



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

ERKLAYLLE GABRIELY CUSTODIO DA SILVA

**AVALIAÇÃO DO ENVELHECIMENTO DO PAPEL PARA DETERMINAR
FALSIFICAÇÕES DE DOCUMENTOS UTILIZANDO ESPECTROSCOPIA
VIBRACIONAL E TÉCNICAS QUIMIOMÉTRICAS**

Recife

2020

ERKLAYLLE GABRIELY CUSTODIO DA SILVA

**AVALIAÇÃO DO ENVELHECIMENTO DO PAPEL PARA DETERMINAR
FALSIFICAÇÕES DE DOCUMENTOS UTILIZANDO ESPECTROSCOPIA
VIBRACIONAL E TÉCNICAS QUIMIOMÉTRICAS**

Dissertação apresentada ao Programa de Pós-Graduação em Química Fundamental, PPGQ, da Universidade Federal de Pernambuco, como requisito parcial para obtenção do título de Mestre em Química.

Área de concentração: Química Analítica

Orientadora: Prof^a Dr^a Maria Fernanda Pimentel Avelar

Coorientadora: Dr^a. Carolina Santos Silva

Recife
2020

Catálogo na fonte
Bibliotecária Nataly Soares Leite Moro, CRB15-861

- S586a Silva, Erklaylle Gabriely Custodio da
Avaliação do envelhecimento do papel para determinar falsificações de documentos utilizando espectroscopia vibracional e técnicas quimiométricas / Erklaylle Gabriely Custodio da Silva. – 2020.
110 f.: il., fig., tab.
- Orientadora: Maria Fernanda Pimentel Avelar.
Dissertação (Mestrado) – Universidade Federal de Pernambuco. CCEN, Química, Recife, 2020.
Inclui referências e apêndices.
1. Química analítica. 2. Envelhecimento. 3. Infravermelho. 4. Papel. I. Avelar, Maria Fernanda Pimentel (orientadora). II. Título.

543

CDD (23. ed.)

UFPE- CCEN 2021 - 134

ERKLAYLLE GABRIELY CUSTODIO DA SILVA

AVALIAÇÃO DO ENVELHECIMENTO DO PAPEL PARA DETERMINAR
FALSIFICAÇÕES DE DOCUMENTOS UTILIZANDO ESPECTROSCOPIA
VIBRACIONAL E TÉCNICAS QUIMIOMÉTRICAS.

Dissertação apresentada ao
Programa de Pós-Graduação no
Departamento de Química
Fundamental da Universidade
Federal de Pernambuco, como
requisito parcial para a obtenção do
título de Mestre em Química.

Aprovada em: 21/10/2020

BANCA EXAMINADORA

Profa. Maria Fernanda Pimentel Avelar (Orientadora)

Universidade Federal de Pernambuco
Departamento de Engenharia Química

Prof. José Licarion Pinto Segundo Neto

Departamento de Química Fundamental
Universidade Federal de Pernambuco

Prof. André Filipe dos Ramos Martins Braz

Slippery Rock University

Profa. Simone da Silva Simões

Departamento de Química
Universidade Estadual da Paraíba

Aos meus pais e minhas orientadoras, por todo incentivo, ensinamento e dedicação.

AGRADECIMENTOS

Primeiramente gostaria de agradecer a minha professora e orientadora Maria Fernanda Pimentel Avelar, pela orientação, paciência, dedicação e conselhos, por todas reuniões e discussões. À minha coorientadora Carolina Santos Silva que me ensinou a dar os primeiros passos na área de pesquisa, pela paciência, dedicação, conselhos, confiança, suporte emocional e por estar sempre de portas abertas para me receber. Gostaria de agradecer a essas duas mulheres que são um exemplo de profissional que tem o amor e a dedicação de passar seus ensinamentos e é uma inspiração para todos a sua volta.

À FACEPE, pela bolsa concedida. E aos órgãos de apoio financeiro NEQUIFOR – CAPES/PRO-FORENSE, ao NUQAAPE/FACEPE, ao INCTAA e ao CNPq. À Universidade Federal de Pernambuco e à Polícia Científica de Pernambuco pelo apoio institucional. Em especial ao Laboratório de Combustíveis e todos os seus integrantes que possibilitaram o desenvolvimento desse trabalho.

Aos meus amigos Aline, Bruno, Flávio, Fernando e os demais amigos do LAC, por todas as conversas e discussões, por todo aprendizado e ensinamento, pelas horas de estudo em grupo. Mas também pelos desabaços, risadas, brincadeiras e por todas as experiências que vivemos juntos. A minha amiga Flávia que estar comigo desde a graduação, pela paciência, pela força durante as aulas que passamos juntas, pelo apoio, pelas conversas e conselhos. Ao meu amigo Matheus pelas noites de estudos passadas em claro, pelo apoio, pelas conversas e por todo suporte quando mais precisei. Acredito que Deus coloca pessoas certas na hora certa, e Agradeço a Ele por me permitir viver essa experiência com vocês.

Gostaria agradecer aos meus pais, Eliene e Gercino, por todo incentivo, apoio, paciência, dedicação e por todo carinho e compreensão que tornaram possível obter essa conquista. A minha irmã que sempre esteve ao meu lado me incentivando e torcendo por mim, obrigada maninha pelas noites me esperando, pelas comidas levada na cama e por sempre acreditar em mim. A minha tia, Elizabeth, por todas as noites passadas em claro, por todas as orações, pela torcida e por sempre me apoiar quando mais preciso, agradeço por ser minha segunda mãe. A Júnior Santos pelo companheirismo, compreensão, carinho, conselhos, incentivo e amor, por sempre estar ao meu lado e torcer por mim.

“[...] We must have perseverance and above all confidence in ourselves. We must believe that we are gifted for something and that this thing must be attained.”
(CURIE, 2015, p.1)

RESUMO

Documentos questionados são aqueles enviados para análise forense, principalmente quando há questionamento judicial. Os resultados dessas análises devem ser objetivos e confiáveis para evitar a punição de inocentes. Por isso é de fundamental importância o estudo da autenticação do papel. A composição do papel é bastante complexa, pois é composto por uma mistura de substâncias e sua degradação pode ser influenciada por diversos fatores. Este trabalho tem como objetivo desenvolver uma metodologia não destrutiva, utilizando a espectroscopia de infravermelho associadas a técnicas quimiométricas, para avaliar o envelhecimento artificial e o envelhecimento natural do papel. Para a realização desse estudo foram utilizados dois conjuntos de amostras. O primeiro conjunto, utilizado no envelhecimento artificial, constou de 10 diferentes tipos e marcas de papéis, submetidas à condição de envelhecimento artificial por aquecimento a 100 °C, por períodos de tempo, que variaram de 10 minutos até 4 meses. O segundo conjunto, utilizado no envelhecimento natural, foi composto de documentos datados de 1998 a 2018. Utilizando análise de componentes principais (PCA), foi possível identificar diferenças entre as marcas dos papéis, além da oxidação da hidroxila da celulose como principal processo de degradação. Foram desenvolvidos modelos de regressão por mínimos quadrados parciais (PLS) para as amostras de ambos os conjuntos de dados, com o objetivo de datação de documentos. Dentre os pré-processamentos tratados, a técnica de Mínimos Quadrados Generalizados Ponderados (GLSW) mostrou-se a mais adequada para a aplicação. Para ambos os tipos de envelhecimentos foi possível diminuir o Bias de Previsão. Foi utilizado o GLSW para o envelhecimento artificial e natural com o $\alpha = 0,032$ e $\alpha = 0,112$, respectivamente. Em seguida, foram comparados ambos os modelos de envelhecimento, mostrando que a análise dos resíduos permite evidenciar as diferenças entre os processos.

Palavras-chaves: envelhecimento; forense; infravermelho; papel; PCA; PLS.

ABSTRACT

Questioned documents are those sent to forensic analysis, especially when it involves legal issues. The results of these analyses must be objective and reliable to avoid the conviction of innocents. In this context, it is important to study paper authenticity. Paper composition is quite complex, as it is composed of a mixture of substances and its degradation process can be influenced by several factors. This work aims to develop a non-destructive methodology, using infrared spectroscopy associated with chemometric techniques, for the study of artificial and natural aging of paper over time. To carry out this study, the samples were divided into two datasets. The first one, corresponding to artificial aging, was composed by 10 different types and brands of papers. Those samples were stored at 100°C and analyzed in different time intervals, varying from 10 minutes up to 4 months. The second dataset correspond to naturally aged documents, and it was composed by documents dating from 1998 and 2018. Using principal component analysis (PCA), it was possible to identify differences among paper brands and, additionally, identify the oxidation of cellulose's hydroxyl as the main degradation process. Partial least squares (PLS) regression models were built for samples from both data sets, with the objective of dating documents. Among the treated pre-processing, the Generalized Weighted Least Squares (GLSW) technique proved to be the most suitable for the application. For both aging processes it was possible to decrease the prediction Bias. GLSW was used for artificial and natural aging with $\alpha = 0.032$ and $\alpha = 0.112$, respectively. Then both aging models were compared. It was possible to identify significant differences between the two processes by assessing the residues.

Keywords: aging; forensic; infrared; paper; PCA; PLS.

LISTA DE FIGURAS

Figura 1 -	Representação da molécula de celulose.	24
Figura 2 -	Oxidação da hidroxila do carbono C(2), C(3) e C(6) sem a abertura do anel de piranose.	30
Figura 3 -	Oxidação da hidroxila do carbono C(2), C(3) e C(1) com a abertura do anel de piranose.	30
Figura 4 -	Vibrações de estiramentos simétrico e assimétrico e deformações angulares para molécula heteronuclear.	33
Figura 5 -	Curva de potencial de um oscilador harmônico e anarmônico.	35
Figura 6 -	Esquema de funcionamento do acessório ATR.	36
Figura 7 -	Correção do espectro utilizando MSC.	40
Figura 8 -	Esquema de cruzamento único e duplo.	47
Figura 9 -	Média dos diferentes tipos de papéis no tempo zero T0 (a) papéis couchês e (b) papéis reciclados e sulfites.	52
Figura 10 -	10. Gráfico da PCA para todos os papéis em todos os tempos de envelhecimento, para as amostras submetidas ao tratamento térmico (a) espectros brutos, (b) espectros pré-processados com centragem na média, (c) escores da PC1 vs PC2 e (d) pesos da PC1 e da PC2.	55
Figura 11 -	Gráfico do efeito do pré-processamento para os papéis couchês C1 (a) espectros brutos e (b) espectros pré-processados com suavização pelo filtro de Savitzky-Golay (com polinômio de 2 ordem e janela de 33 pontos), SNV e centragem na média.	57
Figura 12 -	Gráfico da PCA para os papéis couchês C1 (a) Variância explicada acumulativa para o modelo, (b) escores da PC1 versus o log do tempo (F1: Folha 1, F2: Folha 2 e F3: Folha 3) e (c) pesos da PC1.	58
Figura 13 -	Gráfico do efeito do pré-processamento para os papéis reciclados R1 (a) espectros brutos e (b) espectros pré-processados com suavização pelo filtro de Savitzky-Golay (com polinômio de 2 ordem e janela de 33 pontos), SNV e centragem na média.	59

- Figura 14 - Gráfico da PCA para os papéis reciclados R1 (a) Variância explicada acumulada para o modelo, (b) gráfico de influencia e resíduo das amostras, (c) escores da PC3 versus a PC4 e (d) pesos da PC3 e da PC4. 60
- Figura 15 - Gráfico para o novo modelo PCA, sem as amostras anômalas, para os papéis reciclados R1 (a) Variância explicada acumulativa para o modelo, (b) escores da PC1 versus o log do tempo (F1: Folha 1, F2: Folha 2 e F3: Folha 3) e (c) pesos da PC1. 61
- Figura 16 - Gráfico do efeito do pré-processamento para os papéis sulfites S1 (a) espectros brutos e (b) espectros pré-processados com suavização pelo filtro de Savitzky-Golay (com polinômio de 2 ordem e janela de 33 pontos), SNV e centragem na média. 63
- Figura 17 - Gráfico da PCA para os papéis sulfites S1 (a) Variância explicada acumulativa para o modelo, (b) escores da PC1 versus o log do tempo (F1: Folha 1, F2: Folha 2 e F3: Folha 3) e (c) pesos da PC1. 64
- Figura 18 - Gráficos das análises dos papéis sulfites, (a) espectros brutos, (b) espectros pré-processados com SNV e centragem na média, (c) escores da PC1 versus PC2 e (d) pesos da PC1 e da PC2. 65
- Figura 19 - Gráficos do modelo PLS para os dados pré-processados com SNV e centragem na média para os papéis envelhecidos artificialmente, (a) gráfico de RMSEC e RMSECV, (b) gráficos de regressão e (c) gráfico dos escores VIPs. 68
- Figura 20 - Gráficos do modelo PLS para os dados pré-processados com SNV e centragem na média e seleção de variáveis (GA) para os papéis envelhecidos artificialmente, (a) Variáveis que aparecem com maior frequência quando o GA é aplicado, (b) gráfico de RMSEC e RMSECV, (c) gráficos de regressão e (d) gráfico dos escores VIPs. 69
- Figura 21 - Gráfico das superfícies para o (a) RMSEC e (b) RMSECV. 70
- Figura 22 - Gráficos do modelo PLS para os dados pré-processados com SNV, GLSW ($\alpha = 0,202$) e centragem na média para os papéis envelhecidos artificialmente (a) Gráfico da superfície para o 71

RMSECV, (b) gráfico de RMSEC e RMSECV, (c) gráficos de regressão e (d) gráfico dos Escores VIPs.

- Figura 23 - Gráficos do modelo PLS para os dados pré-processados com SNV, GLSW ($\alpha = 0,032$) e centragem na média para os papéis envelhecidos artificialmente, (a) Gráfico da superfície para o RMSECV, (b) gráfico de RMSEC e RMSECV, (c) gráficos de regressão e (d) gráfico dos Escores VIPs. 72
- Figura 24 - Média dos espectros brutos dos papéis naturalmente envelhecidos por ano. 73
- Figura 25 - Gráficos da PCA para os papéis envelhecidos naturalmente, (a) espectros brutos, (b) espectros pré-processados, (c) escores da PC1 versus PC2 e (d) pesos da PC1 e da PC2. 74
- Figura 26 - Gráficos do modelo PLS para os dados pré-processados com SNV e centragem na média para os papéis envelhecidos naturalmente (a) gráfico de RMSEC e RMSECV, (b) gráficos de regressão e (c) gráfico dos escores VIPs. 77
- Figura 27 - Gráficos do modelo PLS para os dados pré-processados com SNV e centragem na média e seleção de variáveis (GA) para os papéis envelhecidos naturalmente (a) Variáveis que aparecem com maior frequência quando o GA é aplicado, (b) gráfico de RMSEC e RMSECV, (c) gráficos de regressão e (d) gráfico dos escores VIPs. 78
- Figura 28 - Gráfico das superfícies de níveis para os papéis envelhecidos naturalmente para o (a) RMCEC e (b) RMSECV. 79
- Figura 29 - Gráficos do modelo PLS para os dados pré-processados com SNV, GLSW ($\alpha = 0,112$) e centragem na média para as amostras envelhecidas naturalmente (a) Gráfico de superfície para o RMSECV, (b) gráfico de RMSEC e RMSECV, (c) gráficos de regressão e (d) gráfico dos Escores VIPs. 80
- Figura 30 - Gráficos do modelo PLS para os dados pré-processados com SNV, GLSW ($\alpha = 0,542$) e centragem na média para as amostras envelhecidas naturalmente (a) Gráfico da superfície para o RMSECV, (b) gráfico de RMSEC e RMSECV, (c) gráficos de 81

regressão e (d) gráfico dos Escores VIPs.

- Figura 31 - rojeção das amostras envelhecidas naturalmente nos modelos PLS construído com as amostras envelhecidas artificialmente (a) Gráfico do valor previsto, (b) gráfico de influencia e resíduo das amostras e (c) gráfico de influencia e resíduo das amostras ampliado. 83
- Figura 32 - Projeção das amostras envelhecidas artificialmente nos modelos PLS construído com as amostras envelhecidas naturalmente (a) Gráfico do valor previsto e (b) gráfico de influencia e resíduo das amostras. 84
- Figura 33 - Gráfico do efeito do pré-processamento para os papéis couchês C2 (a) espectros brutos e (b) espectros pré-processados. 94
- Figura 34 - Gráfico da PCA para os papéis couchês C2 (a) Variância explicada acumulativa para o modelo, (b) escores da PC1 versus o log do tempo e (c) pesos da PC1. 95
- Figura 35 - Gráfico do efeito do pré-processamento para os papéis reciclados R2 (a) espectros brutos e (b) espectros pré-processados. 96
- Figura 36 - Gráfico da PCA para os papéis reciclados R2 (a) Variância explicada acumulada para o modelo, (b) gráfico de influencia e resíduo das amostras, (c) escores da PC3 versus a PC4 e (d) pesos da PC3 e da PC4. 97
- Figura 37 - Gráfico para o novo modelo PCA, sem as amostras anômalas, para os papéis reciclados R2 (a) Variância explicada acumulativa para o modelo, (b) escores da PC1 versus o log do tempo e (c) pesos da PC1. 98
- Figura 38 - Gráfico do efeito do pré-processamento para os papéis reciclados R3 (a) espectros brutos e (b) espectros pré-processados. 99
- Figura 39 - Gráfico da PCA para os papéis reciclados R3 (a) Variância explicada acumulativa para o modelo, (b) escores da PC1 versus o log do tempo e (c) pesos da PC1. 100
- Figura 40 - Gráfico do efeito do pré-processamento para os papéis sulfites S2 (a) espectros brutos e (b) espectros pré-processados. 101
- Figura 41 - Gráfico da PCA para os papéis sulfites S2 (a) Variância explicada 102

acumulativa para o modelo, (b) escores da PC1 versus o log do tempo e (c) pesos da PC1.

- Figura 42 - Gráfico do efeito do pré-processamento para os papéis sulfites S3 103
(a) espectros brutos e (b) espectros pré-processados.
- Figura 43 - Gráfico da PCA para os papéis sulfites S3 (a) Variância explicada 104
acumulativa para o modelo, (b) escores da PC1 versus o log do tempo e (c) pesos da PC1.
- Figura 44 - Gráfico do efeito do pré-processamento para os papéis sulfites S4 105
(a) espectros brutos e (b) espectros pré-processados.
- Figura 45 - Gráfico da PCA para os papéis sulfites S4 (a) Variância explicada 106
acumulativa para o modelo, (b) escores da PC1 versus o log do tempo e (c) pesos da PC1.
- Figura 46 - Gráfico do efeito do pré-processamento para os papéis sulfites S5 107
(a) espectros brutos e (b) espectros pré-processados.
- Figura 47 - Gráfico da PCA para os papéis sulfites S5 (a) Variância explicada 108
acumulativa para o modelo, (b) escores da PC1 versus o log do tempo e (c) pesos da PC1.
- Figura 48 - Gráfico de contorno das superfícies para os papéis envelhecidos 109
artificialmente para o (a) RMCEC e (b) RMSECV.
- Figura 49 - Gráfico de contorno das superfícies para os papéis envelhecidos 110
naturalmente para o (a) RMCEC e (b) RMSECV.

LISTA DE TABELAS

Tabela 1 -	Especificações dos papéis.	49
Tabela 2 -	Atribuições de bandas para a região de impressão digital da celulose.	53
Tabela 3 -	Resumo dos resultados dos modelos PLS com os diferentes pré-processamentos.	67
Tabela 4 -	Resumo dos resultados dos modelos PLS com os diferentes pré-processamentos para os papéis envelhecidos naturalmente.	75

LISTA DE ABREVIATURAS E SIGLAS

ATR	Refletância Total Atenuada (<i>Attenuated Total Reflectance</i>).
DA	Análise Discriminante (<i>discriminant analysis</i>)
DHAP	2,5 dihidroxiacetofenona
DP	Grau de Polimerização (DP: <i>Degree of Polimerization</i>)
DRIFTS	Refletância Difusa com Transformada de Fourier (DRIFTS: <i>Diffuse reflectance infrared Fourier transform spectroscopy</i>)
EDXRF	Fluorescência de Raios X por Energia Dispersiva (<i>Energy Dispersive X-ray Fluorescence</i>)
FIR	Infravermelho Distante (FIR: <i>Far Infrared</i>)
FTIR	Espectroscopia de Infravermelho com Transformada de Fourier (FTIR: <i>Fourier Transform Infrared</i>)
GA	Algoritmo genético (<i>Genetic algorithm</i>)
GLSW	Mínimo quadrados Generalizados Ponderados (<i>Generalized Least Squares Weighting</i>).
GPC	Cromatografia de Permeação em Gel (<i>Gel Permeation Chromatography</i>).
ICP-MS	Espectrometria de massa acoplada a plasma indutivo (<i>Inductively coupled plasma mass spectrometry</i>).
iPLS	Mínimos Quadrados Parciais por Intervalos (<i>Interval Partial Least Square</i>)
IR	Infravermelho (<i>Infrared</i>)
IRMS	Espectrometria de massa de razão isotópica (<i>Isotope ratio mass spectrometers</i>)
LA-ICP-MS	Ablação a laser acoplado espectrometria de massa acoplada a plasma indutivo (<i>Laser ablation inductively coupled mass spectrometry</i>)
LDA	Análise Discriminante Linear (<i>Linear Discriminant Analysis</i>)
LS-SVM	Máquina de Vetores de Suporte a Mínimos Quadrados (<i>Least Squares-Support Vector Machine</i>)

MC	Centragem na Média (<i>Mean Center</i>)
MIR	Infravermelho Médio (<i>Middle Infrared</i>)
MSC	Correção Multiplicativa de Sinais (<i>Multiplicative Scattering Correction</i>)
NIR	Infravermelho Próximo (<i>Near Infrared</i>)
OLS	Mínimos Quadrados Ortogonais (<i>Orthogonal Least Squares</i>)
PCA	Análise dos Componentes Principais (<i>Principal Component Analysis</i>)
PCA – DA	Análise Discriminante por Análise dos Componentes Principais (<i>Principal Component Analysis with Discriminant Analysis</i>)
PCA-LDA	Análise Discriminante Linear com a análise dos componentes principais (<i>Principal Component Analysis-Linear Discriminant Analysis</i>)
PLS – DA	Análise Discriminante por Mínimos Quadrados Parciais (<i>Partial Least Squares Discriminant Analysis</i>)
PLS – LDA	Análise Discriminante Parcial por Mínimos Quadrados Lineares (<i>Partial Least Squares-Linear Discriminant Analysis</i>)
PS	Spray de Papel (<i>Paper spray</i>)
SIMCA	Modelagem Independente Flexível por Analogias de Classe (SIMCA: <i>Soft Independent Modeling of Class Analogies</i>)
SNV	Padronização Normal do Sinal (<i>Standard Normal Variate</i>)
UV/VIS	Ultravioleta/Visível
VIP	Importância de Variáveis na Projeção (<i>Variable Importance in the Projection</i>)
XRD	Difração de Raios X (<i>X-ray diffraction</i>)
XRF	Fluorescência de raios X (<i>X-ray fluorescence</i>)

SUMÁRIO

1	INTRODUÇÃO	19
2	HIPÓTESE E OBJETIVOS	23
2.1	OBJETIVOS ESPECÍFICOS	23
3	FUNDAMENTAÇÃO TEÓRICA	24
3.1	COMPOSIÇÃO DOS PAPÉIS	24
3.2	ENVELHECIMENTO ARTIFICIAL DO PAPEL	25
3.2.1	O tempo no envelhecimento artificial	28
3.2.2	Mecanismo de degradação por energia térmica	29
3.3	ENVELHECIMENTO NATURAL	31
3.4	ESPECTROSCOPIA DO INFRAVERMELHO	32
3.5	QUIMIOMETRIA	36
3.5.1	Técnicas de pré-processamento	37
3.5.1.1	Suavização pelo método de Savitzky-Golay	37
3.5.1.2	Derivada	38
3.5.1.3	Padronização Normal do Sinal	38
3.5.1.4	Correção Multiplicativa Sinais	39
3.5.1.5	Centragem na Média	40
3.5.1.6	Mínimos Quadrados Generalizados Ponderados	40
3.5.2	Análise dos Componentes Principais	42
3.5.3	Regressão por Mínimos Quadrados Parciais	43
3.5.4	Seleção de variáveis por algoritmo genético	46
4	MATERIAIS E MÉTODOS	48
4.1	AMOSTRAS	48
4.1.1	Amostras Envelhecidas Artificialmente	48
4.1.2	Amostras Envelhecidas Naturalmente	49

4.2	AQUISIÇÃO DOS ESPECTROS	50
4.3	TRATAMENTO DE DADOS	50
5	RESULTADOS E DISCUSSÃO	52
5.1	ENVELHECIMENTO ARTIFICIAL	52
5.1.1	Atribuições de bandas dos espectros	52
5.1.2	Análise dos componentes principais	53
5.1.3	Regressão por Mínimos Quadrados Parciais	66
5.2	ENVELHECIMENTO NATURAL	72
5.2.1	Atribuição de Bandas	72
5.2.2	Análise dos Componentes Principais	73
5.2.3	Regressão por Mínimos Quadrados Parciais	75
5.3	ENVELHECIMENTO NATURAL VS ARTIFICIAL	81
5.3.1	Projetando as amostras de envelhecimento natural no modelo PLS das amostras envelhecidas artificialmente	82
5.3.2	Projetando as amostras de envelhecimento artificial no modelo PLS das amostras envelhecidas naturalmente	84
6	PERSPECTIVA FUTURA	86
7	CONCLUSÃO	88
	REFERÊNCIAS	89
	APÊNDICE A – ANÁLISE DOS COMPONENTES PRINCIPAIS PARA OS PAPÉIS COUCHÊS	94
	APÊNDICE B – ANÁLISE DOS COMPONENTES PRINCIPAIS PARA OS PAPÉIS RECICLADOS	96
	APÊNDICE C – ANÁLISE DOS COMPONENTES PRINCIPAIS PARA OS PAPÉIS SULFITES	101
	APÊNDICE D – GRÁFICOS DE CONTORNO DAS SUPERFÍCIES DE RESPOSTA	109

1 INTRODUÇÃO

O papel é um objeto largamente utilizado no cotidiano, pois toda a vida de um indivíduo é documentada. Esses documentos podem ser apresentados na forma de cartas, envelopes, certificados, livros, cartão de identificação, contratos, testamentos, entre outros. Quando um documento é enviado para análise forense, é considerado um documento questionado. Nesse contexto é de fundamental importância atestar a legitimidade de determinados documentos, principalmente quando há questionamento Judicial. Os resultados das análises devem ser objetivos e confiáveis, para atestar a infração e evitar a punições de inocentes. Essas falsificações, manipulações e adulterações de documentos são estudadas pela documentoscopia (SIEGEL, 2016; BRUNELLE, 2003).

Na documentoscopia uma das grandes dificuldades encontradas é o estudo do envelhecimento do papel, pois ele é composto por uma mistura complexa de componentes que podem sofrer degradação devido a diversos fatores. Nesse contexto, a maioria dos trabalhos encontrado na literatura está relacionada com a análise de tinta de caneta e/ou impressões e apenas 7 % dos trabalhos utilizam o papel como amostra. Dentre esses trabalhos, é possível encontrar duas abordagens diferentes. A primeira, mais comum, tem o objetivo de caracterizar e discriminar documentos ou lotes de papéis, enquanto o segundo estuda a degradação natural ou artificial do papel ao longo do tempo (CALCERRADA e GARCÍA-RUIZ, 2015).

Diversas técnicas vêm sendo utilizadas para a discriminação e caracterização dos papéis, entre elas, tem-se: espectrometria de massa acoplada a plasma indutivo (ICP-MS: *Inductively coupled plasma mass spectrometry*), Fluorescência de raios X (XRF: *X-ray fluorescence*), ablação a laser acoplado ao ICP-MS (LA – ICP – MS: *Laser ablation inductively coupled mass spectrometry*) e espectrometria de massa de razão isotópica (IRMS: *Isotope ratio mass spectrometers*), difração de Raios-X e espectroscopia na região do infravermelho (IR: *Infrared*) (SPENCE *et al*, 2000; VAN ES *et al*, 2009; CAUSIN *et al*, 2010).

Com o passar do tempo e com o desenvolvimento tecnológico foi aumentando o interesse na utilização das técnicas espectroscópicas, como o infravermelho, para a

caracterização e discriminação de diferentes marcas de papéis. Isso ocorreu porque a técnica em questão é não destrutiva e para casos de contestação na justiça, onde as amostras são evidências, é importante a preservação dos documentos. As técnicas espectroscópicas, geralmente fornecem como resultados espectros de natureza multivariada, e conseqüentemente a obtenção de informações relevantes para o estudo não é uma tarefa trivial. Nesse contexto, é necessário o uso de determinadas estratégias, para extrair informações relevantes ao estudo, como o uso da quimiometria. A quimiometria é uma área que envolve técnicas que utilizam a matemática e a estatística para a obtenção de informações químicas relevantes a partir dos dados multivariados (FERREIRA, 2015).

Dentre os vários grupos da análise multivariada, temos as técnicas da análise exploratória, de calibração e de classificação. O primeiro é análise exploratória, cuja ferramenta mais utilizada é a análise dos componentes principais (PCA: *Principal Component Analysis*). O segundo grupo compreende as técnicas de calibração, que têm o objetivo de determinar quantitativamente uma propriedade de interesse. E por fim, têm-se as técnicas de classificação, na qual um modelo é construído com o objetivo de atribuir uma amostra desconhecida a uma classe previamente modelada (BRERETON, R., 2003; BEEBE e KOWALSKI, 1987).

A quimiometria vem sendo aplicada a dados de natureza multivariada, como os espectros de infravermelho próximo e médio com transformada de Fourier (FTIR: *Fourier transform infrared*). Esses espectros podem ser obtidos utilizando diversos acessórios, como o de refletância total atenuada (ATR: *Attenuated Total Reflectance*), refletância difusa (DRIFTS: *Diffuse reflectance infrared Fourier transform spectroscopy*), entre outros. Em estudos cujos objetivos são abordar a discriminação e a caracterização dos papéis, podem ser construídos modelos de classificação como a análise discriminante linear (PCA-LDA: *principal component analysis-linear discriminant analysis*), análise discriminante por mínimos quadrados parciais (PLS-DA: *partial least squares discriminant analysis*), SIMCA (SIMCA: *soft independent modeling of class analogies*), máquina de vetores de suporte por mínimos quadrados (LS-SVM: *least squares-support vector machine*) e a análise discriminante linear por mínimos quadrados parciais (PLS-LDA: *partial least squares-linear discriminant analysis*) para discriminar diferentes tipos e marcas de papéis (KHER *et al*, 2001; XIA *et al*, 2019).

Embora o estudo da discriminação de diferentes tipos de papéis seja bastante importante para a determinação de fraudes, um dos grandes desafios da documentoscopia ainda consiste na datação de documentos. Determinar a autenticidade de um documento de acordo com a suposta data de preparação ainda é um dos grandes desafios para os peritos da documentoscopia. Principalmente para obter resultados objetivos, confiáveis e que não dependa de cada interpretação de cada perito. Na literatura é possível encontrar trabalhos que estudam o envelhecimento natural ou artificial do papel ao longo do tempo, essa é a segunda área de estudo do papel, como já mencionado.

Risoluti e colaboradores no ano de 2018 estudaram o envelhecimento natural do papel. As amostras utilizadas foram livros e folhas datadas do século XVI ao XXI e as análises foram realizadas em um equipamento portátil no infravermelho próximo. Os espectros obtidos foram submetidos a uma análise quimiométrica, especificamente a PCA, na qual foi possível distinguir corretamente os papéis com base no século ao qual pertencem. Os resultados mostraram que a utilização do equipamento MicroNIR para distinguir papéis de acordo com o período de fabricação é promissora e é possível analisar documentos de origem desconhecida, além de ser uma técnica totalmente portátil e não destrutiva. Nesse contexto, é necessário salientar a importância da representatividade das amostras, pois durante a fabricação dos papéis a sua composição muda ao longo dos anos, de modo que essa diferença, encontrada na PCA, pode estar relacionada a composição do papel e não ao processo de degradação (RISOLUTI et al, 2018).

Zieba-Palus e colaboradores (2020) estudaram o envelhecimento artificial do papel. Para esse estudo foram utilizadas como amostras 5 diferentes marcas de papéis, submetidas a um envelhecimento artificial em uma câmara, com umidade relativa abaixo de 65% e temperatura de 90 °C. Essas amostras foram avaliadas durante um período de tempo de 38 dias. As análises foram realizadas por espectroscopia FTIR na região do infravermelho médio e UV/VIS. As amostras de papéis envelhecidas foram diferenciadas analisando determinadas regiões dos espectros, associadas à correlação 2D e a PCA. A mudança espectral relacionada à mudança do papel ao longo do tempo foi observada nas faixas de 1000 - 1120 cm^{-1} , 2860 - 2950 cm^{-1} e 1300 - 1380 cm^{-1} no infravermelho. Para os espectros UV / VIS

foi possível observar mudanças na intensidade das bandas 280 e 350 nm durante o envelhecimento. Ambas as técnicas oferecem informações complementares úteis na análise do papel para fins forenses, possibilitando a identificação de um papel ou documento específico. Esse grupo de pesquisa utilizou a correlação 2D e a PCA para determinar a faixa espectral relacionada a informação da mudança do papel ao longo do tempo, mas não foram construídos modelos de calibração voltados para a datação (ZIEBA-PALUS *et al*, 2020).

Nos últimos anos, nosso grupo de pesquisa tem se voltado para o estudo de datação de documentos. Silva e colaboradores realizaram um estudo preliminar sobre o envelhecimento natural de amostras de papéis, datados de 1986 a 2012 e espectroscopia do infravermelho médio. Com o objetivo de contornar o problema da variabilidade das amostras foi utilizada uma diversidade de amostras para cada ano e foi realizada uma calibração multivariada voltada para datação (SILVA *et al*, 2018).

A autora desta dissertação, em estudo prévio, avaliou o envelhecimento artificial de diferentes tipos de papéis submetidos a diferentes condições de envelhecimento utilizando a espectroscopia na região do infravermelho e quimiometria. As amostras utilizadas consistiram em papéis sulfites, couchês e reciclados, e foram submetidas a 4 condições de envelhecimento, alta temperatura, radiação UV, radiação solar e ao abrigo de luz. Os espectros foram analisados utilizando a PCA. Os resultados obtidos mostraram que a utilização da espectroscopia IR é promissora para o estudo da degradação do papel. A partir desses estudos observou-se que um dos desafios desta análise está relacionado à grande variabilidade da composição nas folhas de papéis e como essas diferenças influenciam o processo de degradação. Além disso, relacionar o envelhecimento artificial e o natural também se mostra bastante complexo. Nesse contexto, o objetivo desse trabalho é o estudo de datação de documentos, para detectar falsificações. Para isso, foram avaliadas e comparadas as amostras de papel envelhecidas de forma natural e artificial (utilizando o tratamento térmico).

