **Riscos e Oportunidades: O Bicudo-do-Algodoeiro**. In: AMPA, APROSOJA-MT, EMBRAPA. *Desafios do cerrado: como sustentar a expansão da produção com produtividade e competitividade - Cuiabá: Associação Mato-grossense dos Produtores de Algodão*. Editora Casa da Árvore. Cap. 3, p, 77-118, 2016.

BENNETT, R.; MORSE, S.; ISMAEL, Y. **The economic impact of genetically modified cotton on South African smallholders: yield, profit and health effects**. *The Journal of Development Studies*, [S.l.], v. 42, n. 4, p. 662-677, 2006.

BERRUETA, L.; ALONSO-SALCES, R.; HEBERGER, K. **Supervised pattern recognition in food analysis**. *Journal of Chromatography*, v.1158, n.1-2, p.196-214,2007

BERTHEAU, Y. et al. **Detection methods and performance criteria for genetically modified organisms**. *Journal of AOAC International*, v.85, p.801-808, 2002

BOCCIA, F. **Os alimentos geneticamente modificados e perspectiva do consumidor**. *Revista Nutricional Agricultura*, São Paulo, v. 20, n. 1, p. 10-22, 2015.

BORDONI, R. et al. **Detection and quantitation of genetically modified maize (Bt-176 Transgenic maize) by applying ligation detection reaction and universal array technology**. *Journal of Agricultural and Food Chemistry*, v.52, p.1049-1054, 2004.

BORÉM, A. **A história da Biotecnologia**. In: *BioTecnologia – Ciência & Desenvolvimento*, 2005. Disponível em: <http://www.biotecnologia.com.br/revista/bio34/bio34.pdf#page=10>

BORÉM, A; MILACH, S.C.K. **Melhoramento de plantas**, 2010. Disponível em: <http://www.biotecnologia.com.br/revista/bio07/encarte7.pdf>, Acessado em: 22 mar 2018.

BOWMAN, D.T; GUTIERREZ, O.A; PERCY, R.G; CALHOUN, D.S; MAY, O.L. **Pedigrees of upland and pima cotton cultivars released between 1970 and 2005**. Mississippi Agricultural & Forestry Experiment Station, Starkville, 58p. (Bulletin 1155), 2006

BRASIL, Lei 11.105, 24 de março de 2005. Disponível em: http://www.planalto.gov.br/ccivil_03/ato2004-2006/2005/lei/l11105.htm. Acesso em: 04 mai 2019.

BRASIL. Ministério de Ciência e Tecnologia. Comissão Técnica Nacional de Biossegurança. Parecer Técnico n. 513/2005. **Liberção comercial de algodão**

geneticamente modificado resistente às principais pragas da Ordem Lepidoptera. 2005. Disponível em: <<http://www.cntbio.mcti.gov.br>>. Acesso em: 09 jan 2020.

BRASILEIRO, A.C.M.; DUSI, D.M.A. Transformação genética de plantas. In: TORRES, A.C. et al. **Cultura de tecidos e transformação genética de plantas.** Brasília: EmbrapaSPI/Embrapa-CNPH, Parte IV, v.2, p.679-735, 1999.

BRERETON, R. G. **Chemometrics: Data analysis for the laboratory and chemical plant** Chichester: John Wiley. *Journal of Chemometrics*, 2003.

BURGER, J.; GOWEN, A. **Data handling in hyperspectral image analysis.** *Chemometrics and Intelligent Laboratory Systems*, v.108, p.13-22, 2011

CAMPBELL, B.T., HINZE, L.L. **Cotton production, processing, and uses of cotton.** B. Singh, editor, Industrial crops and uses. CABI, UK. p. 259–276, 2010.

CAMPBELL, B.T. et al. **Genetic Analysis of Cottonseed Protein and Oil in a Diverse Cotton Germplasm.** Crop Science, vol. 56, sep–oct 2016.

CARRER, H.; BARBOSA, A. L.; RAMIRO, D. A. **Biotecnologia na agricultura.** *Estudos Avançados*, v.24, n.70, p.149-164, 2010.

CEITIL, L. O. **Process Analytical Technology in Active Pharmaceutical Ingredients Production.** *Dissertação para a obtenção do Grau de Doutor em Engenharia Química.* Lisboa: Universidade Técnica de Lisboa - Instituto Superior Técnico, 2007.

