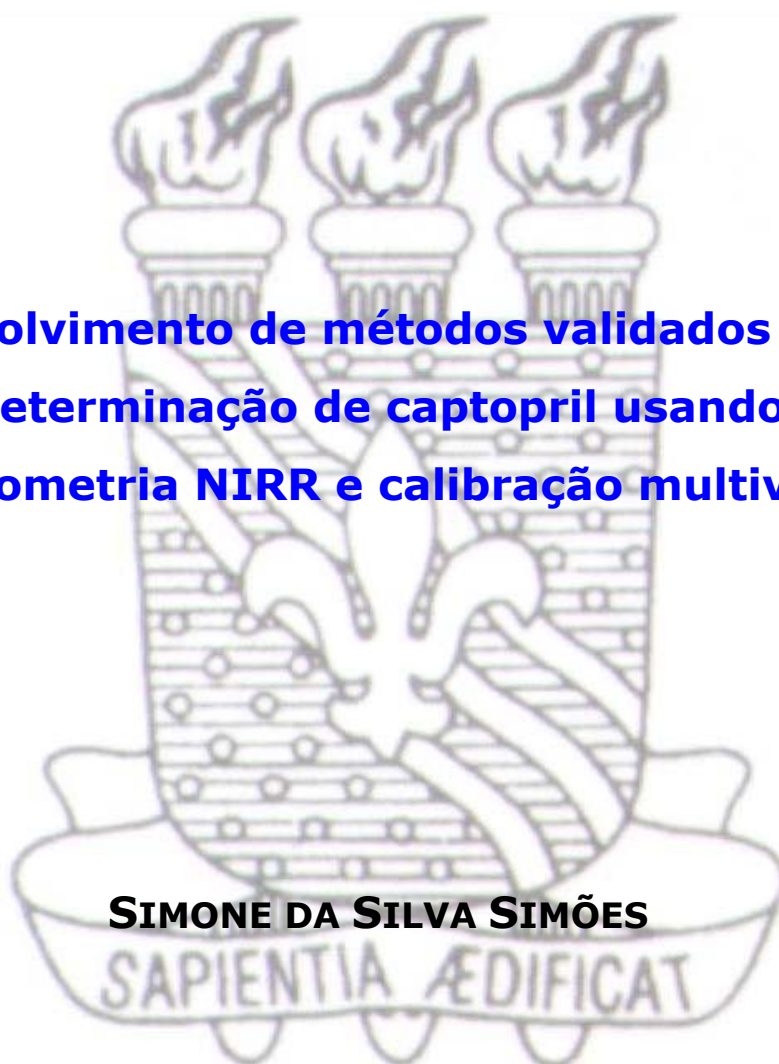




**UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
DEPARTAMENTO DE QUÍMICA
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA**

TESE DE DOUTORADO

**Desenvolvimento de métodos validados para a
determinação de captopril usando
espectrometria NIRR e calibração multivariada**



SIMONE DA SILVA SIMÕES

*João Pessoa – PB - Brasil
04/2008*



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

TESE DE DOUTORADO

**Desenvolvimento de métodos validados para a
determinação de captopril usando
espectrometria NIRR e calibração multivariada**

SIMONE DA SILVA SIMÕES*

Tese apresentada como requisito para obtenção
do título de Doutor em Química pela
Universidade Federal da Paraíba.

**Orientador: Prof. Dr. Mário César Ugulino de Araújo
Co-Orientador: Dr. Wallace Duarte Fragoso**

***Bolsista (CAPES)**

*João Pessoa – PB - Brasil
04/2008*

S593d Simões, Simone da Silva.

Desenvolvimento de métodos validados para a determinação de captopril usando espectrometria NIR e calibração multivariada/ Simone da Silva Simões.- João Pessoa, 2008.

83p.: il.

Orientador: Mário César Ugulino de Araújo

Co-orientador: Wallace Duarte Fragoso

Tese (doutorado) – UFPB/CCEN.

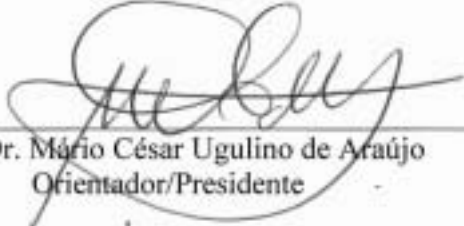
1. Química Analítica. 2. Captopril. 3. Validação. 4. Espectrometria NIR. 5. Seleção de variáveis. 6. MLR-SPA.


UFPB/BC


CDU: 543(043)


Desenvolvimento De Metodologias Validadas para a Determinação de Captopril Usando Espectrometria NIR, PLS-1 e MLR-SPA.

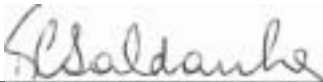
Aprovada pela banca examinadora:

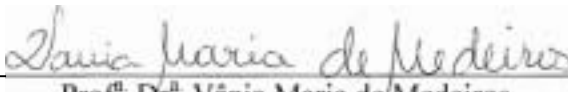

Prof. Dr. Mário César Ugulino de Araújo
Orientador/Presidente


Prof. Dr. Wallace Duarte Fragoso
2º. Orientador


Prof. Dr. Antonio Cícero de Sousa
Examinador


Profª Drª Maria Fernanda Pimentel
Examinadora


Profª Drª Teresa Cristina Bezerra Saldanha
Examinadora


Profª Drª Vânia Maria de Medeiros
Examinadora

*A vida é uma pedra de amolar: desgasta-nos ou afia-nos,
conforme o metal de que somos feitos.*

(George Bernard Shaw)

Às pessoas responsáveis pelo que há de bom em mim.

Especialmente, minha mãe D. Marlene

E meu pai, Seu Aauto

(in memoriam).

Com amor dedico.

AGRADECIMENTOS

- A Deus, que sempre me acompanha e me fortalece.
- Aos meus familiares pela paciência e compreensão.
- À minha mãe e ao seu colo, poço perene de energia e amor.
- Aos amigos do LAQA pelo companheirismo e apoio nos momentos difíceis. Em particular a Pedro Germano e Sófacles Figueiredo.
- A Elaine Cristina companheira de luta, com quem eu sempre pude partilhar minhas angústias.
- A Claudete, amiga sempre fiel e companheira.
- À Professora Teresa Saldanha, por me ensinar a, nem sempre fácil, arte do otimismo.
- Ao meu orientador Mário Ugulino pela orientação e apoio.
- Ao Professor Célio Pasquini e aos demais professores e alunos do Grupo de Instrumentação Analítica - GIA/UNICAMP, pela acolhida durante o desenvolvimento da parte experimental.
- Ao Prof. Dr. Edvan Cirino pela colaboração e discussões.
- A Wallace Fragoso.
- Ao diretor do Laboratório Farmacêutico de Estado de Pernambuco – LAFEPE, Davi Santana, e as funcionárias do departamento de P&D, em especial Zênia, Ayla e Flávia Moraes.
- À Prof^a. Olga da Coordenação de Meio-ambiente/CEFET-PB.
- Aos professores e funcionários do DQ/UFPB.
- E a CAPES pela bolsa concedida e auxílio nas missões de estudo.

SUMÁRIO

LISTA DE FIGURAS	viii
ÍNDICE DE TABELAS	ix
LISTA DE ABREVIATURAS E SIGLAS	x
RESUMO	xi
ABSTRACT	xii
Capítulo 1 <i>Introdução & Objetivos</i>	1
1.1 O captopril	2
1.2 Qualidade de fármacos	4
1.3 Espectroscopia NIR	6
1.3.1 <i>Tipos de medidas na espectrometria NIR</i>	9
1.3.2 <i>Aplicações da espectroscopia NIR</i>	12
1.3.3 <i>Espectrometria NIR na indústria farmacêutica</i>	13
1.4 Quimiometria	14
1.4.1 <i>Planejamento de experimentos</i>	15
1.4.1.1 <i>Planejamentos gerados via computador</i>	16
1.4.2 <i>Métodos de calibração multivariada</i>	17
1.4.2.1 <i>Regressão por mínimos quadrados parciais (PLS)</i>	19
1.4.2.2 <i>Regressão linear múltipla (MLR)</i>	20
1.4.3 <i>Algoritmos para seleção variáveis e amostras</i>	21
1.4.3.1 <i>Algoritmo das Projeções Sucessivas</i>	22
1.4.3.2 <i>Algoritmo Kennard-Stone</i>	25
1.4.3.3 <i>Algoritmo SPXY</i>	26
1.5 Validação analítica e figuras de mérito	27
1.5.1 <i>Recomendações da norma ASTM E 1655 05 para calibração multivariada usando a espectrometria NIR</i>	29
1.5.1.1 <i>Recomendações gerais para a construção do modelo de calibração multivariada usando a espectrometria NIR</i>	30
1.5.1.2 <i>Recomendações para a seleção das amostras do conjunto de calibração</i>	31
1.5.1.3 <i>Recomendações para a seleção das amostras do conjunto de validação</i>	33
1.5.1.4 <i>Questionário para avaliação do modelo de calibração multivariada</i>	35
1.6 Preparação e seleção das amostras de calibração	38
1.7 Detecção de amostras anômalas	40

1.8	Objetivos.....	43
1.8.1	Objetivo geral	43
1.8.2	Objetivos específicos	43
Capítulo 2	<i>Experimental</i>	44
2.1	As amostras.....	45
2.1.1	<i>Comprimidos sintéticos</i>	45
2.1.1.1	Procedimento de preparação dos comprimidos sintéticos ..	47
2.1.2	<i>Amostras analisadas</i>	48
2.2	Registro dos espectros NIRR	48
2.1.3	<i>Conjuntos de calibração e validação</i>	50
2.1.4	<i>Valores de referência dos comprimidos do LAFEPE</i>	50
2.3	Programas estatísticos.....	50
Capítulo 3	<i>Resultados & Discussão</i>	52
3.1	Pré-processamento dos espectros NIRR do captopril	53
3.1.1	<i>Escolha da faixa espectral de trabalho</i>	53
3.1.2	<i>Estudo da necessidade de pré-procesamento</i>	54
3.2	Construção dos modelos de calibração multivariada.....	57
3.2.1	<i>Detecção de amostras anômalas</i>	57
3.2.2	<i>Modelagem PLS-1 e MLR-SPA usando um conjunto de calibração contendo apenas amostras sintéticas</i>	58
3.2.3	<i>Modelagem PLS-1 e MLR-SPA usando um conjunto de calibração com amostras sintéticas e comerciais</i>	60
3.2.4	<i>Modelagem PLS-1 e MLR-SPA com inserção de amostras comerciais no conjunto de calibração</i>	62
3.3	Validação analítica dos modelos PLS-1 e MLR-SPA de acordo com a norma ASTM E 1655 05.....	65
3.3.1	<i>Questionário da ASTM E 1655 05 para avaliação dos modelos PLS-1 e MLR-SPA construídos</i>	68
Capítulo 4	<i>Conclusões & Propostas Futuras</i>	70
4.1	Conclusões	71
4.2	Propostas Futuras	72
	<i>Referências Bibliográficas</i>	74

LISTA DE FIGURAS

Figura 1: Forma estrutural do captopril.	3
Figura 2: Espectros NIR do clorofórmio obtido usando um espectrômetro NIR com transformada de Fourier, CCl ₄ como referência, 100 varreduras com uma resolução de 4 cm ⁻¹ e um caminho óptico de (a) 1 mm e (b) 50 mm.	8
Figura 3: Modos de medidas na região NIR. (a) Transmitância, (b) transflectância e (c) refletância difusa ^[9]	9
Figura 4: Organização dos dados na calibração multivariada.	18
Figura 5: Visão espacial do funcionamento do SPA. (a) Projeção no plano ortogonal a x ₃ e (b) visão espacial dos vetores no plano horizontal.	23
Figura 6: Gráfico da concentração residual de Student versus influência das amostras.	41
Figura 7: Prensa (a) e forma (b) <i>labmade</i> usada para prensar os comprimidos.	47
Figura 8: Espectrofotômetro NIR, modelo ABB BOMEM.	49
Figura 9: Suporte para centralizar a amostra sob a fonte de radiação... ..	49
Figura 10: Espectro NIRR de um comprimido de Captopril.	53
Figura 11: Faixa espectral de trabalho.	54
Figura 12: Espectros NIRR sem pré-processamento dos comprimidos de captopril sintéticos (em azul) e comerciais (em vermelho).....	55
Figura 13: Um espectro NIRR representativo de um comprimido antes (a) e após (b) a suavização Savitzky-Golay com uma janela de 9 pontos	56
Figura 14: Variância residual em X versus leverage das amostras do modelo PLS-1 construído usando os dados de comprimidos sintéticos (a) e dos comerciais (b).	58
Figura 15: Gráfico das concentrações medidas versus as concentrações previstas na construção do modelo de calibração usando o método de validação cruzada.	59
Figura 16: Resíduo de concentração de Student versus a influência das amostras após a inserção de (a) 3, (b) 6; (c) 9, (d) 12, (e) 15, (f) 18 amostras comerciais no conjunto de calibração.....	62
Figura 17: Gráfico das concentrações medidas versus as concentrações previstas na construção do modelo de calibração usando o método de validação cruzada, com a inserção de 15 amostras comerciais no conjunto de calibração.	63
Figura 18: Indicação da posição em um espectro representativo das variáveis selecionadas durante a construção do modelo MLR-SPA... ..	64

Figura 19: Concentração estimada, por comprimido, das 32 amostras do conjunto de previsão usando o modelo PLS-1 construído com o conjunto de calibração contendo 35 amostras sintéticas e 15 amostras comerciais..... 66

ÍNDICE DE TABELAS

Tabela 1: Nomes comerciais, fabricantes e dosagens mais comuns no mercado para os comprimidos contendo o princípio ativo captopril	3
Tabela 2: Planejamento de misturas D-otimizado de misturas para a preparação das amostras sintéticas.....	45
Tabela 3: Valores de RMSEP e número de variáveis latentes utilizadas ou comprimentos de onda utilizados nas modelagens PLS-1 e MLR-SPA, respectivamente.	59
Tabela 4: Valores de RMSEP para os modelos PLS-1 e MLR-SPA.	64
Tabela 5: Valores de $t_{\text{calculado}}$ e $t_{\text{crítico}}$ para os modelos PLS-1 e MLR-SPA. ($t_{\text{crítico}}=1,6955$)	65
Tabela 6: Parâmetros para avaliação dos modelos PLS-1 e MLR-SPA. ...	67

LISTA DE ABREVIATURAS E SIGLAS

ANVISA - Agência Nacional de Vigilância Sanitária
ASTM - *American Society for Testing and Materials*
BPF - Boas Práticas de Fabricação.
ECA - Enzima Conversora da Angiotensina
EMA - Agência Europeia para Avaliação de Produtos Médicos
GA - *Genetic Algorithm*
 h_i - influência das amostras
HPLC - *High Performance Liquid Chromatography*
ISO – *International Organization of Standardization*
KS - Algoritmo Kennard-Stone
LAFEPE - Laboratório Farmacêutico do Estado de Pernambuco
MIR - *Medium Infrared*
MLR - *Multiple Linear Regression*
NAS - *Net Analyte Signal*
NIRR - *Near Infrared Diffuse Reflectance*
NIRS - *Near Infrared Spectrometry*
PASG - Grupo de Ciências Analíticas e Farmacêuticas
PLS - *Partial Least Square*
PRESS - *Prediction Residual Error Sum of Squares*
 rcS_i - resíduo de concentração de Student
RENAME - Relação Nacional de Medicamentos Essenciais
RMSEP - *Root Mean Square Error of Prediction*
RMSSR - *Root Mean Square Spectral Residuals*
RS - *Random Search*
SECV - *Standard Error of Cross Validation*
SPA - *Sucessive Projections Algorithm*
SPXY - *Sample set Partitioning based on joint **X-y** distances*

RESUMO

Título: Desenvolvimento de métodos validados para a determinação de captopril usando espectrometria NIRR e calibração multivariada

Autor: Simone da Silva Simões

Neste trabalho foram desenvolvidos dois métodos para determinação do teor de captopril em comprimidos empregando a espectrometria NIRR (*Near Infrared Diffuse Reflectance*) associada a técnicas de calibração multivariada. O primeiro é baseado em PLS-1 (*Partial Least Squares*) e usa todo o espectro e o segundo se baseia em MLR (*Multiple Linear Regression*) e aplica o algoritmo das projeções sucessivas (SPA) para seleção de variáveis de modo a minimizar problemas de colinearidade. Na construção dos modelos foram usados comprimidos sintéticos, preparados a partir de um planejamento D-otimizado, em uma faixa de concentrações de captopril mais ampla que a dos comprimidos produzidos pelo LAFEPE (Laboratório Farmacêutico do Estado de Pernambuco). Essas amostras sintéticas são necessárias porque, em geral, um conjunto de calibração constituído apenas de amostras da linha de produção pode resultar em modelos com falta de ajuste, uma vez que estas amostras não cobrem uma faixa de concentração adequada para a modelagem. Por outro lado, como as amostras sintéticas não apresentam as variações físicas inerentes às amostras da linha de produção, algumas destas também foram incluídas no conjunto de calibração. O número ideal de amostras da linha de produção a serem incluídas foi estabelecido usando o algoritmo SPXY e gráficos de influência. Foram registrados espectros NIRR, na faixa de 4000 a 14000 cm^{-1} usando um espectrômetro BOMEM, com transformada de Fourier com acessório para sólidos e teflon como branco. O método de referência para determinação das concentrações de captopril nas amostras industriais foi baseado na HPLC (*High Performance Liquid Chromatography*). Os novos métodos NIRR-PLS-1 e NIRR-MLR-SPA ora apresentados são rápidos, precisos e não consomem reagentes nocivos ao meio ambiente. Entretanto, sua principal e notável vantagem é a não necessidade de solubilização das amostras o que possibilita o controle do teor de captopril em todos os comprimidos sem destruição dos mesmos, como ocorre numa análise por HPLC. Estes métodos foram avaliados e validados com base em algumas figuras de mérito da norma ASTM E1655 00 o que permite que sejam certificados *a posteriori* pelo LAFEPE que tem interesse em implantá-los em seu processo de controle de qualidade. Ambos os métodos apresentaram bom desempenho. No entanto, apenas o método NIRR-MLR-SPA pode ser facilmente adaptado a instrumentos NIR muito mais baratos como os fotômetros a base de diodos emissores de luz (LED's), ao controle *in situ* (ou *multisitu*) e *on line* dos comprimidos durante todo o processo de produção.

Palavras-chave: Captopril, validação, espectrometria NIRR, PLS-1, seleção de variáveis, MLR-SPA.

ABSTRACT

Title: Development of validated methods for the determination of captopril using NIRR spectrometry and calibration multivariate

Author: Simone Simoes da Silva

In this work two methods were developed for determining the content of captopril tablets in employing the NIRR spectrometry (Near Infrared Diffuse Reflectance) associated with calibration multivariate techniques. The first one is based on PLS-1 (Partial Least Squares) and uses the entire spectrum and the second is based on MLR (Multiple Linear Regression) and applies the algorithm of successive projections (SPA) for selection of variables in order to minimize problems of colinearity. In the construction of the models synthetic tablets prepared from a planning D-optimized were used, with a range of concentrations of captopril wider than the tablets produced by LAFEPE (Pharmaceutical Laboratory of the State of Pernambuco). These synthetic samples are necessary because, in general, a set of calibration samples consisted only of the production line can result in models with lack of adjustment, as these samples do not cover a range of concentration appropriate for modeling. Moreover, as these synthetic samples do not have the physical variations inherent in samples of the production line, some of these were also included in the set of calibration. The ideal number of samples of the production line to be included was established using the algorithm SPXY and influence plot. NIRR spectra were recorded in the range of 4000 to 14000 cm^{-1} using a BOMEM spectrometer, with Fourier transform with accessory for solids and Teflon as a blank. The reference method for determining concentrations of captopril in the samples was based on industrial HPLC (High Performance Liquid Chromatography). The new methods NIRR-PLS-1 and NIRR-MLR-SPA now submitted are fast, accurate and do not consume reagents harmful to the environment. Meanwhile, its main and remarkable advantage is not to need the solubilization of samples which makes the control the content of captopril in all tablets possible, without destroying them, as it happens in an analysis by HPLC. These methods were tested and validated on the basis of some figures of merit of the standard 00 the ASTM E1655, what allows them to be certified later by LAFEPE that has interest in implanting them in its quality control process. Both methods showed good performance. However, only the method NIRR-MLR-SPA can be easily adapted to much cheaper NIR instruments like the light emitting-based diodes (LED's) photometers, to the *in situ* (or *multisitu*) and online control of tablets throughout the production process.

Keywords: captopril, validation, NIRR spectrometry, PLS-1, selection of variables, MLR-SPA.



Capítulo 1

Introdução & Objetivos

1.1 O captopril

Os agentes anti-hipertensivos podem ser classificados de acordo com seus mecanismos de ação como^[1]:

- i) diuréticos;
- ii) agentes simpatolíticos;
- iii) vasodilatadores;
- iv) bloqueadores dos canais de Cálcio e
- v) inibidores da Enzima Conversora da Angiotensina (ECA ou ACE) e antagonistas do receptor da angiotensina.

O Captopril faz parte do grupo de anti-hipertensivos classificados como Inibidores da Enzima Conversora da Angiotensina (ACE ou ECA) e representam um grande avanço no tratamento da hipertensão. O primeiro inibidor da ECA a ser comercializado foi o Captopril e vários outros foram posteriormente desenvolvidos, como o Enalapril, Lisinopril, Perindopril, Ramipril, Quinalapril etc. Porém, entre estes, o captopril é o anti-hipertensivo que possui menor custo, sendo produzido em larga escala pelos laboratórios oficiais brasileiros e faz parte da Relação Nacional de Medicamentos Essenciais – RENAME^[2].

Em 1949, Maurício Rocha e Silva descobriu que o veneno da jararaca em contato com plasma de cão disparava a produção de um hormônio, a bradicinina, regulador da pressão arterial. Seu aluno, Sérgio Ferreira, descobriu no veneno a substância que foi a base do anti-hipertensivo captopril, que acabou desenvolvido por uma indústria farmacêutica estrangeira, já que o Brasil não tinha condições de fazê-lo. O nome e fórmula química do captopril são, respectivamente, 1-[(2S)-3-Mercapto-2metilpropionil]-L-prolina e $C_9H_{15}NO_3S$ e sua fórmula estrutural é apresentada na **Figura 1**.

