



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

PEDRO QUEIROZ GUIMARÃES

**METODOLOGIA BASEADA EM NIRS E QUIMIOMETRIA PARA A
DETERMINAÇÃO DE PARÂMETROS DE QUALIDADE
DA QUITOSANA PARA FINS BIOMÉDICOS**

**CAMPINA GRANDE
2017**

PEDRO QUEIROZ GUIMARÃES

**METODOLOGIA BASEADA EM NIRS E QUIMIOMETRIA PARA A
DETERMINAÇÃO DE PARÂMETROS DE QUALIDADE
DA QUITOSANA PARA FINS BIOMÉDICOS**

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

Área de concentração: Química Analítica.

Orientadora: Profa. Dra. Simone da Silva Simões.

Coorientador: Prof. Dr. Marcus Vinícius Lia Fook.

**CAMPINA GRANDE
2017**

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

G963m Guimarães, Pedro Queiroz.
Metodologia baseada em NIRS e Quimiometria para a determinação de parâmetros de qualidade da quitosana para fins biomédicos [manuscrito] / Pedro Queiroz Guimarães. - 2017
75 p. : il. colorido.

Digitado.

Dissertação (Mestrado em Química - Mestrado) - Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2017.

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

"Coorientação: Prof. Dr. Marcus Vinícius Lia Fook, UFCG - Universidade Federal de Campina Grande"

1. Quitosana. 2. Espectroscopia no infravermelho próximo.
3. Quimiometria. 4. SPXY.

21. ed. CDD 660

PEDRO QUEIROZ GUIMARÃES

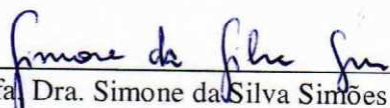
METODOLOGIA BASEADA EM NIRS E QUIMIOMETRIA PARA A
DETERMINAÇÃO DE PARÂMETROS DE QUALIDADE
DA QUITOSANA PARA FINS BIOMÉDICOS

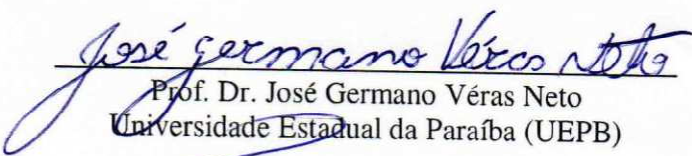
Dissertação apresentada ao Programa de Pós-Graduação em Química da Universidade Estadual da Paraíba, como requisito parcial à obtenção do título de Mestre em Química.

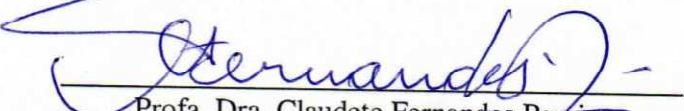
Área de concentração: Química Analítica.

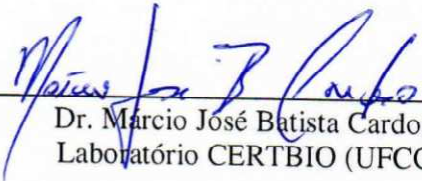
Aprovada em: 25/09/2017.

BANCA EXAMINADORA


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


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


Prof.ª Dra. Claudete Fernandes Pereira
Universidade Federal de Pernambuco (UFPE)


Dr. Marcio José Batista Cardoso
Laboratório CERTBIO (UFPE)

Dedico este trabalho a minha querida esposa Maria Geiza, pelo seu amor, companheirismo, carinho e dedicação ofertados a minha pessoa. Não apenas durante o desenvolvimento deste trabalho, mas em todos os momentos que precisei.

AGRADECIMENTOS

Ao Deus, Todo Poderoso, pelas maravilhas que tem feito em minha vida e por me fazer persistir sempre, mesmo diante de dificuldades colossais.

A minha querida Orientadora e Professora, Simone da Silva Simões, pelo apoio, estímulo, atenção, confiança e paciência que sempre teve para comigo durante a orientação deste trabalho. E também por ter me apresentado à filosofia do “caminho do meio”.

Aos meus queridos pais (Joselina e Heleno) e também aos meus queridos irmãos (Auri, Alexandre, Zé Queiroz, Maria Aparecida, Maria José e Umberto), pela compreensão, estímulo, carinho e por sempre acreditarem em meu potencial.

A minha adorável esposa Maria Geiza, pelo seu imenso amor, carinho, companheirismo, paciência, apoio incondicional e por ter sido o meu Anjo da Guarda nos momentos solitários de laboratório.

Ao coordenador do CERTBIO, Professor Marcus Vinícius Lia Fook, pelo apoio, estímulo, atenção, conselhos e ensinamentos a mim ofertados. E também por ter contribuído na construção deste trabalho como Coorientador.

Aos caros colegas Sistema de Gestão do CERTBIO, Francisco Antonio, Hugo Yves, Suyana, Dailma, Tânia, Livia, Gislaine e Professora Roberta, pelo apoio e compreensão durante os momentos em que precisei me dedicar exclusivamente ao mestrado.

A Milena Costa (Xuxinha) e sua mãe no CERTBIO (Paulinha), pelos conselhos e amizade. Agradeço também a Seu Sérgio, pelos célebres momentos de descontração.

Aos caros colegas do Laboratório de Processamento III, Bianca, Camila, Raid e Vítor, pelo apoio desde a produção às análises das amostras de quitosana. Em especial, agradeço a Eunice Paloma, pelas enormes contribuições neste trabalho. Agradeço ainda, a Márcio, Filipe e Erivan Junior, pelo apoio e colaborações na construção deste trabalho.

Aos caros colegas pioneiros do Mestrado em Química da UEPB, Danúbio, Júlio, Adriana (Fessora), Érica, Kleilton, Zé Roberto, Jamire, Ana Zélia, Sheila, Iranilma e Rayssa, pelo apoio e amizade construída durante o curso.

Ao PPGQ-UEPB, em especial aos Professores Mary, Rodrigo, Ricardo, Germano e Juracy, pelas contribuições ofertadas à minha formação.

A CAPES pelo apoio financeiro e ao Laboratório CERTBIO por proporcionar toda a estrutura necessária ao desenvolvimento desta pesquisa.

De maneira geral, agradeço a todos que contribuíram de forma direta ou indireta para o desenvolvimento deste trabalho.

“O impossível não é um fato: é uma opinião.”

(Mario Sergio Cortella)

RESUMO

A quitosana é um biomaterial que tem como principais características de qualidade a massa molar (MM) e o grau de desacetilação (GD), que influenciam praticamente todas as suas propriedades funcionais. Deste modo, é imprescindível a determinação de ambas para o fornecimento de uma matéria-prima de qualidade. As metodologias padrão utilizadas para determinação destes parâmetros são a viscosimetria e a espectroscopia de infravermelho médio, que, apesar de precisas e exatas, apresentam algumas dificuldades operacionais. Uma alternativa viável para contornar esses problemas, é o desenvolvimento de metodologias baseadas na espectroscopia no infravermelho próximo (NIR) e quimiometria, proposta neste trabalho. Para o desenvolvimento do modelo multivariado foi necessária a ampliação da variação dos parâmetros de interesse. Para isto, foram produzidos dez lotes de quitosana, variando-se os tempos de desacetilação em: 3, 4, 5, 6 e 7 horas, sendo produzidas 5 amostras por lote e 50 amostras no total. Todas as mostras foram caracterizadas em termos de GD e MM de acordo com as metodologias de referências. As mesmas amostras também foram analisadas por espectroscopia NIR. Os espectros NIR das amostras foram registrados, em triplicatas, na faixa espectral de 9.000 a 4.000 cm^{-1} , utilizando-se 32 varreduras e resolução de 8 cm^{-1} , totalizando 150 espectros. Para a construção dos modelos, vários pré-processamentos espectrais foram avaliados em relação a capacidade preditiva. Os conjuntos de calibração (100 amostras) e predição (50 amostras) foram selecionados com o auxílio do algoritmo SPXY. Também foi avaliada e comparada a capacidade preditiva dos modelos construídos utilizando toda a faixa espectral de trabalho com aqueles construídos utilizando as variáveis selecionadas por algoritmos de seleção de variáveis. A avaliação da capacidade preditiva dos modelos foi realizada pela análise de figuras de mérito. Com base nestes parâmetros foi constatado que o melhor pré-processamento espectral foi a 1ª derivada com janela de 5 e polinômio de 1ª ordem para o GD. Para a MM, a melhor performance preditiva foi mostrada pelo EMSC. De forma geral, foi constatado que os modelos construídos utilizando os coeficientes de regressão gerados pelo teste de incerteza de Martens (coeficientes Jack-Knife) apresentavam melhor performance preditiva que os modelos construídos com todas as variáveis espectrais ou com as variáveis espectrais selecionadas pelo Algoritmo das Projeções Sucessivas (SPA). Os erros de predição obtidos para GD e MM foram 1,85% e 29,08 KDa, respectivamente. O erro obtido para GD é menor que o permitido para o método de referência. No entanto para a massa molar, o modelo não mostrou um desempenho satisfatório. Deste modo, fica clara a viabilidade da aplicação das metodologias baseadas na espectroscopia NIR e quimiometria para a determinação do GD na quitosana para fins biomédicos, produzida pelo CERTBIO.

Palavras-Chave: Quitosana. Espectroscopia no infravermelho próximo. Quimiometria. SPXY.

ABSTRACT

Chitosan is a biomaterial in which the main quality characteristics are molar mass (MM) and degree of deacetylation (DD), that influence almost all of its functional properties. Thus, it is essential to determine both, to the supply of a quality raw material. The standard methodologies used to determine these parameters are viscosimetry and medium infrared spectroscopy, which, although accurate, present some operational difficulties. A feasible alternative to overcome these problems is the development of methodologies based on the near infrared (NIR) spectroscopy and chemometrics, proposed in this work. To develop the multivariate model, it was necessary to increase the variation of the parameters of interest. Thus, ten batches of chitosan were produced varying the deacetylation times in 3, 4, 5, 6 and 7 hours, producing 5 samples per batch and 50 samples in total. All samples were characterized in terms of DD and MM, according to the reference methodologies. The same samples were also analyzed by NIR spectroscopy. The NIR spectra of the samples were recorded in triplicates in the spectral range of 9,000 to 4,000 cm^{-1} , using 32 scans and 8 cm^{-1} resolution, totalizing 150 spectra. To build the models, several spectral pre-processings were evaluated in relation to the predictive capacity. The calibration (100 samples) and prediction (50 samples) sets were selected with the assistance of the SPXY algorithm. The predictive capacity of the built models using the full spectral range of work was also evaluated and compared with those built using the variables selected by algorithms of variable selection. The evaluation of the predictive capacity of the models was performed by the analysis of figures of merit. Based on these parameters, it was verified that the best spectral pre-processing was the 1st derivative with window of 5 and 1st order polynomial for the DD. For the MM, the best predictive performance was shown by EMSC. In general, it was verified that the models built using the regression coefficients generated by the Martens' Uncertainty Test (Jack-knife coefficients), presented better predictive performance than the models built with all the spectral variables or with the spectral variables selected by the Successive Projections Algorithm (SPA). The prediction errors obtained for DD and MM were 1.85% and 29.08 KDa, respectively. The error obtained for DD is smaller than the allowed for the reference method. However, for the molar mass, the model did not show satisfactory performance. Therefore, it is clear the viability of the applied methodologies based on NIR spectroscopy and chemometrics to determine the DD in chitosan for biomedical purposes, produced by CERTBIO.

Keywords: Chitosan. Near Infrared Spectroscopy. Chemometrics. SPXY.

LISTA DE ILUSTRAÇÕES

Figura 1 - Estrutura da quitina.....	16
Figura 2 - Estruturas polimórficas da quitina.	17
Figura 3 - Representação esquemática dos biopolímeros quitosana ($n_1 > n_2$) e quitina ($n_2 > n_1$).	19
Figura 4 - Ilustração dos possíveis sítios reacionais da quitina e quitosana.....	19
Figura 5 - Organização dos dados para calibração multivariada.....	31
Figura 6 - Representação genérica do método PLS.....	36
Figura 7 - Exemplo do SPA com $J = 5$, $Mcal = 3$ e $k(0) = 3$. Resultado da primeira projeção: $k(1) = 1$	38
Figura 8 - Fluxograma esquemático das etapas gerais de extração de quitina e obtenção de quitosana.....	47
Figura 9 - Fluxograma das etapas empregadas na construção dos modelos multivariados.	52
Figura 10 - Gráficos da faixa de variação da (a) MM e (b) do GD em relação aos lotes de quitosana produzidos.	54
Figura 11 - Espectros NIR originais das 50 amostras de quitosana, em triplicata e, em detalhe, o espectro médio da triplicata de uma das amostras.....	56
Figura 12 - Gráfico de influência: (a) influência em X e (b) em Y, para o GD; (c) influência em X e (d) em Y, para a MM.	57
Figura 13 - Gráfico de <i>scores</i> para os 150 espectros das amostras de quitosana (a) para os dados brutos, (b) para os dados derivados e (c) para EMSC.....	59
Figura 14 - Variáveis espectrais selecionadas pelo SPA para (a) GD e (b) MM.	66
Figura 15 - Gráfico dos valores preditos pelo modelo D1-J5-POL1-JK versus valores determinados pelo método de referência para o GD.	67
Figura 16 - Gráfico dos valores preditos pelo modelo EMSC-JK versus valores determinados pelo método de referência para a MM.....	67

LISTA DE TABELAS

Tabela 1 - Regiões do espectro infravermelho.	24
Tabela 2 - Principais aplicações da espectroscopia de infravermelho.....	25
Tabela 3 - Etapas do processo padrão de produção da quitosana no CERTBIO e as variações intencionais promovidas na etapa de desacetilação.....	48
Tabela 4 - Valores de referência para o grau de desacetilação (GD) das 50 amostras de quitosana.....	53
Tabela 5 - Valores de referência para a massa molar (MM) das 50 amostras de quitosana.....	54
Tabela 6 – Pré-processamentos espectrais e figuras de mérito para o GD.....	61
Tabela 7 – Pré-processamentos espectrais e figuras de mérito para a MM.....	62
Tabela 8 - Figuras de mérito para a GD empregando-se os algoritmos Jack-Knife e SPA, respectivamente.	63
Tabela 9 - Figuras de mérito para a MM empregando-se os algoritmos Jack-Knife e SPA, respectivamente.	64

LISTA DE ABREVIATURAS E SIGLAS

ASTM	<i>American Society for Testing and Materials</i>
CV	<i>Cross Validation</i>
EMSC	<i>Extended Multiplicative Scatter Correction</i>
FIR	<i>Far-Infrared</i>
FTIR	<i>Fourier Transform Infrared Spectroscopy</i>
GA	<i>Genetic Algorithm</i>
GD	Grau de Desacetilação
iPLS	<i>Interval Partial Least Squares Regression</i>
KS	<i>Kennard-Stone Algorithm</i>
MHS	Mark-Houwink-Sakurada
MIR	<i>Mid-Infrared</i>
MLR	<i>Multiple Linear Regression</i>
MM	Massa Molar
MSC	<i>Multiplicative Scatter Correction</i>
MVC1	<i>Multivariate Calibration 1</i>
NIPALS	<i>Non Linear Iterative Partial Least Squares Algorithm</i>
NIR	<i>Near-infrared</i>
PC	<i>Principal Component</i>
PCA	<i>Principal Component Analysis</i>
PCR	<i>Principal Component Regression</i>
PLS	<i>Partial Least Squares</i>
PLSR	<i>Partial Least Squares Regression</i>
RMSE	<i>Root Mean Square Error</i>
RMSEC	<i>Root Mean Square Error of Calibration</i>
RMSECV	<i>Root Mean Square Error of Cross Validation</i>
RMSEP	<i>Root Mean Square Error of Prediction</i>
RS	<i>Random Search</i>
SNV	<i>Standard Normal Variate</i>
SPA	<i>Sucessive Projections Algorithm</i>
SPXY	<i>Sample set Partitioning based on joint x–y distances</i>
SVD	<i>singular Value Decomposition</i>

SUMÁRIO

1 INTRODUÇÃO	13
2 OBJETIVOS	15
2.1 Objetivo Geral	15
2.2 Objetivos Específicos	15
3 REVISÃO DE LITERATURA	16
3.1 Quitina	16
3.2 Quitosana	18
3.3 Metodologias Convencionais de Caracterização da Quitosana para Fins Biomédicos	21
3.3.1 Massa molar (MM)	22
3.3.2 Grau de desacetilação (GD)	23
3.4 Espectroscopia no Infravermelho	24
3.4.1 Espectroscopia no infravermelho médio (MIR)	25
3.4.2 Espectroscopia no infravermelho próximo (NIR)	27
3.5 Quimiometria	28
3.5.1 Análise multivariada	28
3.5.1.1 Calibração multivariada	29
3.5.1.2 Regressão linear múltipla (MLR)	32
3.5.1.3 Regressão por componentes principais (PCR)	33
3.5.1.4 Análise de componentes principais (PCA)	33
3.5.1.5 Regressão por mínimos quadrados parciais (PLS)	35
3.6 Seleção de Variáveis e Amostras	36
3.6.1 Algoritmos para seleção de variáveis	37
3.6.1.1 Algoritmo das projeções sucessivas (SPA)	37
3.6.1.2 Teste de incerteza de Martens	40
3.6.2 Algoritmos para seleção de amostras	41

3.6.2.1 Algoritmo Kennard-Stone (KS).....	41
3.6.2.2 Algoritmo SPXY	42
3.7 Pré-processamentos de Dados	42
3.8 Cálculo dos Erros	45
3.9 Coeficiente de Correlação	45
4 METODOLOGIA.....	47
4.1 Amostras.....	47
4.2 Determinação dos Parâmetros de Referência	48
4.2.1 Determinação da massa molar (MM)	48
4.2.2 Determinação do grau de desacetilação (GD)	50
4.3 Registros dos Espectros NIR	50
4.4 Construção e Validação dos Modelos	51
4.4.1 Etapas para a construção dos modelos	51
5. RESULTADOS E DISCUSSÃO.....	53
5.1 Amostras e Valores de Referência	53
5.2 Construção dos modelos de calibração multivariada.....	56
6 CONCLUSÕES.....	68
7 PERSPECTIVAS.....	69
REFERÊNCIAS.....	70

1 INTRODUÇÃO

No cenário atual, a quitosana e seus derivados vêm sendo bastante estudados e sua utilização como biomaterial tem sido cada vez mais difundido em diversas áreas como: alimentação, cosméticos, agricultura, proteção ambiental, tratamento de água e efluentes industriais, biomedicina e farmácia. (ANTONINO, 2016; NGO et al., 2015) Seu emprego como biomaterial está diretamente relacionado com alguns de efeitos biológicos característicos, tais como antioxidante, antialérgico, anti-inflamatório, anticoagulante, anticancerígeno e antibacteriano. (NGO et al., 2015) Sendo assim, a quitosana apresenta um enorme potencial para substituir materiais convencionais em muitas aplicações biomédicas e farmacêuticas. Uma vez que os materiais sintéticos típicos, normalmente empregados para tais finalidades, podem ser tóxicos, não renováveis e possuir via de degradação desconhecida, implicando em metabólitos prejudiciais ao meio ambiente. (WEINHOLD et al., 2009)

O termo quitosana não é usado para descrever apenas um produto específico, mas sim uma série de polímeros de quitosana, com diferentes massas molares (50 KDa a 2.000 KDa) e graus de desacetilação maiores que 50%. (BEZERRA, 2011; VEIGA, 2012) Segundo Bezerra (2015), não foi estabelecido ainda um padrão dos parâmetros físicos e químicos para este polímero. Além disso, diferentes condições experimentais podem acarretar em alterações no produto final.

A quitosana disponível no mercado, na maioria das vezes, não possui informações precisas no que diz respeito às suas características (BEZERRA, 2015; RAAFAT; SAHL, 2009) e apresenta vários: graus de pureza, distribuições de massa molar, comprimentos de cadeia, graus de desacetilação, densidades e distribuições de cargas, formas de sais, viscosidades e valores de retenção de água. Essa variabilidade afeta diretamente as suas propriedades físico-químicas, refletindo consideravelmente em quase todas as suas aplicações. (RAAFAT; SAHL, 2009)

Para que uma matéria-prima possa ser empregada na confecção de produtos biomédicos, faz-se necessário a completa caracterização da mesma. A fim de atender tanto as boas práticas de fabricação, como também, garantir principalmente a segurança dos seus futuros usuários por meio do oferecimento de um produto de qualidade. Para atender a tais exigências, Antonino (2016) realizou estudos na quitosana produzida pelo Laboratório de Avaliação e Desenvolvimento de Biomateriais do Nordeste - CERTBIO/UFCG, caracterizando-a quanto à massa molar e ao grau de desacetilação. Em que a determinação do

primeiro parâmetro se deu por meio da viscosimetria intrínseca e a do segundo baseou-se nas técnicas de espectroscopia de infravermelho com transformada de Fourier (FTIR) e espectroscopia UV-Visível.

Para a técnica de FTIR foram empregadas três metodologias (BRUGNEROTTO et al., 2001; DOMSZY; ROBERTS, 1985; ROBERTS, 1992). Sendo indicada como metodologia padrão a metodologia proposta por Brugnerotto et al. (2001). Com relação à espectroscopia UV-Visível, foi utilizada a metodologia da primeira derivada proposta por Muzzarelli e Rocchetti (1985), pela qual foram obtidos resultados satisfatórios e bem próximos da metodologia padrão escolhida para a FTIR. (ANTONINO, 2016) No entanto, o emprego desta última técnica requer um pouco mais de trabalho devido à preparação de soluções padrão para a construção de curvas de calibração e solubilização de amostras. Desse modo, a metodologia preconizada por Brugnerotto et al. (2001) é mais indicada para a determinação do grau de desacetilação em análise de rotina. Porém, apesar dessa técnica ser uma ferramenta laboratorial bastante versátil e sensível, a mesma ainda apresenta algumas dificuldades na caracterização do grau de desacetilação da quitosana, como a correta identificação e delimitação das linhas de base utilizadas no cálculo das áreas integradas dos picos centrais característicos dos grupamentos amida III (1320 cm^{-1}) e CH_2 (1420 cm^{-1}).

Assim como ocorre com a metodologia UV-Visível, a determinação da massa molar pela viscosimetria demanda bastante tempo na solubilização das amostras de quitosana, preparação de padrões característicos e construção de curvas de calibração.

Uma alternativa viável, rápida, bastante prática, eficiente e não destrutiva para a determinação da massa molar viscosimétrica e do grau de desacetilação na produção da quitosana, seria a utilização da espectroscopia de infravermelho próximo (NIR) aliada a técnicas quimiométricas de calibração multivariada. Neste sentido, a técnica de regressão por mínimos quadrados parciais (PLS), combinada a NIR, tem se mostrado uma excelente alternativa para determinação de parâmetros de qualidade.

