

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE QUÍMICA

NÁTALI DA SILVA SARMENTO

**CLASSIFICAÇÃO DE EMBALAGENS DE PAPEL POR INFRAVERMELHO E
QUIMIOMETRIA**

Porto Alegre
2021

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE QUÍMICA

NÁTALI DA SILVA SARMENTO

**CLASSIFICAÇÃO DE EMBALAGENS DE PAPEL POR INFRAVERMELHO E
QUIMIOMETRIA**

Trabalho de conclusão apresentado junto à atividade de ensino “Projeto Tecnológico – Química” do curso de Química Industrial, como requisito parcial para a obtenção do grau de Bacharel em Química Industrial.

Orientador: Prof. Dr. Marco Flôres Ferrão

Coorientadores: Dr. Marcelo C. A. Marcelo e Dr. Samuel Kaiser

Porto Alegre
2021

AGRADECIMENTOS

Aos meus pais, Alessandro Sarmiento e Tatiane Sarmiento, por serem minha base, por terem lutado por um futuro melhor para mim e para o meu irmão desde tão jovens, vibrado com os meus sucessos e me apoiado em todas as dificuldades. Sem vocês nada disso seria possível.

Ao meu irmão e melhor amigo, Eduardo Sarmiento, por ter sido meu fiel ouvinte, sempre com uma palavra de acalento e uma história para me arrancar gargalhadas.

Ao meu companheiro, Gabriel Zimmermann, por toda paciência, zelo, suporte e carinho no semestre mais complicado da nossa trajetória.

Ao meu time, C&CD, que facilitou a sobrevivência por mais de um ano em meio à pandemia. Minha gratidão infinita às minhas amigas Julia de Azevedo, Luciana Canova, Liane Bokowski, Liliane Porte, Camila Assis, Laura Vendrame e Fernanda Lini, e aos meus amigos Maurílio Nespeca, Thiago Shwanz e Jailson Dias, por atribuírem leveza e crescimento aos meus dias de trabalho, pelas conversas e risadas. E também por me ouvirem reclamar muito.

Aos meus mentores, Marcelo Marcelo e Samuel Kaiser, por terem me apresentado a quimiometria e feito esse trabalho acontecer, por toda orientação, apoio e conversas.

Aos meus “quimigos”, que acompanharam essa história desde o princípio. À Letícia e à Nathália, minhas parceiras maravilhosas de TM3, que foram minhas primeiras e melhores amigas na universidade, fundamentais para que eu tivesse chegado até aqui. Ao Augusto, meu quase irmão, meu amigo de todas as horas (principalmente das piores), por sempre enxergar o melhor em mim mesmo quando eu mesma não conseguia. Ao André, por sempre divertir meus dias, mesmo os mais cansativos, além de toda a ajuda ao longo da graduação. À Jaíne e à Lilian, por serem sempre tão solícitas e motivadoras. O grupo da quarentena me motivou demais, muito obrigada!

A toda minha nada pequena família, avós, bisavós, tios, tias, primas, primos e afilhada, por serem um porto seguro e estarem na torcida, comemorem junto comigo.

Finalmente, a mim mesma, que tive a resiliência e a dedicação para sair do ensino público da periferia para obtenção do grau de bacharela em Química Industrial na Universidade Federal do Rio Grande do Sul.

RESUMO

Embalagens de papel são amplamente utilizadas comercialmente, possuindo diferentes propriedades e sendo produzidas por diferentes processos de acordo com o propósito da sua aplicação. Dessa forma, cada embalagem de papel possui características químicas particulares que permitem uma análise de sua composição e o emprego de métodos quimiométricos para rastrear e identificar propriedades de interesse do material, como seu fabricante e o tipo de papel utilizado. Foi utilizada neste trabalho a técnica de espectroscopia no infravermelho próximo por transformada de Fourier. Ela permite uma análise rápida, barata e não destrutiva da composição química da embalagem de papel, que é majoritariamente composta por celulose, lignina e hemiceluloses. Para realização das classificações utilizando o espectro de infravermelho, foram empregadas as ferramentas quimiométricas de análise multivariada como a Análise por Componentes Principais (PCA, do inglês *Principal Component Analysis*) e Análise Discriminante por Mínimos Quadrados Parciais (PLS-DA, do inglês *Partial Least Square Discriminatory Analysis*). As amostras analisadas para a construção do conjunto de calibração foram fornecidas por diferentes empresas produtoras de papel e cedidas pela BAT Brasil, e as amostras utilizadas na validação externa foram obtidas no comércio local e também fornecidas pela BAT Brasil. Os resultados obtidos por este método mostram que é possível classificar com 100 % de acurácia os tipos de papel contidos na calibração, validação interna e externa e indicar as amostras que estão fora do escopo do modelo. Baseando-se nestes resultados, a ferramenta desenvolvida foi promissora e teve capacidade de classificar diferentes tipos de embalagens de papeis, podendo ser empregada em indústrias que produzem e que recebem este material, assim como ser utilizadas para gerar inteligência de negócio.

Palavras-Chave: Análise Discriminatória por Mínimos Quadrados Parciais (PLS-DA). Análise por Componentes Principais (PCA). Classificação de Embalagens. Espectroscopia no Infravermelho. Quimiometria.

ABSTRACT

Paper packaging are widely used commercially. They have different properties and suffer different process that change according to the application purpose. Each packing and paper type contains particular chemical information that allows a chemical analysis associated with chemometrics to track and identify the material characteristics of interest, as well as they manufacturer and paper type used. Near-infrared spectroscopy by Fourier Transform technique was used on this project, enabling a quick, cheap, easy and non-destructive method, since the paper chemical properties are mostly derived from cellulose, lignin and hemicelluloses, which are organic compounds easily detected by the proposed method. Classification was accomplished using infrared spectra through chemometrics tools such as Principal Component Analysis (PCA) and Partial Least Square Discriminant Analysis (PLS-DA). Calibration set analyzed samples were obtained through different paper producing companies and provided by BAT Brazil, as well as the ones used on external validation was obtained from local market. Results showed paper type classification with 100 % accuracy and indicated out of model scope samples. Based on the results, Developed tool is very promising, it was able to classify different paper types and can be used by paper producing industries as well as companies that receive this material.

Keywords: Partial Least Square Discriminant Analysis (PLS-DA). Principal Component Analysis (PCA). Packaging Classification. Infrared Spectroscopy. Chemometrics.

LISTA DE FIGURAS

- FIGURA 1 – Teste de permutação no software SIMCA, com 100 permutações realizadas para o teste estatístico, onde os quadrados em azul representam as amostras originais e os círculos em verde representam as amostras permutadas. O gráfico mostra o cenário de um modelo ajustado, onde todas as amostras permutadas seguem distribuição semelhante..... 22
- FIGURA 2 – Espectros brutos de FT-NIR coletados em absorbância, nos números de onda de 10.000 cm^{-1} até 4000 cm^{-1} de todo o conjunto amostral, incluindo amostras da análise exploratória e do conjunto de calibração. 28
- FIGURA 3 – Espectros de FT-NIR, pré-processados utilizando SNV, coletados em absorbância, nos números de onda de 7.100 cm^{-1} até 4000 cm^{-1} de todo o conjunto amostral, incluindo amostras da análise exploratória e do conjunto de calibração. 29
- FIGURA 4 – Espectros de FT-NIR, pré-processados utilizando primeira derivada de Savitzky-Golay, com janela de 15 pontos, polinômio de ordem 2, e ordem de derivada igual a 1, coletados em absorbância, nos números de onda de 7.100 cm^{-1} até 4000 cm^{-1} de todo o conjunto amostral, incluindo amostras da análise exploratória e do conjunto de calibração. 29
- FIGURA 5 – Exemplos de estruturas químicas das macromoléculas que compõem o papel majoritariamente: Lignina (A), Celulose (B) e Hemicelulose (C), evidenciando os diferentes grupos funcionais da lignina, e as unidades repetitivas da celulose e hemicelulose, que são polímeros naturais que variam de acordo com o tipo de planta e outros aspectos. 31
- FIGURA 6 – Escores da PCA realizada com as duas primeiras componentes principais, PC1 com 36,6 % da variância explicada e PC2 com 14,6 %; com elipse de confiança de 95 %, referentes aos tipos de papel das embalagens após receber a

impressão, onde os círculos vermelhos representam o tipo de papel I, os azuis escuros o tipo H, os verdes o tipo C, os amarelos o tipo A e os azuis claros o tipo B..... 33

FIGURA 7 – Escores da PCA realizada com as componentes principais, PC1 com 36,6 % da variância explicada e PC3 com 10,5 %; com elipse de confiança de 95 %, referentes aos tipos de papel das embalagens após receber a impressão, onde os círculos vermelhos representam o tipo de papel I, os azuis escuros o tipo H, os verdes o tipo C, os amarelos o tipo A e os azuis claros o tipo B. 33

FIGURA 8 – Gráfico da PCA realizada com as componentes principais, PC2 com 14,6 % da variância explicada e PC3 com 10,5 %; com elipse de confiança de 95 %, apresentando os escores referentes aos tipos de papel das embalagens após receber a impressão, onde os círculos vermelhos representam o tipo de papel I, os azuis escuros o tipo H, os verdes o tipo C, os amarelos o tipo A e os azuis claros o tipo B. 34

FIGURA 9 – Gráfico da PCA realizada com as componentes principais, PC1 com 38,2 % da variância explicada e PC2 com 15,9 %; com elipse de confiança de 95 %, apresentando os escores referentes aos tipos de papel das embalagens sem impressão, onde os círculos vermelhos representam o tipo de papel I, os azuis escuros o tipo H, os verdes o tipo C, os amarelos o tipo A e os azuis claros o tipo B..... 35

FIGURA 10 – Gráfico da PCA realizada com as componentes principais, PC1 com 38,2 % da variância explicada e PC3 com 11,5 %; com elipse de confiança de 95 %, apresentando os escores referentes aos tipos de papel das embalagens sem impressão, onde os círculos vermelhos representam o tipo de papel I, os azuis escuros o tipo H, os verdes o tipo C, os amarelos o tipo A e os azuis claros o tipo B..... 35

FIGURA 11 – Gráfico da PCA realizada com as componentes principais, PC2 com 15,2 % da variância explicada e PC3 com 11,5 %; com elipse de confiança de 95 %, apresentando os escores referentes aos tipos de papel das embalagens sem impressão, onde os círculos vermelhos representam o tipo de papel I, os azuis escuros o tipo H, os verdes o tipo C, os amarelos o tipo A e os azuis claros o tipo B..... 36

FIGURA 12 – Gráfico da PCA realizada com as componentes principais, PC2 com 15,9 % da variância explicada e PC4 com 3,79 %; com elipse de confiança de 95 %, apresentando os escores referentes aos tipos de papel das embalagens sem impressão, onde os círculos vermelhos representam o tipo de papel I, os azuis escuros o tipo H, os verdes o tipo C, os amarelos o tipo A e os azuis claros o tipo B..... 36

FIGURA 13 – Gráfico da PCA realizada com as componentes principais, PC3 com 11,5 % da variância explicada e PC4 com 3,79 %; com elipse de confiança de 95 %, apresentando os escores referentes aos tipos de papel das embalagens sem impressão, onde os círculos vermelhos representam o tipo de papel I, os azuis escuros o tipo H, os verdes o tipo C, os amarelos o tipo A e os azuis claros o tipo B..... 37

FIGURA 14 – Gráfico da PCA realizada com as duas primeiras componentes principais, onde PC1 possui 61,1 % da variância explicada e PC2 14,9 %, com elipse de confiança de 95 %, apresentando os escores referentes ao papel do tipo I, onde os círculos amarelos representam a embalagem analisada com a área impressa, e os azuis a embalagem analisada com a área em branco..... 38

FIGURA 15 – Gráfico de DmodX com intervalo de confiança de 95 % das amostras analisadas na face impressa e não impressa, demonstrando que as amostras analisadas com a face impressa apresentam resíduo acima do limite crítico estabelecido..... 39

- FIGURA 16 – Gráfico de T^2 de Hotelling com intervalo de confiança de 95 % das amostras analisadas na face impressa e não impressa, demonstrando que as amostras analisadas com a face impressa apresentam resíduo acima do limite crítico estabelecido..... 39
- FIGURA 17 – Gráfico da PCA realizada com as duas primeiras componentes principais, onde PC1 contém 25,1 % da variância explicada e PC2 12,0 %, com elipse de confiança de 95 %, apresentando os escores referentes ao papel I impresso, dividido em 3 partes: Início do rolo, representado pelos escores em azul; Meio do rolo, representado pelos escores em vermelho e Fim do rolo, representado pelos escores em verde..... 41
- FIGURA 18 – Gráfico da PCA realizada com as duas primeiras componentes principais, onde PC1 contém 9,68 % da variância explicada e PC2 8,47 %, com elipse de confiança de 95 %, apresentando os escores referentes ao papel I sem impressão, dividido em 3 partes: Início do rolo, representado pelos escores em azul; Meio do rolo, representado pelos escores em vermelho e Fim do rolo, representado pelos escores em verde. 41
- FIGURA 19 – Gráfico da PCA realizada com as duas primeiras componentes principais, onde PC1 contém 13,2 % da variância explicada e PC2 8,47 % com elipse de confiança de 95 %, apresentando os escores referentes ao papel B sem impressão, dividido em 3 partes: Início da embalagem, representado pelos escores em azul; meio da embalagem, representado pelos escores em vermelho e fim da embalagem, representado pelos escores em verde..... 42
- FIGURA 20 – Gráfico da PCA realizada com as duas primeiras componentes principais, onde PC1 contém 32,2 % da variância explicada e PC2 11,3 %, com elipse de confiança de 95 %, apresentando os escores referentes ao papel B impresso, dividido em 3 partes: Início da embalagem, representado pelos escores em verde; Meio da embalagem representado pelos escores em azul e Fim da embalagem representado pelos escores em vermelho..... 42

FIGURA 21 – Gráfico da PCA realizada com as duas primeiras componentes principais, onde PC1 contém 92,6 % da variância explicada e PC2 6,94 %, com elipse de confiança de 95 %, apresentando os escores referentes ao tipo I impresso e não impresso dobrado uma, duas, três e quatro vezes ao meio antes da leitura no equipamento. O escores em verde representam a embalagem dobrada uma vez, tanto branca quanto impressa, em azul embalagem dobrada duas vezes, tanto branca quanto impressa, em amarelo embalagem dobrada três vezes, tanto branca quanto impressa, e em vermelho embalagem dobrada quatro vezes, tanto branca quanto impressa. 44

FIGURA 22 – Gráfico da PCA realizada com as duas primeiras componentes principais, com elipse de confiança de 95 %, apresentando os escores referentes a todos os tipos de embalagens de papel analisados, evidenciando a diferença de coloração, explicada pela PC 1, e a diferença de gramatura, explicada pela PC 2, onde os escores em verde representam o tipo K que possui a coloração preta, os escores em amarelo representam as embalagens de alta gramatura (de 200 gsm a 270 gsm), os escores em vermelho representam as de gramatura intermediária (140 gsm) e por fim os escores em laranja representam as embalagens de papel de fina espessura (75 gsm). 45

FIGURA 23 – Fluxograma criado para facilitar a classificação de amostras, subdividindo a classificação em três modelos. O fluxograma mostra a sequência que uma amostra deve ser submetida para uma classificação correta do tipo de embalagem de papel, passando primeiramente pelo modelo de coloração, após pelo modelo de gramatura para então ser classificada como o tipo de papel da embalagem. 46

LISTA DE TABELAS

- TABELA 1 – Descrição das amostras de embalagem de papel utilizados na análise exploratória preliminar em relação aos tipos de papel, gramatura e estágio da embalagem de papel, ou seja, com ou sem impressão. 24
- TABELA 2 – Descrição das amostras de embalagem de papel utilizados na análise discriminante, em relação aos tipos de papel e gramatura. Todas as embalagens descritas receberam impressão. 25
- TABELA 3 – Matriz de confusão do modelo de classificação de coloração do material, demonstrando classificação correta para todas as amostras do conjunto de validação externa. 47
- TABELA 4 – Matriz de confusão do modelo de classificação de gramatura das embalagens de papel, demonstrando classificação correta para todas as amostras do conjunto de validação externa, previamente classificadas como coloração branca. 48
- TABELA 5 – Matriz de confusão do modelo de classificação de tipos de embalagens de papel, demonstrando classificação correta para todas as amostras do conjunto de validação externa, previamente classificadas como coloração branca e gramatura de 200 gsm a 270 gsm. 49
- TABELA 6 – Matriz de confusão do modelo de classificação de tipos das embalagens de papel, demonstrando classificação correta para todas as amostras do conjunto de validação externa, previamente classificadas como coloração branca e gramatura de 75 gsm. 50
- TABELA 7 – Resultados da predição das amostras de 1 a 11 referentes à validação externa do modelo de classificação de embalagens de papeis, mostrando que 8 das 11 amostras foram classificadas corretamente, enquanto que as outras 3

apresentaram resíduos estatísticos altos devido ao seu material não estar
contido na calibração do modelo. 51

LISTA DE ABREVIATURAS E SIGLAS

FIR	Infravermelho Distante
FN	Falso Negativo
FP	Falso Positivo
FT	Transformada de Fourier
IR	Infravermelho
LVs	Variáveis Latentes
MIR	Infravermelho Médio
NIR	Infravermelho Próximo
NV	Negativo Verdadeiro
PC	Componente Principal
PCA	Análise por Componentes Principais
PLS	Mínimos Quadrados Parciais
PLS-DA	Análise Discriminante por Mínimos Quadrados Parciais
PV	Positivo Verdadeiro
SNV	Padronização Normal de Sinal
UV	Unidade de Variância

SUMÁRIO

1	INTRODUÇÃO	13
1.1	Propriedades do papel.....	13
1.2	Embalagens de papel.....	13
1.3	Espectroscopia no infravermelho.....	14
1.4	Análise multivariada	14
1.4.1	Pré-processamento de dados.....	16
1.4.2	Análise por Componentes Principais (PCA)	18
1.4.3	Análise Discriminante por Mínimos Quadrados Parciais (PLS-DA).....	20
2	OBJETIVO E PROPOSTA TECNOLÓGICA.....	23
3	MATERIAIS E MÉTODOS	23
3.1	Amostras.....	23
3.2	Instrumento	25
3.3	Análise de dados	26
3.3.1	Análise de Dados para PCA	26
3.3.2	Análise de Dados para PLS-DA	27
4	RESULTADOS E DISCUSSÕES	28
4.1	Interpretação química.....	28
4.2	Desenvolvimento de metodologia analítica	32
4.2.1	Teste da impressão.....	32
4.2.2	Teste da área de análise	37
4.2.3	Teste de Homogeneidade.....	40
4.2.4	Teste de dobradura.....	43
4.3	Análise Exploratória	44
4.4	Modelos de Classificação	46
4.4.1	Validação Interna.....	46
4.4.2	Validação externa	50
5	CONCLUSÕES.....	53
	REFERÊNCIAS BIBLIOGRÁFICAS	54