2 HIPÓTESE E OBJETIVOS

As mudanças do papel ao longo do tempo, após exposição a diferentes condições de armazenamento, são causadas por modificações, na molécula de celulose, dentre outras. Essas modificações podem não ser facilmente identificadas por inspeção visual do espectro, sendo necessário o uso de técnicas mais avançadas para extrair a informações úteis. Nesse contexto, esse trabalho tem como hipóteses: (1) o processo de degradação do papel causa mudanças na molécula de celulose que podem ser detectadas por IR e modeladas com técnicas de reconhecimento de padrões supervisionadas; (2) O envelhecimento artificial dos papéis pode simular o processo de envelhecimento natural e, portanto, forjar a autenticidade de um documento.

O objetivo deste trabalho é desenvolver uma metodologia não destrutiva, utilizando a espectroscopia de infravermelho médio associada a técnicas quimiométricas, para o estudo do envelhecimento artificial e natural do papel, com o intuito de determinar a datação de documentos e o potencial dessa metodologia para determinar falsificações de documentos.

2.1 OBJETIVOS ESPECÍFICOS

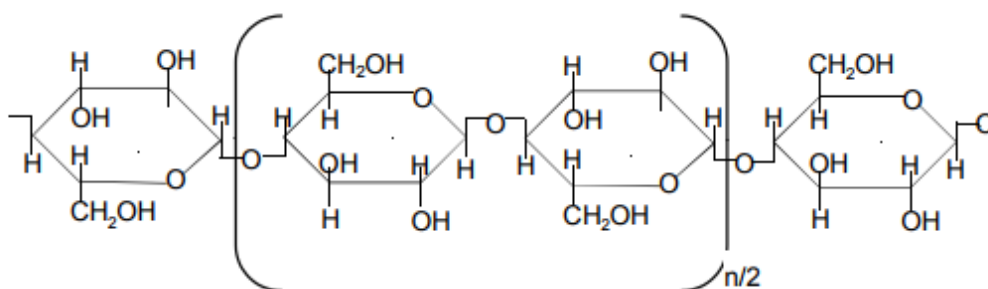
- Utilizar a espectroscopia de infravermelho médio para adquirir os espectros das amostras envelhecidas artificialmente e naturalmente.
- Avaliar o envelhecimento artificial por tratamento térmico e natural do papel utilizando a técnica de análise por componentes principais.
- Construir modelos PLS para as amostras envelhecidas naturalmente e artificialmente avaliando técnicas de pré-processamentos e de seleção de variáveis.
- Comparar o envelhecimento natural e artificial a partir dos modelos PLS visando identificar falsificações de documentos.

3 FUNDAMENTAÇÃO TEÓRICA

3.1 COMPOSIÇÃO DOS PAPÉIS

O papel é composto por uma mistura complexa de componentes, tais como a celulose, hemicelulose e lignina, que são os principais constituintes da madeira, mas também é possível encontrar alguns compostos inorgânicos, como o carbonato de cálcio, CaCO_3 e a caulinita, $\text{Al}_2\text{Si}_2\text{O}_5(\text{OH})_4$. Esses compostos são frequentemente utilizados para o revestimento do papel no processo de fabricação para ajustar a coloração, brilho, opacidade, a suavidade e a receptividade de tintas de impressão (CAUSIN et al, 2010 e UDRISTIOIU et al, 2012). Dentre esses compostos, a celulose é encontrada em maior quantidade. É um polímero não ramificados (Figura 1), constituído por moléculas de glicose ligadas por ligação 1,4- β -glucosídica. O número de unidades de glicose é variável e depende de dois fatores: da matéria prima e do processo de extração. O grau de organização das unidades glicosídica é expresso pelo grau de polimerização, o DP (DP: *Degree of Polimerization*), que sofre mudanças devido ao processo de degradação, entre outros fatores.

Figura 1 - Representação da molécula de celulose



Fonte: Adaptado de Margutti *et al* (2001).

O estudo do papel é bastante complexo, pois o mesmo pode sofrer degradação devido a fatores externos, como o ambiente de armazenamento, a temperatura e a umidade, mas também devido a fatores internos, tais como, a presença de íons

metálicos e a degradação da lignina. A degradação da celulose, pode seguir diversos mecanismos tais quais, a biodegradação, a fotodegradação, a hidrólise ácida e a oxidação. Existem diversas aplicações para o estudo da degradação do papel, como avaliação da cinética, mecanismo de degradação, resistência, durabilidade, preservação do papel e também a datação de documentos. Nesse contexto, o estudo do papel vem ganhando espaço ao longo dos anos. Na literatura pode-se encontrar trabalhos que estudam o envelhecimento artificial ou natural do papel e dificilmente encontram-se trabalhos que comparam ambos os envelhecimentos.

3.2 ENVELHECIMENTO ARTIFICIAL DO PAPEL

Existem três principais objetivos em um estudo de envelhecimento artificial, o primeiro é determinar a estabilidade química e a duração física dos materiais analisados, classificação e combinação dos materiais em um curto período de tempo. O segundo é prever a durabilidade do material de interesse sob condição de uso esperada. E o terceiro objetivo é determinar o(s) mecanismo(s) de degradação do material e conseqüentemente desenvolver técnicas para monitorar esses mecanismos e estender a vida útil do material (FELLER, 1994).

Zou e colaboradores estudaram o envelhecimento artificial de papéis com o objetivo de determinar o mecanismo de degradação e a sua cinética. Os papéis de celulose pura com diferentes pH foram submetidas ao envelhecimento controlando a temperatura e a umidade. Durante esse estudo foram realizados diversos testes físicos e químicos, chegando à conclusão que há uma perda da resistência e a fragilização do papel durante o envelhecimento. Essa ocorrência é devido à despolimerização da celulose causada pela hidrólise catalisada por ácido que acontece de uma maneira aleatória e segue uma cinética de primeira ordem de acordo com os autores (ZOU *et al*, 1994; ZOU *et al*, 1996; ZOU *et al*, 1996).

Outro grupo de pesquisa em 2001 estudou o envelhecimento do papel celulósico por espectroscopia de infravermelho próximo por transformada de Fourier, com o objetivo de avaliar a condição de isolamento dos papéis utilizados em

transformadores elétricos. As amostras de papéis foram submetidas à alta temperatura 120°C, 140°C e 160°C. Os papéis envelhecidos puderam ser discriminados dos papéis novos devido a uma banda localizada em aproximadamente 1710 cm⁻¹ que é atribuído a carbonila/carboxil. Analisando essa mesma banda foi possível determinar que o envelhecimento segue uma cinética de primeira ordem e tem energia cinética de 98 KJ/mol. Para esse estudo também foram utilizadas técnicas quimiométricas como a PCA (ALI *et al*, 2001). Embora os dois trabalhos citados acima utilizem metodologias diferentes (o primeiro utiliza técnicas destrutivas e o segundo faz uso de técnicas espectroscópicas), ambos possuem a mesma finalidade, o estudo do mecanismo de degradação e a cinética dos papéis, mas nenhum deles avalia o envelhecimento artificial voltado à datação de documentos.

Pesquisas mostram que o envelhecimento artificial que mais se assemelha ao envelhecimento natural são os que submetem as amostras a elevadas temperaturas, fluxo de luz e concentração de gases e poluentes atmosféricos comuns (FELLER, 1994; MARGUTTI *et al*, 2001; AREA e CHERADAME, 2011). Um programa no Instituto de Pesquisa de Padrões (ISR: *Institute for Standards Research*) da Sociedade Americana para testes e materiais (ASTM: American Society for testing and materials) tem desenvolvido testes na área do envelhecimento artificial e natural do papel para que seja possível desenvolver padrões de envelhecimento. Os testes de envelhecimento artificial consistem em concentrar a umidade em torno do papel e submeter a elevadas temperaturas selando as amostras dentro de tubos de vidro herméticos. Esses tubos retêm os produtos de degradação, simulando folhas envelhecidas em livros. Os papéis envelhecidos em tubos herméticos a 100 °C durante 5 dias equivale a folhas individuais envelhecidas a 90 °C e 50% de umidade em forno úmido por 30 dias (PORCK e TEYGELER, 2000).

O interesse no estudo do envelhecimento artificial do papel empregando tratamento térmico tem aumentado. Kacik e colaboradores em 2009 submeteram amostras de papel de jornal a 98 °C durante 60 dias. As análises foram realizadas por viscosimetria, cromatografia de permeação em gel (GPC: *Gel Permeation Chromatography*) e determinação de sacarídeos. A partir dos resultados das análises foi possível determinar que no início da degradação a reação de maior influência é a hidrólise seguida da oxidação. Foi possível encontrar um coeficiente

de determinação $R^2 = 0,88$, entre o tempo de envelhecimento e o grau de despolimerização. Esse grupo de pesquisa utilizou técnicas destrutivas para o estudo do envelhecimento do papel, com o objetivo de estudar os mecanismos de degradação. A partir desse estudo é possível observar um interesse em relacionar o envelhecimento com o tempo de exposição das amostras à condição de envelhecimento (KAČÍK *et al*, 2009).

Em 2016, Hajji e colaboradores utilizaram espectroscopia no infravermelho médio com o acessório ATR, difração de raios X (XRD: *X-ray diffraction*) e fluorescência de raios X por energia dispersiva (EDXRF: *Energy Dispersive X-ray Fluorescence*) para monitorar as mudanças em papéis restaurados e expostos a condições extremas de armazenamento. Para simular essas condições foram realizados dois testes de envelhecimento acelerado; o primeiro por calor seco a 90 °C e o segundo por calor úmido a 90 °C e 100% de umidade relativa, durante um período de 28 dias. Como amostras, foram utilizados manuscritos restaurados do Marrocos, datados dos séculos XVI, XVII, XVIII e XIV e papel japonês usado para restauração. Pela análise ATR-FTIR foi possível identificar modificações estruturais da celulose, observar a diminuição da água residual nas amostras envelhecidas em calor seco e determinar o mecanismo de degradação por oxidação das amostras envelhecidas submetidas ao calor úmido. A técnica XRD foi usada para observar a diminuição da cristalinidade da celulose. A partir dos resultados do EDXRF foi possível observar que as duas condições de envelhecimentos artificiais afetaram a composição elementar do papel, especialmente na quantidade de Cálcio (HAJJI *et al*, 2016). Também durante o ano de 2016, outro grupo de pesquisa estudou o envelhecimento artificial do papel em uma câmara climática a 90 °C e 65% de umidade durante 35 dias. As análises foram realizadas por espectroscopia no IR e Raman e as amostras utilizadas foram papéis de escritórios de diferentes fabricantes. Foram analisados os espectros obtidos e a correlação 2D (os espectros são transformados em mapas 2D e mostram uma relação entre as diferentes regiões espectrais) para a espectroscopia de infravermelho, chegando à conclusão que os mecanismos de degradação que ocorre no papel são clivagem da cadeia de celulose e oxidação (ZIEBA-PALUS *et al*, 2016). O primeiro grupo de pesquisa realizou um estudo com o objetivo de avaliar os dois tipos de envelhecimentos artificiais, sendo possível observar as principais mudanças espectrais. O segundo grupo estudou o mecanismo de degradação para

a condição de envelhecimento analisada. Mas nenhum deles realizou um estudo relacionado o envelhecimento com o tempo de exposição das amostras.

Sabendo que não só a temperatura interfere no processo de degradação, mas também o tempo, umidade relativa e a presença de poluentes, Schedl e colaboradores em 2017 estudaram a detecção de um cromóforo responsável pelo amarelamento do papel, 2,5 dihidroxiacetofenona (DHAP). Para essa análise foi utilizado spray de papel (PS) acoplado à espectrometria de massa, sendo possível detectar o DHAP em amostras de documentos históricos. Foi investigado também o envelhecimento artificial do papel para determinar quais fatores influenciam na degradação. Para esse estudo foi realizado um planejamento fatorial completo, os fatores analisados foram: temperatura, tempo, umidade relativa e a presença de íons de ferro. Os fatores que apresentaram maior influência foram a temperatura e umidade relativa, mas as interações entre os fatores também foram significativas (SCHEDL et al, 2017). O diferencial desse trabalho foi a realização de um planejamento fatorial para determinar os fatores, e as suas possíveis interações, que mais influenciam no processo de envelhecimento artificial pela presença do DHAP, embora sem focar na datação de documentos.

3.2.1 O tempo no envelhecimento artificial

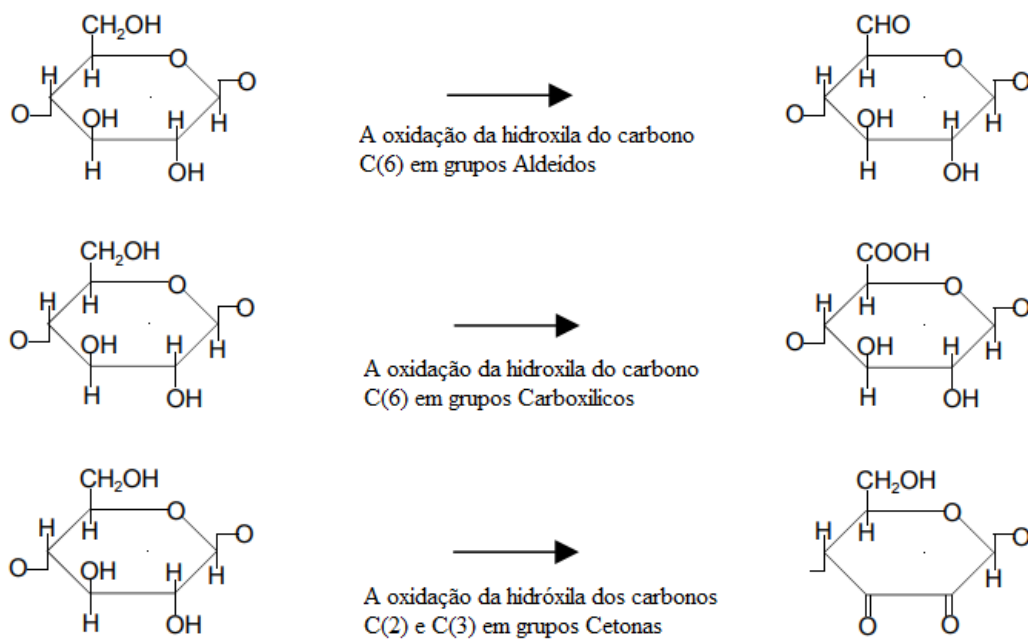
O envelhecimento artificial tem como objetivo simular o processo de envelhecimento por meio de uma degradação acelerada. Assim, é preciso estimar uma determinada propriedade em um curto período de tempo que esteja associada com esse processo. O tempo é medido em relação ao período de exposição das amostras à condição de envelhecimento, podendo ser medido em segundos, minutos, horas, dias ou meses. As mudanças na propriedade de interesse são normalmente estimadas em uma escala linear, como por exemplo, horas de exposição do papel a alta temperatura. Pode-se também ser utilizada uma escala logarítmica, para linearizar funções. Nos estudos de envelhecimento artificial o objetivo de utilizar uma escala de tempo é relacionar a uma escala semelhante de

um envelhecimento natural (FELLER, 1994). No presente trabalho, o tempo de exposição das amostras é medido em minutos e é utilizada uma escala logarítmica.

3.2.2 Mecanismo de degradação por energia térmica

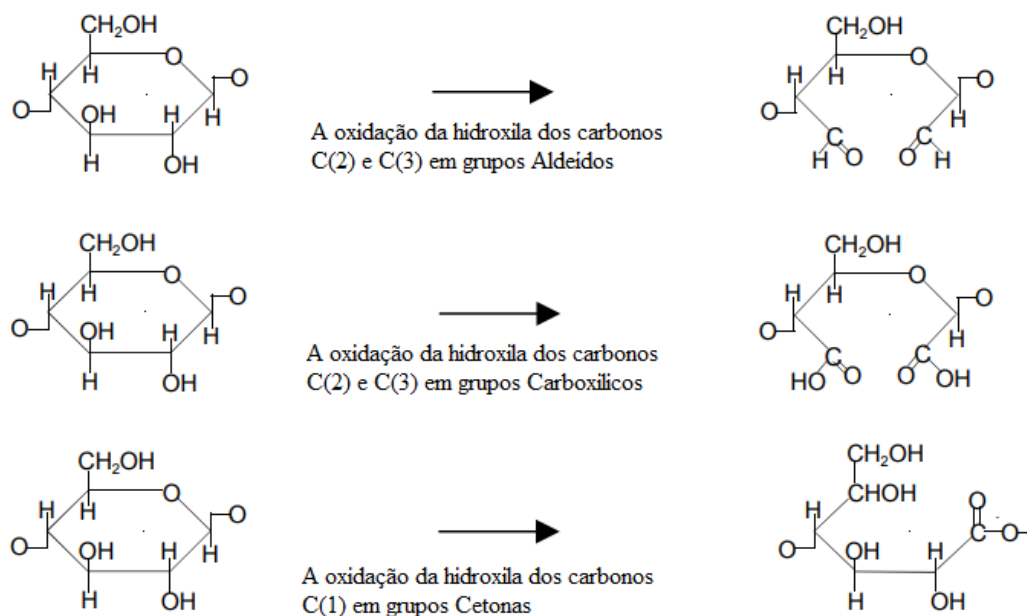
Ao submeter amostras de papéis à condição de envelhecimento por tratamento térmico, o mecanismo de degradação que pode ocorrer é a degradação puramente térmica, chamada de termólise ou degradação termolítica, na qual o oxigênio não está envolvido. No entanto, é necessário levar em consideração a presença do oxigênio atmosférico, ou seja, deterioração termo-oxidativa. Essa degradação é a mais provável de ocorrer quando as amostras são expostas à altas temperaturas na presença de oxigênio ou ar. O oxigênio da atmosfera, vai atuar como um agente oxidante que inicia a reação a partir dos radicais $O\cdot$ e $O_2\cdot$. Os radicais podem atacar as hidroxilas presentes nos carbonos C(2), C(3) e C(6) e sofrem a oxidação nas unidades do anel de piranose, formando novos grupos de carbonila e carbóxilas, podendo ocorrer ou não a abertura do anel. Os novos grupos formados são cromóforos, capazes de absorver radiação visível e são responsáveis pelo amarelecimento do papel. A hidroxila do carbono C(6) sofre oxidação formando grupos aldeídos ou carboxílicos; por outro lado, as hidroxilas dos carbonos C(2) e C(3) podem sofrer oxidação formando cetonas, sem a abertura do anel (Figura 2). Para ocorrer à clivagem do anel, os radicais de oxigênio podem atacar o carbono C(1), que está envolvido na ligação glicosídica, levando à sua clivagem, ou também pode ocorrer o ataque aos carbonos C(2) e C(3) formando grupos aldeídos ou carboxílicos (Figura 3). Outro fator a ser considerado é a presença de água no sistema, que normalmente é monitorada. Quando esse monitoramento é realizado, é medida a umidade presente no ambiente. Dessa forma, também pode ocorrer a hidrólise, embora a oxidação e a hidrólise estejam relacionadas, pois exercem efeitos catalíticos um sobre o outro.

Figura 2 - Oxidação da hidroxila do carbono C(2), C(3) e C(6) sem a abertura do anel de piranose



Fonte: Adaptado de Margutti *et al* (2001).

Figura 3 - Oxidação da hidroxila do carbono C(2), C(3) e C(1) com a abertura do anel de piranose



Fonte: Adaptado de Margutti *et al* (2001).

3.3 ENVELHECIMENTO NATURAL

A ASTM também possui um projeto para o estudo do envelhecimento natural do papel, na qual 10 instituições norte-americanas em diferentes regiões iram armazenar durante 100 anos volumes de 50 tipos de papéis. Durante esses anos folhas de papéis serão testadas em relação a durabilidade óptica e física (PORCK e TEYGELER, 2000). Nesse contexto é possível encontrar diversos trabalhos com o objetivo do estudo do envelhecimento natural dos papéis.

Em 2007, Trafela e colaboradores estudaram a datação de documentos históricos (de 1650 a 2005) utilizando o FTIR próximo e médio. As amostras também foram analisadas quanto ao teor de cinzas, teor de lignina, pH, grau de polimerização (DP: *Degree of Polimerization*) e teor de alumínio usando métodos analíticos de rotina. Utilizando os espectros obtidos foram construídos modelos PLS satisfatórios para todas as propriedades, exceto o teor de alumínio. Esse grupo de pesquisa estudou o envelhecimento natural do papel, analisando diversos fatores, incluindo a datação. Durante o estudo da datação de documentos é importante salientar a necessidade da representatividade das amostras, pois as diferenças observadas nas folhas de papéis ao longo dos anos podem estar associadas à composição dos papéis e não ao envelhecimento das folhas (TRAFELA *et al*, 2007). Em 2017, Martinez e colaboradores (MARTÍNEZ *et al*; 2017) estudaram o envelhecimento natural do papel por fluorescência induzida por laser. As amostras utilizadas foram documentos datados de 1730 a 2009, analisados na região espectral de 540 a 750 nm. Foi possível observar uma mudança espectral do papel ao longo do tempo relacionada com a celulose, hemicelulose e lignina. Essas mudanças apresentaram na forma de deslocamento de bandas e mudanças nas intensidades dos espectros de emissão.

Em 2018 Silva *et al* estudaram o envelhecimento natural do papel por espectroscopia FTIR na região do infravermelho médio associada a técnicas quimiométricas para prever a idade de documentos. Foi realizado o estudo do envelhecimento natural, as amostras utilizadas foram documentos datados de 1986 a 2012. Para esse estudo foram realizados modelos PLS com diferentes pré-processamentos, entre eles o mínimo quadrados generalizados ponderados (GLSW: *Generalized Least Squares Weighting*) e mínimos quadrados ortogonais (OLS:

Orthogonal Least Squares). Outra abordagem utilizada foi à seleção de variável, por esse motivo foram construídos modelos de espaço PLS (sPLS). Esses modelos foram construídos com a finalidade de reduzir a variabilidade existente nas folhas de papéis, principalmente nas amostras de papéis do mesmo ano, para que a variabilidade do papel em relação ao tempo fique evidente. Os modelos construídos apresentaram valores de RMSECV e RMSEP, de cerca de 4 anos, sendo o melhor modelo aquele utilizando o pré-processamento OSC (SILVA *et al*, 2018). Com o objetivo de contornar o problema da representatividade foi utilizada uma grande quantidade e diversidade de amostras para cada ano e foi realizada uma calibração multivariada voltada a datação. A partir desse trabalho surgiu o interesse de comparar o envelhecimento artificial e natural, com o objetivo de identificar falsificações de documentos.

3.4 ESPECTROSCOPIA DO INFRAVERMELHO

A partir dos trabalhos citados acima é possível observar o uso crescente da espectroscopia para o estudo da degradação do papel. Nesse contexto, a espectroscopia de infravermelho é uma técnica de espectroscopia vibracional baseada em absorção molecular. Quando a energia é absorvida, ela provoca transições na molécula de um estado de energia vibracional ou rotacional para outro mais energético. A região do IR pode ser dividida em três sub-regiões. Infravermelho próximo (NIR: *Near Infrared*) compreendida entre 12800 a 4000 cm^{-1} , infravermelho médio (MIR: *Middle Infrared*) região entre 4000 a 200 cm^{-1} e infravermelho distante (FIR: *Far Infrared*) compreendida entre 200 a 10 cm^{-1} . Com o avanço tecnológico vem aumentando o interesse na utilização da espectroscopia vibracional que além de serem técnicas mais rápidas, precisam de nenhuma ou de pouca preparação de amostras (SKOOG *et al*, 2018; SIESLER *et al*, 2002; BURNS e CIURCZAK, 2007).

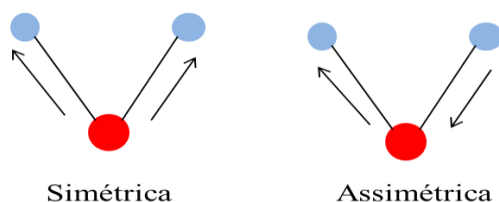
Os espectros da região NIR fornecem majoritariamente bandas de combinações e sobretons, é bastante utilizada para análise de compostos que contém grupos funcionais de hidrogênio ligado ao carbono, nitrogênio e oxigênio. Os dados obtidos no MIR fornecem informações referentes às transições fundamentais dos grupos

funcionais, apresentando, em geral, absorções com maior intensidade que os espectros NIR. A região de FIR é bastante utilizada para estudos com ligas metálicas, pois absorção de átomos de metal e dos ligantes orgânicos e inorgânicos geralmente ocorre nessa região (BURNS e CIURCZAK, 2008; SKOOG *et al*, 2018).

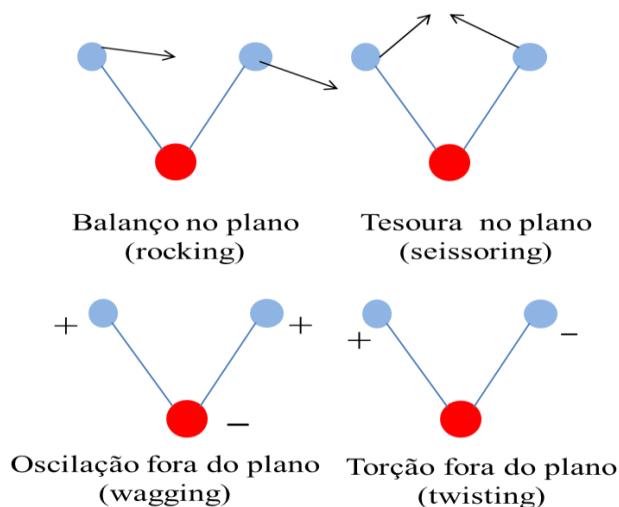
Todas as moléculas possuem vibrações e rotações próprias. Para moléculas cujas vibrações resultam em variação do momento de dipolo e, conseqüentemente variação do campo elétrico da molécula, a radiação infravermelha incidida pode ser absorvida se a frequência da radiação coincide com a frequência vibracional natural da molécula, provocando alteração na amplitude do movimento. A Figura 4 mostra um exemplo de vibrações e deformações angulares para uma molécula triatômica e heteronuclear (WORKMAN e WEYER, 2007; SKOOG *et al*, 2018).

Figura 4 - Vibrações de estiramentos simétrico e assimétrico e deformações angulares para molécula heteronuclear

a) Vibrações de estiramento



a) Vibrações de deformação angular

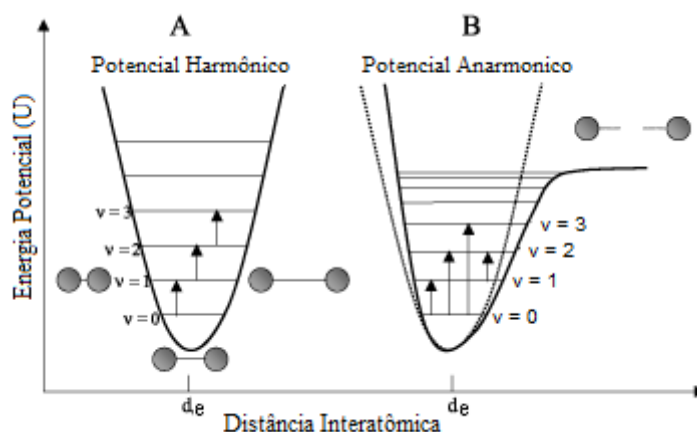


Fonte: Adaptado do Skoog *et al* (2018).

Para entender as absorções moleculares foi proposto inicialmente o modelo do oscilador harmônico, na qual a ligação entre dois átomos heterogêneos se assemelha a um sistema com uma mola que ligam duas massas diferentes (m_1 e m_2). Nesse contexto, a força de ligação, ou seja, a frequência de vibração dependerá das massas dos dois átomos e da constante de força da ligação que une as duas massas. Esse modelo explica apenas as vibrações fundamentais, pois permite apenas transições entre níveis adjacentes, representado pela Figura 5. De acordo com a distribuição de Boltzman, em condições do ambiente, a maioria das moléculas se encontra no estado fundamental localizado no nível $v = 0$, logo as transições permitidas são para o nível $v = 1$. Devido ao fato de, neste modelo, os níveis de energia serem igualmente espaçados seria observado um único pico para cada tipo de vibração molecular. Além disso, a maioria das bandas de absorção das vibrações fundamentais de compostos químicas ocorre na região de MIR (4000 a 200 cm^{-1}), por isso essa região é bastante utilizada para determinação das estruturas, a partir da identificação das funções orgânicas características (BURNS e CIURCZAK, 2007; SKOOG *et al*, 2018; PASQUINI, 2013).

Embora o modelo do oscilador harmônico explique bem as transições fundamentais, ocorrem fenômenos que esse modelo não explica. Para contornar esse problema tem-se o modelo do oscilador anarmônico, que leva em consideração as forças repulsivas entre os átomos e a possibilidade de dissociação da ligação. Nesse contexto foi observado que quando os dois átomos se aproximam, existem forças repulsivas entre as nuvens eletrônicas, fazendo com que a energia potencial aumente mais rapidamente do que prediz o modelo de oscilador harmônico. Por outro lado, existe a possibilidade de dissociação da ligação dos dois átomos. Com o aumento da distância entre eles, ocorre uma diminuição dessa força restauradora e consequentemente uma diminuição na energia potencial. De acordo com esse modelo a energia potencial vibracional não varia periodicamente com a variação da distância entre os núcleos, explicando as transições entre níveis de energias adjacentes, com regra de seleção incluindo as transições fundamentais ($\Delta v = \pm 1$), e também os sobretons e as bandas de combinações ($\Delta v = \pm 2$ ou $\Delta v = \pm 3$), geralmente observados na região NIR (BURNS e CIURCZAK, 2007; SKOOG *et al*, 2018; PASQUINI, 2013).

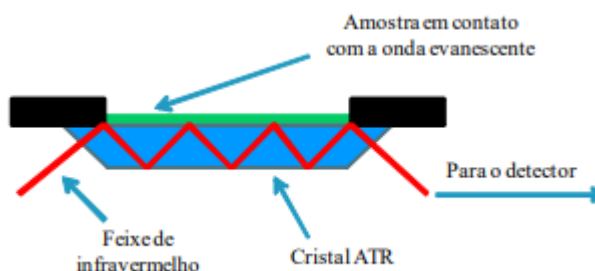
Figura 5 - Curva de potencial de um oscilador harmônico e anarmônico



Fonte: Adaptado de Pasquini (2013).

Em geral, as técnicas utilizadas para obter os espectros de infravermelho diretamente em sólidos pode ser refletância especular, refletância difusa e refletância total atenuada (ATR: *Attenuated Total Reflectance*). No presente trabalho foi utilizada a espectroscopia de infravermelho com o acessório de ATR, que também pode ser utilizado para amostras de filmes, pastas, pós e líquidas. Nessa técnica um feixe de radiação é incidido sobre um cristal, em geral seleneto de zinco, germânio ou diamante, e essa radiação é totalmente refletida na região interna do cristal. Durante o processo de reflexão a radiação penetra na amostra com uma pequena profundidade, mas é suficiente para fornecer informações químicas sobre a amostra analisada. Essa profundidade depende do comprimento de onda, índice de refração dos dois materiais e do ângulo do feixe em relação à interface. Essa radiação penetrante é chamada de onda evanescente. Nos comprimentos de onda que a amostra absorve a radiação, ocorre a atenuação do feixe, por isso essa técnica é chamada de ATR. Na Figura 6 tem-se um esquema de funcionamento do acessório ATR (SKOOG *et al*, 2018).

Figura 6 - Esquema de funcionamento do acessório ATR



Fonte: Aliske (2010).

3.5 QUIMIOMETRIA

Com o avanço da tecnologia é possível obter uma grande quantidade de dados em um período curto de tempo, e um dos grandes desafios é a obtenção de informação útil. Além disso, os sinais registrados são de natureza multivariada, por isso é de fundamental importância determinar quais variáveis contêm a informação relevante para o estudo em questão. Para contornar esse problema utiliza-se métodos matemáticos e estatísticos associados a dados químicos, conhecidas como técnicas quimiométricas (NETO e MOITE, 1997; BEEBE e KOWALSKI, 1987; SIEGEL, 2016).

Essas ferramentas permitem acessar informação relevante do conjunto de dados que podem não ser triviais. Considerando dados espectroscópicos, essas ferramentas são, baseadas na lei de Beer, representada na sua forma multivariada pela Equação 1, que parte do pressuposto que há uma relação linear entre a intensidade de absorção e a concentração do(s) analito(s). A matriz **X**, contém os espectros adquiridos, e pode ser decomposta em um produto de matrizes, a matriz **C** é a concentração do analito, a matriz **S^T** é a matriz das intensidades de sinal do(s) composto(s) puro(s) e a matriz residual **E**.

$$\mathbf{X} = \mathbf{CS}^T + \mathbf{E} \quad \text{Equação 1}$$

As técnicas quimiométricas de reconhecimento de padrões podem ser supervisionadas ou não supervisionadas. Nos métodos supervisionados é

necessária uma informação prévia sobre as amostras (informação que se deseja prever), e o modelo é construído baseado nessas informações, com o objetivo de treinar o modelo, e em seguida é aplicado a um conjunto de dados de predição. Por outro lado, nos métodos não supervisionados não é necessário haver informação inicial de qualquer natureza. Esses métodos identificam similaridades, diferenças, existência de agrupamentos e tendências naturais das amostras (FERREIRA, 2015). Antes de construir qualquer modelo é necessária a análise dos espectros brutos, para tomar algumas decisões, tais como a utilização de técnicas de pré-processamento com o objetivo de remover variações que não estejam associadas ao analito.

3.5.1 Técnicas de pré-processamento

São chamadas de técnicas de pré-processamento qualquer etapa que antecede a análise multivariada. Os espectros obtidos em um determinado equipamento contêm, além da informação do analito de interesse, informações que não são relevantes para o estudo tais como, erros aleatórios, erros sistemáticos e variações relacionadas a fenômenos físicos. Essas informações irrelevantes podem estar associadas ao equipamento, à técnica de aquisição dos espectros e às condições experimentais. Nesse contexto, são utilizadas ferramentas matemáticas para retirar ou minimizar essas informações que podem interferir na construção do modelo. As técnicas de pré-processamento podem ser aplicadas nas amostras ou nas variáveis. Os pré-processamentos de amostras são aqueles aplicados na direção das linhas da matriz de dados, tais como: suavização, derivadas, padronização pelo desvio padrão, correção multiplicativa de sinal. Já as técnicas de pré-processamento das variáveis são aplicadas na direção das colunas da matriz de dados, tem-se seleção de variáveis, como o algoritmo genético e a centragem na média (FERREIRA, 2015; BEEBE, 1987).