No caso da quitosana, estudos recentes realizados por Dimzon (2015) mostraram que a determinação do grau de desacetilação da quitosana (75% a 95%) por espectroscopia de infravermelho médio (MIR) em conjunto com PLS, apresentou resultados comparáveis aos obtidos por titulação potenciométrica que é uma metodologia de referência para a determinação deste parâmetro. Portanto, estes resultados sugerem que a combinação NIR-PLS seja uma metodologia altamente viável para monitorar a qualidade de lotes de quitosana produzidos em escala industrial.

2 OBJETIVOS

2.1 Objetivo Geral

Desenvolver uma metodologia baseada na espectroscopia no infravermelho próximo (NIR) e calibração multivariada destinada à determinação da massa molar e do grau de desacetilação da quitosana, que possa ser facilmente implementada em análises de rotina.

2.2 Objetivos Específicos

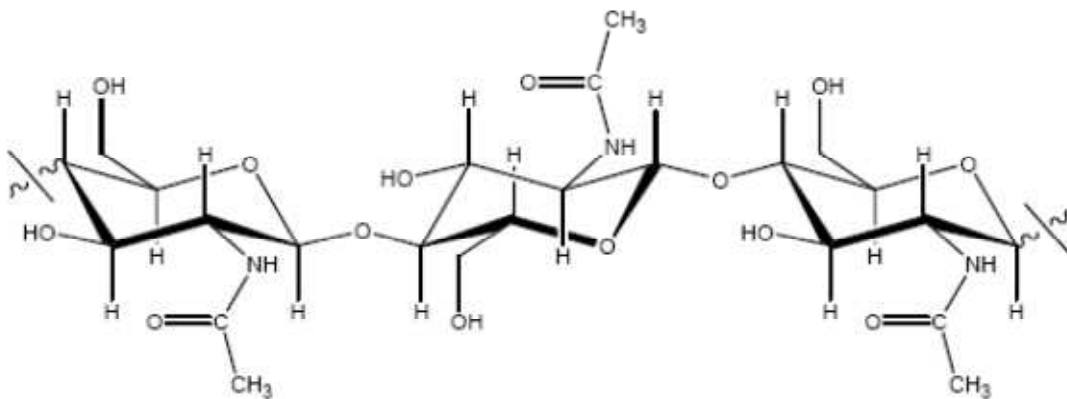
- a) Ampliar a faixa de variação da massa molar e do grau de desacetilação, coletando amostras de quitosana durante a produção em cinco tempos distintos;
- b) Comparar o desempenho dos modelos PLS em função do pré-processamento espectral empregado;
- c) Desenvolver modelos de calibração multivariada utilizando PLS e MLR para a determinação da massa molar e do grau de desacetilação da quitosana.
- d) Testar o desempenho dos algoritmos de seleção de variáveis (Jack-Knife e SPA) para seleção das variáveis espectrais mais informativas nos modelos de calibração multivariada;
- e) Comparar os desempenhos preditivos dos modelos PLS e MLR construídos.

3 REVISÃO DE LITERATURA

3.1 Quitina

A Quitina é um polissacarídeo de cadeia linear que ocorre naturalmente e encontra-se amplamente distribuído ao longo de todo planeta. É constituído por resíduos de 2-acetoamido-2-desoxi-D-glicopirranose (N-acetil-D-glucosamina – GlcNAc), unidos por ligações β (1 \rightarrow 4) e possui massas moleculares variáveis, cuja estrutura está representada na Figura 1. (ANTONINO, 2016; AZUMA et al., 2015; FONSECA, 2016; PRASHANTH; THARANATHAN, 2007; OLIVEIRA, 2011; SENRA, 2015)

Figura 1 - Estrutura da quitina.



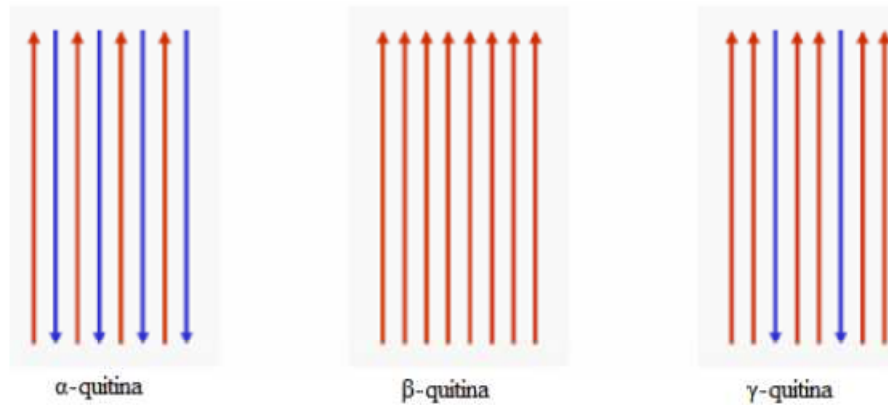
Fonte: ANTONINO, 2007.

Esse biopolímero ocorre naturalmente em exoesqueletos de muitos invertebrados, como crustáceos, moluscos, insetos e aracnídeos. (OLIVEIRA, 2011; SENRA, 2015) Sua presença ainda tem sido observada em algas marinhas e esponjas (poríferas). (OLIVEIRA, 2011) Apesar da ocorrência natural, a quitina não costuma aparecer isoladamente nessas fontes, apresentando-se complexada com outras macromoléculas, como proteínas e até mesmo compostos inorgânicos como carbonato de cálcio. Na literatura apenas dois tipos de diatomáceas, *Thalassiosira fluviatilis* e *Cyclotella cryptica*, apresentam a quitina na sua forma isolada. (ANTONINO, 2016)

De acordo com a função que desempenha, nos diferentes organismos de origem, a quitina pode ocorrer em três arranjos estruturais distintos (Figura 2), os quais correspondem às variedades polimórficas (α , β e γ). (ANTONINO, 2016; BERGER, 2013; DENARI, 2014; FACCHINATTO, 2016; OLIVEIRA, 2011; PEPINO, 2016) A forma α é responsável por

atribuir extrema resistência mecânica a determinadas partes dos organismos, geralmente é encontrada em cutículas de artrópodes, sendo associada com proteínas e/ou materiais inorgânicos. As formas β e γ , além de resistência mecânica proporcionam também flexibilidade. (BERGER, 2013)

Figura 2 - Estruturas polimórficas da quitina.



Fonte: ANTONINO, 2007.

Dentre as três estruturas polimórficas, a α -quitina (estrutura ortorrômbica) é a mais abundante apresentando 92% de cristalinidade contra 72% da β -quitina (estrutura monoclinica). Uma análise detalhada da terceira estrutura polimórfica (γ -quitina) indica que esta pode ser considerada apenas uma variação entre as formas α e β ou até mesmo uma combinação destes arranjos. Apesar de sua ocorrência natural muito rara, pode ser encontrada em alguns tipos de fungos. (FACCHINATTO, 2016; OLIVEIRA, 2011)

A grande abundância da α -quitina e as suas fontes naturais bastante acessíveis (carapaças de caranguejos e camarões), contribuem também para que esta seja a forma polimorfa mais estudada. Seu arranjo antiparalelo de cadeias favorece a formação de um complexo sistema de ligações de hidrogênio, envolvendo cadeias poliméricas adjacentes. Esse tipo de configuração tem sido considerado o arranjo polimorfo mais estável, uma vez que a conversão dos arranjos beta e gama no tipo alfa aparentam ser irreversíveis. (OLIVEIRA, 2011)

No caso da β -quitina, as cadeias pertencentes a diferentes lamelas são dispostas paralelamente, dificultando a formação de ligações intermoleculares de hidrogênio entre cadeias de lamelas adjacentes. Isto implica em um empacotamento menos denso do material. (BERGER, 2013; BEZERRA, 2015; OLIVEIRA, 2011)

Na γ -quitina, as cadeias de duas lamelas em configuração paralela são intercaladas por lamelas dispostas em cadeias antiparalelas. Esse tipo de configuração também é considerado na literatura como uma distorção das duas estruturas anteriores e, ao contrário do que ocorre com a estrutura alfa, é a menos estudada. (BERGER, 2013; BEZERRA, 2015; OLIVEIRA, 2011)

Do ponto de vista químico, a afinidade da quitina frente a solventes é extremamente limitada, dificultando, por exemplo, a formação de soluções aquosas. A literatura reporta a solubilidade deste biopolímero em solventes fortes como HCl e H₂SO₄ em elevadas concentrações. Porém, a sua massa molar diminui a partir da dissolução nesses solventes, principalmente pelos efeitos de degradação ocorridos na cadeia polimérica.

Essa dificuldade de solubilização da quitina, conseqüentemente, reduz sua aplicabilidade em pesquisas biomédicas. Diante deste problema, um dos maiores interesses em relação a este biopolímero é a obtenção de seu derivado desacetilado denominado quitosana, o qual se destaca em relação ao seu precursor por possuir maior reatividade e melhor solubilidade, aumentando assim o seu potencial em modificações e processamento. (ANTONINO, 2016; PEPINO, 2016)

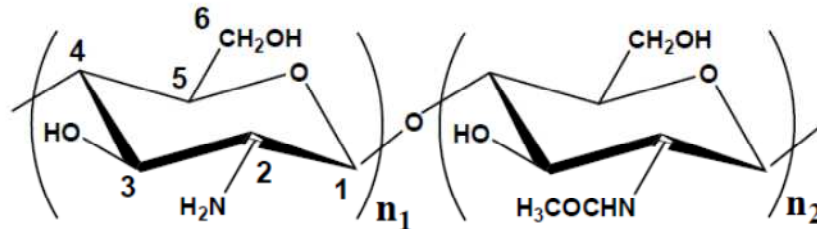
3.2 Quitosana

Quitosana é nome dado ao heteropolissacarídeo derivado do processo de desacetilação química ou enzimática da quitina. Em termos de estrutura química, tanto a quitina como a quitosana são formadas por duas unidades monoméricas distintas (2-acetamido-2-desoxi-D-glicopirranose e 2-amino-2-desoxi-D-glicopirranose) que correspondem às unidades acetiladas e desacetiladas, respectivamente. Na formação das cadeias poliméricas, tais unidades monoméricas são interligadas por meio de ligações glicosídicas do tipo β (1 \rightarrow 4) e distribuídas de forma aleatória (distribuição de Bernoulli). (ANTONINO, 2016; BERGER, 2013; BEZERRA, 2011; MENDES, 2016; OLIVEIRA, 2011; PENICHE; BECHERÁN, 2015) Essa distribuição aleatória e a coexistência de unidades acetiladas e desacetiladas, fazem com que ambos os biopolímeros sejam classificados como copolímeros estatísticos. (OLIVEIRA, 2011)

A grande similaridade entre as estruturas químicas desses biopolímeros os caracteriza como membros de uma mesma família, dentre os quais a quitosana é aquele que apresenta maior percentagem de unidades desacetiladas (Figura 3). Em outras palavras o termo

quitosana é a denominação genérica atribuída a uma série de quitinas parcialmente desacetiladas.

Figura 3 - Representação esquemática dos biopolímeros quitosana ($n_1 > n_2$) e quitina ($n_2 > n_1$).

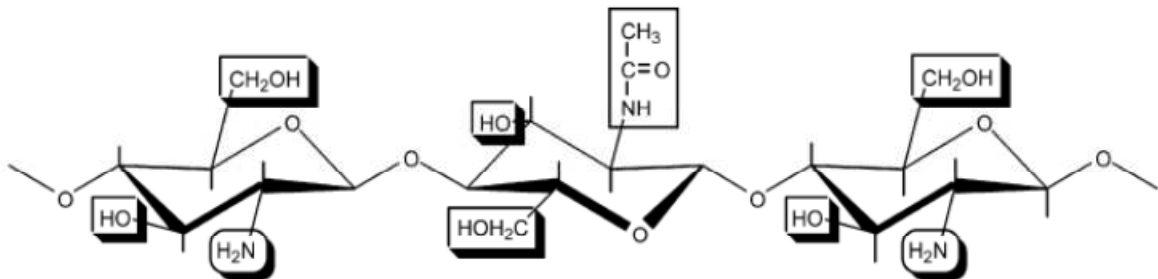


Fonte: OLIVEIRA, 2011.

A quitosana apresenta uma série de características importantes, tais como: excelente biocompatibilidade, quase nenhuma toxicidade, elevada bioatividade, biodegradabilidade, reatividade do grupo amino desacetilado, permeabilidade seletiva, ação polieletrólítica, atividade antimicrobiana, habilidade em formar gel e filme, habilidade de quelação e capacidade adsortiva. (BERGER, 2013)

A quitina e seu derivado quitosana são muito interessantes do ponto de vista da funcionalidade do grupamento amino, que pode ser adequadamente modificado para conferir propriedades desejadas e funções biológicas distintas para ambas, incluindo a solubilidade. Além desse grupamento, as duas hidroxilas existentes nas estruturas também podem sofrer modificações químicas e contribuir para o aumento da solubilidade. Os possíveis sítios reacionais para essas biomoléculas estão ilustrados na Figura 4. Através dos quais, estas podem sofrer reações de eterificação, esterificação, reticulação, copolimerização de enxerto, entre outras. (PILLAI; PAUL; SHARMA, 2009)

Figura 4 - Ilustração dos possíveis sítios reacionais da quitina e quitosana.



Fonte: PILLAI; PAUL; SHARMA, 2009.

Diferenças apresentadas na morfologia semicristalina da quitina denotam que as quitosanas obtidas em uma reação de estado sólido têm uma distribuição heterogênea de grupos N-acetil ao longo das cadeias moleculares. Outras diferenças no processo de produção, como por exemplo, temperatura, concentração alcalina, razão entre soluções alcalinas e matérias-primas, também podem contribuir para que a composição final do produto resulte em uma mistura de quitosanas, variando em massa molar e grau de N-acetilação. Tais misturas podem inclusive conter impurezas como metais pesados, resíduos proteicos e resíduos ácidos/alcalinos. (KUMIRSKA et al., 2010; WEINHOLD et al., 2009) Desse modo, sugere-se que a quitosana seja caracterizada de acordo com suas propriedades intrínsecas como: pureza, massa molar, viscosidade e grau de desacetilação, sendo esses fatores determinantes nas características finais do produto. Como também, em aplicações posteriores que utilizem esse biomaterial. (BEZERRA, 2011)

Dentre as propriedades citadas anteriormente, a massa molar e o grau de desacetilação são as propriedades físico-químicas mais importantes da quitosana, uma vez que esses parâmetros determinam a qualidade da quitosana em suas diversas aplicações. (GAMAL et al., 2016) A primeira tem uma influência direta nas propriedades visco elásticas de suas soluções e formas coloidais hidratadas. Já a segunda influencia praticamente em todas as suas propriedades funcionais, a exemplo da solubilidade, nível de intumescimento em água, susceptibilidade à biodegradação, bioatividade e biocompatibilidade. (LIZARDI-MENDOZA; MONAL, 2016)

Com base na massa molar, a quitosana recebe as seguintes classificações: massa molar baixa (50 - 190 KDa), massa molar média (190 - 310 KDa) e massa molar alta (>310 KDa). (SIGMA-ALDRICH, 2017) Já o grau de acetilação representa a fração ou percentual de unidades glicosídicas N-acetiladas na quitina ou quitosana. Sua representação se dá por meio do termo FA (quando nos referimos à fração mássica) ou DA (quando o representamos em termos percentuais). Comumente o termo DA é substituído por sua forma complementar (GD) que é conhecido como grau de desacetilação e representa o percentual de grupos N-acetil da cadeia polimérica que foram convertidos em grupos amino primários. De forma simplificada, esse parâmetro pode ser representado como $(1 - FA)$ ou $(GD = 100 - DA)$, caso seja escolhida a representação percentual. (PENICHE; BECHERÁN, 2015; PIRES, 2010)

A obtenção de quitosana a nível industrial é feita por meio da desacetilação parcial ou completa da quitina nos estados sólido (processo heterogêneo) e dissolvido (processo homogêneo) em condições alcalinas ou por hidrólise enzimática (usando quitina desacetilase).

(KUMIRSKA et al., 2010) A variação da fonte e processamento para obtenção da quitosana implica em grandes diferenças na qualidade e propriedades dos produtos. Sendo assim, comercialmente esse biopolímero pode ser encontrado com vários graus de pureza, massa molar e grau de desacetilação. (ANTONINO, 2016)

3.3 Metodologias Convencionais de Caracterização da Quitosana para Fins Biomédicos

Para garantir que materiais destinados às aplicações biomédicas satisfaçam os requisitos de segurança e qualidade, vários parâmetros devem ser levados em consideração. No caso da quitosana, esses incluem sua própria identidade como substância, caracterização e testes físico-químicos, perfil de impurezas e também testes relacionados ao seu desempenho. (ASTM F2103-11, 2011)

O conhecimento das propriedades físicas e químicas da quitosana, tais como: o grau de desacetilação, a massa molar (ou viscosidade), a concentração, e assim por diante, ajudarão na escolha do produto que melhor se adequa às necessidades dos consumidores finais. Conhecer esses parâmetros garante ainda aos seus usuários a obtenção de materiais similares aos que já estão sendo utilizados, junto aos fornecedores, auxiliando também na documentação de sua formulação, enquanto matéria-prima, ou na documentação de dispositivos derivados. Finalmente, a caracterização da quitosana permitirá que a funcionalidade desta matéria-prima se adequa à aplicação ou ao produto final. (ASTM F2103-11, 2011)

De acordo com a Norma ASTM F2103-11 (2011), que estabelece métodos para caracterização e teste de sais de quitosana como matérias primas para uso em aplicações biomédicas e em produtos médicos para engenharia de tecidos, a lista dos principais parâmetros de caracterização que estão diretamente relacionados à funcionalidade da quitosana ou de seus sais e que servem para avaliar a consistência dos mesmos, engloba os seguintes parâmetros:

- a) Identidade (quitosana/sais de quitosana);
- b) Composição e estrutura sequencial;
- c) Grau de desacetilação;
- d) Massa molar;
- e) Polidispersividade;
- f) Viscosidade em solução aquosa;
- g) Conteúdo de matéria seca;

- h) Teor de cinzas;
- i) Teor de insolúveis;
- j) Perfil de impurezas (conteúdo de endotoxinas, conteúdo de proteínas, conteúdo de metais pesados e segurança microbiológica);
- k) Biocompatibilidade.

Neste trabalho, será dada ênfase apenas ao grau de desacetilação e a massa molar, pois esses apresentam efeitos marcantes sobre as propriedades funcionais da quitosana, desde as amostras em estado sólido até suas soluções. (LIZARDI-MENDOZA; MONAL, 2016) Tais parâmetros influenciam tanto nas propriedades físico-químicas quanto nas biológicas como solubilidade, estabilidade, biodegradabilidade e baixa toxicidade. (ANTONINO, 2016)

3.3.1 Massa molar (MM)

A Massa molar (MM) da quitosana definirá certas características de desempenho, a exemplo da viscosidade. Além disso, a forma como esse parâmetro encontra-se distribuído na matriz polimérica (polidispersividade), acaba refletindo também em suas propriedades físicas, químicas e biológicas. Como exemplo, propriedades mecânicas de hidrogéis, porosidade de membranas, arcabouços (scaffolds), microcápsulas, atividade antimicrobiana e cicatrizante, entre outras. Desse modo, esse parâmetro afeta diretamente o emprego da quitosana em diversas aplicações como biomaterial. (ASTM F2103-11, 2011; PENICHE; BECHERÁN, 2015)

Devido à natureza polidispersa das quitosanas comerciais em relação à MM, esse parâmetro pode ser expresso como média numérica (M_N) ou média ponderada (M_W). Sendo assim, os principais métodos empregados na determinação da MM para a quitosana são os mesmos empregados para qualquer polímero, inclusive para o seu precursor quitina. Tais métodos são: a dispersão da luz, a cromatografia de permeação em gel, a osmometria, a ultracentrifugação por equilíbrio de sedimentação e, principalmente, a viscosimetria. (ASTM F2103-11, 2011; PENICHE; BECHERÁN, 2015)

A determinação da MM por viscosimetria é feita com base na viscosidade intrínseca (η), a qual descreve a capacidade que um polímero possui em formar soluções viscosas na água, sendo diretamente proporcional à massa molar média do polímero em questão. A viscosidade intrínseca é uma característica do polímero sob condições especificadas de

solvente e temperatura, não dependendo da concentração. Esse tipo de viscosidade está diretamente relacionado com a MM de um polímero através da equação de Mark-Houwink-Sakurada (MHS), Equação 1 (ASTM F2103-11, 2011):

$$n = K \cdot M_w^\alpha \quad \text{Equação (1)}$$

Em que:

n = Viscosidade intrínseca;

K = Constante viscosimétrica;

α = Constante empírica que descreve a conformação do polímero;

M_w = Massa molar viscosimétrica média.

Uma vez conhecida n e de posse dos valores das constantes K e α , determina-se então M_w para uma dada amostra, por meio da manipulação logarítmica da Equação 1: $\log [\eta] = \log K + \alpha (\log M_w)$. (ASTM F2103-11, 2011)

Experimentalmente n é determinada medindo-se a viscosidade relativa num viscosímetro capilar de Ubbelohde. As medições devem ser realizadas num solvente contendo NaCl 0,1 mol.L⁻¹ (um sal monovalente não gelificante) a uma temperatura constante de 20 °C e a uma concentração suficientemente baixa de quitosana. Sempre que possível, a operação e aquisição de dados automáticos são preferidos. (ASTM F2103-11, 2011)

3.3.2 Grau de desacetilação (GD)

Existem várias metodologias disponíveis para a determinação do grau de desacetilação da quitosana (GD). A Norma ASTM F2103-11 (2011) traz como metodologias padrão a Espectroscopia de Ressonância Magnética Nuclear de alta resolução de ¹H e ¹³C (RMN) e a Espectroscopia UV. Sendo essa última fundamentada no estudo proposto por Muzzarelli et al. (1985) que quantifica o número de grupos funcionais amina existentes no biopolímero. Para isto, o método utiliza uma curva analítica, construída a partir de concentrações variáveis de N-acetilglucosamina. O GD é calculado a partir de registros da primeira derivada dos espectros UV de N-acetilglucosamina e amostras de quitosana a 202 nm. Apesar desta norma de referência apresentar em seu texto apenas essas duas metodologias, o próprio documento menciona ainda que a determinação do GD da quitosana não deva ficar restrito às

metodologias apresentadas ao longo do mesmo. A norma sugere a utilização de outras técnicas que melhor se adequem a cada laboratório ou controle de qualidade na indústria.