1 INTRODUÇÃO

1.1 PROPRIEDADES DO PAPEL

O papel é, por definição, uma folha fina, contendo um emaranhado de fibras de origem vegetal, majoritariamente árvores como o Eucalipto e o Pínus no Brasil, mas também existem materiais de fontes alternativas como cana de açúcar, linho, entre outras.¹ As propriedades do material dependem de uma série de fatores das fibras vegetais, como a sua fonte, o seu tipo, como é extraída da fonte, a quantidade de tratamento que são expostas durante o processo de polpação, como são combinadas, o número de camadas usadas para fazer uma folha de papel, quais aditivos são usados no processo, se o papel ou papelão é branqueado, revestido ou combinado com outros materiais.¹

É possível perceber o quanto as fibras são importantes analisando as influências nas propriedades do papel, e a escolha de sua fonte deve ser feita visando a características desejadas no produto. As fibras que compõem um papel devem ser de dois tipos: as fibras de madeira macia (fibras longas) e as de madeira rígida (fibras curtas)² obtidas no Brasil respectivamente do Eucalipto e do Pínus, árvores de crescimento rápido e cultivadas em plantios florestais. Ou seja, as florestas nativas não são exploradas para este fim.³

O processamento da polpa possui grande relevância para a propriedade de um papel e consiste em dois principais processos: mecânico, onde se utiliza trituradores, e químico. Ainda existem dois métodos intermediários, os processos termomecânicos e químico-termomecânico ou semi-químico. O processo mecânico é o mais barato e mais rápido, pois consiste em moer mecanicamente as lascas de madeira previamente cortadas e lavadas; o processo químico, por outro lado, é o método mais caro, e ocorre através de uma digestão das lascas de madeira, até a separação das ligninas e impurezas das fibras de interesse.¹

1.2 EMBALAGENS DE PAPEL

Desde a pré-história, com os primeiros humanos, a utilização de embalagens se mostrava fundamental frente à necessidade de proteger os alimentos, carregar ferramentas ou agrupar pertences. Logicamente, a primeira ideia de embalagem era rudimentar, e passou por

inúmeras transformações e evoluções através dos séculos, em questão de materiais e tecnologias.⁴

Um dos pontos chave para que se iniciassem os investimentos em melhoria nas embalagens foi a Revolução Industrial, onde a produção em massa demandou em escala proporcional a venda e transporte dos produtos, exigindo uma atenção para as quantidades medidas de forma padronizada de determinado produto, ou seja, as embalagens de tamanhos padronizados começaram a ser utilizadas.⁴ Além disso, com o crescimento das indústrias, a produção aumentando cada vez mais, e por consequência, o surgimento de concorrência entre produtos, a embalagem adquiriu uma nova e importante função, informar a marca proveniente e características do produto contido no seu interior.

Ao longo das décadas, a sociedade se transformou e novas tendências mundiais foram originadas, e algumas delas diretamente correlacionadas à globalização. As embalagens e seus formatos, materiais e *designs* acompanharam, e acompanham até hoje, tais mudanças, como, por exemplo, a migração das embalagens de papel e de vidro para as embalagens plásticas, em meados de 1980, devido ao custo associado à produção da matéria prima. No momento atual, século 21, há uma forte inclinação para embalagens sustentáveis, retornáveis ou reutilizáveis e rapidamente biodegradáveis, ou seja, o vidro e o papel retornam como a maior tendência do setor.

Assim como o papelão, o papel é uma das principais matérias primas empregadas na fabricação de embalagens para os mais diversos produtos⁴, passando pelo setor de cosméticos até o setor alimentício, e do ponto de vista sustentável, a utilização desses materiais tende apenas a crescer.

1.3 ESPECTROSCOPIA NO INFRAVERMELHO

A espectroscopia no infravermelho é uma técnica quanti e qualitativa, amplamente empregada na determinação de compostos químicos, sejam eles de origem orgânica ou inorgânica, em estado sólido, gasoso ou líquido, de forma rápida, robusta e não destrutiva.⁵ A radiação infravermelha está compreendida no espectro eletromagnético entre a região do visível e das micro-ondas, na faixa de 780 nm a 100.000 nm.⁶ O espectro de infravermelho é dividido entre 3 regiões, o infravermelho próximo (NIR) com comprimento de onda de 780 nm

a 2.500 nm, o infravermelho médio (MIR) de 2.500 nm a 5.000 nm e o infravermelho distante (FIR), que vai de 5.000 nm a 100.000 nm.⁶

Esta técnica baseia-se no fenômeno de absorção de radiação infravermelha pelas moléculas que convertem a radiação em energia de vibração e rotação dos seus grupos funcionais, e essa frequência de vibração pode ser associada a um tipo específico de ligação química, possibilitando a identificação e quantificação dessa molécula.⁷ A radiação absorvida pelas moléculas resulta em estiramentos das ligações, vibrações e deformações angulares e axiais. As vibrações que ocasionam a variação do momento dipolo da molécula são observadas no espectro de infravermelho, pois nessa situação o campo elétrico alternado da radiação consegue interagir com a molécula e absorver em um determinado comprimento de onda compreendido no espectro.⁷⁻⁸ Na espectroscopia de infravermelho próximo, o espectro obtido é em forma de bandas alargadas e sobrepostas de sobretons e bandas de combinação.

A espectroscopia no infravermelho utiliza métodos para a aquisição dos espectros de infravermelho, como a reflexão e a transmissão. Com o método de reflexão, os espectrômetros são capazes de medir de forma rápida uma grande faixa espectral através de interferômetros empregando a transformada de Fourier (FT).⁹

A espectroscopia NIR e a quimiometria são uma combinação harmônica, pois com a quimiometria é possível extrair inúmeras informações contidas nos espectros NIR, que possui bandas largas e sobrepostas, permitindo a identificação e a quantificação de diversos parâmetros em diferentes matrizes.¹⁰

1.4 ANÁLISE MULTIVARIADA

A análise multivariada de dados químicos, também conhecida como quimiometria, é uma ferramenta matemática e estatística de enorme peso na área da química, pois, como citado anteriormente, permite extrair e maximizar informações acerca de padrões existentes entre as respostas de diferentes análises químicas, como na espectroscopia de infravermelho. A quimiometria subdivide-se em três grandes áreas: planejamento de experimentos, análise exploratória (ou reconhecimento de padrões) e calibração multivariada.¹²

Planejamento experimental consiste em avaliar ou otimizar um produto ou processo, através de um método multivariado onde todos os fatores que tem-se interesse em avaliar são alterados simultaneamente em um conjunto de experimentos. Dessa forma, é possível definir

o menor número de experimentos necessários para extrair o máximo de informação dos dados, reduzindo tempo e custos.¹³

A análise exploratória ocorre a partir de uma grande quantidade de informações (dados químicos, neste caso) sobre as amostras de interesse, onde se pretende encontrar agrupamentos dessas amostras que apresentem similaridade, e dessa forma detectar tendências nos dados.¹² Na calibração multivariada, o objetivo é criar um modelo que relacione as medidas realizadas nas amostras com uma determinada propriedade de interesse, seja ela qualitativa (classificação) ou quantitativa (regressão).¹²

1.4.1 Pré-processamento de dados

A etapa de pré-processamento dos dados, que nada mais é do que transformações e ajustes matemáticos de forma prévia a exploração dos dados, é fundamental para o sucesso da análise multivariada. Os principais objetivos da aplicação das técnicas de pré-processamento são eliminar ou ao menos minimizar as fontes de interferências, ruídos instrumentais ou informações não relevantes para tornar a matriz de dados mais ajustada. Assim, é possível realizar uma análise exploratória, classificação ou regressão com o conjunto de dados com clareza.^{12, 14}

As técnicas de pré-processamento mais amplamente utilizadas podem ser divididas em duas categorias principais: os métodos de correção de dispersão da luz e os métodos de derivadas espectrais.¹⁵ Em dados de infravermelho, é comum o emprego dos métodos de normalização dos espectros, aplicar a derivada de Savitzky-Golay de primeira ou segunda ordem e centrar os dados na média.

Emprega-se como método de correção e normalização a Padronização Normal de Sinal (SNV, do inglês “Standard Normal Variate”). Ele corrige a variação entre as amostras gerada pela dispersão da luz e ajusta as mudanças de linha de base (diferentes intensidades), a partir da seguinte Equação 1:

$$X_{\text{corrigido}} = \frac{X_{\text{original}} - a_0}{a_1} \quad (1)$$

onde $X_{\text{corrigido}}$ é o espectro obtido após o pré-processamento, X_{original} é o espectro bruto, a_0 é o valor médio da amostra que vai ser corrigida, e a_1 é o desvio padrão da amostra.¹⁵

A primeira derivada de Savitzky-Golay tem como função reduzir o efeito da relação sinal-ruído nos espectros e a remoção da linha de base, estimando a diferença entre dois pontos subsequentes no espectro de infravermelho, como mostrado na Equação 2.¹⁵ Nesse método de pré-processamento uma janela é selecionada, contendo um número de variáveis, e os pontos nessa janela são utilizados para determinar o valor no seu ponto central e, assim, o tamanho da janela influencia diretamente o resultado da derivação.

$$x'_i = x_i - x_{i-1} \quad (2)$$

onde x'_i é o ponto derivado, x_i é um ponto qualquer no espectro, e x_{i-1} é um ponto exatamente ao lado do ponto escolhido.

A segunda derivada de Savitzky-Golay, além de corrigir o efeito de linha de base, também remove a tendência linear, e é estimada calculando a diferença entre dois pontos sucessivos dos espectros já corrigidos pela derivada de primeira ordem, segundo a Equação 3 mostrada abaixo.¹⁵

$$x''_i = x'_i - x'_{i-1} = x_{i-1} - 2.x_i + x_{i+1} \quad (3)$$

Por fim, o último pré-processamento dos dados foi a centralização na média e auto escalamento dos espectros, utilizando o método de Unidade de Variância (UV) que centraliza o dado e o atribui pesos com base na Equação 4.¹⁶⁻¹⁷

$$\frac{1}{DP_j} \quad (4)$$

onde DP_j é o desvio padrão da variável j calculado em torno da média.

A utilização do autoescalamento dos dados não é comumente empregada em dados de infravermelho e uma das explicações é a quantidade de ruído que esse tipo de dado pode apresentar. Porém, Nespeca et al.¹⁴ observaram que os espectros infravermelho na região do infravermelho próximo possuem absorções de diferentes magnitudes devido aos sobretons, que são níveis mais energéticos que por sua vez apresentam menor intensidade, ou seja, com o autoescalamento é possível dar maior peso a variáveis que possuem menor intensidade, e assim melhorar o método.¹⁴

A centralização na média consiste em calcular a média das intensidades para cada comprimento de onda e subtrair cada uma das intensidades do valor médio fazendo com que cada variável passe a ter média igual a zero. Dessa forma, as coordenadas vão para o centro dos dados, fazendo com que as diferenças nas intensidades das variáveis sejam visualizadas de forma mais fácil. A Equação 5 é empregada para a centralização dos dados.

$$X_{(centradonamédia)} = X_{(original)} - X_{(média)} \quad (5)$$

onde $X_{(centradonamédia)}$ é a intensidade corrigida com média igual a zero, $X_{(original)}$ é a variável antes do processamento e $X_{(média)}$ é o valor médio das intensidades.

1.4.2 Análise por Componentes Principais (PCA)

Na espectroscopia de infravermelho sem a análise multivariada, as amostras são como pontos localizados em um espaço com “n” dimensões, sendo esse “n” o número de variáveis (comprimentos de onda).¹² A análise por componentes principais (PCA) é um método não supervisionado que permite expressar as informações mais significativas contidas originalmente na matriz de dados em um número menor de novas variáveis, denominados componentes principais (PC), sendo útil na verificação de tendências, identificação comportamentos incomuns, entre outras aplicações.⁶ Expressando de forma matemática, na PCA, a matriz de dados originais, chamada aqui matriz “**X**”, é decomposta em um produto de duas outras matrizes, como mostrada na Equação 6.^{11, 12, 18}

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E} \quad (6)$$

onde **X** é a matriz de dados, **T** são os escores, **P** são os pesos e **E** é a matriz de erros.

A matriz **P**, ou seja, a matriz de pesos carrega a informação sobre como as variáveis da matriz original estão relacionadas às componentes, enquanto que a matriz **T**, matriz dos escores, contém as coordenadas das amostras nos eixos das componentes.¹⁹ O produto das matrizes **T** e **P** produz uma aproximação da matriz de dados **X**, tendo em vista que a matriz **T** possui o número de linhas igual ao número de linhas da matriz **X**, sendo usualmente o número de amostras e a matriz **P** por sua vez possui o mesmo número de colunas, sendo geralmente o

número de variáveis.¹⁹ Observando do ponto de vista geométrico, esse produto de matrizes \mathbf{T} e \mathbf{P} gera um novo sistema de coordenadas, segundo Teófilo e colaboradores, como as amostras têm coordenadas no espaço original, definidas pelas variáveis originais, elas também têm coordenadas em relação aos novos eixos, que são os escores.¹⁹

A primeira componente principal (chamada de PC1) apresenta-se no sentido da maior variação no conjunto de dados originais \mathbf{X} , enquanto que a segunda (PC2) que é ortogonal à primeira descreve o maior percentual da variação que não foi possível explicar através da PC1 e assim sucessivamente com as outras PCs.

Realizando uma avaliação dos pesos é possível observar quais das variáveis são responsáveis pelos agrupamentos visualizados.¹² Através de uma análise crítica do gráfico de escores e também do gráfico de pesos, pode-se avaliar quais componentes principais, são responsáveis pelas características observadas entre as amostras e os agrupamentos.¹⁸

É possível identificar amostras que fogem do padrão estabelecido pelo modelo da PCA avaliando o gráfico da PCA e verificando se existem amostras fora da elipse de confiança de 95 %, e através do resíduo T^2 de Hotelling. O resíduo T^2 de Hotelling é um dos testes de hipótese multivariada mais utilizados, medindo o quão longe se encontra uma amostra observada do centro do modelo, sendo calculado através da Equação 7:

$$T_i^2 = \frac{t_i^T (T^T T)^{-1} t_i}{I-1} \quad (7)$$

onde \mathbf{T} é a matriz dos escores ($I \times R$) de todas as amostras de calibração, e t_i é um vetor dos escores \mathbf{R} até a amostra i .

A amostra é considerada fora da curva e deve ser excluída quando ultrapassa três vezes mais o valor crítico de T^2 , para o intervalo de confiança definido, usualmente 95 % ou 99 %.

1.4.3 Análise Discriminante por Mínimos Quadrados Parciais (PLS-DA)

A análise discriminante por mínimos quadrados parciais (PLS-DA) têm como objetivo classificar as amostras em classes pré-definidas, diferente da análise exploratória, onde o objetivo é visualizar os padrões entre as propriedades graficamente.¹⁴ Esta é uma ferramenta de classificação linear baseada na regressão mínimos quadrados parciais (PLS) que encontra

variáveis latentes com uma covariância máxima entre as informações químicas.²⁰ Entretanto, ao invés de utilizar a propriedade de interesse como um vetor dependente (\mathbf{Y}), são usadas as classes pré-definidas de forma codificada.¹⁴

O método PLS faz com que a relação entre a variável dependente e os escores seja maximizada, portanto, as variáveis latentes (LVs) representam as direções que melhor discriminam as classes.¹⁴ Os pesos (*loadings*) são os coeficientes de variáveis nas combinações lineares que determinam os LVs, logo, podem ser interpretados como a influência de cada variável em cada LV, enquanto que os escores representam as coordenadas das amostras no espaço de projeção.²¹

A PLS-DA não modela somente as variáveis, diferentemente da PCA, visto que o modelo assume a distribuição dos erros entre o conjunto de referência (classes) e as variáveis, ou seja, o método modela dois conjuntos de Equações 8 e 9 distintamente.²²

$$\mathbf{X} = \mathbf{TP} + \mathbf{E} \quad (8)$$

$$c = \mathbf{T}q + f \quad (9)$$

onde \mathbf{c} é um vetor de comprimento \mathbf{I} , que representa um rótulo numérico para cada amostra de acordo com sua classe a ser determinada; q é um parâmetro análogo ao peso na PCA, equivalente ao número de LV; e \mathbf{f} assim como \mathbf{E} , representa uma matriz residual.^{6, 22}

Quando o sistema estudado apresenta n classes a serem determinadas, o vetor \mathbf{c} é calculado da seguinte maneira:⁶

$$c = n * \mathbf{I} \quad (10)$$

onde \mathbf{I} é a matriz identidade.

A escolha da quantidade de variáveis latentes para a PLS-DA é importante para evitar tanto sub-ajuste quanto o sobre ajuste do modelo. O sub-ajuste é um problema na modelagem pois ocorre quando não há dado o suficiente para explicar toda informação. Já o sobre-ajuste é o contrário, ocorrendo quando há inclusão de informação demais no modelo, levando a falsa impressão de que o modelo está classificando corretamente quando na verdade a correlação obtida não está ajustada com aquela que se deseja observar.²³

Para avaliar a existência de sobre-ajuste nos modelos, é empregado o teste de permutação, criado em 1930 por Fischer, que consiste em alterar os elementos que compõem a matriz fazendo com que haja a geração de uma nova combinação, avaliando dessa forma se o comportamento do modelo segue o mesmo, eliminando a dúvida de que haja qualquer tendência. Matematicamente, o teste segue alguns passos. Primeiramente, calcula-se a diferença entre a média das amostras, e logo em seguida agrupa-se todos esses dados; aleatoriamente, a permutação ocorre retirando uma amostra do conjunto e calculando novamente sua média em relação às amostras que permaneceram no grupo de dados, e assim sucessivamente dependendo do número de permutações indicado, rearranjando os dados remanescentes a cada retirada de uma amostra; dessa forma, o teste segue até a comprovação ou não da hipótese de que as amostras das classes comparadas têm distribuição semelhante, e que o modelo tem capacidade preditiva.²⁴ Graficamente, podemos observar a permutação como mostrado na Figura 1, que evidencia o comportamento ideal de um modelo ajustado, onde as amostras originais e permutadas seguem distribuição análoga. Observando o gráfico, identificamos o sobreajuste quando as amostras permutadas, quadrados azuis, ficam acima da linha de interseção, de valor zero, fazendo com que a linha de regressão se torne horizontal.