3.5.1.1 Suavização pelo método de Savitzky-Golay

Esse método de suavização tem o objetivo de minimizar o ruído presente no conjunto de dados, reduzindo a variação do componente aleatório e aumentando a razão sinal/ruído. O funcionamento desse método consiste na determinação de uma janela de pontos adjacentes no espectro e de um polinômio de grau n , que é ajustado para os pontos dessa janela. A janela é móvel e percorre todo o espectro. A escolha desses dois parâmetros (o grau de polinômio e o número de pontos da janela) deve ser feita cuidadosamente. Com o aumento da janela o pré-processamento filtra mais ruído e ocorre uma maior suavização do espectro. Por outro lado, pode eliminar bandas importantes (SAVITZKY, 1964; FERREIRA, 2015).

3.5.1.2 Derivada

Esse pré-processamento tem o objetivo de corrigir efeitos aditivos e multiplicativos dos espectros, além disso, podem destacar picos e bandas que não estão evidentes. Os efeitos aditivos ocorrem quando há um deslocamento constante positivo ou negativo dos espectros. Para corrigi-los pode-se utilizar a 1ª derivada do espectro. Já os efeitos multiplicativos estão presentes quando os espectros apresentam uma inclinação de linha de base que pode ser corrigida tomando a 2ª derivada do espectro. Esse método também pode ser precedido de uma etapa de suavização com o filtro de Savitzky-Golay. Três parâmetros precisam ser ajustados: o grau de polinômio, o número de pontos da janela e a ordem da derivada. É necessário estar atento ao uso desses pré-processamentos pois, pode aumentar significativamente a quantidade de ruídos (SAVITZKY, 1964; FERREIRA, 2015).

3.5.1.3 Padronização Normal do Sinal

Tem o objetivo de corrigir efeitos aditivos e multiplicativos geralmente causados por interferência devido ao espalhamento da radiação, bastante comum em espectro

de refletância de pós, devido principalmente a diferenças no tamanho de partícula. Matematicamente a Padronização Normal do Sinal (SNV: *Standard Normal Variate*) realiza a centragem na média, na qual o valor médio de todas as variáveis em uma amostra é subtraído de cada variável individual e em seguida uma normalização pelo desvio padrão das amostras, como mostra a Equação 2. Onde x_{ijSNV} é o elemento da matriz corrigida, \bar{x}_i é a média das intensidades das variáveis na amostra i e s_i é o desvio padrão das intensidades das variáveis na amostra i . Esse pré-processamento atribui as variáveis que apresentam um maior desvio da média uma maior ponderação (FERREIRA et al., 2009; RINNAN et al., 2009; SIEGEL, 2016).

$$x_{ij \text{ Corr}} = \frac{x_{ij} - \bar{x}_i}{s_i} \quad \text{Equação 2}$$

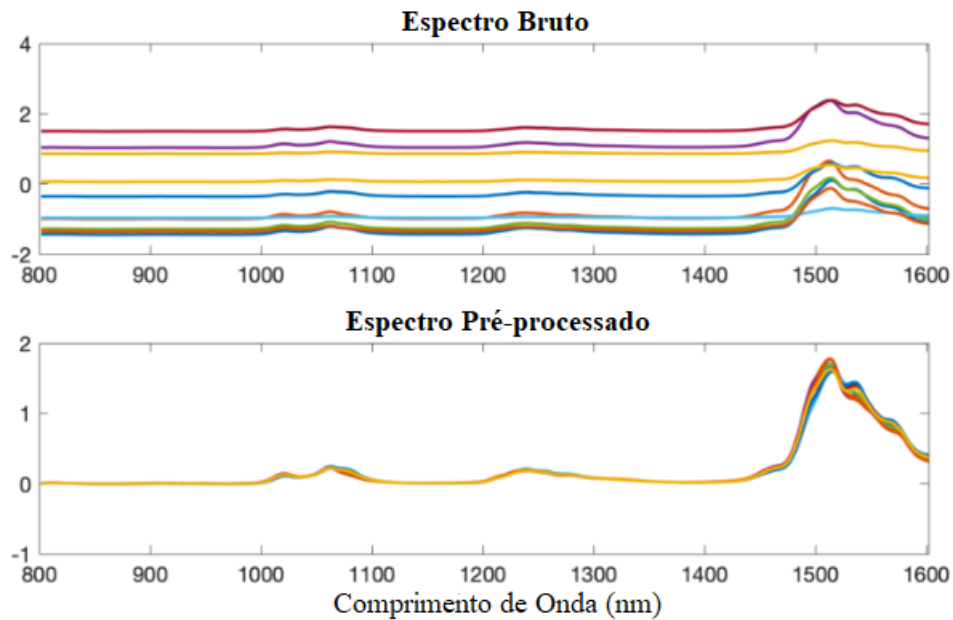
3.5.1.4 Correção Multiplicativa Sinais

A Correção Multiplicativa de Sinais (MSC: *Multiplicative Scattering Correction*) também tem o objetivo de corrigir efeitos causados pelo espalhamento da radiação, removendo os efeitos aditivos e multiplicativos. Esse pré-processamento é obtido através da regressão dos espectros medidos por um espectro de referência, geralmente utiliza-se o espectro médio. A partir da regressão são determinados os coeficientes de correção que são a inclinação e o intercepto. As Equações 3 e 4 representam esse pré-processamento matematicamente, na qual, \mathbf{X} é a matriz dos espectros medidos, \mathbf{X}_{ref} é o espectro de referência, \mathbf{E} é a matriz residual, \mathbf{X}_{corr} é matriz com os espectros corrigidos. Os parâmetros escalares \mathbf{a} e \mathbf{b} variam de amostras para amostras corrigindo o efeito aditivo e multiplicativo, respectivamente. A Figura 7 mostra um espectro corrigido utilizando o MSC (FERREIRA et al., 2009; RINNAN et al., 2009).

$$\mathbf{X} = \mathbf{a} + \mathbf{bX}_{\text{ref}} + \mathbf{E} \quad \text{Equação 3}$$

$$\mathbf{X}_{\text{MSC}} = \frac{\mathbf{X} - \mathbf{a}}{\mathbf{b}} \quad \text{Equação 4}$$

Figura 7 - Correção do espectro utilizando MSC



Fonte: Adaptado de Wise *et al* (2016).

3.5.1.5 Centragem na Média

A centragem na média (MC: *Mean Center*) é aplicado nas variáveis, com o objetivo de realizar uma translação para que o centro do conjunto de dados esteja na origem dos eixos, de modo que os desvios sejam computados em relação à média. Matematicamente, cada valor da matriz é subtraído pela média da coluna na qual se encontra, representado pela equação 5. Onde x_{ijCorr} é o elemento da matriz corrigida, x_{ij} é o i-ésimo elemento da j-ésima coluna e \bar{x}_j é a média dos elementos da j-ésima coluna.

$$x_{ijCorr} = x_{ij} - \bar{x}_j \quad \text{Equação 5}$$

3.5.1.6 Mínimos Quadrados Generalizados Ponderados

O pré-processamento dos Mínimos Quadrados Generalizados Ponderados (GLSW: Generalized Least Squares Weighting) atua como um filtro cujo objetivo é reduzir diferenças entre determinadas amostras. Essa técnica estima as diferenças entre amostras de um grupo que deveriam ser similares e utiliza esses dados para minimizar a atuação dos interferentes. Matematicamente, uma matriz contendo a diferença dos espectros (\mathbf{X}_{dif}) é calculada e removida dos dados originais. Uma forma de calcular essa matriz é utilizando os valores de \mathbf{Y} nos modelos de regressão, pois amostras com o mesmo valor de \mathbf{Y} deveriam possuir espectros semelhantes no conjunto de dado \mathbf{X} , logo \mathbf{X}_{dif} deve ser ortogonal a \mathbf{Y} (SILVA, 2018; ZORZETTI *et al*, 2011).

Para a obtenção da matriz \mathbf{X}_{dif} , primeiramente a matriz \mathbf{X} é organizada a partir da matriz \mathbf{Y} de modo que amostras similares ficassem próximas, ou seja, as linhas das matrizes \mathbf{X} e \mathbf{Y} são organizadas em ordem crescente de valor de \mathbf{Y} , quando têm-se apenas uma propriedade de interesse utiliza-se o vetor \mathbf{y} . Em seguida são determinadas as diferenças entre as amostras próximas utilizando a derivada de Savitzky-Golay aplicando em cada coluna da matriz \mathbf{X} organizada, produzindo a matriz \mathbf{X}_{dif} . A mesma derivada é aplicada ao vetor \mathbf{y} , calculando o \mathbf{y}_{dif} . Devido ao fato das derivadas utilizarem uma janela de pontos que podem ser alteradas, a matriz \mathbf{X}_{dif} pode conter informações que não devem ser removidas. Isso ocorre pois, algumas das diferenças das linhas podem ter sido feitas em grupo de amostras com valores diferentes de \mathbf{y} . Nesse contexto é calculada uma matriz de pesos \mathbf{W} utilizando o vetor \mathbf{y}_{dif} e o seu desvio padrão. Essa matriz de pesos é utilizada para calcular a matriz de covariância da matriz \mathbf{X}_{dif} (Equação 6). Essa matriz de covariância é submetida à decomposição por valores singulares, representado pela Equação 7. A matriz de valores singulares \mathbf{S} é utilizada para determinar a matriz de filtro \mathbf{G} , como mostra a Equação 8 e 9 (SILVA *et al*, 2018; ZORZETTI *et al*, 2011).

$$\mathbf{C} = \mathbf{X}_{\text{dif}}^2 \mathbf{W}^2 \mathbf{X}_{\text{dif}} \quad \text{Equação 6}$$

$$\mathbf{C} = \mathbf{V} \mathbf{S}^2 \mathbf{V}^T \quad \text{Equação 7}$$

$$\mathbf{D} = \sqrt{\frac{\mathbf{S}^2}{\alpha} + \mathbf{D}_1} \quad \text{Equação 8}$$

$$\mathbf{G} = \mathbf{V} \mathbf{D}^{-1} \mathbf{V}^T \quad \text{Equação 9}$$

O α é um parâmetro que mede o efeito do filtro nas matrizes originais. Quando α é grande, o efeito filtro diminui, retirando menos diferenças entre os espectros. Por outro lado, quando o valor de α é pequeno o efeito do filtro aumenta. O valor de α deve ser escolhido com bastante cuidado para não remover variabilidade importante para o estudo em questão. A matriz \mathbf{D}_1 é uma matriz diagonal composta apenas de 1 e \mathbf{V} são os autovalores obtidos na decomposição (SILVA *et al*, 2018; ZORZETTI *et al*, 2011).

3.5.2 Análise dos Componentes Principais

A análise dos componentes principais (PCA: *Principal Component Analysis*) é uma ferramenta de análise exploratória não supervisionada, com o objetivo de reduzir a dimensionalidade do conjunto de dados maximizando a variância. Isso permite encontrar similaridades e diferenças naturais entre as amostras e determinar quais variáveis mais contribuem para a variabilidade. A PCA gera um novo espaço de variáveis ortogonais, essas variáveis são chamadas de componentes principais (PC: *Principal Component*), formadas a partir de combinações lineares das variáveis originais e tem a direção de maior variabilidade dos dados. Matematicamente, o conjunto de dados \mathbf{X} é decomposto em uma multiplicação de duas matrizes, escores \mathbf{T} e pesos \mathbf{P}^T , mais a matriz de erro \mathbf{E} , representado pela Equação 10 (BRO e SMILDE, 2014; FERREIRA; 2015).

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

A matriz de escores \mathbf{T} são as coordenadas das amostras no novo sistema de eixo formados e possui dimensão $n \times k$. A matriz de pesos \mathbf{P}^T são os coeficientes da combinação linear, geometricamente falando são os cossenos dos ângulos entre as variáveis originais e as componentes principais e possui dimensão $k \times \lambda$. A matriz de erro \mathbf{E} possui as informações que não são explicadas pelo modelo, com dimensão da matriz \mathbf{X} , $n \times \lambda$. Onde n é o número de amostras, k o número de componentes principais e λ o número de variáveis originais. Todas as PCs são ortogonais entre si. A PC1 explica a maior variabilidade do conjunto de dados; já a PC2 é ortogonal à primeira, e explica a maior variabilidade do conjunto de dados que a primeira PC não

explica e assim sucessivamente. As primeiras PCs possuem maior variabilidade dos dados, nesse contexto a variabilidade explicada em uma PC diminui quando o número de PC cresce. Definir o número de componentes principais em um modelo é bastante importante, pois influencia diretamente na matriz de resíduo **E**, e consequentemente na determinação de amostras anômalas (WISE *et al*, 2006; FERREIRA; 2015).

Analisando a matriz de resíduos é possível determinar alguns parâmetros importantes, tais como, T^2 de Hotelling e Q residual. O T^2 de Hotelling é uma extensão do teste t podendo ser aplicado também aos escores da PCA, como mostra a equação 11 (BRO e SMILDE, 2014).

$$T_i^2 = \frac{\mathbf{t}_i^T (\mathbf{T}^T \mathbf{T})^{-1} \mathbf{t}_i}{\mathbf{I} - 1} \quad \text{Equação 11}$$

Onde T^2 é o parâmetro T^2 de Hotelling, **T** é a matriz de escores de todas as amostras ($I \times R$), \mathbf{t}_i é um vetor dos escores da i-ésima amostra ($R \times 1$). A partir dos resultados do T^2 de Hotelling é possível a construção dos limites de confiança que são bastante importantes para a determinação de amostras anômalas. Outro parâmetro que também deve ser analisado é o Q residual. Matematicamente, O Q residual é a soma dos quadrados dos resíduos de cada amostra. Esse parâmetro tem o objetivo de determinar as amostras que não são bem representadas pelo modelo. O gráfico do Q residual versus o T^2 de Hotelling é chamado de gráfico de Influência (BRO; SMILDE, 2014).

3.5.3 Regressão por Mínimos Quadrados Parciais

A regressão por mínimos quadrados (PLS: *Partial least squares*) é uma técnica de calibração multivariada, que tem como objetivo a construção de um modelo que é capaz de determinar um parâmetro de interesse a partir de uma matriz de dados espectrais. Para realizar essa análise são necessárias uma matriz de dados **X**, e também um vetor **y** ou uma matriz **Y** que contenha a(s) propriedade(s) de interesse, que se deseja prever. Quando apenas um parâmetro é de interesse tem-se o vetor **y**

com dimensão $n \times 1$. Por outro lado, quando se tem m propriedades de interesse utilizasse, utiliza-se uma matriz \mathbf{Y} $n \times m$ (FERREIRA; 2015).

O modelo PLS não leva em consideração apenas a variabilidade presente na matriz \mathbf{X} , como ocorre na PCA, mas sim, a correlação entre \mathbf{X} e \mathbf{Y} . Nesse contexto, é realizada uma decomposição para a matriz \mathbf{X} e para a matriz \mathbf{Y} , onde \mathbf{T} e \mathbf{U} são os escores para \mathbf{X} e \mathbf{Y} , respectivamente, representada pelas Equações 12 e 13. No entanto, diferentemente da PCA, é determinada uma relação entre os escores \mathbf{T} e \mathbf{U} , para maximizar a correlação entre \mathbf{X} e \mathbf{Y} , como mostra a Equação 14.

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E} \quad \text{Equação 12}$$

$$\mathbf{Y} = \mathbf{UL}^T + \mathbf{F} \quad \text{Equação 13}$$

$$\mathbf{U} = \mathbf{TW} \quad \text{Equação 14}$$

Em seguida, é utilizada essa correlação encontrada na Equação 14, para determinar os coeficientes de regressão para o modelo $\hat{\mathbf{B}}$, que serão utilizados para prever a(s) propriedade(s) de interesse a partir dos seus dados espectrais, representados pela Equação 15 e 16. As novas variáveis não são mais ortogonais entre si e são chamadas de variáveis latentes (LV: *Latent Variables*), pois não estão mais no sentido de maior variabilidade dos dados e sim na direção de maior correlação entre as matrizes \mathbf{X} e \mathbf{Y} (FERREIRA; 2015; WOLD et al; 2001; BEBEE et al, 1987).

$$\hat{\mathbf{B}} = \mathbf{P}(\mathbf{P}^t\mathbf{P})^{-1}\mathbf{WL} \quad \text{Equação 15}$$

$$\hat{\mathbf{y}} = \mathbf{XB} \quad \text{Equação 16}$$

Para verificar o desempenho dos modelos construídos, têm-se algumas figuras de mérito tais como o SEP, definido como o desvio padrão dos resíduos previstos, representada pela Equação 17 e 18 (WISE et al, 2006; NAES et al, 2004).

$$\text{SEP} = \sqrt{\sum_{i=1}^{N_p} (\hat{y}_i - y_i - \text{BIAS})^2 / (N_p - 1)} \quad \text{Equação 17}$$

$$BIAS = \sqrt{\sum_{i=1}^{N_P} (\hat{y}_i - y_i) / N_P} \quad \text{Equação 18}$$

Onde \hat{y}_i é o valor da variável y estimado pelo modelo para a i -ésima amostra, y_i é o valor conhecido para a i -ésima amostra, N_P é o número total de amostras para o conjunto de previsão. A relação entre o SEP e o BIAS é chamado de erro médio quadrático de previsão (RMSEP: *Root Mean Square Error of Prediction*) representado pela Equação 18 (WISE *et al*, 2006; NAES *et al*, 2004).

$$RMSEP^2 = SEP^2 + BIAS^2 \quad \text{Equação 19}$$

Para entender o significado desses erros é necessário definir alguns conceitos como a precisão e exatidão. A precisão mede a diferença entre medições repetidas, enquanto a exatidão mede a diferença entre o valor de y medido e o estimado pelo modelo. Nesse contexto o SEP mede a presença do erro aleatório e consequentemente a precisão, enquanto o BIAS é uma medida do erro sistemático e, portanto, está relacionado com a exatidão (WISE *et al*, 2006; NAES *et al*, 2004).

Antes de avaliar a capacidade preditiva dos modelos representados pelas Equações 17, 18 e 19, é necessário avaliar o ajuste do modelo para os dados de calibração e validação cruzada. Essa análise incluirá um valor de erro médio quadrático de validação cruzada (RMSECV: *Root Mean Square Error of Cross-Validation*) junto com o valor de erro médio quadrático de calibração (RMSEC: *Root Mean Square Error of Calibration*). Esses erros são bastante úteis para determinar o número ideal de variáveis latentes em um modelo de calibração (WISE *et al*, 2006; NAES *et al*, 2004).

Após a construção dos modelos PLS é possível observar as variáveis que tiveram maior influência para a construção desses modelos. Essa análise é realizada a partir dos escores de importância da variável na projeção (VIP: *Variable Importance in Projection*). Os escores VIPs estimam a importância das variáveis usadas no modelo PLS e são calculados a partir dos coeficientes de regressão. De uma forma generalizada, pontuações próximas ou maiores que 1 podem ser consideradas importante para o modelo em estudo, enquanto variáveis com

pontuações significativamente menores que 1 são menos importantes e podem ser excluídas do modelo (WISE *et al*, 2006).

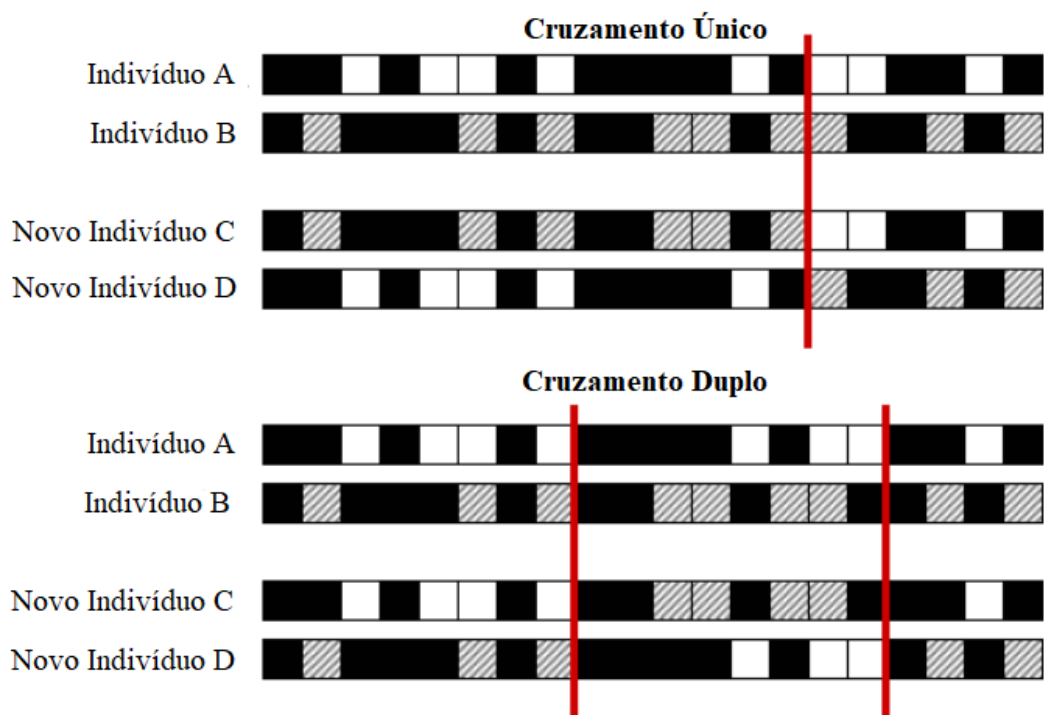
3.5.4 Seleção de variáveis por algoritmo genético

O algoritmo genético (GA: *Genetic Algorithms*) poder ser utilizado com o objetivo de identificar um subconjunto de variáveis que melhor descreve um modelo de regressão. Os modelos de regressão se beneficiam com esse método, pois algumas variáveis podem ser irrelevantes, conter ruído ou sinal de interferência que prejudicam o modelo. Essa técnica é inspirada na teoria da evolução biológica e seleção natural da “sobrevivência do mais apto”, na qual apenas as variáveis dos modelos com alto desempenho tem maior probabilidade de serem incluídas no conjunto de variáveis de modelos subsequentes. Além disso, também pode ser realizada mutação, na qual variáveis são adicionadas ou retiradas em um indivíduo sem ser diretamente herdada da geração anterior. Nesse contexto, a seleção randômica da população e a mutação garante o efeito estocástico do algoritmo. (WISE *et al*, 2006; MEHMOOD *et al*, 2012).

No contexto da calibração, o algoritmo genético em geral utiliza uma abordagem iterativa para determinar o subconjunto de variáveis que fornece o menor erro médio quadrático da validação cruzada (RMSECV: *root-mean-square error of cross validation*). O primeiro passo é determinar a população, que consiste em um grande número de conjuntos de variáveis aleatórias que são usadas para calcular o RMSECV para um modelo de regressão, como o PLS. Cada conjunto de variável é chamado de indivíduo, e são representados por números binários determinando as variáveis inclusas ou não naquela etapa. O segundo passo é descartar o conjunto de variável que obtiveram o maior RMSECV, reduzindo a população pela metade, Para substituir esses indivíduos, o GA gera novos indivíduos por dois métodos: cruzamento único ou duplo, representado pela Figura 8. Para o cruzamento único, as variáveis de dois indivíduos (A e B) são divididos em um ponto aleatório e uma parte de A é trocada por uma parte de B gerando os indivíduos C e D. Por outro lado, o cruzamento duplo seleciona 2 pontos aleatórios das variáveis e a parte do

meio de A e B são trocadas. Além disso, todos os genes podem sofrer mutações, permitindo adicionar ou remover variáveis que podem estar super ou sub-representadas na população. Depois que o tamanho da população for restaurado o processo retorna para a avaliação dos modelos a partir do RMSECV. Esse método é iterativo, os critérios de paradas podem ser um número finito de iterações ou uma determinada porcentagem dos indivíduos da população estar usando subconjunto de variáveis idênticos (WISE *et al*, 2006; MEHMOOD *et al*, 2012).

Figura 8 - Esquema de cruzamento único e duplo



Fonte: Adaptado de Wise *et al* (2006).

4 MATERIAIS E MÉTODOS

4.1 AMOSTRAS

Para esse estudo foram utilizados dois conjuntos de amostras. O primeiro, composto por amostras envelhecidas artificialmente e o segundo com amostras envelhecidas naturalmente.

4.1.1 Amostras Envelhecidas Artificialmente

Foram comprados, no comércio de Recife/PE, Brasil, 10 diferentes tipos e marcas de papéis, entre elas 3 marcas de papéis reciclado, 2 de papéis couchê e 5 marcas de papéis sulfite, com as especificações mostradas na Tabela 1. Para cada marca foram selecionadas, aleatoriamente, três folhas, a partir das quais foram cortados quadrados do centro de cada folha com 4 cm de comprimento e altura. Foram adquiridos 2 espectros de cada amostra. Em seguida, essas amostras foram submetidas ao tratamento térmico a 100 °C, para isso foi utilizado uma estufa de marca Olide fcz. Os espectros foram monitorados nos seguintes períodos de tempo: 10 minutos, 30 minutos, 1 hora, 2 horas, 3 horas, 24 horas, 48 horas, 72 horas, 96 horas, 1 Semana, 2 semanas, 3 semanas, 4 semanas, 5 semanas, 6 semanas, 7 semanas, 8 semanas, 2 meses e 3 meses. Foram também obtidos espectros antes do tratamento térmico (tempo zero).

Tabela 1 - Especificações dos papéis

Marca	Modelo	Tipo	Gramatura	Observações	Sigla
Syspaper	Syspaper	Couche	180 g/m ²	Resistente a água; secagem rápida; A4 21 x 29.7 cm	C1
Syspaper	Syspaper	Couche	180 g/m ²	Resistente a água; secagem rápida; A4 21 x 29.7 cm	C2
Chamex	Chamex	Reciclado	75 g/m ²	A4 21 x 29.7 cm	R1
Jandaia	Turma do Jandainha	Reciclado	75 g/m ²	A4 21 x 29.7 cm; para pintar, trabalhos em geral, colar, desenhar, etc	R2
Jandaia	Eco millennium multiuso	Reciclado	75 g/m ²	A4 21 x 29.7 cm; para pintar, trabalhos em geral, colar, desenhar, etc	R3
Chamex	Chamex Super	Sulfite	90 g/m ²	A4 21 x 29.7 cm; mais encorpado	S1
Chamex	Chamex Multi	Sulfite	75 g/m ²	A4 21 x 29.7 cm; para o dia a dia	S2
Chamex	Chamex office	Sulfite	75 g/m ²	A4 21 x 29.7 cm; uso profissional	S3
Suzano	Suzano Report	Sulfite	75 g/m ²	A4 21 x 29.7 cm	S4
Jandaia	Turma do Jandainha	Sulfite	120 g/m ²	A4 21 x 29.7 cm; para pintar, trabalhos em geral, colar, desenhar, etc	S5

Fonte: A autora (2020).

4.1.2 Amostras Envelhecidas Naturalmente

Para construção desse conjunto de dados, foram utilizados documentos de 20 anos diferentes (1998, 1999, 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017 e 2018) que estavam armazenados no mesmo local, sob as mesmas condições. Essas amostras eram papéis sulfites, especificamente provas e trabalhos armazenados em envelopes, ao abrigo de luz, em uma sala na UFPE, a temperatura do ambiente, cujo intervalo pode ter variado de 22 a 35 °C sem controle de umidade. Para cada ano foram

selecionados 3 documentos e de cada documento 2 folhas foram utilizadas, obtendo um espectro por folha.

4.2 AQUISIÇÃO DOS ESPECTROS

Para a aquisição dos espectros foi utilizado o espectrômetro *Spectrum frontier* da *Perkin Elmer*, com o acessório de Refletância Total Atenuada (ATR: *Attenuated Total Reflectance*). A região espectral utilizada foi a região do MIR, os parâmetros utilizados foram: faixa espectral entre 4000 e 650 cm^{-1} , resolução de 4 cm^{-1} , intervalo de dados de 1 cm^{-1} e 16 varreduras. As análises foram realizadas diretamente no papel, sem danificar as amostras e sem necessitar de nenhum preparo.

4.3 TRATAMENTO DE DADOS

Para o tratamento dos dados foram utilizadas técnicas quimiométricas, tais como pré-processamentos, seleção de variáveis, modelos PCA e PLS. Para esse estudo, foram utilizados os softwares Matlab e o PLS Toolbox da Eigenvector. Para as amostras envelhecidas artificialmente foram realizados modelos PCA para avaliar o processo de degradação para cada tipo de papel, e avaliar a variabilidade presente nas marcas de papéis sulfites. Para a construção desses modelos diversos pré-processamentos foram utilizados, tais como, SNV, MSC, derivada, suavização e centragem na média. Em seguida foram separados os conjuntos de dados de calibração e previsão. O conjunto de previsão é composto por todas as amostras de uma marca e as amostras de outras marcas compuseram o conjunto de calibração. Para determinar a marca que iria compor o conjunto de previsão, foi realizado a PCA apenas com os papéis sulfites e foi selecionada a marca cujas amostras estavam localizadas no centro do modelo. Os modelos PLS foram utilizados com o objetivo de prever o tempo de exposição ao envelhecimento artificial. Para isso foi utilizado os seguinte pré-processamento e seleção de variáveis: (i) SNV + MC, (ii) SNV + MC + GA, (iii) SNV + GLSW + MC.

Para as amostras envelhecidas naturalmente, a PCA foi utilizada para determinar semelhanças e diferenças entre as amostras. Em seguida foram determinados os conjuntos de dados de calibração e previsão. Utilizou-se um documento de cada ano para compor o conjunto de previsão, que foi selecionado aleatoriamente, e o restante foi utilizado no conjunto de calibração. Os modelos PLS foram utilizados para a datação do documento. Os modelos PLS foram construídos com o seguinte pré-processamento e seleção de variáveis: (i) SNV + MC, (ii) SNV + MC + GA, (iii) SNV + GLSW + MC.

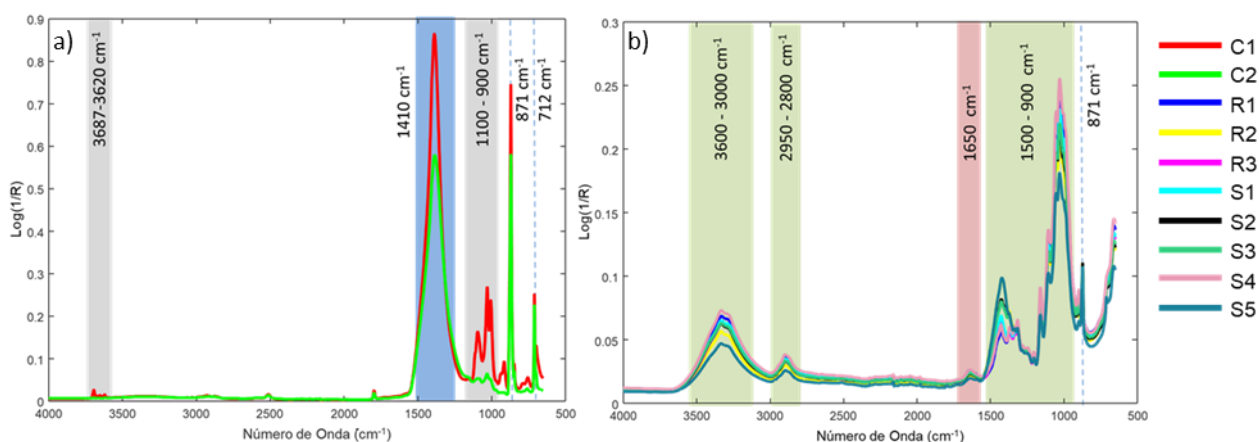
5 RESULTADOS E DISCUSSÃO

5.1 ENVELHECIMENTO ARTIFICIAL

5.1.1 Atribuições de bandas dos espectros

Após a aquisição dos espectros, foi realizada a atribuição de bandas para os espectros obtidos no tempo zero (T0), antes de ser submetido às condições de envelhecimento artificial. Na Figura 9 tem-se a média dos espectros no tempo T0, onde estão destacadas as principais bandas características.

Figura 9 - Média dos diferentes tipos de papéis no tempo zero T0 (a) papéis couchês (b) papéis reciclados e sulfites



Fonte: A autora (2020).

Analisando a Figura 9 observou-se que os espectros dos papéis couchês (C1 e C2) são diferentes dos espectros dos papéis reciclados e sulfites. Nos papéis couchês observou-se a presença da banda em torno de 1410 cm^{-1} e dois picos em torno de 871 e 712 cm^{-1} , relacionados com o estiramento assimétrico do C-O, estiramento simétrico de C-O e deformação de flexão no plano presente em ligações O-C-O, respectivamente. Essas bandas estão relacionadas com o carbonato de cálcio, CaCO_3 . As bandas que se encontram nos intervalos de 3687 a 3620 cm^{-1} e

entre 1100 a 900 cm^{-1} , mais evidentes nos papéis C1, ocorrem devido à presença da caulinita, $\text{Al}_2\text{Si}_2\text{O}_5(\text{OH})_4$. O CaCO_3 e a $\text{Al}_2\text{Si}_2\text{O}_5(\text{OH})_4$ são compostos inorgânicos frequentemente utilizados para o revestimento do papel (CAUSIN et al, 2010 e UDRISTIOIU et al, 2012).

Para os papéis reciclados e sulfites é possível perceber a semelhança entre seus espectros. A banda que se encontra entre 3600 e 3000 cm^{-1} está associada à vibração de estiramento de hidrogênio em grupos de O-H da celulose. Já a absorção em aproximadamente 2890 cm^{-1} é devido ao estiramento simétrico de C-H em ligações CH, CH_2 e CH_3 da celulose (CAUSIN et al, 2010). A banda que se encontra em torno de 1640 cm^{-1} é devido à presença de água adsorvida, enquanto que as bandas entre 1500 e 900 cm^{-1} constituem a região de impressão digital, e estão detalhadas na Tabela 2 (HAJJI et al, 2016; ZIEBA-PALUS et al, 2016). Já a banda em torno de 871 cm^{-1} é devido ao estiramento simétrico de C-O, do carbonato de cálcio (CAUSIN et al, 2010).