Como técnicas alternativas para a determinação do GD, Peniche e Becherán (2015) citam, em seu estudo, a potenciometria, a análise elementar, a análise térmica, a cromatografia de permeação em gel, o dicroísmo circular e a espectroscopia no infravermelho, proposta por Brugnerotto et al. (2001) Esta última, foi comparada à metodologia de espectroscopia UV-Visível proposta por Muzzarelli e Rocchetti (1985), a qual foi utilizada no trabalho desenvolvido por Antonino (2016) para a caracterização das amostras de quitosana produzidas no CERTBIO. No entanto, a maior praticidade da metodologia proposta por Brugnerotto et al. (2001), fez com que a mesma viesse a ser adotada como a metodologia padrão por este laboratório.

3.4 Espectroscopia no Infravermelho

A região do infravermelho do espectro eletromagnético compreende a radiação com números de onda que varia entre (12.800 e 10 cm^{-1}) ou comprimentos de onda de ($0,78$ a $1.000\text{ }\mu\text{m}$). Tanto do ponto de vista das aplicações como dos instrumentos, é conveniente dividir a região de infravermelho em três regiões, a saber, infravermelho próximo (do Inglês, *Near-Infrared* – NIR), médio (do Inglês, *Mid-Infrared* – MIR) e distante (do Inglês, *Far-Infrared* – FIR) (SKOOG; HOLLER; CROUCH, 2009), conforme Tabela 1.

Tabela 1 - Regiões do espectro infravermelho.

Região do Infravermelho	Comprimento de onda (λ), μm	Número de onda (ν), cm^{-1}	Frequências (ν), Hz
Próximo	0,78 a 2,5	12.800 a 4.000	$3,8 \times 10^{14}$ a $1,2 \times 10^{14}$
Médio	2,5 a 50	4.000 a 200	$1,2 \times 10^{14}$ a $6,0 \times 10^{12}$
Distante	50 a 1.000	200 a 10	$6,0 \times 10^{12}$ a $3,0 \times 10^{11}$
A mais utilizada	2,5 a 15	4.000 a 670	$6,0 \times 10^{12}$ a $2,0 \times 10^{13}$

Fonte: SKOOG; HOLLER; CROUCH, 2009.

Dentre essas regiões espectrais, a mais utilizada é a região do MIR, principalmente a faixa que se estende de cerca de 670 até 4.000 cm^{-1} ($2,5$ e $14,9\text{ }\mu\text{m}$). Essa faixa é caracterizada por utilizar espectros de absorção, reflexão e emissão para determinações qualitativas e quantitativas de espécies orgânicas e bioquímicas. A região NIR entre 4.000 e 14.000 cm^{-1} ($0,75$ e $2,5\text{ }\mu\text{m}$), também é muito útil para a determinação quantitativa rotineira

de espécies, tais como água, dióxido de carbono, enxofre, hidrocarbonetos de massas molares baixas, nitrogênio e aminas, além de muitos outros compostos simples de interesse na agricultura e na indústria. Muitas vezes, essas determinações são baseadas em medições da reflectância difusa de amostras sólidas ou líquidas, sem tratamento prévio, ou em estudos de absorção de gases. Já a região FIR (15 a 1.000 μm), tem sua principal utilização voltada à determinação da estrutura de espécies inorgânicas e organometálicas com base em medições de absorção. (SKOOG; HOLLER; CROUCH, 2009) Para uma melhor compreensão do assunto, são apresentadas na Tabela 2 as aplicações da espectrometria no infravermelho de acordo com cada região espectral.

Tabela 2 - Principais aplicações da espectroscopia de infravermelho.

Regiões	Tipo de medição	Tipo de análise	Tipo de amostra
NIR	Reflectância difusa	Quantitativa	Materiais comerciais sólidos ou líquidos
	Absorção	Quantitativa	Misturas gasosas
MIR	Absorção	Qualitativa	Sólidos, líquidos ou gases puros
		Quantitativa	Misturas complexas de gases, líquidos ou sólidos
		Cromatográfica	Misturas complexas de gases, líquidos ou sólidos
	Reflectância	Quantitativa	Sólidos ou líquidos puros
	Emissão	Quantitativa	Amostras atmosféricas
FIR	Absorção	Quantitativa	Espécies inorgânicas puras ou organometálicas

Fonte: SKOOG; HOLLER; CROUCH, 2009.

3.4.1 Espectroscopia no infravermelho médio (MIR)

O espectro no infravermelho médio (4.000 – 200 cm^{-1}) pode ser dividido em aproximadamente quatro regiões e a natureza de uma frequência de grupo pode ser determinada pela região em que está localizada. As regiões são generalizadas da seguinte forma (SKOOG; HOLLER; CROUCH, 2009; STUART, 2004):

- Região de estiramento X-H (4.000 – 2.500 cm^{-1});
- Região de ligação tripla (2.500 – 2.000 cm^{-1});
- Região de ligação dupla (2.000 – 1.500 cm^{-1});
- Região de impressão digital (1.500 – 600 cm^{-1}).

As vibrações fundamentais na região (4.000 – 2.500 cm^{-1}) são devidas aos estiramentos O–H, C–H e N–H. O estiramento O–H produz uma banda larga que ocorre na

faixa de $(3.700 - 3600 \text{ cm}^{-1})$. Já o estiramento de N–H é usualmente observado entre $(3.400 \text{ e } 3.300 \text{ cm}^{-1})$. Esta absorção é muito mais nítida do que o estiramento O–H e pode, por conseguinte, ser diferenciada. As bandas de estiramento C–H de compostos alifáticos ocorrem no intervalo de $(3.000 - 2.850 \text{ cm}^{-1})$. Se a ligação C–H for adjacente a uma ligação dupla ou anel aromático, o número de onda de seu estiramento aumenta, absorvendo entre $(3.100 \text{ e } 3.000 \text{ cm}^{-1})$. (STUART, 2004)

As absorções de estiramento de ligação tripla situam-se na região das constantes de força altas das ligações de $(2.500 - 2.000 \text{ cm}^{-1})$. As ligações C≡C absorvem entre $(2.300 \text{ e } 2.050 \text{ cm}^{-1})$, enquanto que o grupo nitrilo (C≡N) ocorre entre $(2.300 \text{ e } 2.200 \text{ cm}^{-1})$. Esses grupos podem ser facilmente diferenciados, pois o estiramento da ligação C≡C é normalmente menos intenso que o estiramento C≡N, considerado de intensidade média. Essas são as absorções mais comuns nesta região, mas podem-se encontrar algumas absorções de estiramento X–H, onde X é um átomo mais massivo, como fósforo ou silício. Neste caso, as absorções ocorrem geralmente perto de $(2.400 \text{ e } 2.200 \text{ cm}^{-1})$, respectivamente. (STUART, 2004)

As bandas principais na região $(2.000 - 1.500 \text{ cm}^{-1})$ são consequência dos estiramentos C=C e C=O, sendo o estiramento carbonílico uma das absorções mais fáceis de reconhecer num espectro infravermelho. Essa banda geralmente se destaca como sendo a banda mais intensa no espectro, e dependendo do tipo de ligação C=O, ocorre na região de $(1.830 - 1.650 \text{ cm}^{-1})$. Observa-se também que os carbonílicos metálicos podem absorver acima de 2.000 cm^{-1} . O estiramento da ligação C=C é muito fraco e ocorre em torno de 1.650 cm^{-1} , mas essa banda é frequentemente ausente por motivos de simetria ou momento dipolar. Nesta região também ocorre o estiramento do tipo C=N, porém, com intensidade mais forte que o estiramento C=C. (STUART, 2004)

Assumiu-se até agora que cada banda num espectro infravermelho pode ser atribuída a uma deformação particular da molécula, o movimento de um grupo de átomos, ou a flexão ou estiramento de uma ligação particular. Isto é possível para muitas bandas, particularmente as vibrações de estiramento de ligações múltiplas que são "bem comportadas". No entanto, muitas vibrações não são tão "bem comportadas" e podem variar por centenas de números de onda, mesmo para moléculas semelhantes. Isto se aplica à maioria das vibrações de estiramento e flexão, que absorvem na região $(1.500 - 650 \text{ cm}^{-1})$, para a qual pequenos efeitos estéticos ou eletrônicos na molécula levam a grandes mudanças. Um espectro de uma molécula pode ter uma centena ou mais bandas de absorção presentes, sendo considerado

como uma "impressão digital" da molécula. Daí o nome dessa região específica do infravermelho médio: região de impressão digital. (STUART, 2004)

3.4.2 Espectroscopia no infravermelho próximo (NIR)

Inicialmente a espectroscopia NIR não foi considerada uma região de interesse analítico, sendo considerada uma extensão da região do visível que não foi explorada até 1970. Os primeiros trabalhos que despertaram interesse na região do infravermelho próximo como ferramenta de análise na indústria, foram desenvolvidos na década de setenta pelo professor Karl Norris, quando este era responsável por um grupo de pesquisa do Departamento de Agricultura dos Estados Unidos. Entretanto, as limitações tecnológicas da época não permitiram o uso mais abrangente desta técnica, como atualmente vem sendo observado em diversos setores industriais. (GRANGEIRO JÚNIOR, 2011)

As absorções observadas na região do infravermelho próximo ($13.000 - 4.000 \text{ cm}^{-1}$) são resultados de sobretons ou combinações dos estiramentos de bandas fundamentais que ocorrem na região de ($3.000 - 1.700 \text{ cm}^{-1}$). Essas combinações de bandas ocorrem normalmente devido aos estiramentos das ligações C–H, N–H ou O–H. Diferente do que ocorre no infravermelho médio, essas bandas são de intensidade fraca, diminuindo aproximadamente 10 vezes entre um sobretom e outro. Na maioria das vezes essas bandas também aparecem sobrepostas, o que as torna menos útil para a análise qualitativa em relação à região do infravermelho médio. No entanto, existem diferenças importantes entre as posições de grupos funcionais no infravermelho próximo que podem ser exploradas para análise quantitativa. (STUART, 2004)

Vários trabalhos científicos vêm sendo desenvolvidos visando o estudo e aplicação da espectroscopia NIR e esta técnica tem ganhado aceitação pela indústria farmacêutica como meio de identificação, quantificação e controle de parâmetros físicos e químicos, desde a matéria-prima até o produto final. (SIMÕES, 2008)

Os instrumentos atualmente empregados na espectroscopia NIR fornecem rapidamente grandes quantidades de dados que exigem processamentos rápidos e eficientes na extração de informações analíticas úteis. Tal exigência torna necessária a utilização de outras técnicas, como por exemplo, as técnicas quimiométricas, que possam trabalhar em conjunto com a NIR potencializando os seus resultados. (GOODARZI et al., 2015)

3.5 Quimiometria

A quimiometria é a área da química que utiliza métodos matemáticos e estatísticos para o tratamento de dados químicos, de forma a extrair maior quantidade de informações e melhores resultados analíticos. (SIMÕES, 2008) Desde o seu surgimento no final dos anos 60 até os dias atuais, foram desenvolvidos muitos métodos que tornaram possível o processamento e interpretação de dados que antes seriam impossíveis de serem analisados. Um exemplo que afirma o sucesso da utilização da quimiometria são as análises realizadas na região NIR, as quais, sem a utilização de modelos de calibração multivariada não possibilitariam determinações quantitativas. (GRANGEIRO JÚNIOR, 2011; SIMÕES, 2008)

A aplicação das técnicas quimiométricas tem se difundido nas mais diversas áreas da Química e ciências afins. O desenvolvimento de novos instrumentos e emprego dos mesmos em laboratórios de pesquisas passou a fornecer uma grande quantidade de dados, o que criou a necessidade de aplicação de ferramentas estatísticas sofisticadas que pudessem ser utilizadas na análise dos mesmos. Esta necessidade, aliada a uma facilidade crescente de acesso a microcomputadores, possibilitou o crescimento da quimiometria. (CABRAL, 2011; GRANGEIRO JÚNIOR, 2011; SIMÕES, 2008)

Com o passar dos anos, a disseminação da quimiometria fez com que a mesma não ficasse restrita apenas a laboratórios de pesquisa, sendo amplamente utilizada em ambientes industriais. Nesse sentido, pode-se destacar a indústria química, em que a associação de ferramentas quimiométricas com técnicas instrumentais, principalmente as espectrométricas e cromatográficas, são empregadas no desenvolvimento de métodos analíticos para controle de qualidade de matérias-primas e controle de processos. (SIMÕES, 2008)

A quimiometria pode ser considerada ainda como um dos mais recentes campos da Química Analítica. (SIMÕES, 2008) Sendo subdividida em outras áreas principais como: processamento de sinais analíticos, planejamento e otimização de experimentos, reconhecimento e classificação de padrões, calibração multivariada, monitoramento e modelagem de processos multivariados e métodos de inteligência artificial.

3.5.1 Análise multivariada

A necessidade de entender as relações entre muitas variáveis faz da análise multivariada uma técnica bastante útil, apesar de complexa. Tendo em vista que a matemática

necessária para fazer inferências de técnicas é mais difícil do que em técnicas univariadas. (FONSECA, 2011)

Os objetivos de investigações científicas que naturalmente levam a métodos multivariados são (FONSECA, 2011):

- a) Redução de dados ou simplificação estrutural;
- b) Triagem e agrupamento;
- c) Investigação da dependência entre variáveis;
- d) Predição;
- e) Construção e teste de hipóteses.

3.5.1.1 Calibração multivariada

A calibração pode ser definida como sendo uma série de operações que estabelecem por meio de condições específicas uma relação entre medidas instrumentais e valores para uma ou mais propriedades de interesse correspondente. (FONSECA, 2011)

Dentre os métodos de calibração existentes, certamente os mais difundidos são ainda os métodos de calibração univariada. Os quais são conhecidos também como “calibração de ordem zero”, ou seja, tem-se apenas uma medida instrumental para cada uma das amostras de calibração. Em outras palavras, para cada amostra tem-se apenas um escalar. Contudo, a aplicação desse tipo de calibração é restrita, pois quando a amostra não é livre de interferentes e sua análise é realizada diretamente por uma metodologia instrumental, provavelmente provocará desvios na determinação da propriedade de interesse e a aplicação deste método de calibração torna-se inviável. (FONSECA, 2011) Sendo assim, uma alternativa viável para contornar este problema é utilização da calibração multivariada. Esse tipo de calibração consiste na construção de modelos matemáticos para relacionar dados de saída de um instrumento (medidas instrumentais multivariadas) às concentrações dos analitos ou a outras propriedades das amostras. (SIMÕES, 2008)

O processo de calibração multivariada é constituído basicamente de três fases: calibração, validação e predição. Na fase de calibração são construídos os modelos matemáticos e na fase de validação esses modelos são avaliados usando-se ferramentas estatísticas de diagnóstico. Na fase de predição, os modelos matemáticos validados são aplicados para prever as concentrações ou propriedades de amostras desconhecidas. (GRANGEIRO JÚNIOR, 2011; SIMÕES, 2008)

Um modelo de calibração, na verdade, é uma função matemática (f) que relaciona dois grupos de variáveis, uma delas denominada dependente, aqui representada por um (vetor \mathbf{y}) e o outro grupo das variáveis independentes a (matriz \mathbf{X}) (FONSECA, 2011; GRANGEIRO JÚNIOR, 2011):

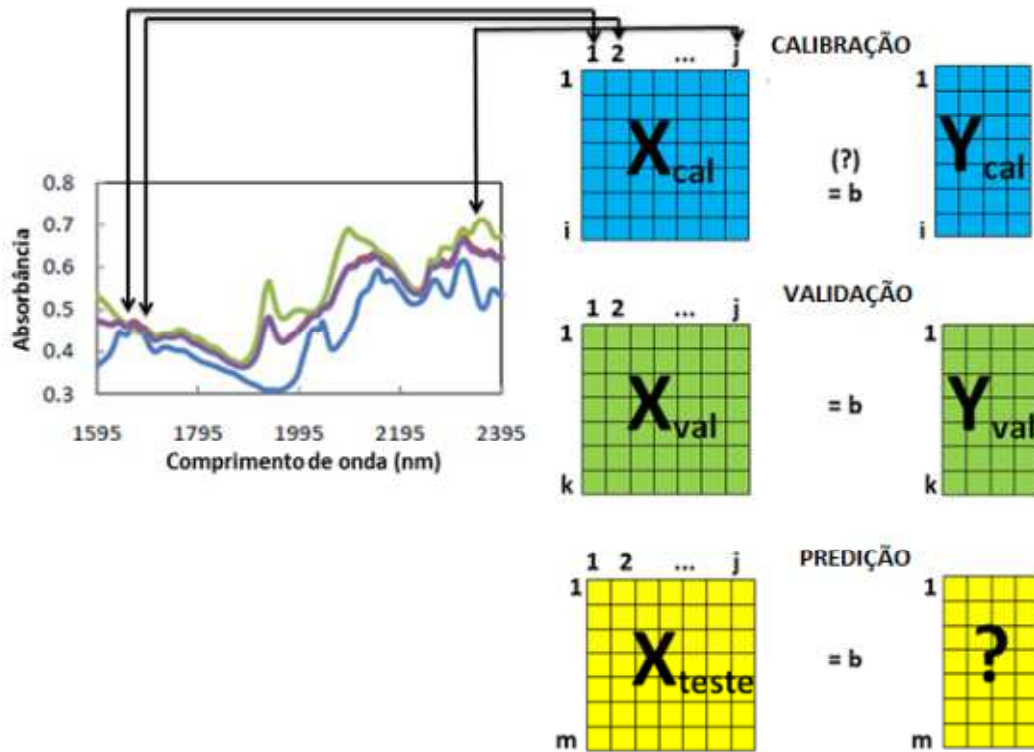
$$\mathbf{y} = f \mathbf{X} = \mathbf{Xb} \quad \text{Equação (2)}$$

Na construção do modelo de calibração o conjunto de dados empregados para esta finalidade é chamado de conjunto de calibração. Os parâmetros do modelo são denominados de estimativa dos coeficientes de regressão (b) determinados matematicamente a partir dos dados experimentais. (FONSECA, 2011; GRANGEIRO JÚNIOR, 2011) Após a construção do modelo de regressão, deve-se utilizar uma série de amostras externas para verificar a capacidade do mesmo em realizar previsões confiáveis, denominado conjunto de predição. (GRANGEIRO JÚNIOR, 2011) Neste conjunto, são utilizadas amostras cujas variáveis dependentes sejam conhecidas para que seja possível estabelecer uma comparação entre os valores preditos pelo modelo e os valores conhecidos previamente através de uma metodologia padrão, o que permitirá a avaliação sobre o desempenho do modelo de calibração proposto. (FONSECA, 2011)

Os dados para a calibração multivariada geralmente são organizados conforme a Figura 5, onde os valores de absorvância dos espectros a cada comprimento de onda correspondem às variáveis independentes (matriz \mathbf{X}), e as concentrações das amostras (ou outro parâmetro de interesse), definida pelo método de referência, indicam as variáveis dependentes (vetor \mathbf{y}). (GRANGEIRO JÚNIOR, 2011)

Espectros de i amostras do conjunto de calibração e de k amostras de validação são obtidos em j valores de comprimento de onda diferentes, formando uma matriz [$\mathbf{X}_{\text{cal}} (i \times j)$] para a fase de calibração e outra matriz [$\mathbf{X}_{\text{val}} (k \times j)$] para a fase de validação. Duas matrizes, $\mathbf{Y}_{\text{cal}} (i \times q)$ e $\mathbf{Y}_{\text{val}} (k \times q)$ também são formadas com q colunas, relacionadas aos diferentes componentes presentes nas amostras e cujos valores de concentração são conhecidos geralmente por métodos de análise de referência. (SIMÕES, 2008)

Figura 5 - Organização dos dados para calibração multivariada.



Fonte: Adaptado de (SIMÕES, 2008).

Existem vários métodos para a construção de modelos de calibração, sendo que a função que ajusta as variáveis dependentes e independentes pode ser linear ou não, dependendo da complexidade do sistema em estudo. (FONSECA, 2011)

Tem sido possível através de métodos quimiométricos, especialmente os métodos de calibração multivariada, a determinação de propriedades químicas e físicas de amostras com ligações ativas na região do Infravermelho e Raman. Isso tem sido feito por métodos com modelagem local, como regressão linear múltipla (do Inglês, *Multiple Linear Regression - MLR*), que utiliza a absorbância em comprimentos de onda selecionados, ou métodos de modelagem global, tais como a regressão por componentes principais (do Inglês, *Principal Component Regression - PCR*) ou mínimos quadrados parciais (do Inglês, *Partial Least Squares - PLS*). (FONSECA, 2011; SIMÕES, 2008)

Comumente, o PLS utiliza toda a faixa espectral para a construção dos modelos de calibração multivariada. Na prática, quando o MLR é aplicado geralmente é necessário o emprego de algoritmos de seleção de variáveis espectrais não redundantes dentro da faixa espectral de trabalho da (matriz \mathbf{X}) para minimizar problemas de multicolinearidade. (SIMÕES, 2008)

3.5.1.2 Regressão linear múltipla (MLR)

Dentre as técnicas de calibração multivariada, a MLR é considerada uma das mais simples e fácil de ser interpretada. As informações disponíveis nas variáveis independentes (matriz \mathbf{X}_{cal}) são aplicadas diretamente na modelagem para prever o parâmetro de interesse, ou seja, as variáveis originais são utilizadas. Durante a aplicação desse modelo, além de todas as informações relevantes, os ruídos também são levados em consideração. (GRANGEIRO JÚNIOR, 2011; SIMÕES, 2008)

Na construção de um modelo MLR cada variável dependente, isto é, cada (vetor \mathbf{y}_{cal}) da (matriz \mathbf{Y}_{cal}), é expressa como uma combinação linear das variáveis independentes da (matriz \mathbf{X}_{cal}) e um vetor (\mathbf{b}_{MLR}) que contém os coeficientes de regressão. A representação matemática dessa combinação linear é dada pela seguinte equação (SIMÕES, 2008):

$$\mathbf{y}_{cal} = \mathbf{X}_{cal}\mathbf{b}_{MLR} + \mathbf{e} \quad \text{Equação (3)}$$

A estimativa do vetor (\mathbf{b}_{MLR}) dos coeficientes lineares é realizada na etapa de calibração, utilizando-se o critério dos “Mínimos Quadrados” que minimiza o vetor dos resíduos ($\mathbf{e} = \mathbf{y}_{cal} - \mathbf{X}_{cal}\mathbf{b}_{MLR}$), sendo calculado da seguinte forma (SIMÕES, 2008).

$$\mathbf{b}_{MLR} = (\mathbf{X}_{cal}^T \mathbf{X}_{cal})^{-1} \mathbf{X}_{cal}^T \mathbf{y}_{cal} \quad \text{Equação (4)}$$

Onde: os índices sobrescritos (-1 e T) representam a inversão e transposição de uma matriz ou vetor, respectivamente.