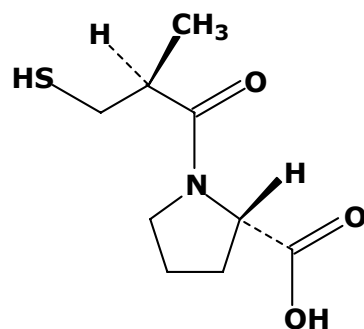


Figura 1: Forma estrutural do captopril.

Este fármaco constitui-se em um pó cristalino, branco ou levemente amarelado, com leve odor característico de sulfeto. É facilmente solúvel em água, etanol, clorofórmio, metanol e em soluções alcalinas^[3]. Apresenta um ponto de fusão entre 104 e 110°C. Este medicamento pode ser encontrado sob várias dosagens e nomes comerciais. Algumas das principais denominações comerciais e dosagens se encontram listadas na **Tabela 1**.

Tabela 1: Nomes comerciais, fabricantes e dosagens mais comuns no mercado para os comprimidos contendo o princípio ativo captopril

Nome comercial (Fabricante)	Dosagem (mg)*
Capoten [®] (Bristol Myers Squibb)	12,5 / 25 / 50
Capotril [®] (Neo-Quimica)	12,5 / 25 / 50
Capril [®] (Teuto Brasileiro)	12,5 / 25 / 50
Captil [®] (Hebron)	12,5 / 25 / 50
Captobel [®] (Sedabel)	12,5 / 25
Captolin [®] (Herald 's)	12,5 / 25
Captomed [®] (Cimed)	12,5 / 25 / 50
Catoprol [®] (Medley)	12,5 / 25 / 50
Capton [®] (Royton)	12,5 / 25 / 50
Captopiril [®] (Bunker)	12,5 / 25 / 50
Captopril (Biosintética, Cazi, EMS, Medley, Merk)	12,5 / 25 / 50
Captopril (LAFEPE)	25
Captopril (Neovita, Novartis)	25 / 50
Captotec [®] (Hexal)	25 / 50
Captrizim [®] (Prodotti)	25 / 50

*Por cada comprimido de 150 mg.

Os comprimidos de captopril utilizados neste trabalho foram aqueles produzidos pelo Laboratório Farmacêutico do Estado de Pernambuco – LAFEPE. O LAFEPE é o segundo maior laboratório público do Brasil. Foi criado em 1966 para produzir medicamentos a baixo custo para as populações de menor poder aquisitivo e tem sede no Recife, capital do Estado de Pernambuco. O LAFEPE produz e comercializa medicamentos destinados às necessidades das políticas de saúde pública e investe na modernização de suas instalações, ao dotar de alta tecnologia seu parque industrial, o que implica na produção de medicamentos de melhor qualidade.

1.2 Qualidade de fármacos

A qualidade vem sendo utilizada como selo de garantia em qualquer produto industrial. Por este motivo, o controle de qualidade dos produtos é indispensável na maioria dos processos industriais e em vários campos de aplicação. Nas indústrias farmacêuticas esse controle é essencial tendo em vista a finalidade dos produtos por ela produzidos. Desvios na qualidade desses produtos podem causar efeitos prejudiciais à saúde dos consumidores, podendo em casos extremos levar à morte.

Segundo a norma NBR ISO 9000:2000, *qualidade é a totalidade de características de uma entidade que lhe confere a capacidade de satisfazer as necessidades explícitas e implícitas*. Segundo essa definição quando se usa o termo “de boa qualidade” pretende-se exprimir a excelência de um produto ou serviço e o cumprimento das especificações, quantitativas ou qualitativas, definidas pelas autoridades competentes. O controle de qualidade consiste em realizar medidas de alguns parâmetros do produto considerado e verificar se os valores obtidos estão de acordo com as especificações pré-estabelecidas.

Para que uma indústria farmacêutica possa produzir e comercializar um fármaco ela precisa cumprir as exigências dos órgãos responsáveis pela fiscalização das características e qualidade dos produtos como a ANVISA (Agência Nacional de Vigilância Sanitária) e as suas sedes regionais de vigilância sanitária. Essas exigências dizem respeito principalmente ao cumprimento das Boas Práticas de Fabricação (BPF's)^[4]. As BPF's são um conjunto de normas específicas, determinadas pelas autoridades sanitárias, referentes à fabricação e ao controle de qualidade do fármaco, a fim de que sejam produzidos medicamentos seguros e eficazes. Entre elas estão o controle de qualidade dos insumos, a validação dos processos de fabricação, as instalações e os equipamentos adequados e o treinamento de pessoal. Além da observação das BPF's, é necessário que os medicamentos produzidos sejam submetidos a testes de biodisponibilidade relativa e de bioequivalência farmacêutica antes de serem disponibilizados aos consumidores.

Determinar todos os parâmetros de qualidade em um processo industrial normalmente envolve um grande número de análises. Deste modo, um laboratório de controle de qualidade dispõe de uma quantidade considerável de técnicas para o controle de rotina e atualmente a maioria destas análises é realizada empregando métodos instrumentais.

As metodologias mais comumente utilizadas no controle de qualidade de fármacos são: a espectrometria UV-Visível e a cromatografia líquida de alta eficiência (do inglês: *High Performance Liquid Chromatography* - HPLC)^[5-8]. Estes métodos são lentos e trabalhosos, pois envolvem muita manipulação analítica (extração/dissolução da amostra, preparação freqüente de reagentes e soluções de calibração, etc) principalmente quando aplicados a produtos farmacêuticos em fase sólida. Além disso, são métodos relativamente dispendiosos, pois requerem o uso de solventes e

reagentes caros, precisam de pessoal qualificado para a manipulação analítica e produzem resíduos prejudiciais ao meio ambiente.

Uma boa alternativa para contornar os problemas apresentados pelas técnicas convencionais é a espectrometria no infravermelho próximo (do inglês: *Near Infrared Spectrometry* - NIRS).

1.3 Espectroscopia NIR

A espectroscopia NIR é uma técnica de análise que utiliza a faixa de comprimentos de onda adjacente ao infravermelho médio e se estende até a região do visível. A região NIR do espectro eletromagnético pode ser definida como a faixa de comprimentos de onda de 780 a 2500 nm. As bandas de absorção mais proeminentes que ocorrem nesta região estão relacionadas a sobretons e combinações de vibrações fundamentais de grupos funcionais **-CH**, **-NH**, **-OH** e **-SH**^[9].

A mecânica quântica, assumindo o modelo do oscilador harmônico, prevê separações iguais entre os níveis vibracionais e impõe a regra de seleção $\Delta v = \pm 1$. Para oscilações anarmônicas, a separação entre os níveis já não é igualmente espaçada e diminui para energias mais altas. Assim, a regra de seleção $\Delta v = \pm 1$ é relaxada e transições com $\Delta v = \pm 2$ ou $\Delta v = \pm 3$ dão origem aos sobretons, que são as transições observadas no NIR. As bandas de combinação aparecem entre 1900 e 2500 nm e são resultantes de interferências construtivas de frequências, isto é, suas frequências são a soma dos múltiplos da cada frequência de interação^[10].

As bandas de absorção NIR são tipicamente largas, sobrepostas e de 10 a 100 vezes mais fracas que suas bandas fundamentais na região do infravermelho médio (do inglês: *Medium Infrared* - MIR). Sua baixa absorvidade, no entanto, permite alta

profundidade de penetração. Este aspecto constitui uma vantagem analítica, pois permite análises diretas de fortes absorventes e nivela amostras com espalhamento, como os líquidos turvos ou sólidos nos modos de refletância ou transmitância sem necessidade de pré-tratamento. A dependência do sinal analítico NIR em relação às propriedades químicas e físicas da amostra, resultante da absorção e do espalhamento, podem ser favoravelmente usados para realizar análises químicas e físicas em uma única medida.

A principal dificuldade para o uso da espectrometria NIR é a complexidade de seus espectros, pois, diferentemente dos espectros na região do infravermelho médio onde as bandas de absorção são diretamente interpretadas devido à especificidade de seus picos, os espectros NIR são de difícil interpretação em virtude da sua natureza (sobretons e bandas de combinação de níveis vibracionais de energia). Um bom exemplo, que ilustra esta complexidade, pode ser observado no espectro NIR da molécula de clorofórmio (**Figura 2**), uma molécula muito simples, que apresenta características espectroscópicas NIR que podem ser atribuídas apenas aos modos de vibração C-H, suas combinações e sobretons.

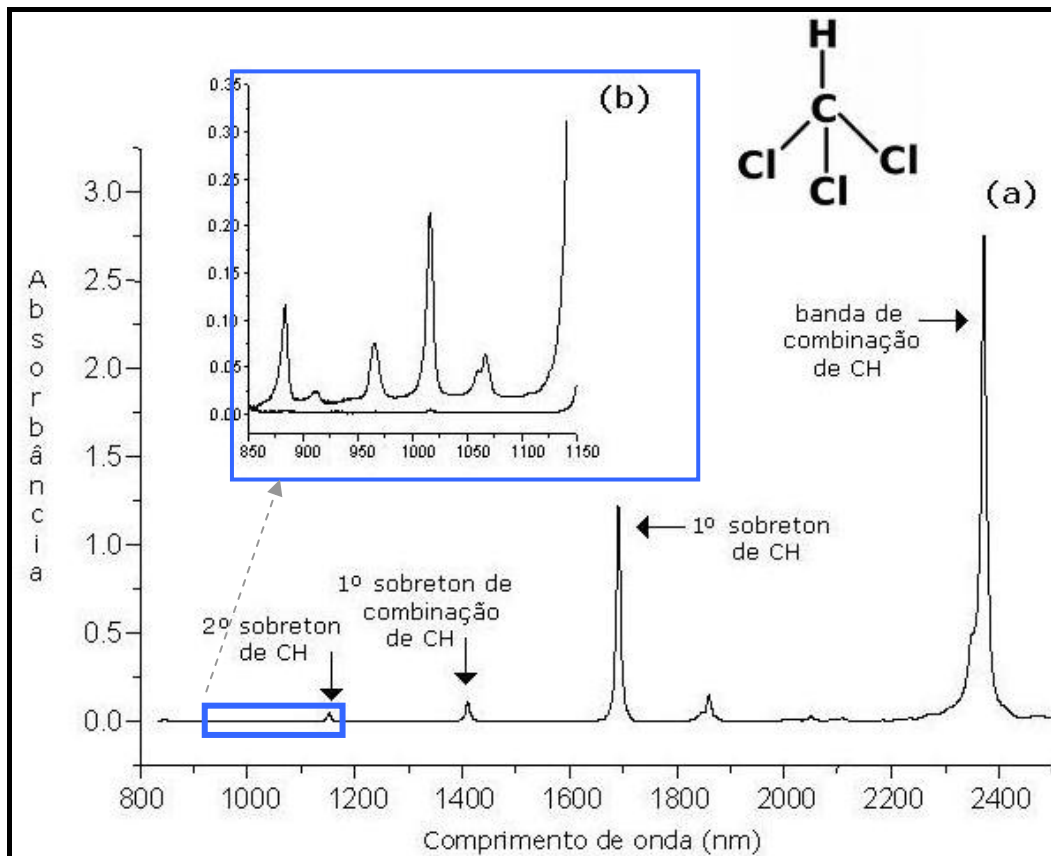


Figura 2: Espectros NIR do clorofórmio obtido usando um espectrômetro NIR com transformada de Fourier, CCl₄ como referência, 100 varreduras com uma resolução de 4 cm⁻¹ e um caminho óptico de (a) 1 mm e (b) 50 mm^[9].

As bandas de absorção devido às combinações e aos 1º e 2º sobretons de ligações de C-H podem ser facilmente observadas na **Figura 2a**. A aparente simplicidade do espectro apresentado na **Figura 2a** é logo descartada se olharmos a **Figura 2b**, que mostra o espectro do CHCl₃ obtido numa região NIR de 850 a 1150 nm com um caminho óptico de 50 mm. No detalhe da **Figura 2b**, cinco bandas de absorção podem ser claramente identificadas na região de comprimento de onda onde, no primeiro espectro (**Figura 2a**), não eram observados picos. Apesar da simplicidade da molécula, são conhecidas 34 bandas de combinações e sobretons do clorofórmio, todas com energias na região NIR^[11]. Além de fornecer muitas informações químicas relevantes, os espectros NIR também se correlacionam com as propriedades físicas das amostras, o que dificulta ainda mais a interpretação direta dos dados obtidos. Deste

modo, torna-se imprescindível o uso de ferramentas matemáticas e estatísticas, ou seja, a Quimiometria, para extrair informações dos espectros NIR das amostras.

1.3.1 Tipos de medidas na espectrometria NIR

No início da descoberta da técnica NIR, acreditava-se que as medidas espectrais podiam ser obtidas através da análise da informação contida em uma porção de radiação refletida de maneira difusa por amostras sólidas através de um fraco sinal de transmitância. Atualmente, a refletância difusa é apenas uma das várias maneiras para se realizar medidas analíticas na região NIR. A **Figura 3** representa as formas mais comuns de medidas realizadas na espectroscopia NIR^[9].

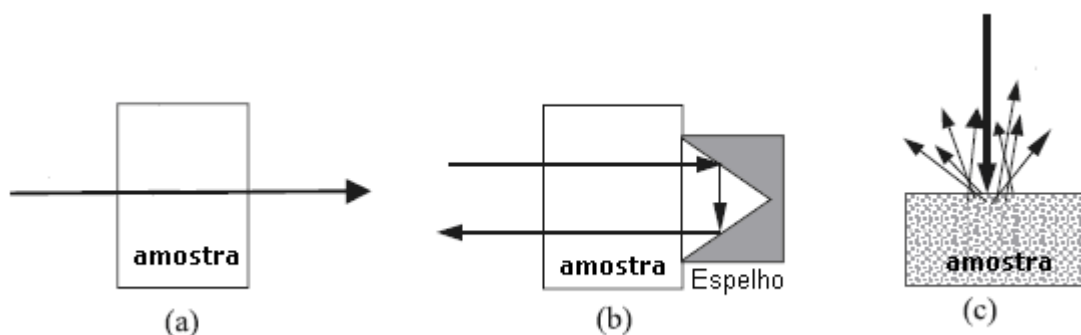


Figura 3: Modos de medidas na região NIR. (a) Transmitância, (b) transflectância e (c) refletância difusa^[9].

Nas medidas de transmitância (**Figura 3a**), cubetas de vidro ou quartzo com caminhos ópticos variando de 1 a 50 mm, são empregadas como suporte transparente para colocação das amostras no momento do registro dos espectros NIR. O caminho óptico é, em princípio, definido de acordo com a região espectral que está sendo estudada.

A **Figura 3b** mostra um modo especial para realizar medidas de transmitância, conhecida como transflectância. Este modo de medida é utilizado quando feixes de fibra óptica são empregados. A

diferença em relação às medidas de transmitância é o caminho óptico duplo pelo qual o feixe de radiação passa duas vezes através da amostra.

As medidas de refletância difusa de amostras sólidas (**Figura 3c**), como mencionado anteriormente, é um modo de medida característico para análises empregando a espectrometria NIR. Este modo de medida é apropriado quando a superfície do material reflectante é pouco absorvente no comprimento de onda incidente. Este tipo de reflexão é a base das medidas de espectroscopia NIR de refletância difusa (do inglês: *Near Infrared Diffuse Reflectance* - NIRDR), no qual o espalhamento e a absorbância de partículas sólidas contribuem para mudanças na intensidade do sinal. Devido à sua complexidade, não há teorias rigorosas sobre a refletância difusa. No entanto, uma série de teorias empíricas foram propostas, entre as quais a mais popular é a teoria de Kubelka-Munk^[12]. No caso de amostras opacas de espessura infinita a função de Kubelka-Munk é descrita como segue, onde R_{∞} é a refletância absoluta da amostra. Esta teoria assume que a radiação incide sobre o meio dispersante sofre um processo de absorção e dispersão, simultaneamente. Deste modo, a radiação refletida pode ser descrita em função das constantes de absorção (k_a) e dispersão (s):

$$f(R_{\infty}) = \frac{(1 - R_{\infty})^2}{2R_{\infty}} = \frac{k_a}{s} \quad (1)$$

Em análises quantitativas, a **Equação 1** pode ser descrita em função da concentração do analito absorvente (c) de acordo com a seguinte equação:

$$f(R_{\infty}) = \frac{(1 - R_{\infty})^2}{2R_{\infty}} = \frac{k_a}{s} = \frac{ac}{s} \quad (2)$$

onde a é a absorptividade molar.

Na prática, ao invés da refletância absoluta R_{∞} é utilizada a refletância relativa R , que é a relação entre as intensidades de luz refletida pela amostra e pelo padrão. Este padrão deve ser um

material estável, com refletância absoluta elevada e relativamente constante na região espectral do NIR, por exemplo, brometo de potássio, teflon, sulfato de bário, óxido de magnésio.

Reescrevendo a **Equação 2** em termos de refletância relativa teremos:

$$f(R) = \frac{(1 - R)^2}{2R} = \frac{ac}{s} \quad (3)$$

Para amostras que seguem a **Equação 3**, o gráfico $f(R)$ em função da concentração é uma linha reta com coeficiente angular igual a a/s . Entretanto, se a matriz também absorve ou se o analito apresenta forte absorção, a refletância difusa não obedece a esta relação linear.

Assim a equação de Kubelka-Munk, assim como a lei de Beer, é limitada e só pode ser aplicada para bandas de absorção de baixa intensidade. Esta limitação ocorre quando se utiliza a espectrometria NIR, pois não é possível separar a absorção do analito da absorção da matriz, ocorrendo desvios de linearidade.

Uma alternativa muito utilizada é a aplicação de uma relação entre a concentração e a refletância relativa, análoga à lei de Beer:

$$\log \frac{R_{padr\tilde{a}o}}{R_{amostra}} = \log \frac{1}{R_{amostra}} + \log R_{padr\tilde{a}o} \cong \frac{ac}{s} \quad (4)$$

Para a radiação monocromática o $\log R_{padr\tilde{a}o}$ é constante e a **Equação 4** pode ser escrita como:

$$A = \log \frac{1}{R} = ac \quad (5)$$

onde A é a absorbância aparente e a uma constante que engloba os termos $1/s$ e $\log R_{padr\tilde{a}o}$. Apesar de não ter as bases teóricas da equação de Kubelka-Munk, esta equação apresenta resultados muito satisfatórios nas aplicações da espectroscopia por refletância difusa.

1.3.2 Aplicações da espectroscopia NIR

As primeiras aplicações analíticas desta técnica ocorreram no final da década de 50 e a primeira aplicação importante foi na década de 60, quando Karl Norris iniciou o seu emprego no estudo de matrizes complexas de origem natural, como, produtos agroindustriais^[12,13]. A partir deste momento, o interesse pela espectrometria NIR cresceu notavelmente e esta técnica foi reconhecida como uma poderosa ferramenta para análises quanti e qualitativas de parâmetros químicos e físicos^[14-18]. Sua propagação nos vários ramos da indústria pode ser atribuída principalmente a sua habilidade para processar amostras sólidas com mínima ou nenhuma manipulação analítica e a sua compatibilidade com sondas de fibra óptica como transmissores de sinais, os quais permitem análises *in situ* e *on line*. Além dessas características, as metodologias baseadas na espectroscopia NIR possuem as seguintes propriedades inerentes:

- *Análise simultânea de vários parâmetros;*
- *São não destrutivas e não invasivas;*
- *Possuem alta velocidade de processamento das informações e rápido fornecimento de resultados quantitativos;*
- *Não consomem reagentes químicos nocivos ao meio ambiente;*
- *São menos laboriosas e de baixo custo.*

Por essas características, o uso da espectrometria NIR atualmente é uma realidade nos mais variados setores da cadeia produtiva, sendo empregada diretamente na linha de produção, no controle de qualidade de matérias primas e produtos, podendo ser aplicada às amostras dos mais variados tipos, tais como produtos da indústria de fármacos^[19,20], de polímeros^[21] e de produtos petroquímicos^[22] e agrícolas em geral^[23].

1.3.3 Espectrometria NIR na indústria farmacêutica

Nas últimas décadas, vários trabalhos científicos vêm sendo desenvolvidos visando ao estudo e à aplicação da espectrometria NIR^[24-26]. Atualmente, esta técnica tem ganhado aceitação pela indústria farmacêutica como meio de identificação^[27,28], quantificação^[28-30] e controle de parâmetros físicos e químicos^[31,32], seja da matéria-prima ou do produto final. Inclusive, várias agências reguladoras, como a Agência Europeia para Avaliação de Produtos médicos (EMA/CVMP/961/01), o Grupo de Ciências Analíticas e Farmacêuticas (PASG) e farmacopéias (como por exemplo: a Farmacopéia Europeia e a Farmacopéia dos Estados Unidos) vêm recomendando as metodologias baseadas na espectrometria NIR como uma boa alternativa analítica para controle de qualidade dos produtos farmacêuticos, devido às suas características vantajosas em relação aos métodos tradicionais. Contudo, sua aceitação e implementação na indústria farmacêutica, bem como em outros setores industriais, tem sofrido restrições devido às exigências para a sua validação, como aquelas expressas na norma ASTM E 1655 05 e outras normas específicas. Nesta norma são descritas as recomendações necessárias para a construção e validação de modelos de calibração multivariada usando espectrometria NIR.

A validação de um método analítico conforme as recomendações de órgãos reguladores é um passo importante para a sua certificação. A partir das figuras de mérito recomendadas para a validação, é possível garantir que a metodologia proposta é confiável para a estimativa das concentrações de amostras desconhecidas e pode ser usada efetivamente para o controle de qualidade de parâmetros químicos e físicos.

1.4 Quimiometria

A Quimiometria pode ser definida como a pesquisa e a utilização de métodos matemáticos e estatísticos para o tratamento de dados químicos de forma a extrair uma maior quantidade de informações e melhores resultados analíticos^[33]. Os métodos utilizados na quimiometria foram, a princípio, desenvolvidos em outras áreas da ciência. Por volta da metade dos anos 60 os métodos instrumentais computadorizados para a análise química passaram a gerar grandes quantidades de dados. Até então, os químicos baseavam suas decisões numa pequena quantidade de dados que, na maioria das vezes eram obtidos de forma lenta e dispendiosa. A partir dos anos 60, com a grande quantidade de dados obtidos de maneira rápida e despendendo menor esforço por parte dos analistas, foi preciso analisar todos esses dados e extrair maior quantidade de informações relevantes. A quimiometria pode ser considerada uma das áreas mais recentes da química analítica, e desde o seu surgimento até os dias atuais foram desenvolvidos muitos métodos que têm tornado possível o processamento e interpretação de dados que antes não poderiam ser analisadas. Graças à quimiometria, tornou-se possível empregar para fins analíticos a região espectral do NIR, de modo que atualmente pode-se afirmar que um equipamento para determinações analíticas por esta técnica baseia-se no *hardware* (instrumento) e *software* (quimiometria).