Tabela 2 - Atribuições de bandas para a região de impressão digital da celulose

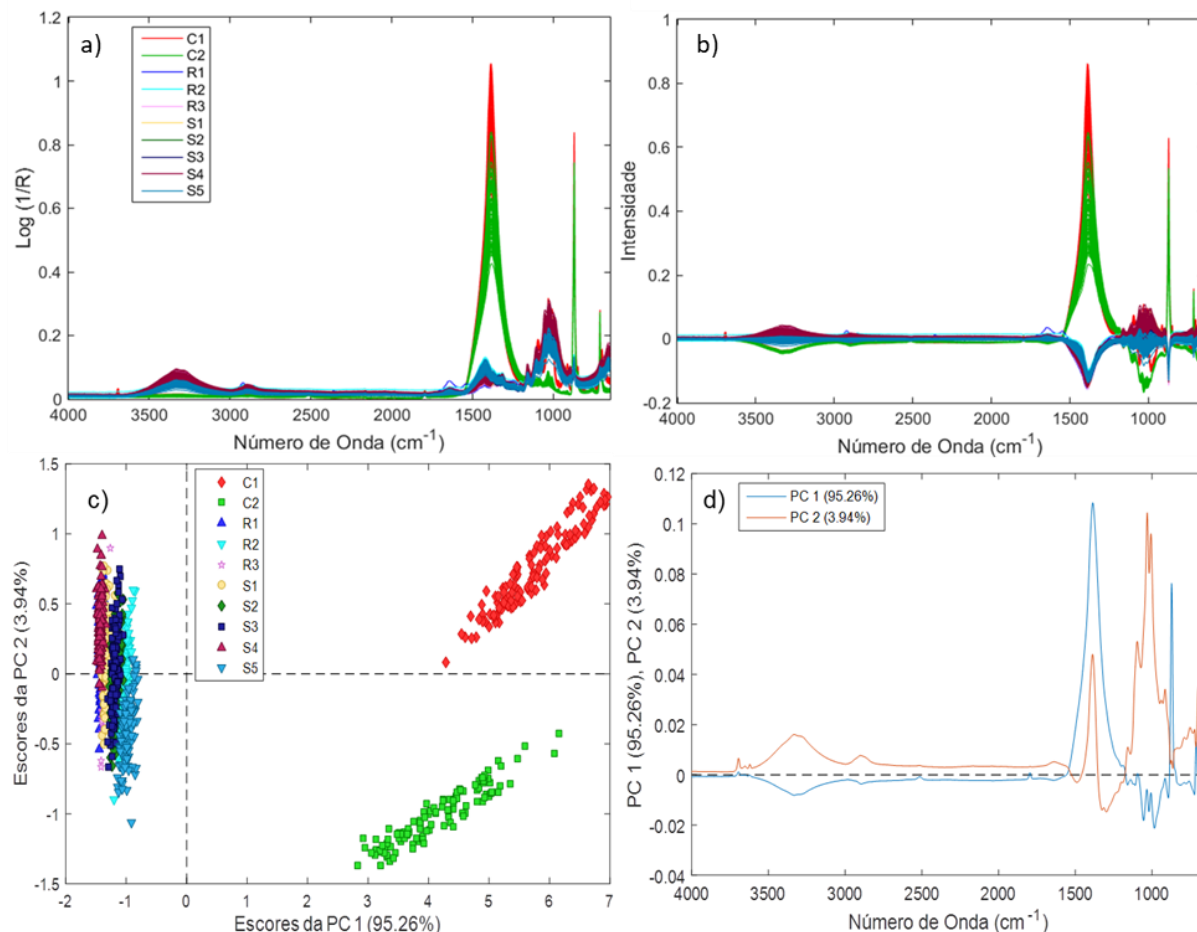
Número de Onda (cm^{-1})	Atribuição de banda
1430	Deformação tesoura de CH_2 , Deformação de CH e flexão no plano de OCH.
1375	Deformação de OH e CH_2 e flexão no plano de CH.
1315	Deformação <i>rocking</i> de CH
1160	Deformação de C-OH e C- CH_2 e estiramento assimétrico de C-O-C
1105	Estiramento glicosídico C-O-C
1055	Estiramento de álcool secundário C-O
1025	C-C, estiramento de C-OH e anel CH.

Fonte: Adaptado de Zieba-Palus et al (2016); Cousin et al (2010).

5.1.2 Análise dos componentes principais

Para as amostras submetidas à condição de envelhecimento por tratamento térmico, foram adquiridos os espectros em todos os tempos de envelhecimentos, obteve-se, portanto uma matriz de dados de 1260 espectros e 3351 variáveis. Entretanto, houve problemas na estufa e as amostras referente ao tempo de 4 meses ficaram comprometidas, fazendo com que seus espectros fossem retirados da matriz de dados, resultando em 1200 espectros e 3351 variáveis. Antes de construir os modelos PLS, foi feito um modelo PCA para estudar a variabilidade do conjunto de dados. Para uma primeira análise, foi utilizado como pré-processamento a centragem na média. Nas Figura 10a e Figura 10b tem-se os espectros brutos e os espectros pré-processados e na Figura 10c e Figura 10d o modelo PCA com os scores e os pesos, respectivamente.

Figura 10 - Gráfico da PCA para todos os papéis em todos os tempos de envelhecimento, para as amostras submetidas ao tratamento térmico (a) espectros brutos, (b) espectros pré-processados com centragem na média, (c) escores da PC1 vs PC2 e (d) pesos da PC1 e da PC2.



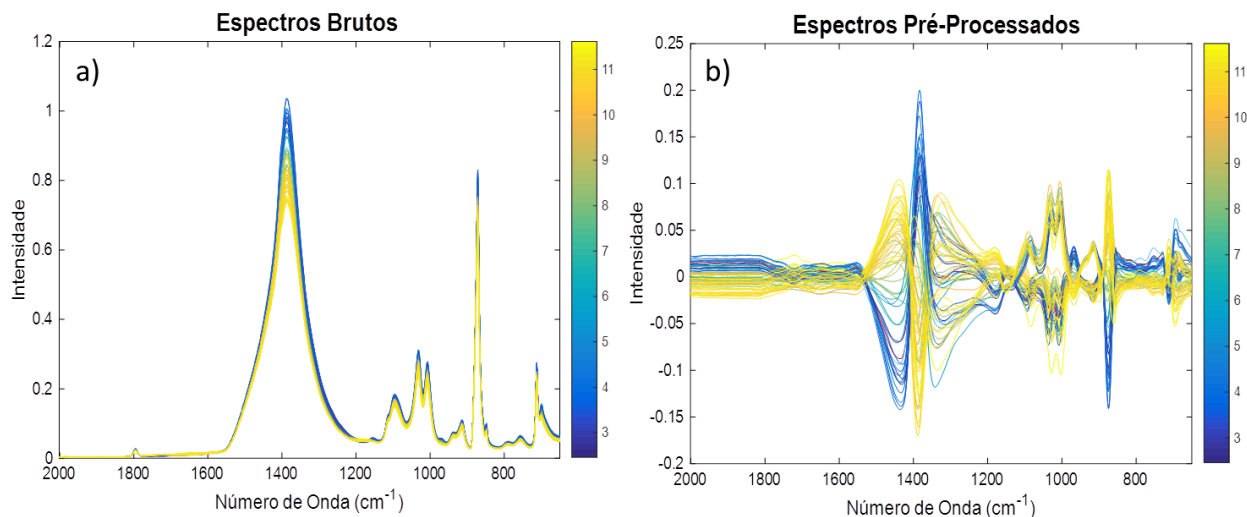
Fonte: A autora (2020).

O modelo PCA com duas componentes principais explica 99,20 % e, a Figura 10c mostra o gráfico dos escores do modelo. A PC1 está relacionada com a diferença entre os papéis couchês, localizados na região positiva dos escores, e os papéis reciclados e sulfites, localizados na região negativa dos escores. Essa diferença pode ser observada na Figura 10d, no gráfico dos pesos, é possível observar que na região positiva da PC1 é referente às absorções que ocorrem nas seguintes regiões espectrais: 1410 cm^{-1} , 871 cm^{-1} , 712 cm^{-1} . Todas essas bandas são referentes à presença do carbonato de cálcio, presentes em concentrações significativas em ambos os papéis couchês. Os valores negativos dos pesos da PC1 estão relacionados a grupos funcionais da celulose (3400 cm^{-1} e 2890 cm^{-1}), e a sua

região de impressão digital (1500 cm^{-1} a 900 cm^{-1}), bastante evidentes nos espectros dos papéis reciclados e sulfites. É possível observar que, para os papéis couchês, não é possível obter informações da celulose com os espectros, provavelmente, devido à camada de revestimento presente, fazendo com que a penetração da radiação nesse tipo de papel não seja suficiente para atravessar o revestimento e fornecer as informações sobre a celulose. O escore da PC2 diferencia as marcas dos papéis couchês. Essa diferença fica evidente analisando a parte positiva dos pesos da PC2, em que a banda de maior contribuição está entre 1200 cm^{-1} e 900 cm^{-1} , região de maior diferença entre os tipos de papéis couchês, que está relacionada com a presença da caulinita presente apenas nos papéis C1.

Para um estudo mais detalhado do processo de degradação de cada tipo de papéis foram construídos modelos PCAs individualmente, para cada marca e tipo de papéis. Para a construção desses modelos foi obtido a média dos dois espectros de uma mesma folha, resultando em um espectro por folha. Em seguida foi realizada a seleção da região espectral de trabalho, de modo que a região espectral utilizada foi de 2000 cm^{-1} a 650 cm^{-1} , pois a maior variação no perfil espectral foi encontrada nessa região, diminuindo significativamente o ruído. Para realizar os modelos PCAs diversos pré-processamentos foram testados, tais como, SNV, MSC, derivada, suavização e centragem na média, os melhores resultados estão apresentados a seguir. A Figura 11 mostra os espectros brutos e os espectros pré-processados com suavização pelo filtro de Savitzky-Golay (com polinômio de 2 ordem e janela de 33 pontos), SNV e centragem na média para os papéis Couchês C1, respectivamente.

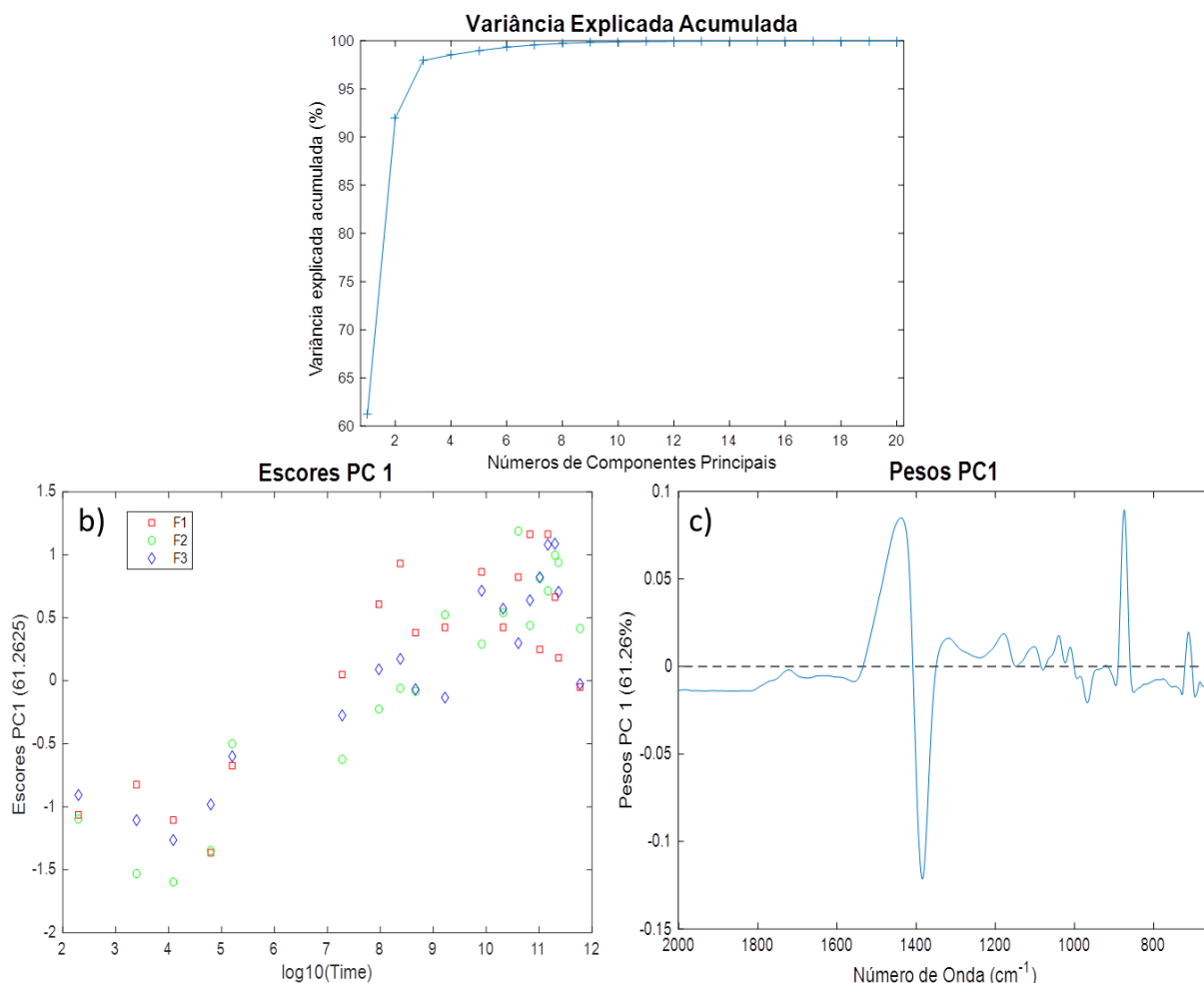
Figura 11 - Gráfico do efeito do pré-processamento para os papéis couchês C1 (a) espectros brutos e (b) espectros pré-processados com suavização pelo filtro de Savitzky-Golay (com polinômio de 2 ordem e janela de 33 pontos), SNV e centragem na média.



Fonte: A autora (2020).

Na Figura 11, as amostras com a coloração azul são as com menos tempo de exposição em uma dada condição de envelhecimento. Por outro lado, as amostras com a coloração amarela são as mais envelhecidas. O tempo de exposição está em minutos e é utilizado uma escala logarítmica. Na Figura 11a é possível observar que, de forma geral, a intensidade das bandas está diminuindo com o aumento do tempo de exposição à condição de envelhecimento. Esse comportamento fica mais evidente na Figura 11b, em torno de 1460 cm^{-1} , 874 cm^{-1} e 712 cm^{-1} . Em seguida, foi construído o modelo PCA para esses dados, com o objetivo de determinar a componente principal que melhor representa a informação da mudança do papel ao longo do tempo. Foram observadas as 5 primeiras PCs *versus* o log do tempo, para cada tipo e marca de papel individualmente. Esse estudo foi realizado detalhadamente no meu trabalho de conclusão de curso (SILVA, 2018). O resultado da PC que contém a informação da mudança do papel ao longo do tempo para o papel couchê C1 foi a PC1, representado na Figura 12.

Figura 12 - Gráfico da PCA para os papéis couchês C1 (a) Variância explicada acumulada para o modelo (b) escores da PC1 versus o log do tempo (F1: Folha 1, F2: Folha 2 e F3: Folha 3) (c) pesos da PC1.

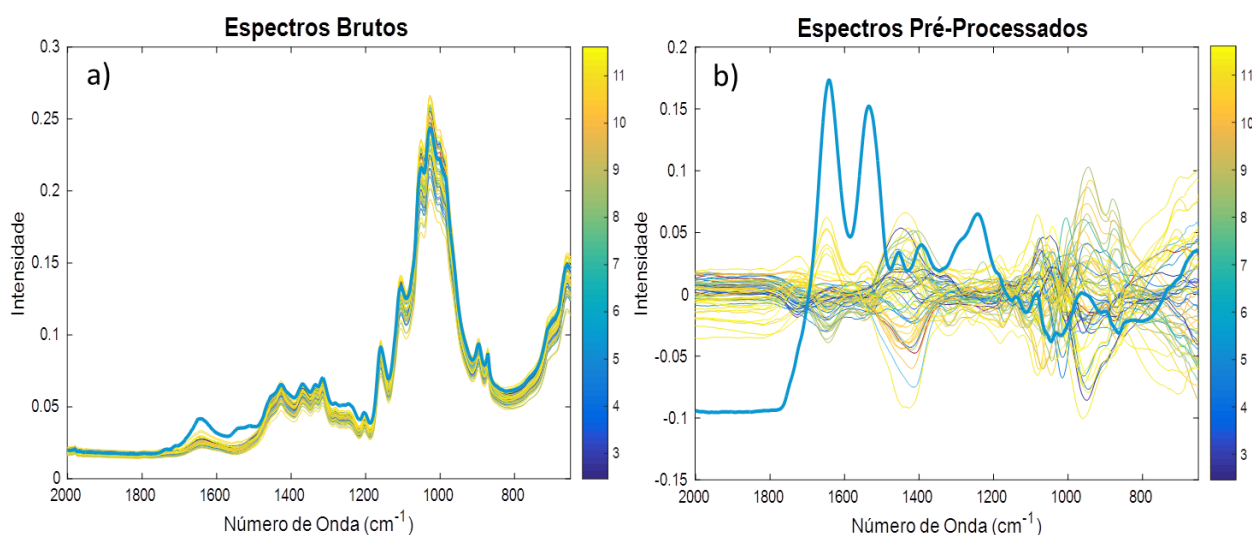


Fonte: A autora (2020).

Analisando o gráfico da variância explicada acumulada (Figura 12a) para o modelo PCA dos papéis couchês C1, as três primeiras componentes principais explicam 97,93 % da variabilidade dos dados. Nesse contexto, a PC que contém a informação da mudança do papel ao longo do tempo é a PC1. Isso é observado na Figura 12b, que mostra os escores da PC1 versus o logaritmo do tempo (em minutos). Analisando esse gráfico é possível observar que, na região negativa, têm-se as amostras com menos tempo de exposição à condição de envelhecimento e as amostras com escores positivos são as mais envelhecidas. Essas diferenças podem ser observadas no gráfico dos pesos da PC1, Figura 12c, na região negativa

(relacionada as amostras com menos tempo exposta a condição de envelhecimento) tem-se se uma banda em torno de 1400 cm^{-1} , que é do carbonato de cálcio. Na região positiva dos pesos, a banda em torno de 1430 cm^{-1} pode ser atribuída à flexão de CH_2 presentes na celulose. Também é possível encontrar bandas em torno de 874 cm^{-1} e 712 cm^{-1} são referentes ao carbonato de cálcio que também estão presentes nas amostras envelhecidas (CAUSIN et al, 2010; ALI et al, 2001). Nesse contexto, é provável que esteja ocorrendo à degradação da camada de revestimento ao longo do tempo, pois para as amostras mais envelhecidas a informação da celulose começa a ser relevante. Analogamente à C1, as amostras do papel C2 também evidenciam o mesmo comportamento (Figura 33 e Figura 34, localizado no Apêndice A). Na Figura 13 têm-se os espectros brutos e pré-processados com suavização pelo filtro de Savitzky-Golay (com polinômio de 2 ordem e janela de 33 pontos), SNV e centragem na média para os papéis reciclados R1.

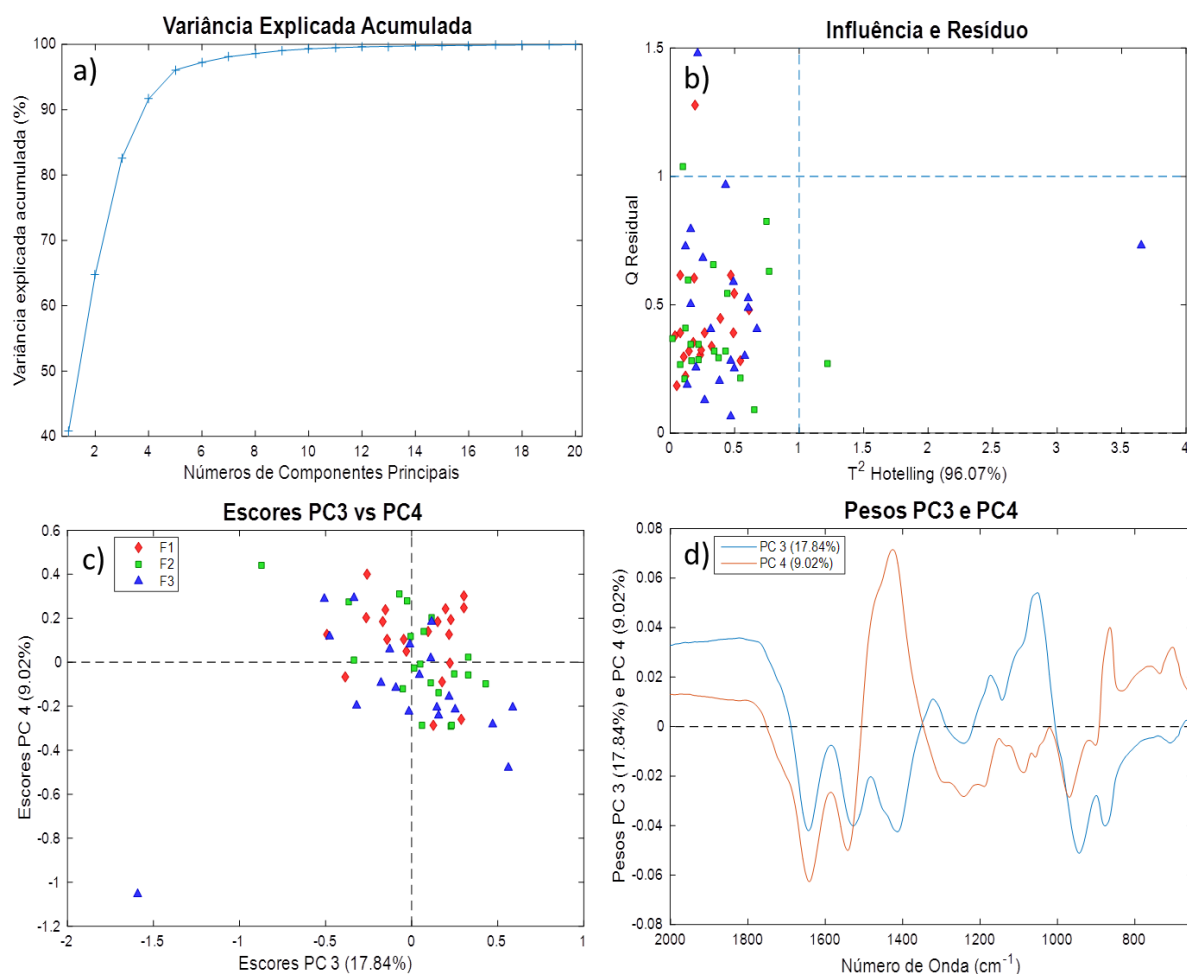
Figura 13 - Gráfico do efeito do pré-processamento para os papéis reciclados R1 (a) espectros brutos e (b) espectros pré-processados com suavização pelo filtro de Savitzky-Golay (com polinômio de 2 ordem e janela de 33 pontos), SNV e centragem na média.



Fonte: A autora (2020).

Na Figura 13a é possível observar um espectro diferente dos demais, essas diferenças são acentuadas quando o pré-processamento é realizado (Figura 13b). O modelo PCA para os papéis reciclados R1 estão representados na Figura 14.

Figura 14 - Gráfico da PCA para os papéis reciclados R1 (a) Variância explicada acumulada para o modelo, (b) gráfico de influência e resíduo das amostras, (c) escores da PC3 versus a PC4 (c) pesos da PC3 e da PC4.

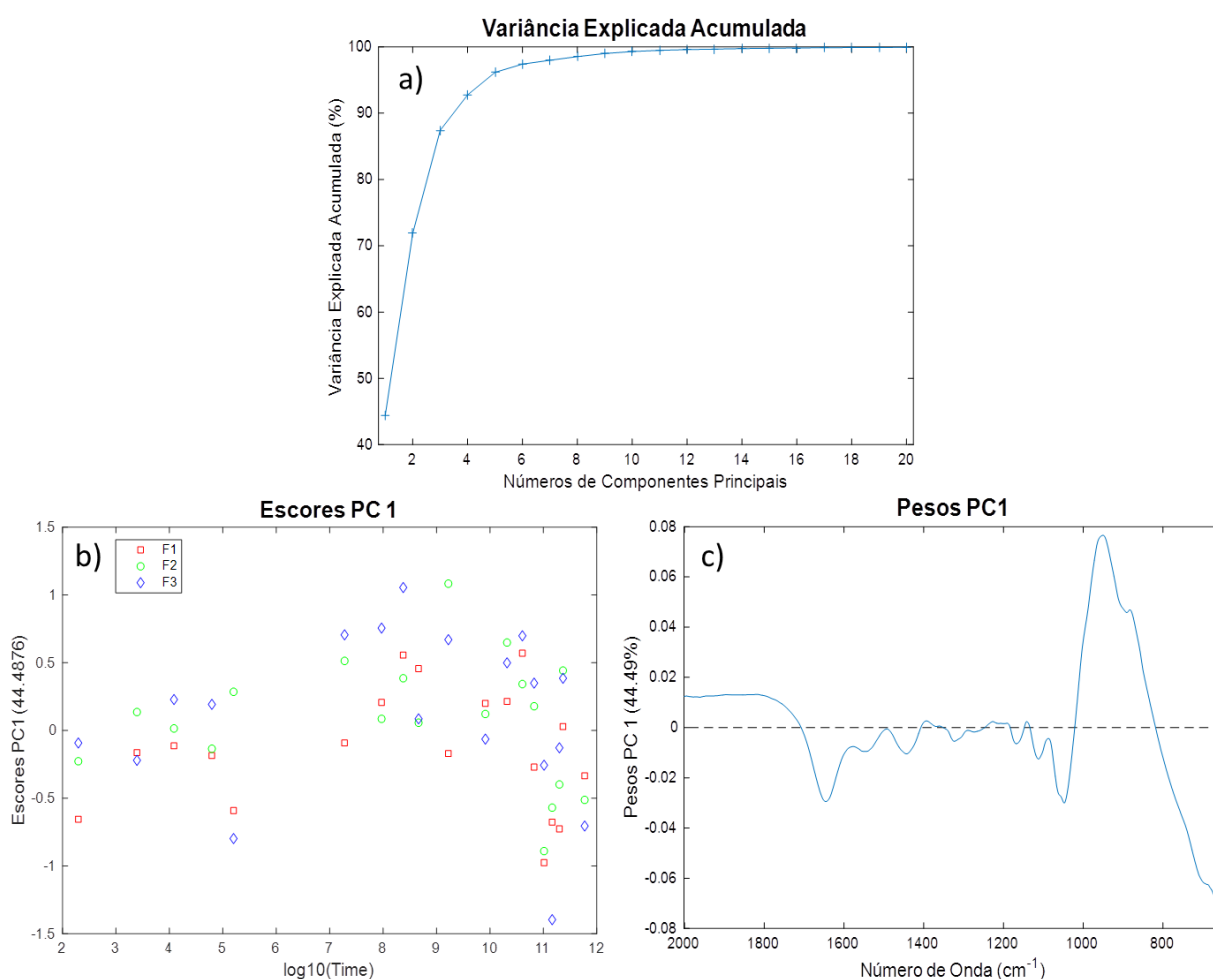


Fonte: A autora (2020).

Analisando o gráfico da variância explicada acumulada (Figura 14a) é possível observar que com 5 PCs o modelo explica 96,07 % dos dados, sugerindo que 5 PCs é suficiente para explicar a maior variabilidade dos dados. A Figura 14b mostra o gráfico de influência e de resíduo das amostras e é possível observar que uma das

amostras possui um T^2 Hotelling alto, indicando ser uma possível amostra anômala. Investigando essa amostra, foi possível observar que a mesma possuía o espectro diferente visto na Figura 13a e Figura 13b e, além disso, a PC3 e a PC4 são utilizadas para explicar a diferença dessa amostra para as demais. Visto que suas replicatas não possuem valores de resíduos semelhantes, confirma-se, portanto, a suspeita de se tratar de uma amostra anômala, justificando a sua remoção. Esse comportamento anômalo pode estar relacionado ao momento da aquisição do espectro devido a algum erro do equipamento ou do operador. Em seguida foi construído um novo modelo PCA sem essa amostra, representado na Figura 15.

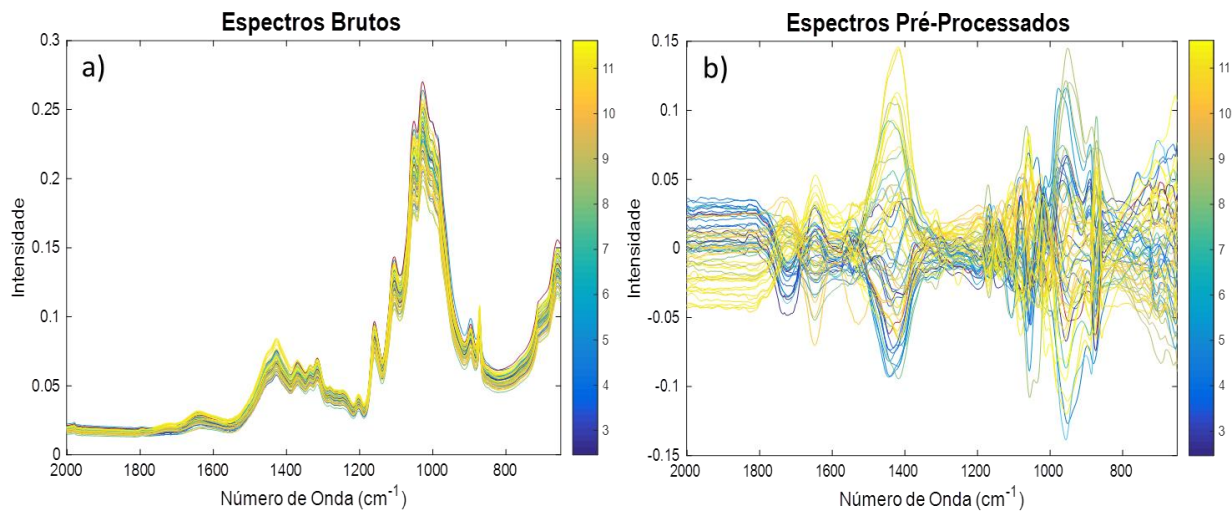
Figura 15 - Gráfico para o novo modelo PCA, sem as amostras anômalas, para os papéis reciclados R1 (a) Variância explicada acumulativa para o modelo (b) escores da PC1 versus o log do tempo (F1: Folha 1, F2: Folha 2 e F3: Folha 3) (c) pesos da PC1.



Fonte: A autora (2020).

A Figura 15a sugere que 5 PCs também são suficientes para explicar a maior variabilidade dos dados, com 96,15 % de variância explicada. Analisando a Figura 15b e as outras 4 componentes principais, não foi possível determinar qual PC que está relacionada a informação das alterações ocorridas no papel ao longo do tempo. Nesse contexto, acredita-se que a variabilidade relacionada à composição do papel é maior que a variabilidade ocasionada pelo envelhecimento do papel ao longo do tempo. Como os papéis reciclados já passaram por um processo de reciclagem, na qual ocorre uma mistura de outros tipos de papéis, é possível que uma folha possua uma complexa composição que é refletida nos perfis espectrais. Outra causa provável de não ser possível associar uma PC ao envelhecimento é que a condição de envelhecimento estudada não seja suficiente para causar mudanças observáveis com o tempo. Os papéis reciclados R2 também apresentaram o mesmo comportamento e um modelo com 5 PCs com 96,58 % de variabilidade explicada pode ser observado, nas Figura 35 e Figura 36 (Apêndice B). Para os papéis reciclados R2 e R3, também não foi possível determinar a PC que contém a informação relacionada a mudança do papel ao longo do tempo (Figura 37 e Figura 39, localizados no Apêndice B). Para investigar o processo de degradação dos papéis sulfites também foi realizado modelo PCA individualmente para cada marca. A Figura 16 mostra os espectros brutos e pré-processados para os papéis sulfites S1.

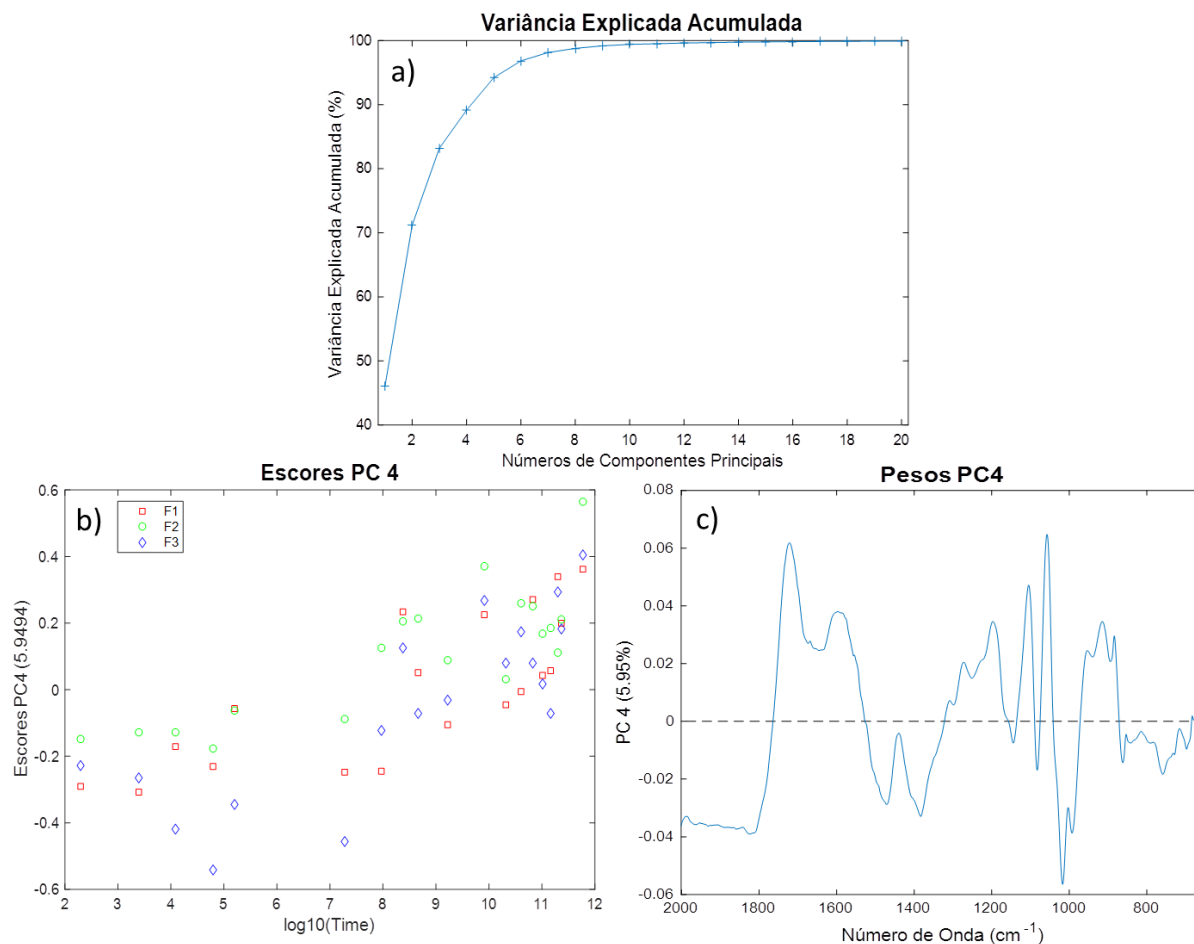
Figura 16 - Gráfico do efeito do pré-processamento para os papéis sulfites S1 (a) espectros brutos e (b) espectros pré-processados com suavização pelo filtro de Savitzky-Golay (com polinômio de 2 ordem e janela de 33 pontos), SNV e centragem na média.



Fonte: A autora (2020).