Uma estimativa para a predição do valor da concentração ou do parâmetro de interesse em uma amostra desconhecida pode ser obtida por:

$$\hat{\mathbf{y}}_{desc} = \mathbf{X}_{desc}\mathbf{b}_{MLR} + \mathbf{e} \quad \text{Equação (5)}$$

A regressão linear múltipla apresenta alguns problemas que limitam sua aplicação. Um desses problemas é que o número de amostras deve ser igual ou superior ao número de variáveis, uma vez que, o modelo consiste na resolução de um sistema de equações lineares simultâneas. Essa condição necessita ser satisfeita caso contrário o sistema torna-se indeterminado. (GRANGEIRO JÚNIOR, 2011; SIMÕES, 2008)

Outro problema desse modelo, é que a matriz ($X_{cal}^T X_{cal}$) pode não apresentar inversa devido a alta correlação entre as variáveis. No entanto, este problema pode ser solucionado pela utilização de um algoritmo de seleção de variáveis. (GRANGEIRO JÚNIOR, 2011; SIMÕES, 2008)

3.5.1.3 Regressão por componentes principais (PCR)

A regressão por componentes principais (PCR) é um método de regressão linear multivariada usado para modelar relações lineares de entrada-saída. A maioria dos métodos de regressão linear se depara com o problema da multicolinearidade que é ocasionado pelas correlações existentes entre as variáveis independentes. Porém, a PCR remove a multicolinearidade por meio de uma técnica estatística multivariada conhecida como análise de componentes principais (do Inglês, *Principal Component Analysis* - PCA). A PCA mapeia os dados de entrada para um conjunto de variáveis independentes denominadas componentes principais (do Inglês, *Principal Component* - PC), que são uma combinação linear de variáveis de entrada. Após esse mapeamento os PC's são ordenados de tal forma que o primeiro PC explica a máxima variância dos dados e os demais PC's contabilizam a variância que não foi explicada por este. (ADUSUMILLI et al., 2015)

3.5.1.4 Análise de componentes principais (PCA)

A análise de componentes principais (PCA) tem por objetivo a redução do número de variáveis originais, substituindo-as por outras variáveis chamadas componentes principais (PC's). A PCA é capaz de decompor uma determinada matriz original em uma soma de matrizes, que são produtos de vetores denominados de *scores* e *loadings* (Equação 6).

$$X = T_1P_1 + T_2P_2 + T_3P_3 + T_4P_4 + T_5P_5 \dots + T_nP_n + E \quad \text{Equação (6)}$$

Onde: **T** e **P** são os *scores* e *loadings*, respectivamente e **E** é a matriz de resíduos.

Sabendo-se que a matriz X é determinada pela Equação 6, realiza-se a decomposição de valores singulares (do Inglês, *singular Value Decomposition* - SVD), obtendo-se em seguida, os valores dos *scores* e *loadings*. (CABRAL, 2011)

$$X = USP^t \quad \text{Equação (7)}$$

Onde: **P** é a matriz dos *loadings* e **US** a matriz dos *scores*. U e P são os vetores ortogonais, enquanto que, **S** é uma matriz diagonal na qual seus elementos são valores singulares e fornece informações sobre a variância que cada componente descreve. Sendo esta variância descrita pela Equação 8 (CABRAL, 2011).

$$\%VAR PC_J = \frac{s_{jj}^2}{\sum_{j=1}^P s_{jj}^2} \quad \text{Equação (8)}$$

A Equação 8 mostra que o denominador se refere à variância total, enquanto que, P indica o número total de valores singulares do conjunto de dados, com o numerador se referindo à variância de cada componente principal. (CABRAL, 2011) Desse modo, a primeira componente principal (PC1) é definida na direção que representa a máxima variância de um conjunto de dados. Enquanto que, a segunda componente principal (PC2) consiste de uma descrição da máxima variância no espaço da primeira componente principal ortogonal, e assim sucessivamente. Cada componente principal representa uma parcela de contribuição em uma fração de variância dos dados. Essas componentes estão num sistema de coordenadas ortogonais entre si, não existindo correlação entre elas. (CABRAL, 2011)

Para a determinação do número adequado de PC's de uma PCA, comumente emprega-se o método da validação cruzada (do Inglês, *Cross Validation* - CV), o qual se baseia na habilidade de predição de um modelo construído por parte de um conjunto de dados, seguido pela predição do restante desse próprio conjunto de dados, que é realizada pelo modelo construído. A CV pode ser realizada em blocos, ou seja, um número determinado de amostras é deixado de fora do processo de construção no modelo e a seguir essas amostras são preditas pelo modelo construído, ou ainda por um caso conhecido como *leave-one-out* (deixe uma fora por vez), em que uma amostra é deixada de fora no processo de construção do modelo, sendo também predita pelo modelo construído em seguida. Em ambos os casos, o processo é repetido até que todas as amostras tenham sido preditas, e assim, o erro quadrático médio da validação cruzada (do Inglês, *Root Mean Square Error of Cross Validation* - RMSECV) possa ser calculado. (FONSECA, 2011)

3.5.1.5 Regressão por mínimos quadrados parciais (PLS)

O PLS ou regressão por PLS (do Inglês, *Partial Least Squares Regression* - PLSR) foi desenvolvido por Herman Wold e Martens em meados de 1964, sendo considerado o método de regressão mais utilizado para a construção de modelos de calibração multivariada a partir de dados de primeira ordem (FONSECA, 2011; SIMÕES, 2008). Para a construção de tais modelos, o método utiliza tanto as informações da matriz de dados espectroscópicos (matriz **X**), como as informações da matriz de propriedades de interesse (matriz **Y**). (SIMÕES, 2008)

A regressão por mínimos quadrados parciais estende o conceito do modelo inverso (propriedade como função da resposta instrumental) trocando as variáveis originais por um subconjunto truncado das variáveis latentes dos dados originais. (BEEB; RANDY; SEASHOLTZ, 1998; OLIVEIRA, 2006) Considerando um caso geral para a determinação de mais de uma espécie de interesse, as matrizes **Xcal** e **Ycal** (Figura 5) são decompostas em suas matrizes de *scores* e *loadings*, respectivamente, sendo esse processo matematicamente representado pelas Equações 9 e 10, respectivamente. (SIMÕES, 2008)

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E} \quad \text{Equação (9)}$$

$$\mathbf{Y} = \mathbf{UQ}^T + \mathbf{F} \quad \text{Equação (10)}$$

Em que: (**T** e **U**) são as matrizes dos *scores*; (**P** e **Q**) são as matrizes dos *loadings* das matrizes (**X** e **Y**), respectivamente. O termo (**E**) representa a matriz de resíduos espectrais, enquanto que, (**F**) representa a matriz dos resíduos de concentração. (SIMÕES, 2008)

A decomposição em *scores* e *loadings* pode ser realizada empregando diferentes tipos de algoritmos, como por exemplo, o NIPALS (do Inglês, *Non Linear Iterative Partial Least Squares Algorithm*). O modelo PLS final consiste em relacionar linearmente os *scores* da (matriz **X**) com os *scores* da (matriz **Y**), de acordo com as Equações 11 e 12, respectivamente. (SIMÕES, 2008)

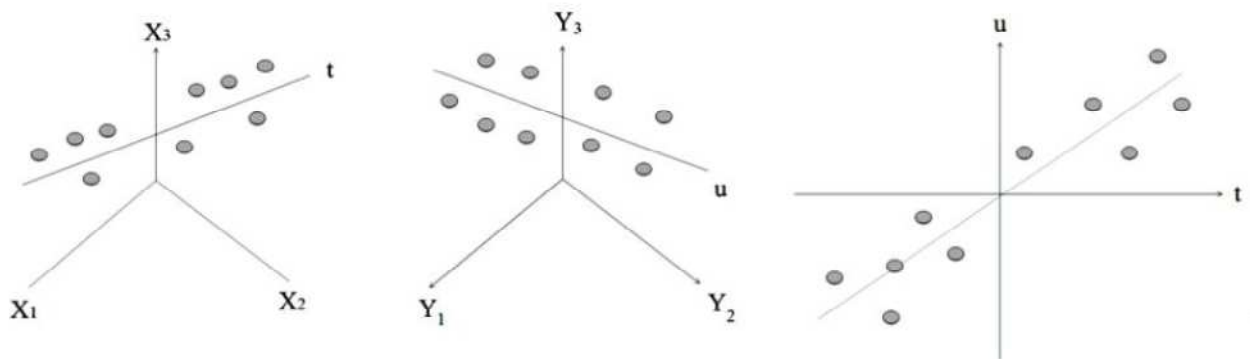
$$\mathbf{U} = \mathbf{BT} + \mathbf{G} \quad \text{Equação (11)}$$

$$\mathbf{Y} = \mathbf{BTQ}^T + \mathbf{H} \quad \text{Equação (12)}$$

Neste caso, (**B**) representa a matriz dos coeficientes de regressão. (**G**) é a matriz de resíduos dos *scores* e (**H**) é a matriz dos resíduos de concentração.

A Figura 6 exemplifica, de forma genérica, como o método PLS processa os dados da matriz. Logo, para a construção desse modelo, a regressão dos mínimos quadrados retira informações de um conjunto de dados da (matriz **X**) correlacionando-os com as informações retiradas do conjunto de dados da (matriz **Y**). (CABRAL, 2011)

Figura 6 - Representação genérica do método PLS.



Fonte: CABRAL, 2011.

3.6 Seleção de Variáveis e Amostras

Para a maioria dos espectros, as medidas em um comprimento de onda estão correlacionadas com as medidas no comprimento de onda vizinho. Sendo assim, a seleção de variáveis que resulte na máxima precisão para construção dos modelos multivariados tem sido uma tarefa desafiadora, principalmente quando os espectros apresentam fortes sobreposições e possuem características distintivas imperceptíveis. (ANDERSEN; BRO, 2010; ARAUJO et al., 2001) Além disso, quando os parâmetros de interesse estão dentro de matrizes complexas, torna-se difícil obter um conjunto representativo de amostras por meio de planejamentos experimentais. Com isso, a divisão do conjunto original de amostras em subconjuntos de calibração e predição tem sido uma etapa de suma importância para a construção de modelos de calibração multivariada. (GRANGEIRO JÚNIOR, 2011)

Diante dessas necessidades, vários algoritmos para seleção de variáveis foram desenvolvidos. A exemplo do algoritmo “*branch and bound*”, frequentemente aplicado em otimização combinatória, o método dos mínimos quadrados parciais por intervalo (do Inglês, *Interval Partial Least Squares Regression* - iPLS), transformada wavelet e o algoritmo

genético (do Inglês, *Genetic Algorithm* - GA). (ARAUJO et al., 2001) Este último, tem sido considerado como uma alternativa bastante viável e flexível, sendo amplamente utilizado. Contudo, devido à sua natureza estocástica, o GA pode não fornecer resultados reprodutíveis. Para contornar este problema, uma alternativa bastante viável é o emprego do algoritmo das projeções sucessivas (do Inglês, *Successive Projections Algorithm* - SPA), em que sua natureza determinística faz com que os seus resultados sejam sempre reprodutíveis. (SIMÕES, 2008)

3.6.1 Algoritmos para seleção de variáveis

3.6.1.1 Algoritmo das projeções sucessivas (SPA)

O algoritmo SPA desenvolvido por Araujo et al. (2001) é considerado um método de seleção direta, isto é, ele começa com um comprimento de onda e vai incorporando novos comprimentos de onda a cada iteração, até que um número especificado N de comprimentos de onda tenha sido atingido. O SPA tem como objetivo selecionar comprimentos de onda cujo conteúdo de informações seja minimamente redundante, para resolver problemas de colinearidade.

Os passos do SPA são descritos abaixo, assumindo que o primeiro comprimento de onda $j(0)$ e o número N são fornecidos. (ARAUJO et al., 2001)

Passo 0: Antes da primeira iteração ($n = 1$), têm-se $\mathbf{X}_i = i$ -ésima coluna de \mathbf{X}_{cal} ; $j = 1, \dots, J$.

Passo 1: Seja S o conjunto de comprimentos de onda que ainda não foram selecionados, isto é, $S = \{p \text{ tal que } 1 \leq j \leq J \text{ e } p \notin \{i(0), \dots, i(n-1)\}\}$.

Passo 2: Calcular a projeção de \mathbf{X}_p no subespaço ortogonal para $\mathbf{X}_{i(n-1)}$ como:

$$\mathbf{P}\mathbf{X}_j = \mathbf{X}_j - (\mathbf{X}_j^T \mathbf{X}_{i(n-1)}) \mathbf{X}_{i(n-1)} (\mathbf{X}_{i(n-1)}^T \mathbf{X}_{i(n-1)})^{-1} \quad \text{Equação (13)}$$

Para todos $p \in S$, onde \mathbf{P} é o operador da projeção.

Passo 3: Faz-se $i(n) = \arg(\max_{j \in S} \mathbf{P}\mathbf{X}_j)$.

Passo 4: Faz-se $\mathbf{X}_p = \mathbf{P}\mathbf{X}_p, j \in S$.

Passo 5: Faz-se $n = n+1$. Se $n < N$, volte para o passo 1.

Fim: Os comprimentos de onda resultantes são $\{i(n); n=0, \dots, N-1\}$.

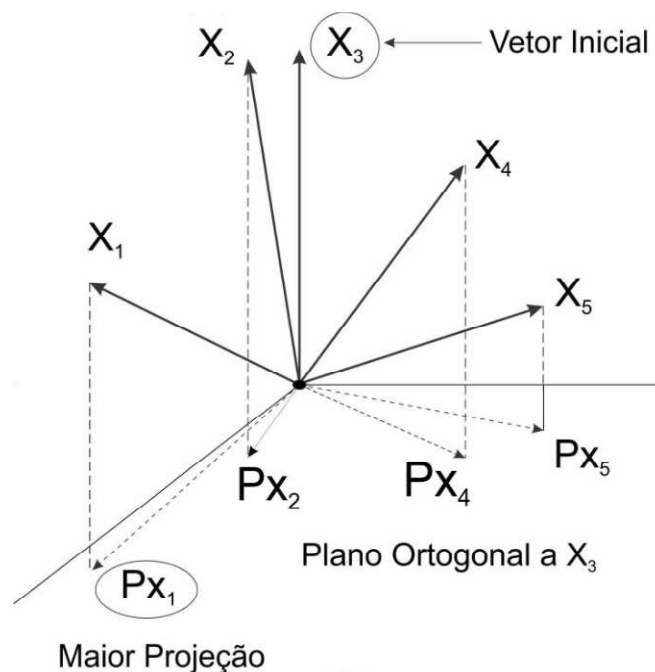
O número de operações de projeções realizadas no processo de seleção pode ser mostrado como $(N - 1) (j - \frac{N}{2})$. Os passos acima que exemplificam a primeira iteração do SPA são mostrados na Figura 7.

Para cada J cadeia construída é possível extrair I_c subconjuntos de variáveis pelo uso de um dos elementos I_c na ordem que eles foram selecionados. Com isso, um total de $J \times I_c$ subconjuntos de variáveis pode ser formado. Na escolha do subconjunto mais apropriado, modelos MLR são construídos e comparados em termos do erro médio quadrático de predição para o conjunto de validação (do Inglês, *Root Mean Square Error of Validation* - RMSEV), em que este é calculado da seguinte forma:

$$RMSEV = \sqrt{\frac{1}{I_V} \sum_{i=1}^{I_V} (y_V^i - \hat{y}_V^i)^2} \quad \text{Equação (14)}$$

Em que: y_V^i e \hat{y}_V^i são os valores de referência e previstos do parâmetro de interesse na i-ésima amostra de validação e I_V é o número de amostras de validação.

Figura 7 - Exemplo do SPA com $J = 5$, $Mcal = 3$ e $k(0) = 3$. Resultado da primeira projeção: $k(1) = 1$.



Fonte: Adaptado de (ARAÚJO et al., 2001).

Definir um conjunto de validação representativo pode não ser uma tarefa simples quando se trabalha com matrizes complexas como óleo e produtos alimentícios. Pois a variabilidade da

composição não pode ser facilmente reproduzida por planejamentos experimentais otimizados. Portanto, o conjunto de validação deve ser extraído das amostras reais disponíveis para a construção do modelo. Diante dessa necessidade, o uso das técnicas de validação cruzada pode ser uma alternativa adequada. (SIMÕES, 2008) Nesse sentido, Galvão et al. (2005) fizeram uma implementação ao algoritmo SPA, onde a técnica de validação cruzada é aplicada para a validação interna do modelo.

Assumindo que os dados de resposta instrumental estão dispostos em uma matriz \mathbf{X} de dimensões (I x J) tal que a j-ésima variável x_j está associada com o j-ésimo vetor coluna x_j e R^1 . Fazendo $M = \min(I - 1, J)$ ser o número máximo de variáveis que pode ser incluído em um modelo SPA com o termo intercepto. O SPA com validação cruzada consiste em três fases: A primeira fase, onde projeções são realizadas na matriz \mathbf{X} , gerando J cadeias de M variáveis cada, sendo processada de maneira idêntica à que ocorre no SPA original. A segunda fase consiste em avaliar os subconjuntos candidatos de variáveis extraídas das cadeias geradas na primeira fase. Por fim, na terceira e última fase realiza-se a eliminação das variáveis que não contribuem significativamente para modelo de predição do MLR. (SIMÕES, 2008)

A Equação 15 apresenta a raiz quadrada do erro médio da validação cruzada (RMSECV) que é o critério de avaliação utilizado para fases 2 e 3.

$$RMSECV = \frac{1}{N_{cal}} \sum_{n=1}^{N_{cal}} (y_{Cal,n} - y_{cal,n})^2 \quad \text{Equação (15)}$$

Em que: $y_{cal,n}$ é o valor de referência do parâmetro em consideração para a enésima amostra do conjunto de calibração, o qual contém N_{cal} amostras.

O valor de $y_{cal,n}$ é obtido por meio do procedimento *leave-one-out*, o qual consiste em remover a i-ésima amostra do conjunto de calibração. Na sequência é gerado um modelo com as amostras restantes para ser aplicado à amostra removida. Este procedimento é aplicado apenas nas fases 2 e 3 do SPA. A Primeira fase (operações de projeção) não é repetida no procedimento de validação cruzada para minimizar o trabalho computacional envolvido no processo. (34)

3.6.1.2 Teste de incerteza de Martens

O teste de incerteza de Martens é um método de teste de significância implementado no software The Unscrambler[®] que permite a avaliação da estabilidade dos modelos PLS, PCR ou PCA. Além disso, essa ferramenta possibilita a identificação de amostras ou variáveis perturbadoras e a seleção das variáveis espectrais (X) mais significativas. O conceito do teste é baseado na validação cruzada (que pode ser completa ou segmentada), princípio Jack-Knife e gráficos de estabilidade. (MARTENS, H.; MARTENS, M., 2000)

No teste de incerteza de Martens é avaliada a incerteza dos parâmetros do modelo individual, em cada segmento de validação cruzada é obtido um conjunto de estimativas de parâmetros do modelo. Cada conjunto de estimativas de parâmetros é baseado em todos os objetos, exceto um ou mais que são mantidos fora neste segmento de validação cruzada. Se um modelo no segmento difere muito do modelo comum construído com base em todos os objetos, isso significa que o(s) objeto(s) mantido(s) fora desse segmento afetou/afetaram seriamente o modelo comum, pressionando algum padrão de variação nos parâmetros do mesmo. Assim, um gráfico de como os parâmetros do modelo são perturbados quando diferentes objetos são mantidos da etapa de calibração nos diferentes segmentos de validação cruzada, mostra a estabilidade do modelo comum contra peculiaridades nos dados de objetos individuais ou segmentos destes. (CAMO S.A., 1998)

Quando a validação cruzada é aplicada na regressão, o ranque da matriz (**A**) é determinado baseado no número de objetos que é deixado fora no modelo individual. A variância aproximada da incerteza nos coeficientes de regressão PLS (**B**) pode ser estimada pelo princípio Jack-Knife de acordo com a Equação 16.

$$S^2\mathbf{B} = \frac{M}{m=1}((\mathbf{B} - \mathbf{B}_m)\mathbf{g})^2 \quad \text{Equação (16)}$$

Onde:

$S^2\mathbf{B}_{(K \times J)}$ = Incerteza da variância estimada de **B**;

$\mathbf{B}_{(K \times J)}$ = O coeficiente de regressão no ranque cross validado (**A**) usando n objetos;

$\mathbf{B}_m_{(K \times J)}$ = O coeficiente de regressão no ranque cross validado (**A**) usando todos os objetos, exceto os objetos excluídos no segmento de validação cruzada (m);

g = coeficiente escalonado (admitiu-se $g = 1$).

3.6.2 Algoritmos para seleção de amostras

3.6.2.1 Algoritmo Kennard-Stone (KS)

A construção de modelos de calibração multivariada a partir de um grande conjunto de amostras é tida como um problema recorrente, tendo em vista que nem sempre os subconjuntos escolhidos apresentam uma representatividade adequada das amostras originais. Nesse sentido, faz-se necessário a utilização de métodos de seleção de variáveis que possam auxiliar de forma prática a definição dos subconjuntos de calibração e validação.

Uma alternativa simples e comumente empregada para esta finalidade é o método de seleção aleatória (do Inglês, *Random Search* - RS), que emprega a distribuição estatística do conjunto inteiro na extração do grupo de dados do conjunto maior. Contudo, este método não garante a representatividade do conjunto de dados, nem previne contra possíveis problemas de extrapolação. Uma boa alternativa para contornar os problemas apresentados pelo RS é o algoritmo Kennard-Stone (do Inglês, *Kennard-Stone Algorithm* - KS) que faz a seleção das amostras com base em suas distâncias euclidianas. (SIMÕES, 2008; SOUZA, 2011)

Este algoritmo tem como objetivo selecionar um subconjunto representativo de um conjunto de N amostras. A fim de garantir uma distribuição uniforme de tal subconjunto ao longo do espaço de dados \mathbf{X} (resposta instrumental), o KS segue um procedimento por etapas no qual novas seleções são tomadas em regiões do espaço longe das amostras já selecionadas. Para isto, o algoritmo emprega as distâncias Euclidianas $d_{\mathbf{x}}(p, q)$ entre os \mathbf{x} -vetores de cada par (p, q) de amostras calculadas como (GALVÃO et al., 2005):

$$d_{\mathbf{x}}(p, q) = \sqrt{\sum_{j=1}^J [x_{p j} - x_{q j}]^2} ; \quad p, q \in [1, N] \quad \text{Equação (17)}$$

Em que os dados espectrais, $x_{p j}$ e $x_{q j}$ são as respostas instrumentais no j-ésimo comprimento de onda para as amostras p e q, respectivamente. Enquanto que J indica o número de comprimentos de onda no espectro. (GALVÃO et al., 2005)

A seleção começa tomando-se o par (p1, p2) de amostras para as quais a distância $d_{\mathbf{x}}(p1, p2)$ é a maior. Em cada iteração subsequente, o algoritmo seleciona a amostra que exibe a maior distância mínima em relação a qualquer amostra já selecionada. Sendo este

procedimento repetido até que o número de amostras especificadas pelo analista seja alcançado. (GALVÃO et al., 2005)

3.6.2.2 Algoritmo SPXY

O algoritmo SPXY (do Inglês, *Sample set Partitioning based on joint x-y distances*) proposto por Galvão et al. (2005) é um método alternativo para a seleção dos conjuntos de calibração, validação e predição. Este método funciona como uma extensão do algoritmo KS e leva em consideração tanto as diferenças de \mathbf{X} quanto a de \mathbf{y} no cálculo das distâncias inter amostras.