CIB - Conselho de Informações sobre Biotecnologia, 20 anos de transgênicos: impactos ambientais, econômicos e sociais no Brasil, 2018.

CONCEIÇÃO, F.R.; MOREIRA, A.N.; BINSFELD, P.C; **Detecção e quantificação de organismos geneticamente modificados em alimentos e ingredientes alimentares.** *Ciência Rural*, Santa Maria v.36, n.1, p.315-324, jan-fev, 2006.

CONCEIÇÃO, F.R. et al. **Detecção de organismos geneticamente modificados.** In: BINSFELD, P.C. Biossegurança em biotecnologia. Rio de Janeiro: Interciência. p.145-169, 2004.

CUI, H.; YE, Z.; XU, L.; FU, X.; FAN, C.; YU, X. **Automatic and Rapid Discrimination of Cotton Genotypes by Near Infrared Spectroscopy and Chemometrics.** *Journal of Analytical Methods in Chemistry*. 2012.

DE BARROS NETO, B.; SCARMINIO, I. S.; BRUNS, R. E. **25 anos de quimiometria no Brasil.** *Química Nova*, v.29, n.6, p.1401-1406, 2006

DE JUAN, A; MAEDER, M; HANCEWICZ, T; DUPONCHEL, L; TAULER, R. **Chemometrics tools for image analysis.** In: H. W. SALZER, R.; SIESLER (Ed.); *Infrared and Raman Spectroscopic Imaging*. p.65–106, 2009.

DE JUAN, A. **Hyperspectral image analysis**. When space meets Chemistry Journal of Chemometrics. p. 1-13, 2017.

DELATORRE, C. A. **Plantas transgênicas: avaliando riscos e desfazendo mitos**. Porto Alegre: Evangraf, p.37, 2005.

EMBRAPA, **Algodão resistente a herbicida será apresentada a produtores cearenses, 2017**. Disponível em: <https://www.embrapa.br/busca-de-noticias/-/noticia/25276303/algodao-resistente-a-herbicida-sera-apresentado-a-produtores-cearenses>. Acesso em: 05 mar 2019.

EMBRAPA. **Contribuição da Embrapa sobre a introdução de eventos de algodoeiros geneticamente modificados para tolerância a herbicidas e para resistência a insetos listados na chamada 02/2007, da Comissão Técnica Nacional de Biossegurança, à luz do conhecimento disponível ao público em 8 de agosto de 2007**. 2007

ENKILDE, K.; JACOBSEN, S.; SØNDERGAARD, I. **Multivariate data analysis of proteome data**. Methods in Molecular Biology, v.355, p.195- 210, 2007.

EWING, G.W. **Métodos instrumentais de análise química**. Tradução: ALBANESE.A.G.; CAMPOS, J.T de S. São Paulo: Blucher, 2011.

FALEIRO, F.G; ANDRADE, S.R.M de; **Biotecnologia, transgênicos e biossegurança**, Planaltina, DF: Embrapa Cerrados, 2009.

FANTAPPIE, M. **Epigenética e Memória celular**. Revista Carbono. Número 03.Dossiê. 2013. Disponível em: <http://www.revistacarbono.com/wp-content/uploads/2013/06/Marcelo-Fantappie-Epigen%C3%A9tica-e-Mem%C3%B3ria-Celular.pdf>.

FEMENIAS, A.; GATIUS, F.; RAMOS, A.J.; SANCHIS, V.; MARÍN, S. **Use of hyperspectral imaging as a tool for Fusarium and deoxynivalenol risk management in cereals: A review**. Food Control. Vol.108, 2020.

FREIRE, E.C. **Algodão no cerrado do Brasil**. Associação brasileira de produtores de algodão. Brasília, p.333-355, 2007.

FREIRE, E.C; Morrelo,C de L; Matos, J.P; Senhorelo, W.L,P. **Desempenho comercial das cultivares BRS Aroeira e BRS Ipê no Estado de Goiás- Safra 2001/2002**, 2002.

GALVÃO, R.K.H; ARAÚJO, M.C.U; JOSÉ, G.E, et al; **A method for calibration and validation subset partitioning**, Talanta, vol.67, p. 736-740, 2005.