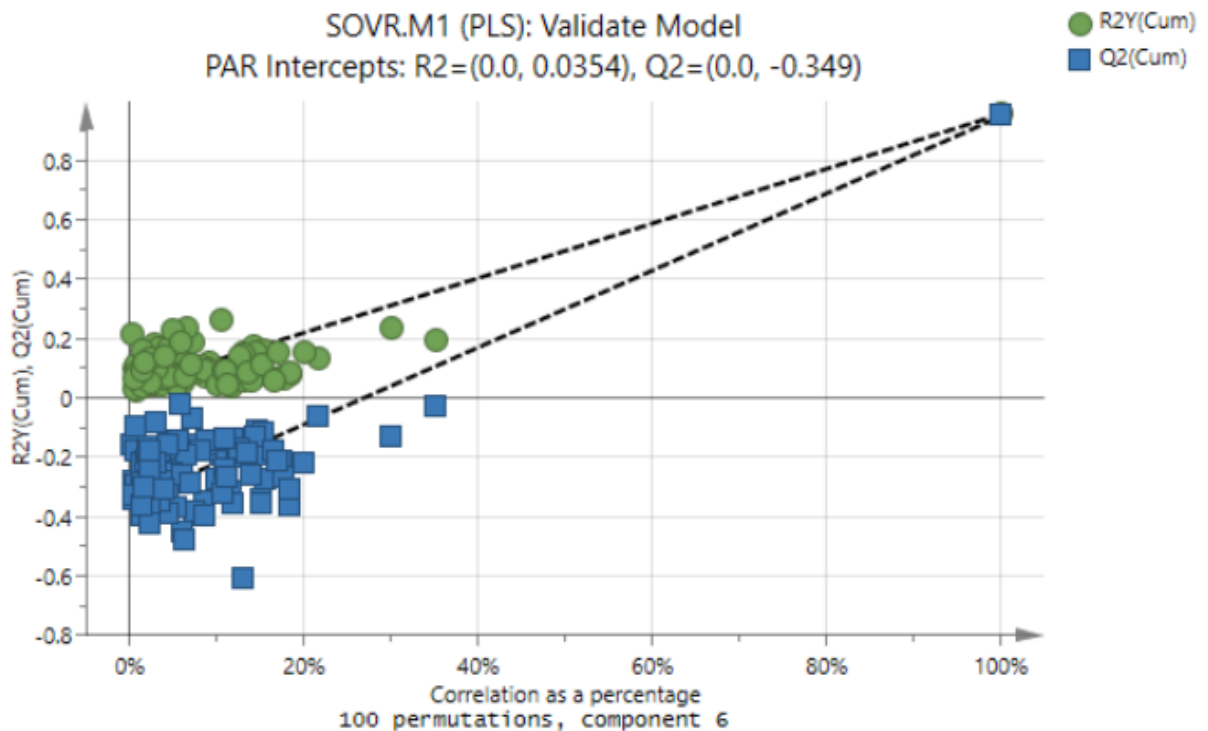


Figura 1 – Teste de permutação no software SIMCA, com 100 permutações realizadas para o teste estatístico, onde os quadrados em azul representam as amostras originais e os círculos em verde representam as amostras permutadas. O gráfico mostra o cenário de um modelo ajustado, onde todas as amostras permutadas seguem distribuição semelhante.

FONTE: SIMCA® user guide, Sartorius.

Outro parâmetro importante, assim como para a PCA, é a identificação de pontos fora da curva, os chamados *outliers*, através da análise dos resíduos Q ou DmodX e o T² de Hotelling. O resíduo Q é empregado na avaliação da conformidade de cada amostra no modelo calibrado, medindo a diferença dessa amostra no dado original e também no dado que foi gerado no modelo.²¹ O resíduo de Hotelling, T², baseia-se na soma do quadrado dos escores normalizados, medindo a distância ou variação de cada amostra em relação ao centro do modelo, como mostrado na Equação 7 previamente para a PCA.²¹ Dessa forma, quando o resíduo DmodX ou T² tem um valor alto, indica que a amostra não está conforme. Para que seja possível a exclusão das amostras com alto resíduo, que exercem uma influência no modelo de PLS-DA, um limite de confiança deve ser estabelecido para ambos os parâmetros, geralmente sendo de 95 % ou 99 %, onde escores que estejam acima desse limite são considerados suspeitos, e os que possuem valor três vezes superior a esse, são considerados fora do escopo do modelo e, portanto, excluídos.

2 OBJETIVO E PROPOSTA TECNOLÓGICA

O presente trabalho teve como objetivo o desenvolvimento e a validação de um método analítico que fosse capaz de diferenciar tipos de papéis empregando análise direta da embalagem de papel por espectroscopia no infravermelho próximo por transformada de Fourier com reflexão total atenuada e utilização da quimiometria.

A proposta tecnológica deste projeto é o desenvolvimento de uma ferramenta da indústria 4.0, de análise rápida, barata e robusta que pode ser empregada nas empresas substituindo a inspeção visual do material das embalagens e atribuindo certeza ao processo, além de ser uma ferramenta útil na análise de custo destes ao comparar as composições químicas e gramaturas.

3 MATERIAIS E MÉTODOS

3.1 AMOSTRAS

Quatrocentas e oitenta e cinco amostras de embalagem de papel de onze tipos diferentes foram analisadas por espectroscopia no infravermelho próximo. Cada tipo de embalagem está relacionado com um fabricante e composição química específicos. Os tipos de papel das embalagens foram identificados pelas letras do alfabeto sequencialmente de A até K.

As embalagens de papel possuem duas faces: a interna, que se tem contato com o produto, e a externa que recebe a impressão. Neste estudo, utilizou-se embalagens de papel sem impressão e com impressão de diferentes tipos de embalagem de papel. Suas diferentes gramaturas também foram avaliadas: 75 gsm, 140 gsm, 200 gsm, 240 gsm e 270 gsm, onde gsm é a medida de gramatura empregada quando trata-se de espessura de papel, que significa o peso em gramas por metro quadrado do material, medido hoje em dia por paquímetros.

Um primeiro conjunto de amostras de embalagem de papel foi utilizado para determinar o método de análise por espectroscopia no infravermelho próximo: influência da impressão, da face da embalagem a ser analisada, gramaturas e outros parâmetros. Para isto, um tipo de embalagem de papel de cada gramatura, com e sem impressão, foi selecionado para a análise exploratória dos dados, totalizando 134 amostras, e estas estão descritas na Tabela 1. Para análise discriminante, foram selecionadas 485 amostras de embalagem de papel entre 11 tipos de papel que estão descritos na Tabela 2, nesta etapa não se utilizou o papel em branco, apenas o impresso.

Todas as amostras utilizadas neste projeto foram fornecidas pela BAT Brasil.

Tabela 1 – Descrição das amostras de embalagem de papel utilizados na análise exploratória preliminar em relação aos tipos de papel, gramatura e estágio da embalagem de papel, ou seja, com ou sem impressão.

Classe de papel	Gramatura (GSM)	Embalagem sem Impressão (Unidade)	Embalagem Impressa (Unidade)
Tipo A	200	10	10
Tipo B	240	13	13
Tipo H	140	10	10
Tipo C	270	10	10
Tipo I	75	23	24

Tabela 2 – Descrição das amostras de embalagem de papel utilizados na análise discriminante, em relação aos tipos de papel e gramatura. Todas as embalagens descritas receberam impressão.

Tipo de papel	Gramatura (gsm)	Embalagem Impressa (Unidade)
Tipo A	200	60
Tipo B	240	120
Tipo C	270	20
Tipo D	225	20
Tipo E	240	145
Tipo F	200	30
Tipo G	240	10
Tipo H	240	10
Tipo I	240	10
Tipo J	140	30
Tipo K	270	30
TOTAL		485

3.2 INSTRUMENTO

As amostras foram analisadas em um espectrômetro de infravermelho próximo (Perkin Elmer, Frontier NIRA II, USA), com detector de arseneto de índio e gálio (InGaAs) e divisor de feixe de CaF₂. A análise foi realizada com 8 varreduras e resolução de 4 cm⁻¹. Os espectros foram obtidos em absorbância na faixa de 4.000 cm⁻¹ a 10.000 cm⁻¹. Realizou-se uma correção de fundo no início de cada dia de análise. As amostras foram posicionadas sob a janela de leitura recebendo o feixe de luz diretamente, não havendo, portanto, preparo de amostra nem sua destruição. Devido ao fato de o papel ser um material de baixa espessura, o

feixe emitido pelo espectrômetro de infravermelho atravessa a amostra, portanto utilizou-se o material para correção de fundo sob a amostra no momento da leitura.

3.3 ANÁLISE DE DADOS

Os dados brutos obtidos pelo equipamento foram importados para tratamento quimiométricos no software SIMCA (versão 15.0.2, Umetrics, USA).

Foram selecionados os comprimentos de onda de 7.100 cm^{-1} até 4000 cm^{-1} na análise exploratória inicial e na classificação por PLS-DA, devido a ausência de bandas e aos ruídos apresentados de 10.000 cm^{-1} até 7.100 cm^{-1} . Os espectros passaram por um pré-processamento utilizando a função derivada de Savitzky-Golay, com uma janela de 15 pontos, polinômio de ordem 2, e ordem de derivada igual a 1. A seguir, empregou-se a função Variação Normal Padrão (sigla SNV do inglês *Standart Normal Variate*) e, por último, as variáveis foram autoescaladas utilizando a função Variância de Unidade (sigla em UV do inglês *Unit Variance*).

Foram realizados a PCA e a PLS-DA com os conjuntos de dados pré-processados.

3.3.1 Análise de Dados para PCA

Para uma análise exploratória, não há uma regra específica para a quantidade de componentes principais, logo, a escolha das componentes principais para a PCA foi realizada analisando o máximo percentual de variância explicada pelo menor número de componentes. Elas são calculadas em ordem decrescente de significância e quantidade de variância, porém contêm informações diferentes e complementares, por isso precisam ser escolhidas com atenção para que não haja perda de informações.

Para a detecção de amostras fora da curva, foi determinada uma elipse limite de 95 %, e amostras fora desta são consideradas suspeitas. Além disso, o resíduo T^2 de Hotelling também é um parâmetro estatístico de detecção de *outliers*.

3.3.2 Análise de Dados para PLS-DA

A escolha das variáveis latentes para a PLS-DA foi realizada através do método de validação cruzada, ou seja, retirando os grupos de validação através do algoritmo de seleção Venetian Blinds²⁵, que faz uma divisão representativa dentro do conjunto total de amostras, separando um conjunto para validação e um para calibração, com 33 e 67 % das amostras, respectivamente.²³ Então o modelo é calibrado com as amostras selecionadas para isso e é realizada a predição do grupo retirado, comparando as classes preditas com aquelas previamente conhecidas. O processo é repetido até que todos os subgrupos tenham sido retirados uma vez para validação, sem que haja repetição de grupos.¹⁶⁻¹⁷

Em relação às amostras não conformes do modelo, foram avaliados os parâmetros T² de Hotelling, com intervalo de confiança de 95 %, e o resíduo DmodX, também com intervalo de confiança de 95 %. Amostras com resíduos T² e DmodX superior a duas vezes o valor crítico foram consideradas *outliers* e removidas do modelo.

Para a classificação dos tipos de papel das embalagens analisadas, utilizou-se os parâmetros numéricos de predição que serão utilizados para avaliar e comparar com os valores gerados pelo modelo em relação a cada amostra predita; Quando o valor predito é menor que 0,35 a amostra é predita como não pertencente a classe, quando o valor está entre 0,35 e 0,65, não é possível concluir a classe e quando o valor é superior a 0,65 é porque a amostra pertence a classe. Visto que foram coletados 3 espectros de cada embalagem, o tipo de papel foi classificado como pertencente a uma classe do modelo quando duas ou mais leituras foram classificadas igualmente.

Na matriz de confusão, que mostra os resultados das classes preditas *versus* classes verdadeiras, surgem quatro parâmetros, o positivo verdadeiro (PV), o negativo verdadeiro (NV), o falso positivo (FP) e o falso negativo (FN), através destes é possível o cálculo estatístico para avaliação do desempenho do modelo em relação às classes preditas. Esses cálculos são acurácia, sensibilidade e especificidade. A Equação 10 mostra como se calcula a acurácia, a Equação 11 mostra como é calculada a sensibilidade e, por último, a Equação 12 mostra o cálculo da especificidade do modelo.

$$Acurácia = \frac{PV+NV}{PV+NV+FP+FN} \quad (11)$$

onde PV é o positivo verdadeiro, NV o negativo verdadeiro, FP o falso positivo, e FN o falso negativo.

$$\text{Sensibilidade} = \frac{PV}{PV+FN} \quad (12)$$

onde PV é o positivo verdadeiro, e FN o falso negativo.

$$\text{Especificidade} = \frac{NV}{NV+FP} \quad (13)$$

onde NV é o negativo verdadeiro e FP o falso positivo.

4 RESULTADOS E DISCUSSÕES

4.1 INTERPRETAÇÃO QUÍMICA

Os espectros de infravermelho brutos de todo conjunto de amostras estão apresentados na Figura 2 e nas figuras 3 e 4 os espectros pré-processados.

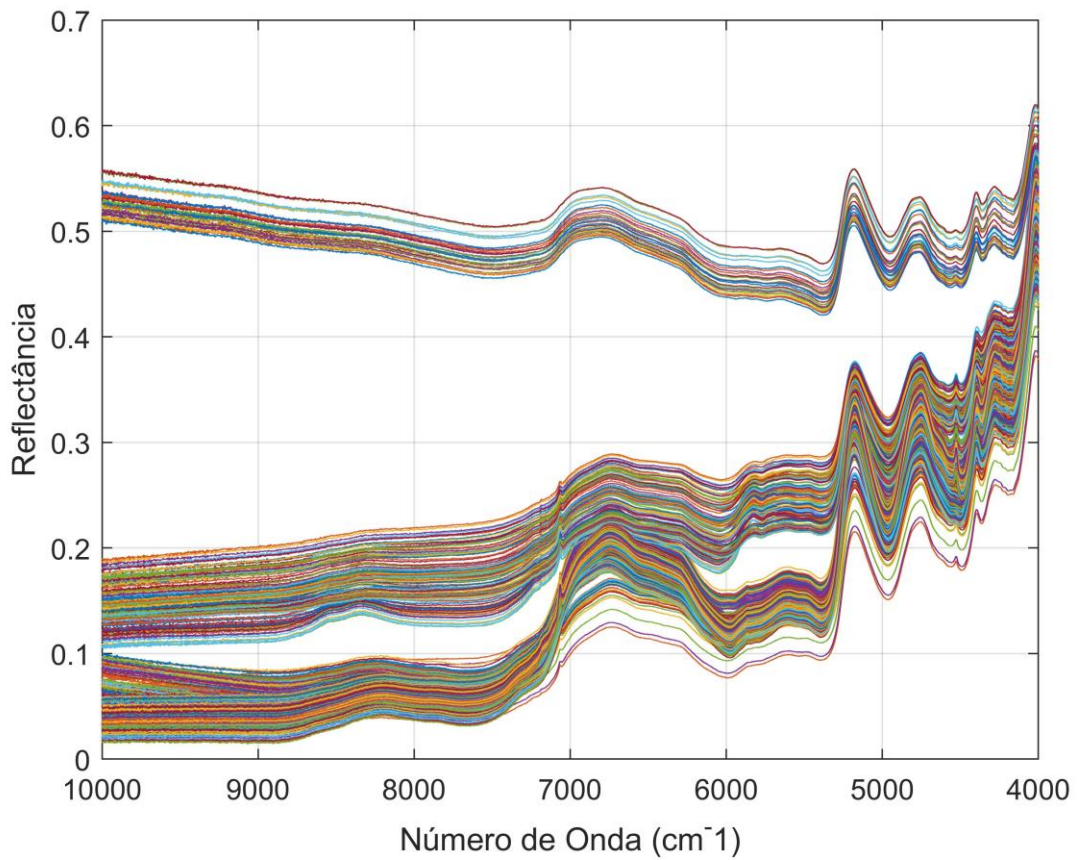


Figura 2 – Espectros brutos de FT-NIR coletados em absorbância, nos números de onda de 10.000 cm^{-1} até 4000 cm^{-1} de todo o conjunto amostral, incluindo amostras da análise exploratória e do conjunto de calibração.

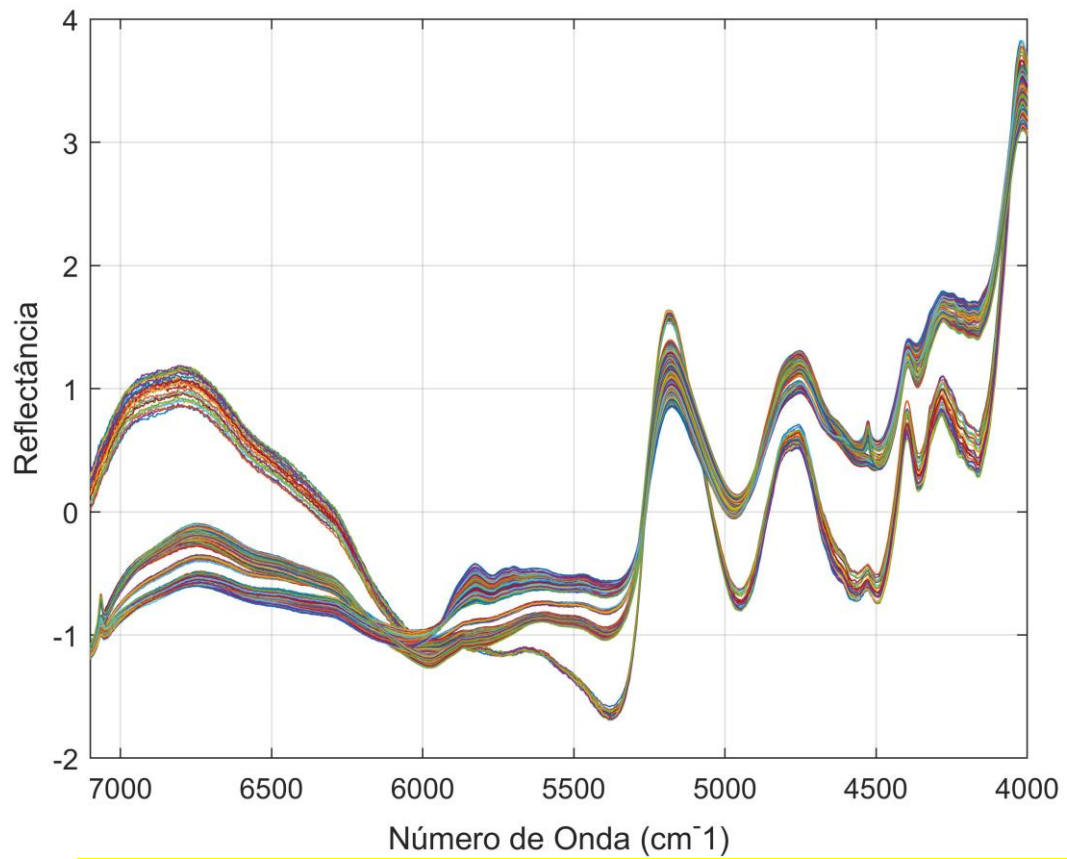


Figura 3 – Espectros de FT-NIR, pré-processados utilizando SNV, coletados em absorvância, nos números de onda de 7.100 cm^{-1} até 4000 cm^{-1} de todo o conjunto amostral, incluindo amostras da análise exploratória e do conjunto de calibração.