A proposta do SPXY consiste em prever a distância definida na Equação 17 com a distância no espaço variável dependente (y) do parâmetro em consideração. Cada distância euclidiana $d_y(p, q)$ pode ser calculada para cada par de amostras p e q como (GALVÃO et al., 2005):

$$d_y(p, q) = \sqrt{(y_p - y_q)^2} = |y_p - y_q| ; \quad p, q \in [1, N] \quad \text{Equação (18)}$$

Para atribuir o mesmo grau de importância à distribuição das amostras nos espaços \mathbf{X} e \mathbf{y} , as distâncias $d_x(p, q)$ e $d_y(p, q)$ são divididas pelos seus valores máximos no conjunto de dados. Desta forma, uma distância d_{xy} normalizada é calculada como (GALVÃO et al., 2005):

$$d_{xy}(p, q) = \frac{d_x(p, q)}{\max_{p, q \in [1, N]} d_x(p, q)} + \frac{d_y(p, q)}{\max_{p, q \in [1, N]} d_y(p, q)} ; \quad p, q \in [1, N] \quad \text{Equação (19)}$$

Um procedimento de seleção por etapas, similar ao algoritmo KS, pode ser aplicado com $d_{xy}(p, q)$ ao invés de $d_x(p, q)$ sozinho. (GALVÃO et al., 2005)

3.7 Pré-processamentos de Dados

Toda medida instrumental é composta por uma porção de sinal referente ao parâmetro de interesse e uma porção de ruído. A presença de ruído ou informação não relacionada ao parâmetro de interesse pode levar a construção de modelos com baixa performance preditiva.

Dessa forma, uma etapa indispensável para a construção de modelos multivariados baseados em NIRS é o estudo da necessidade de pré-processamento espectral.

Os pré-processamentos espectrais como variação padrão normal (do Inglês, *Standard Normal Variate* - SNV) e correção de dispersão multiplicativa (do Inglês, *Multiplicative Scatter Correction* - MSC) são ambos utilizados com a finalidade de remover ou minimizar os efeitos de espalhamento dos espectros de reflexão, os quais podem ser atribuídos a dispersão da radiação no momento em que esta interage com as amostras. (FEARN et al., 2009)

O MSC, introduzido por Martens et. al (1983) e aprimorado por Geladi et. al (1985), realiza a correção dos espectros utilizando coeficientes de correção gerados a partir da regressão por mínimos quadrados de um espectro de referência. (HELLAND; NAES; ISAKSSON, 1995) Na maioria das aplicações, o espectro médio do conjunto de calibração é adotado como o espectro de referência. No entanto, um espectro de referência genérico também pode ser aplicado. No artigo original de Martens et. al (1983), foi sugerido usar apenas as partes do eixo espectral que não incluam informações relevantes (linha de base). Embora isso seja considerado como um bom senso espectroscópico torna-se difícil determinar essas regiões na prática, especialmente nas medidas NIR, onde os sinais de diferentes componentes químicos estão fortemente sobrepostos e correlacionados, e pouca ou nenhuma linha de base verdadeira é encontrada. Esta é a razão pela qual, na maioria dos casos, todo o espectro é usado para encontrar os parâmetros de correção escalar no MSC. (RINNAN; BERG; ENGELSEN, 2009)

Para corrigir o espalhamento espectral, também pode-se empregar a forma expandida do MSC (do Inglês, *Extended Multiplicative Scatter Correction* – EMSC). Esta expansão inclui tanto o ajuste polinomial de segunda ordem quanto o espectro de referência, o ajuste de uma linha de base no eixo do comprimento de onda e as utilizações do conhecimento a priori a partir dos espectros de interesse ou interferentes espectrais. (RINNAN; BERG; ENGELSEN, 2009) O EMSC não se limita a apenas remover efeitos multiplicativos e aditivos dos espectros. Ele também permite uma separação dos efeitos físicos da dispersão da luz dos efeitos químicos da absorvância da luz nos espectros. (CAMO S.A., 1998)

Os conceitos de correção de espalhamento espectral por trás do SNV são os mesmos que para o MSC, exceto o espectro de referência comum que não é necessário. A diferença prática é que o SNV padroniza cada espectro usando apenas os seus próprios dados como a média e o desvio padrão, não utiliza, portanto, o espectro médio de nenhum conjunto. Contudo, como esse pré-processamento não envolve um ajuste de mínimos quadrados na

estimativa dos parâmetros de correção, ele pode ser sensível a entradas ruidosas nos espectros. Para contornar esse problema, Guo et al. (1999) sugeriu usar outros parâmetros estatísticos mais robustos como a média da faixa do quartil interno e o desvio padrão do quartil interno para o processamento do SNV. Isto seria bastante apropriado para os casos de espectros ruidosos, como aqueles obtidos em aplicações NIR ultrarrápidas online, onde as medidas robustas seriam menos afetadas por ruídos pontuais (por exemplo, reflexões seletivas de partículas de onda em fluxos líquidos). De forma geral, o SNV fornece resultados muito parecidos ao MSC, porém a opção por um ou outro está ligado a uma questão de escolha por parte dos seus usuários. (CAMO S.A., 1998; RINNAN; BERG; ENGELSEN, 2009)

Outro método bastante utilizado no pré-processamento de dados espectrais, sobretudo quando se deseja melhorar a razão sinal/ruído, é a suavização Savitzky-Golay (SG). (SAVITZKY; GOLAY, 1964) Esse método tem como princípio ajustar os dados a polinômios de baixa ordem, para isto substitui pontos centrais de janelas que possuem tamanhos pré-estabelecidos por pontos que são estimados pelo polinômio. Durante a SG deve atentar para a quantidade de pontos escolhidos por janelas, pois janelas com muitos pontos levam a distorção de dados e pode acarretar em perdas de informações importantes. Por outro lado, se a quantidade de pontos for subestimada, a eliminação do ruído pode não ser satisfatória. Outro parâmetro a ser levado em consideração é a escolha da ordem do polinômio que será combinado às janelas escolhidas. Desse modo, na SG recomenda-se a realização de diversas combinações de janelas e ordens de polinômios a fim de proporcionar a escolha dos melhores resultados. (BEEB; RANDY; SEASHOLTZ, 1998; RINNAN; BERG; ENGELSEN, 2009)

Além dos efeitos de espalhamento, os espectros NIR também podem apresentar outras anomalias como deslocamento de linha de base, o que exige a utilização de métodos adicionais aos anteriormente apresentados para a correção desses desvios e, conseqüentemente, melhorar as performances dos modelos multivariados. Para entender a esta finalidade, um método bastante utilizado é a derivação baseada na suavização Savitzky-Golay (SAVITZKY; GOLAY, 1964). Para encontrar a derivada num determinado ponto central, um polinômio é montado em uma janela simétrica nos dados brutos. Quando os parâmetros para este polinômio são calculados, a derivada de qualquer ordem dessa função pode ser facilmente encontrada analiticamente e esse valor é posteriormente usado como estimativa da derivada para este ponto central. Esta operação é aplicada em todos os pontos nos espectros de forma sequencial. O número de pontos utilizados para calcular o polinômio (tamanho da janela) e o grau ajustado do polinômio são duas decisões que precisam ser feitas. A derivada

mais alta que pode ser determinada depende do grau do polinômio usado durante o ajuste (isto é, um polinômio de terceira ordem pode ser usado para estimar até a derivada de terceira ordem). Do mesmo modo que a suavização, a derivação Savitzky-Golay requer bastante atenção na escolha de tamanho das janelas e das ordens dos polinômios para que informações importantes utilizadas na construção dos modelos não se percam nem sejam distorcidas. (BEEB; RANDY; SEASHOLTZ, 1998; RINNAN; BERG; ENGELSEN, 2009)

3.8 Cálculo dos Erros

A eficiência do modelo de calibração multivariada pode ser analisada pelo cálculo dos valores de raiz do erro médio quadrático (RMSE) e erro relativo (E). Esses valores expressam a exatidão do modelo, ou seja, a proximidade entre o valor calculado pelo modelo ($y_{predição}$) e o valor verdadeiro ou obtido por um método de referência (y_{real}). Os erros são definidos como (FONSECA, 2011):

$$RMSE = \frac{\sqrt{\sum_{i=1}^n (y_{predição} - y_{real})^2}}{n} \quad \text{Equação (20)}$$

Em que: (n) é número de amostras.

$$E = \frac{\sqrt{(y_{predição} - y_{real})^2}}{(y_{real})^2} \cdot 100 \quad \text{Equação (21)}$$

O RMSEC é o RMSE calculado a partir das amostras de calibração, isto é, uma medida do erro na modelagem; RMSECV a partir das amostras da validação cruzada e RMSEP das amostras do conjunto de predição. (CABRAL, 2011; FONSECA, 2011; GRANGEIRO JÚNIOR, 2011)

3.9 Coeficiente de Correlação

Uma simples associação entre as variáveis (X e Y) produz um par de valores ou, graficamente, aponta esta dispersão em uma linha reta. Uma pequena dispersão muito próxima à linha indica forte associação. A medida numérica desta relação é chamada Coeficiente de Correlação da Amostra. Também conhecido como Coeficiente de Correlação

de Pearson, em homenagem ao seu criador, sendo simbolizado por (R). (FONSECA, 2011)
Observando-se a Equação 22, infere-se que o valor de (R) varia no intervalo $(-1,0 \leq R \leq 1,0)$.

$$R = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad \text{Equação (22)}$$

Onde: $(x_1, y_1), \dots, (x_n, y_n)$ são os (n) pares de observações.

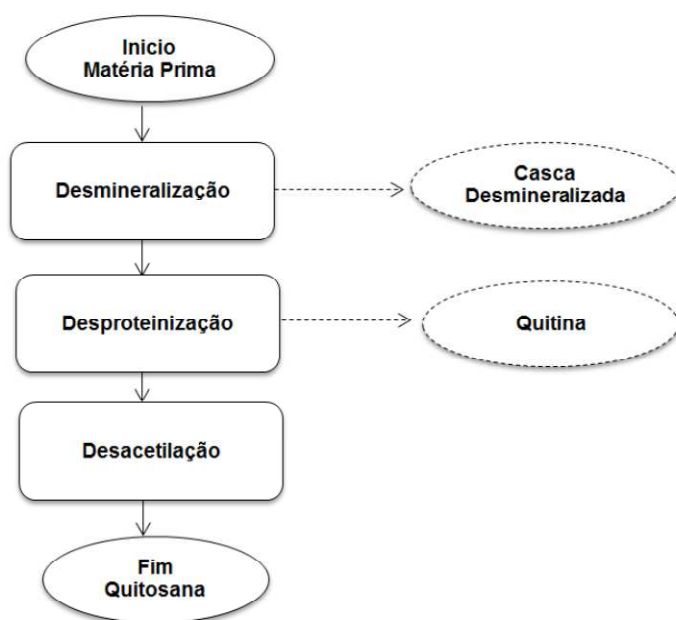
Valores próximos a zero indicam falta de correlação e valores próximos a $(-1,0$ e $1,0)$ indicam alta correlação. (FONSECA, 2011)

4 METODOLOGIA

4.1 Amostras

As amostras de quitosana utilizadas neste trabalho tiveram como matéria-prima o exoesqueleto de camarão (*Litopenaeus vannamei* Boone) e foram produzidas no Laboratório de Processamento III do CERTBIO de acordo com a metodologia de referência. Cujas etapas gerais de produção são ilustradas no fluxograma da Figura 8.

Figura 8 - Fluxograma esquemático das etapas gerais de extração de quitina e obtenção de quitosana.



Fonte: CERTBIO, 2017.

Na metodologia padrão de produção da quitosana pelo CERTBIO, o ajuste de tempo das etapas de desmineralização, desproteíntização e desacetilação são respectivamente: 30 min, 2 e 6 horas.

Como o intuito deste trabalho é desenvolver e validar uma metodologia alternativa para a determinação da massa molar e do grau de desacetilação da quitosana foram promovidas variações intencionais nos tempos da etapa de desacetilação, que é a etapa que está diretamente relacionada a esses dois parâmetros de caracterização, visando assim a ampliação da faixa de trabalho.

Para tanto, foram produzidos 10 lotes de quitosana variando-se os tempos de desacetilação em: 3, 4, 5, 6 e 7 horas. Sendo produzidas 5 amostras por lote e 50 amostras no

total. As etapas de produção da quitosana e as variações de tempos realizadas na etapa de desacetilação são ilustradas na Tabela 3.

Tabela 3 - Etapas do processo padrão de produção da quitosana no CERTBIO e as variações intencionais promovidas na etapa de desacetilação.

DM (desmineralização)	DP (desproteínização)	DA (desacetilação)
		3 horas
		4 horas
30 minutos	2 horas	5 horas
		6 horas (tempo padrão)
		7 horas

Fonte: Adaptado de (CERTBIO, 2017).

4.2 Determinação dos Parâmetros de Referência

A determinação dos parâmetros de referência para as amostras de quitosana foi realizada com base nas metodologias convencionais utilizadas pelo Laboratório CERTBIO. Na determinação da massa molar reproduziu-se o protocolo descrito na ASTM F2103-11 (2011), com a substituição do solvente padrão NaCl 0,1 mol.L⁻¹ por uma solução tampão composta de ácido acético e acetato de sódio 0,3 mol.L⁻¹/0,2 mol.L⁻¹, conforme o estudo de Rinaudo et al. (1993). Já a determinação do grau de desacetilação foi com base na metodologia proposta por Brugnerotto et al. (2001).

4.2.1 Determinação da massa molar (MM)

A massa molar (MM) das amostras foi determinada pelo método da viscosimetria intrínseca (ASTM F2103-11, 2011), tendo como base a equação de MHS (Equação 1). Foi utilizado um viscosímetro capilar de Ubbelohde em um banho termostático de silicone a 25°C. Como solvente padrão utilizou-se uma solução tampão de ácido acético e acetato de sódio 0,3 mol.L⁻¹/0,2 mol.L⁻¹. Os valores das constantes viscosimétricas utilizados foram aqueles descritos por Rinaudo et al. (1993) em seu estudo ($K = 0,076 \text{ mL.g}^{-1}$ e $\alpha = 0,76$). As soluções de quitosana foram preparadas na concentração de (1,10 g.L⁻¹), dissolvendo-se 0,0275 g das amostras em 25 mL do solvente padrão.

Em seguida, foram medidos os tempos de escoamento em triplicata tanto para o solvente padrão (t_0) quanto para as soluções de quitosana com suas respectivas diluições (t). Em que, estas diluições foram obtidas adicionando-se o solvente padrão a um volume de 12 mL da solução de quitosana contidos no viscosímetro. Ou seja, as diluições foram realizadas no mesmo recipiente obedecendo-se a seguinte sequência:

- **1ª diluição:** 1 mL de solução padrão;
- **2ª diluição:** 1,5 mL de solução padrão;
- **3ª diluição:** 2 mL de solução padrão;
- **4ª diluição:** 2 mL de solução padrão.

A partir da razão entre os tempos de escoamento (t/t_0), obtiveram-se os valores das viscosidades relativas (n_{rel}), conforme a Equação 23.

$$n_{rel} = \frac{n}{n_0} = \frac{t\rho}{t_0\rho_0} \cong \frac{t}{t_0}, \rho \approx \rho_0 \quad \text{Equação (23)}$$

A partir das viscosidades relativas e com o auxílio da Equação 24, obtiveram-se as viscosidades específicas (n_{sp}).

$$n_{sp} = \frac{n-n_0}{n_0} = \frac{n}{n_0} - 1 \quad \text{Equação (24)}$$

As razões entre as viscosidades específicas e as concentrações das soluções de quitosana, permitiram realizar os cálculos das viscosidades reduzidas (n_{red}), conforme Equação 25.

$$n_{red} = \frac{n_{sp}}{c} = \frac{n_{rel}-1}{c} = \frac{\left(\frac{n}{n_0}-1\right)}{c} \quad \text{Equação (25)}$$

Em seguida, construiu-se uma curva de calibração da viscosidade reduzida (n_{red}) em função da concentração (C). A partir dessa curva e com ajuste linear adequado, extrapolou-se o valor de $[n]$ no eixo das ordenadas. Por último, os valores das massas molares viscosimétricas médias das amostras (M_W) foram determinadas com base na Equação 1.

4.2.2 Determinação do grau de desacetilação (GD)

O grau de desacetilação (GD) das amostras foi determinado a partir da técnica espectroscópica de infravermelho com transformada de Fourier (FTIR).

As análises foram realizadas num espectrômetro FTIR do fabricante Perkin Elmer (modelo Spectrum 400). As amostras de quitosana foram preparadas na forma de pastilhas de brometo de potássio (KBr), na proporção de 1/6 (amostra/KBr). Sendo utilizados como massa padrão, 20 mg para as amostras de quitosana e 120 mg para o KBr. Em seguida, registraram-se os espectros em triplicata das amostras, utilizando-se os seguintes parâmetros:

- Faixa de varredura de 4.000 a 400 cm^{-1} ;
- 16 varreduras (scans);
- Resolução de 4 cm^{-1} .

Após a aquisição dos espectros, empregou-se a metodologia proposta por Brugnerotto et al. (2001) para o cálculo do GD. Onde foram utilizadas as linhas de base para o cálculo das áreas integradas dos picos centrais dos grupamentos amida III (1320 cm^{-1}) e CH_2 (1420 cm^{-1}), conforme a Equação 26.

$$\%GD = 100 - \left[\frac{\frac{A_{1320}}{0,03133} - 0,3822}{A_{1420}} \right] \quad \text{Equação (26)}$$

Em que:

A_{1320} = Área integrada do pico central do grupamento amida III (1320 cm^{-1});

A_{1420} = Área integrada do pico central do grupamento CH_2 (1420 cm^{-1}).

4.3 Registros dos Espectros NIR

Os espectros NIR de todas as amostras de quitosana em pó foram registrados no modo de reflectância difusa no Laboratório CERTBIO, empregando-se o mesmo equipamento utilizado na determinação do GD. A leitura das amostras foi realizada em triplicatas, utilizando-se os seguintes parâmetros:

- Faixa de varredura de 9.000 a 4.000 cm^{-1} ;
- 32 varreduras;

- Resolução de 8 cm^{-1} .

4.4 Construção e Validação dos Modelos

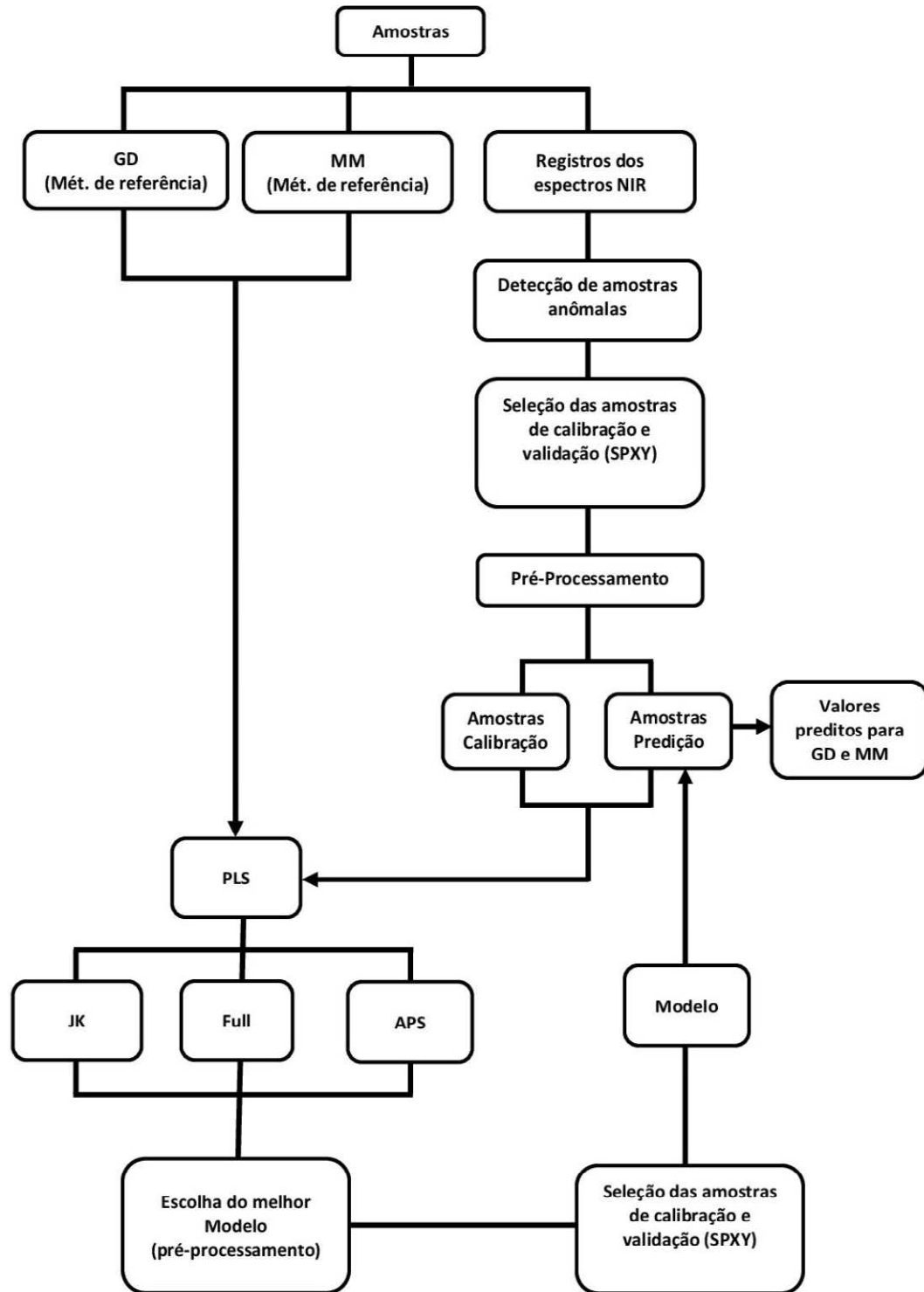
O programa UNSCRAMBLER[®] v9.7 da CAMO S.A. foi empregado para pré-processamento dos espectros e construção dos modelos PLS-1. Na seleção das amostras para os conjuntos de calibração e predição foi utilizado o algoritmo SPXY (ARAUJO et al., 2001). A seleção de variáveis foi realizada utilizando o algoritmo SPA (GALVÃO et al., 2005), ambos os algoritmos foram executados empregando-se o ambiente MATLAB[®], versão 6.5.