A Figura 16b mostra os espectros pré-processados para os papéis sulfites S1. O pré-processamento utilizado foi suavização pelo método de Savitzky-Golay (com polinômio de 2 ordem e janela de 33 pontos), SNV e centragem na média. Em seguida, foi construído o modelo PCA representado pela Figura 17.

Figura 17 - Gráfico da PCA para os papéis sulfites S1 (a) Variância explicada acumulativa para o modelo (b) escores da PC1 versus o log do tempo (F1: Folha 1, F2: Folha 2 e F3: Folha 3) (c) pesos da PC1.



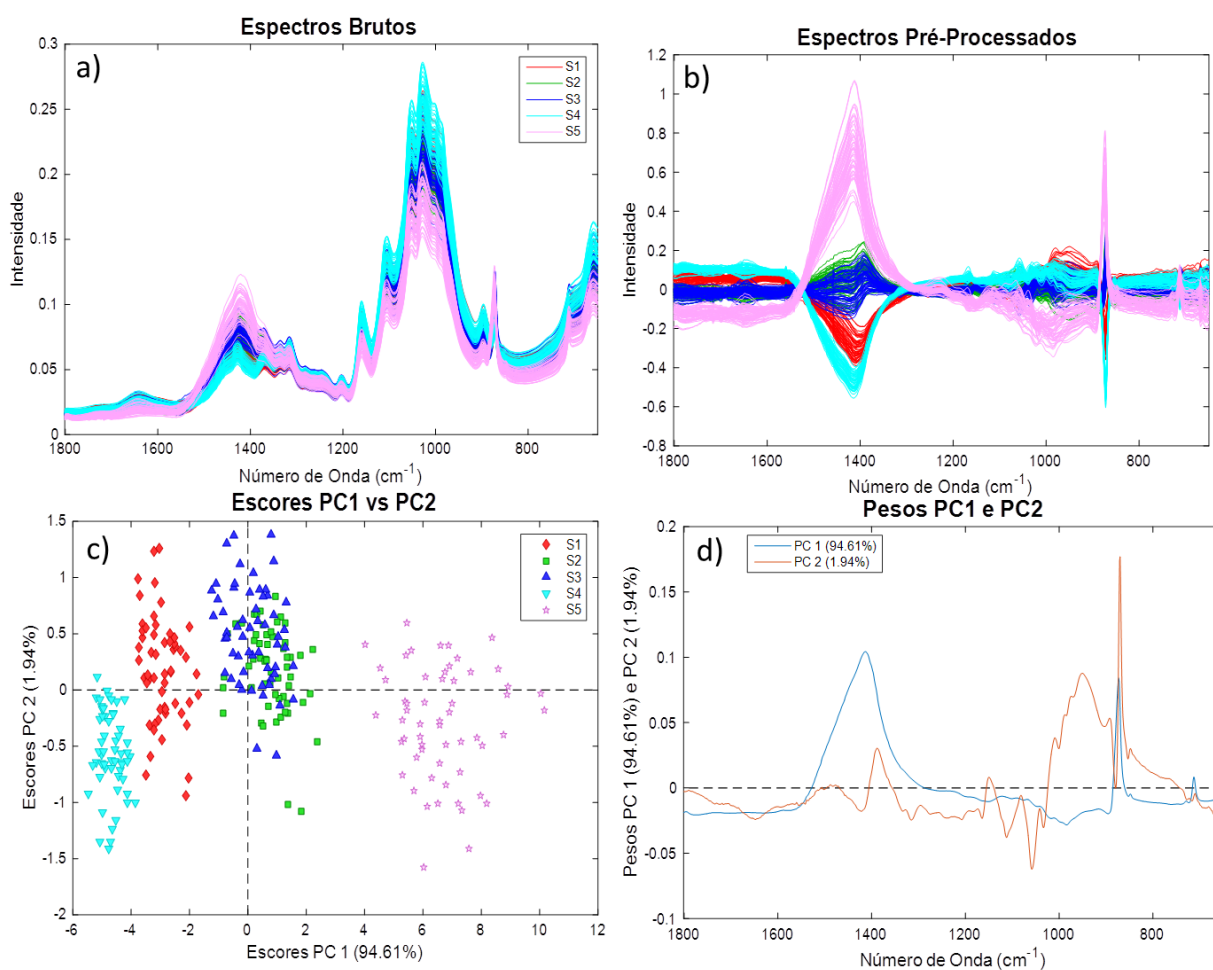
Fonte: A autora (2020).

Analisando a PCA para os papéis sulfites é possível observar que 6 PCs explicam 96,84 % da variabilidade dos dados (Figura 17a). Nesse contexto, foi possível determinar a PC que contém a informação da mudança do papel ao longo do tempo, a PC4 com 5,06 % de variância explicada. Analisando o gráfico dos escores, Figura 17b, observou-se que as amostras com menos tempo possuem escores negativos e as amostras mais envelhecidas estão localizadas na região positiva da PC4. No gráfico dos pesos (Figura 17c), é possível observar a região espectral que possui contribuição positiva. Com maior intensidade, tem-se a absorção em 1722 cm⁻¹, que pode ser atribuída ao estiramento de C=O, presente em ácidos carboxílicos. Esse resultado indica que pode estar ocorrendo a oxidação

da hidroxila da celulose durante o processo de degradação do papel. Esse comportamento foi observado em todas as marcas de papéis sulfites, esses resultados estão presentes da Figura 40 à Figura 46, localizadas no Apêndice C (CAUSIN *et al*, 2010).

Antes de construir os modelos PLS, foi avaliado o modelo PCA para todas 5 marcas de papéis sulfites. Para esse estudo foi selecionada uma nova região de trabalho, compreendida entre 1800 cm^{-1} a 650 cm^{-1} , que contém a maior variação no perfil espectral (Figura 18a).

Figura 18 - Gráficos das análises dos papéis sulfites, (a) espectros brutos, (b) espectros pré-processados com SNV e centragem na média (c) escores da PC1 versus PC2 (d) pesos da PC1 e da PC2.



Fonte: A autora (2020).

Na Figura 18b tem-se os espectros pré-processado com SNV e centragem na média. É possível observar que as bandas em torno de 1410 cm^{-1} , 871 cm^{-1} difere os papéis S4 dos S5. Em seguida um modelo PCA foi construído com 4 PCs que explica 98,73 % da variabilidades dos dados. Analisando o gráfico dos escores (Figura 18c), é possível observar que a maior variabilidade dos dados, PC1, é devido à diferença entre os papéis S1 e S4, localizados na região negativa dos escores da PC1, e os papéis S5, com escores positivos. Analisando o gráfico dos pesos para a PC1 (Figura 18d), a região positiva está relacionada às bandas que estavam em evidência no pré-processamento que são características do carbonato de cálcio. Isso indica que a diferença dos papéis pode estar associada com a quantidade de revestimento que, no papel S5. Os papéis S1 e S4 possuem uma tendência de separação mesmo ambos localizados na região negativa da PC1. Não foi possível observar diferenças entre os papéis S2 e S3, isso pode ter ocorrido devido ao fato de serem da mesma marca, mudando apenas a finalidade de uso.

Esses resultados foram obtidos durante o trabalho de iniciação científica da autora, onde foi estudado o processo de degradação dos diferentes tipos de papéis submetidos ao tratamento térmico. Concluindo que a utilização da espectroscopia na região do infravermelho médio associadas a técnicas quimiométricas se mostrou promissora para o estudo da degradação do papel. Analisando todos os tipos de papéis foi possível observar que os papéis couchês são bastante diferentes dos papéis reciclados e sulfites, diferenciando devido à presença de uma intensa camada de revestimento para os papéis couchês. Para as amostras submetidas à condição de envelhecimento acelerado a alta temperatura, pode estar ocorrendo à degradação da camada de revestimento para os papéis couchês. Para os papéis sulfites pode estar ocorrendo a oxidação da hidroxila da celulose, pois foi possível observar, na PCA, absorção em 1722 cm^{-1} para as amostras mais envelhecidas, que pode ser atribuída ao estiramento de C=O presente em ácidos carboxílicos, produto da oxidação. A partir desses resultados surgiu o interesse de utilizar a calibração multivariada para prever o tempo de exposição à condição de envelhecimento.

5.1.3 Regressão por Mínimos Quadrados Parciais

Para realizar a construção do modelo PLS utilizou-se apenas os papéis sulfites, uma vez que na análise exploratória, os papéis reciclados não mostram variações com o tempo e as variações nos papéis couchês estão muito relacionadas ao revestimento. As amostras de papéis sulfites foram separadas em dois conjuntos, um conjunto de calibração e um de previsão. Para isso, foi selecionada uma marca para compor o conjunto de previsão e as outras quatro restantes pra calibração. Para realizar essa escolha a Figura 18c foi analisada, é possível observar as amostras dos papéis S3 estão localizadas no centro da PC1, ou seja, sua variabilidade está explicada pelo modelo. Assim, os papéis sulfites S3 foram selecionados para o conjunto de previsão. Os modelos PLS foram realizados com 4 diferentes pré-processamentos e seleção de variáveis (i) SNV + MC (Figura 19), (ii) SNV + MC + GA (Figura 20), (iii) SNV + GLSW ($\alpha = 0,202$) + MC (Figura 22) e (iv) SNV + GLSW ($\alpha = 0,032$) + MC (Figura 23). Um resumo das principais figuras de méritos pode ser observado na Tabela 3.

Tabela 3 - Resumo das figuras de méritos dos modelos PLS com os diferentes pré-processamentos

Pré-processamento	LV	RMSEC	RMSECV	RMSEP	Bias _{cal}	Bias _{cv}	Bias _{pred}	R ² _{cal}	R ² _{cv}	R ² _{pred}
SNV + MC	6	1,02	1,08	2,20	0,000	0,005	-2,006	0,88	0,87	0,91
SNV + GA + MC	4	1,61	1,70	1,47	0,000	-0,005	-0,743	0,70	0,67	0,82
SNV + GLSW (0,202) + MC	6	0,67	0,79	1,26	0,000	0,000	-1,104	0,95	0,93	0,96
SNV + GLSW (0,032) + MC	6	0,41	0,59	0,55	0,000	0,010	-0,140	0,98	0,99	0,97

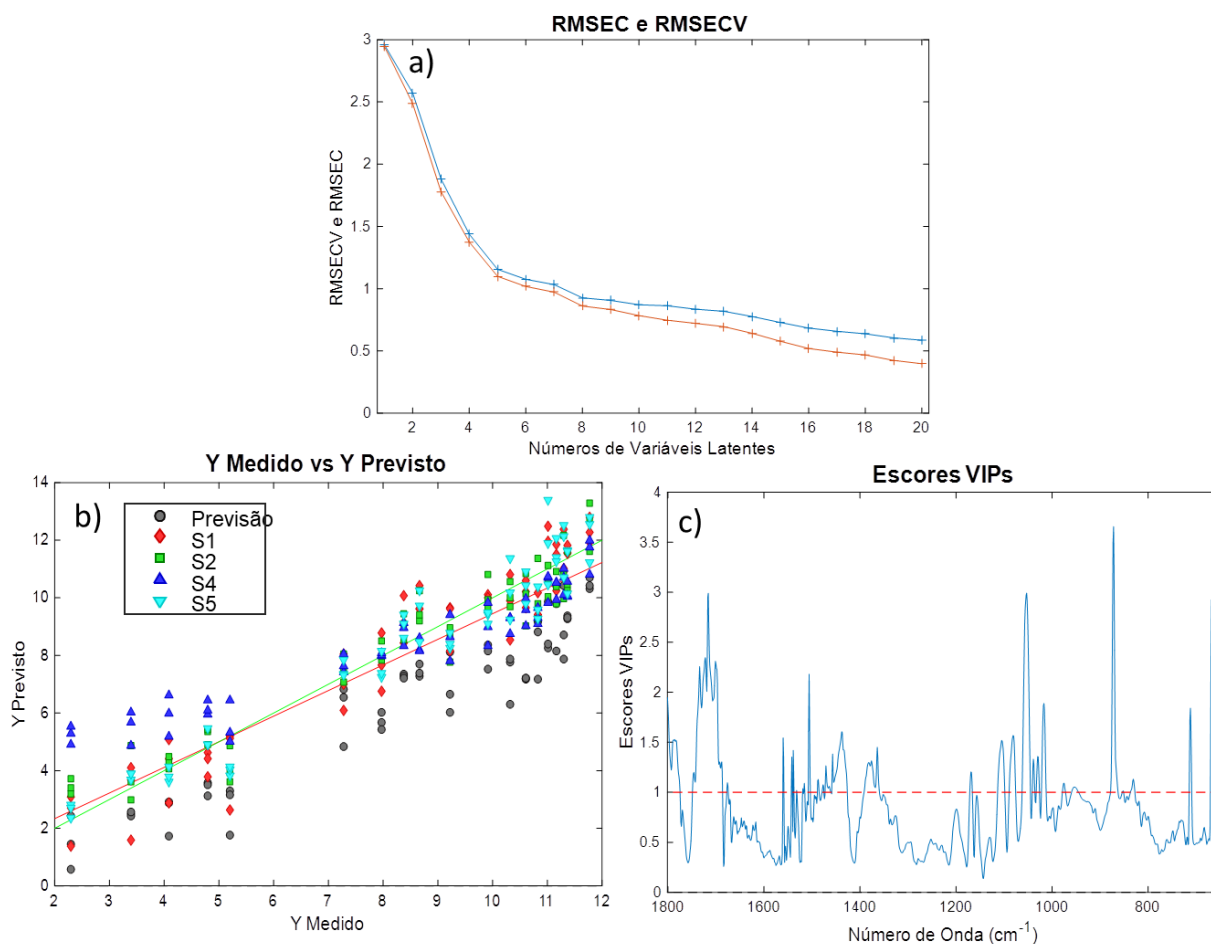
*SNV, MCS, GLSW e GA são os pré-processamento usados; entre parêntesis o valor de α (para GLSW); LV = número de variáveis latentes; RMSEC, RMSECV e RMSEP; Bias_{cal}, Bias_{cv} e Bias_{pred} = Erro de Bias para a calibração, a validação cruzada e a previsão, respectivamente; R²_{cal} R²_{cv} e R²_{pred}, coeficiente de determinação para a calibração, validação cruzada e previsão.

Fonte: A autora (2020).

O modelo PLS utilizando o pré-processamento SNV e centragem na média foi construído utilizando 6 variáveis latentes, representado pela Figura 19a. Analisando a Figura 19b e os dados da Tabela 3 o erro de maior influencia é o Bias de previsão, para tentar minimizar esse erro foi utilizado o pré-processamento SNV, centragem na média e seleção de variáveis (GA), representado na Figura 20. A partir da Figura 20a é possível observar que as variáveis que apareceram com mais frequência

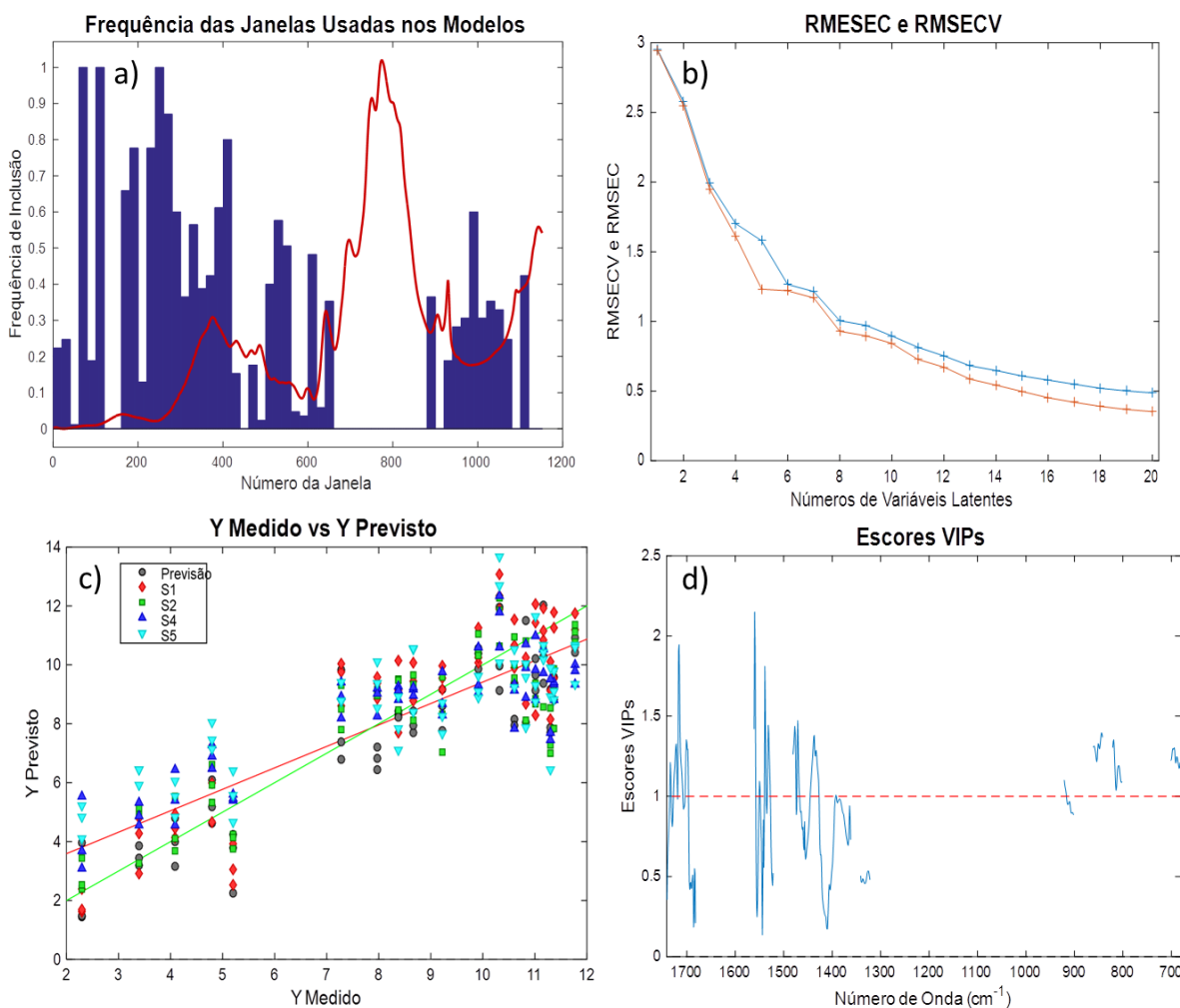
durante a utilização do GA foram a região de 1800 cm^{-1} a 1200 cm^{-1} , esse comportamento já era esperado, pois como analisado na Figura 17, essa região contém a informação da banda característica da degradação. Analisando esses dados é possível observar que o uso da seleção de variáveis diminui a quantidade de variáveis latentes necessárias para explicar o modelo e também houve uma diminuição do Bias de previsão. Por outro lado, o modelo de calibração apresentou valores aparentemente mais baixos de R^2_{CV} que diminuiu de 0,87 para 0,67, prejudicando assim o modelo.

Figura 19 - Gráficos do modelo PLS para os dados pré-processados com SNV e centragem na média para os papéis envelhecidos artificialmente, (a) gráfico de RMSEC e RMSECV (b) gráficos de regressão (c) gráfico dos escores VIPs.



Fonte: A autora (2020).

Figura 20 - Gráficos do modelo PLS para os dados pré-processados com SNV e centragem na média e seleção de variáveis (GA) para os papéis envelhecidos artificialmente, (a) Variáveis que aparecem com maior frequência quando o GA é aplicado, (b) gráfico de RMSEC e RMSECV, (c) gráficos de regressão e (d) gráfico dos escores VIPs.

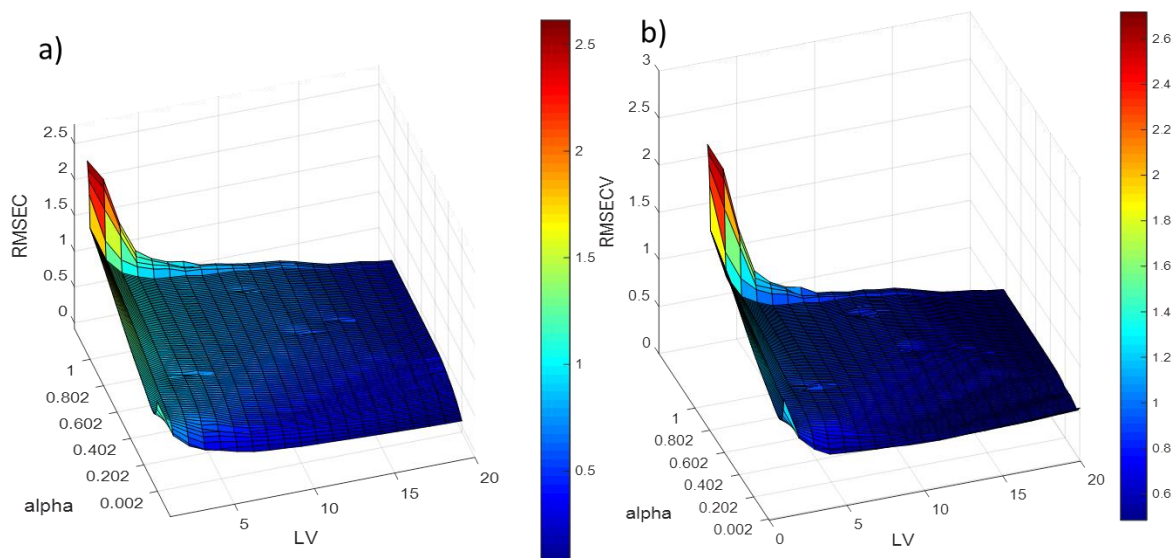


Fonte: A autora (2020).

O modelo PLS também foi testado com outros pré-processamentos SNV, GLSW e a centragem na média. Para utilizar o GLSW é necessário a determinação do parâmetro de regularização responsável pela ponderação do filtro (α), quanto menor o valor de α , mais interferentes o GLSW filtra e, ao aumentar o valor do α , a ação do filtro é diminuída. Para determinar o valor adequado para α , foi construída uma superfície mostrando a variação do RMEC e RMSECV em função das variáveis latentes e do valor de α , como mostra a Figura 21. Os gráficos de contorno das superfícies estão representados na Figura 48 (Apêndice D). A partir desses dados

foram selecionados dois valores de α : 0,202 e 0,032. Esses valores foram escolhidos por possuírem RMSECV relativamente baixo e poucas LV para o modelo, esses resultados estão representados na Figura 22 e Figura 23.

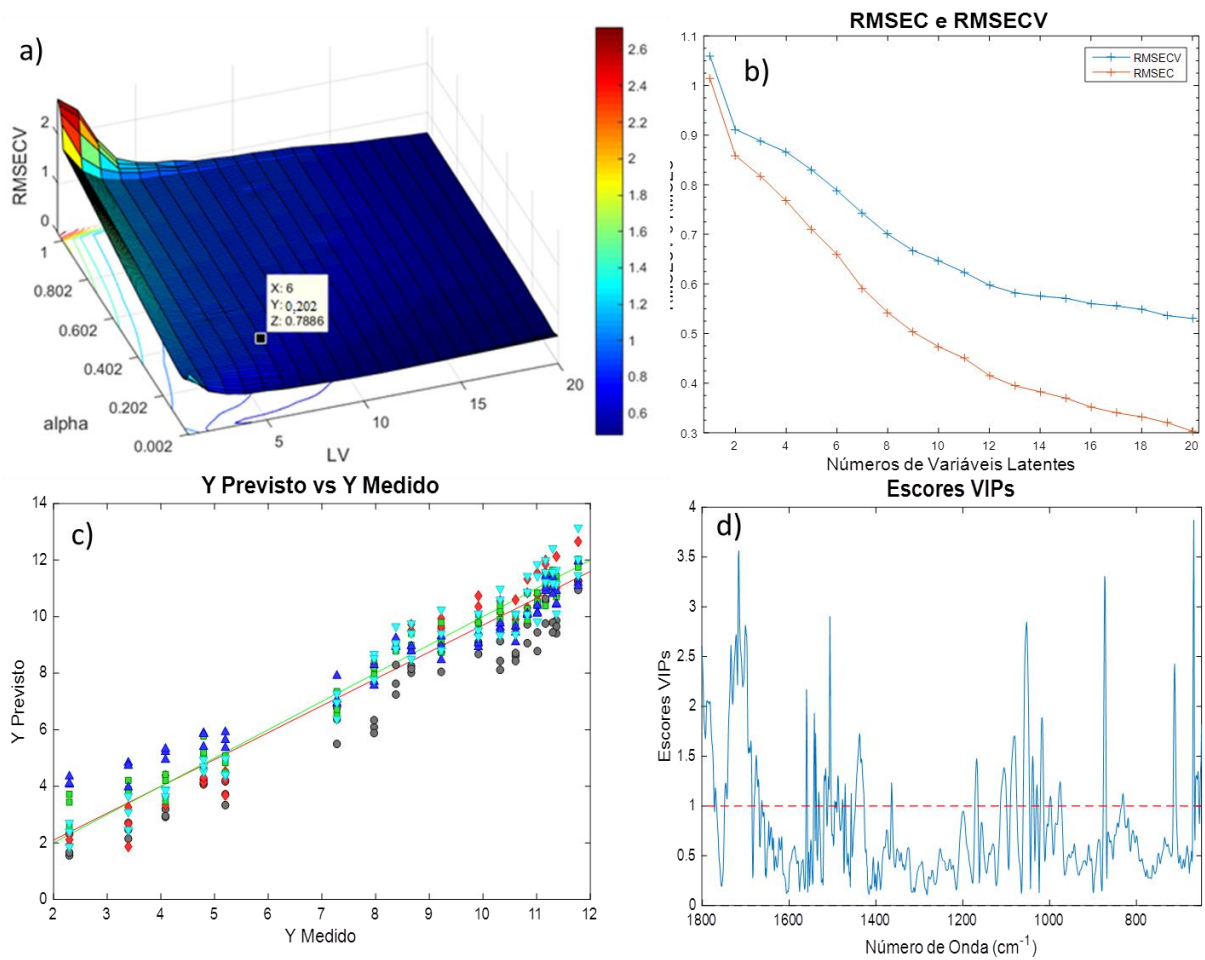
Figura 21 - Gráfico das superfícies para o (a) RMCEC e (b) RMSECV.



Fonte: A autora (2020).

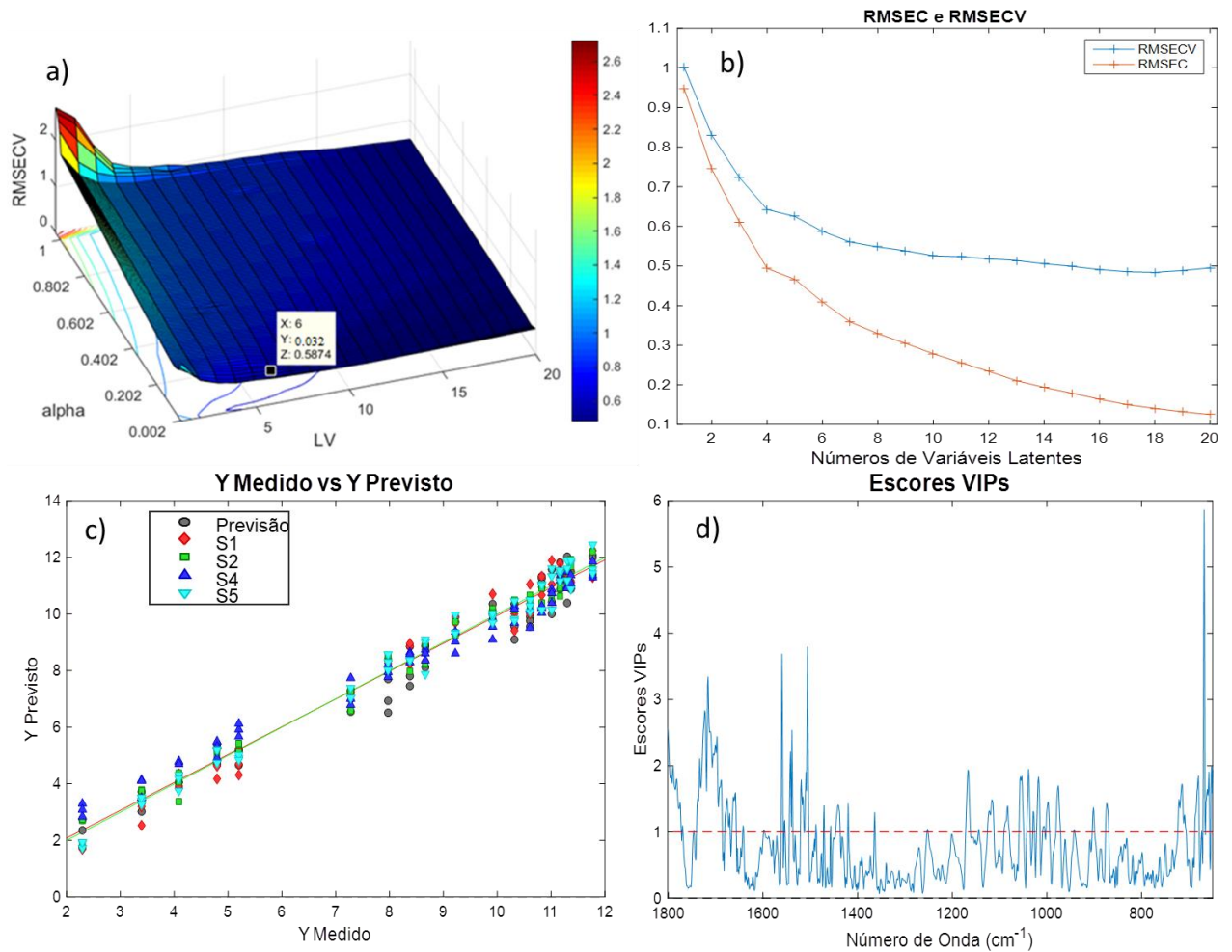
Analisando os modelos utilizando o filtro GLSW, Figura 22 e Figura 23, é possível observar uma melhora significativa na medida em que o valor de α diminui. Essa melhora pode ser observada principalmente para o bias de previsão, onde com $\alpha = 0,202$ é -1,10 e para $\alpha = 0,032$ é -0,14. Essa melhora também é observada para os outros parâmetros do modelo de regressão. Em suma, a variabilidade presentes nas diferentes marcas influenciam na construção dos modelos PLS analisados ao longo do tempo, para reduzir essa variabilidade foi utilizado o filtro GLSW e o melhor resultado foi para o menor valor de α que é 0,032. Analisando o Escores VIPs para todos os quatros modelos PLS construídos (Figura 19d, Figura 20d, Figura 22d e Figura 23d) é possível observar que todos os modelos reconheceram a banda em torno de 1722 cm^{-1} como relevantes para o modelo, como visto anteriormente pode ser atribuída ao estiramento de C=O presente em ácidos carboxílicos, produto do processo de degradação.

Figura 22 - Gráficos do modelo PLS para os dados pré-processados com SNV, GLSW ($\alpha = 0,202$) e centragem 0 na média para os papéis envelhecidos artificialmente (a) Gráfico da superfície para o RMSECV (b) gráfico de RMSEC e RMSECV (c) gráficos de regressão, (d) gráfico de regressão e (d) gráfico dos Escores VIPs.



Fonte: A autora (2020).

Figura 23 - Gráficos do modelo PLS para os dados pré-processados com SNV, GLSW ($\alpha = 0,032$) e centragem na média para os papéis envelhecidos artificialmente, (a) Gráfico da superfície para o RMSECV (b) gráfico de RMSEC e RMSECV, (c) gráficos de regressão e (d) gráfico dos Escores VIPs.



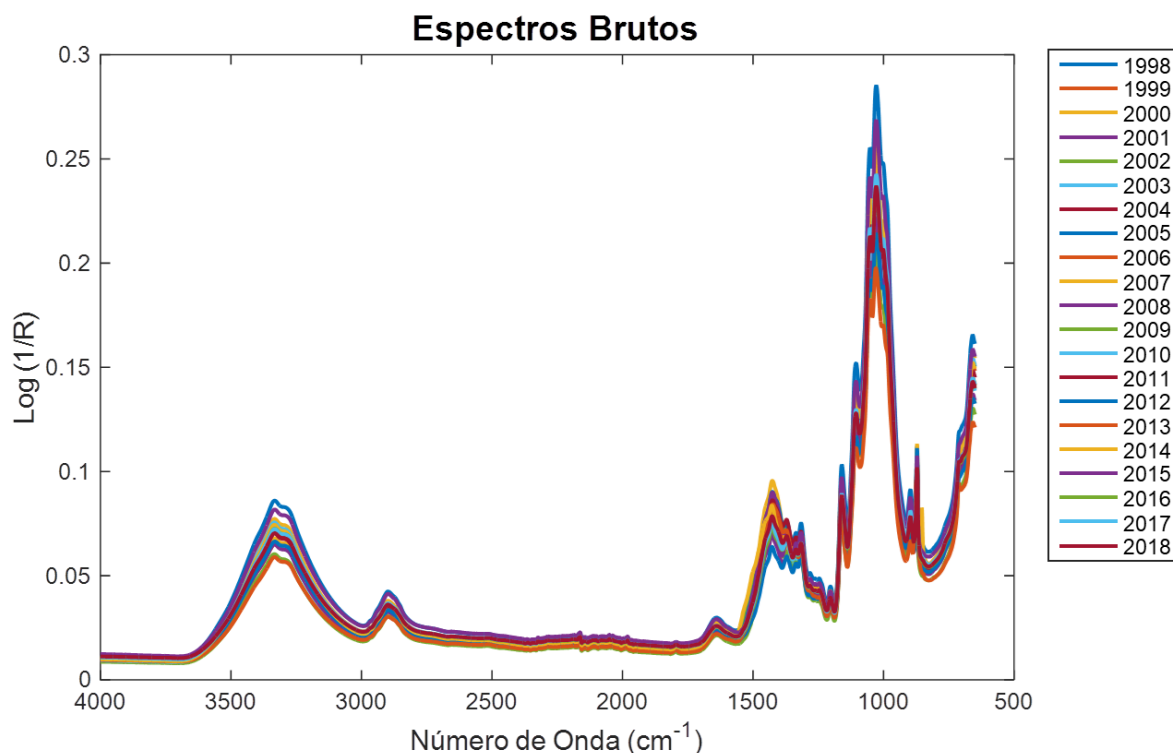
Fonte: A autora (2020).

5.2 ENVELHECIMENTO NATURAL

5.2.1 Atribuição de Bandas

A Figura 24 mostra a média dos espectros brutos por anos para os papéis envelhecidos naturalmente, de forma que as amostras estão rotuladas em datas do período de 1998 a 2018.

Figura 24 - Média dos espectros brutos dos papéis naturalmente envelhecidos por ano.



