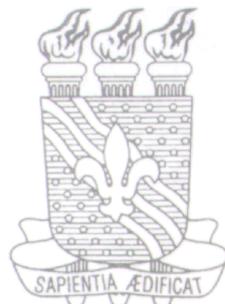


**UNIVERSIDADE FEDERAL DA PARAÍBA**  
**CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA**  
**DEPARTAMENTO DE QUÍMICA**



**DISSERTAÇÃO DE MESTRADO**

*“Desenvolvimento de uma Metodologia para o Controle de Qualidade  
de Medicamento Injetável, sem Violação da Ampola, usando  
Espectroscopia NIR e Técnicas Quimiométricas”*

Por:

*Fátima Aparecida Castriani Sanches*

**João Pessoa/PB - Brasil**

**2009**



UNIVERSIDADE FEDERAL DA PARAÍBA  
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA  
DEPARTAMENTO DE QUÍMICA

*“Desenvolvimento de uma Metodologia para o Controle de Qualidade  
de Medicamento Injetável, sem Violação da Ampola, usando  
Espectroscopia NIR e Técnicas Quimiométricas”*

*Fátima Aparecida Castriani Sanches*

Dissertação submetida ao programa de pós-graduação em química, da Universidade Federal da Paraíba, como requisito parcial à obtenção do título de Mestre em Química, área de concentração em “Química Analítica”.

**1º Orientador (a):** Prof. Dr. Mário César Ugulino de Araújo

**2º Orientador (a):** Prof<sup>a</sup>. Dra. Elaine Cristina Lima do Nascimento

Bolsista: 

**João Pessoa – Novembro/2009.**

S211d Sanches, Fátima Aparecida Castriani.

*Desenvolvimento de uma metodologia para o controle de qualidade de medicamento injetável, sem violação da ampola, usando espectroscopia NIR e técnicas quimiométricas / Fátima Aparecida Castriani Sanches.- João Pessoa, 2009.*

72f. : il.

Orientadores: Mário César Ugulino de Araújo, Elaine Cristina Lima do Nascimento

Dissertação (Mestrado) – UFPB/CCEN

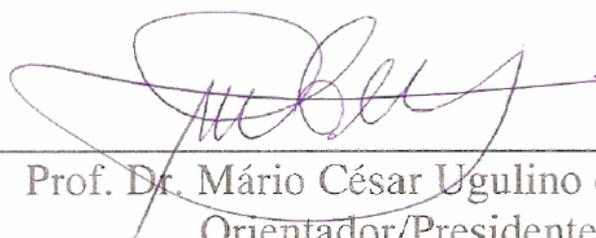
1. Química Analítica. 2. Medicamento injetável – controle de qualidade. 3. Espectroscopia. 4. NIR. 5. Quimiometria.

UFPB/BC

CDU: 543(043)

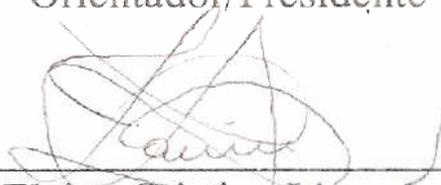
# Desenvolvimento de uma Metodologia para o Controle de Qualidade de Medicamento Injetável, sem Violação da Ampola, Usando Espectroscopia NIR e Técnicas Quimiométricas.

Aprovada pela banca examinadora:



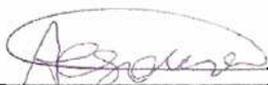
---

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



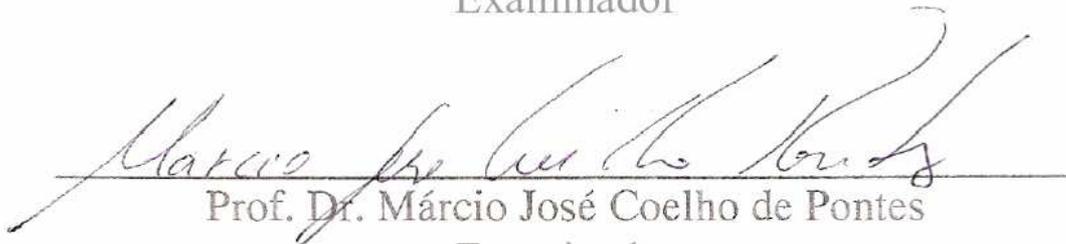
---

Profa. Dra. Elaine Cristina Lima do Nascimento  
2ª. Orientadora



---

Prof. Dr. Antonio Cícero de Souza  
Examinador



---

Prof. Dr. Márcio José Coelho de Pontes  
Examinador

*À Deus, pai misericordioso, que em sua  
infinita bondade, nunca me desamparasse.  
À minha querida família, principalmente,  
à meus amados filhos, como incentivo,  
com muito amor,  
dedico .*

# AGRADECIMENTOS

---

---

Ao meu bondoso e amado Pai, Senhor e Deus meu, por tudo.

Ao professor Mário Ugulino, por me acolher no LAQA; pelas orientações ministradas na vida acadêmica e mestrado; pelos conhecimentos transmitidos; pelos conselhos; pelo respeito e amizade cedidos.

À professora Regiane de Cássia, por ter me iniciado na pesquisa científica; por estar sempre por perto; pelo carinho, respeito, conhecimento e amizade cedidos.

À Elaine Cristina e à Claudete Fernandes, pela co-orientação em fases distintas de minha dissertação; pela ajuda prestada e conhecimentos transmitidos; pela amizade.

Ao *CNPq*, pela bolsa concedida.

Às indústrias farmacêuticas: *Isofarma* e *Farmace*.

Aos farmacêuticos dos hospitais da cidade de João Pessoa, em especial, à Adelaide e à Lorena.

Aos professores Edvan Cirino, Teresa Saldanha e Célio Pasquini, pelos conhecimentos transmitidos; pelas conversas enriquecedoras em momentos de dúvidas e de idéias eminentes.

Aos professores e amigos do Departamento de Farmácia, Departamento de Química e da Pós Graduação, pelos conhecimentos ministrados. E aos amigos de outros laboratórios de pesquisa, principalmente, à estimada amiga Ladjane Sodré.

Aos amigos do LAQA de ontem e de hoje, principalmente à Josenita e ao Willieme, por me aturarem e me aconselharem em diversos momentos.

Aos meus pais, em especial à minha querida mãe, uma verdadeira fonte de inspiração, por quem sinto muito orgulho, admiração e respeito.

À minha irmã Madalena, por tudo. Saiba que hoje eu cheguei até aqui, literalmente, graças a você.

À Nira e João, pelo carinho, amor e respeito que sempre nos dedicaram.

À minha sogra, pela ajuda prestada com as crianças e por me ajudar em momentos de precisão.

Ao meu esposo Valter Adriano, pela paciência e sabedoria transmitidas em momentos rudes. Pelo incentivo prestado em momentos de desânimo e tristeza. Pelo infindo amor a mim concedido.

Aos meus queridos filhos, pela compreensão de tantos momentos ausentes; pelos abraços e beijinhos carinhosos; por me darem essa alegria inigualável de ser sua mãe.

Carinhosamente, eu agradeço.

## **O sábio que sabia quase tudo.**

Certa vez um sábio, que atravessava um rio de barco, perguntou ao barqueiro: "Diga-me uma coisa, você conhece botânica?" O barqueiro olhou surpreso para o sábio e respondeu: "Não, senhor, não sei o que é isso". E o sábio continuou: "Você não sabe botânica, a ciência que estuda as plantas? Que pena! Você perdeu parte da sua vida". O barqueiro continuava remando rio a dentro, e, em seguida o sábio perguntou se ele conhecia um pouco de astronomia. O pobre barqueiro coçou a cabeça e disse: "Não senhor; não sei o que é Astronomia". "Astronomia é a ciência que estuda os astros, o espaço, as estrelas", explicou o grande sábio. "Que pena! Você perdeu parte de sua vida sem saber o que é Astronomia". E o sábio foi perguntando a respeito de cada ciência: física, química, teologia, filosofia. De nada o barqueiro sabia. E o sábio sempre terminava o seu refrão. "Que pena! Você perdeu parte de sua vida". De repente, o barco bateu contra uma pedra, partiu-se e começou a afundar; e o barqueiro perguntou ao sábio: "O senhor sabe nadar?" "Não, não sei" respondeu o sábio. "Que pena, o senhor perdeu toda a sua vida".

Existem muitas pessoas preparadas por cursos acadêmicos, conquistaram seus grandes títulos, falam diversos idiomas, são eruditos de nossa época, mas sem conhecer o Senhor Jesus, nada o salvará na hora da morte. De nada valem milhares de livros se o leitor não fizer uso da Bíblia, a carta de Deus para os homens.

Sábio é aquele que vive com Cristo neste mundo e no além.

*Emerson Alves*

---

---

ÍNDICE DE FIGURAS .....	iii
ÍNDICE DE TABELAS .....	vi
LISTA DE ABREVIATURAS E SIGLAS .....	vii
RESUMO .....	viii
ABSTRACT .....	ix
<b>CAPÍTULO I .....</b>	<b>.....</b>
<b>1. INTRODUÇÃO .....</b>	<b>2</b>
1.1 Caracterização do problema .....	2
1.2 Medicamentos injetáveis .....	4
1.3 Dipirona sódica .....	5
1.4 Metodologias de análise da dipirona sódica injetável .....	7
1.5 Espectrometria na região do infravermelho próximo (NIR) .....	8
1.6 Quimiometria .....	12
1.7 Métodos de reconhecimento de padrão .....	16
1.7.1 PCA .....	16
1.7.2 SIMCA .....	19
1.7.2.i Análise Screening .....	22
1.8 Calibração multivariada .....	22
1.8.1 PLS .....	25
1.8.2 MLR .....	27
1.8.2.1 ASA .....	28
1.8.2.2 GA .....	29
1.8.2.3 SPA .....	29
1.8.2.4 SW .....	30
1.9 Objetivo .....	31
1.9.1 Objetivos específicos .....	31
<b>CAPÍTULO II .....</b>	<b>.....</b>
<b>2. EXPERIMENTAL .....</b>	<b>33</b>
2.1 Metodologia e estratégia de ação .....	33
2.1.1 Obtenção das Amostras e preparação das soluções padrão .....	33
2.1.2 Instrumentação .....	34
2.1.3 Confecção dos suportes para os medicamentos .....	34

---

---

2.1.4 Estudos físicos das ampolas de vidro .....	35
2.1.5 Aquisição dos espectros .....	36
2.1.6 Programas estatísticos .....	37
<b>CAPÍTULO III .....</b>	
<b>3. RESULTADOS E DISCUSSÕES .....</b>	<b>39</b>
3.1 Seleção da região espectral de trabalho .....	39
3.2 Atribuição das bandas dos espectros NIR .....	41
3.3 Pré-Processamento dos dados espectrais.....	42
3.4 Aplicação de uma análise de componentes principais (PCA).....	42
3.5 Análise Screening .....	44
3.6 Resultado da aplicação dos modelos .....	48
3.6.1 Modelo PLS1 .....	49
3.6.2 Modelo MLR-SPA .....	50
3.6.3 Modelo MLR-ASA .....	52
3.6.4 Modelo MLR-SW .....	55
3.6.5 Modelo MLR-GA .....	57
<b>CAPÍTULO IV .....</b>	
<b>4. CONCLUSÕES.....</b>	<b>61</b>
4.1 Propostas futuras.....	62
<b>CAPÍTULO V .....</b>	
<b>5. REFERÊNCIAS BIBLIOGRÁFICAS.....</b>	<b>64</b>

---



---

<b>Figura 1</b>	Fórmula estrutural da dipirona sódica.....	<b>6</b>
<b>Figura 2</b>	Determinação da transmitância de uma amostra.....	<b>10</b>
<b>Figura 3</b>	Exemplo da conversão das variáveis em novos eixos no espaço multidimensional.....	<b>16</b>
<b>Figura 4</b>	Exemplo de modelos <i>SIMCA</i> .....	<b>19</b>
<b>Figura 5</b>	Predição com um modelo <i>SIMCA</i> .....	<b>20</b>
<b>Figura 6</b>	Representação de matrizes de dados (X e Y).....	<b>24</b>
<b>Figura 7</b>	Espectrômetro <i>Perkin Elmer FT-IR</i> modelo <i>GX</i> .....	<b>34</b>
<b>Figura 8</b>	Suporte das ampolas de medicamentos adaptado no compartimento de medidas do espectrômetro <i>NIR</i> .....	<b>35</b>
<b>Figura 9</b>	Estudo de absorção da radiação <i>NIR</i> pelas ampolas de vidro.....	<b>35</b>
<b>Figura 10</b>	Sistema para registro dos espectros <i>NIR</i> dos medicamentos injetáveis sem violação da ampola. <b>A</b> : Espectrômetro <i>NIR</i> ; <b>B</b> : Suporte das ampolas acoplado ao suporte do espectrofotômetro; <b>C</b> : Amostra de dipirona sódica injetável afixada em seu suporte.....	<b>36</b>
<b>Figura 11</b>	Registro do sinal do branco (ampola com água para injetáveis).....	<b>39</b>
<b>Figura 12</b>	Espectros bruto das amostras e padrões na região <i>NIR</i> entre 666 a 2500 nm.....	<b>39</b>
<b>Figura 13</b>	Espectros brutos das amostras e padrões na região <i>NIR</i> entre 884 a 1320 nm.....	<b>40</b>
<b>Figura 14</b>	Atribuição das bandas dos espectro- <i>NIR</i> para uma amostra de dipirona sódica injetável.....	<b>41</b>
<b>Figura 15</b>	Perfis das curvas resultante do pré-processamento em relação aos espectros originais.....	<b>42</b>
<b>Figura 16</b>	Gráfico dos escores da <i>PCA</i> dos espectros pré-processados das amostras comerciais e de padrões de dipirona sódica. Onde <b>P48</b> (em vermelho), <b>P50</b> (em verde), <b>P52</b> (em rosa), indica os padrões nas respectivas concentrações de 48, 50 e 52% (m/v), e <b>A</b> (em azul) indica as amostras comerciais.....	<b>43</b>
<b>Figura 17</b>	Gráfico de influência das amostras comerciais ( <b>A</b> ) e dos padrões de dipirona sódica ( <b>P</b> ).....	<b>44</b>

---



---

<b>Figura 18</b> Gráfico da variância explicada de X versus o número de componentes do modelo <i>SIMCA</i> .....	<b>45</b>
<b>Figura 19</b> Gráfico de <i>Coomans</i> do modelo <i>SIMCA</i> . Onde, em vermelho estão as amostras do conjunto de treinamento e, em verde estão as amostras do conjunto de predição.....	<b>46</b>
<b>Figura 20</b> Gráfico de <i>Coomans</i> para o modelo <i>SIMCA</i> construído para a faixa de 45 a 55% (m/v). Onde, em vermelho estão as amostras do conjunto de calibração/validação e, em verde estão as amostras do conjunto de predição.....	<b>48</b>
<b>Figura 21</b> Gráfico do valor predito versus o valor de referência para o conjunto de validação do modelo <i>PLSI</i> .....	<b>49</b>
<b>Figura 22</b> Gráfico do valor predito versus o valor de referência para o conjunto de predição do modelo <i>PLSI</i> .....	<b>50</b>
<b>Figura 23</b> Variação do <i>RMSEV</i> (% m/v) em função do número de variáveis inclusas no modelo <i>MLR-SPA</i> .....	<b>50</b>
<b>Figura 24</b> Gráfico do espectro <i>NIR</i> de uma amostra de dipirona sódica, indicando aproximadamente a localização das variáveis selecionadas pelo algoritmo <i>SPA</i> .....	<b>51</b>
<b>Figura 25</b> Gráfico do valor predito versus o valor de referência para o conjunto de validação do modelo <i>MLR-SPA</i> .....	<b>52</b>
<b>Figura 26</b> Gráfico do valor predito versus o valor de referência para o conjunto de predição do modelo <i>MLR-SPA</i> .....	<b>52</b>
<b>Figura 27</b> Variação do <i>RMSEV</i> (% m/v) em função do número de variáveis espectrais selecionadas pelo <i>ASA</i> .....	<b>53</b>
<b>Figura 28</b> Gráfico do espectro <i>NIR</i> de uma amostra de dipirona sódica, indicando aproximadamente a localização variáveis selecionadas pelo algoritmo <i>ASA</i> .....	<b>53</b>
<b>Figura 29</b> Gráfico do valor predito versus o valor de referência para o conjunto de validação do modelo <i>MLR-ASA</i> .....	<b>54</b>
<b>Figura 30</b> Gráfico do valor predito versus o valor de referência para o conjunto de predição do modelo <i>MLR-ASA</i> .....	<b>54</b>
<b>Figura 31</b> Gráfico do espectro <i>NIR</i> de uma amostra de dipirona sódica, indicando aproximadamente a localização variáveis selecionadas pelo modelo <i>MLR-SW</i> .....	<b>55</b>

---

---

<b>Figura 31</b> Gráfico do valor predito versus o valor de referência para o conjunto de validação do modelo <i>MLR-SW</i> .....	<b>56</b>
<b>Figura 33</b> Gráfico do valor predito versus o valor de referência para o conjunto de predição do modelo <i>MLR-SW</i> .....	<b>56</b>
<b>Figura 34</b> Gráfico do espectro <i>NIR</i> de uma amostra de dipirona sódica, indicando aproximadamente a localização variáveis selecionadas pelo modelo <i>MLR-GA</i> .....	<b>57</b>
<b>Figura 35</b> Gráfico do valor predito versus o valor de referência para o conjunto de validação do modelo <i>MLR-GA</i> .....	<b>58</b>
<b>Figura 36</b> Gráfico do valor predito versus o valor de referência para o conjunto de predição do modelo <i>MLR-GA</i> .....	<b>58</b>

---

---

<b>Tabela 1.</b> Regiões espectrais do Infravermelho.....	9
<b>Tabela 2.</b> Número de amostras respectivas concentrações, lotes e mês/ano de validade.....	33
<b>Tabela 3.</b> Análise <i>Screening</i> de todas as amostras de dipirona do conjunto de validação e previsão.....	46
<b>Tabela 4.</b> Análise <i>Screening</i> das amostras de dipirona comerciais do conjunto predição, usando o modelo <i>SIMCA</i> construído para a faixa de 45 a 55 % (m/v).....	47
<b>Tabela 5.</b> Valores de <i>RMSEV</i> , <i>RMSEP</i> e correlação linear de validação e de previsão dos modelos. ....	59

---

---

<i>AINES</i>	Antiinflamatórios Não Esteróide
<i>ANVISA</i>	Agência Nacional de Vigilância Sanitária
<i>ASA</i>	Algoritmo da Busca Angular
<i>CLAE</i>	Cromatografia Líquida de Alta Eficiência
$F_{Cal}$	Valor Calculado para o Teste F
$F_{Crit}$	Valor Crítico para o Teste F
<i>FT-IR</i>	Infravermelho-Transformada de <i>Fourier</i>
<i>GA</i>	Algoritmo Genético
<i>HCA</i>	Análise Hierárquica de Classes
<i>ICP-OES</i>	Espectroscopia de Emissão Ótica por Plasma Indutivamente Acoplado
<i>iPLS</i>	Regressão por Mínimos Quadrados Parciais por Intervalo
<i>K-NN</i>	Proporção do Vizinho mais Próximo
<i>KS</i>	<i>Kennard-Stone</i>
<i>LEDs</i>	Diodos Emissores de Luz
<i>MLR</i>	Regressão Linear Multipla
<i>MSC</i>	Correção de Espalhamento Multiplicativo
<i>NIR</i>	Infravermelho Próximo
<i>PC</i>	Componentes Principais
<i>PCA</i>	Análise por Componentes Principais
<i>PLS</i>	Regressão por Mínimos Quadrados Parciais
<i>QSAR</i>	Relação Quantitativa de Estrutura/Atividade
<i>RMSEV</i>	Raiz Quadrada do Erro Médio Quadrático de Validação
<i>RMSEP</i>	Raiz Quadrada do Erro Médio Quadrático de Previsão
<i>SIMCA</i>	Modelagem Independente Flexível por Analogias de Classes
<i>SPA</i>	Algoritmo das Projeções Sucessivas
<i>SPXY</i>	Partição do Conjunto de Amostras Baseado nas Distancias Conjuntas X-y
<i>SW</i>	<i>Stepwise</i>
<i>VIF</i>	Fator de Inflação da Variância
<i>UV-VIS</i>	Ultravioleta e Visível
<i>MIR</i>	Infravermelho Médio
<i>FAR</i>	Infravermelho Distante
<i>LW-NIR</i>	Infravermelho Próximo de Ondas Longas
<i>SW-NIR</i>	Infravermelho Próximo de Ondas Curtas

---

---

## RESUMO

**Título:** *Desenvolvimento de uma Metodologia para o Controle de Qualidade de Medicamento Injetável, sem Violação da Ampola, usando Espectroscopia NIR e Técnicas Quimiométricas.*

**Autor:** Fátima Aparecida Castriani Sanches

**1º Orientador (a):** Prof. Dr. Mário César Ugulino de Araújo

**2º Orientador (a):** Profa. Dra. Elaine Cristina Lima do Nascimento

Neste trabalho, é proposta uma nova metodologia que utiliza a combinação da espectrometria no infravermelho próximo (*NIR*) com métodos quimiométricos de análise multivariada para a determinação do teor de princípio ativo em um medicamento de uso parenteral. O medicamento utilizado neste estudo foi a dipirona sódica injetável. Para construção dos modelos de calibração multivariada foram empregadas amostras padrão de dipirona sódica preparadas no laboratório e amostras comerciais. Os espectros *NIR* foram registrados, pré-tratados e utilizados para a construção de modelos de reconhecimento de padrões e calibração multivariada. Para a análise inicial dos dados, uma análise exploratória foi realizada a partir de modelos *PCA*. Em seguida, modelos *SIMCA* foram elaborados e utilizados para uma análise *Screening* das amostras, empregando um nível de confiança de 95%. Essas análises serviram para constatar que todas as amostras (comerciais e preparadas no laboratório) poderiam ser utilizadas na composição dos conjuntos de calibração, validação e predição, empregados no desenvolvimento da metodologia. Para a seleção das amostras dos conjuntos utilizou-se o algoritmo *SPXY*. Inicialmente, o conjunto de calibração contendo amostras comerciais e as preparadas no laboratório foi empregado na construção de modelos de calibração multivariada *PLS1*, *MLR-SPA*, *MLR-SW*, *MLR-ASA* e *MLR-GA*. Esses modelos foram validados usando um conjunto de teste e foram aplicados para a determinação do teor de dipirona usando um conjunto de predição. Os resultados mostraram que todos os modelos apresentaram erros de previsão satisfatórios com relação ao método de referência, sendo bastante promissores para a predição do teor de dipirona sódica em medicamentos injetáveis. Como a metodologia proposta utiliza um processo não invasivo, não é necessário um tratamento prévio das amostras, nenhum reagente foi usado, e uma frequência analítica relativamente elevada de análise é alcançada.