4.4.1 Etapas para a construção dos modelos

Após o registro dos espectros NIR, seguiu-se com a escolha a faixa espectral de trabalho. Neste momento realizou-se uma análise prévia dos espectros e chegou-se a conclusão de que era possível trabalhar com a faixa de varredura completa (9.000 a 4.000 cm^{-1}), pois os espectros não apresentavam regiões com ruídos excessivos que pudessem comprometer de forma significativa os desenvolvimentos dos modelos multivariados. Em seguida, os dados foram avaliados quanto à presença de amostras anômalas tanto em relação ao ruído espectral (matriz **X**) quanto a erros na determinação dos parâmetros estudados (vetor **y**), usando o gráfico de influência versus resíduo de Student (ADUSUMILLI et al., 2015). Após esta etapa, as amostras dos conjuntos de calibração e validação externa foram escolhidas utilizando-se o algoritmo SPXY (ARAUJO et al., 2001). Na etapa seguinte estudaram-se quais os pré-processamentos espectrais forneceram os melhores modelos, baseando-se em parâmetros de validação do modelo como RMSEC, RMSECV, r^2 , correlação, bias, slope, offset e número de variáveis espectrais utilizadas para a construção do modelo. Como pré-tratamentos espectrais, foram testados: correção de linha de base, MSC, EMSC, SNV, derivadas, suavização e a combinação de alguns destes. Em seguida, foram construídos modelos PLS-1 para cada parâmetro estudado individualmente. A capacidade preditiva dos modelos construídos foi avaliada com base em seus parâmetros de validação. Para avaliar a influência do pré-processamento na capacidade preditiva dos modelos, uma nova seleção de amostras de calibração e predição foi realizada.

Uma melhor compreensão das etapas empregadas na construção dos modelos multivariados pode ser feita com base no fluxograma apresentado na Figura 9.

Figura 9 - Fluxograma das etapas empregadas na construção dos modelos multivariados.



5. RESULTADOS E DISCUSSÃO

5.1 Amostras e Valores de Referência

A construção de um modelo de calibração multivariada tem como meta a obtenção de modelos capazes de prever com precisão as características e/ou propriedades de amostras desconhecidas. (BLANCO; VILLARROYA, 2002) Para isso é necessário que se tenha uma variabilidade em torno do valor (médio ou de referência) do parâmetro estudado. De acordo com a norma ASTM E 1655-05 (2012), que versa sobre as análises quantitativas multivariadas envolvendo infravermelho, o conjunto de calibração deve representar todas as variações químicas ou físicas normalmente encontradas na análise de rotina da amostra estudada. Além disso, as amostras do conjunto de calibração devem abranger exemplos de todos componentes químicos que se espera que estejam presentes nas amostras que devem ser analisadas usando o modelo, de forma a evitar a extrapolação do mesmo. Para atender a esta demanda foram produzidos em laboratório, 10 lotes, cada um contendo 5 amostras de quitosana, totalizando 50 amostras, variando-se o GD na faixa de 69,8% a 91,9% e a MM de 93,9 KDa a 434,9 KDa. De acordo com as Tabelas 4 e 5 e os gráficos (a) e (b) apresentados na Figura 10, pode-se notar que os parâmetros estudados têm uma variabilidade adequada para a construção de um modelo de calibração multivariada.

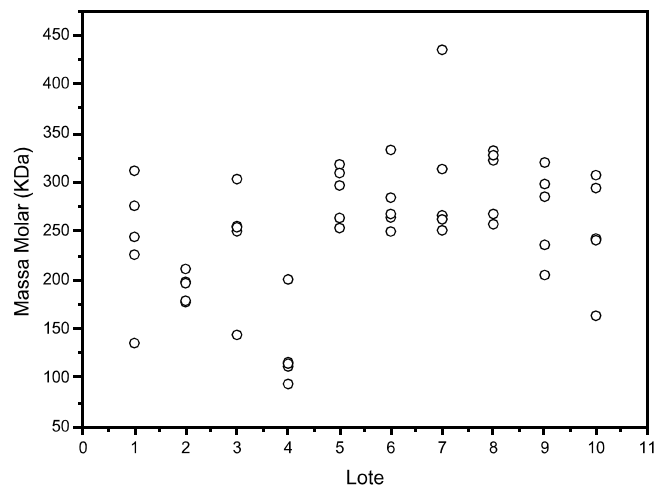
Tabela 4 - Valores de referência para o grau de desacetilação (GD) das 50 amostras de quitosana.

Lote	Amostras de quitosana				
	3 h	4 h	5 h	6 h	7 h
1	77,2%	86,9%	83,0%	82,4%	91,5%
2	78,3%	78,9%	80,0%	80,6%	81,1%
3	74,7%	69,8%	73,3%	75,1%	80,6%
4	79,5%	84,3%	85,0%	84,5%	84,7%
5	77,8%	84,5%	84,9%	83,7%	82,9%
6	77,7%	80,2%	86,5%	85,1%	84,7%
7	80,2%	80,8%	83,0%	89,3%	89,5%
8	88,0%	89,7%	89,8%	91,4%	91,9%
9	76,4%	77,7%	80,9%	83,8%	86,2%
10	76,5%	78,9%	80,4%	86,8%	87,7%

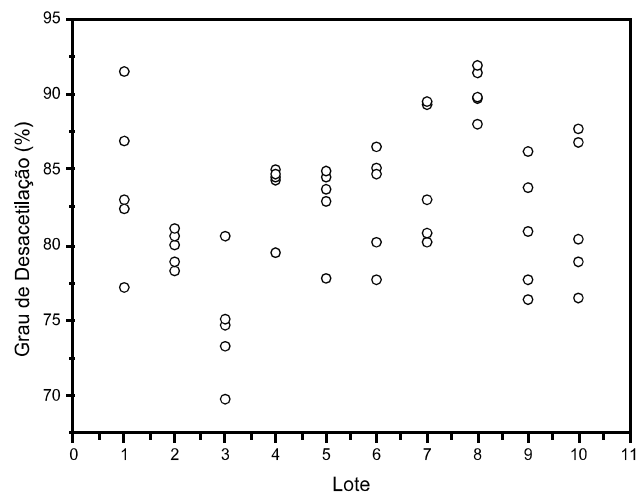
Tabela 5 - Valores de referência para a massa molar (MM) das 50 amostras de quitosana.

Lote	Amostras de quitosana				
	3 h	4 h	5 h	6 h	7 h
1	225,9 KDa	311,6 KDa	243,9 KDa	275,8 KDa	135,6 KDa
2	198,3 KDa	211,4 KDa	196,8 KDa	177,5 KDa	178,9 KDa
3	302,9 KDa	255,0 KDa	249,7 KDa	253,8 KDa	144,1 KDa
4	200,5 KDa	116,2 KDa	111,8 KDa	114,6 KDa	93,9 KDa
5	318,0 KDa	309,1 KDa	296,4 KDa	263,3 KDa	253,0 KDa
6	263,7 KDa	267,4 KDa	249,4 KDa	332,7 KDa	284,2 KDa
7	434,9 KDa	266,0 KDa	313,2 KDa	261,9 KDa	250,7 KDa
8	332,0 KDa	322,1 KDa	327,3 KDa	256,9 KDa	267,4 KDa
9	320,0 KDa	297,9 KDa	285,1 KDa	235,9 KDa	205,2 KDa
10	293,8 KDa	306,9 KDa	242,3 KDa	240,7 KDa	163,6 KDa

Figura 10 - Gráficos da faixa de variação da (a) MM e (b) do GD em relação aos lotes de quitosana produzidos.



(a)



(b)

Com relação aos resultados da MM, observou-se que estes se mostraram dispersos entre as três faixas de massas molares da quitosana (baixa, média e alta), o que contribuiu de forma positiva para a construção dos modelos multivariados. Contudo, a faixa de massa molar média (190 - 310 KDa) foi contemplada pela maioria das amostras, confirmando os resultados apresentados por Antonino (2016) que classificou a quitosana produzida no CERTBIO como sendo de massa molar média.

Outro ponto relevante é o valor apresentado pela amostra 3 h do lote 7 (434,9 KDa), que se destacou dos demais por está situado aproximadamente 100 KDa acima do segundo maior valor. Esse comportamento provavelmente está associado à formação excessiva de espuma no ensaio de viscosimetria por parte desta amostra, a qual comprometeu a repetibilidade dos tempos de escoamento, inclusive dos ensaios subsequentes necessários à confirmação do resultado. Porém, mesmo se destacando das demais amostras, de acordo com a avaliação estatística para a construção dos modelos multivariados ela não foi considerada uma amostra anômala.

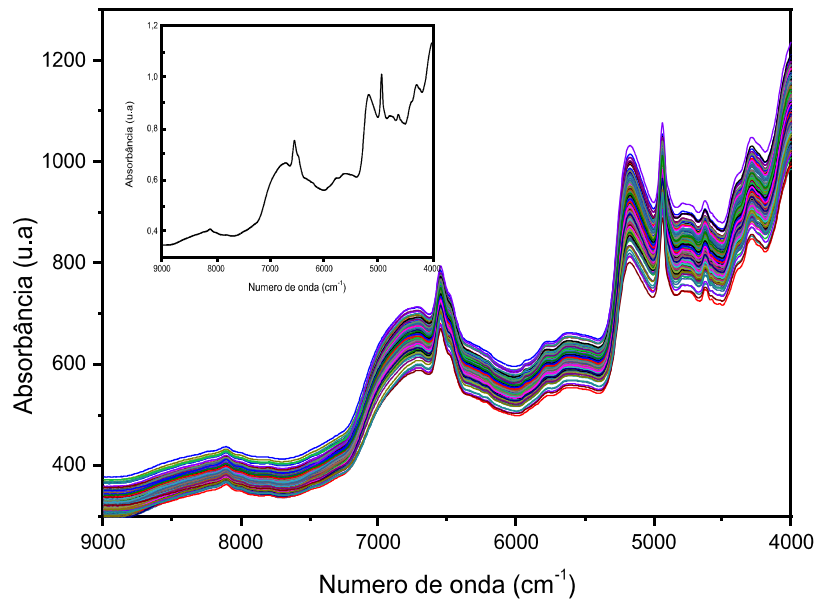
Analisando-se os valores de GD para as amostras de quitosana, percebe-se que poucos valores se apresentaram em consenso com o GD definido no estudo de Antonino (2016) como padrão para as quitosanas produzidas no CERTBIO ($GD > 90\%$). Este fato pode ser atribuído às variações ocorridas no processo de produção como: variações de temperatura, não homogeneidade do meio reacional, pontos de coletas das amostras e diferentes equipamentos utilizados. Além disso, alguns lotes apresentaram relativo escurecimento sugerindo que tenham sofrido degradações durante a produção, o que pode ter levado a alterações significativas das suas propriedades, refletindo inclusive na MM e no GD.

Sobre as prováveis mudanças que podem ocorrer na produção da quitosana, Kumirska et al. (2010) e Weinhold et al. (2009) relataram que diferenças no processo de produção, como por exemplo, temperatura, concentração alcalina, razão entre soluções alcalinas e matérias-primas, também podem contribuir para que a composição final do produto resulte em uma mistura de quitosanas, variando em MM e GD. Esses produtos podem inclusive conter impurezas, tais como metais pesados, resíduos proteicos e resíduos ácidos/alcalinos. Porém, as variações aleatórias apresentadas na produção das amostras não representaram problemas na construção dos modelos multivariados, uma vez que essas variações, aliadas as modificações intencionais no tempo de desacetilação, contribuíram para a construção de modelos mais robustos.

5.2 Construção dos modelos de calibração multivariada

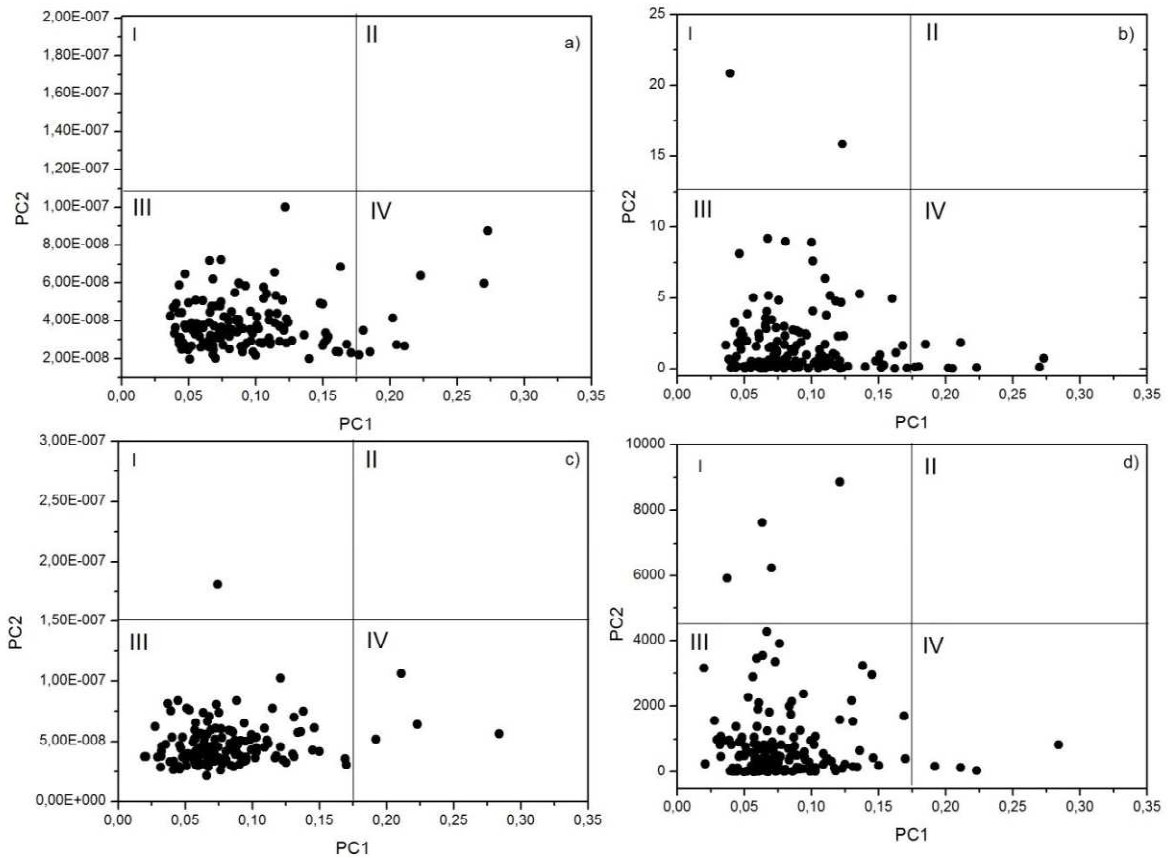
A Figura 11 apresenta o gráfico dos espectros NIR originais (sem pré-processamento) das 50 amostras de quitosana, em triplicata e, o detalhe para o espectro médio da triplicata de uma das amostras.

Figura 11 - Espectros NIR originais das 50 amostras de quitosana, em triplicata e, em detalhe, o espectro médio da triplicata de uma das amostras.



O gráfico de influência versus resíduo de Student foi utilizado para verificar a presença de amostras com comportamento anômalo tanto em relação aos espectros (X) quanto em relação ao método de referência (Y), Figura 12. De acordo com este estudo não houveram amostras anômalas, o que foi constatado pela ausência de amostras no quadrante II dos gráficos citados, ou seja, nenhuma amostra apresentou alta influência e alto resíduo simultaneamente.

Figura 12 - Gráfico de influência: (a) influência em X e (b) em Y, para o GD; (c) influência em X e (d) em Y, para a MM.



Outra etapa importante para a construção do modelo multivariado é a seleção dos conjuntos de calibração e validação, que neste trabalho foi feita com o auxílio do algoritmo SPXY (GALVÃO et al., 2005). Inicialmente, a seleção desses conjuntos foi realizada com os dados brutos e posteriormente os dados foram tratados para a construção dos modelos. Após a escolha do melhor pré-processamento, a seleção dos conjuntos de calibração e validação externa (ou predição) foi refeita para verificar a influência destes na seleção das amostras e consequentemente na qualidade do modelo. Foi avaliada qual a melhor estratégia a partir da comparação das figuras de mérito (Tabelas 6 e 7).

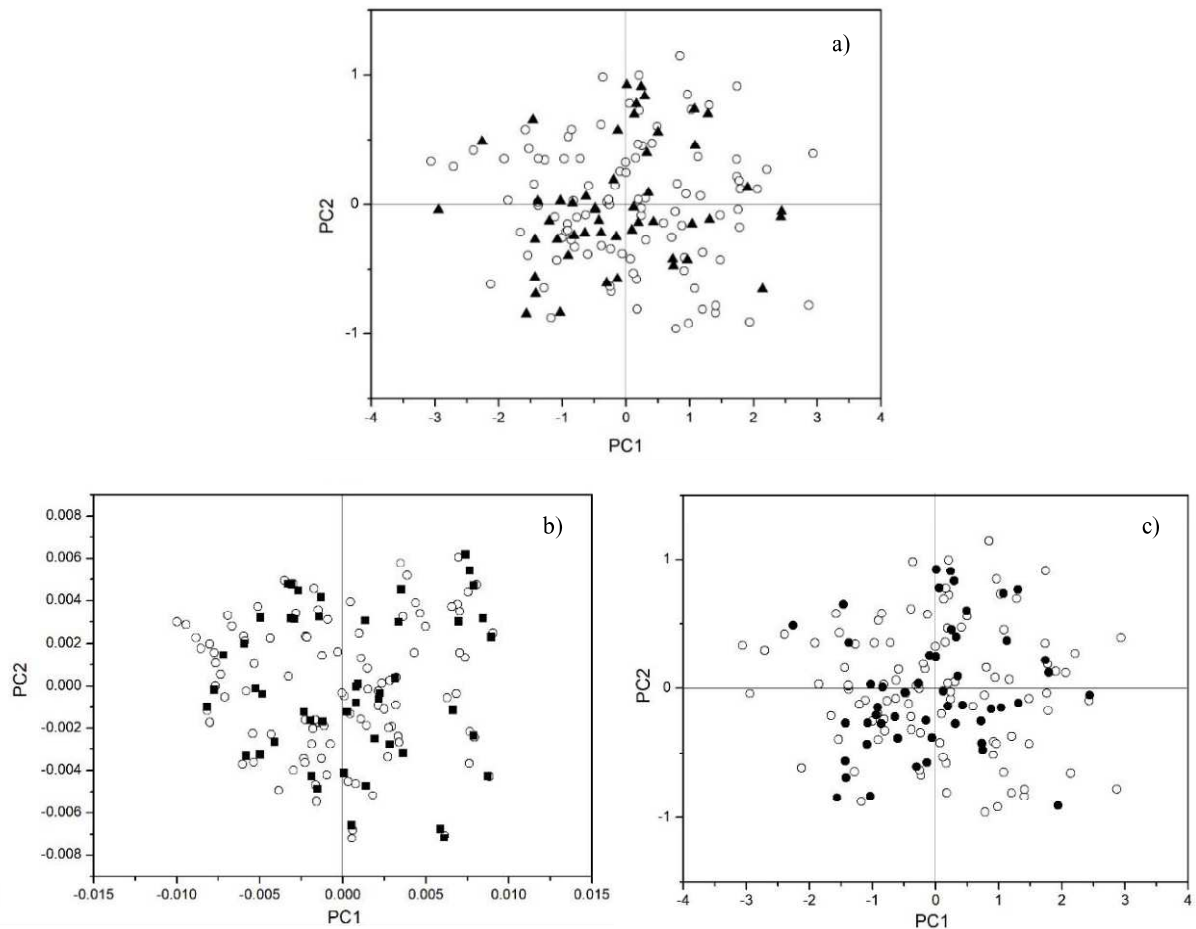
O número de amostras do conjunto de calibração para modelos baseados em NIRS, de acordo com a ASTM E 1655-05 (2012), depende da complexidade das amostras que estão sendo analisadas. Para amostras pouco complexas, que contêm poucos componentes variando em concentração, um conjunto de calibração relativamente pequeno é necessário para definir a relação entre as variáveis espectrais e a propriedade de interesse. Para amostras mais complexas, onde um maior número de componentes varia, é necessário um maior número de amostras de calibração para o desenvolvimento do modelo. A amostra estudada neste trabalho

pode ser considerada uma amostra simples em relação à composição. Deste modo, um pequeno número de amostras de calibração foi requerido para a construção dos modelos. No entanto, a confirmação de que o conjunto de amostras de calibração é ou não adequado para a construção dos modelos, só pode ser constatada após os mesmos terem sido desenvolvidos e a estimativa do número de variáveis latentes necessárias para estes, determinada.

Para estimar o número de variáveis latentes necessárias para a construção dos modelos, foi construído um modelo PLS preliminar utilizando os espectros brutos de todas as amostras. O número ideal de variáveis latentes necessárias para a construção deste modelo preliminar foi 11, o que pode estar relacionado a variabilidade das condições experimentais nas quais foram realizadas as sínteses. Neste caso, o conjunto de calibração deve conter 6 vezes o número de variáveis latentes (VL), o que indica que o número mínimo de amostras de calibração seria 66. No caso de modelos centrados na média, como os desenvolvidos neste trabalho, o número de amostras de calibração deve ser $6(VL+1)$, que levaria a um número mínimo de 72 amostras. Neste trabalho, o conjunto de calibração foi composto por 100 espectros e o de validação externa por 50 espectros, selecionados pelo algoritmo SPXY.

A Figura 13 mostra o gráfico de *scores* para os 150 espectros das amostras de quitosana para os dados brutos e pré-processados. Nela pode-se observar como ocorreu a distribuição das amostras de calibração (○) e de validação externa (▲ ou ● ou ■) selecionadas com o auxílio do algoritmo SPXY. Este algoritmo tem como objetivo selecionar as amostras de forma uniforme ao longo do espaço definido pelas mesmas, garantido a robustez do modelo, e evitar seleção de valores extremos, prevenindo extrapolações na etapa de predição.

Figura 13 - Gráfico de *scores* para os 150 espectros das amostras de quitosana (a) para os dados brutos, (b) para os dados derivados e (c) para EMSC.



Em relação aos espectros da Figura 11, a princípio não foi possível notar a presença de ruído espectral. No entanto, ao se observar o espectro médio em destaque, notou-se a presença de deslocamento de linha de base no eixo Y, o qual poderia comprometer os modelos construídos a partir dos espectros brutos. Sendo necessária a utilização de pré-processamentos espectrais que pudessem avaliar o impacto dessa anomalia na construção dos modelos. Neste trabalho, além correção de linha de base, foram estudados outros pré-processamentos espectrais como MSC, EMSC, SNV, derivadas e a combinação de alguns destes. A escolha do melhor pré-processamento foi realizada com base nos valores de RMSEC, RMSECV, r^2 , correlação, bias, slope, offset e número de variáveis utilizadas para a construção do modelo, descritas nas Tabelas 6 e 7.