Na Indústria Química em geral, a associação de ferramentas quimiométricas com técnicas instrumentais, principalmente as espectrométricas e cromatográficas, é empregada no desenvolvimento de métodos analíticos para controle de qualidade^[34], qualificação de matéria prima^[35] e controle de processos^[34,36].

Com a disseminação da quimiometria foram desenvolvidas novas ferramentas para tratamento de dados encontrando aplicações

distintas conforme o objetivo do estudo, como por exemplo, o planejamento de experimentos, a otimização de processos, a classificação de dados e as determinações quantitativas usando a calibração multivariada.

1.4.1 Planejamento de experimentos

As ferramentas de planejamento de experimentos são utilizadas para planejar os experimentos de modo que as informações desejadas sejam obtidas^[37].

Os planejamentos de misturas apresentam uma significativa diferença em relação a outros tipos de planejamentos. Quando variáveis de misturas estão envolvidas em uma otimização, o resultado depende da proporção em que esses componentes se encontram e seus níveis não podem ser variados sem levar em conta os outros componentes, ou seja, as propriedades de uma mistura são determinadas pelas proporções de seus ingredientes e não por seus valores absolutos. A quantidade de cada componente do sistema deve ser tratada como uma variável de mistura, que não é independente das demais, pois o somatório das proporções dos seus componentes deve ser igual à unidade. Matematicamente isso pode ser descrito como:

$$\sum_{i=1}^q x_i = x_1 + x_2 + \dots + x_q = 1 \quad (6)$$

onde q é o número de componentes da mistura e x_i são as quantidades dos componentes na mistura.

Essa equação retira um grau de liberdade das proporções. Para especificar a composição da mistura, é necessário apenas fixar as proporções de $q-1$ componentes. A proporção do último componente será sempre o que falta para completar 100%. Devido a essa característica, a modelagem das propriedades químicas e físicas dos sistemas de misturas deve ser feita de maneira multivariada, pois

a restrição experimental imposta pela **Equação 6** torna necessário adaptar os métodos do planejamento fatorial e da análise de superfícies de resposta aos problemas específicos das misturas^[37]. Esses métodos modificados têm sido largamente aplicados na pesquisa, engenharia e principalmente na indústria.

1.4.1.1 Planejamentos gerados via computador

Muito do desenvolvimento dos planejamentos otimizados gerados via computador é uma consequência do trabalho de Kiefer (1959,1961) e de Kiefer e Wofowitz (1959) sobre a teoria dos planejamentos otimizados. Planejamento otimizado é o “melhor” planejamento com respeito a alguns critérios^[38]. Para construir este tipo de planejamento são necessários programas computacionais e o procedimento usual consiste em:

- *Especificar o planejamento;*
- *Determinar a região de interesse;*
- *Selecionar o número de seqüências a serem feitas;*
- *Especificar o critério ótimo;*
- *Escolher os pontos do modelo para o conjunto de pontos candidatos de interesse do experimentador.*

Tipicamente, os pontos candidatos são uma grade de pontos espaçados sobre uma região viável do planejamento.

Existem vários critérios populares para a construção de planejamentos otimizados. O mais largamente utilizado talvez seja o critério D-otimizado. Um planejamento é dito D-otimizado se $(\mathbf{X}'\mathbf{X})^{-1}$ é minimizado. Isso mostra que um modelo D-otimizado minimiza o volume da região de confiança conjunta no vetor dos coeficientes de regressão. Uma medida da eficiência relativa entre dois planejamentos que usam o critério D-otimizado é dada por^[38]:

$$D_e = \left(\frac{(\mathbf{X}'_2 \mathbf{X}_2)^{-1}}{(\mathbf{X}'_1 \mathbf{X}_1)^{-1}} \right)^{1/p} \quad (7)$$

onde \mathbf{X}_1 e \mathbf{X}_2 são as matrizes \mathbf{X} para os dois planejamentos e p é o número de parâmetros dos planejamentos.

1.4.2 Métodos de calibração multivariada

Entre os métodos quimiométricos os que mais se aplicam a Química Analítica são os de calibração multivariada. A calibração multivariada pode ser definida como a 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^[39].

O processo de calibração multivariada é constituído basicamente de três fases: *a calibração, a validação e a previsão*. Na fase de calibração são construídos os modelos matemáticos e na fase de validação estes modelos são avaliados usando ferramentas estatísticas de diagnóstico. Na fase de previsão, os modelos matemáticos validados são aplicados para prever as concentrações ou propriedades de amostras desconhecidas.

A organização dos dados na calibração multivariada, ilustrada a partir da análise de dados espectroquímicos, é mostrada na **Figura 4**.

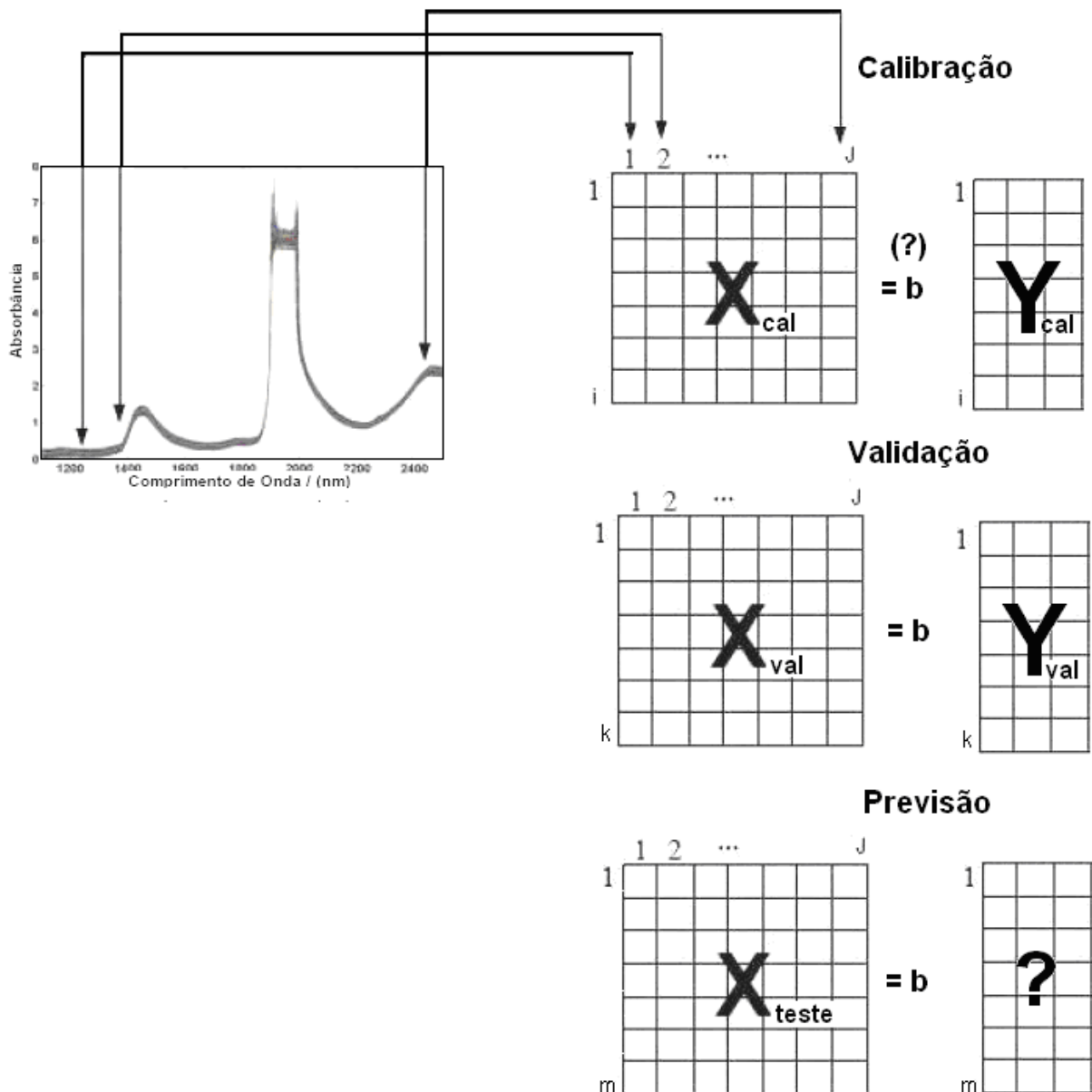


Figura 4: Organização dos dados na calibração multivariada.

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 X_{cal} ($i \times j$) para a fase de calibração e outra matriz X_{val} ($k \times j$) para a fase de validação. Duas matrizes Y_{cal} ($i \times q$) e Y_{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.

Uma variedade de métodos de regressão vem sendo utilizada em Química Analítica para a construção de modelos de calibração multivariada e dentre estes, os mais empregados têm sido a regressão por mínimos quadrados parciais (PLS, do inglês *Partial Least Square*) e a regressão linear múltipla (MLR, *Multiple Linear Regression*), que são métodos usados sempre para a construção de modelos lineares. Geralmente, 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 **X** para minimizar problemas de multicolinearidade^[40].

1.4.2.1 Regressão por mínimos quadrados parciais (PLS)

O método de calibração multivariada PLS foi desenvolvido por Herman Wold em meados de 1964^[41]. Para a construção dos modelos, o método utiliza tanto às informações da matriz de dados espectroscópicos (matriz **X**), como as informações da matriz de concentrações (matriz **Y**).

A regressão por mínimos quadrados parciais estende o conceito do modelo inverso (propriedade como função da resposta instrumental) trocando as variáveis originais por um sub-conjunto truncado das variáveis latentes dos dados originais^[42,43]. Considerando um caso geral para a determinação de mais de uma espécie de interesse, as matrizes **X_{cal}** e **Y_{cal}** (**Figura 4**) são decompostas em suas matrizes de pesos e escores^[33], respectivamente, como mostrado nas **Equações 8 e 9**.

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

$$\mathbf{Y} = \mathbf{UQ}^T + \mathbf{F} \quad (9)$$

Onde: **T** e **U** são as matrizes dos escores e **P** e **Q** são as matrizes dos pesos das matrizes **X** e **Y**, respectivamente. **E** é a matriz de resíduos espectrais e **F** é a matriz dos resíduos de concentração.

A decomposição em escores e pesos pode ser realizada empregando diferentes tipos de algoritmos. Todavia, eles produzem no final resultados similares. Um exemplo desses algoritmos é o NIPALS (do inglês, *Non-linear Iterative Partial Least Squares*)^[44].

O modelo final em PLS consiste em relacionar linearmente os escores da matriz **X** com os escores da matriz **Y**, de acordo com as seguintes equações:

$$\mathbf{U} = \mathbf{BT} + \mathbf{G} \quad (10)$$

$$\mathbf{Y} = \mathbf{BTQ}^T + \mathbf{H} \quad (11)$$

onde: **B** é a matriz dos coeficientes de regressão; **G** é a matriz de resíduos dos escores e **H** a matriz de resíduos de concentração.

1.4.2.2 Regressão linear múltipla (MLR)

Na construção de um modelo MLR cada variável dependente, isto é cada vetor **y_{cal}** da matriz **Y_{cal}**, é expressa como uma combinação linear das variáveis independentes da matriz **X_{cal}** e um vetor, **b_{MLR}**, que contém os coeficientes de regressão e é dada pela seguinte equação:

$$\mathbf{y}_{cal} = \mathbf{X}_{cal} \mathbf{b}_{MLR} + \mathbf{e} \quad (12)$$

O vetor **b_{MLR}** dos coeficientes lineares é estimado, na etapa de calibração, utilizando o critério dos "Mínimos Quadrados" que minimiza o vetor dos resíduos (**e = y_{cal} - X_{cal}b_{MLR}**) e é calculado por:

$$\mathbf{b}_{MLR} = \left(\mathbf{X}_{cal}^T \mathbf{X}_{cal} \right)^{-1} \mathbf{X}_{cal}^T \mathbf{y}_{cal} \quad (13)$$

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

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

$$\hat{\mathbf{y}}_{desc} = \mathbf{X}_{desc} \mathbf{b}_{MLR} + \mathbf{e} \quad (14)$$

A regressão linear múltipla apresenta alguns problemas que limitam sua aplicação. Um destes 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.

O outro problema do MLR é que a matriz $(\mathbf{X}_{cal}^T \mathbf{X}_{cal})$ pode não apresentar inversa devido a alta correlação entre as variáveis^[45]. Este problema pode ser solucionado usando um algoritmo de seleção de variáveis.

1.4.3 Algoritmos para seleção variáveis e amostras

O problema da seleção de variáveis em MLR tem sido assunto de pesquisas há mais de trinta anos^[46] e continua despertando interesse da comunidade científica^[47-50]. Neste contexto, vários algoritmos foram desenvolvidos, como por exemplo, o algoritmo "branch and bound", comumente aplicado em otimização combinatória^[51], o "generalized simulated annealing"^[52], transformada *wavelet*^[53,54], o algoritmo genético (*Genetic Algorithm - GA*)^[55-58], entre outros. Entre as diferentes estratégias de seleção de variáveis, o GA tem sido uma alternativa largamente usada; no entanto, devido à sua natureza estocástica, os resultados da seleção de variáveis feitos por essa técnica podem não ser reproduzíveis. Um método alternativo ao GA é o algoritmo das projeções sucessivas (*Successive Projections Algorithm - SPA*), que possui natureza determinística, ou seja, os resultados da seleção de variáveis feita por esta técnica são sempre reproduzíveis.

1.4.3.1 Algoritmo das Projeções Sucessivas

O algoritmo SPA, proposto por Araújo et al^[47], é um método iterativo de seleção de variáveis que opera na matriz de respostas instrumentais. Seu propósito é selecionar comprimentos de onda cujo conteúdo de informação seja minimamente redundante para resolver problemas de multicolinearidade. A essência do algoritmo SPA consiste de operações de projeção feitas na matriz de calibração \mathbf{X}_{cal} ($i \times j$), cujas linhas e colunas correspondem às amostras de calibração i e variáveis espectrais j , respectivamente.

Os passos para o SPA são descritos a seguir, assumindo que o primeiro comprimento de onda $j(0)$ e o número N são dados^[48].

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: S é 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:

$$\mathbf{P}\mathbf{x}_j = \mathbf{X}_j - (\mathbf{X}_j^T \mathbf{X}_{i(n-1)}) \mathbf{X}_{i(n-1)} (\mathbf{X}_{i(n-1)}^T \mathbf{X}_{i(n-1)})^{-1} \quad (15)$$

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

Passo 3: Faz-se $i(n)=\arg(\max \|\mathbf{P}\mathbf{x}_p\|, j \in S)$.

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 mostrada como $(N-1)(j - \frac{N}{2})$. Os passos acima são exemplificados na **Figura 5**, que ilustra a primeira iteração do SPA.

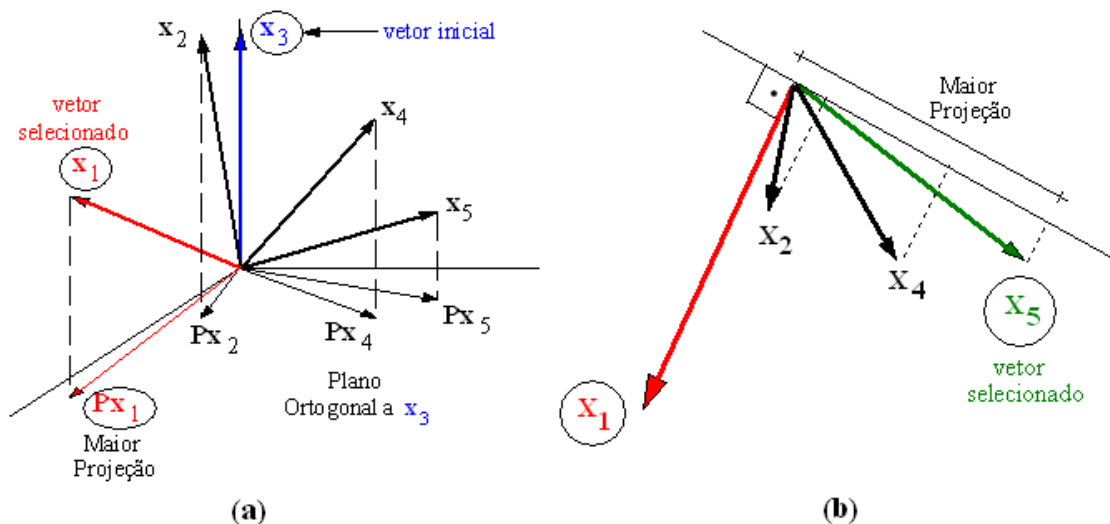


Figura 5: Visão espacial do funcionamento do SPA. (a) Projeção no plano ortogonal a x_3 e (b) visão espacial dos vetores no plano horizontal.

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. Assim, um total de $J \times I_c$ subconjuntos de variáveis pode ser formado. Para escolher o subconjunto mais apropriado, modelos MLR são construídos e comparados em termos do erro médio quadrático de previsão para o conjunto de validação (RMSEV), o qual é calculado por:

$$RMSEV = \sqrt{\frac{1}{I_v} \sum_{i=1}^{I_v} (y_v^i - \hat{y}_v^i)^2} \quad (16)$$

onde: 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.

A definição de um conjunto de validação representativo pode ser uma tarefa não trivial, este problema se torna mais aparente em aplicações analíticas envolvendo matrizes complexas, como óleo e produtos alimentícios, nos quais a variabilidade da composição não pode ser facilmente reproduzida por planejamentos experimentais otimizados. Neste caso o conjunto de validação deve ser extraído das amostras reais disponíveis para a construção do modelo. Neste

contexto, o uso das técnicas de validação cruzada pode ser uma alternativa adequada.

Recentemente, Galvão et al^[49] fizeram uma implementação ao algoritmo SPA, onde a técnica de validação cruzada é aplicada para a validação interna do modelo.

Supondo que os dados de resposta instrumental estão dispostos em uma matriz \mathbf{X} de dimensões (I x J) tal que a *jésima* variável x_j esta associada com o *jésimo* vetor coluna $x_j \in \mathbb{R}^I$. 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 feitas na matriz \mathbf{X} , gerando J cadeias de M variáveis cada é realizada de maneira idêntica a que ocorre no SPA original.

A segunda fase do SPA consiste em avaliar os subconjuntos candidatos de variáveis extraídas das cadeias geradas na primeira fase. A terceira e última fase é o procedimento de eliminação das variáveis que não contribuem significativamente para a habilidade da predição do modelo MLR.

O critério da avaliação das fases 2 e 3 é a raiz quadrada do erro médio da validação cruzada (*Root Mean Square Error of Cross-Validation* – RMSECV).

$$RMSECV = \sqrt{\frac{1}{N_{cal}} \sum_{n=1}^{N_{cal}} (y_{cal,n} - \hat{y}_{cal,n})^2} \quad (17)$$

onde $y_{cal,n}$ é o valor de referência do parâmetro em consideração para a *nésima* amostra do conjunto de calibração, o qual contém N_{cal} amostras.

O valor de $\hat{y}_{cal,n}$ é obtido pelo procedimento *leave-one-out*, o qual consiste em remover a *iésima* amostra do conjunto de calibração. Em seguida é construído o modelo com as amostras restantes, e este é então aplicado a amostra removida. Este procedimento é aplicado na fase 2 e 3 do SPA. A fase 1 (operações de

projeção) não é repetida no procedimento de validação cruzada para moderar o trabalho computacional envolvido no processo.

1.4.3.2 Algoritmo Kennard-Stone

Outro problema importante na construção de modelos de calibração multivariada é a seleção de um subconjunto representativo a partir de um conjunto grande de amostras^[59-62]. Entre os métodos utilizados para este fim está o método de seleção aleatória (do inglês: *Random Search* - RS)^[63] que é uma técnica popular devido à sua simplicidade e também porque o grupo de dados é extraído do conjunto maior seguindo a distribuição estatística do conjunto inteiro. Porém, a RS não garante a representatividade do conjunto, nem previne problemas de extrapolação^[64]. Uma alternativa para contornar as deficiências do RS que é freqüentemente empregada, é o algoritmo Kennard-Stone (KS)^[65]. O KS visa cobrir o espaço multidimensional de uma maneira uniforme, maximizando as distâncias Euclidianas entre os vetores das respostas instrumentais (\mathbf{X}) das amostras selecionadas^[65]. Em um estudo de classificação de redes neurais feito por Wu e colaboradores^[62], o KS foi apontado como superior a RS^[64].

O algoritmo KS tem como objetivo selecionar um subconjunto significativo em um conjunto de N amostras. Para assegurar a uniformidade de distribuição de cada subconjunto ao longo do espaço de dados \mathbf{X} (respostas instrumentais), o KS segue o procedimento *stepwise* no qual uma nova seleção é feita nas regiões do espaço longe das amostras já selecionadas. Para este propósito, o algoritmo emprega a distância Euclidiana $d_{\mathbf{x}}(a,b)$ entre os vetores \mathbf{X} de cada par (a,b) de amostras calculado como:

$$d_{\mathbf{x}}(a, b) = \sqrt{\sum_{j=1}^J [X_a(j) - X_b(j)]^2}; \quad a, b \in [1, M] \quad (18)$$

Para dados espectrais, $X_a(j)$ e $X_b(j)$ são as respostas instrumentais no j -ésimo comprimento de onda para as amostras a e b , respectivamente, e j denota o número de comprimentos de onda no espectro. A seleção começa tomando o par (a_1, a_2) de amostras nas quais a distância $d_x(a_1, a_2)$ é maior. A cada iteração subsequente, o algoritmo seleciona a amostra que exibe a maior distância mínima em relação a uma amostra já selecionada. Cada procedimento é repetido até que o número de amostras especificado pelo analista seja alcançado.

1.4.3.3 Algoritmo SPXY

Um método alternativo para divisão de conjuntos de calibração e validação foi proposto por Galvão et al.^[49], o método denominado SPXY (do inglês: *Sample set Partitioning based on joint \mathbf{X} - \mathbf{y} distances*), é uma extensão do algoritmo KS que leva em consideração tanto as diferenças de \mathbf{X} quanto a de \mathbf{y} no cálculo das distâncias inter-amostras. Este algoritmo foi empregado inicialmente em um problema de calibração envolvendo análises por espectrometria NIR de amostras de diesel e três parâmetros de qualidade foram determinados. O SPXY foi comparado com o KS e com o RS para divisão do conjunto de dados em calibração e validação para a regressão PLS. As performances dos modelos resultantes foram comparadas em termos do menor erro médio de previsão (do inglês: *Root Mean Square Error of Prediction* - RMSEP) calculado para o conjunto de previsão não incluído no processo de modelagem. Para assegurar a independência de cada conjunto, as amostras de previsão são aleatoriamente extraídas do conjunto inicial de dados.