**Palavras Chaves:** Controle de qualidade, medicamentos injetáveis, espectroscopia *NIR*, quimiometria.

---

---

## ABSTRACT

**Title:** *Development of a Methodology for Quality Control of Injection Drug, without violation of the ampoule, using NIR spectroscopy and chemometric techniques.*

In this work, a new methodology is proposed that uses a combination of Near Infrared Spectroscopy (NIR) with chemometric methods of multivariate analysis for the determination of active ingredient in parenteral drug. The drug used in this study was dipyron. To create the models of multivariate calibration, standard samples of dipyron prepared in laboratory and commercial samples were used. NIR spectra were recorded, pre-processed and used to build models of pattern recognition and multivariate calibration. For the initial analysis of data, an exploratory analysis was performed from PCA models. Then, SIMCA models were developed and used for Screening analysis of the samples, at a confidence level of 95%. These tests served to confirm which set of samples (commercial and prepared in the laboratory) could be used in the composition of sets of calibration, validation and prediction employed in the development of the methodology. For the selection of samples from the sets we used the SPXY algorithm. Initially, the calibration set containing commercial and prepared in the laboratory samples was used to build PLS1, MLR-SPA, SW-MLR, MLR-ASA and MLR-GA multivariate calibration models. These models were validated using a test set and applied to the determination of dipyron using a prediction set. The results showed that all models presented satisfactory prediction errors with respect to the reference method, and it is very promising for predicting the dipyron content in drugs injection. As the proposed methodology uses a non invasive process it is not necessary a previous treatment of the samples, no reagent was used, and a relatively high analytical rate is achieved.

Keywords: quality control, injectable drugs, NIR spectroscopy, chemometrics.

# CAPÍTULO I

## INTRODUÇÃO

---



## 1. INTRODUÇÃO

### 1.1 Caracterização do problema

A administração parenteral de medicamentos é uma prática muito comum na rotina hospitalar<sup>[1]</sup>. Estima-se que cerca de 40% de todos os medicamentos dispensados nos hospitais sejam injetáveis, sendo esta, uma das formas que exige maior qualidade e integridade dos medicamentos administrados, visto que dela dependem a saúde e até a vida dos pacientes. Sabe-se, porém, que as condições de transporte e armazenamento dos medicamentos (por exemplo, exposição a luz, umidade, etc.), na maioria dos hospitais brasileiros, estão longe das ideais preconizadas pelos fabricantes e pelos órgãos de fiscalização<sup>[2]</sup>. Este problema acarreta graves conseqüências na qualidade do medicamento, mesmo que este esteja dentro do prazo de validade<sup>[3]</sup>.

Embora a maioria das indústrias farmacêuticas operem com padrões de qualidade estabelecidos internacionalmente, muitos dos medicamentos produzidos ainda chegam a alguns mercados fora desses padrões<sup>[4]</sup>. Além disso, há o caso de indústrias que, visando maiores lucros, falsificam uma diversa gama de medicamentos. A falsificação pode ser realizada pela adulteração das formulações farmacêuticas, seja pela remoção total ou parcial do princípio ativo, pela troca deste por outras substâncias, ou pela troca ou remoção de excipientes caros importantes na formulação. Há casos de falsificação que ocorrem por meio de roubo de cargas ou lotes de medicamentos, seguido de adulteração no nome da indústria produtora, ou no prazo de validade do medicamento, ou ainda pela inserção de substâncias ilícitas ao próprio medicamento (esses casos de adulteração já ocorreram com o medicamento Viagra®, o que fez com que o fabricante *Pfizer* mudasse a embalagem do produto)<sup>[5]</sup>.

A falsificação e o roubo de medicamentos são problemas mundiais<sup>[6]</sup>. Estas práticas são tão lucrativas que grandes quadrilhas tem se formado e se especializado nessas operações<sup>[7]</sup>.

Entre 1997 e 1998, o Ministério da Saúde chegou a registrar 172 casos de falsificação, provocando questionamentos na sociedade sobre a qualidade e a segurança dos produtos comercializados nacionalmente<sup>[8]</sup>. Várias medidas foram adotadas pelo governo a partir dessa crise, destacando a modificação da lei que tornou a falsificação um crime hediondo e a aprovação da Política Nacional de Medicamentos, que garante à população o acesso a medicamentos seguros e eficazes, ações reforçadas com a criação da Agência Nacional de Vigilância Sanitária (ANVISA) em 1999. Logo após a criação da ANVISA, entre 1999 e 2003, apenas sete falsificações foram confirmadas<sup>[8]</sup>. Entretanto, as apreensões de remédios falsos e contrabandeados cresceram 315% de 2005 a 2008<sup>[7]</sup>, e esses produtos chegaram a farmácias e a hospitais públicos e privados<sup>[7]</sup>.

---

---

De acordo com dados do Ministério da Justiça, a adulteração de medicamentos movimentou 10 bilhões de dólares ao ano<sup>[9]</sup>. Para se ter idéia da gravidade do problema, de janeiro a março de 2009, a ANVISA retirou do mercado mais de 170 toneladas de medicamentos, sem registros, falsificados ou com prazo de validade vencido<sup>[9,10]</sup>. E ao final do primeiro semestre de 2009, a ANVISA em conjunto com a Polícia Federal apreenderam 316 toneladas de medicamentos<sup>[7]</sup>. Um aumento expressivo quando comparado a 2008, onde foi apreendido 45,5 toneladas<sup>[7]</sup>. Dentre os medicamentos que estão sendo falsificados estão: Glivec®, Voltaren®, Ponstan®, Terramicina®, Viagra®, Cialis®, Citotec®, Novalgina®, entre outros<sup>[6,7]</sup>.

Para se determinar todos os parâmetros de qualidade em um processo industrial, normalmente, precisa-se de um grande número de análises. Atualmente, a maioria destas análises é realizada empregando métodos instrumentais, sendo as metodologias mais comumente utilizadas no controle de qualidade de fármacos: a espectrometria na região do ultravioleta e visível (UV-Vis) e a cromatografia líquida de alta eficiência (CLAE)<sup>[11]</sup>. Estes métodos são lentos e trabalhosos, pois envolvem muita manipulação analítica (extração da amostra, preparação freqüente de reagentes e soluções de calibração, etc), são relativamente dispendiosos, pois requerem o uso de solventes e reagentes caros, precisam de pessoal qualificado para a manipulação analítica, produzem resíduos prejudiciais ao meio ambiente, são invasivas e destroem as amostras durante as análises. Em função de seu custo de operação e de manutenção, essas análises são realizadas apenas em forma de amostragem, tanto nas etapas de produção (doseamento e produto final), como na fiscalização. Deste modo não se tem um acompanhamento da vida útil do medicamento, desde a sua produção até o momento imediatamente anterior a sua administração no paciente. E é justamente nesse ínterim que os medicamentos estão sujeitos a decomposição e degradação do seu princípio ativo, o que leva à perda de suas propriedades terapêuticas<sup>[12]</sup>.

Este trabalho propõe o desenvolvimento de uma metodologia analítica para o controle de qualidade de medicamentos injetáveis líquidos, sem a violação da embalagem, utilizando espectrometria na região do infravermelho próximo (NIR)<sup>[13-17]</sup> e técnicas quimiométricas<sup>[18,19]</sup>. O fármaco escolhido como ponto de partida para este estudo foi a **dipirona sódica injetável**, por se tratar de um medicamento comercial de alto consumo na clínica médico/hospitalar devido a sua alta eficácia no tratamento de dores agudas, fracas e medianas, seu baixo custo, ampla e ubíqua disponibilidade, e boa margem de segurança.

A técnica proposta permite realizar medições de parâmetros de qualidade deste medicamento injetável, de forma não destrutiva/não invasiva<sup>[15]</sup>, sem qualquer violação física ou química das amostras em suas embalagens, permitindo a identificação de possíveis produtos

adulterados e/ou não conformes com os valores especificados pelas indústrias farmacêuticas produtoras.

Com esta metodologia espera-se oferecer à sociedade brasileira uma maior garantia de seus direitos à saúde e promoção da mesma, contando com um controle mais rápido e efetivo de seus medicamentos.

## *1.2 Medicamentos injetáveis*

Consideram-se preparações injetáveis as soluções, suspensões e emulsões estéreis de substâncias medicamentosas em veículos aquosos, oleosos ou outros apropriados para serem administrados por via parentérica<sup>[20]</sup>. Essa forma de administração de fármacos é muito eficiente, pois disponibiliza o princípio ativo diretamente na corrente sanguínea proporcionando uma ação terapêutica mais rápida e eficaz<sup>[12]</sup>.

Embora em pleno século XVII tenham sido feitas aplicações medicamentosas por via intravenosa no homem, relatando-se mesmo casos de transfusão de braço a braço, que foram proibidas em 1668<sup>[20,21]</sup>, pode-se dizer que a utilização da via parenteral só se difundiu depois dos trabalhos de Wood<sup>[22-27]</sup>, que em 1856, injetou pela primeira vez uma solução de sulfato de atropina mediante o uso de um aparelho também novo, a seringa de Pravaz<sup>[22-28]</sup>. A técnica de injeção chegou três anos mais tarde à França, onde a sua expansão se procedeu a partir dos trabalhos realizados por Behier<sup>[20,21,28]</sup>.

Posteriormente, Limousin sugeriu o condicionamento físico em pequenos recipientes e a sua esterilização pelo calor, tendo a nova forma farmacêutica sido aceita em todo o mundo e passando a vir inscrita em todas as farmacopéias<sup>[20,21]</sup>.

As preparações injetáveis são medicamentos que se introduzem diretamente nos tecidos, portanto exercem uma ação geral. Assim, compreende-se que a sua obtenção impõe condições rigorosas de controle, cuja dupla finalidade é a verificação da sua perfeita tolerância para os tecidos e inocuidade total no organismo<sup>[20,21,29]</sup>. Naturalmente que a primeira propriedade que todas as preparações injetáveis devem apresentar é a de serem estéreis<sup>[20,21,29-31]</sup>. Além desta característica, há, consoante o tipo de injetável e a via de administração, outras qualidades extremamente importantes. Assim, as preparações em veículo aquoso (soluções, dispersões coloidais, suspensões e emulsões de fase aquosa externa) devem ser isotônicas, apresentar pH próximo da neutralidade e ser apirogênicas, enquanto que as que são feitas utilizando um veículo oleoso deverão ser preparadas com óleos isentos de acidez<sup>[20,21,29]</sup>.

Para detectar os potenciais problemas, a indústria farmacêutica realiza frequentemente testes sobre a preparação, reconstituição ou estabilidade dos medicamentos injetáveis, para o controle de qualidade dos medicamentos, para assegurar uma formulação correta e eficaz. Dentre os testes realizados, encontramos: teste de esterilidade (contagem microbiológica e pesquisa de patógenos), doseamento, determinação da massa média, dureza, determinação do volume médio, friabilidade, pH, teste de pirogênios, teste de desintegração, teste de dissolução, toxicidade aguda, uniformidade de doses unitárias, umidade, viscosidade e identificação da substância ativa<sup>[20,21,29]</sup>. O uso destas informações é a chave que assegura a integridade da dosagem das formas farmacêuticas.

Quando se prepara medicamentos injetáveis é de particular importância a escolha de recipiente adequado que irá contê-los. Como é sabido, as ampolas ou frascos que se empregam são geralmente de vidro, eventualmente substituído por diversos materiais plásticos, como o polietileno ou o cloreto de polivinilo.

Os recipientes de vidro utilizados podem ser incolores ou corados. O vidro corado destina-se a proteger os medicamentos da ação da luz, sendo indicado quando o medicamento se altera nessas circunstâncias. A coloração é conferida aos vidros pela junção de diversos metais (crômio – que confere a cor verde, e o ferro e manganésio – que originam cores que vão do âmbar ao amarelo). Nas modernas farmacopéias e formulários indica-se o vidro de cor âmbar, desde que este satisfaça os ensaios de transparência<sup>[20,21]</sup>.

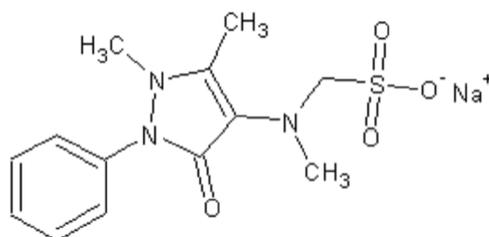
Qualquer que seja, o material utilizado para a confecção dos recipientes de acondicionamento, não deve reagir com os líquidos medicamentosos neles contidos. Isto significa que esses materiais, além de outras qualidades, devem apresentar alta resistência aos líquidos medicamentosos o que, na maioria dos casos, traduz-se por uma elevada resistência hidrolítica. Deve-se ainda considerar o seu tapamento, que geralmente, é feito por fusão, quando se trata de vidro, podendo ainda ser realizado mediante rolhas de borracha, de teflon, de elastômeros, etc.<sup>[20,21,29]</sup>.

### ***1.3 Dipirona sódica***

A dipirona sódica ou Metamizol é um analgésico, antipirético, antiinflamatório sintético extraído do alcatrão (anilina), pertencente ao grupo da pirazolona (antipirina, aminopirina, fenilbutazona, oxfenbutazona)<sup>[32]</sup>. É classificada como analgésico-antiinflamatório não esteroideal (AINE), dotada de efeitos diversos em função da dose empregada que pode ser baixa ( $< 10 \text{ mg.kg}^{-1}$ ) para ação antipirética; mediana ( $15\text{-}30 \text{ mg.kg}^{-1}$ ) para ação analgésica e alta ( $> 50 \text{ mg.kg}^{-1}$ ) para

ações antiinflamatória e antiespasmódica. Em ratos, em doses muito elevadas ( $> 200 \text{ mg.kg}^{-1}$ ) apresenta efeito anticonvulsivante [32-35].

Sua fórmula molecular é  $\text{C}_{13}\text{H}_{17}\text{N}_3\text{O}_4\text{S}$ , sua Massa Molar é  $311,35 \text{ g.mol}^{-1}$ , e seu nome químico é [(2,3-dihidro-1,5-dimetil-3-oxo-2-fenil-1H-pirazol-4-il) metilamino] metanosulfonato. A **Figura 1** ilustra sua fórmula estrutural.



**Figura 1** – Fórmula estrutural da dipirona sódica.

Foi introduzida no Brasil em 1922, sob o nome de *novalgina*®. Sua comercialização cresceu mundialmente até a década de 70, quando foram relatados graves casos de agranulocitose: uma condição clínica de déficit imunológico potencialmente fatal (7 a 25%), que exigia suspensão do tratamento e atendimento médico de urgência com potentes antibióticos. A maioria dos países do Hemisfério Norte, Suécia (1974), EUA (1977), Austrália e Japão proscreeveram a dipirona do receituário a não ser no mercado veterinário, por haver disponibilidade no receituário médico de AINES substitutos menos tóxicos para tratamento da dor (paracetamol, salicilatos, ceterolac, clomifênio, etc.) [41]. No entanto, o Brasil e diversos outros países não suspenderam sua comercialização e mantiveram seu uso até os dias atuais. Atualmente, a incidência estimada da agranulocitose é aceitável, pois oscila entre 0,2 a 2 casos por milhão de pessoas por dia de uso. Além disso, há outras drogas com maior risco de provocar agranulocitose como o antipsicótico clozapina com incidência de 8 casos para cada 1000 pacientes após 6 meses de tratamento [32,36-40].

Ela continua sendo prescrita na Rússia, Brasil, Índia e em muitos outros países sul-americanos, caribenhos, africanos e asiáticos. A dipirona (*Hoechst & Merck*) é um dos principais analgésicos da terapêutica brasileira [41], ocupando 31,8% do mercado. O paracetamol vem em segunda colocação com 29,7%, e o ácido acetilsalicílico em terceiro lugar com 27,1% [41].

A dipirona é o analgésico de preferência em 50% dos usuários de hospitais públicos em São Paulo [42]. A maior parte das vendas (em torno de 80%) ocorre sem prescrição médica, representando um papel importante na condenável auto-medicação, apesar da ANVISA incluí-la na lista de tarja vermelha (venda sob prescrição médica) [32,40,42,43].

É comercializada em diferentes formulações farmacêuticas (comprimidos, gotas, supositórios e injetáveis), podendo ser administrada pelas vias: Oral, retal, intramuscular (glúteo) e intravenosa.

Sabe-se que, dentre essas formulações citadas, a que exige uma atenção especial e maior cuidado é a injetável, que devido à via de administração, oferece maior risco a saúde.

#### *1.4 Metodologias de análise da dipirona sódica injetável*

O método oficial (farmacopeico) de análise de dipirona é o Método Clássico de Titulação Iodométrica<sup>[44,45]</sup>. Neste método, a reação química se processa lentamente e a análise torna-se trabalhosa porque há a necessidade de se manter uma temperatura abaixo de 15°C (banho de gelo); ele é dispendioso pois gera gasto com os reagentes químicos necessários para a preparação e padronização das soluções de trabalho. Além disso, a solução de iodo apresenta um problema de instabilidade por exposição ao ar livre.

Diversos procedimentos são descritos na literatura para determinação de dipirona em formulações farmacêuticas. Vassileva-Alexandrova e Shishmanov<sup>[46]</sup> propuseram um procedimento para determinação de dipirona baseado na reação desse fármaco com o reagente cloreto de neotetrazólio, produzindo um composto amarelo solúvel em meio de etanol-HCl. O método mostrou-se bastante seletivo e preciso na determinação de dipirona em produtos farmacêuticos.

Sakiara *et al.*<sup>[47]</sup> desenvolveram um método baseado na oxidação seletiva da dipirona na presença de ácido sulfúrico, promovendo a formação de formaldeído, que reage com o ácido cromotrópico produzindo um composto violeta, monitorado espectrofotometricamente em 575 nm.

Pereira *et al.*<sup>[48]</sup> desenvolveram um sistema por injeção em fluxo para a geração de íons  $I_3^-$ . Os íons triiodeto gerados em linha, ao reagirem com a dipirona injetada no sistema, oxidam o grupo metanossulfônico do fármaco a sulfato. O consumo de íons  $I_3^-$  pela dipirona provoca um decréscimo da absorvância do complexo triiodeto-amido que é monitorado espectrofotometricamente em 580 nm.

Senyuva *et al.*<sup>[49]</sup> utilizaram a Cromatografia Líquida de fase invertida para a determinação de dipirona em formulações sólidas e líquidas. Já Perez-Ruiz *et al.*<sup>[50]</sup>, se valeram da análise em fluxo com determinação fluorométrica.

A determinação espectrofotométrica proposta por Qureshi *et al.*<sup>[51]</sup> baseia-se na redução de íons Fe(III) por ação da dipirona e subsequente complexação dos íons Fe(II) com 1,10- fenantrolina, que produz um complexo vermelho, monitorado em 500 nm. Esse mesmo grupo de pesquisadores desenvolveu outro procedimento onde a dipirona foi oxidada por uma solução de iodato de potássio

em meio ácido, produzindo um composto de coloração amarela, monitorado espectrofotometricamente em 460 nm<sup>[52]</sup>.

Marcolino-Jr *et al.*<sup>[53]</sup>, realizaram a determinação turbidimétrica de dipirona em fluxo utilizando um reator contendo cloreto de prata imobilizado em resina poliéster, Neste trabalho, um reator em fase sólida contendo AgCl foi inserido em um sistema de análise por injeção em fluxo de linha única, para atuar como fonte de íons Ag<sup>+</sup>, e aplicado na determinação de dipirona (esse procedimento baseou-se na propriedade redutora da dipirona). Os íons Ag<sup>+</sup> contidos no reator eram reduzidos pela dipirona a Ag<sup>0</sup> que era então, detectada turbidimetricamente em 425 nm, e a concentração de dipirona injetada no sistema era proporcional à quantidade de Ag<sup>0</sup> gerada.

Vários outros métodos são descritos na literatura para determinação de dipirona por detecção espectrofotométrica<sup>[54-58]</sup>, fluorimétrica<sup>[50]</sup>, eletroquímica<sup>[59-61]</sup>, por quimiluminescência<sup>[62-64]</sup> e CLAE<sup>[65-69]</sup>. Entretanto, alguns desses procedimentos envolvem diversas etapas analíticas, são de alto custo, enquanto outros são trabalhosos e/ou requerem profissionais com boa capacitação para o seu desenvolvimento.