Outra estratégia para melhorar a performance preditiva é utilizar ferramentas para a seleção das variáveis que estão mais relacionadas ao parâmetro de interesse. Neste sentido, foram comparadas as performances de modelos construídos com todas as variáveis espectrais

e modelos construídos com as variáveis selecionadas pelo teste de incerteza de Martens, utilizando o algoritmo Jack-Knife.

Os melhores pré-processamentos foram a primeira derivada com polinômio de primeira ordem e janela de 5 pontos (D1-J5-POL1), a primeira derivada com polinômio de segunda ordem e janela de 5 pontos (D1-J5-POL2) e 9 pontos (D1-J9-POL2) para o GD. Já para a MM os melhores pré-processamentos foram o EMSC e a combinação do MSC com a suavização Savitzky-Golay com janela de 5 pontos.

Foi verificado que, de forma geral, os modelos construídos com os coeficientes Jack-Knife apresentaram melhor capacidade preditiva que aqueles construídos com todas as variáveis espectrais. Outro aspecto importante observado para tais modelos é que eles necessitam de menor número de variáveis latentes para sua construção. O que indica que os modelos construídos (para o fim desejado) com os coeficientes Jack-Knife são mais parcimoniosos que aqueles construídos com todas variáveis espectrais.

Diante do exposto, outra técnica de seleção de variáveis foi aplicada empregando-se o algoritmo das projeções sucessivas (SPA) e a capacidade preditiva dos modelos construídos com as variáveis selecionadas pela mesma foi comparada a dos modelos construídos com os coeficientes Jack-Knife.

Tabela 6 – Pré-processamentos espectrais e figuras de mérito para o GD.

Pré-processamento	VL	Etapa	Figuras de Mérito (GD)						
			n	RMSE	Correlação	R ²	Slope	Offset	Bias
Nenhum	13	Cal	100	1,24	0,9691	0,9393	0,9393	5,00	3,532e-05
		CV	50	2,71	0,8424	0,7135	0,7457	20,99	0,0348
BO	12	Cal	100	1,31	0,9653	0,9318	0,9318	5,62	-2,678e-05
		CV	50	2,71	0,8417	0,7130	0,7362	21,79	0,0554
BO (JK)	10	Cal	100	1,36	0,9623	0,9623	0,9261	6,09	-2,747e-06
		CV	50	1,85	0,9295	0,8664	0,8750	10,33	0,0305
LBC	14	Cal	100	1,03	0,9787	0,9579	0,9579	3,46	-7,019e-06
		CV	50	2,71	0,8427	0,7135	0,7514	20,49	0,0064
LBC (JK)	9	Cal	100	1,46	0,9564	0,9145	0,9147	7,03	-9,155e-06
		CV	50	1,86	0,9291	0,8659	0,8705	10,67	0,0088
LBC+BO	14	Cal	100	1,03	0,9788	0,9581	0,9581	3,45	1,038e-05
		CV	50	2,71	0,8425	0,7136	0,7474	20,84	0,0300
LBC+BO (JK)	8	Cal	100	1,54	0,9517	0,9058	0,9058	7,76	-1,335e-05
		CV	50	1,92	0,9243	0,8570	0,8613	11,44	0,0148
MSC	11	Cal	100	1,38	0,9608	0,9232	0,9232	6,32	3,525e-05
		CV	50	2,75	0,8380	0,7062	0,7382	21,62	0,0466
MSC (JK)	8	Cal	100	1,57	0,9494	0,9013	0,9013	8,13	-5,035e-06
		CV	50	1,99	0,9179	0,8454	0,8550	11,94	-0,0019
EMSC	11	Cal	100	1,26	0,9680	0,9371	0,9371	5,18	-4,051e-05
		CV	50	2,73	0,8398	0,7089	0,7460	20,93	0,0046
EMSC (JK)	8	Cal	100	1,50	0,9543	0,9107	0,9107	7,36	-3,052e-05
		CV	50	1,94	0,9224	0,8535	0,8665	11,00	0,0040
SNV	11	Cal	100	1,40	0,9601	0,9219	0,9219	6,44	1,030e-05
		CV	50	2,76	0,8359	0,7025	0,7364	21,78	0,0588
SNV (JK)	8	Cal	100	1,60	0,9475	0,8977	0,8977	8,43	3,662e-06
		CV	50	2,03	0,9145	0,8393	0,8520	12,21	0,0084
D1-J5-POL1	5	Cal	100	1,32	0,9647	0,9307	0,9307	5,71	2,441e-06
		CV	50	2,54	0,8631	0,7496	0,7242	22,71	-0,0170
D1-J5-POL1 (JK)	4	Cal	100	1,54	0,9515	0,9054	0,9054	7,79	1,907e-06
		CV	50	1,83	0,9314	0,8699	0,8507	12,29	-0,0125
D1-J11-POL1	6	Cal	100	1,23	0,9692	0,9394	0,9394	4,99	2,441e-06
		CV	50	2,64	0,8502	0,7282	0,7335	21,99	0,0313
D1-J11-POL1 (JK)	5	Cal	100	1,50	0,9541	0,9104	0,9104	7,38	2,441e-06
		CV	50	1,92	0,9238	0,8564	0,8570	11,78	-5,676e-05
D1-J5-POL2	5	Cal	100	1,32	0,9647	0,9307	0,9307	5,71	2,441e-06
		CV	50	2,54	0,8632	0,7496	0,7242	22,71	-0,0170
D1-J5-POL2 (JK)	4	Cal	100	1,54	0,9515	0,9054	0,9054	7,79	1,907e-06
		CV	50	1,83	0,9314	0,8699	0,8507	12,29	-0,0125
D1-J9-POL2	5	Cal	100	1,60	0,9476	0,8980	0,8980	8,41	2,365e-06
		CV	50	2,64	0,8507	0,7288	0,7078	24,08	0,0032
D1-J9-POL2 (JK)	4	Cal	100	1,80	0,9331	0,8708	0,8708	10,65	1,755e-06
		CV	50	2,06	0,9121	0,8352	0,8212	14,72	-0,0068
SG-J5-POL1+MSC	13	Cal	100	1,26	0,9678	0,9367	0,9367	5,22	3,075e-05
		CV	50	2,83	0,8284	0,6879	0,7431	21,17	0,0025
SG-J5-POL1+MSC (JK)	9	Cal	100	1,70	0,9406	0,8847	0,8847	9,50	7,050e-05
		CV	50	2,15	0,9038	0,8201	0,8338	13,69	-0,0044

VL: Variáveis Latentes ou espectrais; **Cal:** Etapa de Calibração; **CV:** Validação Cruzada; **n:** Número de amostras por conjuntos de calibração e validação, respectivamente; **JK:** modelo construído com os coeficientes Jack-Knife; **BO:** Baseline Offset; **LBC:** Linear Baseline Correction; **D1-J5-POL1:** 1ª Derivada-Janela de 5 Pontos-Polinômio de 1ª Ordem; **D1-J11-POL1:** 1ª Derivada-Janela de 11 Pontos-Polinômio de 1ª Ordem; **D1-J5-POL2:** 1ª Derivada-Janela de 5 Pontos-Polinômio de 2ª Ordem; **D1-J9-POL2:** 1ª Derivada-Janela de 9 Pontos-Polinômio de 2ª Ordem; **SG-J5-POL+MSC:** Suavização (Savitzky-Golay)-janela de 5 pontos combinada ao MSC.

Tabela 7 – Pré-processamentos espectrais e figuras de mérito para a MM.

Pré-processamento	VL	Figuras de Mérito (MM)							
		Etapa	n	RMSE	Correlação	R ²	Slope	Offset	Bias
Nenhum	11	Cal	100	30,16	0,8961	0,8030	0,8030	49,10	0,0006
		Val	50	42,11	0,7883	0,6210	0,6797	80,46	0,6461
BO	11	Cal	100	23,97	0,9301	0,8651	0,8651	33,93	9,750e-05
		Val	50	43,82	0,7520	0,5581	0,6601	86,94	1,4719
BO (JK)	9	Cal	100	25,40	0,9211	0,8484	0,8484	38,12	-0,0005
		Val	50	34,69	0,8493	0,7230	0,7741	57,37	0,5608
LBC	8	Cal	100	35,24	0,8416	0,7084	0,7084	73,34	5,096e-05
		Val	50	44,33	0,7392	0,5476	0,6124	98,10	0,6257
LBC (JK)	8	Cal	100	32,30	0,8689	0,7549	0,7549	61,63	7,767e-05
		Val	50	40,44	0,7878	0,6236	0,6737	82,76	0,7061
LBC+BO	9	Cal	100	30,56	0,8835	0,7806	0,7806	55,16	8,377e-05
		Val	50	43,54	0,7515	0,5637	0,6385	92,14	1,2179
LBC+BO (JK)	9	Cal	100	26,90	0,9111	0,8301	0,8301	42,74	0,0002
		Val	50	35,82	0,8372	0,7046	0,7404	65,90	0,6184
MSC	10	Cal	100	26,40	0,9145	0,8363	0,8363	41,17	-0,0002
		Val	50	42,25	0,7686	0,5890	0,6674	84,40	0,7531
MSC (JK)	8	Cal	100	26,60	0,9132	0,8339	0,8339	41,78	-0,0004
		Val	50	32,66	0,8662	0,7544	0,7750	56,82	0,2374
EMSC	7	Cal	100	33,61	0,8571	0,7347	0,7347	66,72	9,323e-05
		Val	50	43,65	0,7490	0,5613	0,6286	94,61	1,2083
EMSC (JK)	6	Cal	100	32,41	0,8679	0,7533	0,7533	62,03	-6,317e-05
		Val	50	37,92	0,8152	0,6690	0,7026	75,55	0,7619
SNV	14	Cal	100	8,85	0,9907	0,9816	0,9816	4,63	0,0002
		Val	50	39,00	0,8082	0,6498	0,7354	67,28	0,7312
SNV (JK)	8	Cal	100	23,11	0,9352	0,8746	0,8746	31,55	-7,553e-05
		Val	50	29,89	0,8892	0,7944	0,8107	47,50	-0,1123
D1-J5-POL1	7	Cal	100	9,54	0,9893	0,9786	0,9786	5,37	-4,272e-06
		Val	50	41,00	0,7789	0,61,30	0,5775	107,03	0,7666
D1-J5-POL1 (JK)	6	Cal	100	11,69	0,9838	0,9679	0,9679	8,08	2,136e-06
		Val	50	21,27	0,9454	0,8958	0,8914	27,34	0,0371
D1-J11-POL1	7	Cal	100	13,19	0,9793	0,9591	0,9591	10,28	-2,167e-05
		Val	50	40,92	0,7791	0,6146	0,6122	98,39	0,8730
D1-J11-POL1 (JK)	6	Cal	100	19,01	0,9566	0,9151	0,9151	21,34	-1,068e-05
		Val	50	26,28	0,9153	0,8410	0,8411	40,02	0,0592
D1-J5-POL2	7	Cal	100	9,54	0,9893	0,9786	0,9786	5,37	-4,272e-06
		Val	50	41,00	0,7789	0,61,30	0,5775	107,03	0,7666
D1-J5-POL2 (JK)	6	Cal	100	11,69	0,9838	0,9679	0,9679	8,08	2,136e-06
		Val	50	21,27	0,9454	0,8958	0,8914	27,34	0,0371
D1-J9-POL2	6	Cal	100	14,31	0,9756	0,9519	0,9519	12,10	7,629e-06
		Val	50	40,39	0,7859	0,6246	0,5993	101,56	0,7852
D1-J9-POL2 (JK)	6	Cal	100	18,45	0,9592	0,9200	0,9200	20,11	2,350e-05
		Val	50	26,81	0,9117	0,8346	0,8370	41,05	0,0696
SG-J5-POL1+MSC	10	Cal	100	28,01	0,9032	0,8158	0,8158	46,33	0,0003
		Val	50	42,15	0,7695	0,5911	0,6663	84,71	0,7790
SG-J5-POL1+MSC (JK)	7	Cal	100	29,24	0,8940	0,7992	0,7992	50,49	0,0003
		Val	50	34,65	0,8480	0,7236	0,7470	63,91	0,2820

VL: Variáveis Latentes ou espectrais; **Cal:** Etapa de Calibração; **CV:** Validação Cruzada; **n:** Número de amostras por conjuntos de calibração e validação, respectivamente; **JK:** modelo construído com os coeficientes Jack-Knife; **BO:** Baseline Offset; **LBC:** Linear Baseline Correction; **D1-J5-POL1:** 1ª Derivada-Janela de 5 Pontos-Polinômio de 1ª Ordem; **D1-J11-POL1:** 1ª Derivada-Janela de 11 Pontos-Polinômio de 1ª Ordem; **D1-J5-POL2:** 1ª Derivada-Janela de 5 Pontos-Polinômio de 2ª Ordem; **D1-J9-POL2:** 1ª Derivada-Janela de 9 Pontos-Polinômio de 2ª Ordem; **SG-J5-POL+MSC:** Suavização (Savitzky-Golay)-janela de 5 pontos combinada ao MSC.

As Tabelas 8 e 9, mostram algumas figuras de mérito para os modelos PLS construídos com os pré-processamentos que apresentaram as melhores performances preditivas para a quantificação do GD e MM, levando em conta além das figuras de mérito a parcimônia, ou seja a construção de modelos mais simples e igualmente precisos e exatos. Estas tabelas também apresentam os resultados obtidos para os modelos usando as variáveis selecionadas pelo algoritmo SPA. Nestes modelos foi feita a seleção dos conjuntos de calibração e validação externa após o pré-processamento.

De acordo com os resultados destas tabelas, podemos constatar que, de forma geral, os modelos construídos com as variáveis selecionadas pelo SPA apresentaram erros de calibração e predição superiores aos modelos PLS construídos com os coeficientes Jack-Knife. No caso da massa molar, o Jack-Knife foi aplicado duas vezes. Os modelos que apresentaram melhor performance foram aqueles construídos com a seleção de amostras de calibração e predição realizada antes da aplicação do pré-processamento espectral, no entanto devido ao erro apresentado na etapa de predição (29,08 KDa) este desempenho não se mostrou suficientemente satisfatório.

Tabela 8 - Figuras de mérito para a GD empregando-se os algoritmos Jack-Knife e SPA, respectivamente.

Pré-processamento	Método para seleção de variáveis	VL	Etapa	RMSE	Correlação	R ²	Slope	Offset	Bias
D1-J5-POL1	-	6	Cal	1,06	0,9759	0,9525	0,9525	3,92	-3,81e-07
			Val Int.	2,35	0,8769	0,7721	0,7353	21,82	-0,0088
			Val Ext.	2,30	0,8991	0,8000	0,7553	20,58	0,3400
D1-J5-POL1	JK	4	Cal	1,34	0,9616	0,9247	0,9247	6,21	-3,05e-07
			Val Int.	1,61	0,9444	0,8934	0,8674	10,92	-0,0152
			Val Ext.	2,21	0,9054	0,8146	0,8097	9,76	0,4201
D1-J5-POL1	SPA	9	Cal	2,81	-	-	-	-	-
			Val Int.	-	-	-	-	-	-
			Val Ext.	3,31	0,7717	-	-	-	-0,1666
D1-J5-POL1 (JK)*	JK	4	Cal	1,54	0,9515	0,9054	0,9054	7,79	1,907e-06
			Val Int.	1,83	0,9314	0,8699	0,8507	12,29	-0,0125
			Val Ext.	1,85	0,9267	0,8549	0,8032	16,41	0,0917
D1-J5-POL1*	SPA	4	Cal	2,80	-	-	-	-	-
			Val Int.	-	-	-	-	-	-
			Val Ext.	3,73	0,6522	-	-	-	-0,04223

VL: Variáveis Latentes ou espectrais; **Cal:** Etapa de Calibração; **Val Int.:** Etapa de Interna; **Val Ext.:** Etapa de Externa; **JK:** modelo construído com os coeficientes Jack-Knife; **D1-J5-POL1:** 1ª Derivada-Janela de 5 Pontos-Polinômio de 1ª Ordem; **D1-J5-POL1 (JK)*** e **D1-J5-POL1*:** modelos construídos com amostras selecionadas antes da aplicação do pré-processamento.

Tabela 9 - Figuras de mérito para a MM empregando-se os algoritmos Jack-Knife e SPA, respectivamente.

Pré-processamento	Método para seleção de variáveis	VL	Etapa	RMSE	Correlação	R ²	Slope	Offset	Bias
EMSC	-	8	Cal	32,87	0,8096	0,7561	0,7561	60,55	2,65-05
			Val int	42,34	0,7769	0,6036	0,6729	81,75	0,5195
			Val ext	46,73	0,7566	0,5615	0,6459	91,53	2,69
EMSC	JK	7	Cal	34,45	0,8215	0,7325	0,7322	66,48	-0,0013
			Val int	38,02	0,8557	0,6803	0,7020	74,05	0,0448
			Val ext	48,52	0,7415	0,5287	0,6347	97,95	6,3037
EMSC	SPA	24	Cal	29,20	-	-	-	-	-
			Val int	49,06	0,7259	-	-	-	-
			Val ext	49,86	0,7529	-	-	-	-
EMSC (JK)*	JK	6	Cal	32,41	0,8679	0,7533	0,7533	62,03	-6,31e-05
			Val int	37,92	0,8152	0,6690	0,7026	75,55	0,7619
			Val ext	29,08	0,7825	0,5943	0,6822	70,12	2,24
EMSC*	SPA	48	Cal	29,20	-	-	-	-	-
			Val int	-	-	-	-	-	-
			Val ext	49,06	0,7865	-	-	-	-8,6867

VL: Variáveis Latentes ou espectrais; **Cal:** Etapa de Calibração; **Val Int.:** Etapa de Interna; **Val Ext.:** Etapa de Externa; **JK:** modelo construído com os coeficientes Jack-Knife; **EMSC (JK)* e EMSC*:** modelos construídos com amostras selecionadas antes da aplicação do pré-processamento.

Levando em conta figuras de mérito como RMSE, r , R^2 , número de variáveis utilizadas, que descrevem a performance preditiva, o modelo D1-J5-POL1-JK foi considerado mais adequado para a predição do parâmetro GD na quitosana produzida para fins biomédicos pelo CERTBIO. Os modelos D1-J9-POL2-JK apresentaram valores de RMSEC, RMSECV, RMSEP e VL comparáveis aos apresentados pelo modelo D1-J5-POL2-JK, no entanto o segundo apresentava melhor correlação entre os espectros e o valor de referência (r), melhor coeficiente de determinação (R^2) e valor de slope mais próximo de 1.

De acordo com o princípio da parcimônia, para o GD foi escolhido como melhor modelo aquele construído com os dados derivados com janela de 5 pontos e polinômio de primeira ordem (D1-J5-POL1) e para a MM aqueles pré-processados com o EMSC.

Comparando-se a performance dos modelos construídos realizando-se a seleção de conjuntos de calibração e predição antes ou após a aplicação do pré-processamento espectral (Tabelas 8 e 9), podemos verificar que a seleção dos conjuntos antes da aplicação do pré-processamento fornece modelos com menores valores de erros de predição. Além disso, este procedimento torna a metodologia mais simples para aplicação nas análises de rotinas de laboratórios.

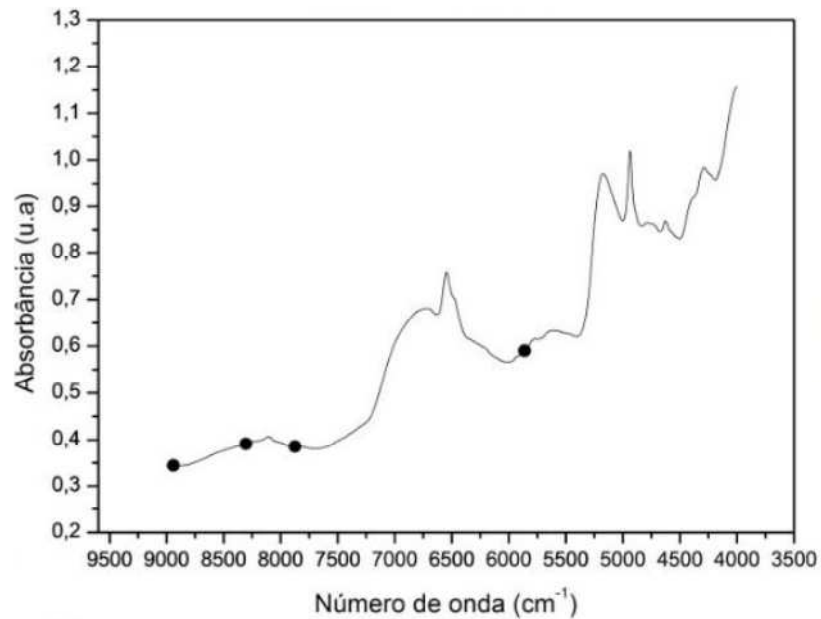
Deste modo, foram considerados como melhores os modelos construídos aplicando-se a seleção das amostras de calibração e predição antes de se pré-processar os dados espectrais. E os resultados apresentados e discutidos a seguir são relativos a estes modelos.

Os melhores pré-processamentos foram utilizados para o desenvolvimento dos modelos MLR/SPA. Os dados comparativos para a performance preditiva dos modelos PLS, construídos com todas as variáveis espectrais (full) e com os coeficientes Jack-Knife (JK), e do MLR/SPA são apresentados nas Tabelas 8 para o GD e 9 para o MM.