Este algoritmo tem como proposta prever a distância definida na **Equação 18** com a distância no espaço variável dependente (y) do parâmetro em consideração. Cada distância $d_y(a, b)$ pode ser calculada para cada par de amostras a e b como:

$$d_y(a,b) = \sqrt{(y_a - y_b)^2} = |y_a - y_b|; a,b \in [1, N] \quad (19)$$

Para assegurar a igualdade de importância para a distribuição das amostras no espaço de \mathbf{X} e \mathbf{y} , as distâncias $d_{\mathbf{x}}(a,b)$ e $d_{\mathbf{y}}(a,b)$ são divididas pelos seus valores máximos no conjunto de dados. Desta maneira, uma distância \mathbf{xy} normalizada é calculada como:

$$d_{xy}(a,b) = \frac{d_x(a,b)}{\max_{a,b \in [1, N]} d_x(a,b)} + \frac{d_y(a,b)}{\max_{a,b \in [1, N]} d_y(a,b)}; \quad a,b \in [1, N] \quad (20)$$

Um procedimento de seleção *stepwise* similar ao algoritmo KS pode ser aplicado com $d_{xy}(a,b)$ ao invés de $d_{\mathbf{x}}(a,b)$ sozinho.

1.5 Validação analítica e figuras de mérito

Sempre que uma nova metodologia analítica é proposta faz-se necessário validá-la e esta validação pode ser feita através de parâmetros conhecidos como figuras de mérito. O número de figuras de mérito para serem determinadas na validação de um método e o nível que deve ser atingido em cada uma delas, depende de sua aplicação, do seu propósito e/ou do órgão de fiscalização a que está sujeito e encontram-se descritas em normas específicas^[67] e guias de validação^[68,69]. Um desses órgãos é a ASTM (do inglês: *American Society for Testing and Materials*) que possui uma norma específica para regular a validação de métodos analíticos baseados em espectrometria NIR.

A validação de uma metodologia NIR é conceitualmente diferente daquela das técnicas convencionais, como as descritas nas farmacopéias internacionais^[6-8], e é feita através de parâmetros quimiométricos. Esses parâmetros podem ser relacionados às características de validação fundamentais que são requeridas para qualquer método analítico. Alguns trabalhos científicos têm sido desenvolvidos e enfocam a determinação de figuras de mérito em

calibração multivariada. Muitas destas figuras de mérito têm seu cálculo estritamente ligado ao conceito de sinal analítico líquido (do inglês: *Net Analyte Signal* – NAS)^[70-72]. As principais figuras de mérito abordadas neste trabalho são:

- **Exatidão:** Este parâmetro expressa a concordância entre o valor estimado e o valor de referência. Em calibração multivariada a exatidão é estimada através da raiz quadrada do erro médio de predição (RMSEP)^[72]:

$$RMSEP = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n_v}} \quad (21)$$

onde n_v é o número de amostras de validação.

- **Seletividade:** Em calibração univariada, os coeficientes de seletividade ($\xi_{a,f}$) são fundamentais para conhecer como a quantidade variável de interferentes na amostra pode influenciar na predição. Os interferentes podem ser adequadamente modelados usando dados multivariados. Esta diferença essencial explica porque a estimativa numérica da seletividade em calibração multivariada é obtida de maneira diferente. Vários critérios foram publicados no passado. O critério de seletividade LBOZ enfoca na descrição de qual parte do sinal medido permanece por quantificação, isto é, o NAS, que na ausência de interferentes é apenas a medida do sinal analítico total. O critério natural para avaliar o impacto de todos os interferentes espectrais simultaneamente, é calculado como segue:

$$\xi_a = \frac{\|s_a^*\|}{\|s_a\|} \quad (22)$$

onde s_a representa o espectro do analito puro e s_a^* é associada a parte não sobreposta^[72]. Como o NAS é apenas uma parte do sinal analítico total, a seletividade é um parâmetro adimensional variando de 0 a 1. A mais baixa seletividade está associada a um alto grau de sobreposição espectral entre o analito e os outros constituintes da amostra.

- **Sensibilidade:** Este parâmetro é definido como o NAS gerado por uma concentração do analito igual à unidade. Para um modelo clássico, sensibilidade é a inclinação do modelo de calibração (pseudo-univariado). Conseqüentemente, sensibilidade é o inverso da inclinação deste gráfico para o modelo inverso. Por conseguinte, em termos do vetor dos coeficientes de regressão, a sensibilidade é dada por $1 / \|b\|$ [72].
- **Limite de detecção:** Limite de detecção é a menor quantidade do analito que pode ser detectada quantitativamente. Este é um parâmetro necessário para a análise de traços e pode ser necessário para testes de uniformidade de conteúdo e dissolução [72].
- **Precisão:** É a medida do erro aleatório associado ao método analítico. Expressa a proximidade de uma série de resultados obtidos a partir de diferentes amostras de um produto perfeitamente homogêneo. Os resultados podem ser expressos em termos do desvio padrão absoluto ou relativo [72].
- **Robustez:** É a avaliação da susceptibilidade do método de análise às variações das condições analíticas, por exemplo, variações de reagentes e variações instrumentais. Caso sejam determinados parâmetros que afetam o método de análise, estes deverão ser controlados e os procedimentos para o desenvolvimento da metodologia deverão incluir recomendações de como mantê-los sob controle [72].

1.5.1 Recomendações da norma ASTM E 1655 05 para calibração multivariada usando a espectrometria NIR

A ASTM é um órgão americano de normatização de vários materiais, produtos, sistemas e serviços. Foi fundada em 1898 nos Estados Unidos por um grupo de cientistas e engenheiros, liderados por Charles Benjamin Dudley, para analisar as freqüentes quebras de trilhos de trens. Como resultado, o grupo desenvolveu uma norma para o aço utilizado nas ferrovias. A ASTM produz normas para

diversas áreas da indústria, sendo muito usadas na padronização de materiais, como ligas de aço, alumínio, polímeros e combustíveis. Além disso, produz normas que indicam procedimentos de análise, como, por exemplo, para a determinação do tamanho de grãos em ensaios metalográficos (ASTM E112) e para obtenção da quantidade de benzeno e tolueno na gasolina de aviação (ASTM Method 3606). As normas ASTM para o desenvolvimento de análises quantitativas usando espectrometria NIR (ASTM E 1655 05) tratam de recomendações relativas elaboração das amostras dos conjuntos de calibração e validação interna, parâmetros estatísticos para avaliar, otimizar e recalibrar modelos de calibração, entre outras.

1.5.1.1 Recomendações gerais para a construção do modelo de calibração multivariada usando a espectrometria NIR

Entre as recomendações da ASTM E 1655 05 para o desenvolvimento de um modelo de calibração adequado para espectrometria NIR podemos citar:

- 1.** Em relação ao número de amostras, recomenda-se seja feita uma calibração exploratória, de modo que um modelo de calibração seja construído com uma quantidade entre 30 a 50 amostras, que deve cobrir a faixa do constituinte ou propriedade de interesse. A faixa de variação da concentração ou propriedade deve ser preferivelmente 5 vezes, e nunca menos que três vezes, o desvio padrão da reprodutibilidade da análise de referência.
- 2.** Quando da coleta dos dados espectrais, variações no tamanho da partícula, na apresentação da amostra e nas condições do processo que são esperadas durante as análises devem ser reproduzidas nas amostras do conjunto de calibração. Vários espectros da mesma amostra, sobre diferentes condições, podem ser empregados se as variações nas condições são antecipadas durante as análises.

- 3.** Análises de referência dessas amostras devem ser feitas usando o método de referência oficialmente aceito. Se a faixa para o componente ou propriedade não é pelo menos 5 vezes o desvio padrão da reprodutibilidade do método de referência, então r análises em replicata devem ser feitas para cada amostra, tal que \sqrt{r} vezes a faixa do componente ou propriedade seja preferencialmente igual a 5 vezes o desvio padrão do método de referência, e nunca menor que três vezes.
- 4.** O modelo de calibração deve ser desenvolvido usando MLR, PCR e/ou PLS e deve ser testado preferencialmente usando métodos de validação cruzada como SECV (Standard Error of Cross Validation) ou PRESS (*Prediction Residual Error Sum of Squares*). No entanto, outras ferramentas estatísticas podem ser usadas para julgar a qualidade global da calibração.
- 5.** Se o valor de SECV obtido da validação cruzada sugere que um modelo de precisão adequada pode ser construído, então amostras adicionais são coletadas com valores de concentração em torno dos valores das amostras do conjunto de calibração e irão servir como conjunto de validação. Espectros dessas amostras são coletados e um modelo final é desenvolvido e validado.

1.5.1.2 Recomendações para a seleção das amostras do conjunto de calibração

Para o desenvolvimento do modelo de calibração multivariada usando espectrometria NIR, o conjunto de calibração ideal deverá:

- 1.** Conter amostras englobando todos os componentes químicos que se espera estar presentes nas amostras que devem ser analisadas pelo modelo, assegurando que essas análises envolvem apenas interpolação do modelo, ou seja, que não ocorra previsões com extrapolação.
- 2.** Conter amostras, em que a faixa de variação na concentração dos componentes químicos exceda a faixa de variação esperada para as

amostras que devem ser analisadas usando o modelo, garantindo assim que as análises envolvem apenas interpolação do modelo.

- 3.** Conter amostras em que as concentrações dos componentes químicos estejam uniformemente distribuídas sobre sua faixa total de variação.
- 4.** Conter um número suficiente de amostras para definir estatisticamente a relação entre as variáveis espectrais e a concentração ou propriedade do componente a ser modelada.
- 5.** Para misturas simples, as amostras de calibração podem geralmente ser preparadas para satisfazer os critérios acima. Para misturas complexas obter um conjunto de calibração ideal é difícil, mas não é impossível. Os testes estatísticos que são usados para detectar amostras anômalas protegem contra conjuntos de calibração não-ideais. Os valores de RMSSR (*Root Mean Square Spectral Residuals*) detectam quando amostras a serem analisadas contêm componentes que não estão representados no conjunto de calibração (*violação do critério 1 acima*). As estatísticas de influência (*leverage*) detectam quando amostras a serem analisadas estão fora da faixa de concentração representada no conjunto de calibração (*violação do critério 2*). A detecção de amostras anômalas durante o desenvolvimento do modelo identifica componentes para os quais as faixas de concentração no conjunto de calibração não são uniformes (*violação do critério 3*).
- 6.** O número de amostras requeridas para calibrar um modelo multivariado utilizando espectrometria NIR depende da complexidade das amostras a serem analisadas. Se as amostras a serem analisadas contêm alguns componentes que variam em concentração, então haverá um pequeno número de variáveis (componentes principais, variáveis latentes, comprimentos de onda), e um conjunto relativamente pequeno de calibração é adequado para definir a relação entre as variáveis e as concentrações ou propriedades. Se um maior número de componentes varia nas amostras a serem analisadas, então um maior número de amostras de calibração é requerido para o desenvolvimento do modelo. Determinar se um conjunto de amostras de calibração é adequado ou não só é possível depois que o modelo for

desenvolvido e de ser feita uma estimativa do número de variáveis espectrais requeridas para o mesmo.

- 7.** Se o modelo multivariado é desenvolvido usando três ou menos variáveis, então o conjunto de calibração deve conter um mínimo de 24 amostras depois da eliminação das amostras anômalas.
- 8.** Se um modelo multivariado é desenvolvido usando $k (>3)$ variáveis então o conjunto de calibração deve conter um mínimo de $6k$ espectros depois da eliminação das amostras anômalas. Se o modelo é centrado na média, um mínimo de $6(k + 1)$ espectros devem restar, após a eliminação das amostras anômalas.
- 9.** Para algumas análises espectroscópicas, é possível calibrar usando misturas preparadas gravimetricamente ou volumetricamente que contenham significativamente menos componentes que as amostras que deverão ser finalmente analisadas. Para esses métodos “substitutos”, a estatística das amostras anômalas aqui descrita não é apropriada, pois amostras reais são, por definição, anômalas em relação a calibrações simplificadas. Assim, métodos “substitutos” não preenchem completamente os requisitos desta prática. Os métodos substitutos, no entanto, devem seguir as recomendações feitas com relação ao número e a faixa de calibração das amostras de calibração.

1.5.1.3 Recomendações para a seleção das amostras do conjunto de validação

A validação de um método multivariado usando espectrometria NIR é realizada pela aplicação do modelo para a análise de um conjunto de v amostras de validação, e estatisticamente comparado aos valores de referência conhecido para essas amostras. A validação requer um teste completo do modelo para assegurar que o seu desempenho atinge as expectativas derivadas das estatísticas do conjunto de calibração. Para a validação de um modelo multivariado o conjunto de amostras de validação

deve obedecer às especificações **1, 2, 3 e 5** descritas para o conjunto de calibração e :

- 4.** Conter um número suficiente de amostras para testar estatisticamente a relação entre as variáveis espectrais e as concentrações ou propriedades dos componentes que foram modeladas.
- 6.** O número de amostras necessárias para validar um modelo multivariado depende da complexidade do modelo. Apenas as amostras cujas análises são interpolações do modelo devem ser usadas no procedimento de validação. Se cinco ou menos variáveis espectrais são usadas no modelo, recomenda-se, então, um mínimo de 20 amostras de interpolação. Se $k > 5$ variáveis espectrais são usadas no modelo, então um mínimo de $4k$ amostras de interpolação devem ser usadas na validação. Além disso, as amostras de validação devem:
 - 6.a** Varrer a faixa de valores de concentração ou de propriedade das amostras de calibração, isto é, a amplitude e o desvio padrão destes valores para as amostras de validação deve ser pelo menos 95% daqueles das amostras de calibração, e os valores de concentração ou propriedade para as amostras de validação. Além disso, os valores de concentração e de propriedades devem estar distribuídos o mais uniformemente quanto possível ao longo da faixa e;
 - 6.b** Cobrir a faixa espectral de variáveis para a qual o modelo foi desenvolvido. Isto é, se a faixa das variáveis espectrais no modelo de calibração é de a até b e o desvio padrão das variáveis espectrais é c , então as variáveis espectrais das amostras de validação devem cobrir pelo menos 95% da faixa de a até b e deve estar distribuída tão uniformemente quanto possível ao longo da faixa de modo que o desvio padrão das variáveis espectrais das amostras de validação seja pelo menos 95% de c .
- 7.** A determinação de quanto o conjunto de validação é adequado irá geralmente depender do conjunto a ser analisado, assim como das variáveis espectrais deste conjunto que possam ser analisadas. Amostras para as quais as análises são extrapolações do modelo não devem ser incluídas no conjunto de validação. Se o conjunto de

validação não obedece aos critérios 6a e 6b, amostras de validação adicionais devem ser empregadas.

1.5.1.4 Questionário para avaliação do modelo de calibração multivariada.

O questionário de avaliação do modelo de calibração foi elaborado pela ASTM para ajudar o usuário a determinar se um modelo de calibração multivariada foi desenvolvido e validado de acordo com a norma ASTM E 1655 05. Se todas as perguntas abaixo forem respondidas afirmativamente, então a calibração multivariada pode ser considerada como desenvolvida e validada de acordo com a norma ASTM E 1655 05.

Perguntas relativas à metodologia matemática usada na calibração:

1. A técnica utilizada na calibração foi MLR, PCR ou PLS-1?
2. A metodologia de calibração é capaz de detectar amostras anômalas utilizando a estatística de influência, h ?
3. A metodologia é capaz de detectar amostras anômalas com base no resíduo espectral?

Perguntas relativas ao modelo de calibração:

4. n_{cal} ($n_{cal} = n^{\circ}$ de amostras de calibração) $> 6k$ ($k = n^{\circ}$ de variáveis espectrais para MLR, ou PC's para PCR ou variáveis latentes para PLS) para modelos não centrados na média) ou $n_{cal} > 6(k+1)$ para modelos centrados na média?
5. Foram utilizadas pelo menos 24 amostras de calibração?

Perguntas relativas à validação do modelo:

6. Um conjunto de amostras de validação foi separado e usado para testar a calibração?

7. As amostras anômalas detectadas com base na influência (distância de Mahalanobis) ou resíduo espectral foram excluídas do conjunto de validação?
8. O número de amostras de validação (n_v) foi maior que $4k$ (para modelos não centrados na média) ou maior que $4(k+1)$ (para modelos centrados na média)?
9. O número de amostra de validação é pelo menos 20?
10. As amostras de validação cobrem 95% da faixa das amostras de calibração?
11. Se o SEC (*Standard Error of Calibration*) é o erro padrão de calibração, 95% dos resultados das amostras de validação estão dentro de $\pm t \cdot \text{SEC} \cdot \sqrt{1+h}$ do valor de referência? (t é o valor *t-Student* para $n-k$ graus de liberdade se os dados não são centrados na média ou $n-k-1$ graus de liberdade se os dados são centrados na média; e h é a influência estatística).

Nota: *Os limites de confiança na estimativa para as amostras de validação devem ser calculados, e uma determinação deve ser feita para avaliar se os valores de referencia individuais das amostras de validação estão dentro da faixa de $\hat{y} - t \cdot \text{SEC} \cdot \sqrt{1+D^2}$ a $\hat{y} + t \cdot \text{SEC} \cdot \sqrt{1+D^2}$ ($D^2=h$). Se mais que 5% dos valores de referência cair fora da faixa, então o limite de confiança baseado no SEC é questionável, e mais testes são requeridos para mostrar a concordância entre o modelo e o método de referência.*

12. Os resultados da validação apresentam tendências estatisticamente não significativas?

Nota: *O teste t é usado para determinar se a estimativa da validação apresenta tendências (bias) estatisticamente significativas. O valor de t é calculado como:*

$$t_{\text{calculado}} = \frac{|\bar{e}_v| \sqrt{d_v}}{SDV} \quad (23)$$

Onde \bar{e}_v é o valor médio das tendências dos valores estimados para o conjunto validação; d_v é o número total de valores de referência que no nosso caso é igual a 1 já que foram usados valores individuais das medidas de referência; SDV é o desvio padrão do erro de validação.

O valor de $t_{\text{calculado}}$ é comparado ao valor de t tabelado para o número de grau de liberdade adequado.

- Se o valor de $t_{\text{calculado}}$ é menor que o valor de t tabelado, então é esperado que as análises baseadas no modelo multivariado apresentem o mesmo resultado médio que aqueles obtidos pelo método de referência, determinando que as análises representam uma interpolação do modelo.
- Se o valor de $t_{\text{calculado}}$ é maior que o valor de t tabelado, existe 95% de probabilidade que a estimativa do modelo multivariado não apresente a mesma média de resultados do método de referência. A validade do modelo multivariado é suspeita neste caso, e mais investigações são necessárias para resolver as prováveis tendências indicadas.

13. A precisão do modelo é determinada usando $am_{\text{test}} \geq k \geq 3$ ($am_{\text{test}} =$ n° de amostras de teste para as quais foram feitas replicatas) e $r \geq 6$ ($r =$ n° de replicatas por amostra)?

Nota: A precisão dos valores estimados por um modelo multivariado usando espectrometria NIR é calculada usando medidas espectrais repetidas. O número de amostras para as quais foram feitas replicatas deve ser no mínimo igual ao número de variáveis usadas no modelo e nunca menor que três. As amostras usadas para as repetições devem cobrir ao menos 95% da faixa de concentração usada no modelo. Quando possível devem ser selecionadas amostras que assegurem que alguma variação em cada variável espectral é exibida entre as amostras. Ao menos 6 espectros devem ser coletados para cada amostra. Os espectros devem ser analisados e os valores devem ser estimados para cada espectro. A média estimada para cada amostra deve ser calculada e o desvio padrão entre os valores estimados deve ser obtido. Se y_{ij} é a estimativa para o j^{th} espectro de r_i total de espectros para a i^{th} amostra, então a média estimada para a amostra é:

$$\bar{\hat{y}}_i = \frac{\sum_{j=1}^{r_i} \hat{y}_{ij}}{r_i} \quad (24)$$

O desvio padrão da replicata estimada é calculado como:

$$\sigma_i = \sqrt{\frac{\sum_{j=1}^{r_i} (\hat{y}_{ij} - \bar{\hat{y}}_i)^2}{r_i - 1}} \quad (25)$$

Um valor χ^2 é calculado usando o valor do desvio padrão calculado:

$$\chi^2 = \frac{2.3026}{c} (r \log \sigma^2 - \sum_{j=1}^{r_i} r_i \log \sigma_i^2) \quad (26)$$

no qual,

$$r = \sum_{i=1}^{am_{teste}} r_i \quad (27)$$

$$\sigma = \sqrt{\frac{1}{r} \sum_{i=1}^{am_{teste}} r_i \sigma_i^2} \quad (28)$$

$$c = 1 + \frac{1}{3(z-1)} \left(\sum_{i=1}^z \frac{1}{r_i} - \frac{1}{r} \right) \quad (29)$$

E z é o número de amostras para a qual as replicatas foram feitas.

O valor de χ^2 calculado é comparado com o valor tabelado para $t-1$ graus de liberdade.

- Se o valor χ^2 calculado é menor que o valor crítico, então todas as variâncias para as medidas replicadas pertencem à mesma população e a média da variância calculada pode ser usada como avaliação da repetitividade da medida no NIR. Uma medida no NIR tem repetibilidade esperada na ordem de $am_{teste} \cdot \sqrt{2\sigma}$.
- Se o valor calculado de χ^2 é maior que o valor chi-quadrado tabelado, então a repetitividade estimada da medida no NIR deve variar com a composição da amostra. Neste caso, a repetitividade esperada da análise no NIR é da ordem de $am_{teste} \cdot \sqrt{2} \cdot \sigma_{max}$, onde σ_{max} é o valor σ_i máximo para as medidas replicadas.

14. Se a calibração inclui pré ou pós-tratamento, eles foram feitos automaticamente?

1.6 Preparação e seleção das amostras de calibração

Um passo essencial para construção de um modelo de calibração multivariado adequado para ser adaptado à linha de produção é a preparação e a seleção das amostras do conjunto de calibração. Em geral, o uso de amostras apenas da linha de produção pode resultar em modelos sub-ajustados (*underfitting*) susceptíveis a erros de previsão por extrapolação. Isto ocorre porque as amostras da linha de produção não cobrem de maneira adequada à faixa de

concentração em que podem ser encontradas as amostras desconhecidas.