Como visto, tanto o método de referência como qualquer outro método de análise conhecido até o momento, exige a violação da embalagem dos medicamentos injetáveis e o consumo deste durante a análise, além de necessitar de muita atenção e manipulação analítica.

Para superar estes inconvenientes, é que se propõe o desenvolvimento de métodos analíticos usando a espectrometria NIR<sup>[13-17]</sup> para a análise deste medicamento. Esta metodologia permite a realização de análises das amostras diretamente em suas embalagens (ampolas ou frascos), evitando a destruição das amostras e diminuindo drasticamente a sua manipulação pelos analistas.

### ***1.5 Espectrometria na região do infravermelho próximo (NIR)***

A radiação NIR compreende a região usualmente delimitada entre os comprimentos de onda de 780 a 2500 nm do espectro eletromagnético. Nesta região as ocorrências de transições eletrônicas são raras. De fato, as ocorrências espectrais observadas nesta região se devem quase que totalmente às bandas de absorção relacionadas aos sobretons de transições fundamentais que ocorrem no infravermelho médio e as combinações de transições que ocorrem também naquela região<sup>[15]</sup>. A **Tabela 1** mostra as regiões espectrais do infravermelho, que é convencionalmente dividido em três regiões: infravermelho próximo (NIR), médio (*Middle Infrared - MIR*) e distante (*Far Infrared - FAR*)<sup>[70]</sup>.

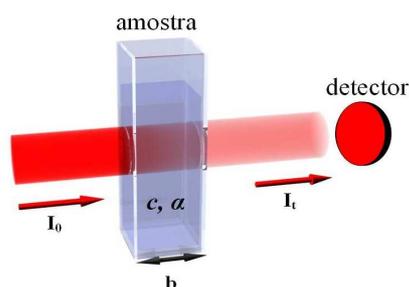
Tabela 1 - Regiões espectrais do Infravermelho.

Região	Intervalos das regiões		
	Comprimentos de Onda ( $\mu\text{m}$ )	Números de Onda ( $\text{cm}^{-1}$ )	Frequência (Hz)
<i>NIR</i>	0,78 a 2,5	12.800 a 4.000	$3,8 \times 10^{14}$ a $1,2 \times 10^{14}$
<i>MIR</i>	2,5 a 50	4.000 a 200	$1,2 \times 10^{14}$ a $6,0 \times 10^{12}$
<i>FAR</i>	50 a 1000	200 a 10	$6,0 \times 10^{12}$ a $3,0 \times 10^{11}$

Em geral, os espectros *NIR* estão associados a grupos funcionais que apresentam ligação muito polarizada como C-H, N-H, O-H e S-H.

As ocorrências espectrais na região *NIR* são caracterizadas por apresentarem intensidade fraca, cerca de 10 a 100 vezes inferiores àquelas observadas no *MIR* e *FAR*<sup>[15]</sup>. No entanto, estas ocorrências são mais susceptíveis ao ambiente dos átomos participantes da ligação química, o que, de certa forma, aumenta a seletividade de técnicas analíticas baseadas na espectroscopia *NIR*<sup>[15]</sup>. A diminuição de sensibilidade devida às transições fracas observadas, não se constitui em grande desvantagem, uma vez que fontes de radiação intensas e detectores de alta eficiência estão disponíveis para uso nesta região de comprimentos de onda. Por outro lado, o fato das ocorrências não apresentarem alta intensidade, permite o uso da técnica em amostras nas quais um alto teor de água se encontra presente. Assim, a espectroscopia *NIR* tem permitido determinar água em glicerina, hidrazina, Freon, membranas orgânicas e acetona, além de constituir um valioso método de análise das misturas de aminas aromáticas<sup>[71]</sup>. A determinação de umidade em amostras de produtos agrícolas é uma aplicação rotineira da técnica *NIR*<sup>[72]</sup>.

A região do *NIR* é acessível com óptica de quartzo, vantagem que se associa à maior sensibilidade dos respectivos detectores e maior intensidade das fontes luminosas<sup>[70]</sup>. Com o uso de instrumental apropriado, as medidas espectroscópicas na região *NIR* podem ser utilizadas analiticamente de forma igual à que se emprega na espectroscopia UV-vis, através de medidas de transmitância/absorbância. A absorção da radiação atenua o feixe de acordo com a lei de Beer<sup>[70]</sup>, ou seja, à medida que a luz atravessa um meio contendo um analito que absorve, um decréscimo de intensidade ocorre na proporção que o analito é excitado (**Figura 2**). A transmitância T da solução é a razão entre a potência radiante do feixe transmitido ( $I_t$ ) e a potência radiante do feixe incidente ( $I_0$ ).



**Figura 2** - Determinação da transmitância de uma amostra. Onde:  $b$  = caminho óptico do meio;  $\alpha$  = absorvidade molar do analito;  $c$  = concentração da espécie absorvente.

A complexidade e o alto grau de sobreposição das bandas que ocorrem na região *NIR* impede, na maioria das vezes, a utilização da absorbância medida em um único comprimento de onda como parâmetro analítico na determinação da concentração de uma dada espécie. No entanto, a relação entre concentração e medida espectroscópica continua sendo regida pela lei de Beer<sup>[70]</sup>.

Para a realização de análises químicas baseadas na espectroscopia *NIR*, dois tipos básicos de equipamentos podem ser utilizados: espectrômetros ou fotômetros. Os espectrômetros *NIR* são equipamentos normalmente grandes, utilizados para registrar espectros de uma amostra em toda a faixa espectral do *NIR*. Por conta dessa característica, são equipamentos caros utilizados por centros de pesquisa ou por laboratórios de indústrias de grande porte financeiro. Logo, não são equipamentos desenvolvidos para a análise de campo. Por sua vez, os fotômetros *NIR* são equipamentos muito mais baratos, porém eles são construídos para se trabalhar com um número discreto de pequenas regiões espectrais (os comprimentos de onda) distribuídas ao longo da faixa do *NIR*. Estes comprimentos de onda são isolados empregando filtros ópticos, cuja banda espectral passante isolada é em geral muito larga, o que prejudica a precisão e a qualidade dos fotômetros *NIR*. Fotômetros *NIR* de boa qualidade e muito mais simples podem ser construídos usando diodos emissores de luz (*LED*)<sup>[73,74]</sup>.

Dispõe-se, atualmente, de um grande número de espectrômetros que operam na faixa do infravermelho próximo. A complexidade pode variar desde os mais simples, que empregam filtros de interferência na seleção de poucos comprimentos de onda, até os mais sofisticados, que empregam as técnicas multiplexadas como os de transformada de *Fourier* e de *Hadamard*. Os instrumentos que empregam filtros óptico-acústicos sintonizáveis têm se tornado populares devido, principalmente, à sua robustez, que possibilita o uso do espectrofotômetro para monitoração em linha nos processos industriais. Os detectores mais empregados são os de sulfeto de chumbo que apresentam desempenho adequado nesta faixa de comprimento de onda, principalmente quando resfriados com o uso de elementos tipo *Peltier*. Os detectores de InAs e GaInAs têm sido bastante

empregados nos instrumentos de varredura mais modernos, devido à exigência de respostas rápidas para permitir varreduras de amplas regiões espectrais a taxas tão elevadas quanto  $500 \text{ nm s}^{-1}$ .

Os espectros *NIR* podem ser relacionados tanto as propriedades químicas qualitativas<sup>[75,76]</sup> e quantitativas<sup>[76-78]</sup>, como as propriedades físicas (viscosidade, densidade, bem como a presença de pequenas partículas sólidas ou qualquer material estranho, perceptível ou não ao olho humano, nos medicamentos injetáveis).

A espectroscopia *NIR* tem sido potencialmente aplicada à solução de diversos problemas analíticos, sendo uma realidade nos mais variados setores da cadeia produtiva. Por exemplo, o Canadá certifica o teor de proteína totais dos grãos de cereais exportados utilizando medidas de espectros *NIR* diretamente nos grãos<sup>[79]</sup> e calibração multivariada, em um procedimento que não requer mais do que alguns minutos para ser efetuado. Comparando este método com o método de referência usado para a mesma finalidade, o Método de *Kjehldal*<sup>[80]</sup>, fica evidente a importância de se desenvolver metodologias analíticas, simples, baratas, que não destruam a amostra e não produzam resíduos químicos prejudiciais ao homem e ao meio-ambiente<sup>[81]</sup>. O método *Kjehldal* usa digestão ácida drástica da amostra empregando mercúrio ou selênio como catalisador, seguida de destilação e a titulação da amônia produzida.

Esta técnica também vem sendo empregada na indústria de fármacos<sup>[15,72]</sup>, polímeros<sup>[15,72]</sup> e produtos petroquímicos<sup>[72,82]</sup>, e produtos agrícolas em geral<sup>[72,82,84]</sup>.

No contexto da análise de medicamentos, a maioria dos trabalhos publicados é com formas farmacêuticas sólidas<sup>[72]</sup>. A popularidade da técnica cresceu muito a partir de 1986<sup>[72]</sup>, ano em que Ciurczak e Maldacker<sup>[85]</sup> publicaram o uso *NIR* na análise de misturas de formulação de comprimidos de aspirina, butalbital, e cafeína, com suas concentrações variando num intervalo de 0%(ausência de princípio ativo) e de 90 a 110%. Neste experimento, espectros de placebos eram subtraídos do espectro da composição completa, gerando espectros muito próximos aos das drogas omitidas.

Em 1997, Candolfi et al.<sup>[86]</sup> estudaram as fontes de variação na medição *NIR* de comprimidos e cápsulas. As fontes de variação estudadas foram a repetibilidade de medição, o posicionamento da amostra, variabilidade do dia-a-dia, objetos de variabilidade dentro de um lote, e de modo contínuo avaliaram a variabilidade dos lotes. Este trabalho deu um dos primeiros indícios da importância do posicionamento da amostra.

O primeiro trabalho publicado sobre análise de comprimidos diretamente em suas embalagens foi o de Dempster et al.<sup>[87]</sup> onde foram utilizadas três configurações de amostragem para investigar a classificação da droga presente. Na primeira configuração era necessária a

remoção do comprimido do blister, na segunda abordagem, os comprimidos foram analisados através da embalagem de plástico usando a janela espectrômetro, e no terceiro arranjo, os comprimidos foram analisados através da embalagem blister de plástico com uma sonda de fibra óptica.

Em 2000 Moffat et al. [88] demonstraram que o método de análise *NIR* preenche os critérios estabelecidos pela ICH-Q2 [89,90]. Critérios como: exatidão, precisão, especificidade, limite de detecção, limite de quantificação, linearidade, intervalo, robustez e adequação do sistema de teste, demonstraram ser adequados para uso em testes de rotina. Este trabalho foi o ponto de partida para a validação da metodologia baseado *NIR* para análise de medicamentos [72].

Atualmente, métodos analíticos baseados na espectroscopia *NIR* estão sendo empregados como método de referência na Europa e em alguns países, onde já estão sendo incorporados à farmacopeia oficial [91].

Embora a espectroscopia *NIR* já esteja se consagrando e sendo aplicada em diversos segmentos da indústria farmacêutica [72], até o presente momento, poucas foram as aplicações publicadas sobre o seu uso para a análise de medicamentos injetáveis [72], e destas nenhuma sugere a análise de injetáveis sem a violação física da ampola.

Devido à grande quantidade de dados gerados, a alta sobreposição espectral e baixa intensidade dos sinais, este estudo é muito complexo, e para que uma metodologia baseada na espectroscopia *NIR* possa ser implementada com eficácia, é necessária a utilização da quimiometria [18,19]. A combinação da espectroscopia *NIR* com técnicas quimiométricas de análise multivariada, tem possibilitado o desenvolvimento de metodologias bastante promissoras para o controle de qualidade de medicamentos injetáveis, entre outros motivos, por determinar tanto as propriedades químicas (qualitativas e quantitativas), como as propriedades físicas das amostras, de forma não-invasiva, pouco laboriosa, rápida e precisa, sem necessidade de qualquer tratamento químico, ou seja, sem destruição das amostras [15,72].

## 1.6 Quimiometria

Quimiometria é a aplicação de métodos estatísticos ou matemáticos em dados de origem química (química orgânica, analítica ou medicinal) de natureza multivariada [92]. A Sociedade Internacional de Quimiometria (*International Chemometrics Society - ICS*) propõe a seguinte definição: “Quimiometria é a ciência relacionada a medidas realizadas em um sistema ou processo químico, obtendo informações sobre o estado do sistema através da aplicação de métodos matemáticos ou estatísticos” [93].

---

---

Muita ênfase tem sido dada aos sistemas multivariados, nos quais se pode medir muitas variáveis simultaneamente, ao se analisar uma amostra qualquer. Nesses sistemas, a conversão da resposta instrumental no dado químico de interesse, requer a utilização de técnicas de estatística multivariada, álgebra matricial e análise numérica. Essas técnicas se constituem no momento na melhor alternativa para a interpretação de dados e para a aquisição do máximo de informação sobre o sistema<sup>[92]</sup>. A organização do conjunto de dados em análises multivariada são geralmente apresentados em uma matriz, onde as linhas desta matriz representam o conjunto das amostras e as colunas, das variáveis medidas. A matriz pode ser representada graficamente num espaço multidimensional, onde cada variável medida corresponde a uma dimensão do espaço e cada amostra um ponto no espaço<sup>[18,94,95]</sup>.

No contexto de análise de medicamentos injetáveis usando espectroscopia *NIR*, dois grupos de técnicas quimiométricas<sup>[18]</sup> podem ser utilizados: O grupo das técnicas de reconhecimento de padrão<sup>[96]</sup> (sessão 1.7) e o grupo das técnicas de calibração multivariada<sup>[97]</sup> (sessão 1.8). Entretanto, devido a natureza dos dados, um pré-processamento pode ser necessário antes de realizar o seu processamento.

O pré-processamento<sup>[18]</sup> pode ser definido como qualquer manipulação matemática feita com o conjunto de dados. Há um número considerável de técnicas disponíveis na espectroscopia *NIR* para ajudar os usuários com a interpretação qualitativa e quantitativa dos espectros<sup>[98]</sup>. O pré-processamento é usado para remover ou reduzir fontes irrelevantes de variação e tratar as matrizes de dados para tornar os cálculos melhor condicionados antes da modelagem. Ao ser aplicado, altera os dados originais. A escolha da técnica de pré-processamento dos dados adequada é uma etapa crítica para o êxito de uma análise de dados<sup>[18]</sup>.

Para que se execute um bom pré-processamento é necessário que haja interação entre pré-processamento e o exame dos dados, e que se conheça as características (como a precisão das análises) e o comportamento dos dados<sup>[18]</sup>.

As técnicas de pré-processamento de dados podem ser divididas basicamente em dois tipos: (a) pré-processamento de dados aplicado às amostras e (b) pré-processamento de dados aplicado às variáveis.

**(a) Pré-processamento de dados aplicado às amostras** – Neste caso são aplicadas técnicas que operem em uma amostra de cada vez e sob todas as variáveis. Dentre as técnicas de pré-processamento nas amostras temos: normalização, ponderação, suavização e correções de linha de base.

---

---

A normalização coloca todas as amostras em uma mesma escala dividindo-as por uma constante, para remover variações sistemáticas das amostras. Nesta técnica três constantes podem ser usadas: normalização por unidade de área (que divide cada elemento do vetor amostra pela sua norma-1), normalização por unidade de comprimento (que divide cada elemento do vetor amostra pela norma-2), e normalização fazendo a máxima intensidade igual a 1 (que divide cada elemento do vetor pela norma infinita, definida como o máximo valor, em absoluto, do vetor) [18]. Um outro tipo de normalização é a Normalização Variável Padrão (SNV). A SNV [99] é uma linha de transformação orientada para os centros das escalas dos espectros individuais. Funciona como um auto-escalonamento nas amostras, onde na escala vertical, cada espectro é centrado em zero e varia de cerca de -2 a 2. Basicamente o SNV uniformiza cada espectro, utilizando apenas os dados de cada espectro individualmente [99].

A ponderação de amostras é uma técnica de pré-processamento que dá a algumas amostras maior peso nas análises em relação a outras, deste modo enfatiza uma amostra sobre as demais. Esta técnica é utilizada para remover variações sistemáticas nas amostras, e para isso, atribui-se as amostras consideradas mais importantes pesos proporcionais a sua importância para o modelo, e multiplica-se cada elemento do vetor amostra pelo seu peso [18].

Já a suavização é uma técnica de pré-processamento utilizada para reduzir a variação randômica presente em um conjunto de dados [100]. Esta técnica é usada para reduzir matematicamente o ruído aleatório e aumentar a relação sinal/ruído. A suavização normalmente reduz o número de variáveis, e pode eventualmente, remover informações úteis.

Existem diversos tipos de suavização, entre eles temos: Suavização por média [18]; por média móvel [18]; por mediana móvel [18]; por polinômio móvel [18]; por transformada de *Fourier* [18]; e por transformada de *Wavelet* [18]. Entretanto, neste trabalho iremos abordar apenas a suavização por polinômio móvel, também conhecida por Método de *Savitzky-Golay* [101].

No Método de *Savitzky-Golay* [101], um polinômio de baixa ordem, pré-definido, é ajustado, por mínimos quadrados, aos pontos de uma janela. Tanto a ordem do polinômio, como o tamanho da janela, são estabelecidos por um usuário. O uso do método de *Savitzky-Golay* resulta na eliminação de  $(\text{tamanho da janela} - 1)/2$  pontos em cada extremidade do vetor. A escolha do número de ponto é uma etapa crítica, uma vez que, ao se escolher um número elevado pontos pode-se remover informações relevantes, e ao se escolher um número pequeno de pontos, o ruído pode ainda permanecer nos dados espectrais.

Além de ruído aleatório, as medidas podem conter variações sistemáticas não relacionadas com a investigação química, denominadas de feições da linha de base, que podem dominar a análise

se não forem removidas. As feições de linha de base podem ser removidas usando modelos explícitos tais como: derivativos (usando derivadas de primeira e segunda ordem), e correção do fator multiplicativo do sinal (*MSC*)<sup>[18]</sup>.

Os métodos derivativos (primeira e segunda derivada) não se baseiam na identificação dos pontos de linha de base, sendo muito útil em casos onde a linha de base é difícil de identificar. Com as operações de derivação dos espectros, as informações contidas nos diferentes comprimentos de onda são geralmente acentuadas<sup>[100]</sup>. Dentre os métodos derivativos conhecidos, pode-se citar: derivação por simples diferença móvel (onde a simples diferença entre pontos adjacentes é usada para estimar a 1ª derivada, a 2ª e sucessivas derivadas); derivação por diferença de média móvel (onde uma janela é selecionada e a diferença entre médias móveis é tomada); e derivação pelo método de *Savitzky-Golay* (onde, na suavização ajusta-se um polinômio a uma janela móvel do sinal resposta e o ponto central da janela é estimado pelo polinômio. Na derivação por esse método é usado o valor estimado pela derivada do polinômio da janela)<sup>[18]</sup>. Deve-se ter cuidado com a qualidade dos espectros aos quais deseja-se aplicar o cálculo das derivadas, uma vez que não só os sinais espectrais, como também os ruídos, tornam-se acentuados.

O *MSC* é uma técnica que pode ser generalizada como uma correção de sinal multiplicativa. Foi desenvolvida para corrigir as variações devidas ao espalhamento da luz para as medidas *NIR* de reflectância difusa<sup>[18,100]</sup>. Mesmo com o uso da técnica, o espectro pré-processado continua a assemelhar-se ao espectro original, não prejudicando a sua interpretação<sup>[18]</sup>.

**(b) Pré-processamento de dados aplicado às variáveis** - Para se realizar este pré-processamento faz-se uso de técnicas que operam em uma variável de cada vez e sob todas as amostras. Usualmente, três tipos de pré-processamento são aplicados: centragem dos dados na média, escalonamento e auto-escalonamento<sup>[18,94]</sup>.

Na centragem dos dados na média<sup>[18]</sup> calcula-se o valor médio para cada variável e o valor de cada elemento da coluna é subtraído pelo seu respectivo valor médio, resultando numa matriz, onde todas as colunas possuem média zero. Como a centragem na média geralmente não prejudica a análise dos dados, e sempre colabora para a resolução de problemas presentes nos mesmos, é por isso, usada como configuração padrão em alguns pacotes computacionais para cálculos estatísticos<sup>[18]</sup>.

No escalonamento<sup>[18]</sup>, cada elemento de uma linha é dividido pelo desvio padrão da sua respectiva variável, de modo que a variância se torne unitária e as variáveis passem a ser expressas em unidades de desvio padrão.

O auto-escalonamento<sup>[18]</sup> consiste em centralizar os dados na média e, em seguida, efetuar o escalonamento. Desta forma, as variáveis terão médias zero e desvios padrão igual a um.

Tanto o escalonamento quanto o auto-escalonamento é utilizado quando se pretende atribuir os mesmos pesos às variáveis do sistema de investigação, mas para dados espectroscópicos, estes dois últimos métodos não são comumente utilizados devido a possibilidade de maximizar informações irrelevantes (ruído)<sup>[18]</sup>.