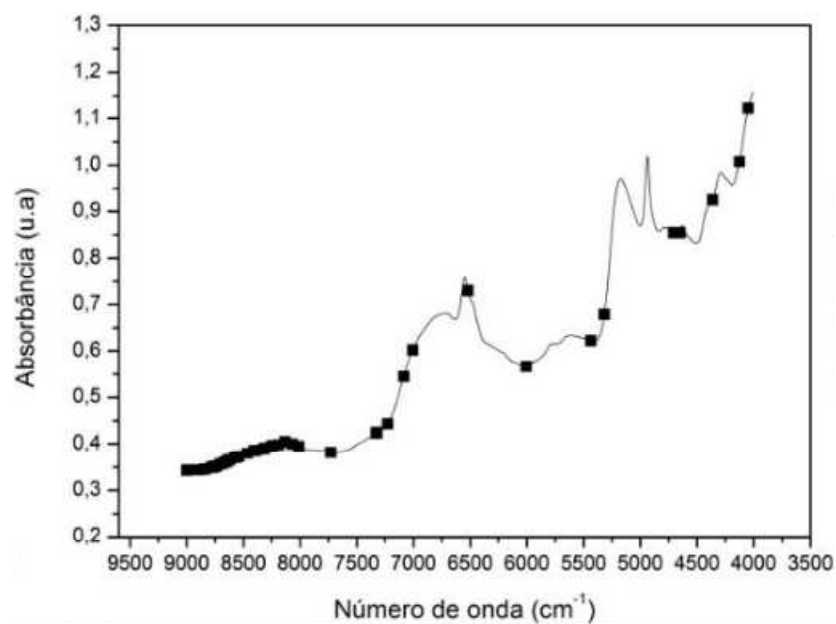
As variáveis selecionadas pelo algoritmo SPA para o modelo destinado a quantificação do GD, Figura 14 (a), foram 8.940 cm^{-1} , 8.302 cm^{-1} , 7.872 cm^{-1} e 5.860 cm^{-1} . Nenhuma das variáveis selecionadas conseguiu ser atribuída a bandas que pudessem estar diretamente relacionadas à conversão dos grupos N-acetil da cadeia polimérica em grupos amino primários e, conseqüentemente com o grau de desacetilação. No entanto, isto não comprometeu o desempenho dos modelos MLR/SPA para a determinação do GD, que apresentaram erros satisfatórios.

Não foi possível realizar a atribuição de bandas para a determinação da MM, pois esta característica está relacionada a parâmetros físicos como tamanho da cadeia polimérica, o que não tem significado químico capaz de estar relacionado a uma banda espectral. Deste modo, as variáveis espectrais selecionadas para a MM estão relacionadas a características físicas da amostra quitosana e do processo de fabricação da mesma. Pode-se observar que foram selecionados números de ondas nas extremidades do espectro, que podem estar relacionados a flutuação de linha de base ou espalhamento do sinal, devido as características físicas da amostra.

Figura 14 - Variáveis espectrais selecionadas pelo SPA para (a) GD e (b) MM.



(a)



(b)

As Figuras 15 e 16 mostram os gráficos dos valores preditos pelo modelo versus valores determinados pelo método de referência para o GD e MM, respectivamente, que demonstram a concordância do valor predito pelo modelo com os valores medidos pelo método de referência. Pode-se observar a distribuição aleatória das amostras em torno da bissetriz, o que mostra que não há tendências nos resultados. Também é possível notar para a MM (Figura 16), que existe uma amostra que apresenta um valor bem maior que as demais

amostras analisadas. No entanto este valor de concentração está contemplado no conjunto de calibração, assegurando que a predição da mesma não se deu por extrapolação do modelo.

Figura 15 - Gráfico dos valores preditos pelo modelo D1-J5-POL1-JK versus valores determinados pelo método de referência para o GD.

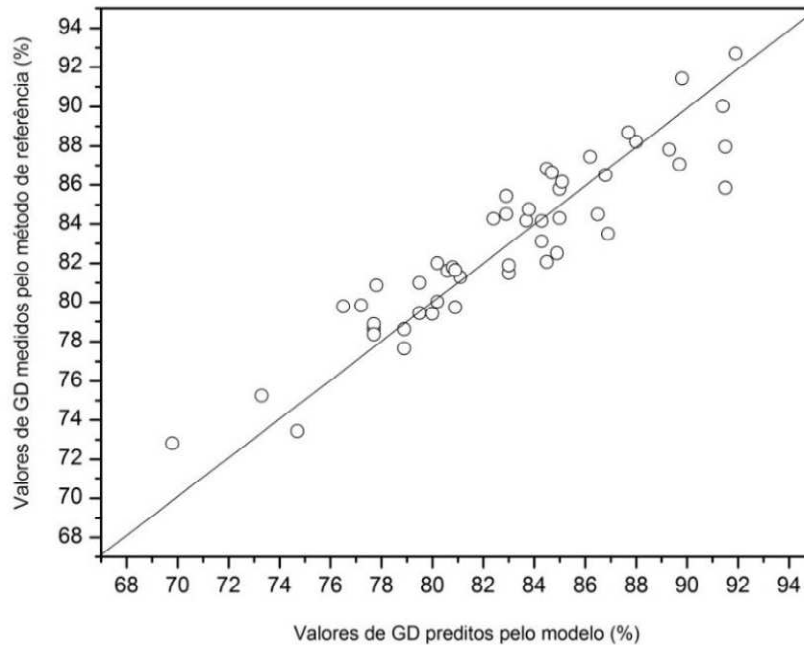
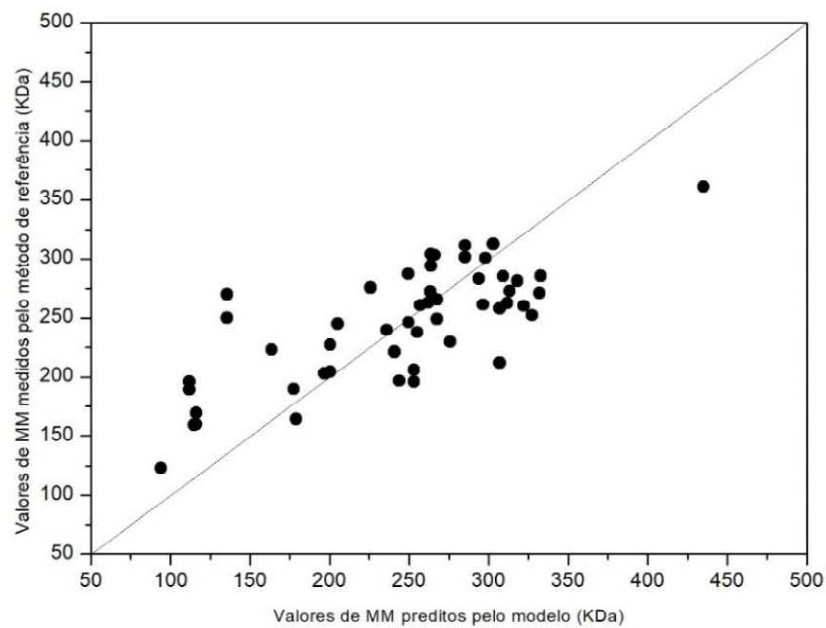


Figura 16 - Gráfico dos valores preditos pelo modelo EMSC-JK versus valores determinados pelo método de referência para a MM.



6 CONCLUSÕES

A espectroscopia NIR aliada a métodos quimiométricos se apresentou como uma boa alternativa. Para o desenvolvimento da metodologia proposta foi necessário a ampliação da faixa de variação dos parâmetros estudados, a fim de proporcionar mais robustez aos modelos. Diversos pré-processamentos foram testados e os que apresentam melhores performances ao serem aplicados aos dados espectrais para o GD foram as combinações: 1ª derivada-janela de 5 pontos-polinômio de primeira; 1ª derivada-janela de 9 pontos-polinômio de segunda. Já para a MM o pré-processamento que apresentou melhor desempenho foi o EMSC. Foi avaliada a influência do pré-processamento sobre a seleção dos conjuntos de calibração e predição, e verificou-se que a seleção desses conjuntos antes da aplicação do pré-processamento fornece modelos com menores erros de predição. Também foi avaliado se a construção de modelos usando algoritmos para seleção de variáveis melhoraria a performance dos modelos construídos com os dados brutos. Para isto comparou-se a capacidade preditiva de modelos construídos com todas as variáveis espectrais (full spectra), com os modelos construídos com os coeficientes de regressão gerados pelo teste de incerteza de Martens (Coeficientes Jack-Knife) e com aqueles construídos com as variáveis selecionadas pelo SPA. De forma geral os modelos construídos utilizando os coeficientes Jack-Knife apresentaram melhor performance preditiva que os full spectra e que os construídos com as variáveis selecionadas pelo SPA. Mostrando melhores valores de RMSE, menor número de variáveis latentes empregadas na construção do modelo, concordância entre os valores de RMSEC e RMSEP, bom ajuste e boa correlação entre os espectros e os valores de referência. O erros obtidos na etapa de predição foram 1,85% para o GD e 29,08 KDa para a MM. No caso do GD, o erro obtido está de acordo com os valores aceitáveis para o método de referência. (DASH et al., 2011) Já o modelo destinado à determinação da massa molar não mostrou desempenho satisfatório. Tendo em vista que a metodologia proposta para a determinação do parâmetro de qualidade GD, utilizando NIRS e PLS-JK, realiza determinações rápidas, pouco onerosas, pouco laboriosas, de fácil aplicabilidade em rotina e que apresenta erro aceitável, fica clara a viabilidade de sua aplicação na caracterização de quitosana para fins biomédicos produzida pelo Laboratório CERTBIO.

7 PERSPECTIVAS

- Testar a metodologia desenvolvida em quitosanas produzidas comercialmente;
- Realizar estudos semelhantes aos que foram desenvolvidos neste trabalho que avaliem o potencial da espectroscopia NIR e das técnicas quimiométricas frente a outras metodologias de referência para os parâmetros estudados;
- Realizar o estudo dos espectros MIR utilizados na determinação do GD pela metodologia de Brugnerotto et al. (2001), avaliando seu potencial para construção de modelos de calibração multivariada;
- Desenvolver modelos baseados na técnica quimiométrica de análise de componentes principais (PCA), visando emprego dos espectros MIR e NIR na classificação de amostras de quitosana quanto as suas faixas de massas molares.
- Avaliar o potencial dos modelos utilizados neste trabalho para o estudo de outros parâmetros de referência da quitosana descritos na Norma ASTM F2103-11 (2011), a exemplo da polidispersividade, viscosidade em solução aquosa, conteúdo de matéria seca, teores de cinzas e de insolúveis.
- Validar o melhor modelo PLS usando MVC1.

REFERÊNCIAS

ADUSUMILLI, S. et al. A novel hybrid approach utilizing Principal Component Regression and Random Forest Regression to bridge the period of GPS outages. **Neurocomputing**, v. 166, p. 185–192, out. 2015.

ANDERSEN, C. M.; BRO, R. Variable selection in regression — a tutorial. *J. Chemometr.* 728–737, 2010.

ANTONINO, N. A. **Otimização do processo de obtenção de quitina e quitosana de exoesqueletos de camarões oriundos da indústria pesqueira paraibana.** 88 f. Dissertação (Mestrado em Química Inorgânica) Universidade Federal da Paraíba. João Pessoa, 2007.

ANTONINO, R. S. C. M. DE Q. **Produção de Quitosana a Partir de Exoesqueleto de Camarão (*Litopenaeus vannamei boone*) para Aplicações Biomédicas.** p. 66, 2016. Dissertação de Mestrado. Universidade Federal de Campina Grande.

ARAUJO, M. C. U. et al. The successive projections algorithm for variable selection in spectroscopic multicomponent analysis. **Chemometrics and Intelligent Laboratory Systems**, v. 57, n. 2, p. 65–73, 2001.

ASTM. **E1655 - 05 Standard Practices for Infrared Multivariate Quantitative Analysis.** West Conshohocken, PA, USA: ASTM International 2012. Disponível em: <<https://www.astm.org/Standards/E1655.htm>>.

ASTM. **F2103-11 Standard Guide for Characterization and Testing of Chitosan Salts as Starting Materials Intended for Use in Biomedical and Tissue- Engineered Medical Product Applications.** West Conshohocken, PA, USA: ASTM International 2011. Disponível em: <<https://www.astm.org/Standards/F2103.htm>>.

AZUMA, K. et al. Chitin, Chitosan, and Its Derivatives for Wound Healing: Old and New Materials. **Journal of Functional Biomaterials**, v. 6, n. 1, p. 104–142, 13 mar. 2015.

BEEB, K. R.; RANDY, J. P.; SEASHOLTZ, M. B. **Chemometrics: A Practical Guide,** Wiley-Interscience, New York, 1998.

BERGER, L. R. R. **Produção e caracterização de quitina e quitosana por *Rhizopus arrhizus* e *Cunninghamella elegans* e aplicação em membranas na remoção de cádmio.** p. 180, 2013. Tese de Doutorado. Universidade Federal de Pernambuco.

BEZERRA, A. M. **Síntese e avaliações físico-químicas e biológicas de derivados de quitosana de alta e baixa massa molecular.** Dissertação de Mestrado. Universidade de São Paulo (USP), 2011.

BEZERRA, A. M. **Prospecção quantitativa e qualitativa de uma nova fonte renovável de quitosana.** Tese de Doutorado. Universidade de São Paulo (USP), 2015.

BLANCO, M.; VILLARROYA, I. NIR spectroscopy: a rapid-response analytical tool. **TrAC - Trends in Analytical Chemistry**, v. 21, n. 4, p. 240–250, 2002.

BRUGNEROTTO, J. et al. An infrared investigation in relation with chitin and chitosan characterization. **Polymer**, v. 42, n. 8, p. 3569–3580, 2001. ISSN 0032-3861.

CABRAL, A. **Construção de modelos multivariados para determinação de lipídios totais e umidade em leite em pó comercial utilizando espectroscopia no infravermelho próximo.** p. 104, 2011. Dissertação de Mestrado. Universidade Federal do Rio Grande do Norte.

CAMO S.A. Manual do Usuário “**The Unscrambler 9.7**”. Noruega, 1998.

DASH, M. et al. Statistical approach to the spectroscopic determination of the deacetylation degree of chitins and chitosans. **Carbohydrate Polymers**, v. 86, n. 1, p. 65–71, ago. 2011.

DENARI, N. S. M. **Biomateriais binários de quitosana/amido e quitosana/gelatina em L-ácido láctico.** p. 137, 2014. Dissertação de Mestrado. Universidade de São Paulo (USP), São Carlos-SP.

DOMSZY, J. G.; ROBERTS, G. A. Evaluation of infrared spectroscopic techniques for analysing chitosan. **Die Makromolekulare Chemie**, v. 186, n. 8, p. 1671-1677, 1985. ISSN 0025-116X.

FACCHINATTO, W. M. **Preparação e Caracterização de Nanosuspensões e Hidrogéis de N,O-metoxipoli(etilenoglicol)-g-quitosana para aplicação em Sistemas de Liberação de Fármacos Antitumorais.** p. 139, 2016. Dissertação de Mestrado. Universidade de São Paulo (USP), São Carlos-SP.

FEARN, T. et al. On the geometry of SNV and MSC. **Chemometrics and Intelligent Laboratory Systems**, v. 96, n. 1, p. 22–26, 15 mar. 2009.

FONSECA, A. C. M. **Processos de obtenção e caracterização físico-química de quitinas e**

quitosanas extráidas dos rejeitos da indústria da região de Cananéia - SP. Dissertação de Mestrado. Instituto de Pesquisas Energéticas e Nucleares (IPEN) - Autarquia associada à Universidade de São paulo (USP), 2016.

FONSECA, F. DOS S. “ **Espectrometria no Infravermelho Médio e Calibração Multivariada por PLS para quantificação de biodiesel metílico e etílico de soja , milho e girassol em misturas com diesel** ”. Dissertação de Mestrado. Universidade Federal de Uberlândia, 2011.

GALVÃO, R. K. H. et al. A method for calibration and validation subset partitioning. **Talanta**, v. 67, n. 4, p. 736–740, 2005.

GAMAL, R. F. et al. Optimization of chitin yield from shrimp shell waste by *Bacillus subtilis* and impact of gamma irradiation on production of low molecular weight chitosan. **International Journal of Biological Macromolecules**, v. 91, p. 598–608, 2016.

GELADI, P.; MACDOUGALL, D.; MARTENS, H. Linearization and Scatter-Correction for Near-Infrared Reflectance Spectra of Meat. **Applied Spectroscopy**, v. 39, n. 3, p. 491–500, 1985.

GOODARZI, M. et al. Multivariate calibration of NIR spectroscopic sensors for continuous glucose monitoring. **TrAC - Trends in Analytical Chemistry**, v. 67, p. 147–158, abr. 2015.

GRANGEIRO JÚNIOR, S. **Aplicação de Métodos Quimiométricos em Análises de Formas Farmacêuticas Sólidas Lamivudina (3TC) + Zidovudina (AZT) Baseados na Espectrofotometria de absorção na região do Ultravioleta e do Infravermelho próximo.** Tese de Doutorado. Universidade Federal de Pernambuco, 2011.

GUO, Q. et al. The robust normal variate transform for pattern recognition with near-infrared data. **Analytica Chimica Acta**, 382.1 (1999): 87-103.

HELLAND, I. S.; NAES, T.; ISAKSSON, T. Related versions of the multiplicative scatter correction method for preprocessing spectroscopic data. **Chemometrics and Intelligent Laboratory Systems**, v. 29, n. 2, p. 233–241, 1 out. 1995.

Laboratório de Avaliação e Desenvolvimento de Biomateriais do Nordeste – CERTBIO. **Procedimento Técnico: PT48 Produção de Quitosana a Partir de Exoesqueleto de Camarão – Rev. 01.** Campina Grande, PB, 2017.

LIZARDI-MENDOZA, J.; MONAL, W. M. A.; VALENCIA, F. M. G. Chemical

Characteristics and Functional Properties of Chitosan. In: BAUTISTA-BAÑOS, S.; ROMANAZZI, G.; JIMÉNEZ-APARICIO, A. (Ed.). **Chitosan in THE PRESERVATION OF AGRICULTURAL COMMODITIES**. 1. ed. [s.l.] Academic Press is an imprint of Elsevier, 2016. p. 384.

KUMIRSKA, J. et al. Application of spectroscopic methods for structural analysis of chitin and chitosan. **Marine Drugs**, v. 8, n. 5, p. 1567- 636, 2010. ISSN 1660-3397 (Electronic) 1660-3397 (Linking). Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/20559489> >.

MARTENS, H.; MARTENS, M. Modified Jack-knife estimation of parameter uncertainty in bilinear modelling by partial least squares regression (PLSR). **Food Quality and Preference**, v. 11, n. 1–2, p. 5–16, 2000.

NASCIMENTO, J.S.; BANDEIRA, M.G.A; MENDES, A.G.G. Estudo Prospectivo da Quitosana : Análise Atráves dos Pedidos de Patentes e Artigos Científicos. **VII International Syposium on Technological Innovation**. Aracajú, SE, Brasil, 2016.

MUZZARELLI, R. A.; ROCCHETTI, R. Determination of the degree of acetylation of chitosans by first derivative ultraviolet spectrophotometry. **Carbohydrate Polymers**, v. 5, n. 6, p. 461-472, 1985. ISSN 0144-8617.

NGO, D.-N. D.-H. et al. Biological effects of chitosan and its derivatives. **Food Hydrocolloids**, v. 51, p. 200-216, 10// 2015. ISSN 0268-005X. Disponível em: < <http://www.sciencedirect.com/science/article/pii/S0268005X15002246> >.

OLIVEIRA, F. C. C.; **Modelos de Calibração Multivariada Associados à Espectroscopia Vibracional para Análise de Misturas Diesel - Óleos Vegetais**. p. 120, 2006. Dissertação de Mestrado. Universidade de Brasília.

OLIVEIRA, M. **O Biopolímero Quitosana, Modificado Quimicamente ou Reticulado com Metais, em Forma de Pó ou Esfera, Aplicado no Estudo Termoquímico da Interação com Cobre e Aminas Alifáticas**. Tese de Doutorado. Universidade Estadual de Campinas (UNICAMP), 2011.

PRASHANTH, K. V. H.; THARANATHAN, R. N. Chitin/chitosan: modifications and their unlimited application potential-an overview. **Trends in Food Science and Technology**, v. 18, n. 3, p. 117–131, 2007.

PENICHE, C.; BECHERÁN, L. Quitosano como plataforma tecnológica en productos biofarmacéuticos y aplicaciones dérmicas. In: SOUSA, H. C. D.; BRAGA, M. E., et al (Ed.).

Biomateriais aplicados ao desenvolvimento de sistemas terapêuticos avançados, 2015. cap. 3, p.111-146. ISBN 9892608801.

PEPINO, R. DE O. **Desenvolvimento de géis e esponjas de quitosana e blendas quitosana/gelatina em ácido adípico**. Dissertação de Mestrado. Universidade de São Paulo (USP), 2016.

PILLAI, C. K. S. K. S.; PAUL, W.; SHARMA, C. P. Chitin and chitosan polymers: Chemistry, solubility and fiber formation. **Progress in Polymer Science (Oxford)**, v. 34, n. 7, p. 641–678, jul. 2009. ISSN 0079-6700.

PIRES, G. **Biomateriais Derivados de Quitosana e Hidroxiapatita com Potencial para Preenchimento Ósseo**. p. 88, 2010. Tese de Doutorado. Universidade Estadual de Campinas (UNICAMP).

RAAFAT, D.; SAHL, H. G. Chitosan and its antimicrobial potential - A critical literature survey. **Microbial Biotechnology**, v. 2, n. 2 SPEC. ISS., p. 186–201, 2009.

RINAUDO, M.; MILAS, M.; LE DUNG, P. Characterization of chitosan. Influence of ionic strength and degree of acetylation on chain expansion. **International Journal of Biological Macromolecules**, v. 15, n. 5, p. 281-285, 1993. ISSN 0141-8130.

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, 1 nov. 2009.

ROBERTS, G. A. F. **Chitin chemistry**. London: Macmillan, 1992. xvi, 350 p. ISBN 0333524179.

SAVITZKY, A.; GOLAY, M. J. E. Smoothing and Differentiation of Data by Simplified least squares procedures. **Analytical chemistry**, v. 36.8, p. 1627–1639, 1964.

SENRA, T. D. A. **N, N, N-trimetilquitosana e N-(2-hidróxi)-propil-3-trimetilamônio quitosana: preparação, caracterização e estudo de suas interações com decanossulfonato de sódio**. 2015. Tese de Doutorado. Universidade de São Paulo (USP), São Carlos-SP.

SIGMA-ALDRICH. **Peso molecular médio de quitosano**. Disponível em: <<http://www.sigmaaldrich.com/catalog/product/aldrich/448877?lang=pt®ion=BR>>. Acesso em: 1 set. 2017.

SIMÕES, S. S. **Desenvolvimento de métodos validados para a determinação de captopril usando espectrometria NIRR e calibração multivariada.** p. 83, 2008. Tese (Doutorado em Química) - Universidade Federal da Paraíba, João Pessoa.

SKOOG, D. A.; HOLLER, F. J.; CROUCH, S. R. **Princípios de análise instrumental.** Porto Alegre: Editora Bookman, 6ª edição, 2009.

SOUZA, J. A. L. **Desenvolvimento de uma metodologia NIR para determinação do produto de degradação do captopril em comprimidos.** Dissertação de Mestrado. Universidade Federal de Pernambuco, 2011.

STUART, B. H. **Infrared Spectroscopy: Fundamentals and Applications.** Wiley, 2005.

WEINHOLD, M. X. et al. Strategy to improve the characterization of chitosan for sustainable biomedical applications: SAR guided multi-dimensional analysis. **Green Chemistry**, v. 11, p. 498, 2009.