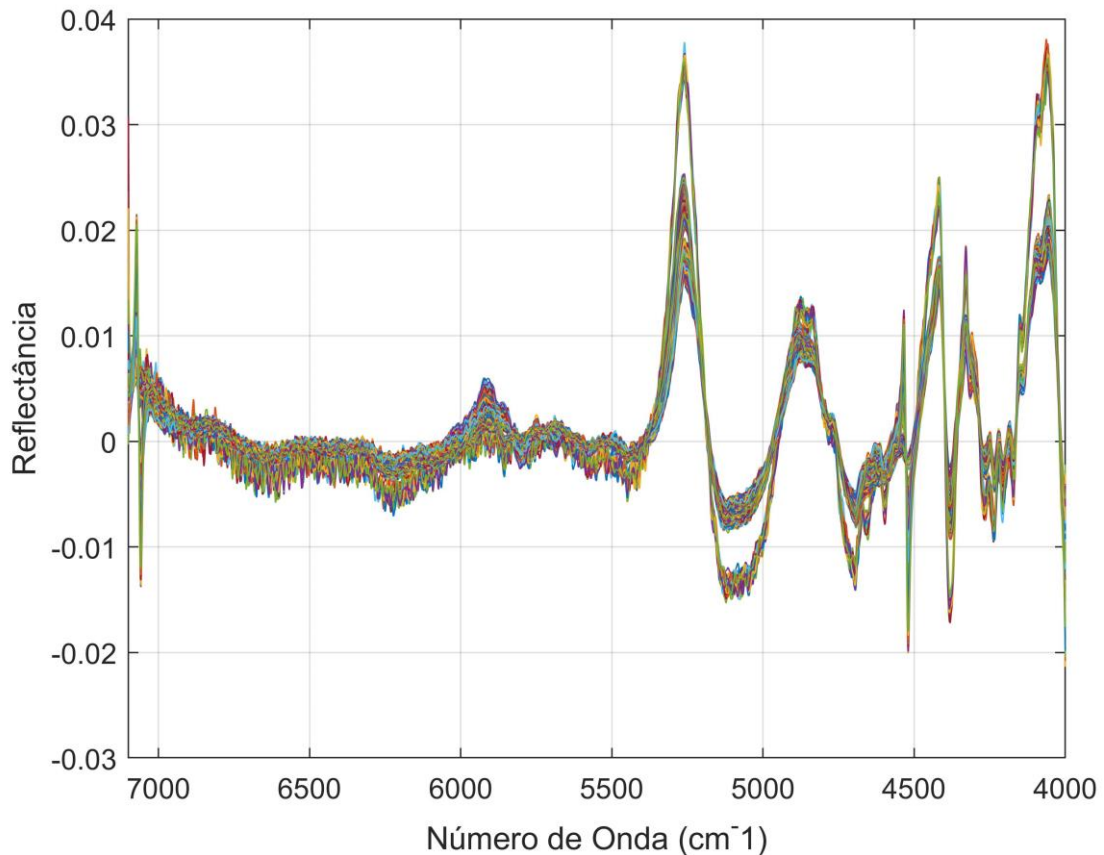


Figura 4 – Espectros de FT-NIR, pré-processados utilizando primeira derivada de Savitzky-Golay, com janela de 15 pontos, polinômio de ordem 2, e ordem de derivada igual a 1, coletados em absorvância, nos números de onda de 7.100 cm^{-1} até 4000 cm^{-1} de todo o conjunto amostral, incluindo amostras da análise exploratória e do conjunto de calibração.

Mesmo sem análise exploratória, é possível observar a diferença notória entre dois grupos de espectros, na Figura 2 e 3, devido a diferença de espessura da embalagem analisada, visto que o conjunto de espectros com maior absorvância entre 6.000 cm^{-1} e 4.000 cm^{-1} refere-se às embalagens com gramatura de 75 gsm , que compreende os tipos de papel I e J. O perfil espectral é muito semelhante, dada a mesma matriz e mesmas macromoléculas na composição, apenas as bandas são menos intensas provavelmente pela baixa espessura do material.

Devido à variabilidade de plantas utilizadas na fabricação do papel, a composição química também não é constante. Entretanto, existem compostos que sempre estarão presentes na sua matriz, como a celulose, a lignina e as hemiceluloses. Essas 3 macromoléculas provavelmente são as principais responsáveis pelas bandas e picos observados no espectro de infravermelho, variando de acordo com a formulação de cada tipo de papel utilizado para a embalagem. As Hemiceluloses e as ligninas são amorfas, enquanto que a celulose nem

sempre. Ela possui, na verdade, ambas as fases tanto cristalina quanto amorfa, portanto para a celulose existe um grau de cristalinidade, que é justamente a razão entre essas duas fases. O grau de cristalinidade é um fator de grande relevância quando se trata desta molécula, pois é diretamente relacionado com as propriedades químicas e físicas da celulose. Ele depende de inúmeros fatores, como a espécie da árvore de onde foi obtida, bem como todo o processo de plantio e ambiente envolvido no crescimento da planta.²⁶ Hemiceluloses e ligninas, para o papel, funcionam como aglutinantes e agentes que favorecem o endurecimento do material. A lignina é uma macromolécula muito ativa quimicamente, mais do que a própria celulose, devido aos seus grupos funcionais, como grupos fenólicos e hidroxilas.²⁷ Entretanto, o uso da lignina na indústria do papel varia de acordo com o interesse do produto final, pois a lignina sofre oxidação e torna o material amarelado com o tempo, e, por este motivo, algumas indústrias processam a polpa para retirada de parte da lignina.²⁶ As estruturas das três macromoléculas citadas estão na Figura 5.

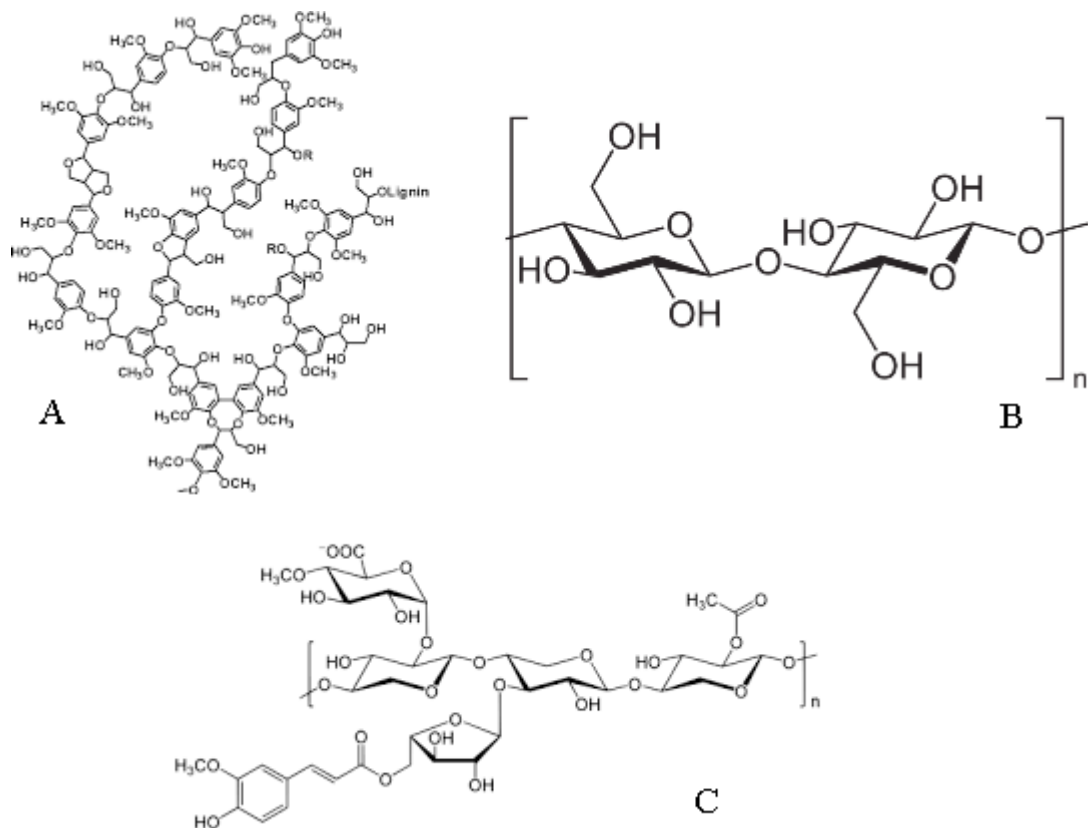


Figura 5 – Exemplos de estruturas químicas das macromoléculas que compõem o papel majoritariamente: Lignina (A), Celulose (B) e Hemicelulose (C), evidenciando os diferentes grupos funcionais da lignina, e as unidades repetitivas da celulose e hemicelulose, que são polímeros naturais que variam de acordo com o tipo de planta e outros aspectos.

A identificação das regiões do espectro realizou-se a partir dos espectros normalizados. Em 7.000 cm^{-1} , onde há a banda referente ao primeiro harmônico de OH, pode corresponder à região de absorção da estrutura amorfa da celulose. Em 5.263 cm^{-1} , tem-se a região referente ao primeiro harmônico de RCOOH e segundo harmônico da C=O, que possivelmente foram originados a partir dos grupos da hemicelulose. Na região aproximadamente 4.800 cm^{-1} , tem-se combinações ROH, possivelmente referentes às 3 macromoléculas do material. Na região entre 4.000 cm^{-1} e 4.700 cm^{-1} , tem-se as bandas referentes às combinações C-H + C-H, CH+C-C, CH, CH₂ e CH₃, podendo também ser resultado das estruturas da celulose, hemicelulose, lignina e outros compostos orgânicos adicionados na receita do papel utilizado na fabricação da embalagem.¹⁷

4.2 DESENVOLVIMENTO DE METODOLOGIA ANALÍTICA

Foram realizadas 4 avaliações de método por análise exploratória, considerando formas de análise da amostra. O objetivo dessas análises exploratórias preliminares é propor um método analítico com capacidade de melhor aquisição das informações químicas necessárias para diferenciação dos tipos de embalagem de papel.

4.2.1 Teste da impressão

Foram analisadas 10 amostras em triplicata de cada um dos cinco tipos de embalagem de papel (A, B, C, H e I) com e sem impressão, totalizando 100 amostras, para avaliação da influência da impressão no espectro de infravermelho do material, e conseqüentemente na sua classificação. As PCAs foram realizadas utilizando 3 e 4 componentes principais, para as embalagens impressas e não impressas, respectivamente.

Na figura 6, referente ao material impresso, infere-se que a PC1 separa os materiais de acordo com suas gramaturas, pois os escores da embalagem de 75 gsm, embalagem de papel do tipo I, estão posicionados bem a esquerda da elipse, os escores da embalagem de 140 gsm, tipo H, estão mais ao centro e por fim os escores do restante dos papeis de maior espessura

mais a direita do gráfico. A PC2 provavelmente é a responsável pela segregação dos tipos de papéis das embalagens, apenas não separando o tipo B e o tipo C, que são mais bem separados quando se compara PC2 e PC3, como mostrado na Figura 8 e devem ter maior semelhança entre suas composições químicas em comparação com o restante das embalagens. A Figura 7, que mostra a PCA com PC1 e PC3, corrobora a ideia que a PC1 está diretamente relacionada com a gramatura do material.

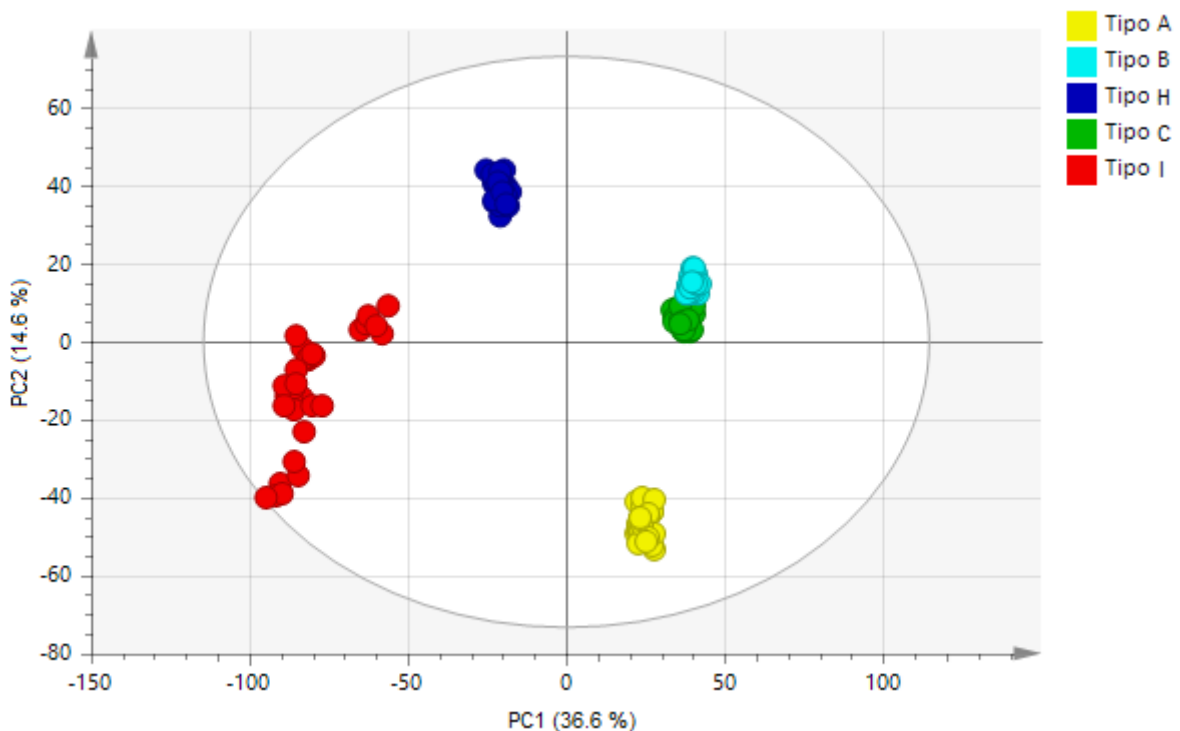


Figura 6 – Escores da PCA realizada com as duas primeiras componentes principais, PC1 com 36,6 % da variância explicada e PC2 com 14,6 %; com elipse de confiança de 95 %, referentes aos tipos de papel das embalagens após receber a impressão, onde os círculos vermelhos representam o tipo de papel I, os azuis escuros o tipo H, os verdes o tipo C, os amarelos o tipo A e os azuis claros o tipo B.

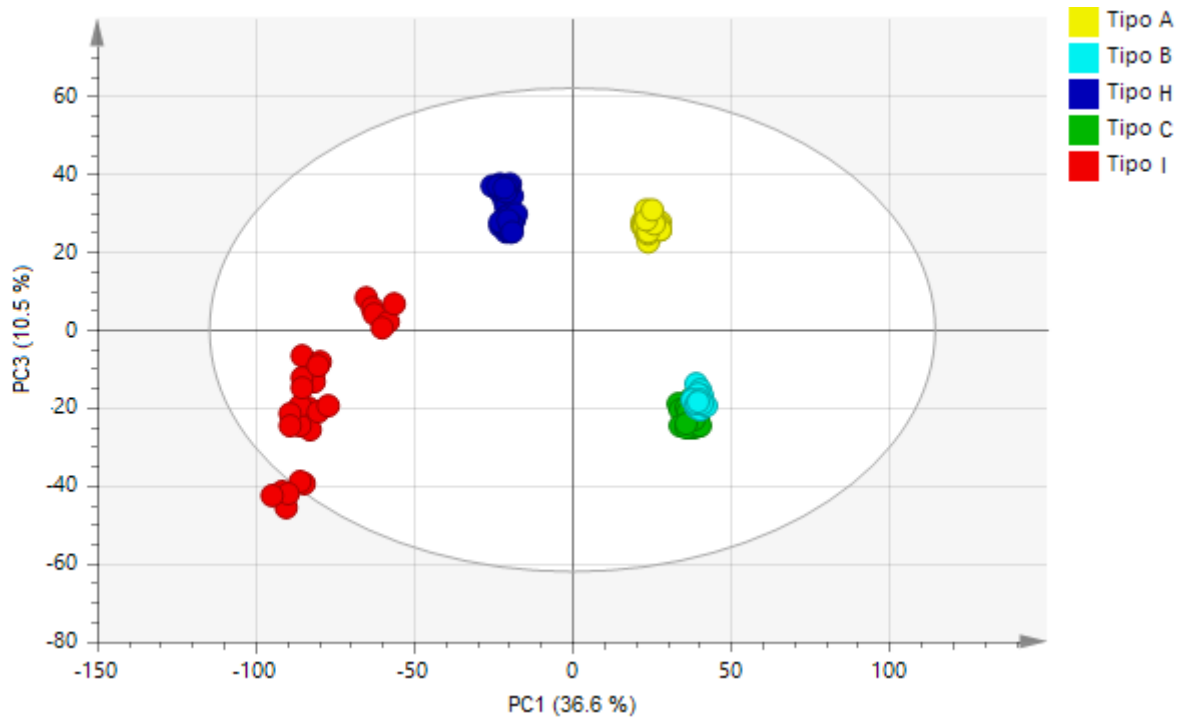


Figura 7 – Escores da PCA realizada com as componentes principais, PC1 com 36,6 % da variância explicada e PC3 com 10,5 %; com elipse de confiança de 95 %, referentes aos tipos de papel das embalagens após receber a impressão, onde os círculos vermelhos representam o tipo de papel I, os azuis escuros o tipo H, os verdes o tipo C, os amarelos o tipo A e os azuis claros o tipo B.