Blanco e Alcalá^[32] descreveram o uso de três alternativas que podem ser utilizadas para a construção de conjunto de calibração. A primeira envolve a preparação de amostras de calibração cobrindo a faixa apropriada de concentração dos diferentes componentes da amostra, usando uma planta piloto do processo de produção industrial. Esse procedimento é complexo e inviável devido à infraestrutura requerida e aos altos custos envolvidos. Além disso, ele produz uma grande quantidade de amostras em uma mesma faixa de concentração que não servirão para a construção do modelo já que bastariam algumas amostras replicatas de cada faixa de concentração. Dessa forma, uma grande quantidade de amostras necessita ser descartada, gerando desperdício de matéria-prima e resíduos para o meio-ambiente. A segunda alternativa consiste em utilizar amostras da produção, super e sub-dosadas com o princípio ativo. No entanto, a alta correlação entre as amostras que resultam deste tipo de procedimento pode levar à resultados espúrios se o modelo for aplicado a amostras que possuam diferentes concentrações de excipientes.

A terceira alternativa relata a preparação de amostras em laboratório, pela mistura do princípio ativo e excipientes em quantidades apropriadas cobrindo a faixa de concentração desejada para cada componente. Os modelos de calibração resultantes do uso das amostras dos conjuntos de calibração preparadas dessa maneira podem ser melhores do que aqueles obtidos com os conjuntos preparados usando as duas alternativas anteriores. Todavia, como as amostras preparadas em laboratório não possuem as variações físicas inerentes às amostras da linha de produção, esses conjuntos resultam em modelos subajustados susceptíveis a erros de previsão. Uma alternativa para superar este problema é incluir neste conjunto

de calibração algumas amostras da linha de produção. Entretanto, um estudo minucioso deve ser feito de modo a selecionar quais e quantas amostras da linha de produção devem ser incluídas.

1.7 Detecção de amostras anômalas

Uma das exigências da ASTM E 1655 05 para a construção do conjunto de calibração multivariada é que ele não possua amostras anômalas. Uma ferramenta utilizada para este fim é o gráfico de influência (*leverage*), que é mostrado na **Figura 6**.

O eixo X da **Figura 6** representa a *leverage* das amostras (h_i) e o eixo Y o resíduo de concentração de Student (rcS_i) calculados pelas **Equações (30)** e **(31)**. As linhas **A** e **B** verticais representam respectivamente a média da *leverage* das amostras e três vezes sua média. As linhas **C** e **E** indicam os níveis de confiança do resíduo espectral de Student ($\pm 2,5$), enquanto a linha **D** representa valor igual a zero deste resíduo. Esses limites são os mesmos adotados por Beebe et al^[73].

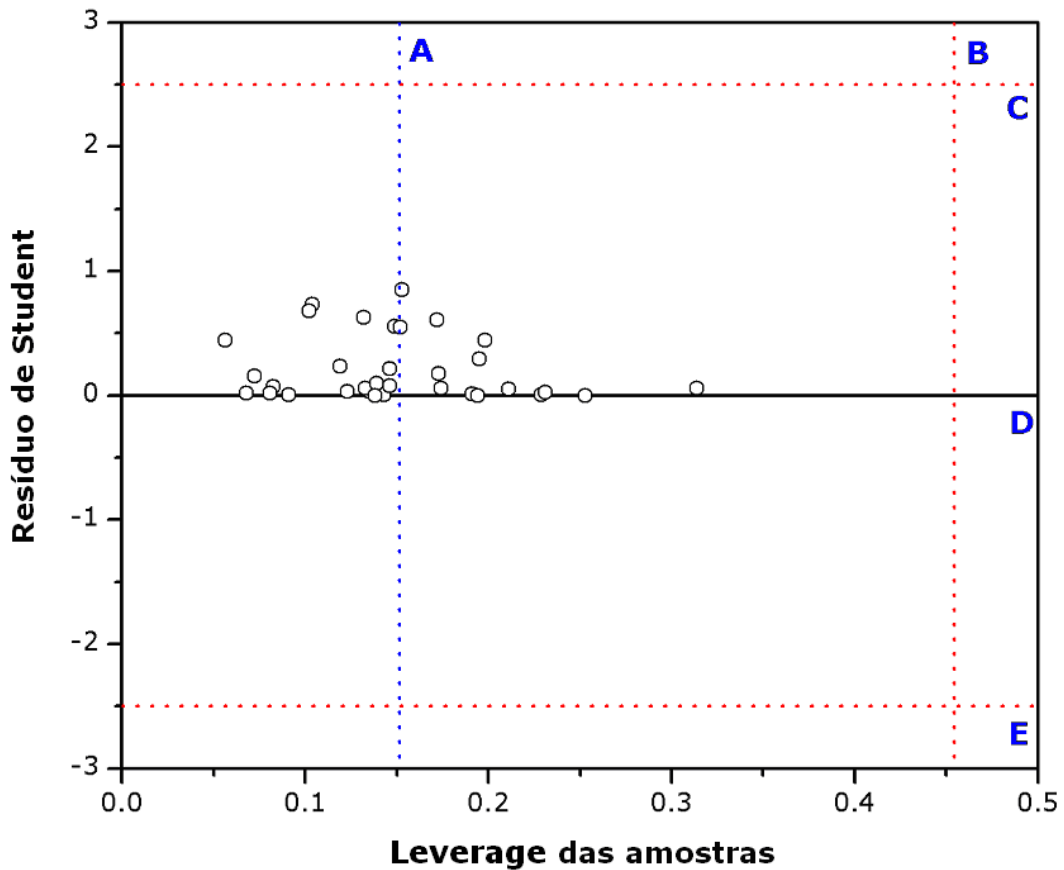


Figura 6: Gráfico da concentração residual de Student versus influência das amostras.

A leverage de uma amostra (h_i) pode ser calculada por^[73]:

$$h_i = \frac{1}{i} + \mathbf{u}_i^T (\mathbf{U}^T \mathbf{U}) \mathbf{u}_i \quad (30)$$

onde i é o número de amostras do conjunto de calibração, \mathbf{u}_i é o vetor de escores da amostra i , e \mathbf{U} é a matriz de escores do conjunto de calibração truncado para o número apropriado de fatores ou variáveis latentes.

A medida da influência de uma amostra em um modelo de regressão, quando calculada para os dados do conjunto de calibração, tem valor máximo igual a 1. Um valor de influência pequeno indica que a amostra influencia pouco na construção do modelo de calibração. Segundo Beebe et al^[73], amostras com valores de

influência que são significativamente maiores que o resto das amostras (isto é, maior que três vezes a média das influências) devem ser inspecionadas com mais cuidado. No entanto, elas não são necessariamente ruins (ou anômalas), elas apenas têm alta influência no modelo. Estas amostras devem ser investigadas com mais cuidado porque as amostras anômalas têm alta influência e alto efeito adverso no modelo. Por outro lado, as amostras com alto resíduo espectral também terão alta influência mas não necessariamente adversa pois podem trazer informação não contida nas demais amostras. Portanto, vale ressaltar que as amostras anômalas (*outliers*) têm sempre alta influência, mas também alto resíduo espectral.

O resíduo de espectral de *Student* (rxS_i) de uma amostra é o resíduo espectral ($x_i - \hat{x}_i$), convertido em unidades de desvio padrão pela equação

$$rxS_i = \frac{x_i - \hat{x}_i}{s\sqrt{1 - h_i}} \quad (31)$$

onde, x_i e \hat{x}_i são respectivamente os espectros conhecidos e previstos pelo modelo construído para a *i*-ésima amostra; h_i é a influência desta amostra calculada pela **Equação 30**, e s é o desvio padrão dos resíduos espectrais das amostras do conjunto de calibração, calculada pela equação abaixo, onde n_{VL} é o número de fatores ou variáveis latentes usados na construção do modelo.

$$s = \sqrt{\frac{\sum_{i=1}^i (x_i - \hat{x}_i)^2}{i - n_{VL} - 1}} \quad (32)$$

O rxS_i mede quão bem os espectros da *i*-ésima amostra está ajustado ao modelo. Um alto valor de rxS_i positivo ou negativo indica amostras que não foram bem ajustadas ao modelo. $|rxS_i| > 2,5$ são considerados altos, pois como eles são expressos em unidade de

desvio padrão de um valor médio, valores além de $\pm 2,5$ são considerados pouco comum segundo as suposições estatísticas usuais.

1.8 Objetivos

1.8.1 Objetivo geral

O objetivo deste trabalho foi desenvolver metodologias analíticas baseadas em espectrometria NIRR e calibração multivariada para a determinação de captopril em comprimidos farmacêuticos produzidos pelo LAFEPE.

1.8.2 Objetivos específicos

- Construir modelos de calibração multivariada baseados em PLS-1 e MLR,
- Aplicar técnicas quimiométricas de planejamento experimental, de seleção de variáveis e de amostras, para construir os modelos de calibração,
- Validar as metodologias desenvolvidas com base na norma ASTM E 1655 05,
- Sugerir qual o melhor método a ser implantado pelo LAFEPE para o controle dos comprimidos de captopril produzidos e comercializados por esta empresa.



Capítulo 2

Experimental

2.1 As amostras

O conjunto de calibração é composto tanto de amostras sintéticas quanto de amostras comerciais. Um total de 35 amostras sintéticas foram preparadas de acordo com um planejamento de misturas D-otimizado (incluindo triplicata autêntica do ponto central). Cada uma dessas 35 amostras sintéticas foi preparada em triplicata. As amostras comerciais são compostas por 50 diferentes lotes de comprimidos produzidos pelo LAFEPE.

A capacidade preditiva dos modelos de calibração multivariada construídos foi avaliada utilizando um conjunto de predição constituído de amostras produzidas comercialmente pelo LAFEPE.

2.1.1 Comprimidos sintéticos

Os comprimidos sintéticos foram elaborados utilizando um planejamento de misturas D-otimizado do pacote quimiométrico Unscrambler[®]. A **tabela 2** mostra o planejamento de misturas, cada linha da tabela representa uma das 35 composições utilizada na preparação dos comprimidos sintéticos. As três últimas linhas da tabela representam as repetições autênticas do ponto central.

Tabela 2: Planejamento de misturas D-otimizado de misturas para a preparação das amostras sintéticas.

Princípio ativo (% massa)	Excipiente A (% massa)	Excipiente B (% massa)	Excipiente C (% massa)	Excipiente D (% massa)	Excipiente E (% massa)
0.200	0.353	0.418	0.012	0.005	0.012
0.133	0.530	0.312	0.008	0.005	0.012
0.200	0.353	0.422	0.008	0.005	0.012
0.200	0.353	0.420	0.012	0.003	0.012
0.133	0.403	0.441	0.008	0.003	0.012
0.200	0.483	0.294	0.008	0.003	0.012
0.133	0.530	0.312	0.012	0.005	0.008
0.185	0.353	0.441	0.008	0.005	0.008
0.133	0.405	0.441	0.008	0.005	0.008
0.200	0.485	0.294	0.008	0.005	0.008
0.133	0.403	0.441	0.012	0.003	0.008
0.153	0.530	0.294	0.012	0.003	0.008
0.133	0.407	0.441	0.008	0.003	0.008

0.200	0.353	0.426	0.008	0.003	0.010
0.200	0.353	0.426	0.010	0.003	0.008
0.200	0.481	0.294	0.012	0.003	0.010
0.200	0.482	0.294	0.012	0.004	0.008
0.200	0.482	0.294	0.008	0.004	0.012
0.200	0.417	0.358	0.012	0.005	0.008
0.200	0.420	0.361	0.008	0.003	0.008
0.133	0.530	0.309	0.012	0.004	0.012
0.133	0.530	0.317	0.008	0.004	0.008
0.133	0.530	0.312	0.010	0.003	0.012
0.133	0.399	0.441	0.012	0.005	0.010
0.133	0.399	0.441	0.010	0.005	0.012
0.155	0.530	0.294	0.008	0.003	0.010
0.182	0.353	0.441	0.012	0.004	0.008
0.156	0.376	0.441	0.012	0.003	0.012
0.174	0.503	0.294	0.012	0.005	0.012
0.200	0.481	0.294	0.010	0.005	0.010
0.133	0.465	0.377	0.012	0.003	0.010
0.166	0.441	0.367	0.008	0.005	0.012
0.167	0.442	0.368	0.010	0.004	0.010
0.167	0.442	0.368	0.010	0.004	0.010
0.167	0.442	0.368	0.010	0.004	0.010

Nesta elaboração, foi levada em conta a formulação (princípio ativo e excipientes) usada pelo LAFEPE para a produção dos comprimidos comerciais com valor nominal de 25 mg de captopril e uma massa total de cerca de 150 mg.

Também foram consideradas as recomendações (item 1.5.1.2) da norma da ASTM E 1655 05 quanto à preparação do conjunto de calibração usado na construção de modelos de calibração multivariada empregando a espectrometria NIR.

De acordo com o planejamento D-otimizado, foi gerado um conjunto de calibração com um total de 35 composições para as amostras sintéticas, incluindo triplicatas autênticas do ponto central. A concentração do princípio ativo captopril e dos excipientes foi variada em $\pm 20\%$ em torno dos valores especificados na formulação do LAFEPE, de modo que as 35 amostras sintéticas cobrem a faixa ampliada de concentração de 20 a 30 mg de captopril, em relação ao valor nominal esperado (25 mg) para os comprimidos comerciais produzidos pela empresa.

2.1.1.1 Procedimento de preparação dos comprimidos sintéticos

Para a produção dos comprimidos sintéticos foram utilizados:

- Almofariz e pistilo para misturar o princípio ativo e os excipientes;
- Prensa e molde (**Figura 7**) para produzir as amostras sintéticas na forma de comprimidos;



Figura 7: Prensa (a) e forma (b) *labmade* usada para prensar os comprimidos.

Na preparação dos comprimidos sintéticos foi adotado sempre o seguinte procedimento:

- Para minimizar o erro na pesagem dos ingredientes, resolveu-se pesar, em separado, as quantidades de excipientes e de princípio ativo para a preparação de 100 comprimidos de cada uma das 35 composições sugeridas pelo planejamento D-otimizado. Cada um dos comprimidos deveria apresentar uma massa total de cerca de 150 mg, como os comprimidos comerciais do LAFEPE. Das massas pesadas foram utilizadas apenas as quantidades relativas à preparação de 10 comprimidos de cada amostra do planejamento, sendo descartado o restante. Isto resulta em um total de 350 comprimidos a serem preparados.

- A quantidade de cada excipiente foi misturada, o mais homogeneamente possível e espalhada sobre as paredes do almofariz, para só depois adicionar o próximo ingrediente de forma sucessiva e aleatória. O princípio ativo foi sempre adicionado por último. Esse cuidado foi tomado para promover a homogeneidade das misturas de ingredientes e, conseqüentemente, a homogeneidade do princípio ativo captopril nos comprimidos;
- A massa pesada de cada um dos 350 comprimidos foi prensada na forma de um comprimido usando uma prensa (a) e uma forma (b) construída no laboratório, conforme ilustrado na **Figura 7**.
- Os 10 comprimidos de cada uma das 35 amostras do planejamento foram acondicionados em recipientes de vidro âmbar e guardados em dessecador até que fossem registrados os espectros NIRR.

Por ter sido assinado um termo de confidencialidade a formulação dos comprimidos será mantida em sigilo.

2.1.2 Amostras analisadas

Foram usadas na construção e validação interna dos modelos 3 comprimidos (repetições não-autênticas) de cada uma das 35 composições sugeridas pelo o planejamento D-otimizado e 50 comprimidos retirados de cada um dos 50 lotes de fabricação diferente. Assim, havia disponível um total de 105 amostras de comprimidos sintéticos e 750 amostras de comprimidos comerciais.

2.2 Registro dos espectros NIRR

Os espectros NIRR das 855 amostras foram registrados na faixa de 4.000 a 14.000 cm^{-1} utilizando um espectrômetro NIR com transformada de Fourier da marca BOMEM, modelo MB 160 D (**Figura 8**).



Figura 8: Espectrofotômetro NIR, modelo ABB BOMEM.

Para o registro dos espectros foi empregado o acessório *Powder Samplr* e um suporte construído no laboratório para adaptação dos comprimidos no percurso óptico (**Figura 9**).

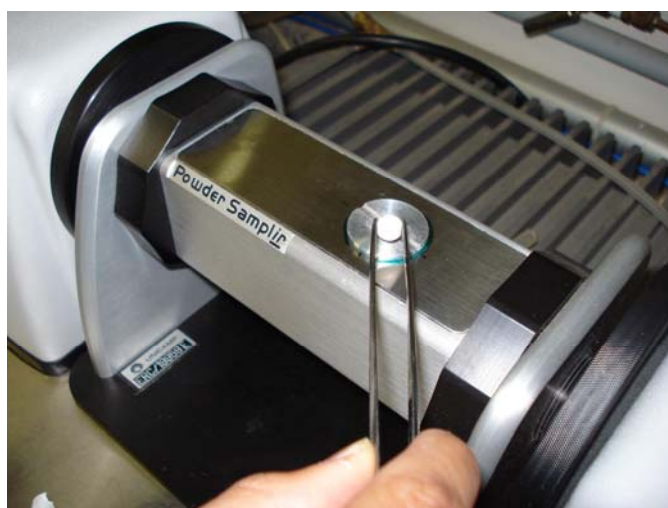


Figura 9: Suporte para centralizar a amostra sob a fonte de radiação.

No registro do espectro do branco foi utilizado Teflon (99% reflective Spectralon, Labsphere me., North Sutton, NH, USA). Cada espectro registrado foi resultado da média de 60 varreduras realizadas utilizando uma resolução de 4 cm^{-1} . Nestas condições, o tempo de

obtenção de um espectro foi de cerca de 60 segundos. Todos os espectros foram registrados em ambiente com temperatura (em torno de 25 °C) e umidade relativa (cerca de 45%) controladas.

Os espectros dos comprimidos sintéticos foram sempre registrados nas duas faces. Todavia, devido à presença de um sulco de meia dosagem em uma das faces dos comprimidos comerciais do LAFEPE, seus espectros foram registrados apenas na face não-sulcada.

2.1.3 Conjuntos de calibração e validação

O conjunto de calibração foi composto de 35 espectros de comprimidos sintéticos, sendo cada um a média de seis espectros obtidos nas duas faces de cada um dos 3 comprimidos das 35 amostras sugeridas pelo o planejamento D-otimizado, e de 18 espectros de comprimidos comerciais do LAFEPE, sendo cada um a média de 15 espectros dos 15 comprimidos retirados de cada um dos 50 lotes de fabricação. Esses 18 espectros foram selecionados usando o algoritmo SPXY.

O conjunto de validação foi composto pelos 32 espectros de comprimidos comerciais do LAFEPE que restaram após a seleção do algoritmo SPXY, sendo cada um a média de 15 espectros dos 15 comprimidos retirados de cada um dos 50 lotes de fabricação diferente.

2.1.4 Valores de referência dos comprimidos do LAFEPE

O teor de captopril dos comprimidos de cada um dos 50 lotes foi determinado pelo LAFEPE empregando o método HPLC de referência, recomendado pela farmacopéia americana. Esses valores de referência foram usados na calibração e predição dos modelos.

2.3 Programas estatísticos

O programa UNSCRAMBLER[®] v9.1 da CAMO Process A.S foi usado para gerar o planejamento de misturas D-Otimizado, para fazer a

regressão em mínimos quadrados parciais e a análise de influência das amostras industriais inseridas no modelo de calibração multivariada.

Os programas SPXY^[49], MLR-SPA^[50] e MCV1^[74] foram elaborados e executados empregando o ambiente MATLAB[®], versão 6.5. Estes programas foram usados, respectivamente, para a seleção de amostras do conjunto de calibração, para seleção de variáveis em MLR e para o cálculo de algumas figuras de mérito analítico.



Capítulo 3

Resultados & Discussão

3.1 Pré-processamento dos espectros NIRR do captopril

Antes de se iniciar a etapa de construção dos modelos quimiométricos, freqüentemente é necessário realizar uma etapa de pré-processamento nos espectros. Esta prática visa remover qualquer informação que não esteja relacionada ao problema químico em estudo, que pode ocorrer devido a vários fatores, tais como: instabilidade do espectrômetro NIR ao longo da faixa espectral de medida, variabilidade das propriedades físicas das amostras, ou ainda a problemas de espalhamento da radiação durante as medidas.

3.1.1 Escolha da faixa espectral de trabalho

Os espectros dos comprimidos de captopril foram registrados na faixa de 4.000 a 14.000 cm^{-1} e um espectro representativo de um comprimido de captopril é mostrado na **Figura 10**.

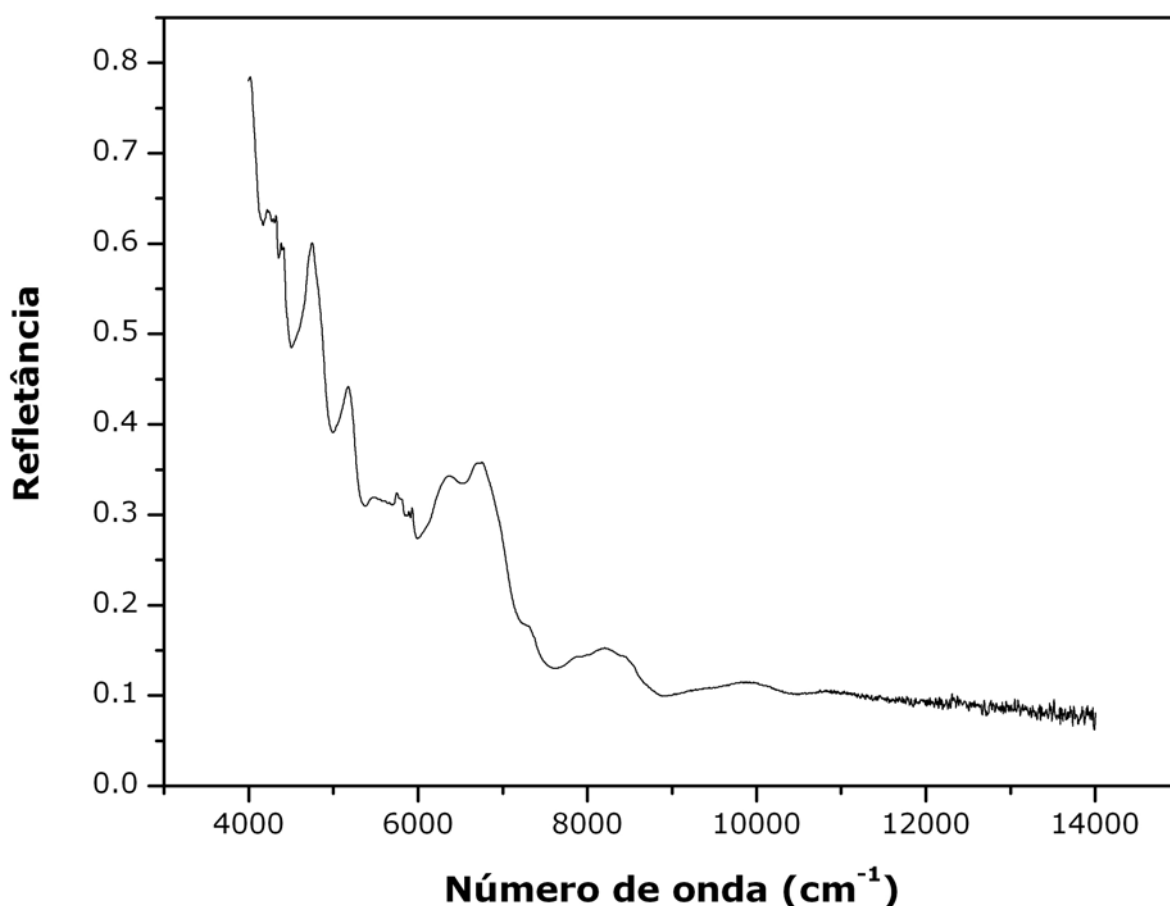


Figura 10: Espectro NIRR de um comprimido de Captopril.