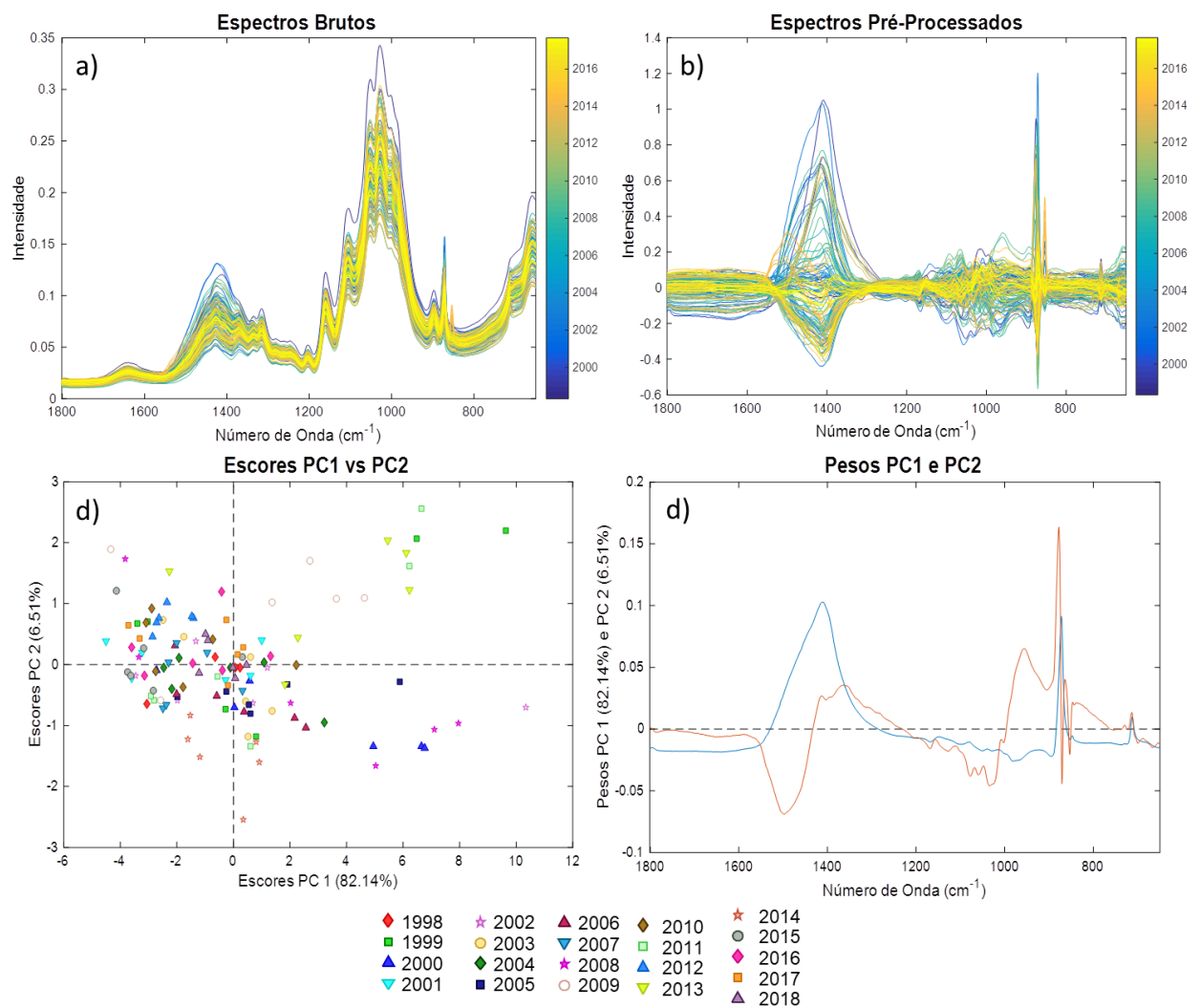
Fonte: A autora (2020).

As bandas presentes são as mesmas encontradas para os papéis sulfites descritos anteriormente. Nesse contexto a região entre 3600 cm^{-1} a 2600 cm^{-1} é atribuída às funções orgânicas da celulose, também é possível observar a presença de água adsorvida, enquanto que as bandas entre 1500 e 900 cm^{-1} constituem a região de impressão digital da celulose, e estão detalhadas na Tabela 2 (ZIEBA-PALUS et al, 2016; COUSIN et al, 2010). A região espectral utilizada para a construção dos modelos foi de 1800 cm^{-1} a 650 cm^{-1} , pois é onde contém a maior variabilidade espectral.

5.2.2 Análise dos Componentes Principais

A matriz de dados utilizada para a análise de componentes principais dos papéis envelhecidos naturalmente, possui 124 espectros e 1135 variáveis. Na Figura 25a têm-se os espectros brutos e na Figura 25b têm-se os espectros pré-processados com SNV e centragem na média.

Figura 25 - Gráficos da PCA para os papéis envelhecidos naturalmente, (a) espectros brutos, (b) espectros pré-processados (c) escores da PC1 versus PC2 (d) pesos da PC1 e da PC2.



Fonte: A autora (2020).

Analisando o gráfico dos escores é possível observar que as amostras estão bem distribuídas, não há formação de agrupamentos naturais. Durante a construção dos modelos PLS foi observado que duas amostras (espectro de uma folha de 1998

e 2002) estavam influenciando negativamente os modelos, mesmo possuindo espectros bastante parecidos com os demais e que não estivesse em evidência na PCA. Essas amostras ficaram evidentes no gráfico de y predito para a validação vs y mensurado, prejudicando assim o R^2_{CV} , então essas amostras foram removidas.

5.2.3 Regressão por Mínimos Quadrados Parciais

Para a construção dos modelos PLSs foi separado o conjunto de calibração e o de previsão. Para cada ano têm-se três documentos e de cada documento duas folhas foram utilizadas, nesse contexto, o conjunto de previsão foi formado por um documento de cada ano, ou seja, duas folhas, esses documentos foram escolhidas de forma aleatória para compor o conjunto de previsão. Os documentos restantes formaram o conjunto de calibração. Os modelos PLS foram realizados com 4 diferentes pré-processamentos, (i) SNV + MC (Figura 26), (ii) SNV + MC + GA (Figura 27), (iii) SNV + GLSW ($\alpha = 0,202$) + MC (Figura 29) e (iv) SNV + GLSW ($\alpha = 0,542$) + MC (Figura 30). Um resumo das principais figuras de méritos resultados obtidos podem ser observados na Tabela 4.

Tabela 4 - Resumo das figuras de méritos dos modelos PLS com os diferentes pré-processamentos para os papéis envelhecidos naturalmente.

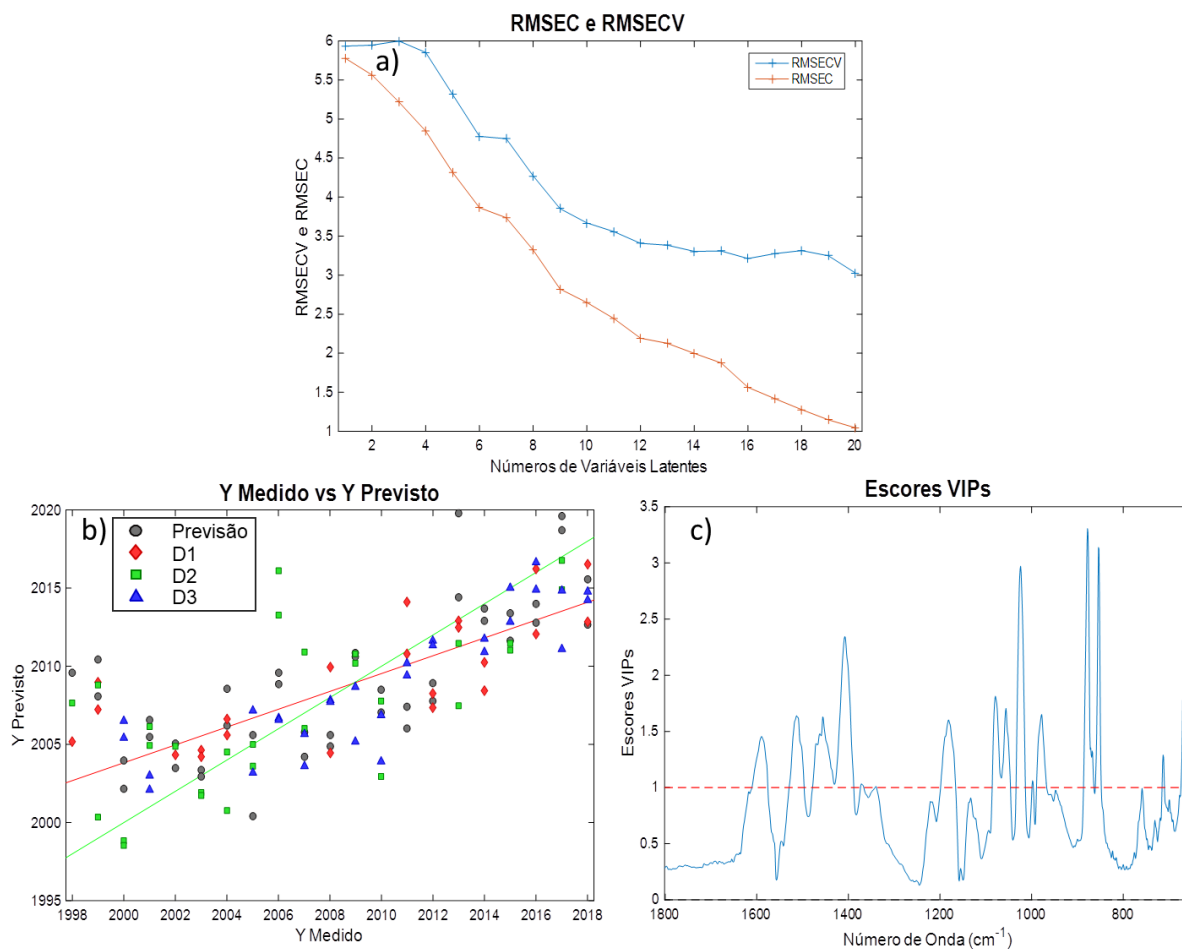
Pré-processamento	LV	RMSEC	RMSECV	RMSEP	Bias _{Cal}	Bias _{CV}	Bias _{Pred}	R ² _{Cal}	R ² _{CV}	R ² _{Pred}
SNV + MC	6	3,95	4,95	4,16	0,000	0,101	0,733	0,56	0,33	0,52
SNV + GA + MC	6	3,91	4,65	3,88	0,000	0,106	0,374	0,56	0,40	0,58
SNV + GLSW (0,112) + MC	6	1,53	3,13	3,00	0,000	0,214	-0,403	0,93	0,73	0,77
SNV + GLSW (0,542) + MC	6	2,45	3,51	3,16	0,000	0,115	-0,744	0,83	0,66	0,74

*SNV, MCS, GLSW e GA são os pré-processamento usados; entre parêntesis o valor de α (para GLSW); LV = número de variáveis latentes; RMSEC, RMSECV e RMSEP; Bias_{cal}, Bias_{CV} e Bias_{Pred} = Erro de Bias para a calibração, a validação cruzada e a previsão, respectivamente; R²_{Cal}, R²_{CV} e R²_{Pred}, coeficiente de determinação para a calibração, validação cruzada e previsão.

Fonte: A autora (2020).

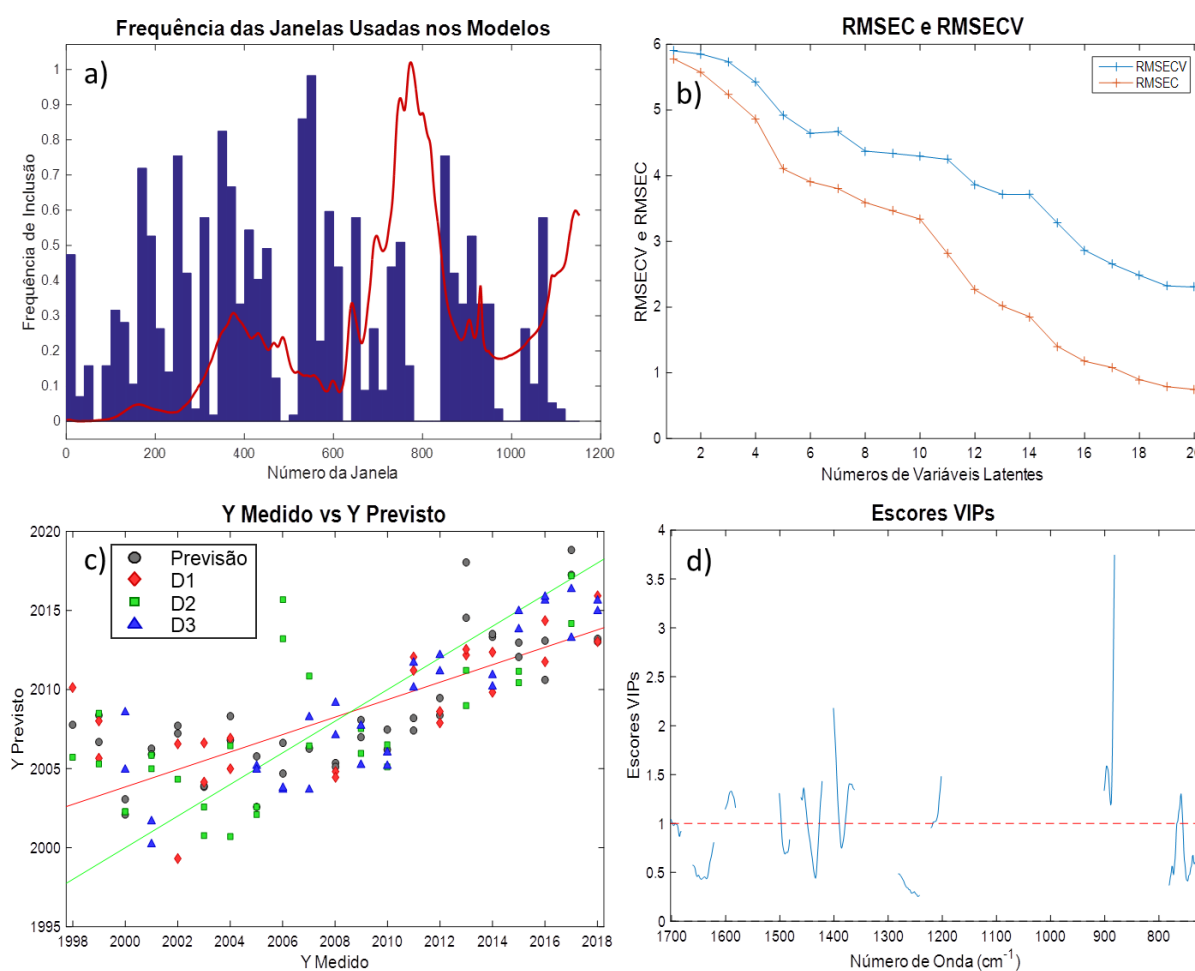
Todos os modelos construídos utilizaram 6VL para explicar a maior variabilidade dos dados. Analisando os dados da Tabela 4, pode-se observar que os modelos PLS que utilizaram o pré-processamento SNV e centragem na média e SNV, GA e centragem na média não foram eficientes para descrever a datação dos documentos. Esse comportamento pode ser observado nos valores de R^2_{CV} , que foram iguais a 0,33 e 0,40 (Tabela 4, Figura 26 e Figura 27). Nesse contexto, foi utilizado o filtro GLSW para remover a variância presente da matriz X que é ortogonal ao vetor do tempo, y , ou seja, a variabilidade dos documentos de um mesmo ano. Para a utilização desse pré-processamento foi necessária a determinação do α . Da mesma forma como foi feito para os papéis artificialmente envelhecidos, foi construída uma superfície para os valores de RMSEC e/ou RMSECV em função do número de variáveis latentes e do valor de α , como mostra a Figura 28. Os gráficos de contorno das superfícies estão representados na Figura 49 (Apêndice D).

Figura 26 - Gráficos do modelo PLS para os dados pré-processados com SNV e centragem na média para os papéis envelhecidos naturalmente (a) gráfico de RMSEC e RMSECV, (b) gráficos de regressão e (c) gráfico dos escores VIPs.



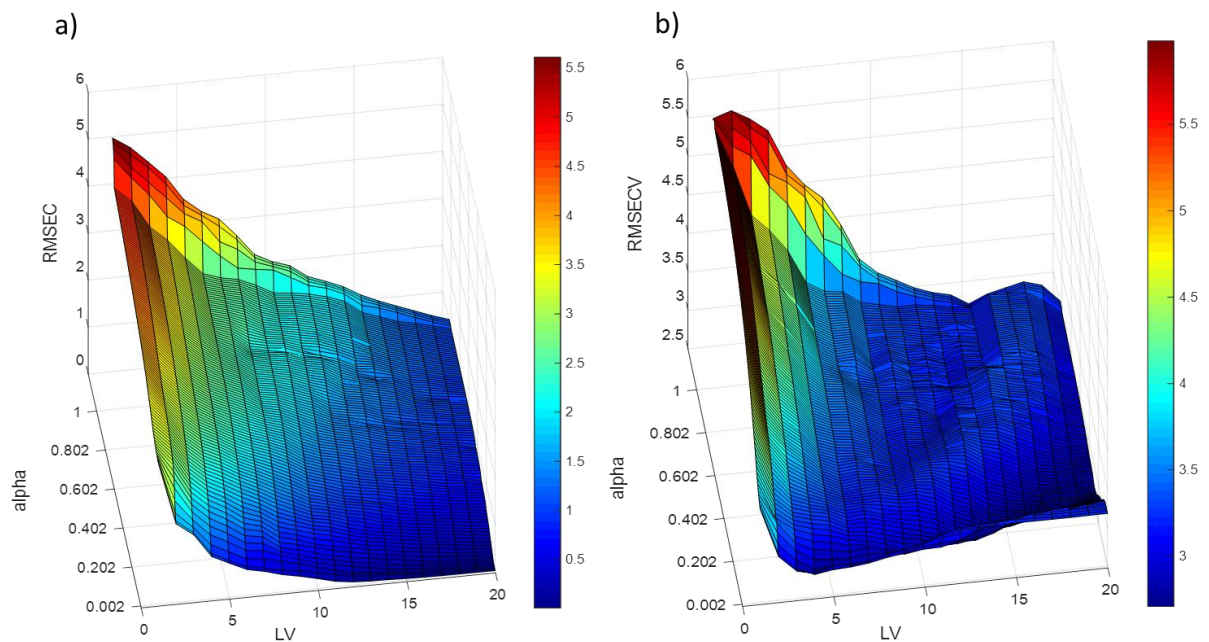
Fonte: A autora (2020).

Figura 27 - Gráficos do modelo PLS para os dados pré-processados com SNV e centragem na média e seleção de variáveis (GA) para os papéis envelhecidos naturalmente (a) Variáveis que aparecem com maior frequência quando o GA é aplicado, (b) gráfico de RMSEC e RMSECV, (c) gráficos de regressão e (d) gráfico dos escores VIPs.



Fonte: A autora (2020).

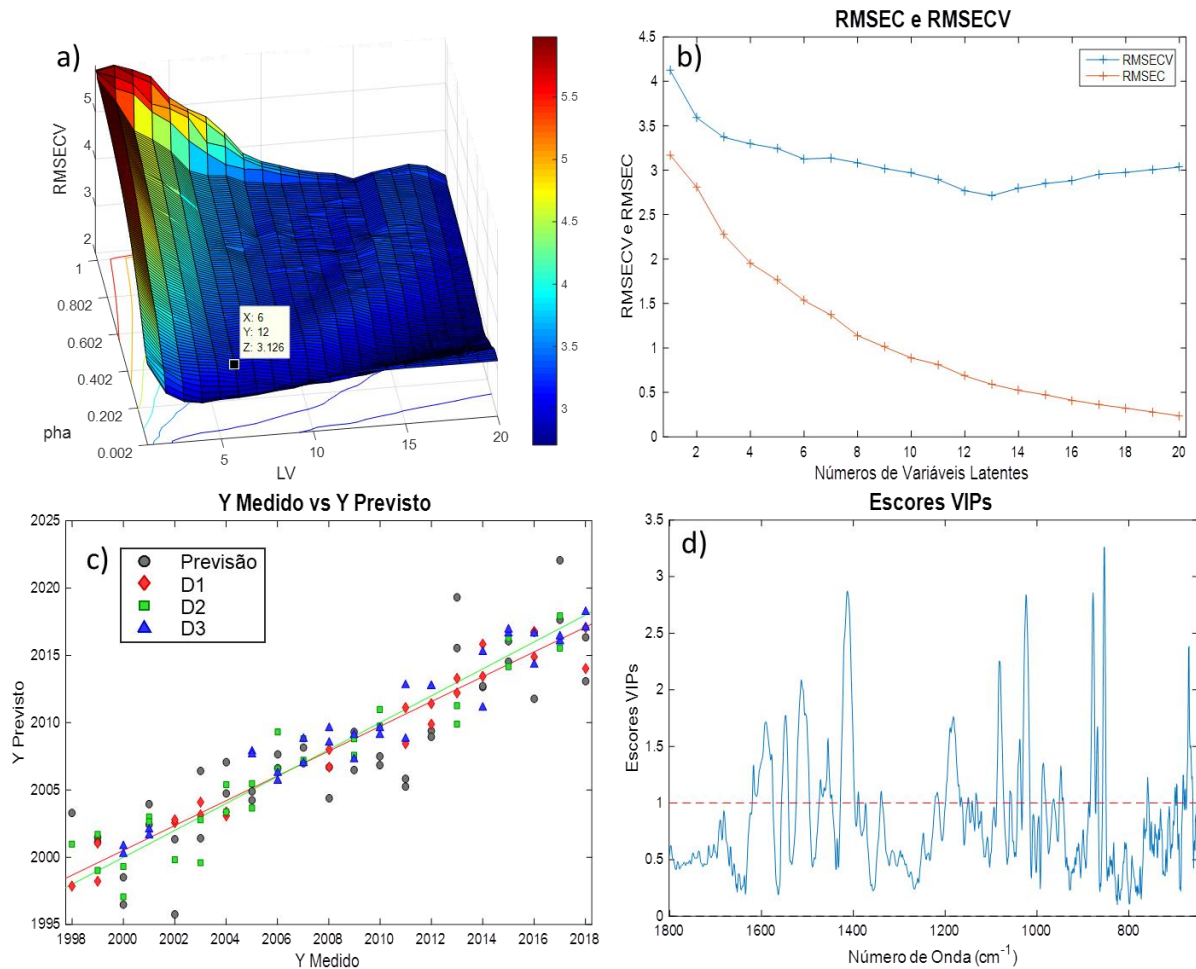
Figura 28 - Gráfico das superfícies de níveis para os papéis envelhecidos naturalmente para o (a) RMSEC e (b) RMSECV.



Fonte: A autora (2020).

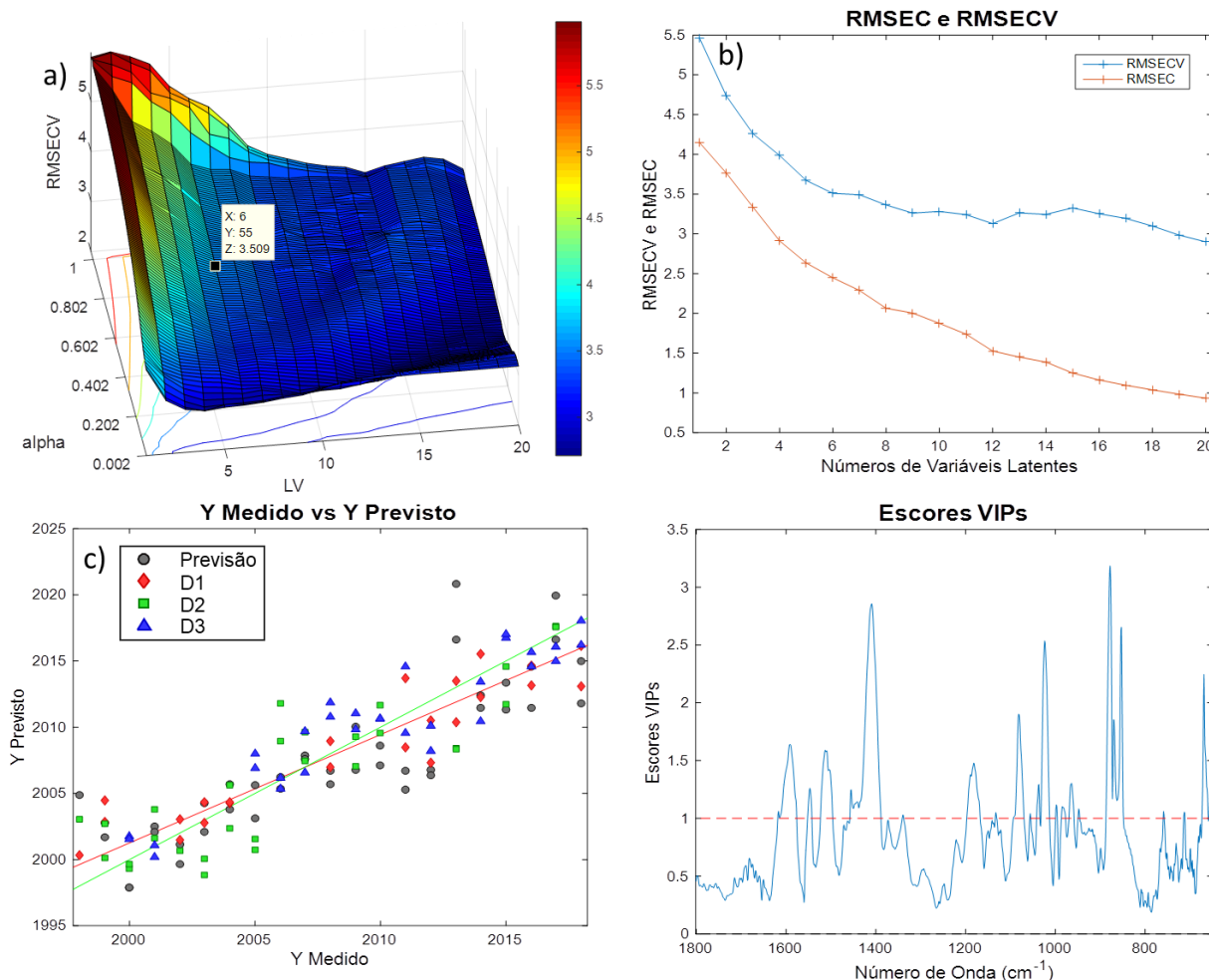
Analisando a Figura 28 foram selecionados dois alfas 0,112 e 0,542, pois possuíam RMSECV relativamente baixos e não precisavam de muitas LV para explicar o modelo (Figura 29 e Figura 30). As figuras de méritos desses modelos estão apresentadas na Tabela 3. Analisando a eficiência do filtro GLSW para reduzir a variabilidade presente no mesmo ano é possível observar uma melhora significativa na medida que o valor de α diminui. Essa melhora pode ser observada comparando o modelo usando $\alpha = 0,112$, cujo R^2_{cv} é igual a 0,73, e o modelo com $\alpha = 0,542$, cujo R^2_{cv} é 0,66. Essa melhora também é observada para os outros parâmetros do modelo de regressão (RMSECV, RMSEP, $Bias_{Pred}$ e R^2_{Pred}). O melhor resultado foi para o menor valor de α que é 0,112.

Figura 29 - Gráficos do modelo PLS para os dados pré-processados com SNV, GLSW ($\alpha = 0,112$) e centragem na média para as amostras envelhecidas naturalmente (a) Gráfico de superfície para o RMSECV, (b) gráfico de RMSEC e RMSECV, (c) gráficos de regressão e (d) gráfico dos Escores VIPs.



Fonte: A autora (2020).

Figura 30 - Gráficos do modelo PLS para os dados pré-processados com SNV, GLSW ($\alpha = 0,542$) e centragem na média para as amostras envelhecidas naturalmente (a) Gráfico da superfície para o RMSECV (b) gráfico de RMSEC e RMSECV (c) gráficos de regressão (d) gráfico dos Escores VIPs.



Fonte: A autora (2020).

5.3 ENVELHECIMENTO NATURAL VS ARTIFICIAL

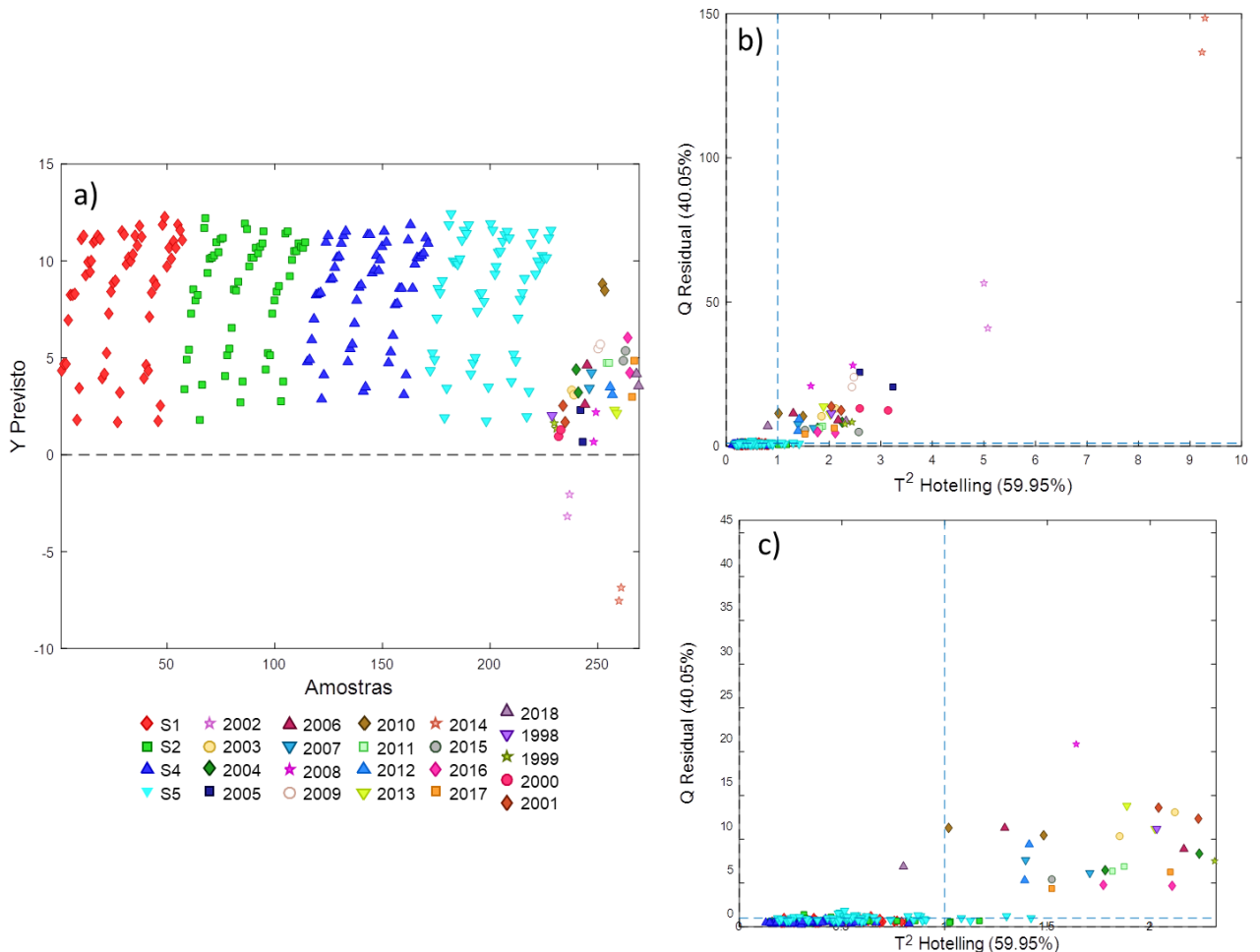
Atestar a legitimidade de documentos é bastante importante no contexto forense. Assim, podem ser utilizados os modelos de datação para identificar fraudes documentais. Deste modo, é importante testar se documentos artificialmente envelhecidos podem ser apresentados como autênticos (naturalmente envelhecidos). Esse estudo pode ser realizado por duas abordagens, a primeira projetando-se amostras de documentos naturalmente envelhecidos no modelo PLS

construído com amostras envelhecidas artificialmente. A segunda abordagem é a projeção de amostras artificialmente envelhecidas em um modelo PLS construído com amostras naturalmente envelhecidas. Essas duas abordagens podem ser usadas para verificar se há compatibilidade entre o tempo de exposição a uma determinada condição de envelhecimento artificial com o envelhecimento natural. Ou seja, se um documento envelhecido artificialmente pode se passar por um documento autêntico de um determinado período de tempo.

5.3.1 Projetando as amostras de envelhecimento natural no modelo PLS das amostras envelhecidas artificialmente

As amostras naturalmente envelhecidas foram projetadas no melhor modelo de calibração do envelhecimento acelerado, ou seja, as que utilizaram como pré-processamento SNV, GLSW ($\alpha = 0,032$) e centragem na média. Os resultados dessa projeção estão presentes na Figura 31, que mostra o modelo de calibração construído com as amostras envelhecidas artificialmente e o conjunto de previsão com as amostras envelhecidas naturalmente.

Figura 31 - Projeção das amostras envelhecidas naturalmente nos modelos PLS construído com as amostras envelhecidas artificialmente (a) Gráfico do valor previsto, (b) gráfico de influência e resíduo das amostras e (c) gráfico de influência e resíduo das amostras ampliado.



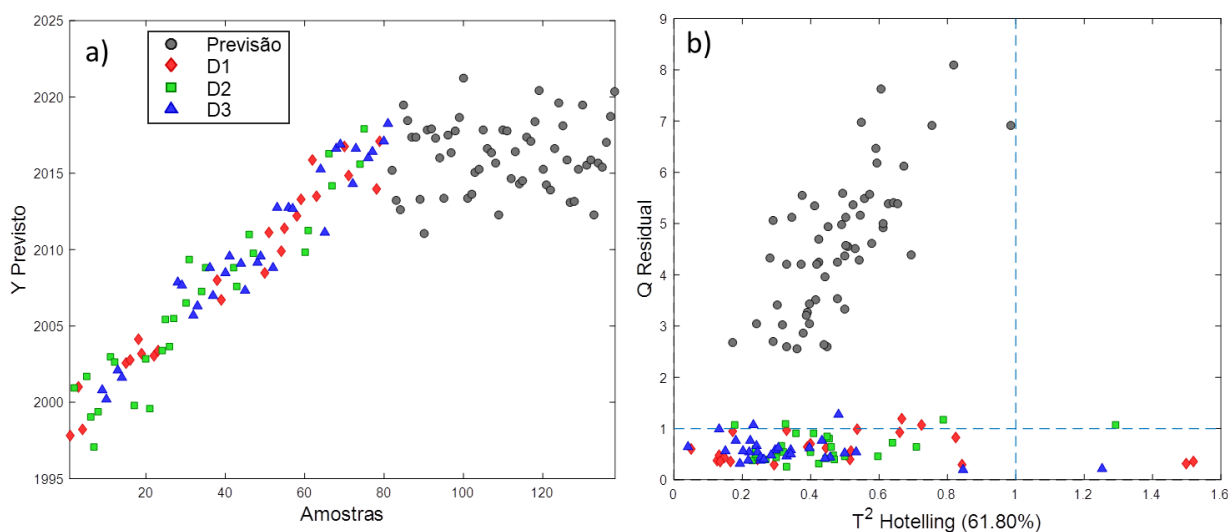
Fonte: A autora (2020).