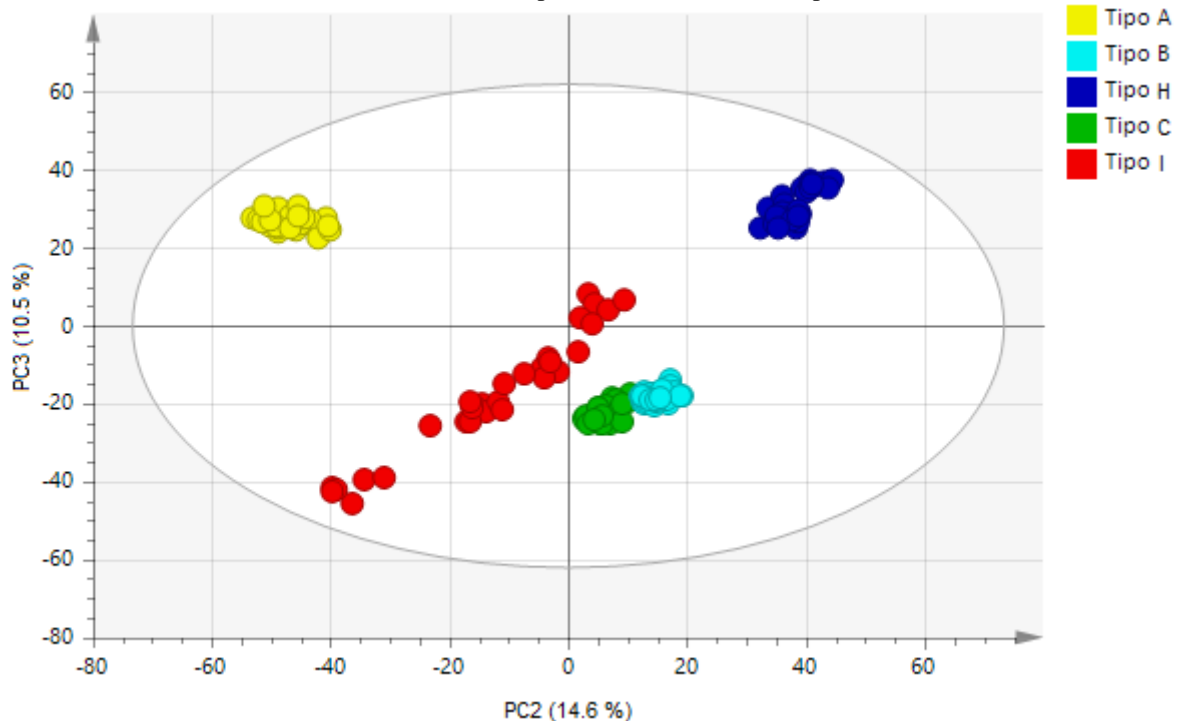


Figura 8 – Gráfico da PCA realizada com as componentes principais, PC2 com 14,6 % da variância explicada e PC3 com 10,5 %; com elipse de confiança de 95 %, apresentando os escores referentes aos tipos de papel das embalagens após receber a impressão, onde os círculos vermelhos representam o tipo de papel I, os azuis escuros o tipo H, os verdes o tipo C, os amarelos o tipo A e os azuis claros o tipo B.

A interpretação das duas primeiras componentes principais para as embalagens que não receberam impressão segue a mesma do caso anterior, ou seja, supõe-se que a PC1 separa os materiais de acordo com suas gramaturas, a PC2 e PC3 como prováveis responsáveis pela segregação dos tipos embalagens, apresentados nas Figuras 9, 10 e 11. Entretanto, com as embalagens de papel em branco a separação entre os tipos B e C foi mais complexa, necessitando da quarta componente principal para separar efetivamente os dois tipos, como mostrado nas Figuras 12 e 13.

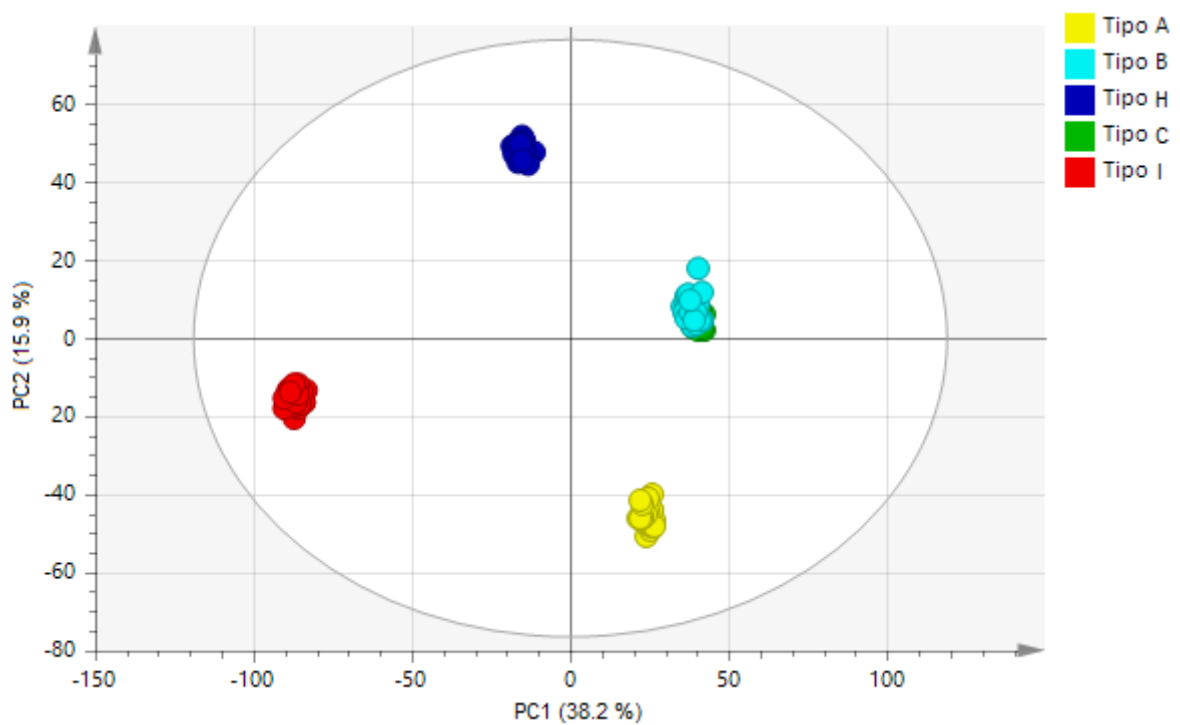


Figura 9 – Gráfico da PCA realizada com as componentes principais, PC1 com 38,2 % da variância explicada e PC2 com 15,9 %; com elipse de confiança de 95 %, apresentando os escores referentes aos tipos de papel das embalagens sem impressão, onde os círculos vermelhos representam o tipo de papel I, os azuis escuros o tipo H, os verdes o tipo C, os amarelos o tipo A e os azuis claros o tipo B.

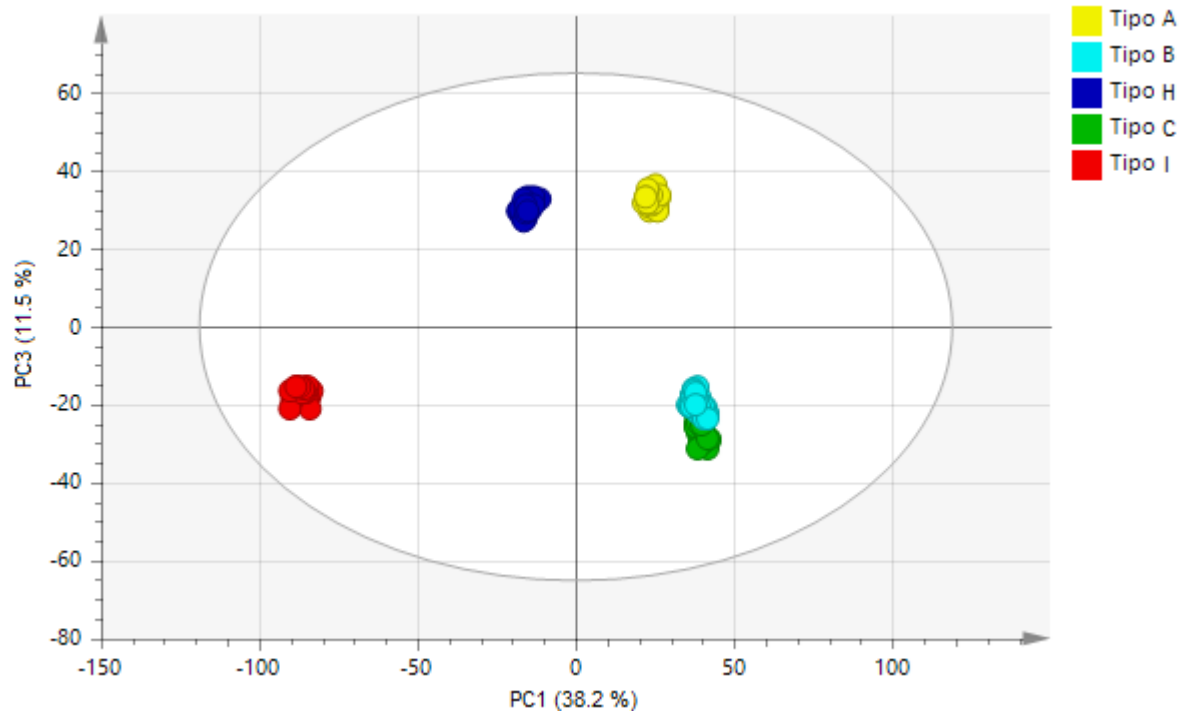


Figura 10 – Gráfico da PCA realizada com as componentes principais, PC1 com 38,2 % da variância explicada e PC3 com 11,5 %; com elipse de confiança de 95 %, apresentando os escores referentes aos tipos de papel das embalagens sem impressão, onde os círculos vermelhos representam o tipo de papel I, os azuis escuros o tipo H, os verdes o tipo C, os amarelos o tipo A e os azuis claros o tipo B.

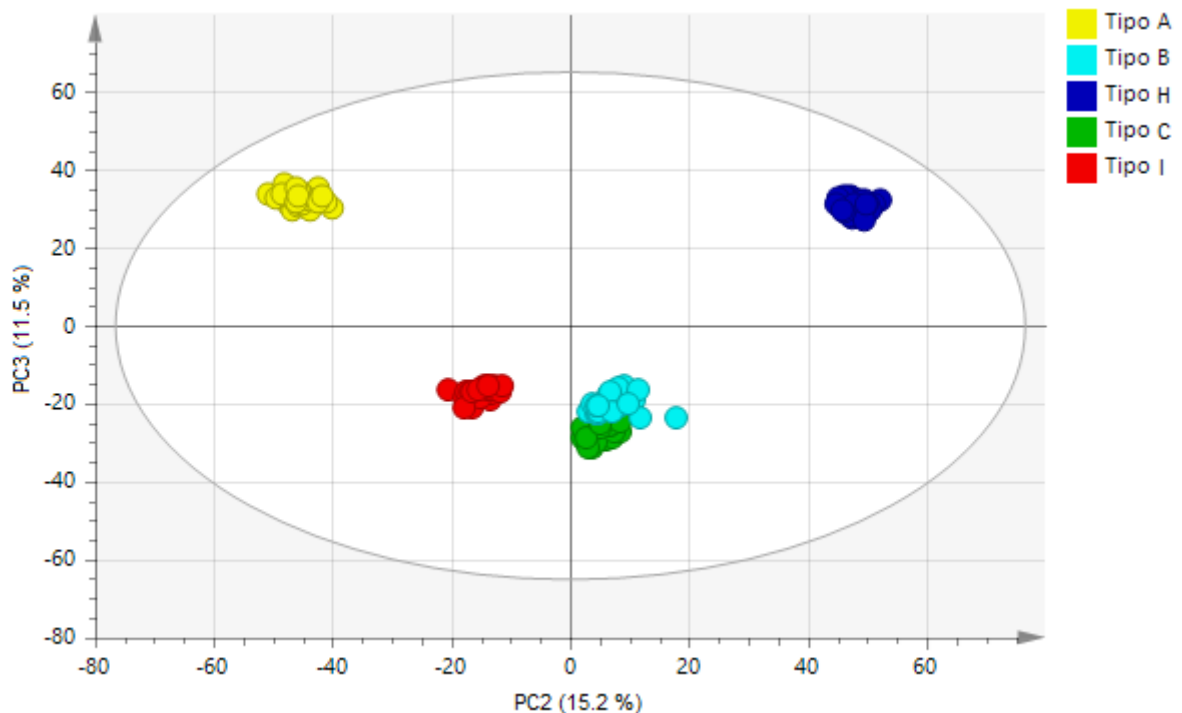


Figura 11 – Gráfico da PCA realizada com as componentes principais, PC2 com 15,2 % da variância explicada e PC3 com 11,5 %; com elipse de confiança de 95 %, apresentando os escores referentes aos tipos de papel das embalagens sem impressão, onde os círculos vermelhos representam o tipo de papel I, os azuis escuros o tipo H, os verdes o tipo C, os amarelos o tipo A e os azuis claros o tipo B.

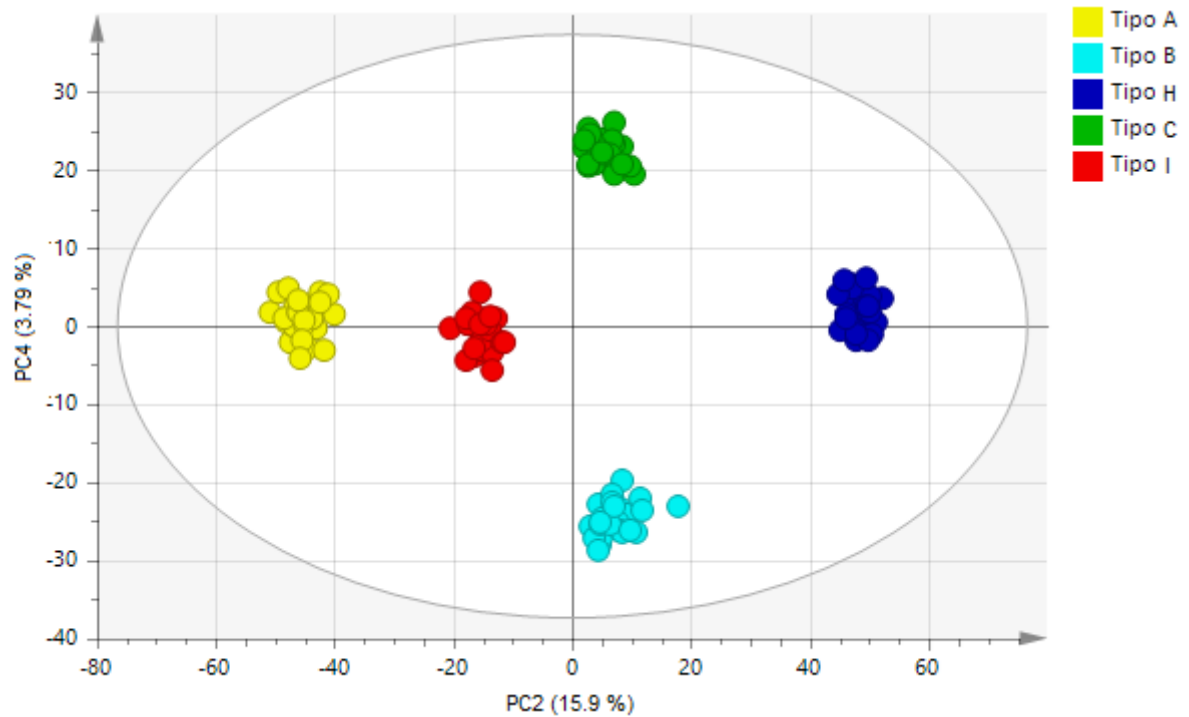


Figura 12 – Gráfico da PCA realizada com as componentes principais, PC2 com 15,9 % da variância explicada e PC4 com 3,79 %; com elipse de confiança de 95 %, apresentando os escores referentes aos tipos de papel das embalagens sem impressão, onde os círculos vermelhos representam o tipo de papel I, os azuis escuros o tipo H, os verdes o tipo C, os amarelos o tipo A e os azuis claros o tipo B.

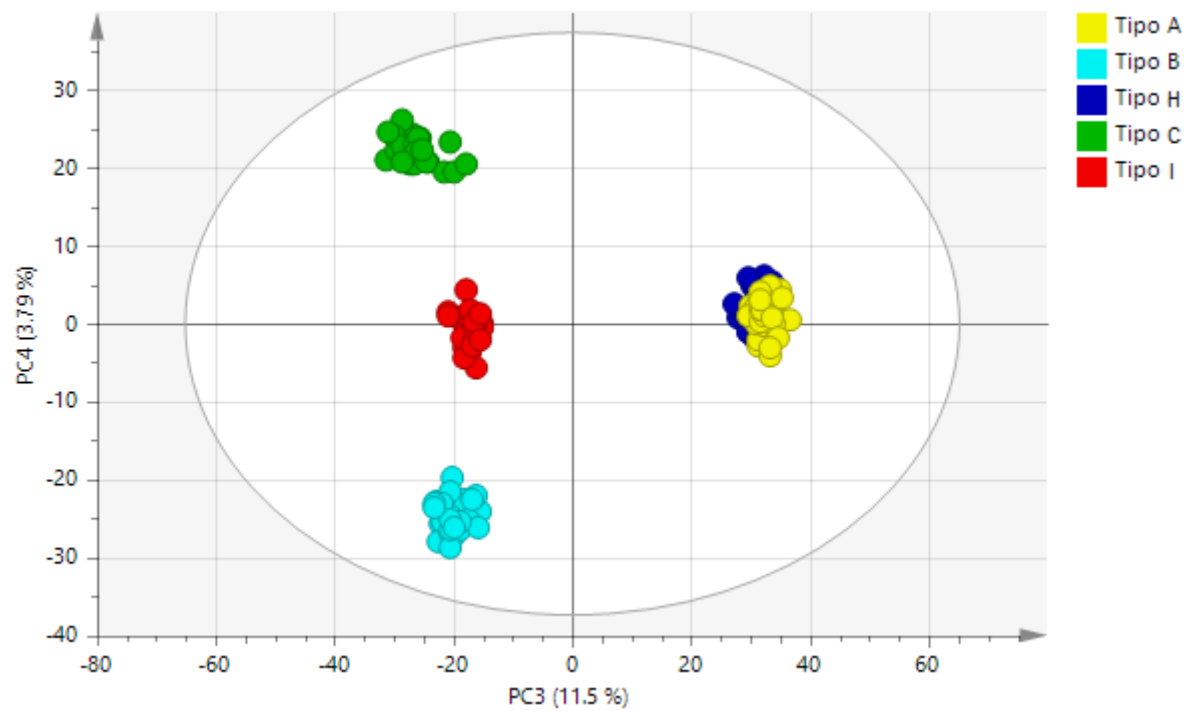


Figura 13 – Gráfico da PCA realizada com as componentes principais, PC3 com 11,5 % da variância explicada e PC4 com 3,79 %; com elipse de confiança de 95 %, apresentando os escores referentes aos tipos de papel das

embalagens sem impressão, onde os círculos vermelhos representam o tipo de papel I, os azuis escuros o tipo H, os verdes o tipo C, os amarelos o tipo A e os azuis claros o tipo B.