Todavia a região do espectro posterior a 9.000 cm^{-1} é bastante ruidosa. Isto ocorre porque nesta região o detector à base de Arseneto de Gálio e Índio (AsGaIn)^[75] do espectrômetro Bomen, com a qual foi registrados os espectros NIRR dos comprimidos de captopril, apresenta uma baixa relação sinal/ruído. Este mesmo problema também ocorre na região de 4.000 a 4.500 cm^{-1} . Assim, estas regiões foram descartadas para evitar à inserção de informação ruidosa aos modelos quimiométricos e a região espectral de 9.000 a 4.500 cm^{-1} , destacada na **Figura 11**, foi adotada como região de trabalho em todo o tratamento quimiométrico daqui por diante.

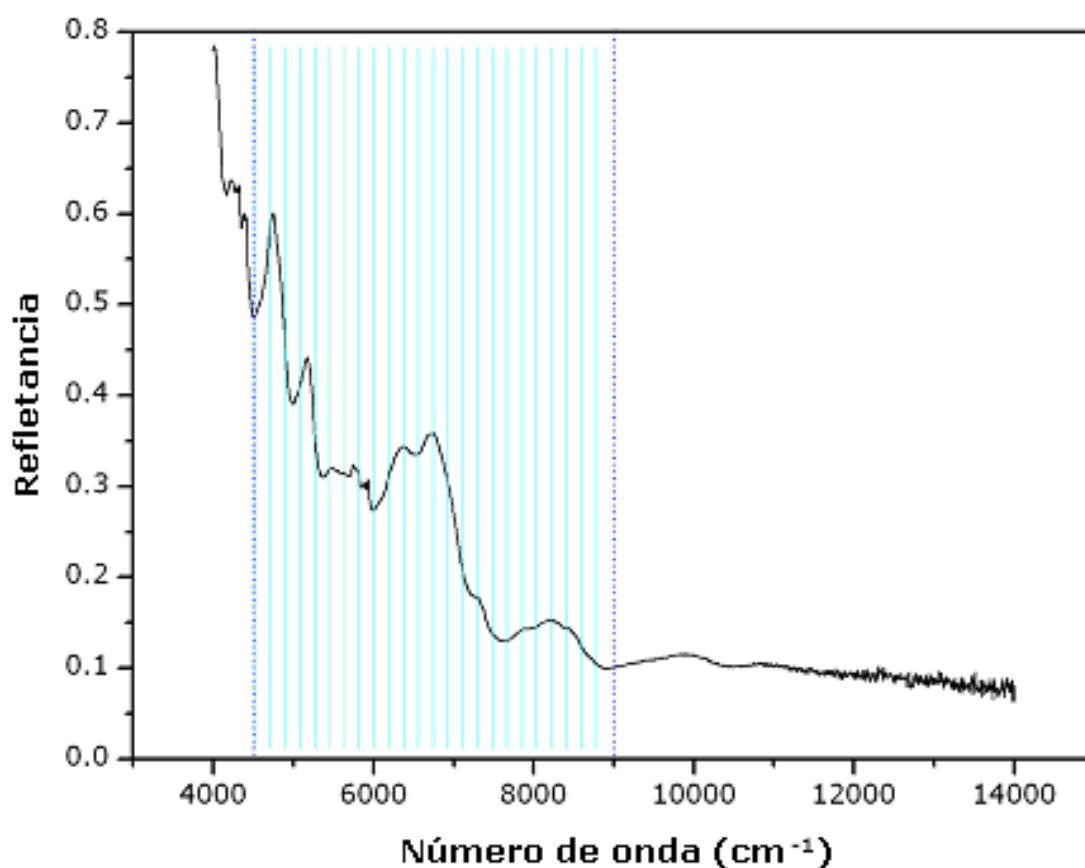


Figura 11: Faixa espectral de trabalho.

3.1.2 Estudo da necessidade de pré-processamento

Escolhida a faixa espectral de trabalho, foi feita uma avaliação quanto à necessidade de algum outro pré-processamento dos espectros NIRR dos comprimidos. Na **Figura 12** são mostrados os espectros NIRR

sem pré-processamento dos 35 comprimidos sintéticos (em azul) e dos 50 comerciais (em vermelho).

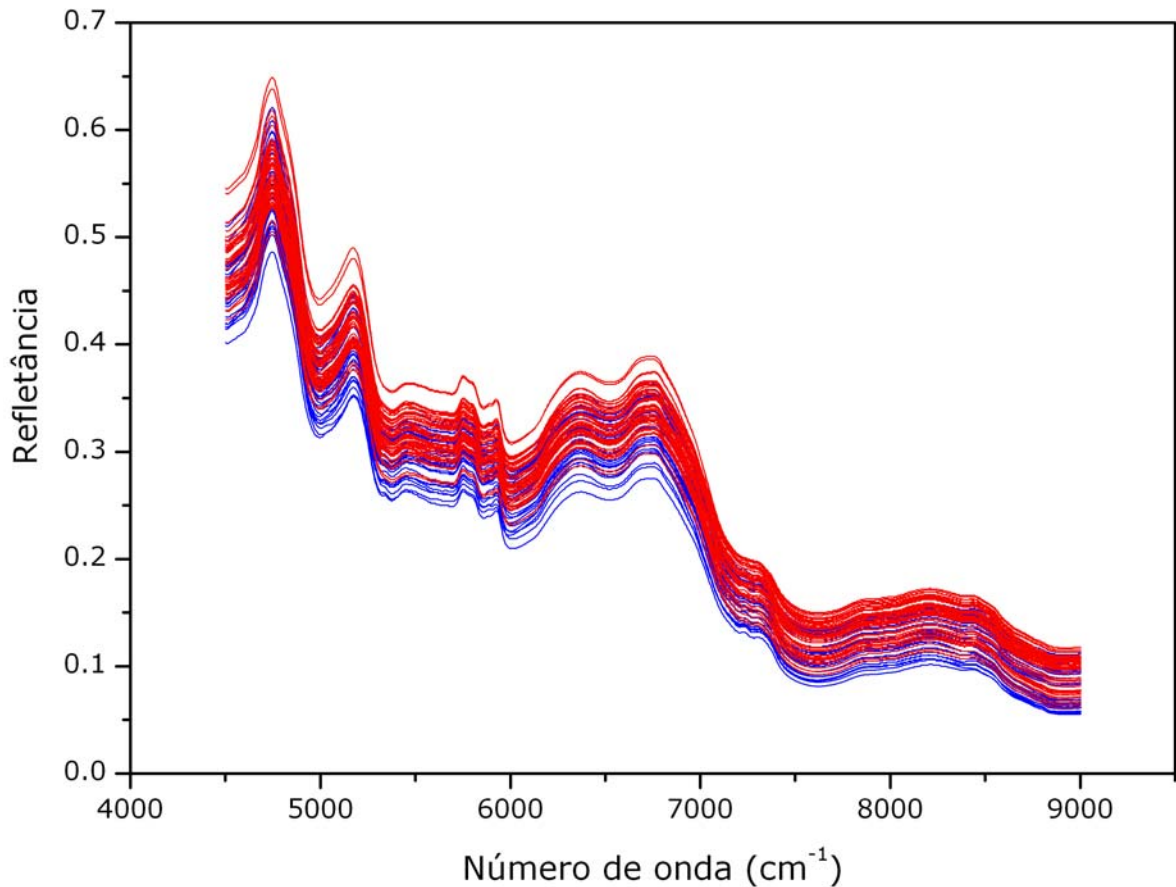


Figura 12: Espectros NIRR sem pré-processamento dos comprimidos de captopril sintéticos (em azul) e comerciais (em vermelho).

Pode-se notar na **Figura 12** que existe um deslocamento, em relação ao eixo de refletância, entre os dois conjuntos de espectros dos comprimidos sintéticos e dos comerciais possivelmente devido às características do processo de fabricação do comprimido, como por exemplo, granulometria e compactação dos comprimidos. De fato, as microcavidades existentes nos comprimidos causam pequenas mudanças no índice de refração, fazendo com que ocorram estes deslocamentos. Tais variações podem ser incluídas nas modelagens quimiométricas. Todavia, verifica-se também a presença de um pequeno componente ruidoso e este necessita ser minimizado empregando um processo de pré-processamento, antes da construção dos modelos quimiométricos. Para isso, testou-se várias técnicas de pré-processamento (por exemplo:

Multiplicative Scatter Correction-MSC, correção de linha de base, derivação, etc) e optou-se por usar a suavização Savitzky-Golay já que este método apresentou os menores erros. A melhor suavização por esse método foi obtida usando uma janela de 9 pontos, após terem sido testadas diferentes janelas. A **Figura 13** mostra que a presença de ruído no espectro NIRR bruto (sem pré-processamento) foi minimizada após o emprego deste pré-processamento. Assim, daqui por diante, na construção dos modelos foram sempre empregados os espectros NIRR suavizados pelo método Savitzky-Golay usando uma janela de 9 pontos.

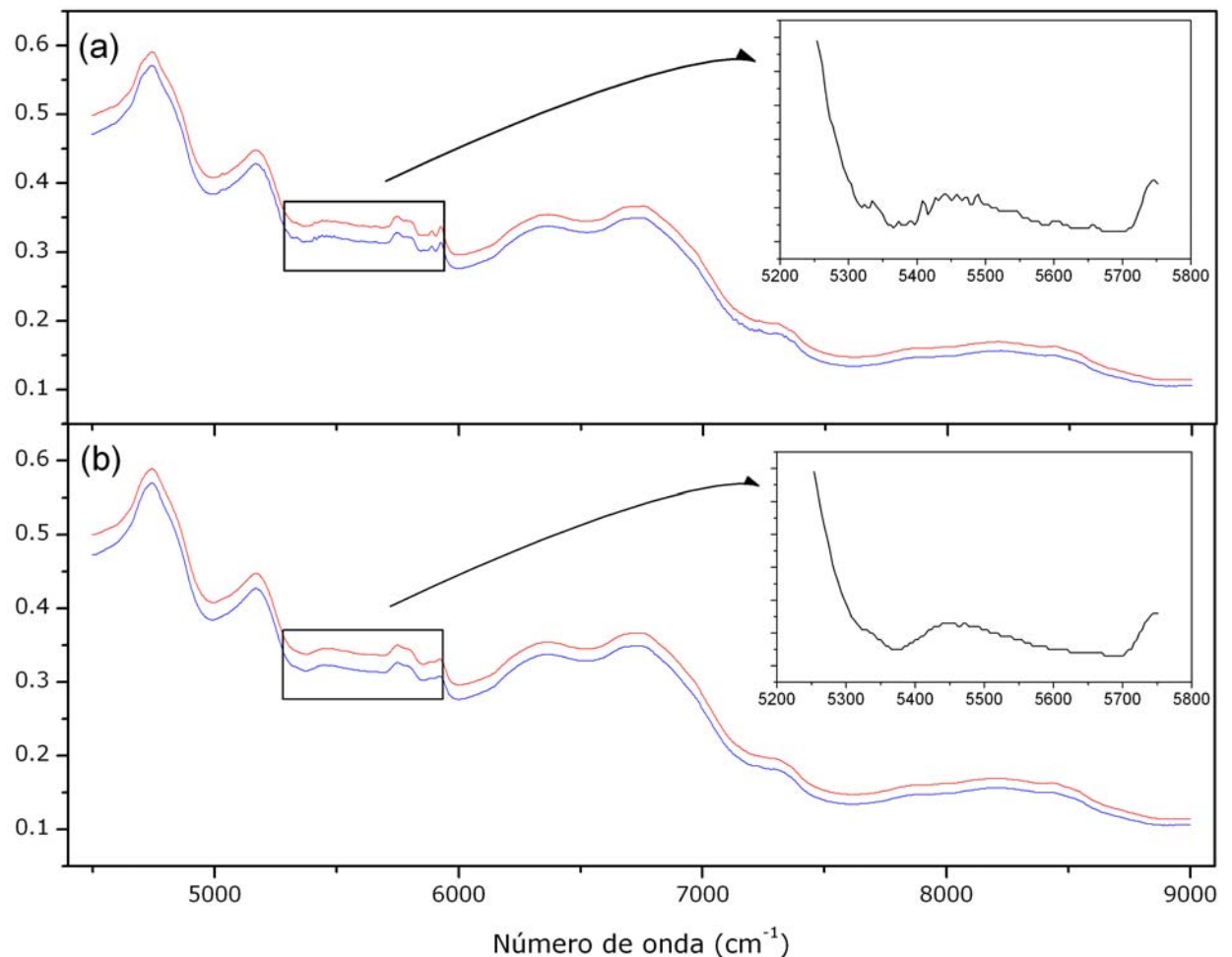


Figura 13: Um espectro NIRR representativo de um comprimido antes (a) e após (b) a suavização Savitzky-Golay com uma janela de 9 pontos. A suavização é mostrada no detalhe das figuras.

3.2 Construção dos modelos de calibração multivariada

Para a construção dos modelos de calibração multivariada foram empregadas e avaliadas duas estratégias. Na primeira estratégia, o conjunto de calibração foi composto apenas de amostras sintéticas e na segunda estratégia, o conjunto de calibração foi constituído de amostras sintéticas e de um número variável (entre 3 a 18) de amostras comerciais. A avaliação da capacidade preditiva dos modelos PLS-1 e MLR-SPA construídos foi feita usando um conjunto de 35 amostras comerciais do LAFEPE, o qual foi chamado de conjunto de predição.

Antes de iniciar efetivamente a modelagem, foi realizado um estudo para verificar a presença de amostras anômalas nos conjunto das amostras sintéticas e das amostras comerciais.

3.2.1 Detecção de amostras anômalas

Dois modelos PLS-1 foram construídos, um usando um conjunto de dados das amostras sintéticas e um outro usando apenas o conjunto de dados das amostras comerciais. Para tanto foram gerados os gráficos do resíduo de concentração de Student versus a influência das amostras para os dois modelos. Estes gráficos são apresentados na Figura 14. As amostras 6 e 26, figura 14(a), possuem uma alta influência no modelo, no entanto não se caracterizam como amostras anômalas pois possuem resíduo espectral dentro dos limites aceitáveis. Deste modo podemos concluir que o conjunto de amostras sintéticas, mostrado na figura 14(a) não possui amostras anômalas. Pelo mesmo raciocínio, podemos afirmar que o conjunto de amostras comerciais, figura 14(b), também não apresenta amostras anômalas.

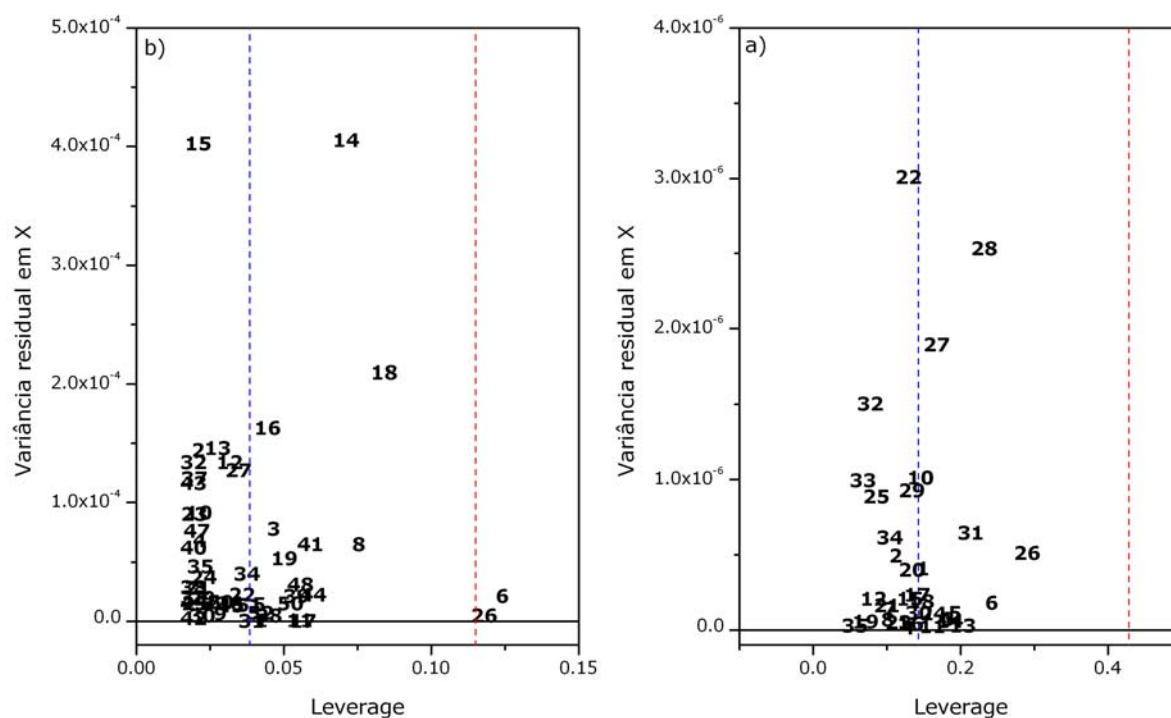


Figura 14: Variância residual em X versus leverage das amostras do modelo PLS-1 construído usando os dados de comprimidos sintéticos (a) e dos comerciais (b). A linha vertical tracejada azul representa a média da influência das amostras e a vermelha três vezes a média da influência.

Deste modo, tendo sido constatada a ausência de amostras anômalas, partiu-se para a construção dos modelos de calibração multivariada empregando PLS-1 e MLR-SPA. Na construção dos modelos PLS-1 e MLR-SPA foi sempre empregado o processo de validação cruzada.

3.2.2 Modelagem PLS-1 e MLR-SPA usando um conjunto de calibração contendo apenas amostras sintéticas

A **figura 15** mostra o gráfico das concentrações medidas versus concentrações previstas para a construção do modelo de calibração usando a técnica de validação cruzada para validação do modelo. Os círculos azuis representam a calibração e os vermelhos a validação. Na construção do modelo de calibração foi obtido um RMSEC de 0,4663 e uma correlação de 0,9942. E na validação do modelo foi obtido um RMSECV 0,5433 e uma correlação de 0,9921.

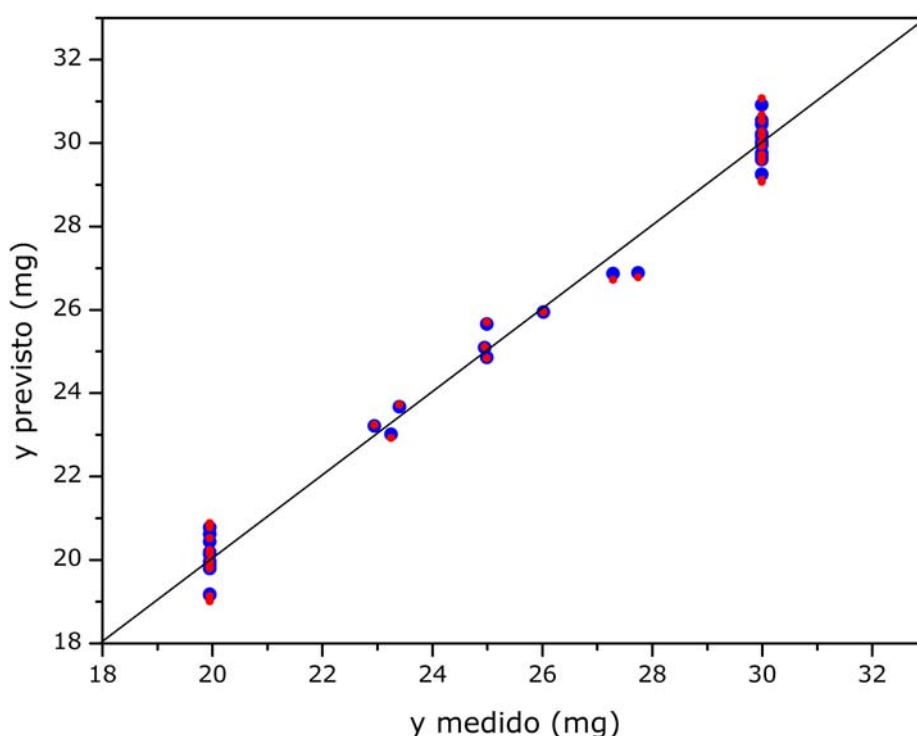


Figura 15: Gráfico das concentrações medidas versus as concentrações previstas na construção do modelo de calibração usando o método de validação cruzada.

A **Tabela 3** apresenta os valores de RMSEP obtidos no conjunto de previsão usando os modelos PLS-1 e MLR-SPA construídos com o conjunto de calibração contendo apenas amostras sintéticas. Pode-se notar que ambos os modelos obtiveram valores de RMSEP dentro dos limites aceitáveis pelo método de referência, de acordo com a Farmacopéia^[5], que estabelece um erro máximo permitido de $\pm 10\%$ do valor nominal (25,00 mg) ou seja, $\pm 2,50$ mg. Vale salientar que o LAFEPE também adota estes limites para o controle do teor de captopril em sua produção.

Tabela 3: Valores de RMSEP e número de variáveis latentes utilizadas ou comprimentos de onda utilizados nas modelagens PLS-1 e MLR-SPA, respectivamente.

Modelo	RMSEP	VL ou Variáveis selecionadas
PLS-1	2.3	4
MLR-SPA	2.0	14

O modelo PLS-1 foi parcimonioso escolhendo 4 variáveis latentes e o modelo MLR-SPA selecionou para a construção do modelo 14 comprimentos de onda dos 1168 originais.