GAO, J. et al. **Application of hyperspectral imaging technology to discriminate different geographical origins of *Jatropha curcas* L. seeds**. Computers and Electronics in Agriculture, v. 99, p.186-193, 2013.

GELADI, P.; KOWALSKI, B. R. **Partial least-squares regression: a tutorial.** *Analytica Chimica Acta*, v. 185, p. 1–17, 1986

GIEHL, G. **A biotecnologia e segurança dos alimentos transgênicos.** Revista *Âmbito Jurídico*, v. 23, n. 5, p. 180-189, 2006.

GIOVANNINI, T.; CONCILLO, L. **PCR detection of genetically modified organisms: a review.** *Starch*, v.54, p.321-327, 2002

GOETZ, A. F. H.; VANE, G., Solomon, T.H.E.; ROCK, B.N. **Imaging spectrometry for earth remote sensing.** *Science*, v. 228, p. 1147-1153, 1985.

GOWEN, A. A.; MARINI, F.; ESQUERRE, C.; O'DONNELL, C. L.; DOWNEY, G.; BURGER, J. **Time series hyperspectral chemical imaging data: Challenges, solutions and applications.** *Analytica Chimica Acta*, v.705, p.272–282, 2011

GOWEN, A.A; O'DONNELL, C.P; CULLEN, P.J; DOWNEY, G; FRIAS, J.M. **Hyperspectral imaging — an emerging process analytical tool for food quality and safety control,** *Trends in Food Science & Technology* v.18, p.590–598, 2007.

GRAHN, H.F.; GELADI, P. **Techniques and Applications of Hyperspectral Image Analysis,** John Wiley, 2007.

GREDILLA, A.; DE VALL EJUELO, S. F-O; AMIGO, J. M.; DE DIEGO, A.; MADARIAGA, J. M. **Unsupervised pattern-recognition techniques to investigate metal pollution in estuaries.** *Trends in Analytical Chemistry*, v.46, p.59-69, 2013.

GRIFFITHS, P.R. **Imaging and Raman instrumentation for mapping and imaging,** in: R. Salzer, H.W. Siesler (Eds.), *Infrared and Raman Spectroscopic Imaging*, Wiley VCH Verlag GmbH&Co. KGaA, Weinheim, p. 3–62, 2011.

HAI-FENG, C; ZI-HONG, Y; LU-XU; XIAN-SHU, F; CUI-WEN, F; XIAO-PING, Y. **Automatic and Rapid Discrimination of Cotton Genotypes by Near Infrared Spectroscopy and Chemometrics.** *Journal of Analytical Methods in Chemistry*, China, 2012.

HAO, Y. et al. **Identification of Rice Varieties and Transgenic Characteristics Based on Near-Infrared Diffuse Reflectance Spectroscopy and Chemometrics.** *Molecules*, 24, 4568, 2019. Disponível em: www.mdpi.com/journal/molecules. Acesso em: 12 jan 2020.

HASLBERGER, A; LEKOAPE, K. **Modern food biotechnology, human health and development: and evidence-based study.** Department of Food Safety, Zoonoses and Foodborne Diseases, World Health Organization. p.32-37, 2005.

HOLST-JENSEN, A. et al. **PCR technology for screening and quantification of genetically modified organisms (GMOs)**. *Analytical and Bioanalytical Chemistry*, v.375, p.985- 993, 2003.

HUANG, X.; YUAN, T.; LIANG, M.; DU, M.; XIA, S.; DITTMAR, R. **Exosomal miR-1290 and miR-375 as prognostic markers in castration-resistant prostate cancer**. *European Urology*, v.67, p.33–41, 2015.

HURBURGH, C.R. et al. **Detection of genetically modified grains by near-infrared spectroscopy**. *Proceedings PITTCON 2000—Science for the 21st Century*, New Orleans, La, 2000.

IQBAL, A.; SUN, D-W.; ALLEN, P. **Review An overview on principle, techniques and application of hyperspectral imaging with special reference to ham quality evaluation and control**. *Food Control*, v.46, p.242-254, 2014

ISAAA. 2018. **Global Status of Commercialized Biotech/GM Crops in 2018: Biotech Crops Continue to Help Meet the Challenges of Increased Population and Climate Change**. ISAAA Brief No. 54. ISAAA: Ithaca, NY. Disponível em: <https://d335luupugsy2.cloudfront.net/cms/files/50569/1566823794Resumo-executivo.pdf>. Acesso em 12 jan 2020

JAMES, C. **Status global das cultvares transgênicas comercializadas**. Serviço Internacional para Aquisição de Aplicações em Agrobiotecnologia (ISAAA), 2013.