### 1.7 Métodos de Reconhecimento de Padrão

As técnicas quimiométricas de reconhecimento de padrões baseiam-se na busca de semelhanças, ou agrupamentos de amostras, a partir da similaridade entre elas ou a partir de diferenças em suas propriedades físicas ou químicas, o que permite também identificação de amostras anômalas. Dentre essas técnicas, estão as **não supervisionadas**<sup>[18]</sup>: Análise de Componentes Principais (*PCA*) e Análise de Agrupamentos Hierárquicos (*HCA*), que servem para a análise exploratória dos dados, ou seja, são usados apenas para examinar similaridades ou diferenças entre amostras, bem como para identificar a formação de padrões no espaço multidimensional.; e as **supervisionadas**<sup>[18]</sup>: Modelagem Independente e Flexível por Analogia de Classe (*SIMCA*) e Método do K-ésimo Vizinho mais Próximo (*KNN*), que servem para classificação das amostras ou agrupamento das mesmas em classes de acordo com suas características<sup>[18,94-96]</sup>. Neste trabalho serão aplicadas apenas as técnicas *PCA* e *SIMCA*.

#### 1.7.1 PCA

A *PCA* é uma ferramenta exploratória projetado por Karl Pearson<sup>[102]</sup> em 1901 para identificar as tendências desconhecidas em um conjunto de dados. É um dos métodos estatísticos mais usados quando se pretendem analisar dados multivariados<sup>[103]</sup>. O algoritmo foi introduzido em 1933 pelo psicólogo H. Hotelling<sup>[104]</sup>, daí às vezes é chamado Transformada Hotelling<sup>[104]</sup>.

A análise por componentes principais (*PCA*) é um dos métodos mais comuns empregados na análise de informações<sup>[92,105]</sup>, sendo principalmente utilizada pela sua capacidade de compressão dos dados em função da existência de correlação entre diversas variáveis medidas, permitindo a transformação de um conjunto de variáveis originais, intercorrelacionadas, num novo conjunto de variáveis não correlacionadas<sup>[106]</sup>. Para isso, faz-se uso de operações matriciais para se obter novas coordenadas das amostras em outro sistema de eixos, denominados de fatores ou componentes principais (*PC*)<sup>[18,94,95,107]</sup>, que seja mais conveniente para a análise de dados. A construção dos

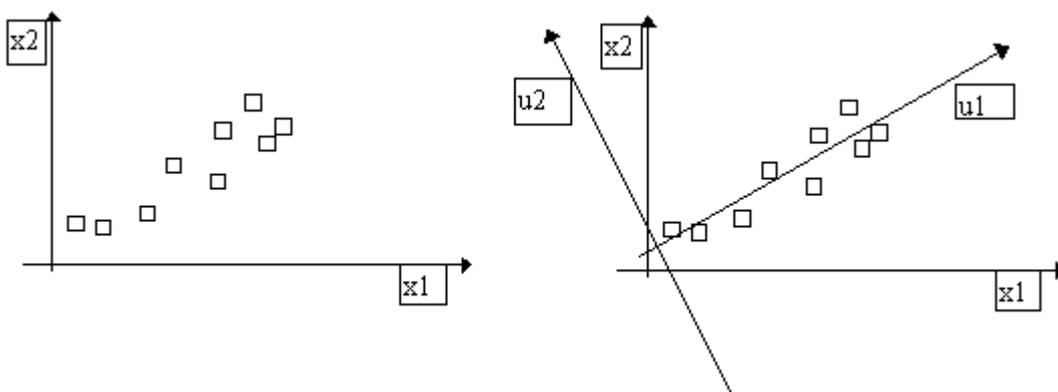
novos eixos é feita a partir da combinação linear do conjunto de variáveis originais, sendo estes eixos sempre ortogonais entre si<sup>[18,94,95,107]</sup>.

Assim, quando aplicamos um algoritmo de *PCA* num conjunto de variáveis, como em espectros no infravermelho, o conjunto original destas variáveis é substituído por um novo conjunto de variáveis (as *PC*), o qual pode ser facilmente reconstruído a partir da combinação linear das variáveis originais (espectros)<sup>[106]</sup>.

A vantagem do uso desta técnica baseia-se num princípio relativamente simples: ao decorrelacionar os dados, elimina-se a parte da informação redundante em cada dimensão. Os benefícios resultantes são vários: os dados podem ser descritos de uma forma mais concisa, certas características escondidas dos dados podem vir a luz depois de transformadas, e a distribuição dos dados pode ser representada (aproximadamente) pelas densidades individuais de cada dimensão<sup>[108]</sup>.

Muitos aplicativos disponíveis utilizam a técnica de decomposição do valor singular (*SVD*)<sup>[109]</sup> para obter as *PC*, sendo neste caso a primeira componente principal (*PC1*) definida na direção (eixo) de maior variância do conjunto de variáveis originais<sup>[106]</sup>. De forma decrescente em termos de variação são definidas as demais componentes principais, porém estas serão sempre ortogonais a *PC1* e entre si. Assim, um sistema que seja reduzido a 3 *PC* (*PC1*, *PC2* e *PC3*) se assemelha ao sistema cartesiano de coordenadas, em que todos os eixos são linearmente independentes<sup>[106]</sup>.

Por exemplo, observe na **Figura 3** um conjunto hipotético de pontos no espaço euclidiano, descrito pelas coordenadas  $x_1$  e  $x_2$ . Para a conversão no novo sistema de coordenadas, a primeira componente principal ( $u_1$ , que é o autovetor com o maior autovalor) corresponde a uma linha que passa através da média e minimiza a soma do erro quadrado com esses pontos. A segunda componente principal ( $u_2$ ) segue o mesmo princípio, sendo ortogonal a primeira componente e apresentando o segundo maior autovalor. A operação prossegue até que todas as novas coordenadas sejam estabelecidas<sup>[109,110]</sup>.



**Figura 3.** Exemplo da conversão das variáveis em novos eixos no espaço multidimensional.

Cada autovalor indica a porção da variância que está correlacionada com cada autovetor<sup>[109,110]</sup>. Assim, a soma de todos os valores próprios é igual à soma do quadrado da distância com os seus pontos de média, dividido pelo número de dimensões. A *PCA*, essencialmente, gira o conjunto de pontos em torno de sua média, a fim de alinhar as primeiras componentes principais<sup>[109,110]</sup>. Isso move a variação tanto quanto possível para as primeiras dimensões (usando uma transformação linear), ficando as primeiras componentes individualmente responsáveis pela variância das observações, e neste sentido, representam-nas mais claramente<sup>[108]</sup>. Estes novos eixos representam as direções com máxima variabilidade, na ordem da maior para menor variância de  $PC_1$  para  $PC_n$ . Os valores restantes nas últimas dimensões tendem a ser altamente correlacionados e podem ser descartados com perda mínima de informação química uma vez que as *PC* iniciais é que são as responsáveis por explicar a maior proporção da variação total associada ao conjunto original<sup>[108]</sup>.

Para um conjunto de espectros *NIR*, inicialmente a matriz de dados  $\mathbf{X}$  ( $m \times n$ ), onde  $m$  corresponde ao número de amostras (espectros) e  $n$  ao número de variáveis (frequências do espectro), é decomposta em 3 outras matrizes,  $\mathbf{U}$ ,  $\mathbf{S}$  e  $\mathbf{V}$ <sup>[92]</sup>, conforme a expressão:

$$\mathbf{X} = \mathbf{USV}^t \quad (1)$$

As colunas de  $\mathbf{U}$  e  $\mathbf{V}$  são ortogonais. A matriz  $\mathbf{V}$  corresponde a contribuição de cada variável nas *PC*, e nada mais é do que o cosseno do ângulo entre o eixo da variável e o eixo da *PC* que varia entre os valores de -1 a 1<sup>[106]</sup>. A matriz  $\mathbf{V}$  é conhecida como a matriz dos pesos, onde a primeira coluna contém os pesos de  $PC_1$  e assim por diante. O produto  $\mathbf{U} \times \mathbf{S}$  corresponde as coordenadas das amostras com relação as *PC* (obtidas pelo produto da matriz de dados pela matriz dos pesos). A matriz  $\mathbf{U}$  é denominada de matriz dos escores e define a nova localização das amostras de  $\mathbf{X}$  no vetor espaço, e  $\mathbf{S}$  é matriz diagonal, cujos elementos (valores singulares) contém informações sobre a quantidade de variância que cada componente principal descreve. A matriz  $\mathbf{S}$  é importante na determinação da dimensionalidade intrínseca da matriz de dados, podendo os analistas definir quantas *PC* ou fatores devem ser utilizados para análises posteriores<sup>[106]</sup>.

A *PCA* é freqüentemente usada para a redução de dimensionalidade por ter a distinção de ser a transformação linear ideal para manter o subespaço que tem a maior variância<sup>[109,110]</sup>.

Com o método *PCA* é possível<sup>[18]</sup>:

- Mostrar de que forma uma amostra é diferente da outra;
- Identificar quais variáveis contribuem mais para estas diferenças;

- Identificar quais variáveis contribuem de forma correlacionadas ou independentemente uma da outra;
- Detectar padrão de amostras (agrupamentos);
- Distinguir informação útil de ruído;
- Identificar amostras anômalas (*outliers*), etc.

### 1.7.2 SIMCA

O *SIMCA* é um método de reconhecimento de padrão supervisionado, utilizado para classificação de amostras futuras. O método *SIMCA*<sup>[111]</sup> usa *PCA* para modelar a forma e a posição do objeto formado pelas amostras no espaço linha para definição de uma classe. No *SIMCA* um modelo multidimensional é construído para cada classe e a classificação de amostras de predição é efetuada determinando dentro de qual classe a amostra está situada. Para isso, um número ótimo de componentes principais é determinado independentemente para cada classe e o modelo final é obtido através da definição da fronteira entre cada *PCA* modelado<sup>[106,112]</sup>. Como consequência, uma hipercaixa com tantas dimensões quantas forem o número de *PC* mais adequados para cada classe é obtida.

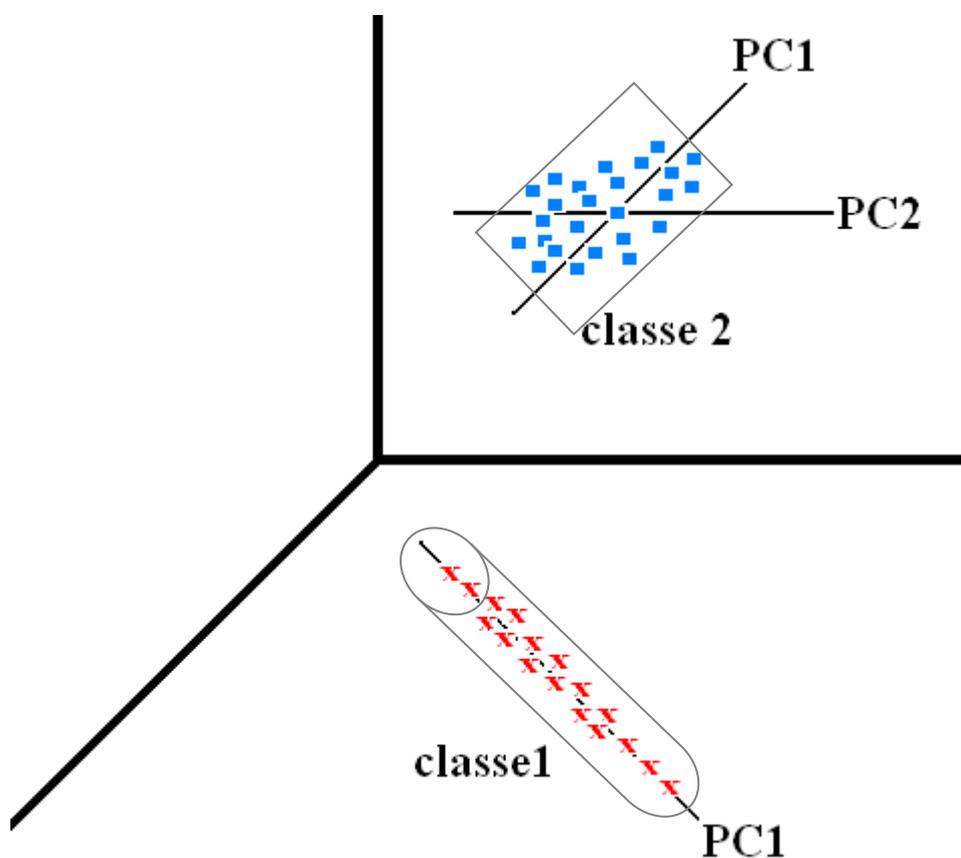
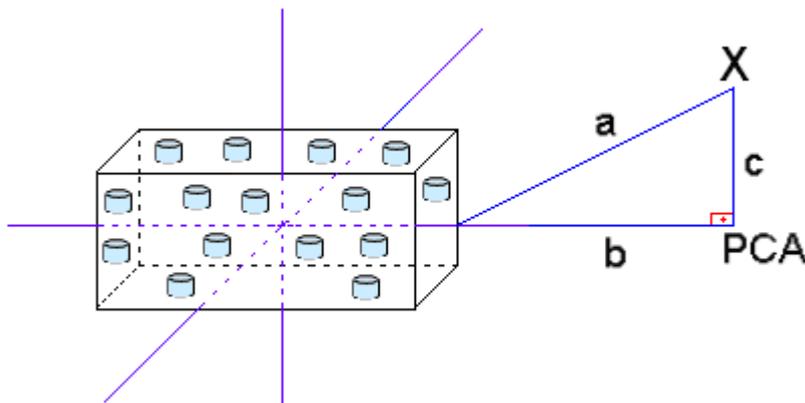


Figura 4. Exemplo de modelos *SIMCA*.

Para modelar ou delimitar as fronteiras onde estão localizadas as amostras, é utilizado um conjunto de treinamento contendo amostras de todas as classes e cuja identidade de cada amostra deste conjunto é conhecida a priori. As fronteiras de cada classe são modeladas ou delimitadas usando apenas as amostras daquela classe conhecida no conjunto de treinamento<sup>[106,112]</sup>. Uma vez definidas as hipercaixas é importante verificar se existem classes com sobreposição. A **Figura 4** apresenta uma simples exemplificação de modelos *SIMCA* hipotéticos, para duas classes de amostras distintas.

O poder de discriminação dos modelos *SIMCA* se baseia na não existência de sobreposição entre duas classes diferentes, bem como na maior distância interclasses possível<sup>[112]</sup>. A classificação de uma amostra externa àquelas empregadas na modelagem *SIMCA* se baseia nas diferentes projeções possíveis no espaço dos escores avaliando-se se a mesma encontra-se dentro das fronteiras definidas para a respectiva classe em todas as projeções obtidas<sup>[112]</sup>. Assim, na predição de um modelo *SIMCA* a amostra é classificada como pertencente à classe se apresentar variância dentro de um valor crítico determinado pelo modelo. Este valor crítico é função do valor da proximidade de uma amostra “X” em relação à fronteira da caixa tridimensional construída pelo modelo *SIMCA*. Para melhor entendimento sobre o cálculo do valor crítico, observe um modelo hipotético na **Figura 5**, onde se pretende fazer a predição de uma amostra X.



**Figura 5.** Predição com um modelo *SIMCA*.

O primeiro passo é encontrar os valores de  $a$ ,  $b$  e  $c$ , onde “ $a$ ” corresponde à proximidade de X em relação à fronteira da caixa tridimensional, “ $b$ ” é a distância entre a fronteira e a projeção de X na PCA e “ $c$ ” o resíduo da PCA<sup>[18, 94,95, 113]</sup>. O valor de  $a$  pode ser obtido pela **Equação 2**.

$$a^2 = b^2 + c^2 \quad (2)$$

---

---

O cálculo das distâncias  $a$ ,  $b$  e  $c$  determina se uma amostra desconhecida pertence à classe definida pelo seu modelo *SIMCA*. Após calcular o valor de  $a$ , este é dividido pela variância da classe para formar um valor calculado,  $F_{\text{cal}}$ . Utiliza-se, então, o teste F onde um valor crítico,  $F_{\text{crit}}$ , é escolhido empiricamente ou a partir de uma tabela do teste F. Se o  $F_{\text{cal}}$  for menor do que o valor crítico adotado, a amostra desconhecida pode ser classificada como pertencente à classe [18, 94,95, 113].

Portanto, a atribuição de uma amostra teste a uma dada classe é baseada na projeção da mesma no espaço dos escores e sua distância das fronteiras da classe em questão. Isto é repetido para todas as classes e no final, se os modelos não tiverem poder de discriminação suficiente, a amostra pode ser membro de mais de uma classe.

Se uma amostra é classificada em mais de uma classe, as medidas ou os modelos *SIMCA* não são potencialmente discriminantes. Se uma amostra não é classificada como membro de uma das classes do conjunto de treinamento, ela pode pertencer a uma outra classe ou pode ainda ser uma amostra anômala em relação ao conjunto de treinamento por ser suficientemente diferente do conjunto de dados. As amostras anômalas podem ocorrer devido a um erro nas medidas, ou devido ao surgimento de uma ou mais amostras que sofreram um fenômeno químico anômalo ou desconhecido [18].

O resultado da classificação *SIMCA* fornecido pelo programa *Unscrambler*® apresenta, como configuração padrão, uma tabela de asteriscos, onde a presença dos mesmos representa que a amostra avaliada está em conformidade com o modelo elaborado, e a ausência representa a não-conformidade [99]. Outra forma de apresentação do resultado por este programa computacional é pela ilustração gráfica, seja pelo gráfico de *Coomans*, pelo gráfico *Si* versus *Hi*, ou pelo gráfico *Si/SO* versus *Hi*. O gráfico de *Coomans* [99] mostra as distâncias ortogonais dos novos objetos para dois modelos diferentes ao mesmo tempo. Ele estabelece os limites de sociedade da classe para ambos os modelos, de modo que se pode ver se uma amostra é susceptível de pertencer a uma classe, a ambas, ou a nenhuma. O gráfico *Si* versus *Hi* [99] mostra os limites utilizados para a classificação, usando o desvio-padrão residual (*Si*) e a distância da nova amostra ao centro do modelo (*Hi*). O gráfico *Si/SO* versus *Hi* [99] mostra o mesmo que o *Si* versus *Hi*, mas as distâncias são medidas em relação à distância média do modelo.

O *SIMCA* é um método bem estabelecido na literatura e é largamente utilizado para classificação de amostras em conjuntos de dados com alta dimensionalidade. Tem sido aplicado com sucesso em diferentes matrizes, incluindo: medicamentos [114], alimentos [115], cosméticos [116], entre outras [117].

Ao se comparar o método *PCA* com o *SIMCA*, pode-se dizer que a *PCA* simplesmente define um sistema de coordenadas no qual as amostras estão representadas, sendo esta utilizada apenas para análise exploratória dos dados, e não para classificação. Já o *SIMCA* utiliza os eixos definidos por estes sistemas de coordenadas, estabelecendo fronteiras sendo usado justamente para classificar amostras. O uso destas técnicas quimiométricas de reconhecimento de padrões permite realizar uma análise *Screening* dos medicamentos analisados, baseadas na busca de semelhanças, ou agrupamentos de amostras, a partir da similaridade entre os espectros *NIR*.

### **1.7.2.i Análise Screening**

A análise *Screening* caracteriza-se como um método que envolve a interação do analista em uma ou mais etapas de um processo analítico auxiliando nas decisões (sim ou não, presente ou ausente, positivo ou negativo, continue ou pare, etc.)<sup>[118-120]</sup>. É uma modalidade que vem ganhando grande destaque na Química Analítica por poder ser incluída dentro do contexto de análise prévia de amostras<sup>[121-123]</sup>. O termo análise *Screening* incorpora um conjunto de interesses analíticos, não sendo possível encontrar na literatura uma definição inequívoca. Contudo, esse termo pode ser genericamente entendido como: um processo preliminar de análise condicionante da escolha e aplicação de rotas ou decisões posteriores<sup>[118-120,123-125]</sup>.

A análise *Screening* ganhou importância devido ao aumento da demanda de análises que operam com redução de custos operacionais, e da necessidade de se reduzir o descarte de reagentes que contaminam significativamente o meio ambiente<sup>[121-123]</sup>. Esta modalidade auxilia a implementação de métodos modernos que forneçam resultados rápidos, que evitem procedimentos analíticos desnecessários e reduzam ao mínimo o número de determinações qualitativas e quantitativas.

## **1.8 Calibração multivariada**

Os métodos de análise multivariada são assim chamados, pois, em caso do emprego das técnicas espectroscópicas no infravermelho, é possível manipular dados de absorvância espectral associados a mais de uma frequência ao mesmo tempo. Estes métodos têm recentemente tornado possível modelar propriedades químicas e físicas de sistemas simples e complexos a partir de seus dados espectroscópicos<sup>[15]</sup>. As análises qualitativas e quantitativas de medicamentos, utilizando espectroscopia na região do infravermelho, expandiram-se a partir do momento em que os dados gerados por um espectrofotômetro FT-IR puderam ser digitalizados, habilitando os métodos estatísticos na resolução de problemas da análise química<sup>[15]</sup>.

---

---

A possibilidade de utilizar várias frequências do espectro tem aumentado o tipo de amostras que podem ser analisadas por espectroscopia no infravermelho<sup>[15,126]</sup>.

O processo de calibração multivariada baseado em dados espectrais consiste em construir um modelo matemático que relacione as medidas de absorvância, ao longo dos espectros, com os parâmetros de interesse (concentrações das espécies em análise, acidez, viscosidade, índice de refração, etc.) nas amostras<sup>[72,92,127]</sup>.

O processo de calibração multivariada é constituído basicamente de três fases: a calibração, a validação e a predição. O objetivo da fase de calibração é produzir um modelo matemático para relacionar a saída do sinal de um instrumento às propriedades das amostras, quer seja relacionar os dados espectrais com os valores de concentração obtidos por um método de referência<sup>[100]</sup>. Na fase de calibração são construídos os modelos matemáticos e na fase de validação estes modelos são validados usando ferramentas estatísticas de diagnóstico (como: o gráfico dos pesos, o gráfico dos resíduos, o gráfico de influência, o teste F, etc...). Na fase de predição, os modelos matemáticos validados são usados para prever as concentrações ou propriedades de amostras desconhecidas, porém de mesma natureza das amostras dos conjuntos de calibração e validação.