A variabilidade encontrada nas folhas de papéis ao longo dos anos é muito grande de modo que o modelo de calibração para as amostras envelhecidas artificialmente em alta temperatura não serve para explicar essas amostras envelhecidas naturalmente. Esse comportamento é observado no gráfico de influência e resíduo das amostras, pois as amostras naturais apresentaram Qresidual e T^2 Hotelling bastante elevados.

5.3.2 Projetando as amostras de envelhecimento artificial no modelo PLS das amostras envelhecidas naturalmente.

Para esse estudo amostras envelhecidas artificialmente foram projetada no melhor modelo de calibração do envelhecimento natural, ou seja, o modelo com pré-processamento SNV, GLSW ($\alpha = 0,112$) e centragem na média. Os resultados dessa projeção estão presentes na Figura 32, na qual se observa o modelo de calibração com as amostras envelhecidas naturalmente e o conjunto de previsão foram com as amostras envelhecidas artificialmente.

Figura 32 - Projeção das amostras envelhecidas artificialmente nos modelos PLS construído com as amostras envelhecidas naturalmente (a) Gráfico do valor previsto (b) gráfico de influência e resíduo das amostras.



Fonte: A autora (2020).

Analisando o gráfico da Figura 32b, é possível observar que as amostras envelhecidas artificialmente possuem um Q residual bastante elevado no gráfico de influências, mostrando que o modelo não é eficiente para descrever essas amostras de previsão. A partir dos gráficos da Figura 31 e da Figura 32, pode-se observar que

o método de projeção de amostras no modelo anteriormente construído não se mostrou eficiente para a comparação dos envelhecimentos naturais e artificial.

Isso significa que não é possível prever a data de um documento original (envelhecidos naturalmente) utilizando um modelo construído com amostras envelhecidas artificialmente. Como também não foi obtido sucesso ao se construir o modelo com amostras envelhecidas naturalmente para prever a datação de amostras envelhecidas artificialmente. No contexto forense, esse estudo mostra que não é possível que uma amostra envelhecida artificialmente seja reconhecida como uma amostra antiga, envelhecidas naturalmente, pois analisando os modelos é possível observar um Q residual elevado para essa amostra. Mostrando ser promissor a utilização da espectroscopia de IR associadas as técnicas quimiométricas para identificar a falsificação de documentos.

6 PERSPECTIVA FUTURA

Como relatado, um dos grandes desafios encontrados no estudo desenvolvido foi a obtenção de um modelo confiável para a datação dos documentos. Para contornar esse problema, propõe-se a continuidade e o aprofundamento do estudo dos modelos construídos com amostras envelhecidas naturalmente, otimizando o parâmetro α do GLSW, a partir das superfícies de respostas construídas no trabalho atual. Também é proposto o estudo de alguns parâmetros como o RMSEC e RMSECV, bias, com a finalidade de minimizar os seus valores e maximizar o R^2 . Esses fatores devem ser levados em consideração para escolher o maior valor de α (que não altere os espectros) e o menor número das variáveis latentes possível. Além disso, testes de significância, como teste F e teste t, poderão ser aplicados para realizar a comparação dos modelos e identificar a significância dos parâmetros analisados, de forma que com toda essa informação, seja possível encontrar um modelo adequado, empregando o GLSW.

Além de otimizar o parâmetro do GLSW, outros pré-processamentos puderam ser testados como a Correção Ortogonal de Sinal (OSC, do inglês *Orthogonal Signal Correction*) que também realiza uma espécie de filtragem multivariada para a construção dos modelos de regressão; bem como outras técnicas de calibração multivariada como a projeções ortogonais a Mínimos Quadrados Parciais (OPLS: *Orthogonal Partial Least Squares*) e técnica de regressão de vetores de suporte (SVR: *Support Vector Regretion*). O SVR é uma técnica conhecida de aprendizagem de máquinas, que é capaz de tratar de modo relativamente rápido problemas de calibração multivariada lineares ou não lineares.

Outra proposta é ampliar o estudo do envelhecimento artificial em diferentes condições para investigar o processo de degradação dos papéis exposto a radiação UV e a radiação solar. Podem ser aplicadas diferentes técnicas quimiométricas para identificar as mudanças nos perfis espectrais, além técnicas de seleção de variáveis, tais como: a Importância de Variáveis na Projeção (VIP: *Variable Importance in the Projection*), Mínimos Quadrados Parciais por Intervalos (iPLS: *Interval Partial Least Square*) e o Algoritmo genético (GA: *Genetic algorithm*).

Outra proposta é estudo do envelhecimento artificial do papel, a partir de um planejamento fatorial, para analisar a influência dos fatores no processo de degradação, tais como, temperatura, radiação UV, umidade e suas possíveis interações. Essas condições de armazenamento e o tempo de exposição seriam analisados para identificar quais fatores e a quantidade de tempo necessária para produzir uma resposta química equivalente ao envelhecimento natural. Além disso, diferentes técnicas espectroscópicas também podem ser empregadas para a análise dos compostos presentes nos papéis, tais como, a espectroscopia de infravermelho próximo (NIR), médio (MIR) e/ou a espectroscopia Raman.

7 CONCLUSÃO

É possível concluir que as mudanças do papel ao longo do tempo e submetidas à condição de envelhecimento artificial causam modificações na molécula de celulose, devido à degradação. A utilização da espectroscopia na região do infravermelho médio associada a técnicas quimiométricas se mostrou promissora para o estudo da degradação do papel. Com a PCA foi possível determinar um possível processo de degradação: para os papéis couchês observou-se a degradação da camada de revestimento e para os papéis sulfites a oxidação. Para o envelhecimento artificial os modelos PLS foram utilizados para prever o tempo de exposição à condição de envelhecimento. Os modelos PLS para as amostras envelhecidas naturalmente foram construídos com o objetivo de datação de documentos. Para ambos os casos a utilização de modelos PLS se mostrou eficiente para descrever o processo de envelhecimento dos papéis, principalmente devido à utilização do pré-processamento GLSW que atua como um filtro minimizando a variabilidade das folhas de papéis do mesmo ano para as amostras envelhecidas naturalmente, e minimizando a diferença entre as marcas de papéis para as amostras envelhecidas artificialmente. A utilização de seleção de variáveis como o GA não se mostrou promissora para o estudo de nenhum tipo de envelhecimento.

Os modelos PLS foram eficientes para explicar o envelhecimento separadamente, mas quando associado um envelhecimento natural com o artificial, a partir das projeções de um tipo de envelhecimento em outro, os resultados não foram promissores para a previsão. Por outro lado, a comparação dos modelos e os resíduos mostram-se promissores para identificar se um documento artificial foi fraudado e apresentado como autêntico. Isso porque as diferenças dos processos de degradação são claras mesmo usando o filtro GLSW, sendo possível identificar que são documentos que passaram por processos diferentes. Em suma, realizar um paralelo entre documentos envelhecidos artificialmente e naturalmente é um processo bastante complexo, não sendo possível construir um modelo com amostras envelhecidas artificialmente por exposição à temperatura para então estimar a data de um documento desconhecido.

REFERÊNCIAS

- ALI, M.; EMSLEY, A.M.; HEMAN, H.; HEYWOOD R.J.; Spectroscopic studies of the ageing of cellulosic paper. **Polymer**. n. 42, p. 2893-2900, 2001.
- ALISKE, M. A. **Medidas de Espectroscopia no Infravermelho Médio para a Determinação do Teor de Biodiesel em Óleo Diesel**. Dissertação de Mestrado. Universidade Federal do Paraná, 2010.
- BEEBE K. R.; KOWALSKI B. R. An Introduction to Multivariate calibration and Analysis. **Analytical Chemistry**. v. 59, p. 1007-1017, 1987.
- BRERETON, R. **Chemometrics: Data Analysis for the Laboratory and Chemical Plant**. West Sussex: John Wiley & Sons Ltd, 2003.
- BRO, Rasmus; SMILDE, Age. Principal component analysis. **Analytical Methods**, n. 6, p. 2812–2831, 2014.
- BRUNELLE, R. L.; CRAWFORD, K. R. **Advances in the Forensic Analysis and Dating of Writing Ink**. Charles C Thomas Pub Ltd, 2003.
- BURNS, D. A.; CIURCZAK, E. W. **Handbook of Near-Infrared Analysis**. 3. Ed, Broken Sound Parkway NW: CRC Press, 2007.
- CALCERRADA, M.; GARCÍA-RUIZ, C. Analysis of questioned documents: a review, **Anal. Chim. Acta**. v. 853, p. 143-166, 2015.
- CAUSIN, V.; MAREGA, C.; MARIGO, A.; CASAMASSIMA, R.; PELUSO, G.; RIPANI, L. Forensic differentiation of paper by X-ray diffraction and infrared spectroscopy, **Forensic Science International**. v. 197, p. 70–74, 2010.
- CURIE, Marie. On The Business Of Life. 2015. Forbes. Disponível em: <https://www.forbes.com/quotes/6520/>. Acesso em: 20 jun. 2020.
- FELLER, R. L. **Accelerated Aging Photochemical and Thermal Aspects**, Research in conservation, 1994.
- FERREIRA, M.M.C.; **Quimiometria: Conceitos, Métodos E Aplicações**. 1.ed. Unicamp, 2015.

HAJJI, L.; BOUKIR, A; ASSOUIK, J; PESSANHA, S; FIGUEIRINHAS, J.L; CARVALHO, M.L. Artificial aging paper to assess long-term effects of conservative treatment. Monitoring by infrared spectroscopy (ATR FTIR), X-ray diffraction (XRD), and energy dispersive X-ray fluorescence (EDXRF). **Microchemical Journal**, v. 124, p. 646-656, 2016.

KAČÍK, F. et al. Cellulose degradation in newsprint paper ageing. **Polymer Degradation and Stability**, v. 94, n. 9, p. 1509–1514, 2009.

KHER, A.; MULHOLLAND, M.; REEDY, B.; MAYNARD, P.; Classification of Document Papers by Infrared Spectroscopy and Multivariate Statistical Techniques. **Applied Spectroscopy**. v. 55, p. 1192-1198, 2001.

MARGUTTI, S.; CONIO, G.; CALVINI, P.; PEDEMONTE, E. Hydrolytic and Oxidative Degradation of Paper. **Restaurator**, v.22, p. 67-83, 2001.

MARTÍNEZ, J. R.; NIETO-VILLENA, A.; CRUZ-MENDOZA J. Á.; ORTEGA-ZARZOSA G.; GUERREIRO A. L.; Monitoring the natural aging degradation of paper by fluorescence. **Journal of Cultural Heritage**, v. 26, p. 22-27, 2017.

MEHMOOD, T.; LILAND, K. H.; SNIPEN, L.; SÆBØ, S. A review of variable selection methods in Partial Least Squares Regression. **Chemometrics and Intelligent Laboratory Systems**, v. 118, p. 62–69, 2012.

NAES, T; ISAKSSON, T; FEARN, T; DAVIES, T. **A user-friendly guide to Multivariate Calibration and Classification**. NIR Publications, 2004.

NETO, M.N.; MOITA, G.C.; Uma introdução à análise exploratória de dados multivariado. **Química Nova**, v.21, n.4, 1997.

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.

PORCK, H. J.; TEYGELER, R. **Preservation Science Survey**, Council on Library and Information Resources Washington, DC, 2000.

RINNAN, Å.; BERG, F. VAN DEN; 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.

RISOLUTI, R.; MATERAZZI, S.; TAU, F.; RUSSO, A.; ROMOLO, F. S.; Towards innovation in paper dating: a MicroNIR analytical platform and chemometrics, **Analyst**, v. 143, p. 4394-4399, 2018.

SAVITZKY, A.; GOLAY, M. Smoothing + differentiation of data by simplified least squares procedures. **Analytical Chemistry**, v. 36, n. 8, p. 16-27, 1964.

SCHEDL A.; ZWECKMAIR T.; KIKUL F.; HENNIGES U.; ROSENAU T.; POTTHAST A.; Aging of paper – Ultra-fast quantification of 2,5-dihydroxyacetophenone, as a key chromophore in cellulose, by reactive paper spray-mass spectrometry. **Talanta**, v. 167, p. 672-680, 2017.

SKOOG, D.; HOLLER, J.; CROUCH, S. **Principles of Instrumental Analysis**. 7th. ed. [s.l.] Cengage Learning, 2018.

SIEGEL, J. A. **Forensic Chemistry Fundamentals and Applications**. West Sussex: John Wiley & Sons Ltd, 2016.

SIESLER, H. W.; OZAKI, Y.; KAWATA, S.; HEISE, H. M. **Near-Infrared Spectroscopy Principles, Instruments, Applications**. 1, Ed, Weinheim: WILEY-VCH, 2002.

SILVA, C.S.; PIMENTEL, M. F.; AMIGO, J. M.; GARCÍA-RUIZ, C.; ORTEGA-OJEDA, F. Chemometric approaches for document dating: Handling paper variability, **Analytica Chimica Acta**, v. 1031, p. 28-29, 2018.

SPENCE, L. D.; BAKER, A. T.; BYRNE, J. P. Characterization of document paper using elemental compositions determined by inductively coupled plasma mass spectrometry, **J. Anal. At. Spectrom.** v. 15, n. 166, p. 813-819, 2000.

TRAFELA, T.; STRLIC, M.; KOLAR, J.; LICHTBLAU, D. A.; ANDERS, M.; MENCIGAR, D.P.; PIHLAR, B. Nondestructive Analysis and Dating of Historical Paper Based on IR Spectroscopy and Chemometric Data Evaluation. **Analytical Chemistry**, v. 79, p. 6319 – 6323, 2007.

UDRIȘTIOIU, F. M.; BUNACIU, A. A.; ABOUL-ENEIN³ H. I.; TANASE¹, I. G. Application of Micro-Raman and FT - IR Spectroscopy in Forensic Analysis of Questioned Documents. **Gazi University Journal of Science**, v.25, p. 371-375, 2012.

VAN ES, A.; KOEIJER, J.; VAN DER PEIJL, G. Discrimination of document paper by XRF, LA-ICP-MS and IRMS using multivariate statistical techniques, **Science and Justice**. v. 49, p. 120–126, 2009.

WISE, B.M.; GALLAGHER, N.B.; BRO, R.; SHAVER, J.; WINDIG, W.; KOCH, R.S. Chemometrics Tutorial for PLS _ Toolbox and Solo, Eigenvector Research, **Inc., Wenatchee**, 2006.

WOLD, S.; SJOSTROMA, M.; ERIKSSON, L. PLS-regression: a basic tool of chemometrics. **Chemometrics and Intelligent Laboratory Systems**, v. 58, p. 109–130, 2001.

WORKMAN j.; WEYER L. **Practical guide to interpretive near-infrared spectroscopy**. Broken Sound Parkway NW: CRC Press, 2007.

XIA, J.; ZHANG, J.; ZHAO, Y.; HUANG, Y.; XIONG, Y.; MIN, S. Fourier transform infrared spectroscopy and chemometrics for the discrimination of paper relic types, **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**. v. 219. p. 8-14, 2019.

ZIĘBA-PALUSA, J.; TRZCIŃSKA, B.; WESEŁUCHA-BIRCZYŃSKA A.; MOSKAL, P.; SACHARZ, J.; The sequence of changes observed during degradation process of paper by the use of UV/VIS and FTIR spectrometry with application of the PCA and 2D correlation method for forensic purposes, **Journal of Molecular Structure**, v. 1205, p. 127651, 2020.

ZIEBA-PALUS, J.; WESELUCHA-BIRCZYNSKA, A.; TRZCINSKA, B.; KOWALSKI, R.; MOSKAL, P.; Analysis of degraded papers by infrared and Raman spectroscopy for forensic purposes, **Journal of Molecular Structure**, v.1140, p. 154 162, 2016.

ZORZETTI, B.M.; SHAVER, J.M.; HARYNUK, J.J. Estimation of age of weathered mixture of volatile organic compounds”, **Analytica Chimica Acta**, v. 694, p. 31-37, 2011.

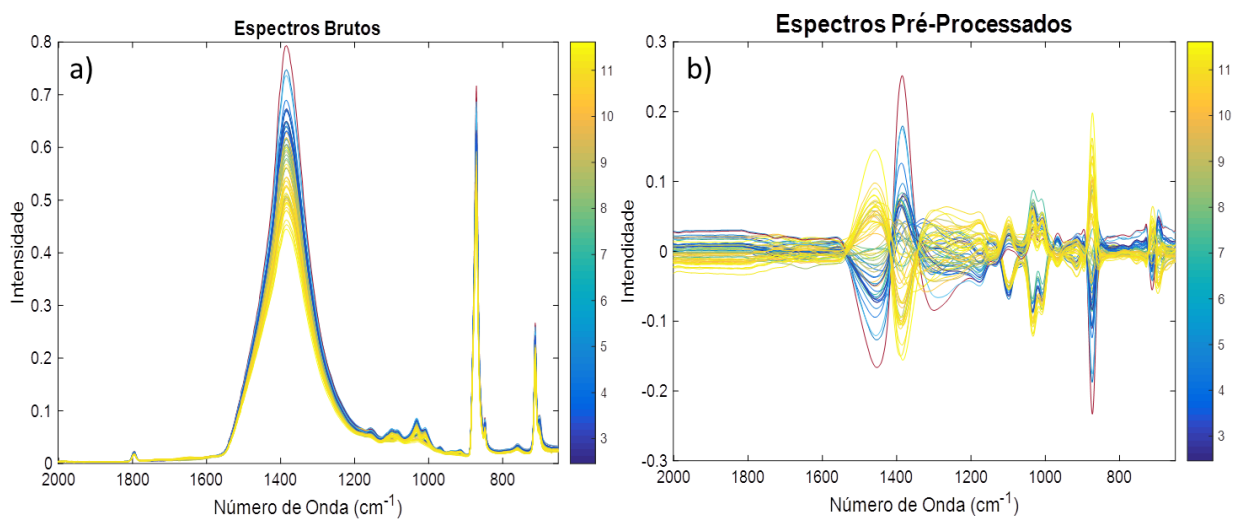
ZOU, X; GURNAGUL, N.; UESAKA, T; BOUCHARD, J. Accelerated aging of papers of pure cellulose: mechanism of cellulose degradation and paper embrittlement. **Polymer Degradation and Stability**. v. 43. p. 393-402, 1994.

ZOU, X; UESAKA, T; GURNAGUL, N. Predictions of paper permanence by accelerated aging I. Kinetic analysis of the aging process. **Cellulose**, 3, 243-267, 1996a.

ZOU, X; UESAKA, T; GURNAGUL, N. Predictions of paper permanence by accelerated aging II. Comparison of the predictions with natural aging results. **Cellulose**, 3, 269-279, 1996b.

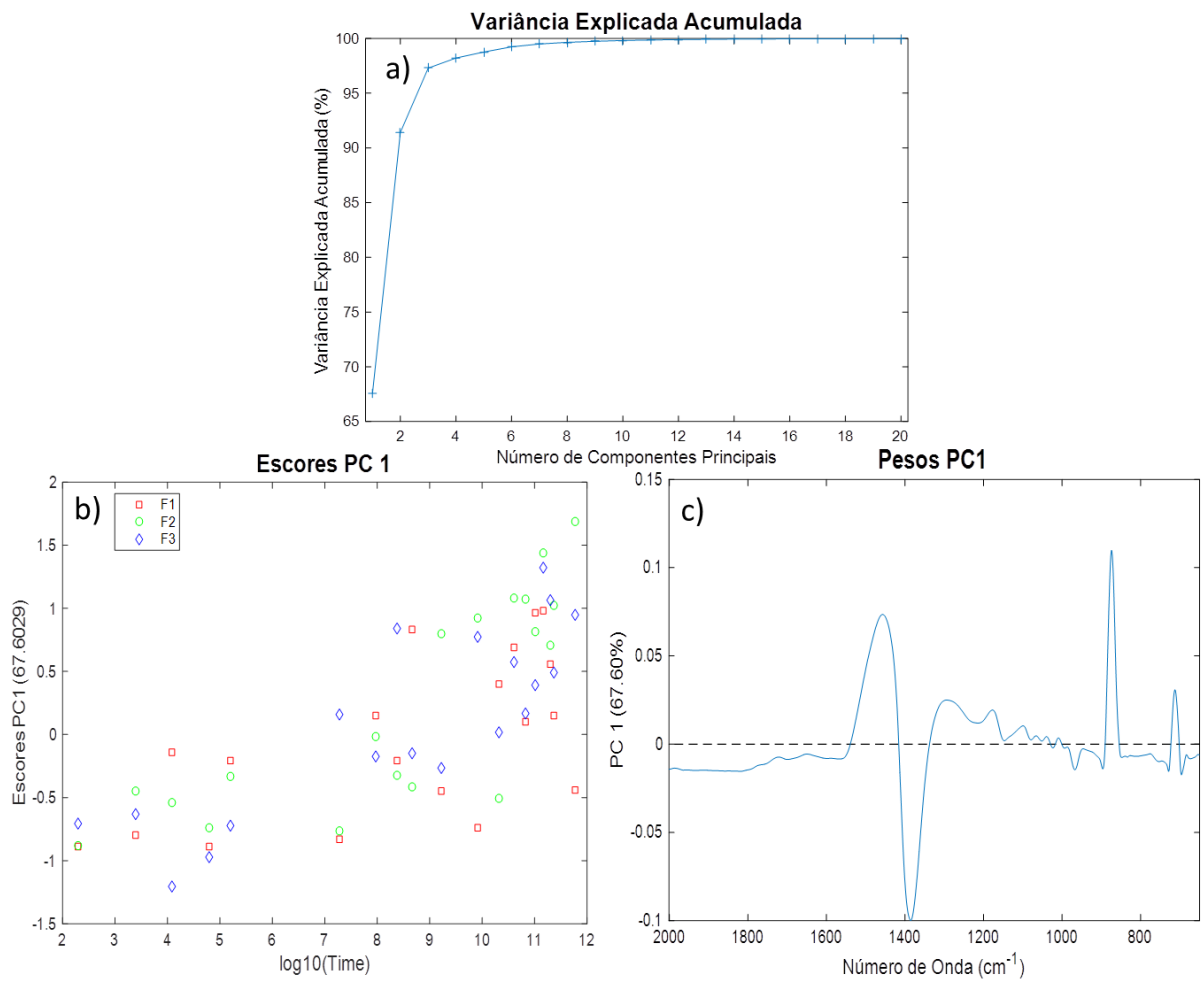
APÊNDICE A – ANÁLISE DOS COMPONENTES PRINCIPAIS PARA OS PAPÉIS COUCHÊS

Figura 33 - Gráfico do efeito do pré-processamento para os papéis couchês C2 (a) espectros brutos e (b) espectros pré-processados.



Fonte: A autora (2020).

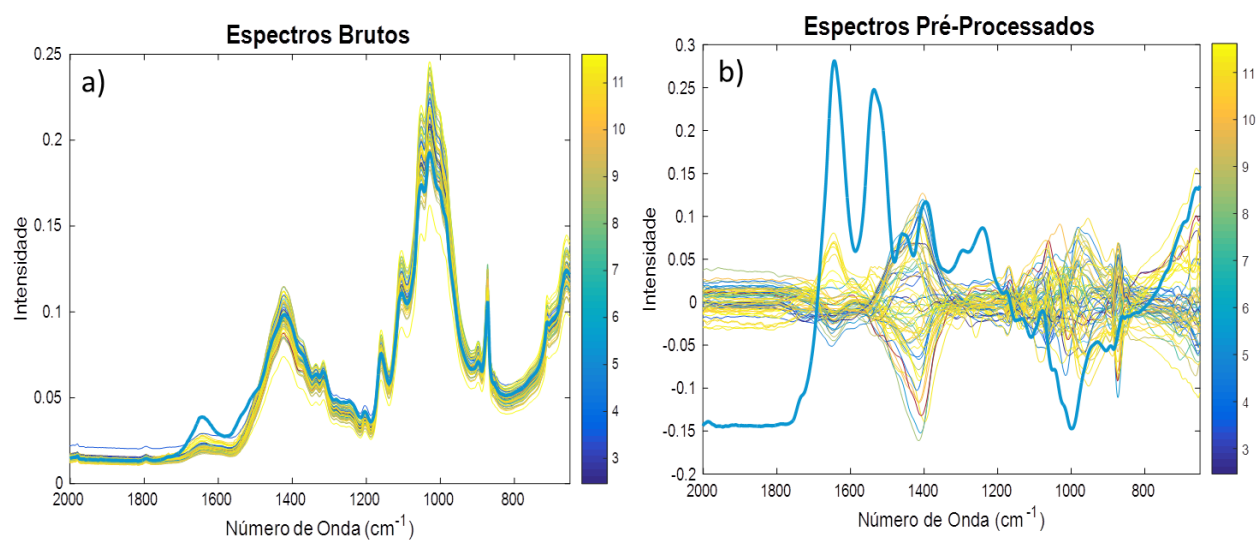
Figura 34 - Gráfico da PCA para os papéis couchês C2 (a) Variância explicada acumulativa para o modelo (b) escores da PC1 versus o log do tempo (c) pesos da PC1.



Fonte: A autora (2020).

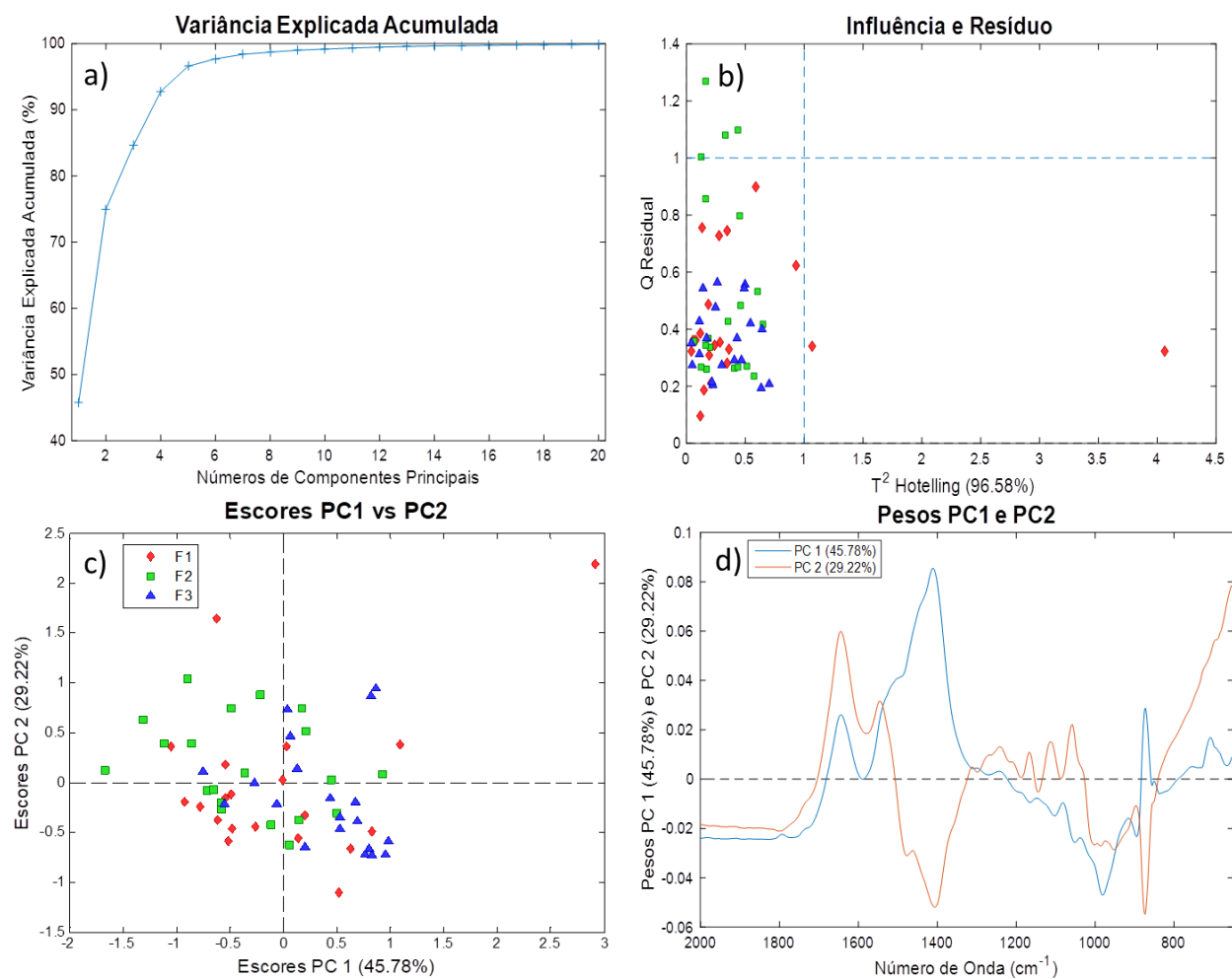
APÊNDICE B – ANÁLISE DOS COMPONENTES PRINCIPAIS PARA OS PAPÉIS RECICLADOS

Figura 35 - Gráfico do efeito do pré-processamento para os papéis reciclados R2 (a) espectros brutos e (b) espectros pré-processados.



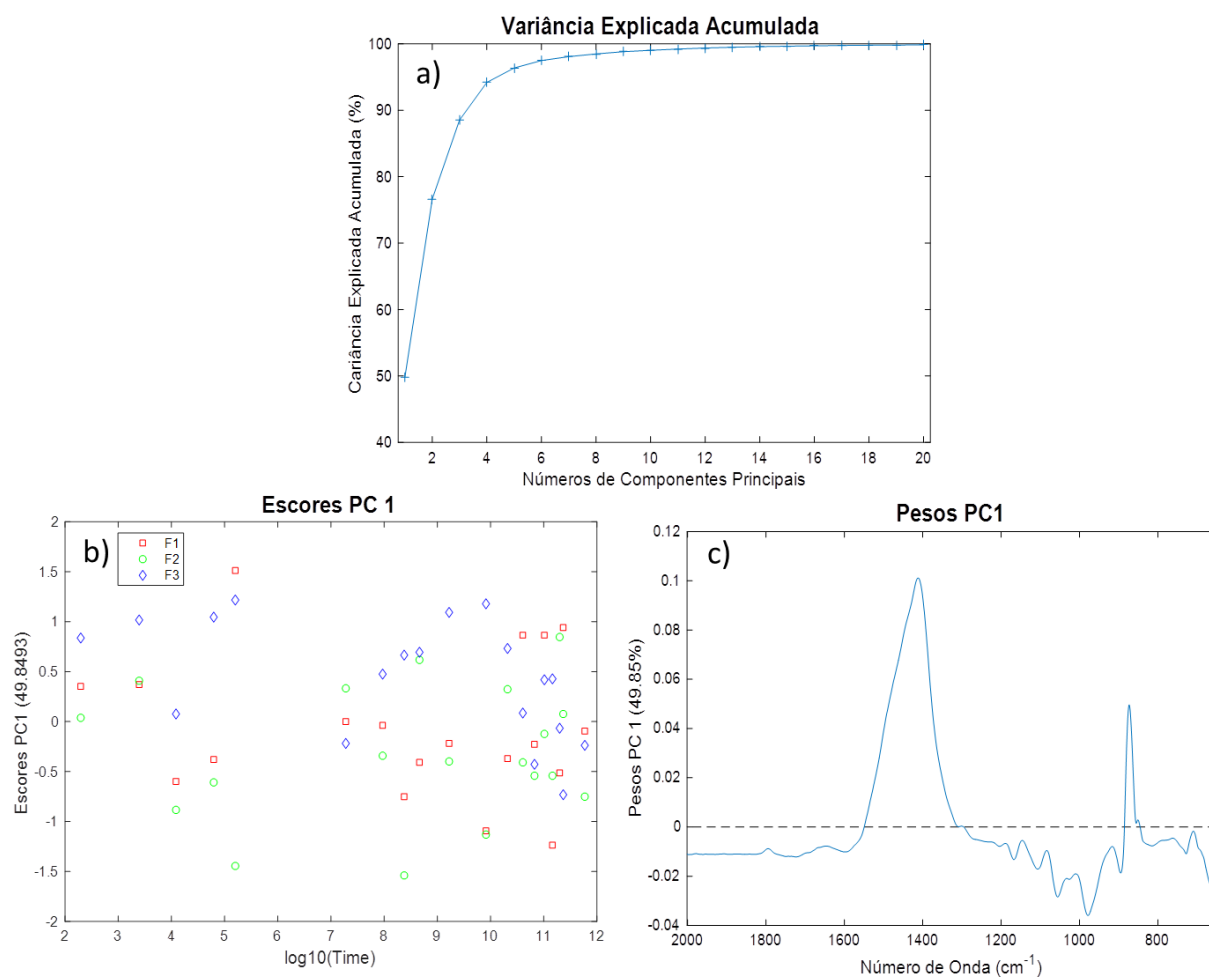
Fonte: A autora (2020).

Figura 36 - Gráfico da PCA para os papéis reciclados R2 (a) Variância explicada acumulada para o modelo, (b) gráfico de influencia e resíduo das amostras, (c) escores da PC3 versus a PC4 (d) pesos da PC3 e da PC4.



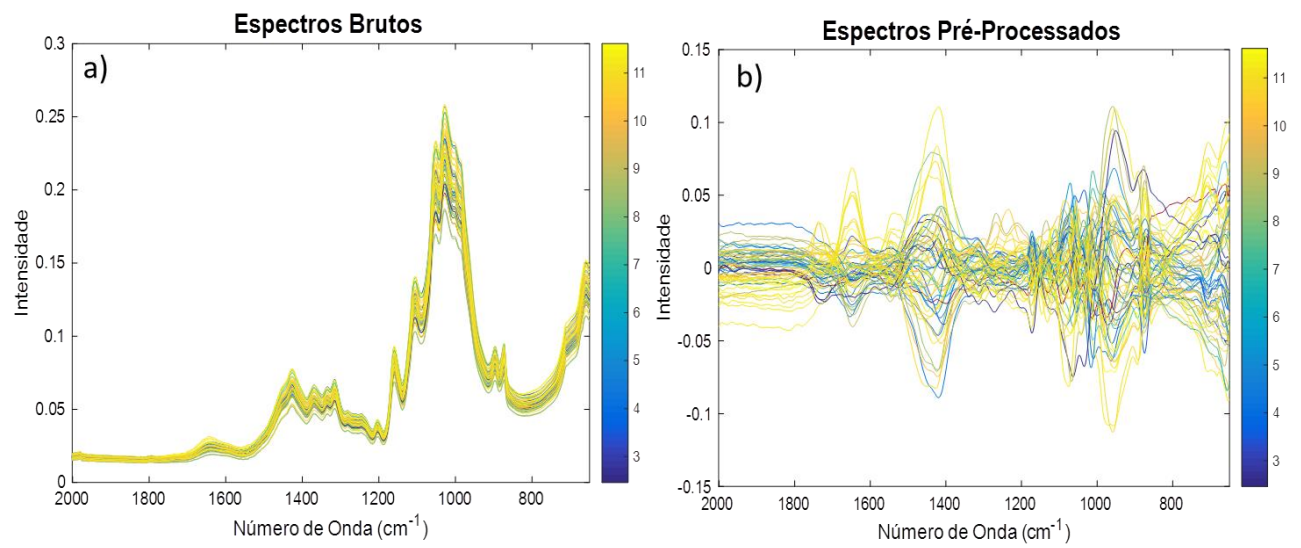
Fonte: A autora (2020).