Pode-se confirmar a partir das PCAs obtidas que é possível a separação por tipos de papel das embalagens com ou sem impressão, visto que o comportamento das componentes principais é muito semelhante. Para este fim ambos os estágios podem ser utilizados, portanto definiu-se a utilização da embalagem impressa para o método de análise, pois este é o formato que as embalagens são comercializadas, indo de encontro com o objetivo do projeto.

4.2.2 Teste da área de análise

Após definição da utilização da embalagem com impressão para o método de análise, é necessário confirmar que a área a ser analisada pode ser a área interna, ou seja, a face que faz contato com o produto. Para tanto, analisou-se 10 amostras com 3 replicatas do tipo I impresso, onde os espectros foram coletados posicionando primeiramente a área contendo a tinta no leitor e posteriormente 10 amostras com 3 replicatas desta vez analisando a face em branco. A PCA foi realizada utilizando um número de componentes principais igual a 2, apresentada na Figura 14. A PC1 separa com mais de 60 % da variância explicada os escores referentes à área impressa e à área sem impressão, mesmo a embalagem sendo composta exatamente do mesmo tipo de papel. Isso dá indícios de que o pigmento da embalagem exerce muita influência no espectro de infravermelho.

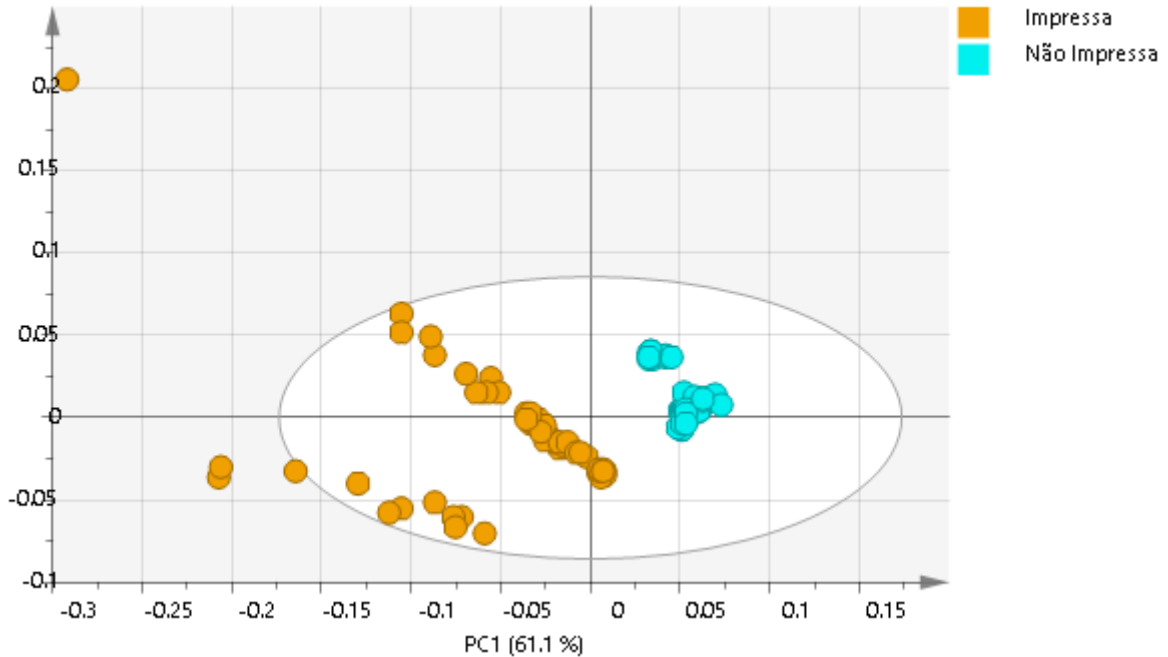


Figura 14 – Gráfico da PCA realizada com as duas primeiras componentes principais, onde PC1 possui 61,1 % da variância explicada e PC2 14,9 %, com elipse de confiança de 95 %, apresentando os escores referentes ao papel do tipo I, onde os círculos amarelos representam a embalagem analisada com a área impressa, e os azuis a embalagem analisada com a área em branco.

Outra informação que pode ser retirada da PCA da Figura 14 é a presença de amostras fora da elipse de confiança, ou seja, demonstrando comportamento anômalo, provavelmente devido à resposta da tinta utilizada na impressão. Para investigação e determinação de que estes escores representam *outliers* foi avaliado os resíduos T^2 de Hotelling e D_{modX} , demonstrado nas Figuras 15 e 16.

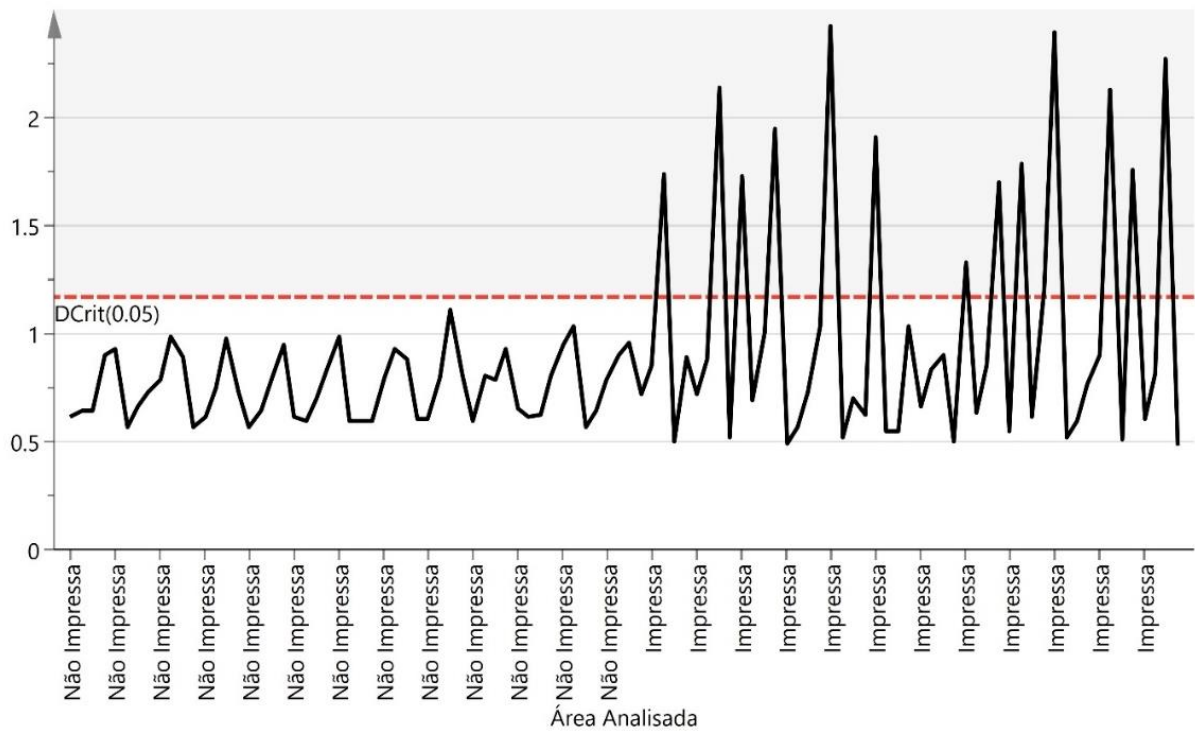


Figura 15 – Gráfico de DmodX com intervalo de confiança de 95 % das amostras analisadas na face impressa e não impressa, demonstrando que as amostras analisadas com a face impressa apresentam resíduo acima do limite crítico estabelecido.

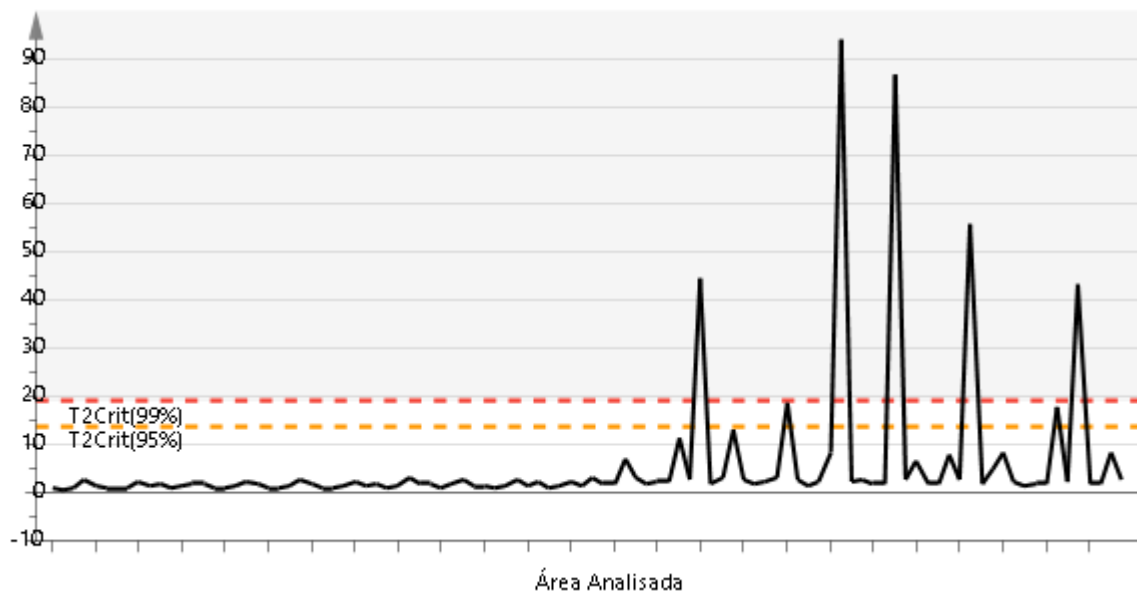


Figura 16 – Gráfico de T^2 de Hotelling com intervalo de confiança de 95 % das amostras analisadas na face impressa e não impressa, demonstrando que as amostras analisadas com a face impressa apresentam resíduo acima do limite crítico estabelecido.

A maioria das amostras que foram analisadas com a superfície impressa posicionada no leitor apresentou resíduo T^2 de Hotelling alto, acima do limite crítico de 95 % e 99 %, algumas amostras inclusive ultrapassando o valor de duas vezes o valor crítico. Portanto, é nítido que através da análise da área contendo tinta há maior risco de amostras com comportamento não conforme, podendo atrapalhar na predição das embalagens. Logo, a face não impressa foi escolhida para o método.

4.2.3 Teste de Homogeneidade

Foram realizadas análises de dois tipos de embalagens de papel (I e B) para avaliar a influência da área de análise direta no espectrômetro de infravermelho. O objetivo foi avaliar a homogeneidade do material impresso e sem impressão além da influência da arte ao se analisar a respectiva área branca da embalagem. Para isto, 3 regiões distintas da embalagem de papel foram realizadas em quintuplicata independente na área não impressa de cada amostra.

Foram realizados quatro conjuntos, de 3 amostras cada, dessa forma: Embalagem de papel do tipo I impresso e não impresso e embalagem de papel do tipo B impresso não impresso. Cada PCA foi realizada utilizando 2 componentes principais, e estão apresentadas abaixo nas Figuras 17, 18, 19 e 20.

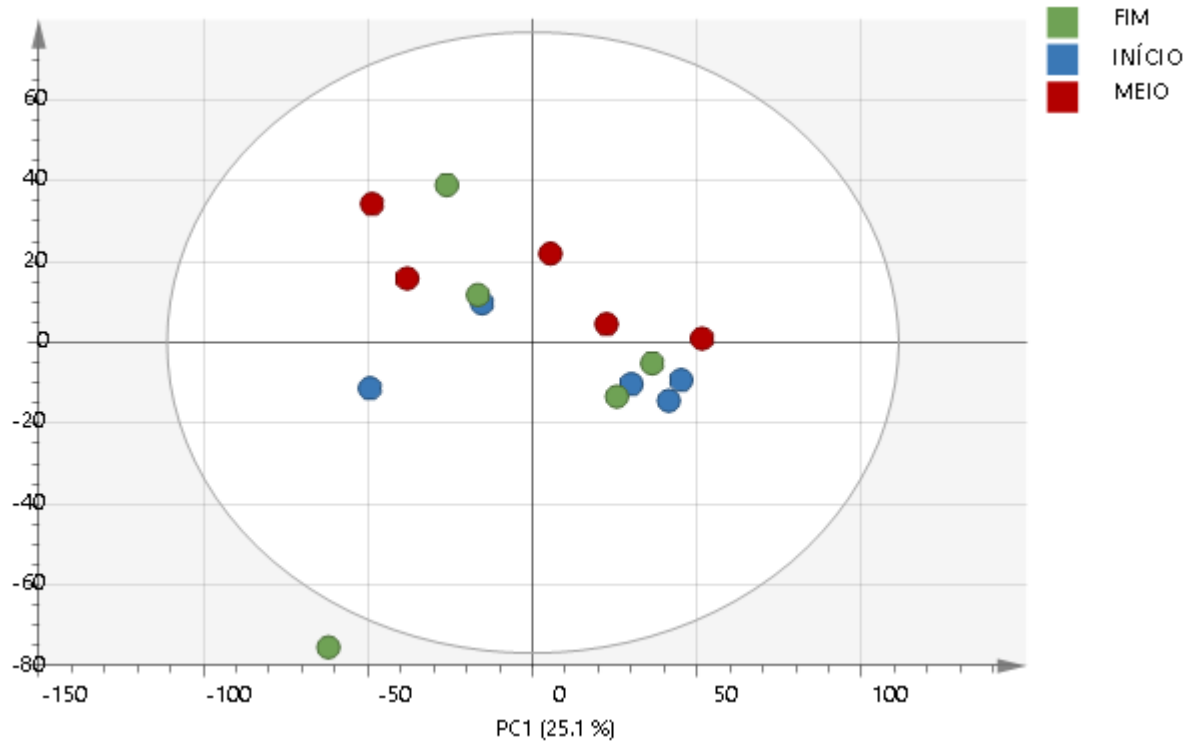


Figura 17 – Gráfico da PCA realizada com as duas primeiras componentes principais, onde PC1 contém 25,1 % da variância explicada e PC2 12,0 %, com elipse de confiança de 95 %, apresentando os escores referentes ao papel I impresso, dividido em 3 partes: Início do rolo, representado pelos escores em azul; Meio do rolo, representado pelos escores em vermelho e Fim do rolo, representado pelos escores em verde.

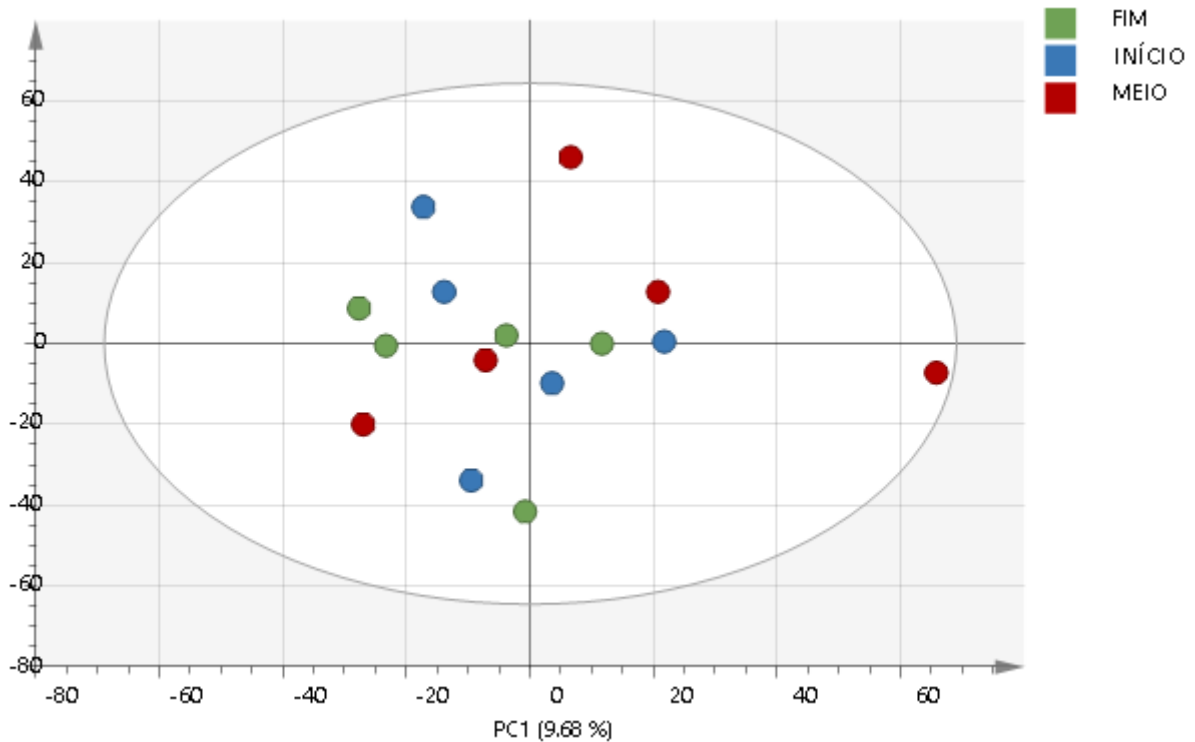


Figura 18 – Gráfico da PCA realizada com as duas primeiras componentes principais, onde PC1 contém 9,68 % da variância explicada e PC2 8,47 %, com elipse de confiança de 95 %, apresentando os escores referentes ao papel I sem impressão, dividido em 3 partes: Início do rolo, representado pelos escores em azul; Meio do rolo, representado pelos escores em vermelho e Fim do rolo, representado pelos escores em verde.