Durante a construção dos modelos de calibração multivariada, duas metodologias podem ser utilizadas: A validação cruzada e a série de teste<sup>[18]</sup>. Na metodologia série de teste usa-se três conjuntos de amostras separadamente, sendo um para calibração, um para a validação e um para a predição do modelo. A escolha da quantidade de amostras que irá compor cada conjunto não é algo pré-definido, podendo-se usar, por exemplo, 50% do banco de dados das amostras para calibrar o modelo, 25% para validação e 25% para predição<sup>[18]</sup>. Já na metodologia de validação cruzada, usa-se dois conjuntos de amostras, sendo um conjunto para calibração e validação do modelo, e um conjunto para predição. Durante a construção do modelo uma amostra é escolhida de forma aleatória e separada do conjunto de calibração e usada para validar o modelo. Esta operação se processa de forma que todas as amostras do conjunto sejam testadas e o melhor resultado seja escolhido. Para esta metodologia também não há uma especificação da quantidade de amostras que se deve utilizar para compor cada conjunto, podendo-se usar, por exemplo, 70% do banco de dados das amostras para calibrar e validar o modelo e 30% para predição<sup>[18]</sup>.

Na construção de modelos de calibração multivariada, os dados espectrais de natureza multivariada (**Figura 6**) são representados numa matriz **X**, de dimensão ( $n \times m$ ), onde  $n$  representa o número de amostras cujos espectros foram medidos e  $m$  representa o número de variáveis ou de comprimentos de ondas em que as absorvâncias foram medidas. Os parâmetros de interesse são

representados pela matriz  $\mathbf{Y}$  de dimensão  $(n \times p)$ , onde  $p$  são os parâmetros a serem estimados pelos modelos construídos<sup>[92]</sup>.

$$\begin{array}{c}
 \mathbf{X} = \left[ \begin{array}{cccc}
 X_{11} & X_{12} & \cdots & X_{1m} \\
 X_{21} & X_{22} & \cdots & X_{2m} \\
 \vdots & \vdots & & \vdots \\
 X_{n1} & X_{n2} & \cdots & X_{nm}
 \end{array} \right] \quad \mathbf{Y} = \left[ \begin{array}{cccc}
 Y_{11} & Y_{12} & \cdots & Y_{1p} \\
 Y_{21} & Y_{22} & \cdots & Y_{2p} \\
 \vdots & \vdots & & \vdots \\
 Y_{n1} & Y_{n2} & \cdots & Y_{np}
 \end{array} \right] \\
 \text{Variáveis independentes} \qquad \qquad \qquad \text{Variáveis dependentes}
 \end{array}$$

Figura 6. Representação de matrizes de dados ( $\mathbf{X}$  e  $\mathbf{Y}$ ).

Os espectros *NIR* obtidos fornecem valores quantitativos dos componentes da amostra analisada. Tais informações químicas precisam ser ajustadas a modelos matemáticos que permitam fazer previsões de casos não experimentados na prática<sup>[100]</sup>, e para isso são empregadas técnicas estatísticas multivariadas. Dentre os métodos mais empregados na calibração multivariada podemos citar a regressão por mínimos quadrados parciais - *PLS* (sessão 1.8.1), e a regressão linear múltipla - *MLR* (sessão 1.8.2).

Para a construção de bons modelos de calibração multivariada, um importante problema deve ser levado em consideração: a seleção de um subconjunto representativo em um conjunto grande de amostras<sup>[128-131]</sup>. Como a construção dos modelos deve ser feita de forma a se obter modelos precisos e parcimoniosos, é necessário escolher as amostras mais significativas de cada conjunto. Usam-se, geralmente, algoritmos para a seleção de amostras (calibração, validação e predição). Dois algoritmos se tornaram de grande valia na seleção desses conjuntos, o Algoritmo *Kennard-Stone* (*KS*)<sup>[132]</sup> e o Algoritmo de partição do conjunto de amostras baseado nas distancias conjuntas  $\mathbf{X}$ - $\mathbf{Y}$  (*SPXY*)<sup>[133]</sup>.

O Algoritmo *KS*<sup>[132]</sup> busca 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 algoritmo *KS* segue o procedimento por etapas de modo que uma nova seleção é feita nas regiões do espaço longe das amostras já selecionadas<sup>[132]</sup>. Para este propósito, o algoritmo emprega a distância Euclidiana  $d_x(a,b)$  entre os vetores  $\mathbf{X}$  de cada par  $(a,b)$  de amostras, como mostrado na **Equação 3**:

$$d_x(a,b) = \sqrt{\sum_{j=1}^J [X_a(j) - X_b(j)]^2}; \quad a,b \in [1,N] \quad (3)$$

Para dados espectrais,  $X_a(j)$  a  $X_b(j)$  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, subseqüentemente, o algoritmo seleciona a amostra que exhibe a maior distância em relação a uma amostra já selecionada. O procedimento é repetido até que o número de amostras requerido seja encontrado<sup>[132]</sup>.

O algoritmo *SPXY*<sup>[133]</sup>, é uma extensão do algoritmo *KS* que leva em consideração tanto as diferenças de  $\mathbf{X}$  quanto as de  $\mathbf{Y}$  no cálculo das distâncias inter-amostras.

Este algoritmo calcula cada distância  $d_x(a, b)$  (Equação 3) e cada distância  $d_y(a, b)$  (Equação 4), soma as distâncias de  $\mathbf{X}$  com  $\mathbf{Y}$  e normaliza dividindo  $d_x(a, b)$  e  $d_y(a, b)$  pelos seus valores máximos no conjunto de dados (Equação 5).

$$d_y(a, b) = \sqrt{(y_a - y_b)^2} = |y_a - y_b|; a, b \in [1, N] \quad (4)$$

$$d_{xy}(a, b) = \frac{d_x(a, b)}{\max_{a, b \in [1, N]} d_x(a, b)} + \frac{d_y(a, b)}{\max_{a, b \in [1, N]} d_y(a, b)}; a, b \in [1, N] \quad (5)$$

### 1.8.1 PLS

A regressão *PLS* é um método de análise de fatores, introduzida por H.Wold<sup>[134]</sup>, que busca estabelecer uma relação entre a matriz de dados de uma dada resposta instrumental, como os espectros *NIR* (matriz  $\mathbf{X}$ ), com a matriz de dados de uma informação de interesse, como a concentração de um determinado analito (matriz  $\mathbf{Y}$ )<sup>[135-137]</sup>. Ambas as matrizes são decompostas em novas variáveis denominadas variáveis latentes, fatores ou componentes principais, a partir de duas matrizes de variações: Pesos e escores. Os espectros originais podem ser considerados como combinações lineares das variações dos espectros (pesos) onde os escores representam a contribuição de cada espectro original<sup>[100]</sup>. O modelo final consiste, basicamente, em se encontrar a melhor relação linear entre os escores das duas matrizes.

De uma forma geral, para se entender o funcionamento do *PLS*<sup>[138]</sup>, consideremos a Equação 6:

$$\mathbf{Y} = \mathbf{X} \mathbf{b} \quad (6)$$

onde  $\mathbf{Y}$  é a propriedade de interesse (concentração, por exemplo),  $\mathbf{b}$  é o vetor de regressão e  $\mathbf{X}$  refere-se aos espectros das amostras representadas por uma matriz onde as colunas são os valores de absorvância para cada comprimento de onda e as linhas referem-se às amostras.

A dificuldade de encontrar o modelo para determinar a propriedade de interesse está em encontrar o valor de  $\mathbf{b}$ , uma vez que a matriz  $\mathbf{X}^T\mathbf{X}$  não é inversível devido à redundância das variáveis. A regressão de componentes principais elimina esta redundância pela construção de uma nova matriz  $\mathbf{U}$  com colunas que são combinações lineares das colunas originais em  $\mathbf{X}$ <sup>[138]</sup>. Esta matriz tem o mesmo número de linhas que  $\mathbf{X}$ , mas um número menor de colunas. Usando a matriz  $\mathbf{U}$ , um novo modelo pode ser escrito (**Equação 7**).

$$\mathbf{y} = \mathbf{U}\mathbf{b} \quad (7)$$

Várias maneiras podem ser utilizadas para construir a matriz  $\mathbf{U}$ . A análise de componentes principais (*PCA*)<sup>[139]</sup> é um eficiente método para encontrar uma combinação linear de variáveis. A matriz  $\mathbf{U}$  é a matriz de escores do *PCA*, a qual define a localização das amostras relativas umas as outras no vetor espaço. Essa matriz de escores pode ser encontrada pela decomposição *SVD*:

$$\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^T \quad (8)$$

onde  $\mathbf{V}$  é a matriz contendo os pesos e  $\mathbf{S}$  é a matriz diagonal com os valores singulares.

Para determinar  $\mathbf{U}$ , pode-se então usar a propriedade ortonormal:

$$\mathbf{U} = \mathbf{R}\mathbf{V}\mathbf{S}^{-1} \quad (9)$$

No PLS, a covariância das medidas com as concentrações são usadas em adição à variância em  $\mathbf{X}$  para gerar  $\mathbf{U}$ . É um processo iterativo, no qual se otimiza ao mesmo tempo a projeção das amostras sobre os pesos para a determinação dos escores e o ajuste por uma função linear dos escores da matriz  $\mathbf{X}$  aos escores de  $\mathbf{y}$  (lembrando que  $\mathbf{y}$  é um vetor de respostas da propriedade de interesse obtida por metodologia padrão, de modo a minimizar os desvios)<sup>[138]</sup>.

A partir da matriz  $\mathbf{U}_{\text{PLS}}$ , é possível resolver a **Equação 10** para o vetor de regressão  $\mathbf{b}$ :

$$\mathbf{y} = \mathbf{U}_{\text{PLS}}\mathbf{b}_{\text{prev}} \quad (10)$$

Construído o modelo de calibração, este é usado para fazer a estimativa do parâmetro em novas amostras. Os procedimentos de seleção do número ótimo de fatores e de validação dos modelos são interligados, sendo o número ótimo de componentes àquele que resultar num modelo com melhor capacidade preditiva<sup>[138]</sup>.

O algoritmo dos mínimos quadrados parciais iterativos não-lineares (*NIPALS*) é um algoritmo eficiente para extrair os vetores de pesos a partir dos espectros, na ordem decrescente de suas contribuições à variância nos espectros de calibração<sup>[140]</sup>, é o adotado pelo programa *Unscrambler*® para realizar este cálculo<sup>[99]</sup>.

Dentre os vários métodos de calibração multivariada, o *PLS* é particularmente útil quando nem todas as substâncias espectroscopicamente ativas são conhecidas e o número de variáveis é grande<sup>[127]</sup>.

A aplicação da quimiometria apresentou um avanço significativo com a utilização do método *PLS*. Com o uso deste método, muitos problemas de cálculo puderam ser superados e variadas aplicações analíticas começaram a surgir<sup>[141]</sup>.

### 1.8.2 MLR

*MLR*<sup>[142]</sup> é uma técnica que busca estabelecer uma relação linear entre o sinal analítico e a concentração aplicando o método dos mínimos quadrados. O modelo *MLR* pode ser obtido a partir de uma matriz  $\mathbf{X}$  de respostas instrumentais, e dos dados de um vetor  $\mathbf{y}$  que contém as concentrações (ou outra propriedade) obtidas por um método de referência das amostras. Cada variável dependente de  $\mathbf{y}$  é expressa como uma combinação linear das variáveis independentes da matriz  $\mathbf{X}$  por intermédio da expressão:

$$\mathbf{y}=\mathbf{X}\mathbf{b} \quad (11)$$

onde o vetor  $\mathbf{b}$  contém os coeficientes da regressão, que são calculados por mínimos quadrados a partir da pseudoinversa de  $\mathbf{X}$ :

$$\mathbf{b}=(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \quad (12)$$

onde -1 e T representam a inversão e transposição da matriz, respectivamente.

Com o modelo determinado, as concentrações de novas amostras podem ser estimadas a partir da seguinte equação:

$$\mathbf{y}_{\text{prev}} = \mathbf{X}^*\mathbf{b} \quad (13)$$

onde  $\mathbf{X}^*$  representa a matriz de dados para as novas amostras.

Os modelos *MLR* são mais simples e mais fáceis de interpretar que os modelos *PLS*, porque os modelos *PLS* executam uma regressão em variáveis latentes que em geral, não têm significado físico. Por outro lado, para a construção de modelos *MLR* o número de amostras deve ser igual ou superior ao número de variáveis<sup>[18]</sup>. Uma vez que o modelo consiste na resolução de um sistema de equações lineares simultâneas, essa condição tem que ser satisfeita. Caso contrário, o sistema torna-se indeterminado. Um problema importante em calibração *MLR* é que a matriz  $(\mathbf{X}^T\mathbf{X})$  pode não ser inversível ou promover a propagação de erros quando existir forte correlação ou multicolinearidade entre as variáveis<sup>[137]</sup>. Uma vez que, a inclusão de variáveis redundantes pode ter vários efeitos potencialmente danosos, ainda que as variáveis adicionais não influenciem diretamente o desempenho do modelo<sup>[18]</sup>. Em função disso, os modelos *MLR* são mais dependentes de uma boa escolha das variáveis espectrais. A seleção dos comprimentos de onda que resultem na máxima precisão ainda é uma tarefa desafiadora, principalmente quando os espectros possuem uma forte sobreposição e têm características distintivas imperceptíveis, como é o caso da espectroscopia UV-Vis.

Várias técnicas de seleção de variáveis têm sido propostas na literatura para selecionar conjuntos de variáveis não redundantes e mais informativas para calibração multivariada, e os algoritmos adotados por nosso grupo de trabalho são: Algoritmo da Busca Angular (*ASA*)<sup>[143]</sup>, Algoritmo Genético (*GA*)<sup>[145]</sup>, Algoritmo das Projeções Sucessivas (*SPA*)<sup>[149]</sup>, Algoritmo *Stepwise* (*SW*)<sup>[157]</sup>.

### 1.8.2.1 *ASA*

O *ASA*<sup>[143]</sup> é uma técnica que busca minimizar a correlação e a multicolinearidade entre as variáveis da matriz de respostas instrumentais. O *ASA* calcula os cossenos dos ângulos entre os vetores associados às variáveis definidas no espaço das amostras e aplica um procedimento mínimo-máximo para construir cadeias de variáveis minimamente correlacionadas. Num conjunto de dados centrados na média (comum nos métodos de regressão multivariada e também utilizado no *ASA*), o valor do cosseno é, matematicamente, igual a correlação entre duas variáveis. Para remover a colinearidade restante entre as variáveis, o *ASA* descarta as variáveis mais colineares por intermédio do Fator de Inflação da Variância (*VIF*) que fornece uma medida da multicolinearidade entre as variáveis, selecionando as que apresentem valores abaixo de um determinado limiar<sup>[143,144]</sup>. A combinação *ASA-VIF* possibilita a seleção de um conjunto de variáveis minimamente correlacionadas e sem multicolinearidade, permitindo a seleção das variáveis mais informativas

---

---

com relação à matriz  $\mathbf{Y}$  por intermédio da menor raiz quadrada do erro médio quadrático de previsão –  $RMSEP$ , para um conjunto independente de amostras.

### 1.8.2.2 GA

O GA<sup>[145]</sup>, muito utilizado em calibração multivariada, é um algoritmo que utiliza operadores matemáticos para simular processos da evolução, inspirados em mecanismos de seleção natural proposta por Charles Darwin. Este algoritmo modela uma solução para um problema específico em uma estrutura de dados como a de um cromossomo e aplica operadores que re-combinam estas estruturas preservando informações críticas, de modo que selecione variáveis aleatoriamente, mantendo-se menos susceptível a soluções locais<sup>[146-148]</sup>.

Numa utilização mais abrangente do termo, um algoritmo genético é qualquer modelo baseado em população que utiliza operadores de seleção e re-combinação para gerar novos pontos amostrais em um espaço de busca<sup>[146-148]</sup>.

Uma implementação do algoritmo genético começa com uma população (geralmente randômica) de cromossomos. Estas estruturas são então avaliadas para gerar oportunidades reprodutivas de forma que, cromossomos que representam uma solução "melhor" tenham maiores chances de se reproduzirem do que os que representam uma solução "pior". A definição de uma solução melhor ou pior é tipicamente relacionada à população atual<sup>[145-148]</sup>. O uso de um gerador randômico para criar uma população inicial de cromossomos antes da otimização serve para evitar influência tendenciosa. A evolução é realizada por meio de gerações. A cada geração, a adaptação de cada solução na população é avaliada, e alguns indivíduos são selecionados para a próxima geração, podendo ainda ser recombinados ou mutados para formar uma nova população de modo a melhorar gradativamente a adaptação. A função de adaptação é avaliada mediante a construção de modelos  $MLR$ , cujo valor é calculado como o inverso do  $RMSEP$  obtido<sup>[145]</sup>. A nova população então é utilizada como entrada para a próxima iteração do algoritmo.

Muitos algoritmos genéticos foram introduzidos por pesquisadores de uma perspectiva experimental, e destes, a maior parte tinham interesse de utilizar o algoritmo genético como uma ferramenta de otimização<sup>[146-148]</sup>.

Entretanto, muito embora o GA realize uma procura aleatória, devido à sua natureza estocástica, normalmente os resultados das seleções das variáveis dependentes não são reprodutíveis.

### 1.8.2.3 SPA

O SPA<sup>[149]</sup> é uma técnica de seleção de variáveis que utiliza operações simples num espaço vetorial para minimizar problemas de colinearidade. Este algoritmo mostrou-se eficiente para seleção de variáveis espectrais, no contexto da calibração multivariada, especificamente quando aplicado ao *MLR*.

O algoritmo *SPA* compreende basicamente três etapas. Inicialmente, o algoritmo seleciona subconjuntos de variáveis com base no critério de minimização da multicolinearidade. Esses subconjuntos são obtidos de acordo com uma seqüência de operações de projeções aplicadas nas colunas da matriz de calibração<sup>[149]</sup>. Posteriormente, o melhor subconjunto é escolhido de acordo com um critério que avalia a habilidade de predição de um modelo *MLR* a partir de uma série de testes, tal como o *PRESS*. A estatística *PRESS*<sup>[137]</sup> é uma medida de validação obtida pela eliminação de cada observação, uma por vez, e prevendo-se esse valor dependente com o modelo de regressão estimado a partir das demais observações. Por fim, o subconjunto escolhido é submetido a um procedimento de eliminação para determinar se alguma variável poderá ser removida sem perda significativa da capacidade de predição, e para isso calcula-se o *RMSEP*<sup>[149]</sup>.

Após a apresentação inicial do *SPA*, muitos trabalhos científicos foram publicados com diferentes modificações e aplicações, incluindo UV-Vis<sup>[149,150]</sup>, *ICP-OES*<sup>[151]</sup>, espectrometria *NIR*<sup>[152-154]</sup>, *QSAR*<sup>[155]</sup> e classificação<sup>[156]</sup>.

### 1.8.2.4 SW

O algoritmo *Stepwise*<sup>[157]</sup> é um método para seleção, que determina dentre um universo de potenciais variáveis explicativas, uma combinação que deve compor o modelo a ser utilizado. Esta técnica remonta à década de 1970<sup>[158]</sup>. Nele, a importância que cada variável independente apresenta dentro de um dado modelo é investigada. Para isso, as variáveis são excluídas ou adicionadas ao modelo de acordo com algum critério pré-estabelecido<sup>[159]</sup>. Algumas versões ligeiramente diferentes deste algoritmo podem ser encontradas na literatura<sup>[157-162]</sup>.

O método *Stepwise* determina que variáveis ou quais variáveis dentre o conjunto de variáveis independentes especificadas serão realmente usadas para a regressão, e em qual ordem elas serão introduzidas, começando com as variáveis forçadas, e continuando com as outras variáveis e conjuntos de variáveis, uma a uma<sup>[157]</sup>. Posteriormente, o algoritmo seleciona, das variáveis preditoras remanescentes, a variável ou conjunto de variáveis que produz a maior redução na variância dos resíduos (não-explicada) da variável dependente, a menos que a sua contribuição ao *F*-calculado total para a regressão permaneça abaixo de um nível especificado. O algoritmo

---

---

avalia depois de cada passo se a contribuição de qualquer variável ou conjunto de variáveis já inclusas cai abaixo de um nível especificado, senão estas serão eliminadas da regressão<sup>[157]</sup>.

Em outras palavras, o algoritmo *Stepwise* procura, passo a passo, uma variável independente não presente na equação de regressão que possui a menor probabilidade de F. As variáveis já na regressão são removidas se a sua probabilidade de F tornar-se muito grande. O algoritmo termina quando não há mais variáveis elegíveis para inclusão ou remoção<sup>[161,162]</sup>.

## ***1.9 Objetivo***

Este trabalho tem como objetivo propor diferentes modelos quimiométricos aplicando seleção de variáveis e calibração multivariada a espectros *NIR*, para a determinação do teor do principio ativo de um medicamento injetável diretamente em seus recipientes, sem violá-los.