Esses resultados podem ser considerados aceitáveis em relação à Farmacopéia. Entretanto os valores de RMSEP encontrados estão muito próximos do limite, possivelmente porque, embora as amostras do conjunto de calibração tenham a mesma matriz química das amostras do conjunto de previsão elas apresentam diferenças físicas, tais como granulometria, grau de compactação, etc, que são importantes para a modelagem.

Por isso, acreditando que os modelos PLS-1 e MLR-SPA podem ser melhorados ainda mais, optou-se por incluir no conjunto de calibração amostras comerciais. Fazendo isso, os novos modelos construídos deverão incluir na modelagem as variações inerentes às diferentes etapas do processo de fabricação dos comprimidos sintéticos e comerciais, as quais não foram incluídas na modelagem anterior, na qual utilizou-se apenas as amostras sintéticas no conjunto de calibração.

3.2.3 Modelagem PLS-1 e MLR-SPA usando um conjunto de calibração com amostras sintéticas e comerciais

Para escolher quais amostras comerciais deveriam ser inseridas no conjunto de calibração foi utilizado o algoritmo SPXY, que na escolha leva em conta tanto as variações (a estatística) da matriz espectral (**X**) quanto da matriz de concentração (**y**). Optou-se por escolher, usando o algoritmo SPXY, quantidades crescentes de amostras a serem adicionadas de 3 em 3. A cada adição de amostras comerciais ao conjunto de calibração, um novo modelo PLS-1 era construído, assim como o seu gráfico do resíduo espectral de Student versus a influência das amostras (**Figura 15**).

Verifica-se na **Figura 15** que à medida que são adicionadas amostras comerciais no conjunto de calibração de 3 até 12, a média da influência (linha tracejada azul) diminui mais significativamente. Pode-se notar que neste caso as amostras selecionadas pelo algoritmo SPXY e inseridas no conjunto de calibração possuem valores de influência menores que a influência média do conjunto, contribuindo assim na diminuição gradativa da influência média. Já a partir da **Figura 15(e)**,

que corresponde à inserção de 15 amostras comerciais no conjunto de calibração, nota-se uma tendência à estabilização da média da influência, indicando que a inserção de novas amostras não mais a afetará significativamente a influência. Deste modo, acreditamos que a inclusão de 15 amostras comerciais é suficiente para a construção de bons modelos de calibração, considerando que as variações inerentes aos processos de fabricação dos comprimidos sintéticos e comerciais foram incluídas na modelagem. Posteriormente, outras ferramentas estatísticas de diagnóstico poderão comprovar se esta hipótese está correta.

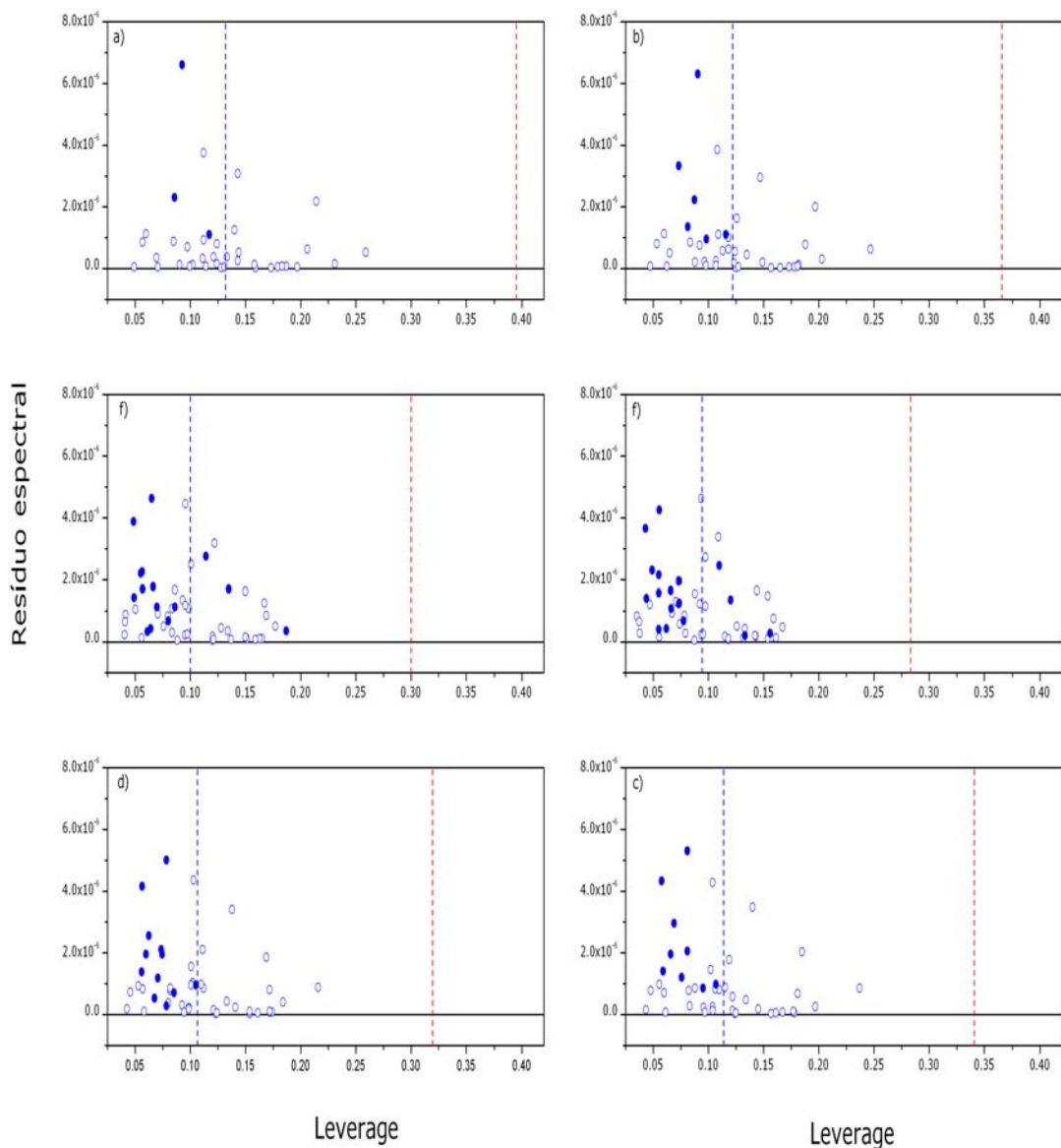


Figura 16: Resíduo de concentração de Student versus a influência das amostras após a inserção de (a) 3, (b) 6; (c) 9, (d) 12, (e) 15, (f) 18 amostras comerciais no conjunto de calibração. A linha vertical tracejada azul representa a média da influência das amostras e a vermelha três vezes a média da influência. A linha horizontal tracejada vermelha indica o valor limite para o resíduo de concentração de Student.

3.2.4 Modelagem PLS-1 e MLR-SPA com inserção de amostras comerciais no conjunto de calibração

Após determinar quais e quantas amostras industriais são suficientes para inserir as variações físicas inerentes aos processos de fabricação dos comprimidos, partiu-se para a construção dos modelos

PLS-1 e MLR-SPA usando o conjunto de calibração contendo 35 comprimidos sintéticos e 15 comprimidos comerciais.

A figura 17 mostra o gráfico das concentrações medidas versus as concentrações previstas na construção do modelo de calibração usando o método de validação cruzada para o modelo construído com as amostras sintéticas e com a inserção de 15 amostras comerciais. O valor de RMSEC obtido para este modelo foi de 1,2059 e a correlação 0,9437, para a validação o RMSECV obtido foi 1.3294.

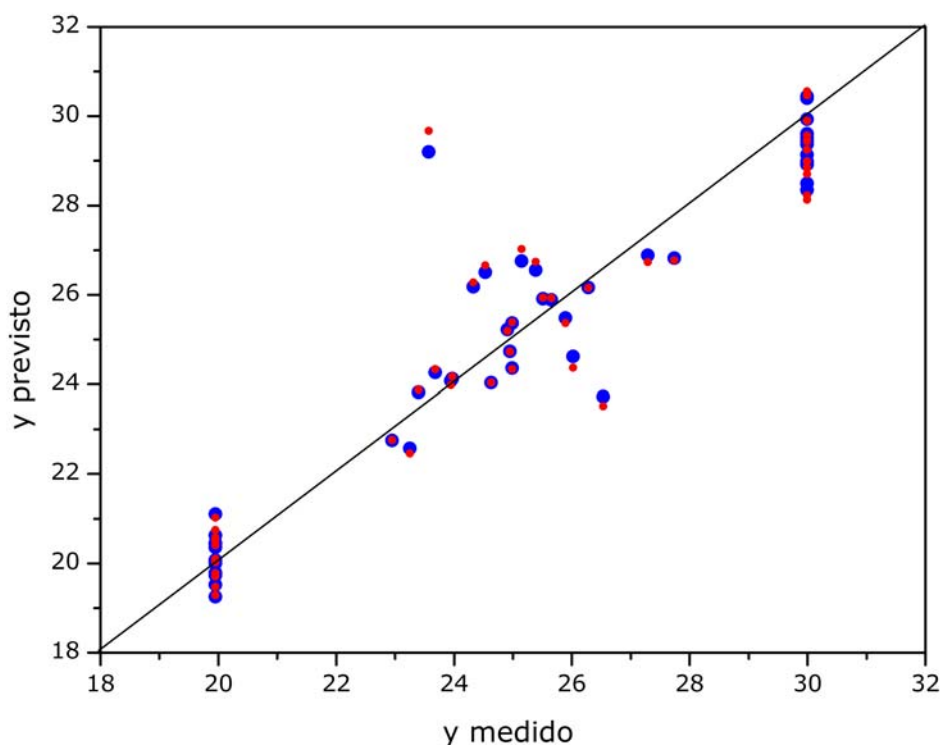


Figura 17: Gráfico das concentrações medidas versus as concentrações previstas na construção do modelo de calibração usando o método de validação cruzada, com a inserção de 15 amostras comerciais no conjunto de calibração.

Para efeito de comparação e diagnóstico, também foram construídos modelos usando conjuntos de calibração composto por 35 amostras sintéticas e com a inserção de 3, 6, 9, 12 e 18 amostras comerciais.

A **Tabela 4** mostra os valores de RMSEP bem como o número de variáveis selecionadas para PLS-1 (VL) e para o MLR-SPA (k).

Tabela 4: Valores de RMSEP para os modelos PLS-1 e MLR-SPA.

Nº de amostras comerciais inseridas no conjunto de calibração	RMSEP		Nº de variáveis selecionadas	
	PLS-1	MLR-SPA	VL* (PLS-1)	K** (MLR-SPA)
0	2,3	2,0	4	14
3	2,2	2,0	4	8
6	1,7	1,7	4	12
9	1,8	1,7	4	16
12	1,7	1,9	4	13
15	1,5	1,4	4	14
18	1,4	1,9	10	18

*Variáveis latentes

**Número de onda

De acordo com os resultados apresentados na **Tabela 4**, os melhores modelos, tanto para o PLS-1 quanto para o MLR-SPA, foram aqueles construídos com o conjunto de calibração contendo 15 amostras comerciais. O modelo PLS-1 apresentou um RMSEP de 1,5 e utilizou 4 variáveis latentes. Já o RMSEP do modelo MLR-SPA foi 1,4 usando 14 números de onda na construção do modelo.

Na **Figura 18** encontra-se indicadas as variáveis selecionadas pelo algoritmo SPA durante a construção do modelo MLR-SPA.

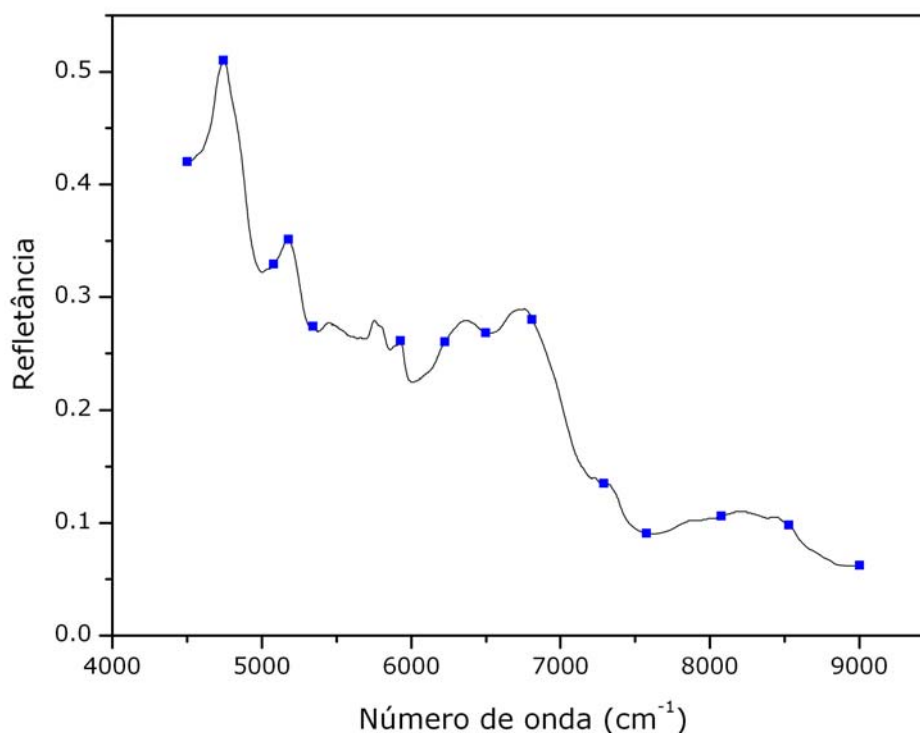


Figura 18: Indicação da posição em um espectro representativo das variáveis selecionadas durante a construção do modelo MLR-SPA.

3.3 Validação analítica dos modelos PLS-1 e MLR-SPA de acordo com a norma ASTM E 1655 05

Um dos requisitos obrigatórios para que um modelo de calibração multivariada seja considerado como construído de acordo com a norma ASTM E 1655 05 é que ele não contenha tendências estatísticas (*bias*) significativas. Para verificar a presença de tendências estatisticamente significativas nos modelos, seguiu-se o procedimento descrito no item 12 da seção 1.5.1.4 desta tese. Os modelos que possuem $t_{\text{calculado}} > t_{\text{crítico}}$ possuem tendências estatísticas significativas, ou seja, existe 95% de probabilidade de que os valores estimados por estes modelos não tenham a mesma média que a estimada pelo método de referência. Deste modo, a estimativa das concentrações feita por estes modelos não seria confiável.

A **Tabela 5** mostra os valores de $t_{\text{crítico}}$ para as 31 amostras ($n_{\text{cal}}-1$) do conjunto de previsão e de $t_{\text{calculado}}$ para os modelos construídos usando o conjunto de calibração com 35 amostras sintéticas e com a inserção de 0, 3, 6, 9, 12, 15 e 18 amostras comerciais. Verifica-se na **Tabela 4** que os modelos PLS-1 e MLR-SPA construídos usando o conjunto de calibração contendo 6 à 18 amostras para o PLS-1 e 15 e 18 para o MLR-SPA amostras comerciais não possuem *bias*. Deste modo, pode-se afirmar que os valores estimados pelos modelos PLS-1 e MLR-SPA construído com o conjunto de calibração contendo 35 amostras sintéticas e 15 amostras comerciais têm a mesma média que a média do método de referência e que as concentrações estimadas por estes modelos podem ser consideradas confiáveis.

Tabela 5: Valores de $t_{\text{calculado}}$ e $t_{\text{crítico}}$ para os modelos PLS-1 e MLR-SPA. ($t_{\text{crítico}}=1,6955$)

Nº de amostras comerciais inseridas no conjunto de calibração	$t_{\text{calculado}}$	
	PLS-1	MLR-SPA
0	8,2	3,6
3	7,5	3,1
6	$4,7 \times 10^{-3}$	1,7
9	1,5	3,0
12	1,4	4,2
15	1,01	1,3
18	0,8	0,6

Outro requisito para a construção de um modelo de calibração multivariada segundo a norma ASTM E 1655 05 é que pelo menos 95% das amostras do conjunto de previsão estejam dentro dos limites definidos no item 11 da seção 1.5.1.4 desta tese. Este requisito foi testado apenas para o modelo PLS-1, pois ainda não se tem disponível esta ferramenta de diagnóstico para o modelo MLR-SPA. Para o PLS-1 este requisito é testado usando o gráfico da concentração estimada por este modelo para as 32 amostras do conjunto de previsão (**Figura 19**). Como podemos perceber na **Figura 19**, apenas 2 amostras estão fora dos limites de confiança, o que corresponde a menos que 5% das amostras do conjunto de previsão.

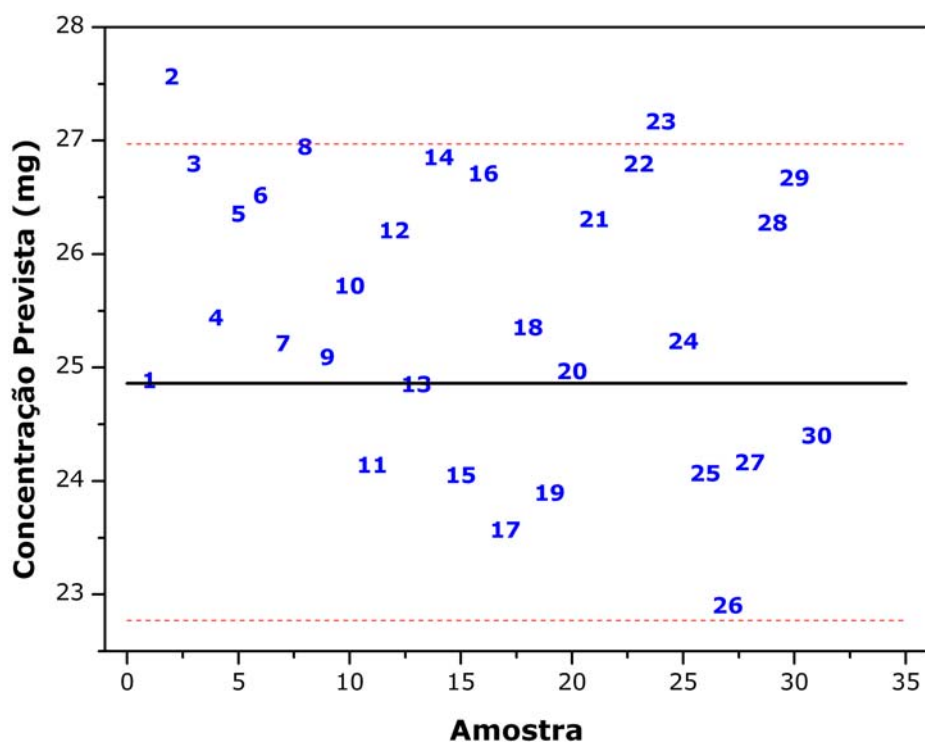


Figura 19: Concentração estimada, por comprimido, das 32 amostras do conjunto de previsão usando o modelo PLS-1 construído com o conjunto de calibração contendo 35 amostras sintéticas e 15 amostras comerciais. A linha horizontal azul representa a média das concentrações obtidas pelo método de referência, e as duas linhas em vermelho representam os limites de confiança estimados segundo a norma ASTM (vide item 11 da seção 1.5.1.4 desta tese).

A precisão dos valores estimados por um modelo multivariado usando espectrometria NIR é calculada usando medidas espectrais

repetidas para várias amostras diferentes. Como a precisão é baseada no desvio padrão das concentrações estimadas pelo modelo, é necessário saber se todas as variâncias para as medidas repetidas pertencem à mesma população. Para isso compara-se um valor de qui-quadrado-calculado (vide item 13 do sessão 1.5.1.4) com o valor qui-quadrado-tabelado para $(am_{teste}-1)$ graus de liberdade. No caso dos modelos aqui construídos, qui-quadrado-calculado foi 14,25 para o PLS-1 e 16,02 para MLR-SPA e o qui-quadrado-tabelado foi 58,12. Como $\chi^2_{calc} < \chi^2_{tab}$ para os 2 métodos, as variâncias das medidas replicadas pertencem à mesma população e a média da variância pode ser usada na avaliação da repetibilidade. Deste modo, a precisão dos modelos PLS-1 é 1,35 e do modelo MLR-SPA é 1,28.

A **Tabela 6** apresenta o resumo dos parâmetros de avaliação dos modelos PLS-1 e MLR-SPA construídos com o conjunto de calibração contendo 35 amostras sintéticas e 15 amostras comerciais. Foram também incluídas nesta tabela algumas figuras de mérito analítico calculadas pelo programa MVC1^[74] para o modelo PLS-1. As figuras de mérito não puderam ser calculadas para o MLR-SPA, porque ainda não se tem disponível uma maneira de calculá-las usando este modelo.

Tabela 6: Parâmetros para avaliação dos modelos PLS-1 e MLR-SPA.

	PLS-1	MLR-SPA
RMSEC	1,2	1,3
RMSEP	1,5	1,4
Precisão	1,3	1,2
$t_{calculado}$	1,0	1,3
Sensibilidade	0,0017	
Seletividade	0,0815	
Limite de detecção	0,2512	
Resíduo médio	$9,6650 \times 10^{-5}$	

Os valores de RMSEC e RMSEP são indicadores da precisão dos métodos desenvolvidos apresentam um nível aceitável de dispersão e uma boa concordância entre si. Os valores de precisão também apresentaram um nível de dispersão aceitável.

Segundo o valor do teste-t as possíveis influências de tendências significativas no modelo e nos parâmetros por ele estimados podem ser desprezadas, uma vez que o valor de $t_{calculado}$ é menor que o valor tabelado da distribuição de t-Student.

O modelo apresentou um baixo valor numérico para a sensibilidade que pode ser atribuído em parte ao pré-tratamento utilizado.

3.3.1 Questionário da ASTM E 1655 05 para avaliação dos modelos PLS-1 e MLR-SPA construídos

Segundo a norma ASTM E 1655 05, se todas as perguntas do questionário abaixo forem respondidas afirmativamente (vide seção 1.5.1.4), os modelos PLS-1 e MLR-SPA podem ser considerados validados.