KALIVAS, J. H.; GEMPERLINE, P. J. Calibration. In: P. GEMPERLINE (Ed.); *Practical Guide to Chemometrics*. 2nd ed., p.105–165, 2006. Boca Raton: Taylor & Francis Group

KENNARD, R. W.; STONE, L. A. **Computer Aided Design of Experiments**, *Technometrics*, v.11, n. 1, p. 137 – 148, Feb. 1969.

KLUMPER, W. & Qaim, M. **A Meta -Analysis of the Impacts of Genetically Modified Crops**. *PLoS ONE*. v.9, p.11, 2014.

KOK, E.J. et al. **DNA methods: critical review of innovative approaches**. *Journal of AOAC International*, v.85, p.797-800, 2002.

KOTHARI, N.; CAMPBELL, B.T.; DEVER, J.K.; HINZE, L.L. **Combining ability and performance of cotton germplasm with diverse seed oil content**. *Crop Sci*. vol.56, p. 1–11, 2015.

KUMAR, N.; BANSAL, A.; SARMA, G. S.; RAWAL, R. K. **Review Chemometrics tools used in analytical chemistry: An overview**. *Talanta*, v.123, p.186–199, 2014.

KUMAR, M.; SINGH, S.P.; KUMAR, M.; KUMAR, A.; KUMAR, S.; KUMARI, P.; **Biosafety Issues in Commercialization and Development of Transgenic Crops. International**. *Journal of Current Microbiology and Applied Sciences*. v.7, n.4, 2018.

LIU, W. et al. **Liquid-phase hybridization based PCR-ELISA for detection of genetically modified organisms in food.** Food Control, v.15, p.303-306, 2004.

LUNA, A.S.; DA SILVA, A.P.; PINHO J.S.A.; FERRÉ, J.; BOQUÉ, R. **Rapid characterization of transgenic and non-transgenic soybean oils by chemometric methods using NIR spectroscopy.** Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy. v. 100, p. 115-119, 2013.

MAGALHÃES, D.F. **Aplicação da espectroscopia de infravermelho próximo na monitorização de processos farmacêuticos.** Lisboa, 2014.

MAGIN, K. et al. **Methods for detection of GMO grain in commerce.** 2000.

MAHGOUB, S.E. O.; NOLLET, L.M.L. **Testing and Analysis of GMO-containing Foods and Fedd.** 1st Edition, CRC Press, 2018.

MARIOTTI, E. **Surface plasmon resonance biosensor for genetically modified organisms detection.** Analytica Chemical Acta, v.453, p.165-172, 2002

MARTENS, H.; NIELSEN, J. P.; ENGELSEN, S. B. (2003). **Light scattering and light absorbance separated by extended multiplicative signal correction. Application to near-infrared transmission analysis of powder mixtures.** Analytical Chemistry, vol. 75(3), p.394–404.2003.

MARIOTTI, E; MINUNNI, M. & MASCINI, M. **Surface plasmon resonance biosensor for genetically modified organisms' detection.** Analytica Chimica Acta, vol. 453, n. 2, p. 165 -172 ,2002.

METROHM. NIR Spectroscopy. **A guide to near-infrared spectroscopic analysis of industrial manufacturing processes.** Herisau, Suíça: Metrohm, 2013.

MIRAGLIA, M. et al. **Detection and traceability of genetically modified organisms in the food production chain.** Food Chemical Toxicology, v.42, p.1157-1180, 2004.

MONSANTO EMEA, MON 88913 Genuity® Roundup Ready® Flex Cotton Second-Generation Herbicide Tolerance Key facts, may, 2015. Disponível em: https://www.europabio.org/sites/default/files/products/information/mon_88913_key_facts_20150506.pdf. Acesso em: 12 jan 2020

MOTA, A. A. R. **Transgenia no Brasil: eventos autorizados e cultivares registradas.** 2011. 124 f. Monografia (Graduação em Agronomia e Medicina Veterinária) – Universidade de Brasília, UNB, Brasília, 2011.