### ***1.9.1 Objetivos específicos***

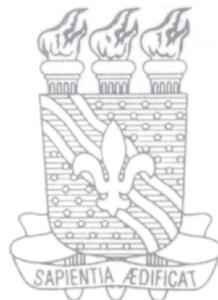
Para que os objetivos fossem atingidos foi necessário:

- Adquirir amostras comerciais de um medicamento injetável utilizado na clínica médico/hospitalar (o medicamento escolhido foi a dipirona sódica injetável);
- Desenvolver uma técnica para aquisição de espectros *NIR* diretamente nas ampolas dos medicamentos injetáveis;
- Elaborar modelos quimiométricos de calibração multivariada a partir dos espectros *NIR* registrados;
- Aplicar os modelos de calibração multivariada em espectros *NIR* de amostras comerciais do medicamento injetável e verificar a capacidade preditiva de cada modelo.

# CAPÍTULO II

## EXPERIMENTAL

---



## 2. EXPERIMENTAL

### 2.1 Metodologia e estratégia de ação

A seguir serão apresentadas informações acerca do método utilizado no desenvolvimento deste trabalho.

#### 2.1.1 Obtenção das amostras e preparação das soluções padrão

O fármaco escolhido para as análises foi a dipirona sódica injetável 1g/2mL (500 mg/mL), comercializada em frasco ampola de 2 mL de coloração âmbar. A opção por esse medicamento se deu pelo fato de se tratar de um medicamento comercial de alto consumo na clínica médico/hospitalar, devido às suas propriedades intrínsecas, tais como: alta eficácia no tratamento de dores agudas, fracas e medianas, baixo custo, ampla e ubíqua disponibilidade, e boa margem de segurança. Além disso, esse medicamento é intensamente comercializado no mercado brasileiro<sup>[41]</sup>.

Em função da extrema dificuldade de se obter uma grande quantidade de amostras de hospitais procedentes de um único laboratório farmacêutico, optou-se pela obtenção das amostras diretamente de uma indústria farmacêutica, dado que neste caso, também teríamos em mãos medicamentos recém produzidos e dentro dos padrões de qualidade requeridos. Assim, foram gentilmente fornecidas pelo fabricante Isofarma Industrial Farmacêutica, com sede na cidade de Euzébio, Estado do Ceará, 64 amostras (8 amostras de 8 lotes diferentes), e seus respectivos laudos. A **Tabela 2** apresenta o número de amostras (**NºA**), com suas respectivas concentrações de dipirona sódica injetável (**CDSI**) determinadas segundo o método de referência oficial, os códigos dos lotes (**CL**), e mês/ano de validade (**M/A**) de acordo com os laudos fornecidos pelo fabricante.

**Tabela 2** – Número de amostras com suas respectivas concentrações, lotes e mês/ano de validade.

Nº A	CDSI (% m/v)	CL	M/A	Nº A	CDSI (% m/v)	CL	M/A
8	50,27	HR541	Dez/2008	8	52,72	HR545	Dez/2008
8	51,67	HR542	Dez/2008	8	50,97	HR546	Dez/2008
8	52,90	HR543	Dez/2008	8	52,02	HR549	Dez/2008
8	49,92	HR544	Dez/2008	8	50,98	HR550	Dez/2008

Para a construção dos modelos quimiométricos, soluções padrão foram preparadas com o princípio ativo: dipirona sódica, de grau analítico (99,2% m/m) adquirido em farmácias de manipulação de João Pessoa-PB.

As soluções padrão foram preparadas por simples diluição do princípio ativo em água de injeção esterilizada (fornecidas pelo mesmo fabricante) nas concentrações de 0,0 (contendo somente água para injetáveis), 10,0; 20,0; 24,0; 28,0; 32,0; 36,0; 40,0; 44,0; 44,5; 45,0; 46,0; 48,0; 50,0; 52,0; 54,0; 54,5; 55,0; 55,5; 56,0 e 60,0% (m/v), totalizando 21 padrões. Após o preparo, as soluções foram acondicionadas em frascos de vidro de coloração âmbar.

Água para injetáveis é uma água estéril, obtida por destilação, em equipamentos cujas partes em contato com a água são de vidro neutro, quartzo ou metal apropriado, podendo ainda, ser obtida por processo equivalente ou superior à destilação na remoção de contaminantes químicos ou microorganismos<sup>[44]</sup>.

As concentrações de 45,0 a 55,0 % (m/v) englobam a faixa de variação das concentrações de dipirona sódica em medicamentos injetáveis aceita pelos órgãos fiscalizadores, que neste caso é de  $\pm 10,0\%$  do valor nominal do princípio ativo do medicamento injetável, que é 50,0% (m/v).

Não foi possível preparar padrões em concentrações superiores a 60,0% (m/v) devido à limitação da solubilidade do princípio ativo.

### 2.1.2 Instrumentação

Para o registro dos espectros das amostras do medicamento injetável dipirona sódica foi utilizado o espectrômetro de Infravermelho com Transformada de Fourier (*FT-IR*) da marca *Perkin Elmer*, modelo *Spectrum*, série *GX*, mostrado na **Figura 7**.



**Figura 7** – Espectrômetro Perkin Elmer *FT-IR* modelo *GX*.

### 2.1.3 Confeção dos suportes para os medicamentos

Para o procedimento das análises *NIR*, era necessário o correto encaixe das ampolas no caminho óptico (fixo) do feixe de radiação eletromagnética para minimizar os problemas de espalhamento e para garantir uma boa reprodutibilidade das medidas. Para isso, tornou-se necessário a construção de suportes (um para cada classe de ampolas, por exemplo, ampolas de: 1, 3

ou 5 mL), confeccionados em teflon afixados em bases de alumínio. Esses suportes se adaptaram perfeitamente ao compartimento de análises do espectrômetro *NIR* (**Figura 8**).



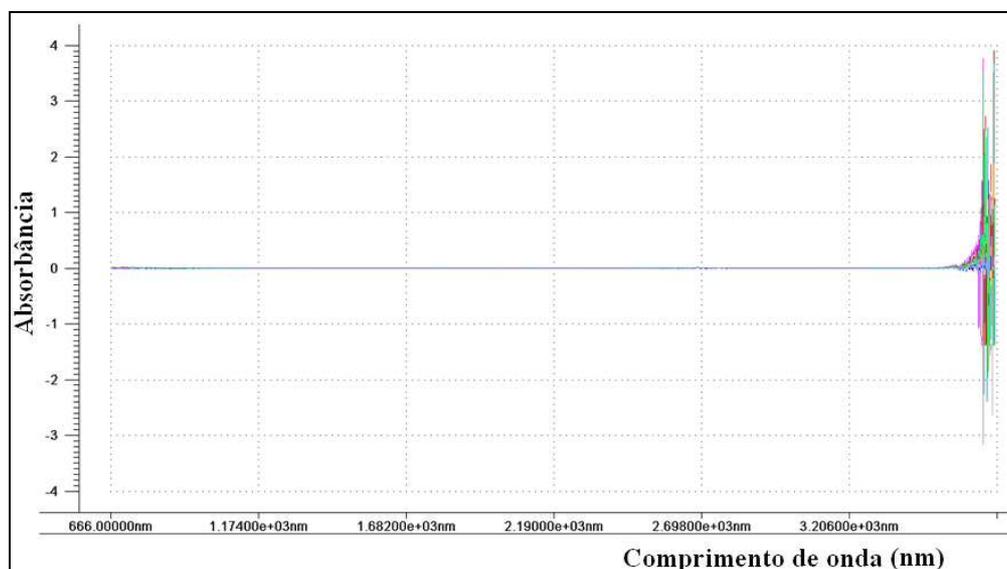
**Figura 8** – Suporte das ampolas de medicamentos adaptado ao compartimento de medidas do espectrômetro *NIR*.

Vale ressaltar que, os suportes foram construídos de tal forma que as rotulações das ampolas não alcançavam o feixe da radiação do instrumento.

#### *2.1.4 Estudos físicos das ampolas de vidro*

Foram realizados testes para avaliar a interação das paredes de vidro da ampola com a radiação eletromagnética na região do *NIR*.

Para os testes, utilizou-se uma ampola vazia, a qual foi previamente marcada em uma extremidade e rotacionada em 360°, para registro dos espectros *NIR* a cada 20°. Conforme se pode verificar na **Figura 9**, nenhuma absorção foi detectada, indicando a não necessidade de se fixar a ampola em uma determinada direção para efetuar as análises.



**Figura 9** – Estudo de absorção da radiação *NIR* pelas ampolas de vidro.

### 2.1.5 Aquisição dos espectros

Quatro ampolas foram previamente abertas, esvaziadas e limpas, para serem usadas como cubetas nas devidas análises. O tamanho do caminho óptico percorrido pelo feixe luminoso foi de aproximadamente 1 cm.

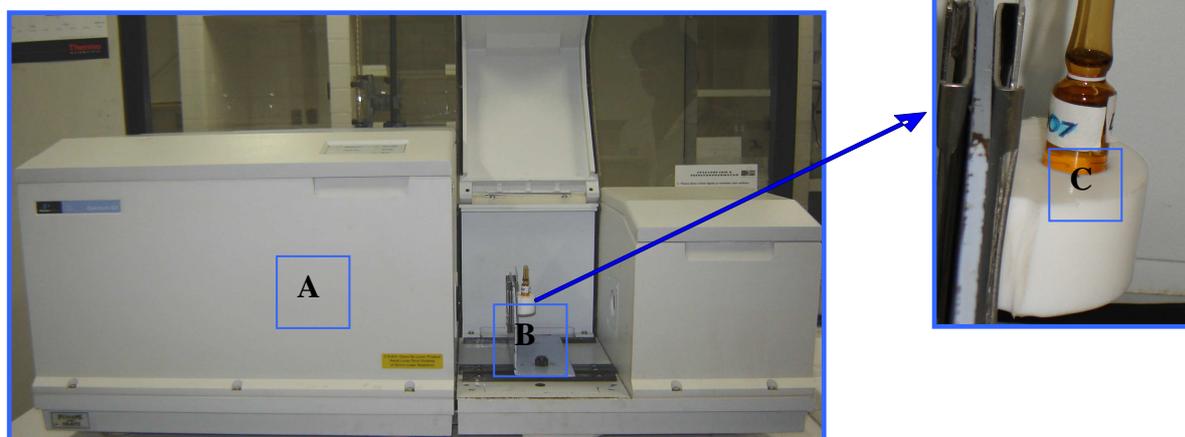
Para o registro do branco, utilizou-se as ampolas vazias reservadas, preenchidas com 2 mL de água para injetáveis.

Para o registro das soluções padrão, utilizou-se as quatro ampolas vazias reservadas, preenchidas com 2mL de solução, para cada amostra padrão. Como se preparou 21 soluções padrão, ao se registrar 4 amostragens de cada padrão, formou-se um conjunto de 84 amostras. Todos os registros dos espectros foram efetuados com três réplicas para cada amostra, totalizando 312 espectros de soluções padrão.

Os espectros *NIR* das amostras comerciais foram registrados incidindo a radiação infravermelha diretamente sobre a ampola de vidro do medicamento injetável devidamente fechada. Os registros dos espectros das amostras comerciais também foram efetuados em triplicata para cada amostra, totalizando 192 espectros.

A totalidade de espectros registrados para serem usados na construção dos modelos, entre amostras comerciais e soluções padrão, foi de 504 espectros. O registro de todas as amostras foi arquivado em um microcomputador para posterior tratamento quimiométrico.

A **Figura 10** mostra os detalhes para o registro dos espectros *NIR* dos medicamentos injetáveis sem a violação da ampola.



**Figura 10** – Sistema para registro dos espectros *NIR* dos medicamentos injetáveis sem violação da ampola. **A:** espectrômetro *NIR*; **B:** suporte das ampolas acoplado ao suporte do espectrofotômetro; **C:** amostra de dipirona sódica injetável afixada em seu suporte.

---

---

Todos os espectros foram registrados na faixa espectral compreendida entre 666 a 2500 nm. Estes espectros foram utilizados para a análise *Screening* e para a construção, validação e predição dos modelos de calibração multivariada.

Todas as análises foram efetuadas em ambiente controlado, onde a temperatura média foi de 26°C e umidade relativa média de 54%. Cada espectro *NIR* registrado foi sempre um espectro médio resultante de 16 varreduras, empregando uma resolução de 4 cm<sup>-1</sup>.

### ***2.1.6 Programas estatísticos***

Para realizar o pré-processamento dos dados e aplicar as técnicas quimiométricas de análise de dados adequadas (as técnicas de reconhecimento de padrões *PCA* e *SIMCA*, e de regressão *PLS*), foi utilizado o programa quimiométrico *The Unscrambler*® 9.7 da CAMO SA. Os algoritmos de seleção de amostras (*SPXY*), e de variáveis (*SW*, *GA*, *SPA* e *ASA*) associadas a regressão *MLR*, foram elaborados e executados em ambiente *Matlab*® 6.5. Para a determinação do número ótimo de fatores foi utilizando o programa *The Unscrambler*® 9.7.

# **CAPÍTULO III**

## **RESULTADOS E DISCUSSÃO**

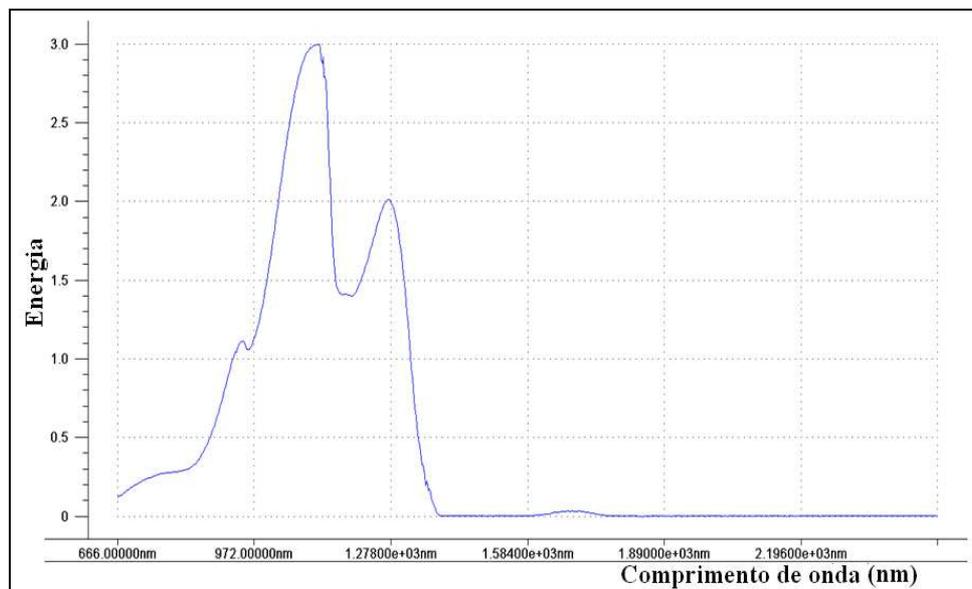
---



### 3. RESULTADOS E DISCUSSÃO

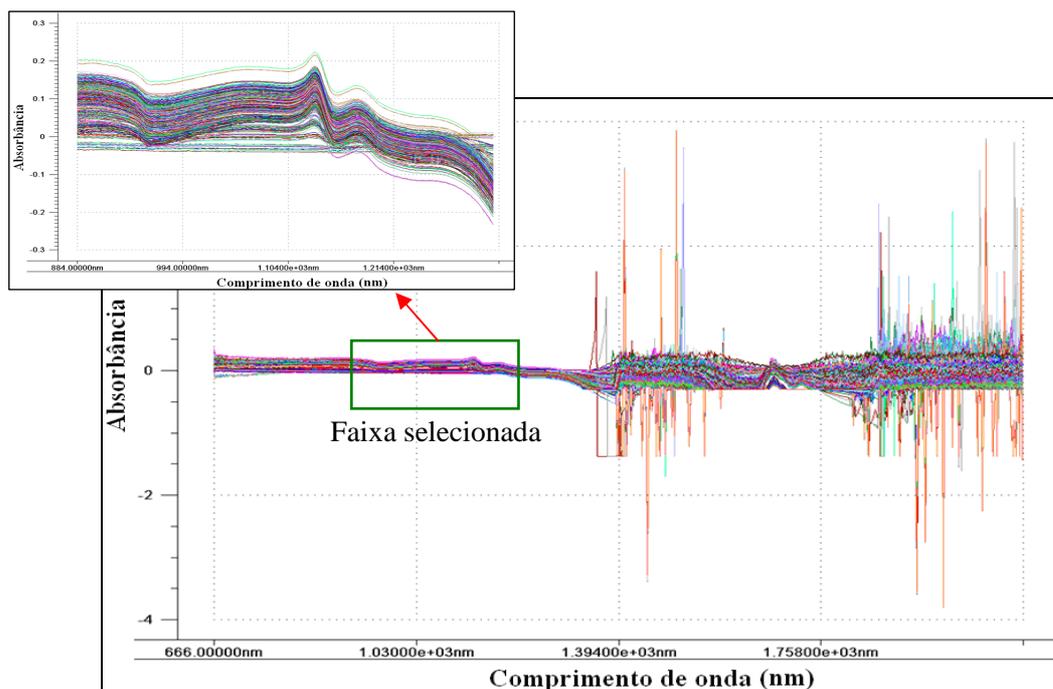
#### 3.1 Seleção da região espectral de trabalho

Para o registro do sinal do branco utilizou-se uma ampola de 2 mL com água para injetáveis. O sinal espectral obtido encontra-se ilustrado na **Figura 11**.



**Figura 11** – Registro do sinal do branco (ampola com água para injetáveis).

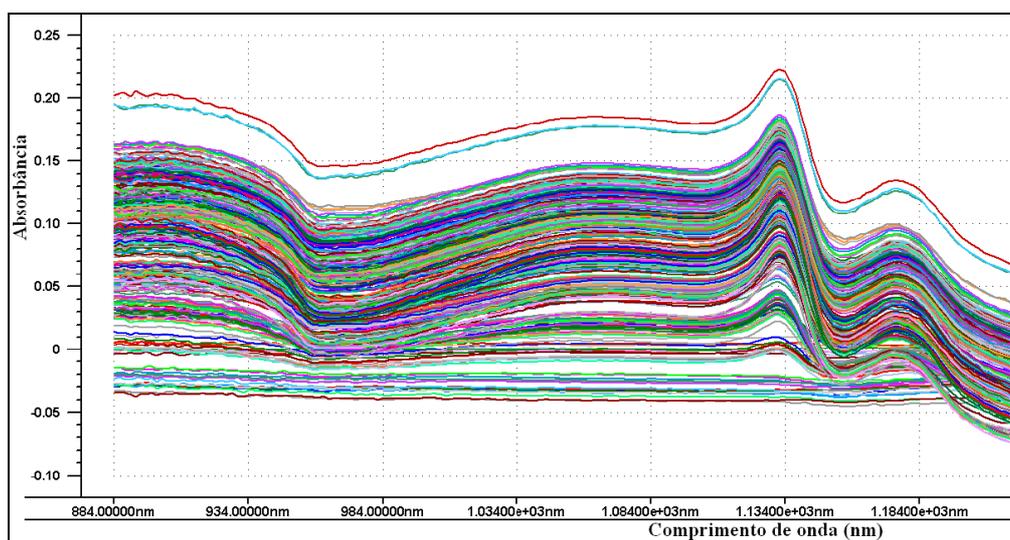
Os espectros *NIR* das amostras do medicamento dipirona sódica injetável comercial e das soluções padrão, na faixa espectral de 600 a 2500 nm, são apresentados na **Figura 12**.



**Figura 12** - Espectros brutos das amostras e dos padrões na região *NIR* entre 666 a 2500 nm.

Na **Figura 12**, observa-se que os valores de absorvância na região compreendida entre 1300 e 2500 nm apresentam grandes saturações, e a região abaixo de 884 nm (666 a 884 nm) encontra-se ruidosa e sem a presença de picos aparentes. Estas regiões foram descartadas *a priori* na definição da região espectral de trabalho.

A região escolhida para análise foi a faixa espectral compreendida entre 884 e 1220 nm, a qual se encontra ilustrada na **Figura 13**. A maior parte desta faixa encontra-se na região do infravermelho próximo de ondas curtas (*Shortwave-NIR - SW-NIR*)<sup>[163,164]</sup>. O *SW-NIR* compreende a região que vai de 800 a 1100 nm ( $12500$  a  $9091$   $\text{cm}^{-1}$ ). A partir desta região encontra-se o infravermelho próximo de ondas longas (*Longwave-NIR - LW-NIR*) onde se insere a faixa de 1100 a 2500 nm ( $9091$  a  $4000$   $\text{cm}^{-1}$ ). Essa região escolhida caracteriza-se pela ocorrência principalmente de sobretons e combinações de transições fundamentais.