QUESTIONÁRIO	PLS-1	MLR-SPA	COMENTÁRIOS
1. Foi utilizada MLR, PCR ou PLS-1?	SIM	SIM	A técnica usada para a construção dos modelos foi PLS-1 e MLR.
2. O método é capaz de detectar amostras anômalas usando a estatística de influência, h ?	SIM	NÃO	A ainda não se tem disponível uma maneira de calculá-las para o modelo MLR-SPA.
3. A metodologia é capaz de detectar amostras anômalas via estatística baseada em resíduo espectral?	SIM	NÃO	
4. $n_c > 6k$ (para modelos não centrados na média) ou $n_c > 6(k+1)$ (para modelos centrados na média)?	SIM	NÃO	' n_c ' foi 50, ' k ' foi 4 para o PLS-1 e 14 para o MLR-SPA e os dados foram centrados na média. Assim, o nº de amostras deveriam ser: 30 para PLS-1 e 90 para MLR-SPA;
5. O nº de amostras de calibração usadas foi pelo menos 24 ?	SIM	SIM	o nº de amostras de calibração é 50, portanto maior que 24.
6. Um conjunto de amostras separado foi usado para testar a validação?	SIM	SIM	Um conjunto de amostras distinto, o conjunto de previsão, foi usado.
7. Amostras anômalas foram excluídas do conjunto teste com base na influência ou resíduo espectral?	SIM	SIM	Não foram detectadas amostras anômalas no conjunto teste.
8. O nº de amostras teste ' n_t ' foi $> 4k$ (para dados não centrados na média) ou $>$ que $4(k+1)$ (para modelos centrados na média)?	SIM	Não	' n_t ' foi 32, ' k ' foi 4 para o PLS-1 e 14 para o MLR-SPA e os dados foram centrados na média. Assim, o nº de amostras deveriam ser: 20 para PLS-1 e 60 para MLR-SPA.
9. O n_t é pelo menos 20?	SIM	SIM	n_t é 32, portanto, maior que 20.
10. As amostras de teste cobrem 95% da faixa de calibração?	SIM	SIM	Todas amostras teste estão dentro da faixa de concentração da calibração.
11. 95% dos resultados das previsões estão dentro do valor de referência	SIM	NÃO	No PLS-1, apenas 2 amostras ficaram fora dos limites de

$\pm tSEC\sqrt{1+h}$?			confiança e ainda não é possível fazer este teste para o MLR-SPA.
12. Os resultados das previsões têm <i>bias</i> insignificantes?	SIM	SIM	$t_{calculado} < t_{crítico}$ para os dois modelos.
13. Foi determinada a precisão do modelo usando $am_{test} \geq k \geq 3$ e $r \geq 6$?	SIM	SIM	A precisão do modelo foi determinada usando 7 amostras para o PLS-1 e 14 para o MLR-SPA e 6 replicatas para cada amostra.
14. Os pré ou pós-tratamentos foram feitos automaticamente?	SIM	SIM	Elas fazem parte do programa usado para construir os modelos.

A partir do questionário de avaliação elaborado pela ASTM, podemos comprovar que a metodologia NIRR-PLS-1 foi e validada de acordo com a norma ASTM E 1655 05 e deste modo pode ser usada como método oficial para controle de qualidade do teor de captopril nos comprimidos fabricados pelo LAFEPE após sua certificação pelos órgãos competentes. No entanto, a metodologia NIRR-MLR-SPA, apesar de ter apresentado um bom desempenho na estimativa das concentrações dos comprimidos do conjunto de previsão, ainda não possui a ferramenta de detecção de amostras anômalas com base na influência e no resíduo espectral de Student e deste modo não pode ser considerado como validado de acordo com a norma ASTM E 1655 05. No entanto, estas ferramentas como também o cálculo das figuras de mérito, estão em desenvolvimento e serão implementadas em um futuro muito próximo.



Capítulo 4

Conclusões & Propostas Futuras

4.1 Conclusões

A espectrometria NIR aliada às técnicas de calibração multivariada PLS-1 e MLR-SPA possibilitou o desenvolvimento de metodologias de análise de comprimidos contendo captopril produzidos comercialmente pelo LAFEPE, que por sua rapidez e qualidade dos resultados, constituem uma alternativa vantajosa aos métodos tradicionais de análise.

Para que os modelos construídos pudessem proporcionar resultados satisfatórios foi necessário que os espectros fossem pré-processados usando a suavização pelo método Savitzky-Golay com uma janela de 9 pontos. Além disso, foi necessária também a preparação de comprimidos sintéticos com teor de captopril em uma faixa de concentração mais ampliada que aquela dos comprimidos produzidos pelo LAFEPE. Os comprimidos sintéticos foram elaborados utilizando um planejamento de misturas D-otimizado. Nesta elaboração, foi levada em conta a formulação (princípio ativo e excipientes) usada pelo LAFEPE para a produção dos comprimidos comerciais com valor nominal de 25 mg de captopril e uma massa total de cerca de 150 mg. Também foram consideradas as recomendações (item 1.5.1.2) da norma da ASTM E1655 05 quanto à preparação do conjunto de calibração usado na construção de modelos de calibração multivariada empregando a espectrometria NIR. A preparação de amostras sintéticas com faixa de concentração ampliada se faz necessária para evitar erros de previsão por extrapolação, causados pelo uso de modelos subjustado que, em geral, são produzidos quando o conjunto de calibração contém apenas amostras da linha de produção. Por outro lado, como as amostras preparadas em laboratório não possuem as variações físicas inerentes às amostras industriais, algumas amostras industriais também necessitam ser incluídas no conjunto de calibração para a construção dos modelos PLS-1 e MLR-SPA. Essas amostras foram

escolhidas a partir de um estudo usando a SPXY e gráficos de influência para avaliar, respectivamente, quais e quantas amostras seriam necessárias para inserir as variações do processo industriais no conjunto de calibração. Os modelos que apresentaram o melhor desempenho na estimativa das concentrações das amostras teste foi aquele construído com um conjunto de calibração contendo 35 amostras sintéticas e 15 amostras comerciais.

A metodologia NIR-PLS-1 foi desenvolvida e validada de acordo com a norma ASTM E 1655 05 e apresentou um bom desempenho para a estimativa do teor de captopril nas amostras de predição, podendo ser implementada como metodologia para o controle de qualidade do teor de captopril em linha de produção, após sua certificação final pelos órgãos reguladores. A metodologia NIR-MLR-SPA não pode ser considerada como validada de acordo com a norma ASTM E 1655 05, pois não possui ferramentas de detecção de amostras anômalas. No entanto, o cálculo das figuras de mérito e a validação completa do modelo MLR-SPA, segundo a norma ASTM E 1655 05, ainda estão em fase de desenvolvimento.

O algoritmo SPA escolheu 14 variáveis espectrais que podem ser usadas para a construção de um fotômetro NIR de baixo custo a base de LEDs para o monitoramento *in situ* e *on line* do teor de captopril nos comprimidos produzidos.

4.2 Propostas Futuras

Como proposta de continuidade desse trabalho pretende-se:

- Implementar o cálculo de figuras de mérito e a identificação de amostras anômalas no algoritmo MLR-SPA para que os modelos desenvolvidos utilizando este algoritmo possam ser considerados como validados segundo a norma ASTM E 1655 05.

- Estudar e implementar as recomendações para validação de métodos analíticos das normas mais específicas para a indústria farmacêutica como a USP/ICH, entre outras.
- Utilizar a metodologia NIR para desenvolver modelos de calibração multivariada para estimar outros parâmetros de qualidade, por exemplo, dureza, desintegração, friabilidade, etc.
- Construir um fotômetro dedicado para a determinação do teor de captopril utilizando as variáveis selecionadas pelo algoritmo MLR-SPA.
- Aplicar a metodologia desenvolvida para outros comprimidos produzidos pelo LAFEPE.



Referências Bibliográficas

- [1] GILMAN, A. G.; HARDMAN, J. G.; LIMBIRD, L. E.; As Bases Farmacológicas da Terapêutica, Goodman & Gilman 10ª edição; McGraw-Hill Interamericana do Brasil, **1996**.
- [2] <http://w3.datasus.gov.br/hiperdia/captopril.php> (**acesso em 03/03/2008**).
- [3] The Pharmaceutical Codex, 1994; European Pharmacopeia, **1997**.
- [4] Resolução ANVISA - RDC nº 210, de **04 de agosto de 2003**.
- [5] United States Pharmacopoeia 25, NF 20, Rockville, p.1147, **2002**.
- [6] British Pharmacopoeia **1998**, CD-ROM, Version 2.0.
- [7] US Pharmacopoeia **2000**, CD-ROM, Version 5.1.
- [8] Real Farmacopeia Española **1997**.
- [9] PASQUINI, C.; Near infrared spectroscopy: Fundamentals, practical aspects and analytical applications; *Journal of the Brazilian Chemical Society*, **14:198, 2003**.
- [10] SKOOG, D.A.; HOLLER, F.J.; NIEMAN, T.A; Princípios de análise Instrumental. 5ª ed., Bookman, Porto Alegre, **2002**.
- [11] MILLER, C.E. In Near-Infrared Technology ; Williams, P.; Norris, K. eds., 2nd ed. , American Association of Cereal Chemistry, Inc.: St. Paul, Minnesota, USA, p. 19, **2001**.
- [12] KUBELKA, P.; MUNK, F.Z.; Tech. Physik., **12:593, 1931**
- [13] BIRTH, G. S.; Agricul. Mark., *appud*, **2:2, 1957**.
- [14] BIRTH, G. S.; NORRIS, K. H., *Food Technol., appud*, **1958**.

- [15] KAROUI, R.; de BAERDEMAEKER, J.; A review of the analytical methods coupled with chemometric tools for the determination of the quality and identity of dairy products, *Food Chemistry*, **102:621, 2007.**
- [16] ESTEBAN-DIEZ, I.; GONZALEZ-SAIZ, J.M.; SAENZ-GONZALEZ, C.; PIZARRO, C.; Coffee varietal differentiation based on near infrared spectroscopy, *Talanta*, **71:221, 2007.**
- [17] BLANCO, M.; VILLAESCUSA, V.; Use of NIR spectroscopy in the production of modified industrial resins, *Talanta*, **71:1333, 2007.**
- [18] TSUCHIKAWA, S.; A review of recent near infrared research for wood and paper, *Applied Spectroscopy Reviews*, **42:43, 2007.**
- [19] LANGERON, Y.; DOUSSOT, M.; HEWSON, D. J.; DUCHENE, J.; Classifying NIR spectra of textile products with kernel methods. *Engineering Applications of Artificial Intelligence*, **20:415, 2007.**
- [20] BLANCO, M.; ALCALA, M.; Simultaneous quantitation of five active principles in a pharmaceutical preparation: Development and validation of a near infrared spectroscopic method. *European Journal of Pharmaceutical Sciences*, **27:280, 2006.**
- [21] QU, H.B.; YANG, H.L.; CHENG, Y.Y.; Fast and nondestructive discrimination of donkeyhide glue by near-infrared spectroscopy, *Spectroscopy and spectral analysis*, **26:60, 2006.**
- [22] PRA, L. D.; CHAMPAGNE, S. D. A comparative study of the electronic structure and spectroelectrochemical properties of electrosynthesized polyaniline films and nanotubes. *Thin Solid Films*, **479:321, 2005.**

- [23] SZYDLOWSKI-ZANIER, N.; BERGER, M.; WAHL, F.; GUILLAUME, D. Performance of a near infrared spectrometer equipped with an autosampling accessory. *Journal of near infrared spectroscopy*, **11:83, 2003**.
- [24] COCCHI, M.; DURANTE, C.; FOCA, G.; MARCHETTI, A.; TASSI, T.; ULRICI, U. Durum wheat adulteration detection by NIR spectroscopy multivariate calibration. *Talanta*, **68:1505, 2006**.
- [25] REICH, G.; Near-infrared spectroscopy and imaging: Basic principles and pharmaceutical applications, *Advanced Drug Delivery Reviews*, **57:1109, 2005**.
- [26] AZZOUZ, T.; TAULER, R.; Application of multivariate curve resolution alternating least squares (MCR-ALS) to the quantitative analysis of pharmaceutical and agricultural samples, *Talanta*, **74:1201, 2008**.
- [27] CIURCZAK, E.W.; Uses of near-infrared spectroscopy in pharmaceutical analysis, *Applied Spectroscopy Reviews*, **23:147, 1987**.
- [28] SCAFI, S. H. F., PASQUINI, C., Identification of counterfeit drugs using near-infrared spectroscopy, *Analyst*, **126:2218, 2001**.
- [29] BLANCO, M., EUSTAQUIO, A., GONZALEZ, J.M., SERRANO, D., Identification and quantitation assays for intact tablets of two related pharmaceutical preparations by reflectance near-infrared spectroscopy: validation of the procedure. *Journal of Pharmaceutical and Biomedical Analysis*, **22:139, 2000**.
- [30] BROAD, N. W., JEE, R. D., MOFFAT, A. C., SMITH, M. R., Application of transmission near-infrared spectroscopy to

uniformity of content testing of intact steroid tablets. *Analyst*, **126:2207, 2001**.

- [31] BLANCO, M.; ALCALA, M.; GONZALEZ, J. M.; TORRASZ, E.; A process analytical technology approach based on near infrared spectroscopy: Tablet hardness, content uniformity, and dissolution test measurements of intact tablets, *Journal of Pharmaceutical Sciences*, **95(10):2137, 2006**.
- [32] OTSUKA, M.; Chemoinformetrical evaluation of granule and tablet properties of pharmaceutical preparations by near-infrared spectroscopy, *Chemometrics and Intelligent Laboratory Systems*, **82:109, 2006**.
- [33] BLANCO, M.; ALCALA, M.; Content uniformity and tablet hardness testing of intact pharmaceutical tablets by near infrared spectroscopy - A contribution to process analytical technologies, *Analytica Chimica Acta*, **557:353, 2006**.
- [34] OTTO, M., Chemometrics – Statistics and Computer Application in Analytical Chemistry, Weinheim, New York : Wiley-VCH, **1999**.
- [35] BLANCO, M.; ALCALA, M.; GONZALEZ, J. M.; TORRASZ, E.; A process analytical technology approach based on near infrared spectroscopy: Tablet hardness, content uniformity, and dissolution test measurements of intact tablets, *Journal of Pharmaceutical Sciences*, **95:2137, 2006**.
- [36] BLANCO, M.; COELLO, J.; EUSTAQUIO, A.; ITURRIAGA, H.; MASPOCH, S.; Analytical control of pharmaceutical production steps by near infrared reflectance spectroscopy, *Analytica Chimica Acta*, **392:237, 1999**.

- [37] FREITAS, M. P.; SABADIN, A.; SILVA, L. M.; GIANNOTTI, F. M.; COUTO, D. A.; TONHI, E.; MEDEIROS, R. S.; COCO, G. L.; RUSSO, V. F. T.; MARTINS, J. A., Prediction of drug dissolution profiles from tablets using NIR diffuse reflectance spectroscopy: A rapid and nondestructive method, *Journal of Pharmaceutical and Biomedical Analysis*, **39:17, 2005**.
- [38] BARROS NETO, B.; SCARMINIO, I.S.; BRUNS, R.E., Como fazer experimentos: pesquisa e desenvolvimento na ciência e na indústria. 3ª edição, *Editores da UNICAMP*, Campinas-SP, **2007**.
- [39] MONTGOMERY, D. C.; Design and Analysis of Experiments; 5th ed.; *Wiley*, New York, **2001**.
- [40] BRERETON, R. G.; Introduction to multivariate calibration in analytical chemistry, *Analyst*. **125:2125, 2000**.
- [41] DRAPER, N. R.; Smith, H. Applied Regression Analysis, 3rd ed.; *Wiley*, New York, **1998**.
- [42] WOLD, S.; Personal memories of the early PLS development; *Chemometrics and Intelligent Laboratory Systems*, 58:83, **2001**.
- [43] BEEB, K.R., KOWALSKI, B.R., An introduction to multivariate calibration analysis, *Analytical Chemistry*, **59:A1007, 1987**.
- [44] GELADI, P., KOWALSKI, B.R., Partial least square regression: a tutorial, *Analytica Chimica Acta*, **185:1, 1986**.
- [45] VANDEGINSTE, B.G.M., MASSART, D.L., BUYDENS, L.M.C., JING, S., LEWI, P.J., SMEYERS-VERBEKE, J., Handbook of Chemometrics and Qualimetrics: Part B. Amsterdam: Elsevier, **1998**.

- [46] MASON, R.L; GUNST, R.F; WEBSTER, J.T; Regression-analysis and problems of multicollinearity. *Communications in Statistics*. **4:277, 1975.**
- [47] ARAUJO, M.C.U., SALDANHA, T.C.B., GALVAO, R.K.H., YONEYAMA, T., CHAME, H.C., VISANI, V., The successive projections algorithm for variable selection in spectroscopic multicomponent analysis, *Chemometrics and Intelligent Laboratory Systems*, **57:65, 2001.**
- [48] GALVÃO, R.K.H., PIMENTEL, M.F., ARAUJO, M.C.U., YONEYAMA, T., VISANI, V., Aspects of the successive projections algorithm for variable selection in multivariate calibration applied to plasma emission spectrometry, *Anal. Chim. Acta*, **443:107, 2001.**
- [49] GALVÃO, R.K.H; ARAUJO, M.C.U; JOSE, G.E, et al.; A method for calibration and validation subset partitioning, *Talanta*, 67:736, **2005.**
- [50] GALVAO, R. K.H.; ARAUJO, M. C. U.; SILVA, E.C.; et al; Cross-validation for the selection of spectral variables using the successive projections algorithm. *Journal of the Brazilian Chemical Society*, **18:580, 2007.**
- [51] SASAKI, K.; KAWATA, S.; MINAMI, S.; Optimal Wavelength Selection for Quantitative Analysis; *Appl. Spectroscopy*, **40:185, 1986.**
- [52] KALIVAS, J. H.; ROBERTS, N.; SUTTER, J. M.; Global Optimization by Simulated Annealing with Wavelength Selection for Ultraviolet-Visible Spectrophotometry; *Anal. Chem.*; **61:2024, 1989.**

- [53] ALSBERG, B. K.; WOODWARD, A. M.; WINSON, M. K.; ROWLAND, J. J.; KELL, D. B.; Variable selection in wavelet regression models; *Anal. Chim. Acta*, **368:29, 1998**.
- [54] JOUAN-RIMBAUD, D.; WALCZACK, B.; POPPI, R. J.; DENOORD, O. E.; MASSART, D. L.; Application of Wavelet Transform To Extract the Relevant Component from Spectral Data for Multivariate Calibration, *Anal. Chem.*; **69:4317, 1997**.
- [55] LUCASIU, C. B.; BECKERS, M. L. M.; KATEMAN, G.; Genetic algorithms in wavelength selection: a comparative study, *Anal. Chim. Acta*, **286:135, 1994**.
- [56] LEARDI, R., Application of a genetic algorithm to feature selection under full validation conditions and to outlier detection, *J. Chemo.*, **8:65, 1994**.
- [57] JOUANRIMBAUD, D.; MASSART, D. L.; LEARDI, R.; NOORD, O. E., Genetic algorithms as a tool for wavelength selection in multivariate calibration, *Anal. Chem.*, **67:4295, 1995**.
- [58] LEARDI, R.; BOGGIA, R.; TERRILE, M.; Genetic algorithms as a strategy for feature selection, *J. Chemom.*, **6:267, 1992**.
- [59] DASZYKOWSKI, M.; WALCZAK, B.; MASSART, D.L. Representative subset selection. *Analytica Chimica Acta*, **468:91, 2002**.
- [60] TOMINAGA, Y. Representative subset selection using genetic algorithms. *Chemometrics. Intelligent Laboratory System*, **43:157, 1998**.
- [61] SALES, F.; RIUS, A.; CALLAO, M.P.; RIUS, F.X.; Standardization of a multivariate calibration model applied to the determination of chromium in tanning sewage, *Talanta*, **52:329, 2000**.

- [62] WU, W.; WALCZAK, B.; MASSART, D.L.; HEUERDING, S.; ERNI, F.; LAST, I.R.; PREBBLE, K.A.; Artificial neural networks in classification of NIR spectral data: Design of the training set. *Chemometrics and Intelligent Laboratory System*, **33:35, 1996**.
- [63] HAMZACEBI, C.; KUTAY, F.; A heuristic approach for finding the global minimum: Adaptive random search technique. *Applied Mathematics and Computation*, **173:1323, 2006**.
- [64] KANDUC, K.R.; ZUPAN, J.; MAJCEN, N.; Separation of data on the training and test set for modelling: a case study for modelling of five colour properties of a white pigment. *Chemometric Intelligent Laboratory System*. **65:221, 2003**.
- [65] KENNARD, R.W.; STONE, L.A.; Computer aided design of experiments, *Technometrics*, **11:137, 1969**.
- [66] CAPITAN-VALLVEY, L.F.; NAVAS, N.; DEL OLMO, M.; CONSONNI, V.; TODESCHINI, R.; Resolution of mixtures of three nonsteroidal anti-inflammatory drugs by fluorescence using partial least squares multivariate calibration with previous wavelength selection by Kohonen artificial neural networks, *Talanta*, **52:1069, 2000**.
- [67] Annual Book of ASTM Standards. Standards practices for infrared, multivariate, quantitative analysis, E1655, vol 03.06. ASTM International, West Conshohocken, Pennsylvania, USA, **2000**.
- [68] Eurachem/Citac – Work Group. Quantifying Uncertainty in Analytical Measurement. 2. ed., **2000**.
- [69] DANZER, K; OTTO, M. Currie, L. Guidelines for calibration in analytical chemistry. Part 2. Multispecies calibration (IUPAC

Technical Report). *Internacional Union of Pure Applied Chemistry*, **76(6):1215, 2004**.

- [70] LORBER, A.; Error propagation and figures of merit for quantification by solving matrix equations, *Analytical Chemistry*, **58:1167, 1986**.
- [71] LORBER, A.; FABER, K.; KOWALSKI, B. R.; Net analyte signal calculation in multivariate calibration, *Analytical Chemistry*, **69:1620, 1997**.
- [72] OLIVIERE, A. C.; FABER, N. M.; FERRÉ, J.; BOQUÉ, R.; KALIVAS, J. H.; MARK, H.; Uncertainty estimation and figures of merit for multivariate calibration; *Pure Appl. Chem.*, **78:663, 2006**.
- [73] BEEBE, K. R.; RANDY, J. P.; SEASHOLTZ, M. B., *Chemometrics: A practical guide*, Wiley-Interscience, New York, **1998**.
- [74] OLIVIERI, A. C.; Goicoechea, H. C.; Iñón, F. A.; MVC1: an integrated MatLab toolbox for first-order multivariate calibration, *Chemometrics and Intelligent Laboratory Systems*, **73:189, 2004**.
- [75] http://www.eosystems.com/pdf/IGA_Series_spectral_responsivity.pdf (**acesso em 26/03/2008**).