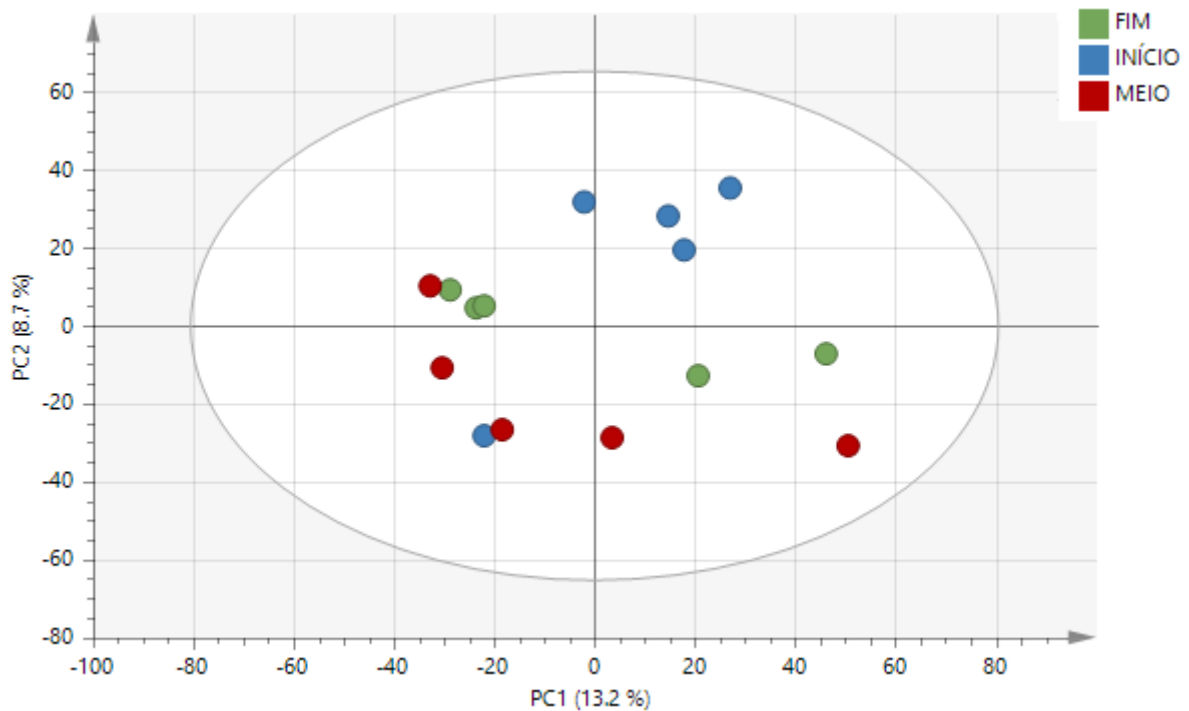


Figura 19 – Gráfico da PCA realizada com as duas primeiras componentes principais, onde PC1 contém 13,2 % da variância explicada e PC2 8,47 % com elipse de confiança de 95 %, apresentando os escores referentes ao papel B com impressão, dividido em 3 partes: Início da embalagem, representado pelos escores em azul; meio da embalagem, representado pelos escores em vermelho e fim da embalagem, representado pelos escores em verde.

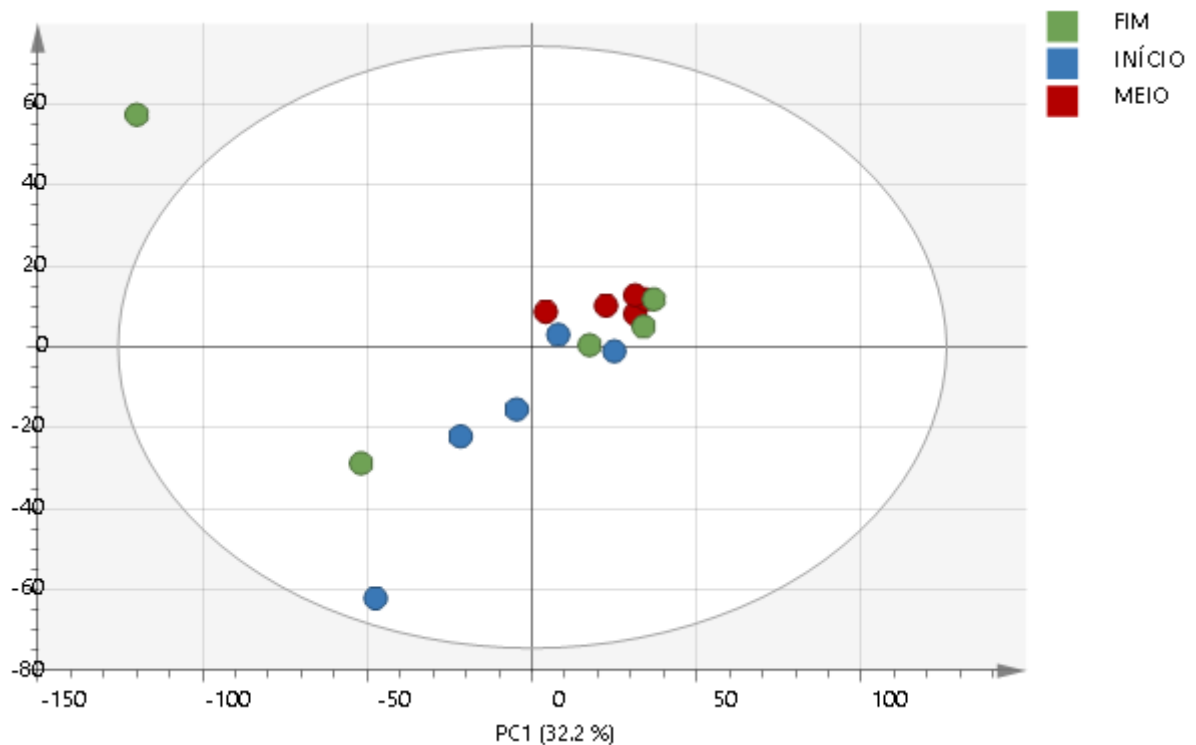


Figura 20 – Gráfico da PCA realizada com as duas primeiras componentes principais, onde PC1 contém 32,2 % da variância explicada e PC2 11,3 % com elipse de confiança de 95 %, apresentando os escores referentes ao papel B sem impressão, dividido em 3 partes: Início da embalagem, representado pelos escores em verde; Meio da embalagem representado pelos escores em azul e Fim da embalagem representado pelos escores em vermelho.

É possível constatar a homogeneidade no papel em branco e impresso devido à inexistência de agrupamentos das amostras do início, meio e fim do rolo de papel para o tipo I, na Figura 18. Já na Figura 19, para o tipo B, ocorre um agrupamento dos escores em azul representado o segmento “INÍCIO”, ou seja, para a embalagem impressa há algumas segregações visíveis entre as três partes do mesmo papel, possivelmente decorrentes da cor de tinta utilizada na outra face do material ou efeito do equipamento. Portanto, definiu-se para análise da embalagem comercial que as leituras sejam feitas em triplicatas, compreendendo as três regiões início, meio e fim.

4.2.4 Teste de dobradura

Testou-se a metodologia de dobrar a embalagem com o objetivo de avaliar se há essa necessidade, visto que a radiação de infravermelho atravessa a embalagem no momento da leitura, o que pode acarretar em perda da sensibilidade da análise. Caso não haja grande variação entre os escores, a etapa de dobradura será dispensada e a embalagem será analisada de forma direta, atribuindo velocidade ao método. Para isso, o tipo I foi utilizado, por ser o de mais baixa gramatura e apresentar maior maleabilidade para dobras.

Foram realizadas leituras com uma, duas, três e quatro dobras ao meio, no mesmo tipo de papel em branco e impresso, e a Figura 21 mostra a PCA realizada para avaliação da metodologia.

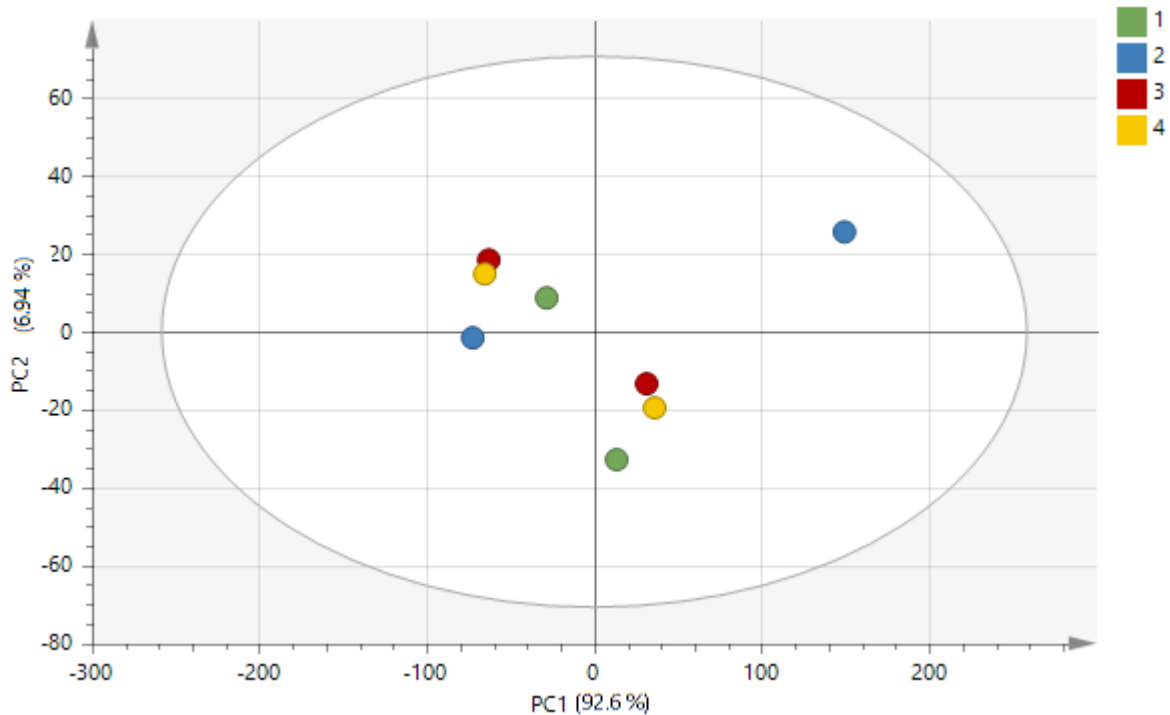


Figura 21 – Gráfico da PCA realizada com as duas primeiras componentes principais, onde PC1 contém 92,6 % da variância explicada e PC2 6,94 %, com elipse de confiança de 95 %, apresentando os escores referentes ao tipo I impresso e não impresso dobrado uma, duas, três e quatro vezes ao meio antes da leitura no equipamento.

O escores em verde representam a embalagem dobrada uma vez, tanto branca quanto impressa, em azul embalagem dobrada duas vezes, tanto branca quanto impressa, em amarelo embalagem dobrada três vezes, tanto branca quanto impressa, e em vermelho embalagem dobrada quatro vezes, tanto branca quanto impressa.

A partir do resultado das PCAs, conclui-se que não há agrupamentos entre a quantidade de dobras, portanto, a necessidade de dobrar o material é desconsiderada.

4.3 ANÁLISE EXPLORATÓRIA

Foram analisadas 485 amostras utilizando a metodologia desenvolvida e descrita anteriormente visando a classificação por tipo de papel de cada embalagem. Primeiramente foi realizada uma análise exploratória por PCA para identificar os principais grupos de amostra antes de realizar o modelo supervisionado. A Figura 22 apresenta os escores nas duas primeiras componentes do PCA, apresentando aproximadamente 50 % da variância total do sistema.

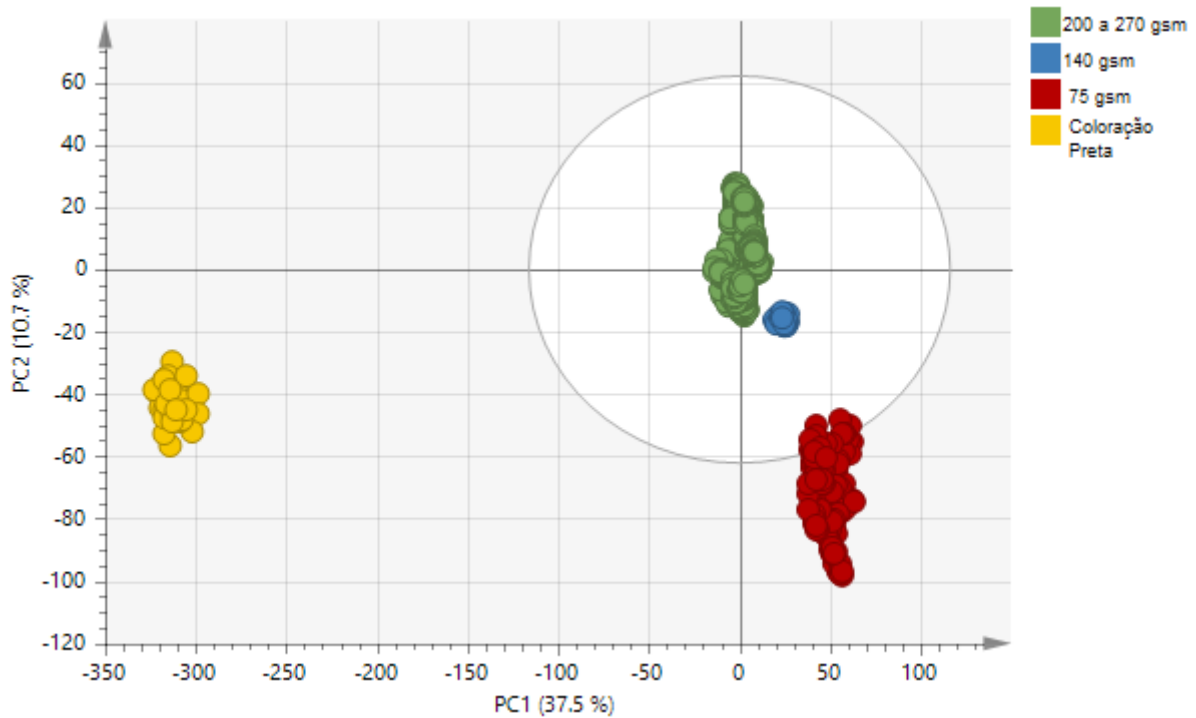


Figura 22 – Gráfico da PCA realizada com as duas primeiras componentes principais, com elipse de confiança de 95 %, apresentando os escores referentes a todos os tipos de embalagens de papel analisados, evidenciando a diferença de coloração, explicada pela PC 1, e a diferença de gramatura, explicada pela PC 2, onde os escores em verde representam o tipo K que possui a coloração preta, os escores em amarelo representam as embalagens de alta gramatura (de 200 gsm a 270 gsm), os escores em vermelho representam as de gramatura intermediária (140 gsm) e por fim os escores em laranja representam as embalagens de papel de fina espessura (75 gsm).

Quatro grupos de embalagens de papel são formados de acordo com a Figura 22, no qual gramatura parece ser um fator determinante para aglomeração destas: um grupo de amostras com elevada gramatura (200 a 270 gsm, amarelo na Figura 22), um grupo com gramaturas intermediárias (150 gsm, vermelho escuro na Figura 22), um grupo com baixa gramatura (75 gsm, laranja na Figura 22) e um grupo de embalagens de papel que se difere do restante. Este grupo, embalagem do tipo K, tem como característica a coloração preta da embalagem de papel na área não impressa e analisada enquanto que as demais embalagens de papel possuem a coloração branca. Esse grupo de coloração preta na área não impressa e analisada difere-se tanto, que se pode considerar a primeira componente apenas para sua separação, contendo assim 37,5 % da variância explicada total do sistema. A separação por gramatura ficou na segunda componente com 10,7 % da variância explicada.

Para facilitar a posterior etapa de classificação e separação das macro diferenças entre as embalagens de papel, foi proposto um fluxo de modelagem com a intenção de separá-los por subgrupos antes de classificá-los por tipo de embalagem de papel. A vantagem deste tipo

de modelagem é a remoção do conjunto de dados amostras muito diferentes evitando assim que sejam considerados *outliers* ou ate mesmo que diminuam a acurácia, sensibilidade e especificidade do modelo para classificação de outras classes.

Tendo isto em vista, o fluxograma da Figura 23 foi proposto para facilitar a classificação posterior: primeiramente o conjunto de dados é submetido a um modelo de classificação de amostras com face interna não impressa de coloração preta contra face interna não impressa de coloração branca; em seguida, as amostras com interior branco não impresso são classificadas de acordo com sua gramatura; e por fim são classificadas de acordo com o seu tipo de embalagem de papel.

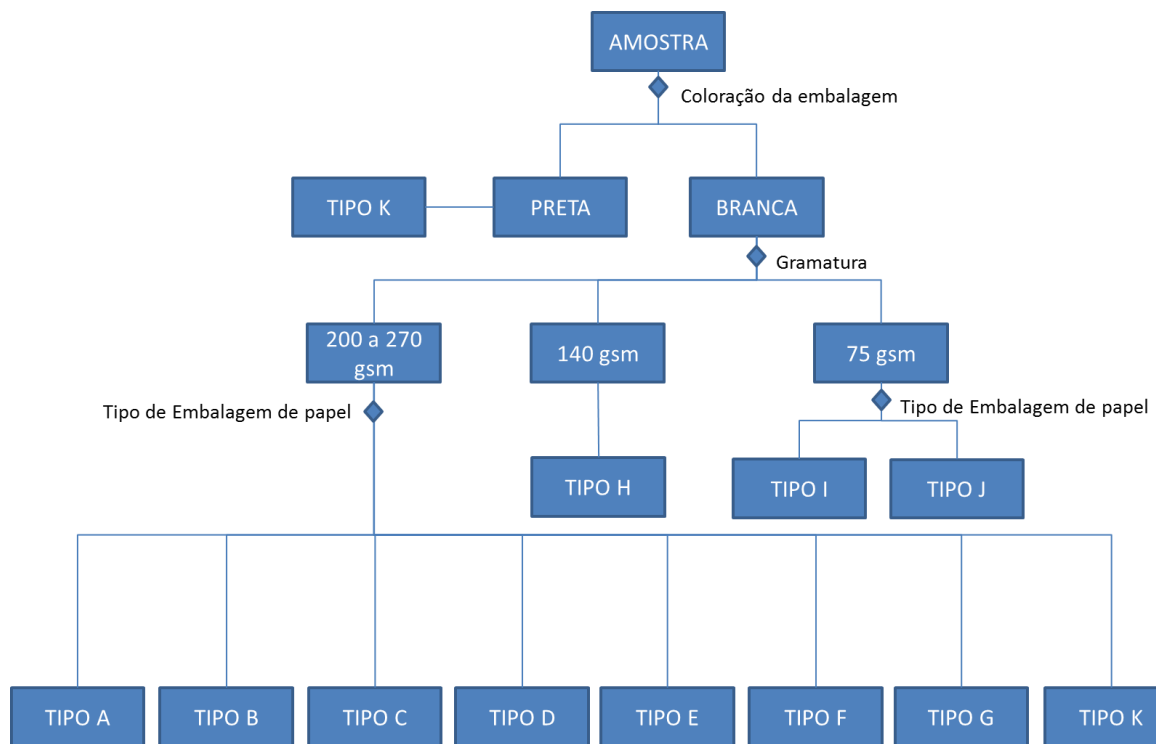


Figura 23 – Fluxograma criado para facilitar a classificação de amostras, subdividindo a classificação em três modelos. O fluxograma mostra a sequência que uma amostra deve ser submetida para uma classificação correta do tipo de embalagem de papel, passando primeiramente pelo modelo de coloração, após pelo modelo de gramatura para então ser classificada como o tipo de papel da embalagem.