**Figura 13** - Espectros brutos das amostras e padrões na região *NIR* entre 884 a 1220 nm.

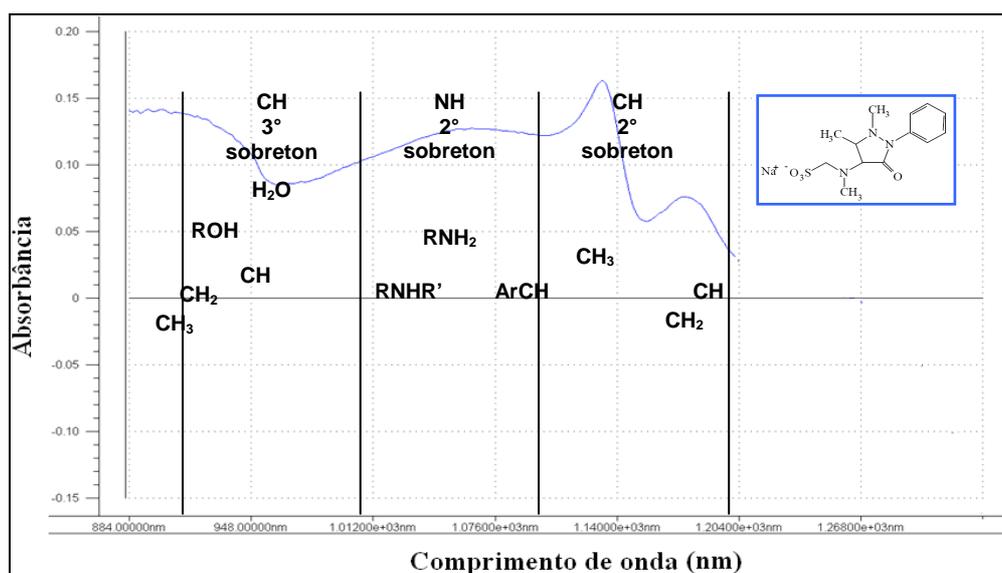
Como se pode observar na **Figura 13**, uma parte do espectro apresenta-se ruidoso, e estes possuem uma grande variação de linha de base entre as amostras. Porém o ruído instrumental é muito comum nessa região espectral, podendo ser facilmente tratado pelo uso de uma técnica de suavização adequada, por exemplo, a suavização Savitzky-Golay<sup>[101]</sup>, que ajusta, pelos mínimos quadrados, um polinômio de baixa ordem aos pontos de uma janela. Neste método de suavização, a escolha do número de pontos utilizado na janela é muito importante, pois um número elevado de pontos pode acarretar perda de informações relevantes, e um número pequeno de pontos acarreta a permanência de ruído nos dados espectrais.

Já o deslocamento de linha de base observado, provavelmente é devido à variação do índice de refração, que neste caso ocorre devido às pequenas variações como: pequenas diferenças na espessura das paredes das ampolas; pequenas variações na transparência da coloração das paredes das ampolas de vidro; além das pequenas diferenças que o feixe luminoso enfrenta em seu percurso, proporcionadas pela mudança do meio (ar/vidro/água). A variação da linha de base também não se torna um problema crítico, uma vez que também existem técnicas que possibilitam sua correção, como por exemplo, os métodos derivativos.

É possível se observar ainda, a presença de algumas absorbâncias negativas na região entre 1195 e 1220 nm. Isso ocorre devido ao fato de a absorbância do sinal do branco ser muito alta em relação à absorbância da amostra. A banda do O-H da água presente no branco ocorre numa região próxima a essa faixa (em aproximadamente 1350 nm), e é tão larga e tão intensa que chega a afetar essa região.

### 3.2 Atribuição das bandas dos espectros NIR

A região espectral compreendida entre, aproximadamente 700 e 1200 nm corresponde à região de segundo sobreton do O-H, terceiro sobreton do N-H e quarto sobreton do C-H. Pode-se verificar na **Figura 14**, que o intervalo espectral que se encontra aproximadamente entre 884 e 1000 nm apresenta uma absorção relativa à combinação de transições fundamentais dos grupos C-H<sub>3</sub>, C-H<sub>2</sub>, C-H e R-OH pertencentes às moléculas de dipirona presentes na amostra <sup>[163]</sup>.

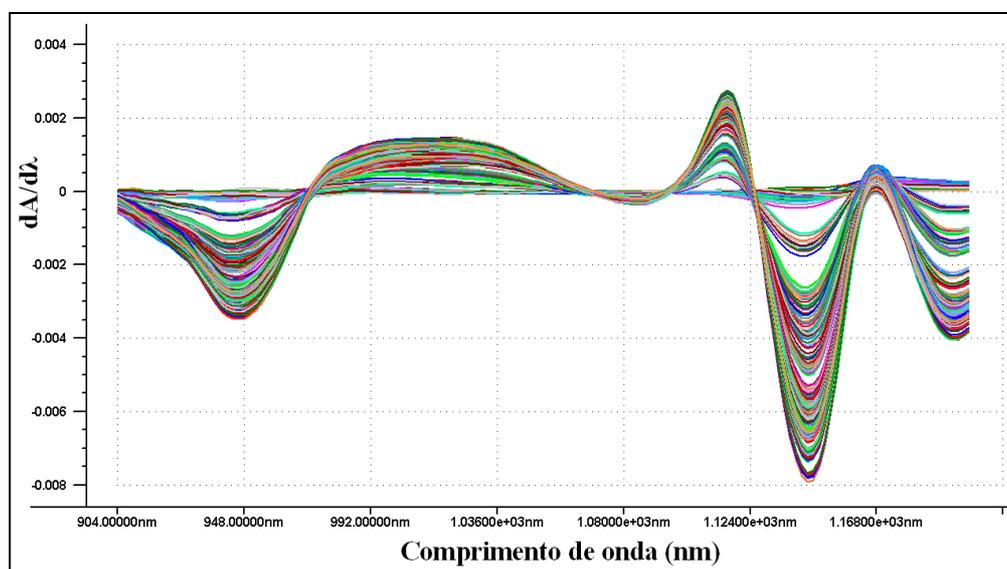


A faixa de absorção que se encontra aproximadamente entre 1000 e 1100 nm corresponde à região de combinação de bandas de estiramento de ligações R-NH-R' e Ar-CH. As bandas de absorção que aparecem entre 1100 e 1200 nm estão associadas à região de combinação de transições fundamentais de C-H, C-H<sub>2</sub> e C-H<sub>3</sub>, que estão associadas às frequências de estiramento destas ligações<sup>[110]</sup>.

### 3.3 Pré-Processamento dos dados espectrais

O padrão do ruído presente nas amostras não é uniforme e não se distribui da mesma maneira em todos os espectros, caracterizando-se como um ruído aleatório. O pré-processamento utilizado deveria ser capaz de remover a maior porção possível desta informação espúria.

O pré-processamento adotado para ser aplicado no conjunto de dados foi a primeira derivada ( $dA/d\lambda$ ) com suavização pelo método *Savitzky-Golay* utilizando uma janela de 21 pontos e polinômio de segunda ordem. Este pré-processamento corrige as variações de linha de base, e minimiza os ruídos espectrais presentes nos dados. A **Figura 15** apresenta o sinal resultante deste procedimento de pré-processamento.



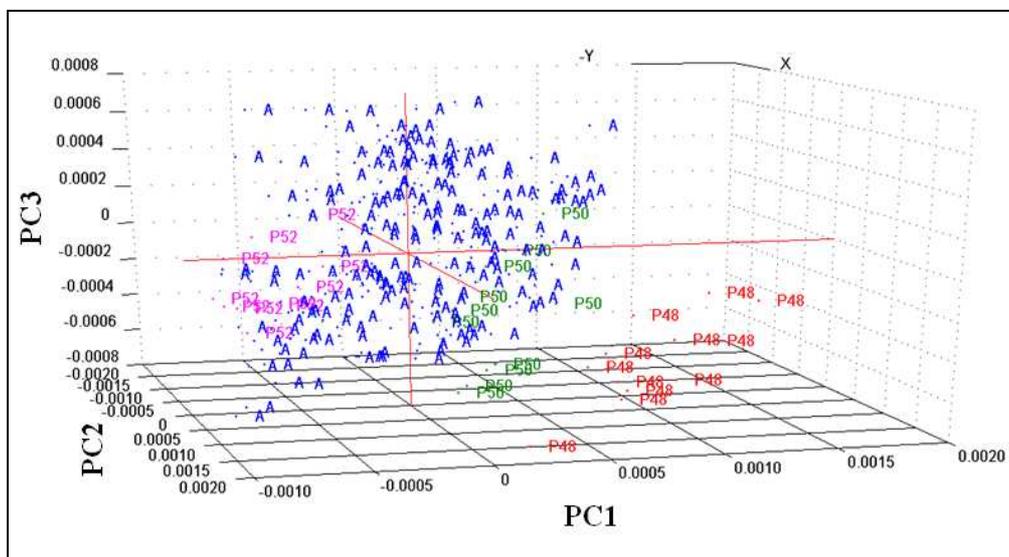
**Figura 15** - Perfis das curvas resultante do pré-processamento em relação aos espectros originais.

### 3.4 Aplicação de uma análise de componentes principais (PCA)

Com efeito, uma vez que se utiliza outro método para a preparação dos padrões, que não seja exatamente àquele com as mesmas condições nas quais as amostras comerciais foram preparadas, pode acontecer de haver fatores que diferenciem significativamente essas amostras. Para se verificar se isso havia ocorrido com o banco de dados gerado, realizou-se uma análise exploratória das

amostras através da construção de um modelo *PCA* com os dados pré-processados, usando três *PC*, com a finalidade de se visualizar a disposição espacial das amostras e avaliar a provável formação de grupos. Se caso houvesse uma diferença realmente significativa entre as amostras comerciais e os padrões preparados no laboratório, observar-se-ia a formação de duas classes distintas de amostras e, portanto, verificar-se-ia a distribuição espacial em dois grupos. Caso isso ocorresse, implicaria na inviabilidade do método, sendo necessária uma reprodução exata das condições utilizadas na indústria farmacêutica, pois indicaria diferenças físico-químicas consideráveis entre as amostras comerciais e os padrões preparados no laboratório.

Para a construção do modelo *PCA* foram usadas todas as amostras comerciais (os 192 espectros das amostras) e as concentrações dos padrões que mais se aproximavam dos valores das concentrações das amostras comerciais adquiridas (os 12 espectros dos padrões na concentração de 48,0% (m/v), os 12 espectros dos padrões na concentração de 50,0% (m/v) e os 12 espectros dos padrões na concentração de 52,0% (m/v)). Como se pode verificar no gráfico tridimensional dos escores obtido pela *PCA*, apresentado na **Figura 16**, essa distribuição em dois grupos não é evidente, e o que se observa é que as amostras comerciais se intercalam entre amostras dos padrões de dipirona, chegando algumas até mesmo a se misturar com os mesmos, indicando uma similaridade entre estas amostras.



**Figura 16** – Gráfico dos escores da *PCA* dos espectros pré-processados das amostras comerciais e de padrões de dipirona sódica. Onde **P48** (em vermelho), **P50** (em verde), **P52** (em rosa), indica os padrões nas respectivas concentrações de 48, 50 e 52% (m/v), e **A** (em azul) indica as amostras comerciais.

A **Figura 17** apresenta o gráfico de influência (para a primeira *PC*), que revela uma variância residual muito baixa entre as amostras (da ordem de  $0,1 \times 10^{-7}$ ) e uma distância da amostra ao centro do modelo também muito pequena (menor que 0.8). Além disso, o

comportamento dessas amostras indica a ausência de amostras anômalas, dado que nenhuma amostra apresenta simultaneamente alto resíduo e alta influência [72].

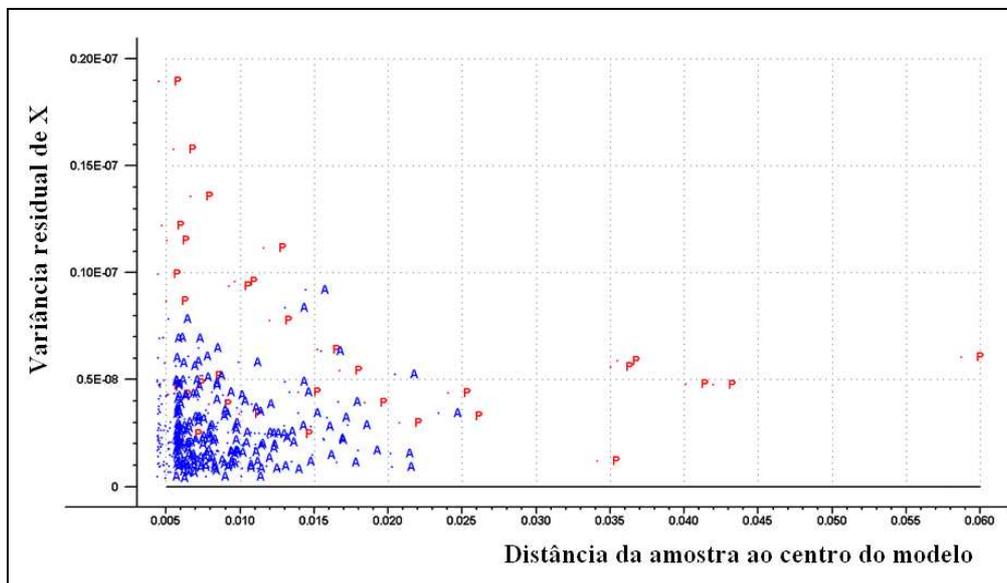


Figura 17 – Gráfico de influência das amostras comerciais (A) e dos padrões de dipirona sódica (P).

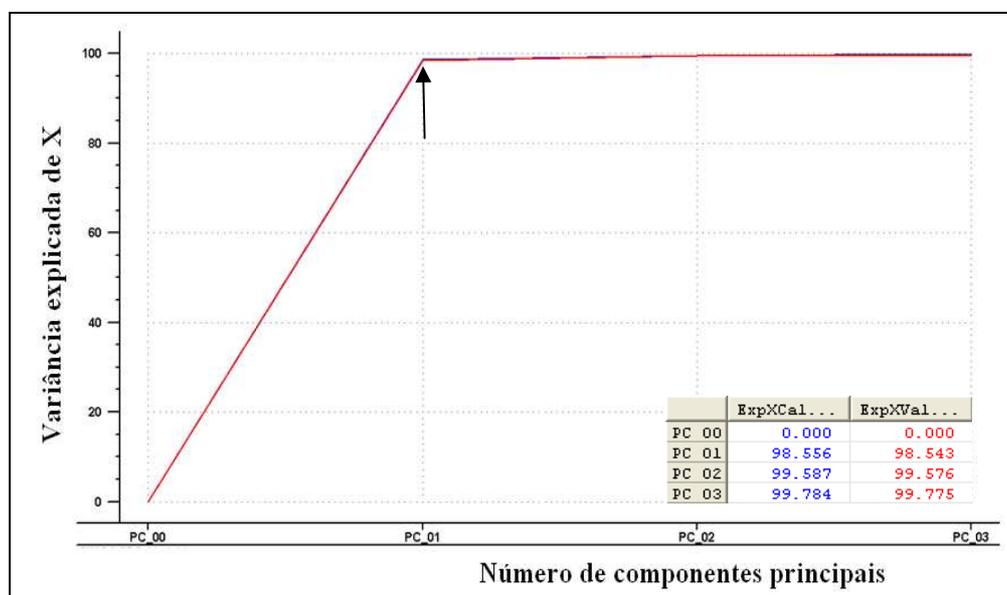
A priori parece não haver diferenças significativas entre as amostras comerciais e os padrões de dipirona sódica preparados no laboratório. Entretanto, para reforçar ainda mais essa hipótese, foi realizada uma análise *Screening* dos dados utilizando modelos *SIMCA*.

### 3.5 Análise Screening

Os espectros *NIR* das amostras de dipirona sódica injetável, como já mencionado, foram pré-processados, e a partir desses espectros pré-processados foram construídos os modelos *SIMCA* com uma parte das amostras dos padrões de dipirona sódica. Estes modelos *SIMCA* foram, então, utilizados para a realização da análise *Screening* de amostras “desconhecidas” (amostras de padrões que não entraram na modelagem e as amostras comerciais de dipirona sódica injetável), para prever se uma determinada amostra pertence ou não à classe de amostras previamente modeladas.

Para este propósito, um conjunto de treinamento foi composto por padrões de dipirona sódica preparados nas concentrações de 0,0 (contendo somente água para injetáveis), 10,0; 20,0; 24,0; 28,0; 32,0; 36,0; 40,0; 44,0; 48,0; 52,0; 56,0 e 60,0% (m/v). A partir deste conjunto foram construídos os modelos *SIMCA*, empregando o processo de validação cruzada para 1, 2 e 3 *PC*. O gráfico da variância explicada em função do número de *PC* (Figura18) foi, então, utilizado como ferramenta de diagnóstico para determinar o número ideal de componentes principais a ser usado na construção deste modelo.

O conjunto reservado para predição contempla os padrões de dipirona sódica nas concentrações de 44,5; 45,0; 46,0; 50,0; 54,0; 54,5; 55,0 e 55,5% (m/v) e também todo o conjunto de amostras comerciais analisadas.



**Figura 18** – Gráfico da variância explicada de X versus o número de componentes do modelo *SIMCA*.

Pela análise do gráfico da variância explicada em função do número de componentes principais, optou-se por utilizar apenas uma *PC* para a construção do modelo *SIMCA*, visto que esta já explica 98,6% da variância de **X** (matriz das variáveis espectrais), e como a *PC2* explica apenas cerca de 1% a mais do que esta, e a *PC3* explica somente cerca de 0,2% a mais que a *PC2*, a inclusão dessas *PC* poderiam prejudicar o modelo. O modelo *SIMCA* gerado foi então utilizado para a análise *Screening* de todas as amostras comerciais e padrões de dipirona sódica (conjunto de treinamento e o de predição), empregando um nível de confiança de 95%.

Na **Tabela 3** são apresentados os resultados do modelo *SIMCA* usado para análise *Screening* de todas as amostras de dipirona sódica. Essa análise demonstrou que o modelo *SIMCA* foi capaz de classificar tanto as amostras de dipirona sódica de sua classe (amostras de padrões preparadas nas mesmas condições), como as amostras da classe formada pelas amostras comerciais de dipirona sódica injetável, ou seja, todas as amostras foram classificadas como pertencentes a uma mesma classe.

Este mesmo resultado também pode ser observado no gráfico de *Coomans* criado a partir do mesmo conjunto de dados, que está ilustrado na **Figura 19**. Neste gráfico, observa-se que nenhuma amostra ficou fora do espaço dimensional criado pelo modelo *SIMCA*.

Tabela 3.- Análise *Screening* de todas as amostras de dipirona do conjunto de validação e predição.