MUNCK, L.; MOLLER, B., JACOBSEN, S.; SONDERGAARD, S. **Near infrared spectra indicate specific mutant endosperm genes and reveal a new mechanism for substituting starch with (1/3, 1/4)-b-glucan in barley.** Journal of Cereal Science, vol.40, p, 213–222, 2004.

OBEID, P.J. et al. **Rapid analysis of genetically modified organisms by in-house developed capillary electrophoresis chip and laser-induced fluorescence system.** Electrophoresis, v.25, p.922-930, 2004.

OZAKI, Y. **Near-infrared spectroscopy - its versatility in analytical chemistry,** Anal. Sci. vol. 28, 2012.

PÁDUA, G.P; VIEIRA, R.D. **Deterioração de sementes de algodão durante o armazenamento.** Revista Brasileira de Sementes. Brasília, v.23, n.2, p.255-262, 2001.

PAPA, G. **Pragas e seu controle.** ALGODÃO: pesquisas e resultados para o campo. Cuiabá-MT: FACUAL, p 2006-239, 2006.

PASQUINI, C. **Near infrared spectroscopy: Fundamentals, practical aspects and analytical applications.** Journal of the Brazilian Chemical Society, v. 14, n. 2, p. 198–219, 2003.

PASQUINI, C. **Near infrared spectroscopy: A mature analytical technique with new perspectives e A review.** Analytica Chimica Acta 1026 p, 8-36, 2018.

PEREIRA, C. M. **A Espectroscopia NIR no Controle de Qualidade de Solventes Industriais.** Dissertação para obtenção do Grau de Mestre em Engenharia Química. Lisboa:Universidade Técnica de Lisboa - Instituto Superior Técnico, 2011.

PESSOA, L. G. A.; SILVA, W. E. B.; RIBEIRO, M. P.; LOUREIRO, E. S.; DIAS, P. M. **Efeito do sal de ácido carboxílico sobre ninfas de Aphis gossypii Glover (1877) (Hemiptera: Aphididae).** Revista de Agricultura Neotropical, Cassilândia-MS, v. 4, Suplemento 1, p. 79-83, dez. 2017.

PETIT, L. et al. **Screening of genetically modified organisms and specific detection of Bt176 maize in flours and starches by PCR-enzyme linked immunosorbent assay,** European Food Research and Technology, v.217, p.83-89, 2003.

POWELL, C. 2015. **How to make a GMO.** Harvard University, The Graduate School of Arts and Sciences Blog, Special Edition on GMOs.2015.
<http://sitn.hms.harvard.edu/flash/2015/how-to-make-a-gmo/>

PRATS-MONTALBÁN, J. M.; DE JUAN, A.; FERRER, A. **Multivariate image analysis: A review with applications.** Chemometrics and Intelligent Laboratory Systems, v.107, n.1, p.1–23, 2011.

RIBEIRO, M.I; COUTO, C; RIBEIRO, R; SILVA, S; **Perception about consumer's knowlegde and confidence in transgenic products.** Revista de Ciências Agrárias., vol. 40, p. 266-273, 2017.