4.4 MODELOS DE CLASSIFICAÇÃO

4.4.1 Validação Interna

Foram construídos, ao total, quatro modelos de classificação. Realizaram-se testes de validação interna, utilizando algoritmo de Kennard Stone ²⁵ para escolha das amostras de

validação, separando um conjunto de 177 amostras para predição e 308 amostras de calibração. Cada amostra foi classificada como sendo de um tipo específico de embalagem de papel quando duas ou mais replicatas foram preditas em uma mesma classe. Ou seja, se cada uma das três replicata fosse classificada como um tipo diferente, o resultado seria inconclusivo.

Resumidamente, a classificação das amostras de predição ocorre primeiramente categorizando a cor de seu material; se esta amostra for identificada como branca, ela será classificada novamente no segundo modelo, especificando se possui gramatura de 200 gsm a 270 gsm ou de 140 gsm ou de 75 gsm; se ela for classificada como, por exemplo, 75 gsm, passará pela classificação de tipo de papel, e então poderá ser finalmente definida como tipo I ou tipo J.

A classificação empregando PLS-DA, com 4 variáveis latentes, no primeiro modelo (Tabela 3) cujo objetivo é a segregação de embalagens de papel de coloração preta daquelas de coloração branca, foi 100 % correta, com 97,4 % de variância explicada em **X** e 97,1 % em **Y**, apresentando acurácia, sensibilidade e especificidade igual a 100 %, pois não houve nenhum FP e FN apenas VP e VN. Os valores críticos de resíduo para T^2 de Hotelling e $DmodX$, foram de 11,43 e 1,12, respectivamente, com intervalo de confiança de 95 %, e nenhuma amostra foi considerada como *outlier*. Como visualizado na PCA da Figura 22, as amostras de coloração preta na parte interna se diferenciam muito do restante, visto que o infravermelho é uma técnica de análise que responde muito bem à cor dos materiais, portanto uma boa classificação era esperada para este modelo, e a expectativa foi cumprida.

Tabela 3 – Matriz de confusão do modelo de classificação de coloração do material, demonstrando classificação correta para todas as amostras do conjunto de validação externa.

	Classe verdadeira		Classe predita	
	Amostras (Un)	Correto	Branca	Preta
Branca	171	100 %	171	0
Preta	6	100 %	0	6
Total	177	100 %		

O segundo modelo (Tabela 4), com 6 variáveis latentes e variância explicada em **X** de 96,7 % e em **Y** de 82 %, separa as amostras baseando-se nas 3 classes de gramatura do material, a primeira classe que engloba gramaturas de 200 a 270 gsm, a segunda de 140 gsm e a terceira de 75 gsm. Obteve-se igualmente a classificação 100 % correta, novamente com

acurácia de 100 %, sensibilidade de 100 % e especificidade igual a 100 %. Os valores críticos de resíduo para T^2 de Hotelling e D_{modX} , foram de 12,72 e 1,12, respectivamente, com intervalo de confiança de 95 %, e nenhuma amostra foi considerada como *outlier*. Assim como no caso das amostras com coloração, a diferença de gramatura também pode ser visualizada pela PCA da Figura 22, trazendo uma expectativa de classificação correta. As diferenças de gramatura são capturadas pela técnica de infravermelho próximo com reflectância total atenuada devido a maior absorção das amostras com maior gramatura, pois menos radiação chega no detector nesse caso, diferentemente das amostras com gramatura mais fina, onde o feixe atravessa com maior facilidade e mais radiação vai para o detector. As amostras apresentam gramaturas com diferenças de aproximadamente o dobro, portanto, é possível observar a separação das três classes de gramaturas com clareza.

Tabela 4 – Matriz de confusão do modelo de classificação de gramatura das embalagens de papel, demonstrando classificação correta para todas as amostras do conjunto de validação externa, previamente classificadas como coloração branca.

Classe verdadeira			Classe predita		
			200 a 270 gsm	140 gsm	75 gsm
Amostras (Un)	Correto				
200 a 270 gsm	153	100 %	153	0	0
140 gsm	4	100 %	0	4	0
75 gsm	14	100 %	0	0	14
Total	171	100 %			

O terceiro, indicado na Tabela 5 e o quarto modelo indicado na Tabela 6 classificam o tipo de embalagem de papel, e novamente todas as amostras foram precisamente classificadas, com 4 e 11 variáveis latentes, respectivamente. Mais uma vez, sensibilidade, acurácia e especificidade foram de 100 %. O terceiro modelo, para embalagens com espessura de 200 gsm a 240 gsm, apresentou variância explicada de 97,7 % em X e 82,4 % em Y , enquanto que o quarto modelo, para embalagens de 75 gsm, teve variância explicada de 99 % em X e 79 % em Y . Os valores críticos de resíduo para T^2 de Hotelling e D_{modX} , foram de 25,55 e 1,13, respectivamente, para o terceiro modelo e de 10,38 e 1,18, sendo T^2 de Hotelling e D_{modX} respectivamente, para o quarto modelo, com intervalo de confiança de 95 %, e nenhuma amostra foi considerada como *outlier*. Não foi construído um modelo específico para embalagens de gramatura 140 gsm, visto que esta não possui variedade o suficiente para a

construção de um modelo. Logo, toda amostra classificada como 140 gsm é automaticamente do tipo H.

Através da análise exploratória realizada, percebe-se que há diferenciação entre tipos de papéis da mesma gramatura pela PCA, e o mesmo comportamento foi observado para a PLS-DA. Possivelmente, essa segregação ocorre devido às diferentes composições químicas de cada tipo de papel utilizado nas embalagens. Um dos motivos pode ser a relação de fibras curtas e fibras longas da celulose, que é um polímero de tamanho de cadeia variável de acordo com o seu tratamento e da planta que é proveniente, utilizado na receita do material. Outra possível razão é o aglutinante empregado, que pode ser lignina ou qualquer outro produto químico que absorva no infravermelho, apresentando influência no espectro da amostra. Qualquer tratamento químico de branqueamento ou adição de compostos que confirmam durabilidade ao material também podem ser fatores influentes na classificação dos diferentes tipos de embalagens de papel.

Tabela 6 – Matriz de confusão do modelo de classificação de tipos das embalagens de papel, demonstrando classificação correta para todas as amostras do conjunto de validação externa, previamente classificadas como coloração branca e gramatura de 75 gsm.

	Classe verdadeira		Classe predita	
	Amostras (Un)	Correto	Tipo I	Tipo J
Tipo I	9	100 %	9	0
Tipo J	5	100 %	0	5
Total	14	100 %		

Observa-se que todos os modelos possuem boa capacidade preditiva, considerando os resultados de validação interna, visto que nenhuma amostra foi classificada erroneamente, ratificando que o fluxo de modelagem proposto e validado através do algoritmo de Kennard Stone foi bem sucedido. Portanto, devido aos indícios de que o modelo é capaz de prever novas amostras, sua aplicação para amostras reais, ou seja, comerciais, é possível.

4.4.2 Validação externa

Além da validação interna, foi realizada uma validação externa com um conjunto de onze amostras comerciais, identificadas sequencialmente de 1 a 11. As amostras foram abertas com cuidado para não rasgar nem comprometer a área necessária para a leitura, então o produto foi retirado e a amostra de embalagem seguiu para a análise.

Todas as amostras tinham o interior branco, as gramaturas de 200 gsm a 270 gsm foram identificadas nas amostras 1, 2, 3, 4, 6, 8, 9, 10 e 11, e 75 gsm nas amostras 5 e 7. Apesar de serem amostras comerciais, os tipos de papel utilizados nas embalagens é conhecido, sendo as amostras 2, 3 e 10 do tipo A, as amostras 8, 9 e 11 do tipo B, a amostra 1 do tipo E, a 5 tipo H, a 7 tipo Y e as amostras 4 e 6 do tipo X. Os tipos X e Y não fazem parte do conjunto de calibração do modelo de classificação.

O fluxograma foi seguido, todas as amostras foram classificadas corretamente quanto à coloração da parte interna e às gramaturas. Todavia, quando preditas pelo modelo de tipo de embalagem de papel, as amostras 4, 6 e 7 apresentaram resíduos estatísticos acima dos valores críticos de T^2 de Hotelling de 25,55 para gramatura de 200 gsm a 270 gsm e 10,38 para 75 gsm, além do valor crítico de DmodX, de 1,13 para 200 gsm a 270 gsm e 1,18 para 75 gsm. A amostra 4 apresentou T^2 de Hotelling e DmodX médios de 42,93 e DmodX de 1,99, quando classificada no modelo de tipo de embalagem de 200 gsm a 270 gsm; a amostra 6 apresentou T^2 de Hotelling

e DmodX médios de 19,59 e 2,37, respectivamente; a amostra 7 apresentou T^2 de Hotelling e DmodX médios de 31,66 e 3,53, respectivamente, e ambas (6 e 7) foram preditas no modelo de tipo de embalagem de 75 gsm. Apesar de não serem classificadas como *outliers*, visto que foi estipulado que uma amostra seria classificada como tal quando seu resíduo estatístico ultrapassasse duas vezes o valor crítico de T^2 e DmodX, essas amostras apresentam resíduo alto e acima do valor limite, enquanto as outras amostras (1, 2, 3, 5, 8, 9, 10 e 11) não apresentam, ou seja, considera-se que as amostras são suspeitas e podem não ser classificadas com certeza.

A Tabela 7 relaciona os tipos reais de embalagem com os tipos preditos pelo modelo de classificação. Mais uma vez demonstrando que possui acurácia, sensibilidade e especificidade de 100 %.

Tabela 7 – Resultados da predição das amostras de 1 a 11 referentes à validação externa do modelo de classificação de embalagens de papéis, mostrando que 8 das 11 amostras foram classificadas corretamente, enquanto que as outras 3 apresentaram resíduos estatísticos altos devido ao seu material não estar contido na calibração do modelo.

Identificação da Amostra	Tipo Predito	Tipo Verdadeiro
Amostra 1	E	E
Amostra 2	A	A
Amostra 3	A	A
Amostra 4	-	X
Amostra 5	H	H
Amostra 6	-	X
Amostra 7	-	Y
Amostra 8	B	B
Amostra 9	B	B
Amostra 10	A	A
Amostra 11	B	B

Os resultados obtidos quanto à classificação correta das amostras cujo tipo de embalagem era conhecido pelo modelo e quanto à identificação através dos resíduos estatísticos daquelas amostras que eram desconhecidas, validam o modelo para predição de embalagens de papel comerciais e isso demonstra o potencial da ferramenta desenvolvida.

5 CONCLUSÕES

O método empregando espectroscopia de infravermelho próximo com reflectância total atenuada associada às técnicas quimiométricas de análise multivariada PCA e PLS-DA demonstrou grande eficácia para classificação de 11 diferentes tipos de papéis. Além dessa classificação, é possível observar a diferenciação entre coloração e gramaturas dos materiais.

O modelo construído empregando PLS-DA demonstrou bom desempenho para classificação de amostras comerciais, como comprovado pela validação interna e externa. Logo, mesmo que a calibração não tenha sido realizada com embalagens comerciais já com o produto acabado em seu interior, a metodologia apresenta aplicabilidade muito satisfatória para estas.

Foi desenvolvida uma ferramenta de análise rápida, barata, robusta e que dispensa preparo de amostra, apresentando acurácia, sensibilidade e especificidade de 100 %, cumprindo o objetivo proposto para o trabalho e apresentando vantagens econômicas e de grande valia para as indústrias com foco em desenvolvimento de tecnologia 4.0.

Sugere-se, para a continuidade do trabalho, a análise de amostras da gramatura 140 gsm, do papel de coloração preta, visto que estas classes não possuíam amostras o suficiente para a construção de um modelo de classificação e das amostras que foram identificadas como fora do escopo, para que na próxima aplicação já estejam dentro das amostras de calibração.

A aplicabilidade dos modelos de classificação depende dos tipos de embalagens de papeis contidos na sua calibração, e necessita de uma atualização para manter a representatividade das amostras. Logo, como perspectiva, novas amostras de classes dentro e fora do escopo devem ser continuamente acrescentadas para manutenção das figuras de mérito dos modelos apresentados.

REFERÊNCIAS

- 1 RILEY, A. Paper and paperboard packaging. *In*: Emblem, A.; Emblem, H. **Packaging technology: Fundamentals, materials and process**. UK: Woodhead Publishing Limited, 2012. Cap. 10, p. 178-239.
- 2 BRITT, K. W. Papermaking. **Enciclopédia Britannica**, Nova York. Disponível em: <<https://www.britannica.com/technology/papermaking/Paper-properties-and-uses> > Acesso em 02 abr. 2021.
- 3 EMBRAPA - Empresa Brasileira de Pesquisa Agropecuária. Brasília. Disponível em: <https://www.embrapa.br/contando-ciencia/arvores/-/asset_publisher/Zd2bjD3HpAAC/content/sabe-quantas-arvores-sao-usadas-para-fazer-papel> Acesso em: 03 abr. 2021.
- 4 EMBLEM, A. Packaging and Society. *In*: EMBLEM, A.; EMBLEM, H. **Packaging technology: Fundamentals, materials and process**. UK: Woodhead Publishing Limited, 2012. Cap. 1, p. 3-9.
- 5 KAROUI, R.; MAZEROLLES, D.; DUFOUR, E. Spectroscopic techniques coupled with chemometric tools for structure and texture determinations in dairy products. **International Dairy Journal**, v. 13, n. 8, p. 607-620, 2003.
- 6 MARCELO, C. A. M. **Identificação da procedência geográfica da erva mate (*Ilex paraguariensis*) mediante análise elementar, reflexão no infravermelho próximo e quimiometria**. 2013. 103 f. Dissertação (Mestrado em Química), Universidade Federal do Rio Grande do Sul, Porto Alegre, 2013.
- 7 SILVERSTEIN, R. M. et al. Espectrometria de infravermelho. *In*: SILVERSTEIN, R. M. et al. **Identificação espectrométrica de compostos orgânicos**. 7. ed. Rio de Janeiro: LTC, 2010. Cap. 2, p. 70-104.
- 8 HOLLER, F. J.; SKOOG, D. A.; CROUCH, S. R. **Princípios de Análise Instrumental**. 6. ed. Porto Alegre: Bookman, 2009. 1055 p.
- 9 BURNS, D. A.; CIURCZAK, E. W. **Handbook of near-infrared analysis**, 3. ed. Boca Raton: CRC Press, 2009. 836 p.
- 10 SOUZA, A. M. Experimento didático de quimiometria para calibração multivariada na determinação de paracetamol em comprimidos comerciais utilizando espectroscopia no infravermelho próximo: um tutorial, parte II. **Química Nova**, v. 36, n. 7, p. 1057-1065, 2013.
- 11 FERRÃO, M. F. **Aplicação de técnicas espectroscópica de reflexão no infravermelho no controle de qualidade de farinha de trigo**. 2000. 139 f. Tese (Doutorado em Ciências), Universidade Estadual de Campinas, Campinas, 2000.

- 12 SOUZA, A. M. de; POPPI, R. J. Experimento didático de quimiometria para análise exploratória de óleos vegetais comestíveis por espectroscopia no infravermelho médio e análise de componentes principais: um tutorial, parte I. **Química Nova**, v. 35, n. 1, p. 223-229, 2012.
- 13 BREITKREITZ, M. C.; SOUZA, A. M. de; POPPI, R. J. Experimento didático de quimiometria para planejamento de experimentos: avaliação das condições experimentais na determinação espectrofotométrica de ferro II com *o*-fenantrolina. Um tutorial, parte III. **Química Nova**, v. 37, n. 3, p. 564-573, 2014.
- 14 NESPECA, G. M. **Aplicação de espectroscopia no infravermelho próximo e análise multivariada para identificação e quantificação de hidrocarbonetos totais do petróleo em solo**. 2018. 218 f. Tese (Doutorado em Química), Universidade Estadual Paulista, Araraquara, 2018.
- 15 RINNAN, A.; van den BERG, F.; ENGELSEN, S. B. Review of the most common pre-processing techniques for near-infrared spectra. **TrAC Trends in Analytical Chemistry**, v. 28, n. 10, p. 1201-1222, 2009.
- 16 SIMCA® 15 User guide - Sartorius Stedim Data Analytics AB, 2020.
- 17 XIAOBO, Z. et al. Variables selection methods in near-infrared spectroscopy. **Analytica Chimica Acta**, v. 667, n. 1-2, p. 14-34, 2010.
- 18 BRO, R.; SMILDE, A. K. Principal component analysis. **Analytical Methods**, v. 6, p. 2812-2831, 2014.
- 19 TEÓFILO, R. F. **Métodos Quimiométricos: Uma Visão Geral - Conceitos básicos de quimiometria**, Universidade Federal de Viçosa, Viçosa, v. 1, 2013.
- 20 XIA, J. et al. Fourier transform infrared spectroscopy and chemometrics for the discrimination of paper relic types. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**, v. 219, n. 5, p. 8-14, 2019.
- 21 BALLABIO, D.; CONSONNI, V. Classification tools in chemistry. Part 1: linear models.PLS-DA. **Analytical Methods**, v. 5, p. 3790-3798, 2013.
- 22 BRERETON, R. G.; LLOYD, G. R. Partial least squares discriminant analysis: taking the magic away. **Journal of Chemometrics**, v. 28, n. 4, p. 213-225.
- 23 ASSIS, C. **Previsão do teor de lignina em cana-de-açúcar usando espectroscopia no infravermelho próximo e métodos quimiométricos**. 2014. 88 f. Dissertação (Mestrado em Agroquímica), Universidade Federal de Viçosa, Minas Gerais, 2014.
- 24 WILCOX, R. R. Comparing two independent groups. *In*: WILCOX, R. R. **Applying Contemporary Statistical Techniques**. Academic Press, p. 237-284, 2003.

25 KENNARD, R. W.; STONE, L. A. Computer aided Design of Experiments. **Technometrics**, v. 11, p. 137-148, 1969.

26 JIN, X.; KAMDEN, P. D. Chemical composition, crystallinity and crystallite cellulose size in *Populus* hybrids and aspen. **Cellulose Chemistry and Technology**, v. 43, n. (7-8), 229-234, 2009.

27 ELENA, U. et al. Chemical modification and characterization of straw lignin. **Cellulose Chemistry and Technology**, v. 43, n. (7-8), p. 7-8, 2009.