Figura 37 - Gráfico para o novo modelo PCA, sem as amostras anômalas, para os papéis reciclados R2 (a) Variância explicada acumulativa para o modelo (b) escores da PC1 versus o log do tempo (c) pesos da PC1.



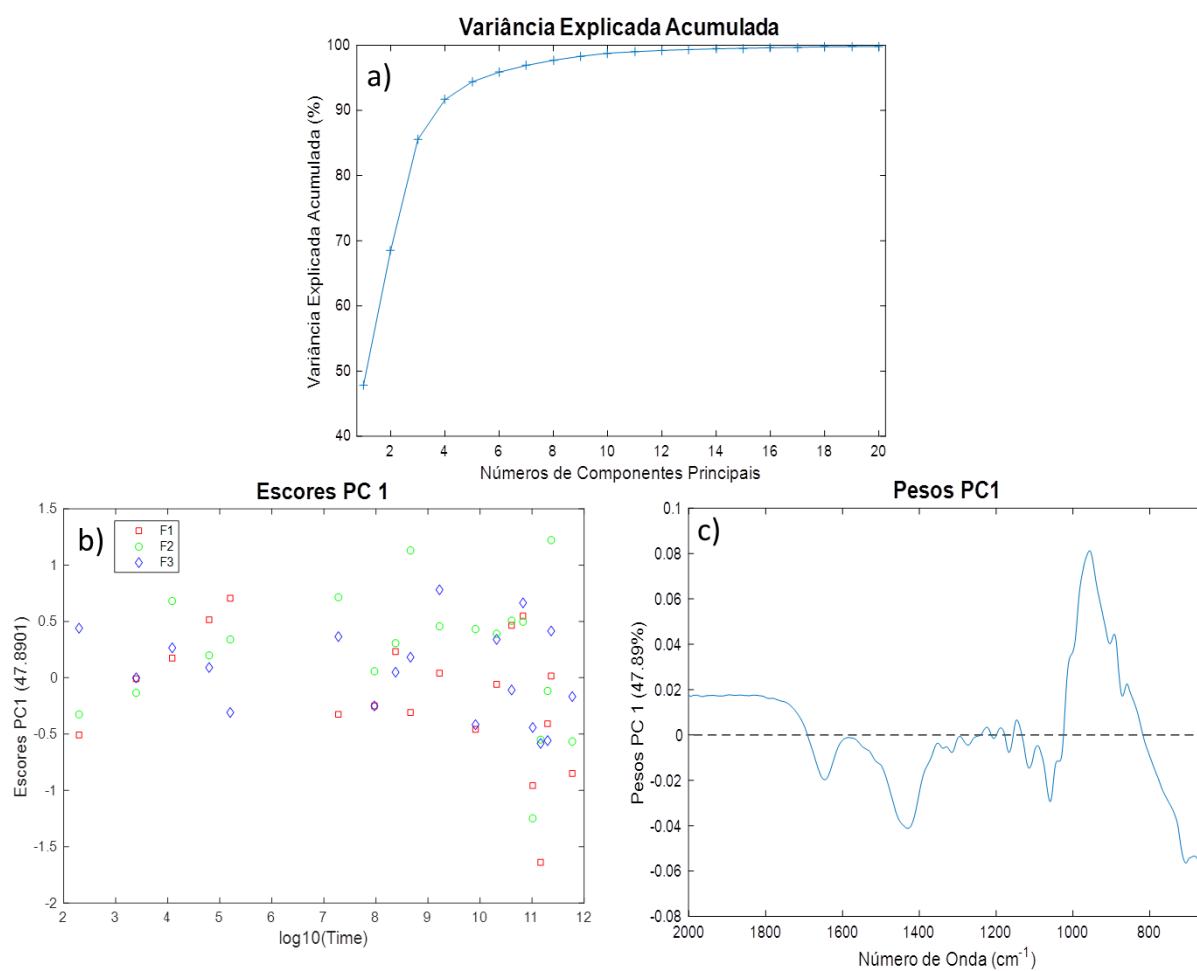
Fonte: A autora (2020).

Figura 38 - Gráfico do efeito do pré-processamento para os papéis reciclados R3 (a) espectros brutos e (b) espectros pré-processados.



Fonte: A autora (2020).

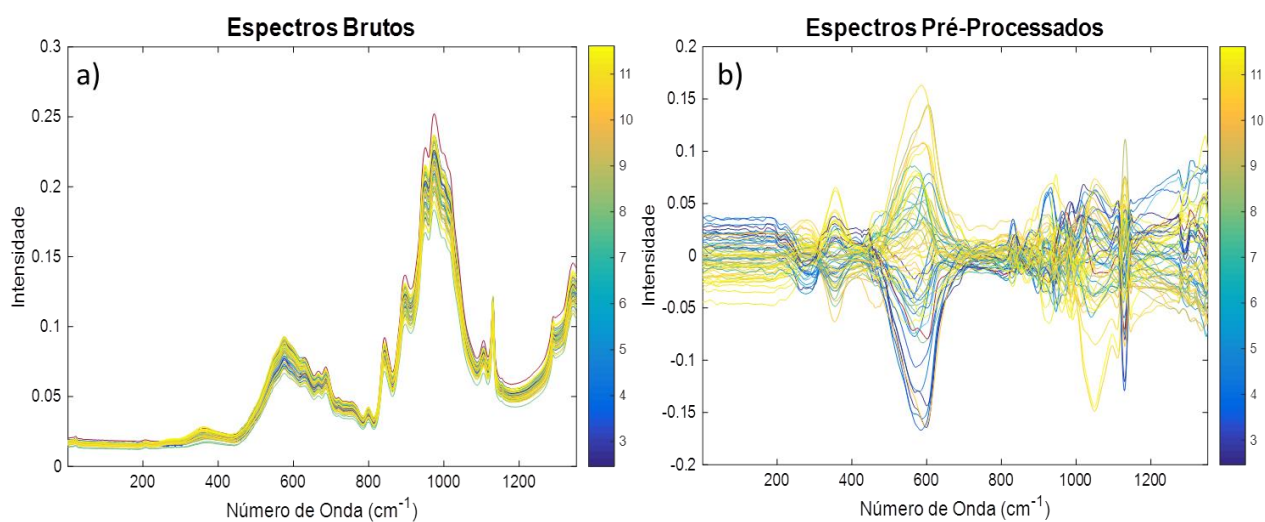
Figura 39 - Gráfico da PCA para os papéis reciclados R3 (a) Variância explicada acumulativa para o modelo (b) escores da PC1 versus o log do tempo (c) pesos da PC1.



Fonte: A autora (2020).

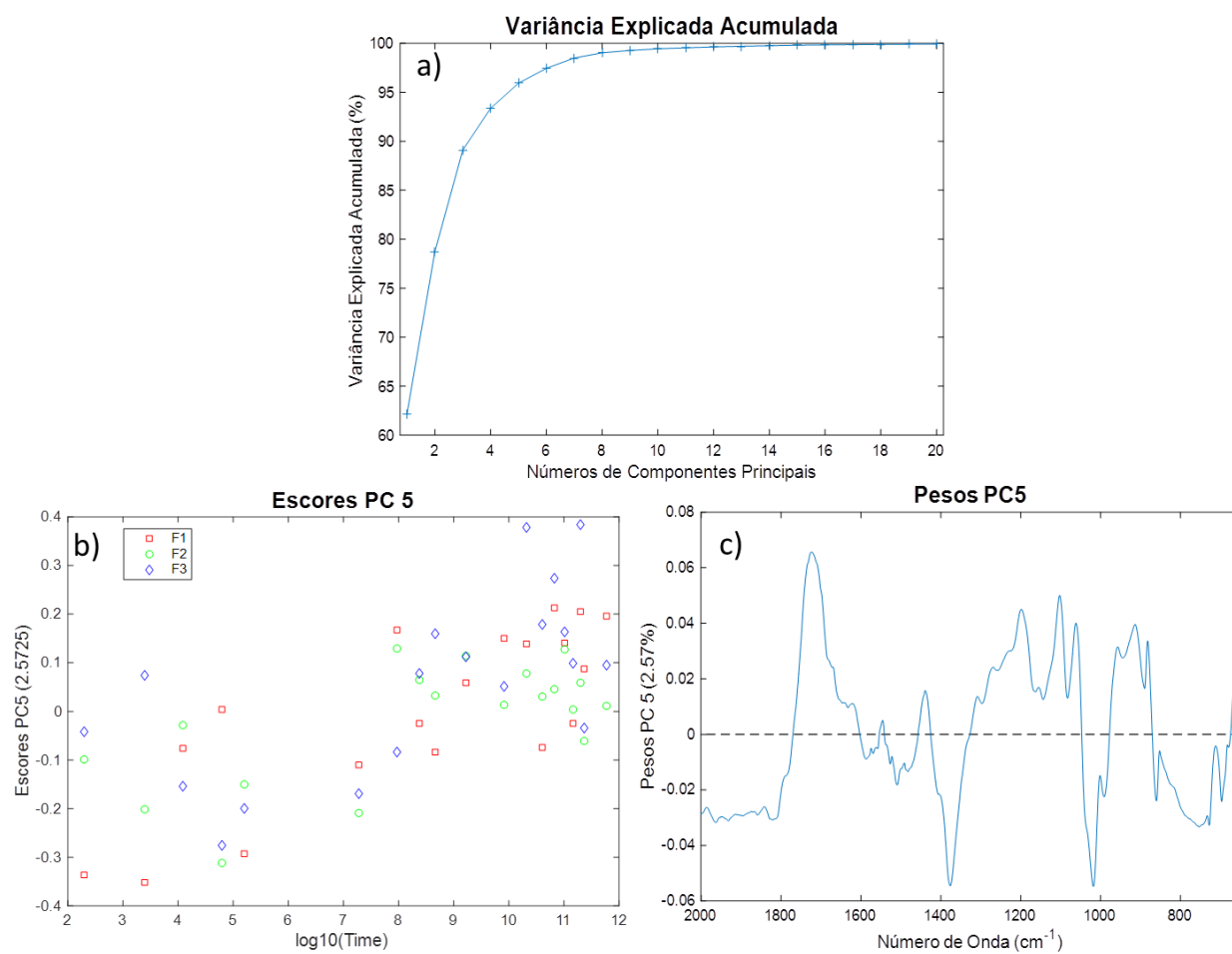
APÊNDICE C – ANÁLISE DOS COMPONENTES PRINCIPAIS PARA OS PAPÉIS SULFITES

Figura 40 - Gráfico do efeito do pré-processamento para os papéis sulfites S2 (a) espectros brutos e (b) espectros pré-processados.



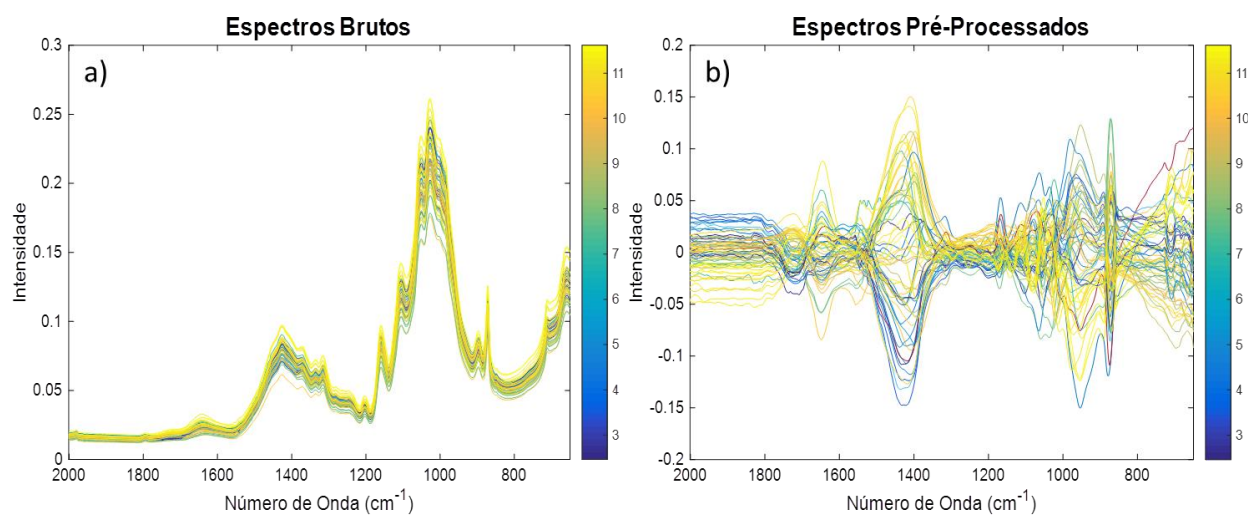
Fonte: A autora (2020).

Figura 41 - Gráfico da PCA para os papéis sulfites S2 (a) Variância explicada acumulativa para o modelo (b) escores da PC1 versus o log do tempo (c) pesos da PC1.



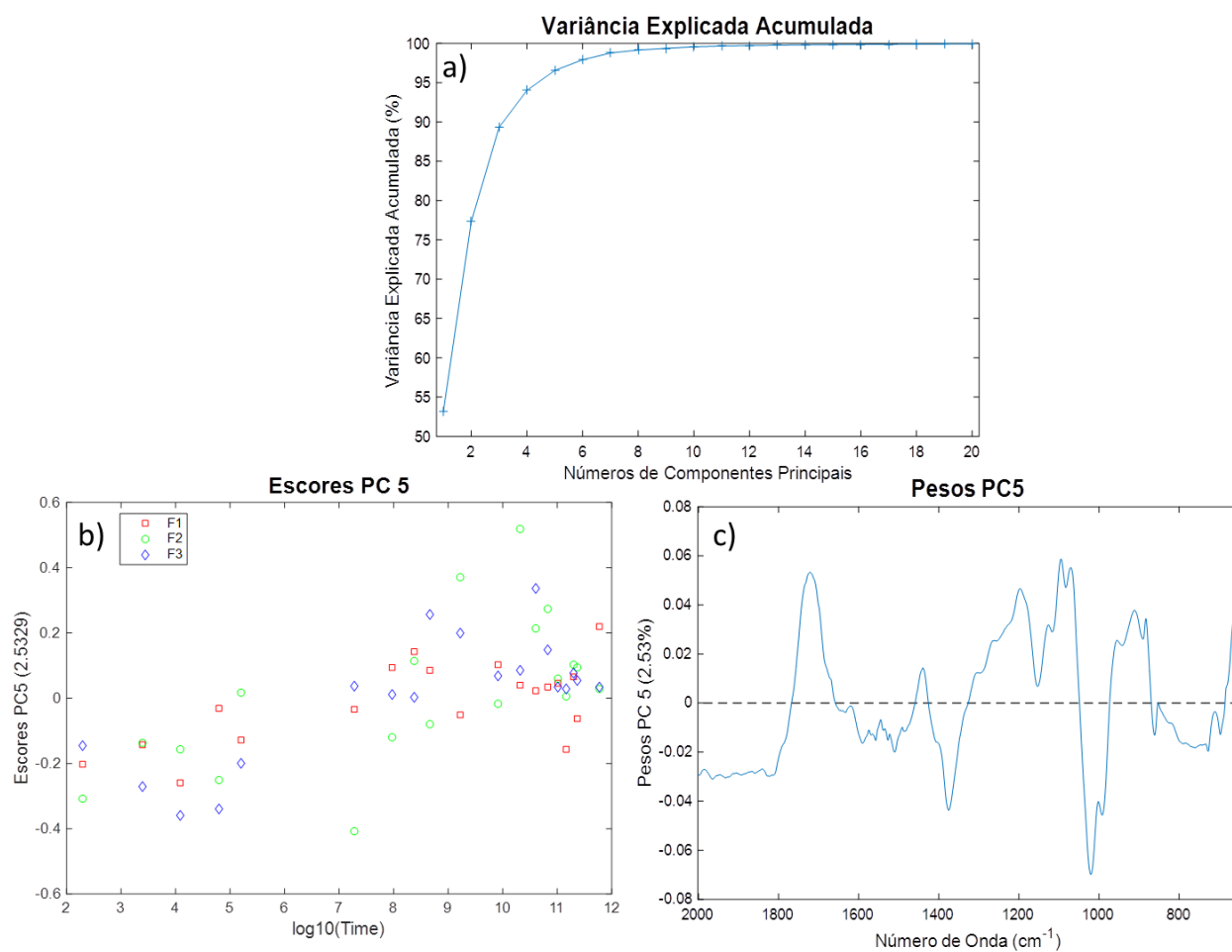
Fonte: A autora (2020).

Figura 42 - Gráfico do efeito do pré-processamento para os papéis sulfites S3 (a) espectros brutos e (b) espectros pré-processados.



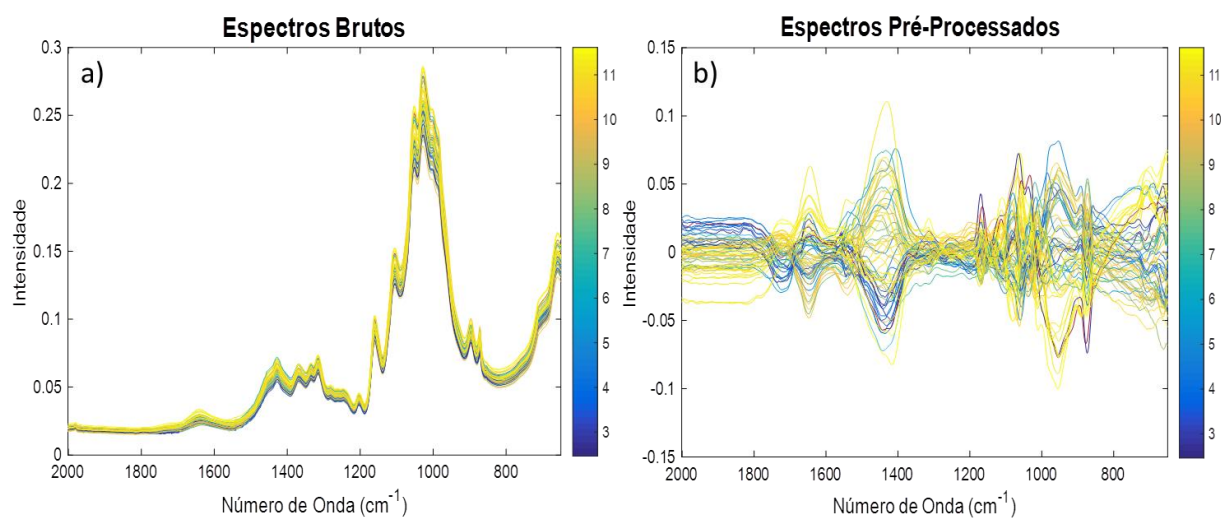
Fonte: A autora (2020).

Figura 43 - Gráfico da PCA para os papéis sulfites S3 (a) Variância explicada acumulativa para o modelo (b) escores da PC1 versus o log do tempo (c) pesos da PC1.



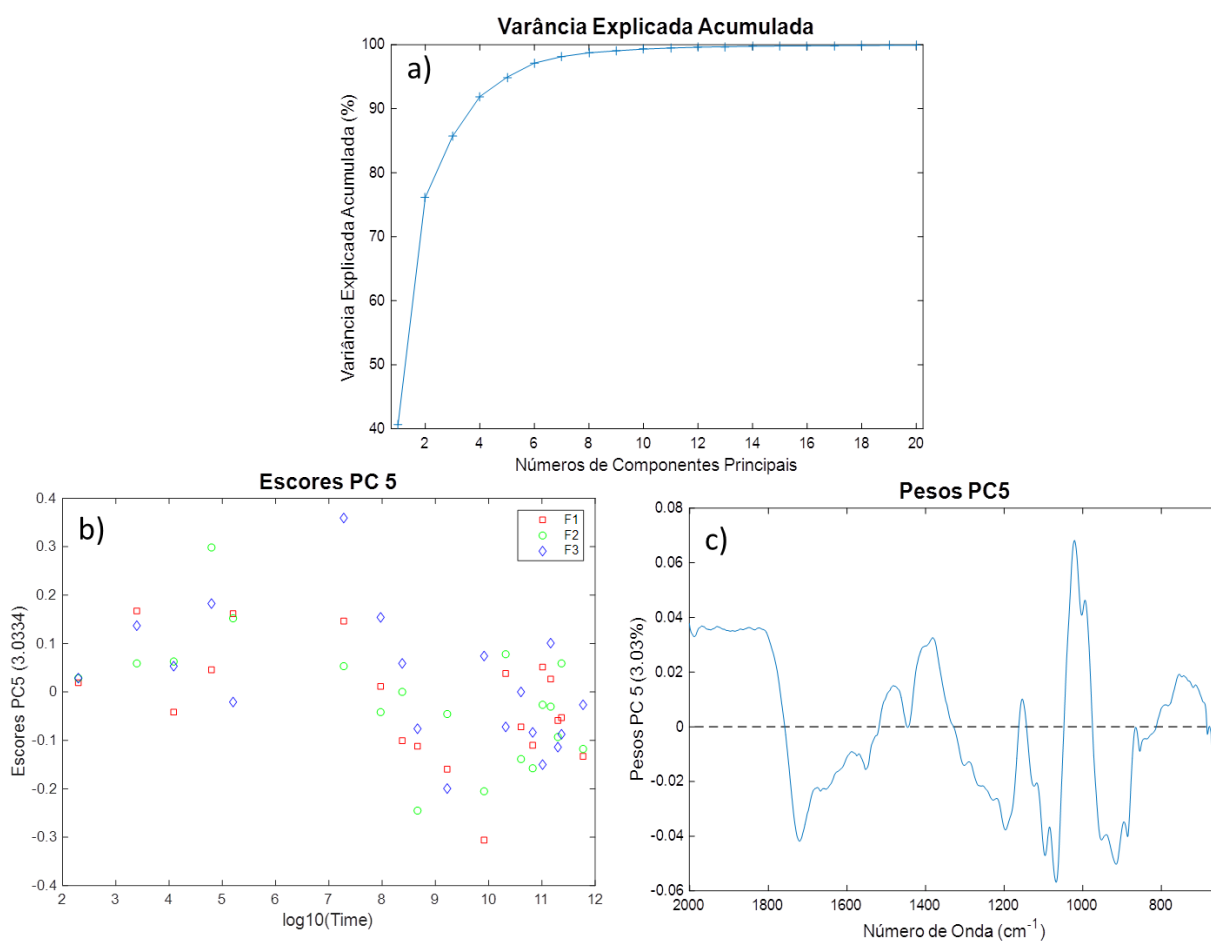
Fonte: A autora (2020).

Figura 44 - Gráfico do efeito do pré-processamento para os papéis sulfites S4 (a) espectros brutos e (b) espectros pré-processados.



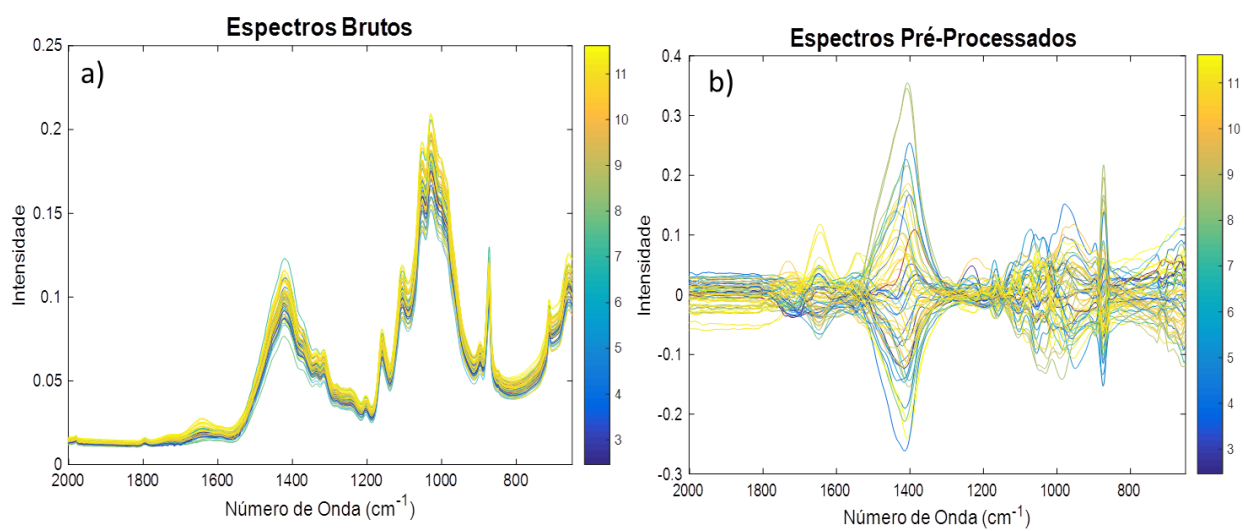
Fonte: A autora (2020).

Figura 45 - Gráfico da PCA para os papéis sulfites S4 (a) Variância explicada acumulativa para o modelo (b) escores da PC1 versus o log do tempo (c) pesos da PC1.



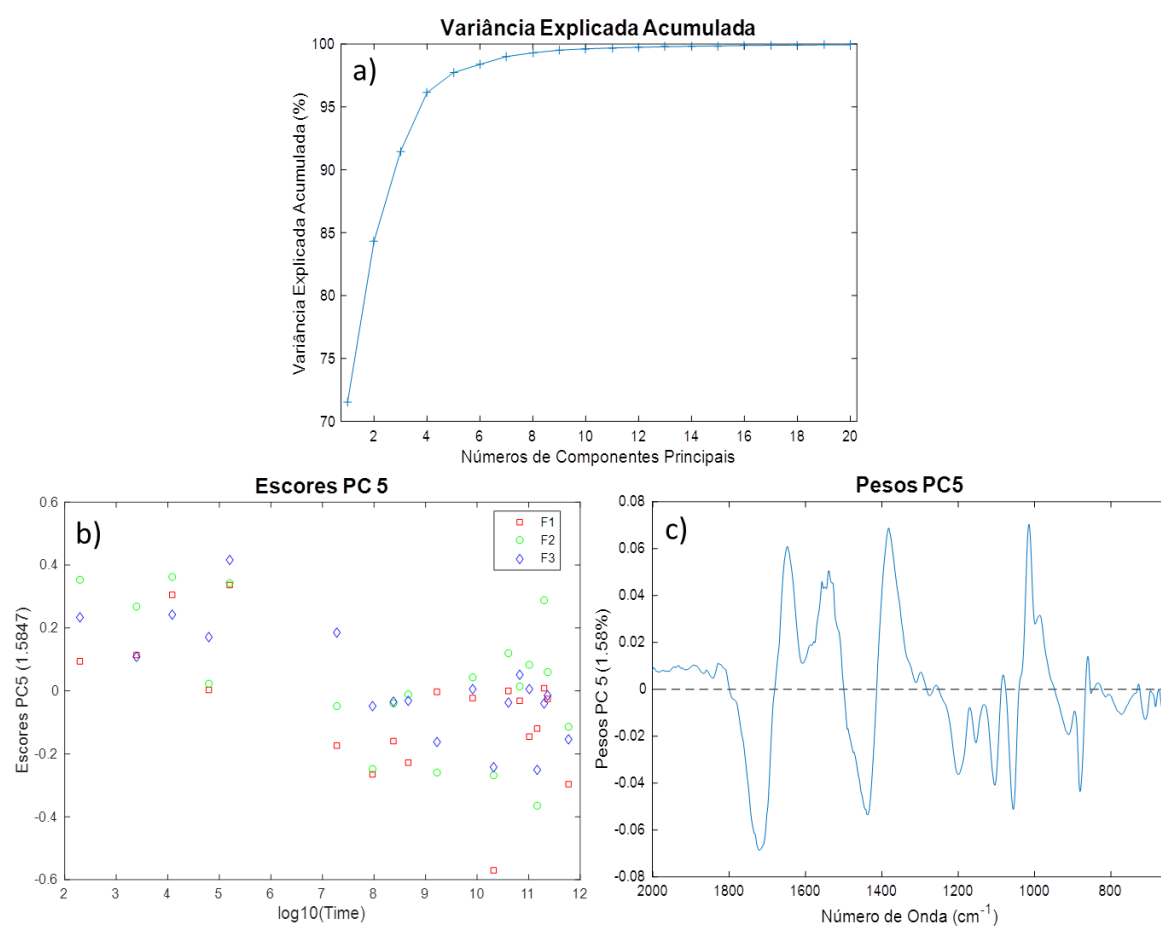
Fonte: A autora (2020).

Figura 46 - Gráfico do efeito do pré-processamento para os papéis sulfites S5 (a) espectros brutos e (b) espectros pré-processados.



Fonte: A autora (2020).

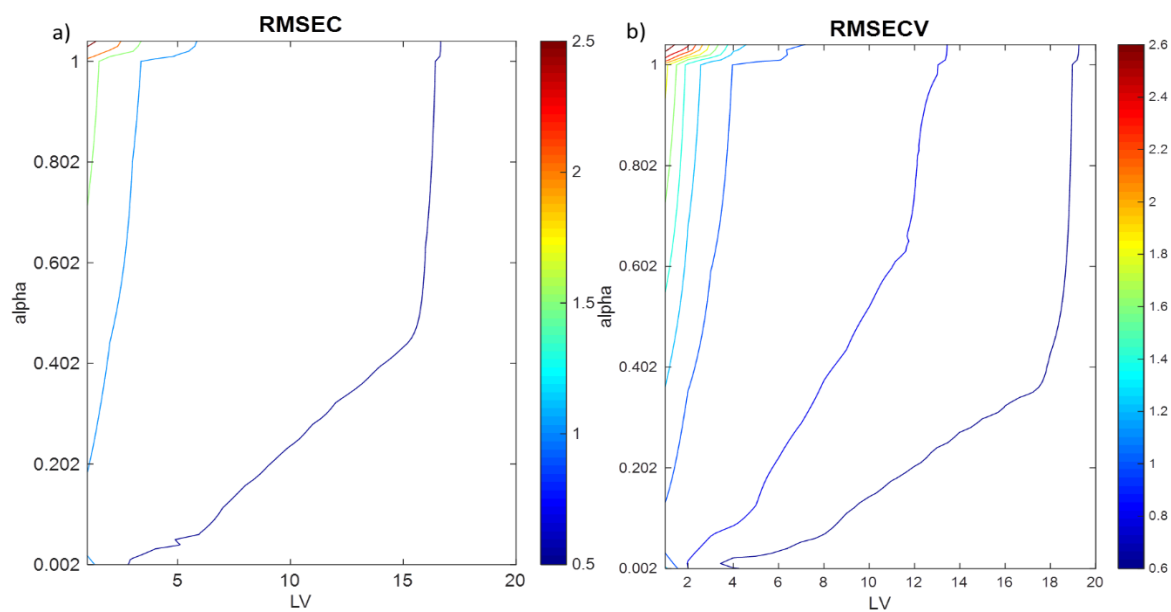
Figura 47 - Gráfico da PCA para os papéis sulfites S5 (a) Variância explicada acumulativa para o modelo (b) escores da PC1 versus o log do tempo (c) pesos da PC1.



Fonte: A autora (2020).

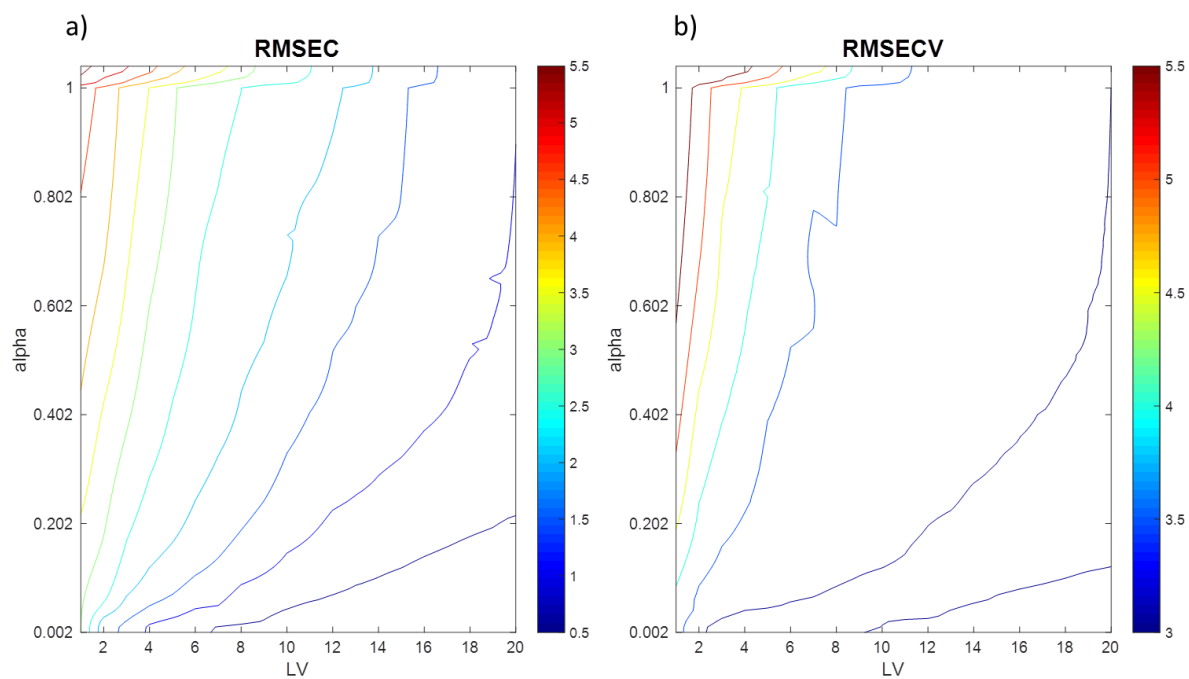
APENDÊCE D – GRÁFICOS DE CONTORNO DAS SUPERFÍCIES DE RESPOSTA

Figura 48 - Gráfico de contorno das superfícies para os papéis envelhecidos artificialmente para o (a) RMSEC e (b) RMSECV



Fonte: A autora (2020).

Figura 49 - Gráfico de contorno das superfícies para os papéis envelhecidos naturalmente para o (a) RMSEC e (b) RMSECV.



Fonte: A autora (2020).