- RICHETTI, A. **Estimativa de custo de produção de algodão, safra 2006/07, para o Mato Grosso e Mato Grosso do Sul.** Comunicado Técnico 125. CPAO, 16p, 2006.
- RODRÍGUES-PULIDO, F.J. et al. **Grape seed characterization by NIR hyperspectral imaging.** Postharvest Biology and Technology, v.76, p.74-82, 2013.
- SALZER, R.; SIESLER, H.W. (Eds.), **Infrared and Raman Spectroscopic Imaging,** Wiley-VCH, Weinheim (Germany), 2009.
- SANFORD, J.C. The biolistic process. Trends in Biotechnology; v .6, p. 299-302, 1988.
- SANTOS, R. L.; TORRES, J. B. **Produção da Proteína Cry1Ac em algodão transgênico e controle de lagartas.** Revista Brasileira de Ciências Agrárias, [S.I.], v. 5, n. 4, p. 509-517, out-dez, 2010.
- SILVA, C.S. **Uso de imagens hiperespectrais na região do infravermelho próximo para identificar fraudes em documentos.** Dissertação de mestrado, Recife, 2013.
- SILVERSTEIN, R. M.; WEBSTER, F. X.; KIEMLE, D. J. **Spectrometric Identification of Organic Compounds.** 7th ed. Danvers: John Wiley & Sons, 2005.
- SILVIE, P.; BÉLOT, J. L.; MICHEL, B. **Manual de identificação das pragas e seus danos no cultivo de algodão.** 2. ed. Cascavel-PR: COODETEC/CIRAD-CA, (Boletim Técnico, 34), p. 120, 2007.
- SINGH, S. **Transgenic cotton-its adoption, threats and challenges ahead: A review Subash Singh.** Journal of Entomology and Zoology Studies, p.1989-1997, 2018.
- SKOOG, D. A.; HOLLER, F. J.; CROUCH, S. R. **Principles of Instrumental Analysis.** Brooks Cole, 2006.
- SKOOG, D.A.; HOLLER, F.J.; NIEMAN, T.A. **Princípios de Análise Instrumental.** 6. Ed. Porto Alegre: Bookman, 2009.
- SOARES-COSTA, A.; SCHNEIDER, V.K.; NAKAYAMA, D.G.; ARENCIBIA, A.D.; HENRIQUE-SILVA, F. **Plantas transgênicas fundamentos, métodos de produção e aplicações.** Biotecnologia Aplicada à Agro&Indústria - Vol. 4. 1ed.: Editora Blucher, p.649-674,2017.
- STAVE, J.W. **Detection of new or modified proteins in novel foods derived from GMO – future needs.** Food Control, v.10, p.361-374, 1999.
- STAVE, J.W. **Protein immunoassay methods for detection of biotech crops: Applications, limitations and practical considerations.** Journal of AOAC International, v.85, p.780-786, 2002.

STEINBERG, F. **Biotecnologia farmacêutica e bioterapia: uma visão geral.** *Society for Pharmaceutical Sciences*, v. 12, n. 3, p. 1-17, 2002.

TERADA, R.; URAWA, H.; INAGAKI, Y.; TSUGANE, K.; IIDA, S. **Efficient gene targeting by homologous recombination in rice.** *Nature Biotechnology*, v.20, p.1030-1034, 2002.

TORRES, A.C. et al. **Bioassay for detection of transgenic soybean seeds tolerant to glyphosate.** *Pesquisa Agropecuária Brasileira*, v.38, p.1053-1057, 2003

VARGAS, B.D, et al. **Biotecnologia e Alimentos Geneticamente Modificados: Uma Revisão.** *Revista Contexto & Saúde*, Editora Unijuí, v. 18, n. 35, p.19-25, jul/dez, 2018.

VIDAL, M.; AMIGO, J.M. **Pre-processing of hyperspectral images. Essential steps before image analysis.** *Chemometrics and Intelligent Laboratory Systems* v.117, p. 138–148, 2012.

VIEIRA, A. **Debates atuais sobre a segurança dos alimentos transgênicos e os direitos dos consumidores.** *Revista dos Tribunais*, v. 23, n. 60, p. 37-57, 2006.

VIEIRA, R. M.; BELTRÃO, N. E. de M. **Produção de sementes de algodoeiro.** In: Beltrão, N. E. de M. *O agronegócio do algodão no Brasil*. Campina Grande: EMBRAPA CNPA. v.1, p.430-453, 1999.

XIE, L.; YING, Y.; YING, T.; YU, H. FU,X. **Discrimination of transgenic tomatoes based on visible/near-infrared spectra,** *Analytica Chimica Acta*, vol. 584, p,379–384, 2007.

YAMAGUCHI, H. et al. **Two detection methods of genetically modified maize and the state of its import into Japan.** *Food Control*, v.14, p.201-206, 2003.

YI, W. S.; et al. Gastric cancer differentiation using Fourier transform near-infrared spectroscopy with unsupervised pattern recognition. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**, v.101, p.127-131, 2013

WHO. **Frequently asked questions on genetically modified foods.** Food safety. 2014. Disponível em: http://www.who.int/foodsafety/areas_work/food-technology/faq-geneticallymodified-food/en/. Acesso em: 07 Mai 2019

WOLD, S. **PLS-regression: a basic tool of chemometrics.** *Chemometrics and Intelligent Laboratory Systems*, v. 58, n. 2, p. 109–130, 2001.

WORKMAN, J.J; WEYER, L. **Practical Guide to Interpretive Near-Infrared Spectroscopy.** CRC Press, 2007.