A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C				
P0 1	*	P20 7	*	P36 1	*	P44 19	*	P46 1	*	P52 7	*	P55 13	*	P56 19	*	A50,27 1	*	A50,97 19	*	A51,67 13	*	A52,72 7	*
P0 2	*	P20 8	*	P36 2	*	P44 20	*	P46 2	*	P52 8	*	P55 14	*	P56 20	*	A50,27 2	*	A50,97 20	*	A51,67 14	*	A52,72 8	*
P0 3	*	P20 9	*	P36 3	*	P44 21	*	P46 3	*	P52 9	*	P55 15	*	P56 21	*	A50,27 3	*	A50,97 21	*	A51,67 15	*	A52,72 9	*
P0 4	*	P20 10	*	P36 4	*	P44 22	*	P46 4	*	P52 10	*	P55 16	*	P56 22	*	A50,27 4	*	A50,97 22	*	A51,67 16	*	A52,72 10	*
P0 5	*	P20 11	*	P36 5	*	P44 23	*	P46 5	*	P52 11	*	P55 17	*	P56 23	*	A50,27 5	*	A50,97 23	*	A51,67 17	*	A52,72 11	*
P0 6	*	P20 12	*	P36 6	*	P44 24	*	P46 6	*	P52 12	*	P55 18	*	P56 24	*	A50,27 6	*	A50,97 24	*	A51,67 18	*	A52,72 12	*
P0 7	*	P24 1	*	P36 7	*	P44,5 1	*	P46 7	*	P54 1	*	P55 19	*	P60 1	*	A50,27 7	*	A50,98 1	*	A51,67 19	*	A52,72 13	*
P0 8	*	P24 2	*	P36 8	*	P44,5 2	*	P46 8	*	P54 2	*	P55 20	*	P60 2	*	A50,27 8	*	A50,98 2	*	A51,67 20	*	A52,72 14	*
P0 9	*	P24 3	*	P36 9	*	P44,5 3	*	P46 9	*	P54 3	*	P55 21	*	P60 3	*	A50,27 9	*	A50,98 3	*	A51,67 21	*	A52,72 15	*
P0 10	*	P24 4	*	P36 10	*	P44,5 4	*	P46 10	*	P54 4	*	P55 22	*	P60 4	*	A50,27 10	*	A50,98 4	*	A51,67 22	*	A52,72 16	*
P0 11	*	P24 5	*	P36 11	*	P44,5 5	*	P46 11	*	P54 5	*	P55 23	*	P60 5	*	A50,27 11	*	A50,98 5	*	A51,67 23	*	A52,72 17	*
P0 12	*	P24 6	*	P36 12	*	P44,5 6	*	P46 12	*	P54 6	*	P55 24	*	P60 6	*	A50,27 12	*	A50,98 6	*	A51,67 24	*	A52,72 18	*
P0 13	*	P24 7	*	P40 1	*	P44,5 7	*	P48 1	*	P54 7	*	P55,5 1	*	P60 7	*	A50,27 13	*	A50,98 7	*	A52,02 1	*	A52,72 19	*
P0 14	*	P24 8	*	P40 2	*	P44,5 8	*	P48 2	*	P54 8	*	P55,5 2	*	P60 8	*	A50,27 14	*	A50,98 8	*	A52,02 2	*	A52,72 20	*
P0 15	*	P24 9	*	P40 3	*	P44,5 9	*	P48 3	*	P54 9	*	P55,5 3	*	P60 9	*	A50,27 15	*	A50,98 9	*	A52,02 3	*	A52,72 21	*
P0 16	*	P24 10	*	P40 4	*	P44,5 10	*	P48 4	*	P54 10	*	P55,5 4	*	P60 10	*	A50,27 16	*	A50,98 10	*	A52,02 4	*	A52,72 22	*
P0 17	*	P24 11	*	P40 5	*	P44,5 11	*	P48 5	*	P54 11	*	P55,5 5	*	P60 11	*	A50,27 17	*	A50,98 11	*	A52,02 5	*	A52,72 23	*
P0 18	*	P24 12	*	P40 6	*	P44,5 12	*	P48 6	*	P54 12	*	P55,5 6	*	P60 12	*	A50,27 18	*	A50,98 12	*	A52,02 6	*	A52,72 24	*
P0 19	*	P28 1	*	P40 7	*	P45 1	*	P48 7	*	P54,5 1	*	P55,5 7	*	A49,92 1	*	A50,27 19	*	A50,98 13	*	A52,02 7	*	A52,90 1	*
P0 20	*	P28 2	*	P40 8	*	P45 2	*	P48 8	*	P54,5 2	*	P55,5 8	*	A49,92 2	*	A50,27 20	*	A50,98 14	*	A52,02 8	*	A52,90 2	*
P0 21	*	P28 3	*	P40 9	*	P45 3	*	P48 9	*	P54,5 3	*	P55,5 9	*	A49,92 3	*	A50,27 21	*	A50,98 15	*	A52,02 9	*	A52,90 3	*
P0 22	*	P28 4	*	P40 10	*	P45 4	*	P48 10	*	P54,5 4	*	P55,5 10	*	A49,92 4	*	A50,27 22	*	A50,98 16	*	A52,02 10	*	A52,90 4	*
P0 23	*	P28 5	*	P40 11	*	P45 5	*	P48 11	*	P54,5 5	*	P55,5 11	*	A49,92 5	*	A50,27 23	*	A50,98 17	*	A52,02 11	*	A52,90 5	*
P0 24	*	P28 6	*	P40 12	*	P45 6	*	P48 12	*	P54,5 6	*	P55,5 12	*	A49,92 6	*	A50,27 24	*	A50,98 18	*	A52,02 12	*	A52,90 6	*
P10 1	*	P28 7	*	P44 1	*	P45 7	*	P50 1	*	P54,5 7	*	P56 1	*	A49,92 7	*	A50,97 1	*	A50,98 19	*	A52,02 13	*	A52,90 7	*
P10 2	*	P28 8	*	P44 2	*	P45 8	*	P50 2	*	P54,5 8	*	P56 2	*	A49,92 8	*	A50,97 2	*	A50,98 20	*	A52,02 14	*	A52,90 8	*
P10 3	*	P28 9	*	P44 3	*	P45 9	*	P50 3	*	P54,5 9	*	P56 3	*	A49,92 9	*	A50,97 3	*	A50,98 21	*	A52,02 15	*	A52,90 9	*
P10 4	*	P28 10	*	P44 4	*	P45 10	*	P50 4	*	P54,5 10	*	P56 4	*	A49,92 10	*	A50,97 4	*	A50,98 22	*	A52,02 16	*	A52,90 10	*
P10 5	*	P28 11	*	P44 5	*	P45 11	*	P50 5	*	P54,5 11	*	P56 5	*	A49,92 11	*	A50,97 5	*	A50,98 23	*	A52,02 17	*	A52,90 11	*
P10 6	*	P28 12	*	P44 6	*	P45 12	*	P50 6	*	P54,5 12	*	P56 6	*	A49,92 12	*	A50,97 6	*	A50,98 24	*	A52,02 18	*	A52,90 12	*
P10 7	*	P32 1	*	P44 7	*	P45 13	*	P50 7	*	P55 1	*	P56 7	*	A49,92 13	*	A50,97 7	*	A51,67 1	*	A52,02 19	*	A52,90 13	*
P10 8	*	P32 2	*	P44 8	*	P45 14	*	P50 8	*	P55 2	*	P56 8	*	A49,92 14	*	A50,97 8	*	A51,67 2	*	A52,02 20	*	A52,90 14	*
P10 9	*	P32 3	*	P44 9	*	P45 15	*	P50 9	*	P55 3	*	P56 9	*	A49,92 15	*	A50,97 9	*	A51,67 3	*	A52,02 21	*	A52,90 15	*
P10 10	*	P32 4	*	P44 10	*	P45 16	*	P50 10	*	P55 4	*	P56 10	*	A49,92 16	*	A50,97 10	*	A51,67 4	*	A52,02 22	*	A52,90 16	*
P10 11	*	P32 5	*	P44 11	*	P45 17	*	P50 11	*	P55 5	*	P56 11	*	A49,92 17	*	A50,97 11	*	A51,67 5	*	A52,02 23	*	A52,90 17	*
P10 12	*	P32 6	*	P44 12	*	P45 18	*	P50 12	*	P55 6	*	P56 12	*	A49,92 18	*	A50,97 12	*	A51,67 6	*	A52,02 24	*	A52,90 18	*
P20 1	*	P32 7	*	P44 13	*	P45 19	*	P52 1	*	P55 7	*	P56 13	*	A49,92 19	*	A50,97 13	*	A51,67 7	*	A52,72 1	*	A52,90 19	*
P20 2	*	P32 8	*	P44 14	*	P45 20	*	P52 2	*	P55 8	*	P56 14	*	A49,92 20	*	A50,97 14	*	A51,67 8	*	A52,72 2	*	A52,90 20	*
P20 3	*	P32 9	*	P44 15	*	P45 21	*	P52 3	*	P55 9	*	P56 15	*	A49,92 21	*	A50,97 15	*	A51,67 9	*	A52,72 3	*	A52,90 21	*
P20 4	*	P32 10	*	P44 16	*	P45 22	*	P52 4	*	P55 10	*	P56 16	*	A49,92 22	*	A50,97 16	*	A51,67 10	*	A52,72 4	*	A52,90 22	*
P20 5	*	P32 11	*	P44 17	*	P45 23	*	P52 5	*	P55 11	*	P56 17	*	A49,92 23	*	A50,97 17	*	A51,67 11	*	A52,72 5	*	A52,90 23	*
P20 6	*	P32 12	*	P44 18	*	P45 24	*	P52 6	*	P55 12	*	P56 18	*	A49,92 24	*	A50,97 18	*	A51,67 12	*	A52,72 6	*	A52,90 24	*

O asterisco (“\*”) na tabela indica que a amostra foi classificada como pertencente à classe.

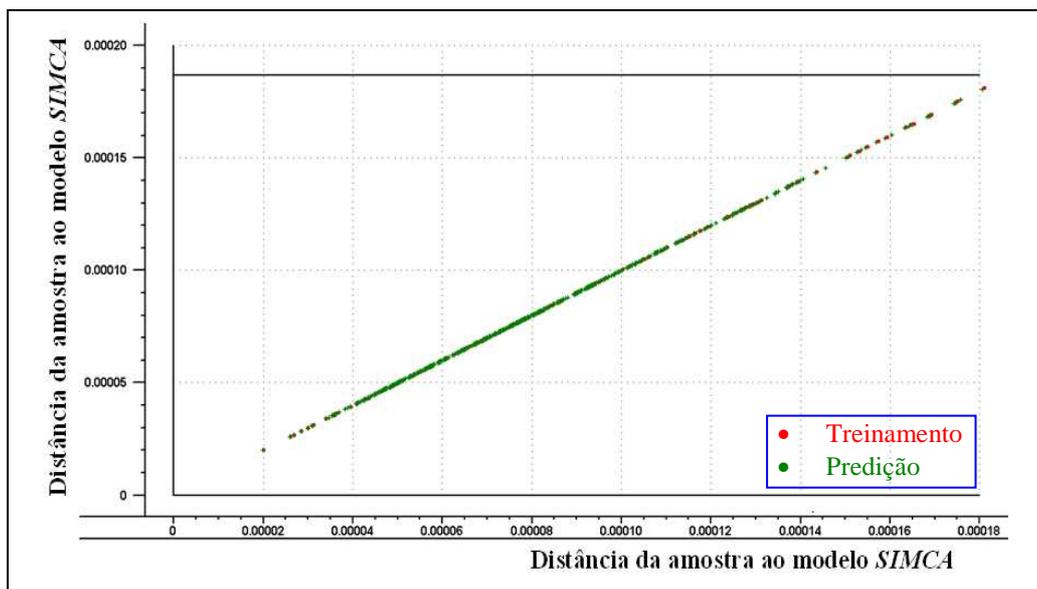


Figura 19 – Gráfico de *Coomans* do modelo *SIMCA*. Onde, em vermelho estão as amostras do conjunto de treinamento e, em verde estão as amostras do conjunto de predição.

Após testar o modelo com toda a faixa de concentração preparadas, optou-se por testar também a faixa de concentração aceita pelos órgãos fiscalizadores que é de  $\pm 10\%$  (m/v). Para isso, usou-se um conjunto de treinamento composto por padrões de dipirona sódica preparados nas concentrações de 45,0; 46,0; 50,0; 54,0; 54,5 e 55,0 % (m/v). Empregou-se o processo de validação cruzada e utilizou-se uma *PC* para a construção do modelo *SIMCA*. O conjunto de predição ficou reservado apenas para as amostras comerciais de dipirona sódica injetável.

A **Tabela 4** apresentada os resultados do modelo *SIMCA* usado para análise *Screening* na faixa de concentração de 45 a 55% de dipirona sódica. Nesta análise o modelo *SIMCA* foi capaz de classificar todas as amostras de dipirona sódica comerciais, com um nível de confiança de 99%.

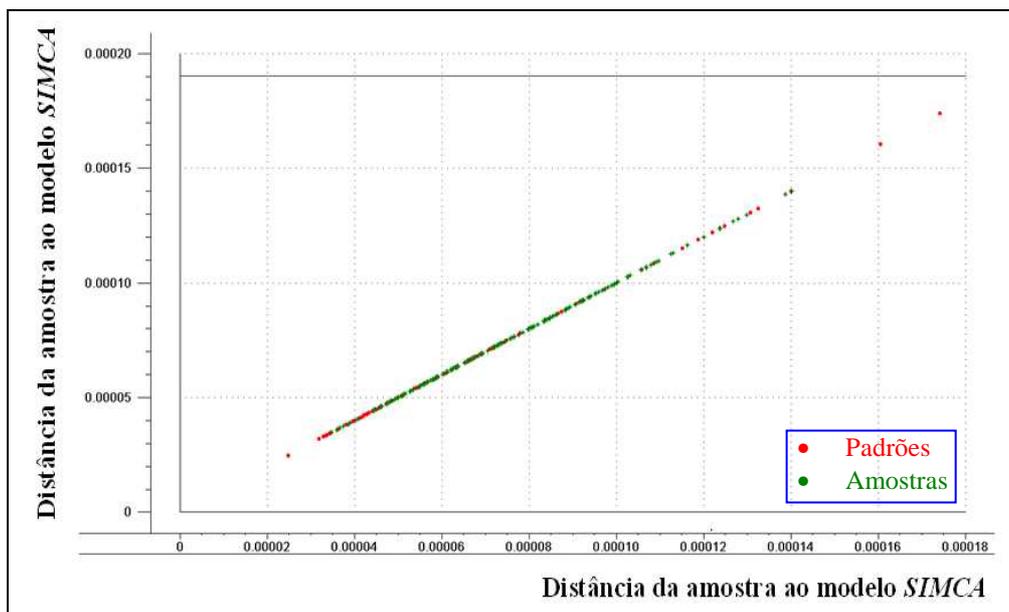
**Tabela 4.-** Análise *Screening* das amostras de dipirona comerciais do conjunto predição, usando o modelo *SIMCA* construído para a faixa de 45 a 55 % (m/v).

A	C	A	C	A	C	A	C	A	C	A	C	A	C	A	C
A49,92 1	*	A50,27 1	*	A50,97 1	*	A50,98 1	*	A51,67 1	*	A52,02 1	*	A52,72 1	*	A52,90 1	*
A49,92 2	*	A50,27 2	*	A50,97 2	*	A50,98 2	*	A51,67 2	*	A52,02 2	*	A52,72 2	*	A52,90 2	*
A49,92 3	*	A50,27 3	*	A50,97 3	*	A50,98 3	*	A51,67 3	*	A52,02 3	*	A52,72 3	*	A52,90 3	*
A49,92 4	*	A50,27 4	*	A50,97 4	*	A50,98 4	*	A51,67 4	*	A52,02 4	*	A52,72 4	*	A52,90 4	*
A49,92 5	*	A50,27 5	*	A50,97 5	*	A50,98 5	*	A51,67 5	*	A52,02 5	*	A52,72 5	*	A52,90 5	*
A49,92 6	*	A50,27 6	*	A50,97 6	*	A50,98 6	*	A51,67 6	*	A52,02 6	*	A52,72 6	*	A52,90 6	*
A49,92 7	*	A50,27 7	*	A50,97 7	*	A50,98 7	*	A51,67 7	*	A52,02 7	*	A52,72 7	*	A52,90 7	*
A49,92 8	*	A50,27 8	*	A50,97 8	*	A50,98 8	*	A51,67 8	*	A52,02 8	*	A52,72 8	*	A52,90 8	*
A49,92 9	*	A50,27 9	*	A50,97 9	*	A50,98 9	*	A51,67 9	*	A52,02 9	*	A52,72 9	*	A52,90 9	*
A49,92 10	*	A50,27 10	*	A50,97 10	*	A50,98 10	*	A51,67 10	*	A52,02 10	*	A52,72 10	*	A52,90 10	*
A49,92 11	*	A50,27 11	*	A50,97 11	*	A50,98 11	*	A51,67 11	*	A52,02 11	*	A52,72 11	*	A52,90 11	*
A49,92 12	*	A50,27 12	*	A50,97 12	*	A50,98 12	*	A51,67 12	*	A52,02 12	*	A52,72 12	*	A52,90 12	*
A49,92 13	*	A50,27 13	*	A50,97 13	*	A50,98 13	*	A51,67 13	*	A52,02 13	*	A52,72 13	*	A52,90 13	*
A49,92 14	*	A50,27 14	*	A50,97 14	*	A50,98 14	*	A51,67 14	*	A52,02 14	*	A52,72 14	*	A52,90 14	*
A49,92 15	*	A50,27 15	*	A50,97 15	*	A50,98 15	*	A51,67 15	*	A52,02 15	*	A52,72 15	*	A52,90 15	*
A49,92 16	*	A50,27 16	*	A50,97 16	*	A50,98 16	*	A51,67 16	*	A52,02 16	*	A52,72 16	*	A52,90 16	*
A49,92 17	*	A50,27 17	*	A50,97 17	*	A50,98 17	*	A51,67 17	*	A52,02 17	*	A52,72 17	*	A52,90 17	*
A49,92 18	*	A50,27 18	*	A50,97 18	*	A50,98 18	*	A51,67 18	*	A52,02 18	*	A52,72 18	*	A52,90 18	*
A49,92 19	*	A50,27 19	*	A50,97 19	*	A50,98 19	*	A51,67 19	*	A52,02 19	*	A52,72 19	*	A52,90 19	*
A49,92 20	*	A50,27 20	*	A50,97 20	*	A50,98 20	*	A51,67 20	*	A52,02 20	*	A52,72 20	*	A52,90 20	*
A49,92 21	*	A50,27 21	*	A50,97 21	*	A50,98 21	*	A51,67 21	*	A52,02 21	*	A52,72 21	*	A52,90 21	*
A49,92 22	*	A50,27 22	*	A50,97 22	*	A50,98 22	*	A51,67 22	*	A52,02 22	*	A52,72 22	*	A52,90 22	*
A49,92 23	*	A50,27 23	*	A50,97 23	*	A50,98 23	*	A51,67 23	*	A52,02 23	*	A52,72 23	*	A52,90 23	*
A49,92 24	*	A50,27 24	*	A50,97 24	*	A50,98 24	*	A51,67 24	*	A52,02 24	*	A52,72 24	*	A52,90 24	*

O asterisco (“\*”) na tabela indica que a amostra foi classificada como pertencente à classe.

O mesmo resultado também foi observado no gráfico de *Coomans*, ilustrado na **Figura 20**, onde se observa que nenhuma amostra ficou fora do espaço dimensional criado pelo modelo *SIMCA*.

A vantagem de se realizar uma análise *Screening* é que essa poderosa técnica permite uma análise prévia dos dados. Com ela torna-se possível identificar anomalias no conjunto de dados, bem como identificar amostras adulteradas. Quando se tem um grande conjunto de amostras, como o utilizado neste trabalho, e se verifica a presença de amostras anômalas nesse conjunto, essas amostras é que seriam primeiramente investigadas, e não todo o banco de dados, o que na prática gera a redução de análises químicas, redução de custos operacionais e de gastos com reagentes.



**Figura 20** – Gráfico de *Coomans* para o modelo *SIMCA* construído para a faixa de 45 a 55% (m/v). Onde, em vermelho estão as amostras do conjunto de calibração/validação e, em verde estão as amostras do conjunto de predição

Os resultados obtidos com a análise *Screening* foram muito satisfatórios, e demonstram que não há diferença significativa entre os conjuntos de amostras preparadas no laboratório e as amostras comerciais, sendo todas amostras apropriadas para a construção e validação dos modelos de calibração para a predição do teor de dipirona sódica em medicamentos injetáveis.

### 3.6 Resultados da aplicação dos modelos

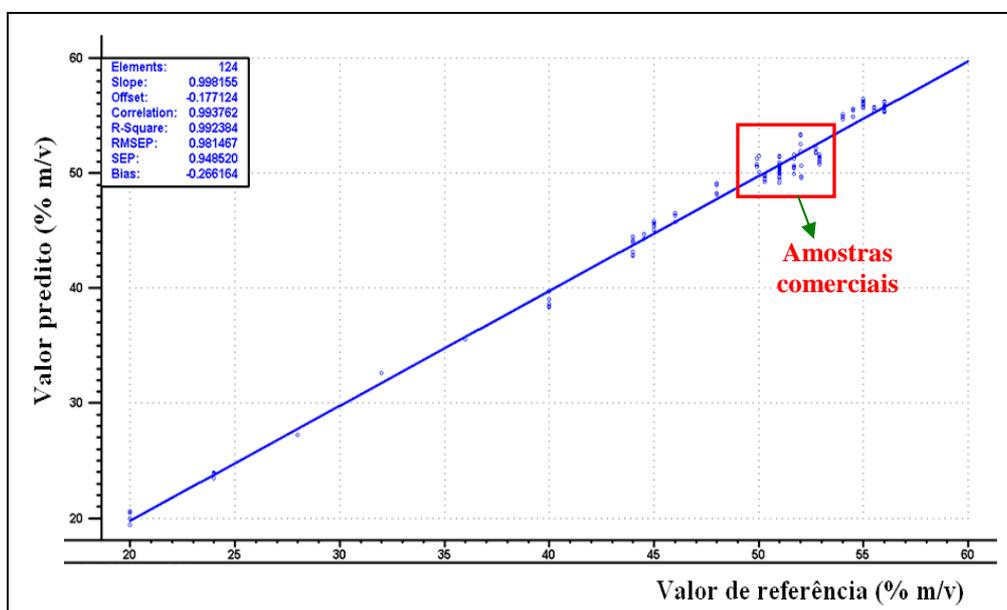
Após pré-processados os dados relativos às amostras e aos padrões de dipirona, foram separados em três grupos: calibração, validação e predição, sendo 252 amostras sintéticas e comerciais separadas para o conjunto de calibração, 126 amostras sintéticas e comerciais separadas para o conjunto de validação e 126 amostras sintéticas e comerciais separadas para o conjunto de predição. As amostras com concentrações extremas (10,0 e 60,0%), bem como as que não contém o princípio ativo (0,0%) só entraram no conjunto de calibração para evitar extrapolação. Para a seleção das amostras de cada conjunto utilizou-se o algoritmo *SPXY*.

Esses conjuntos foram utilizados para a elaboração de modelos de calibração (*PLS1*, *MLR-SPA*, *MLR-ASA*, *MLR-GA* e *MLR-SW*) com o objetivo de utilizá-los para a determinação da concentração de dipirona (predição) nas amostras reais e em alguns padrões que simulavam a presença de amostras fora das especificações no conjunto.

### 3.6.1 Modelo PLS1

Para a construção do modelo *PLS1*, utilizou-se apenas uma variável latente, uma vez que esta explica 99,2% da variância. Além disso, acreditava-se que o parâmetro que representasse a maior variância, e portanto estivesse variando na primeira *PC*, fosse a concentração do princípio ativo. Essa afirmativa foi reforçada com os resultados obtidos pelo modelo de regressão.

Para a validação do modelo *PLS1* adotou-se o método série de teste. A **Figura 21** ilustra o gráfico do valor predito versus o valor de referência das concentrações de dipirona sódica para o conjunto de validação obtido pelo modelo de calibração *PLS1*. Neste gráfico observa-se uma boa correlação linear (no valor de 0,9938), e um *RMSEV* de 0,98 % (m/v).



**Figura 21** - Gráfico do valor predito versus o valor de referência para o conjunto de validação do modelo *PLS1*.

O modelo elaborado (calibrado e validado) foi, então, aplicado as amostras comerciais de medicamentos injetáveis e nos padrões de concentrações de 20,0 a 56,0% (m/v), para a predição das concentrações de dipirona sódica. Os resultados da predição dos valores das concentrações estão resumidos no gráfico da **Figura 22**. O valor do coeficiente de correlação linear obtido e o *RMSEP* foram respectivamente, 0,9904 e 1,07 % (m/v).

Tanto as amostras do conjunto de validação como as amostras do conjunto de predição apresentaram-se distribuídas de forma aleatória em torno da bissetriz (**Figura 21** e **Figura 22**), indicando ausência de erro sistemático no modelo *PLS1*. O erro encontrado para predição da concentração de dipirona nas amostras é muito baixo, com relação ao método de referência.

Portanto, pode-se afirmar que o modelo *PLS1* é bastante promissor para a determinação de dipirona sódica neste medicamento injetável comercial.

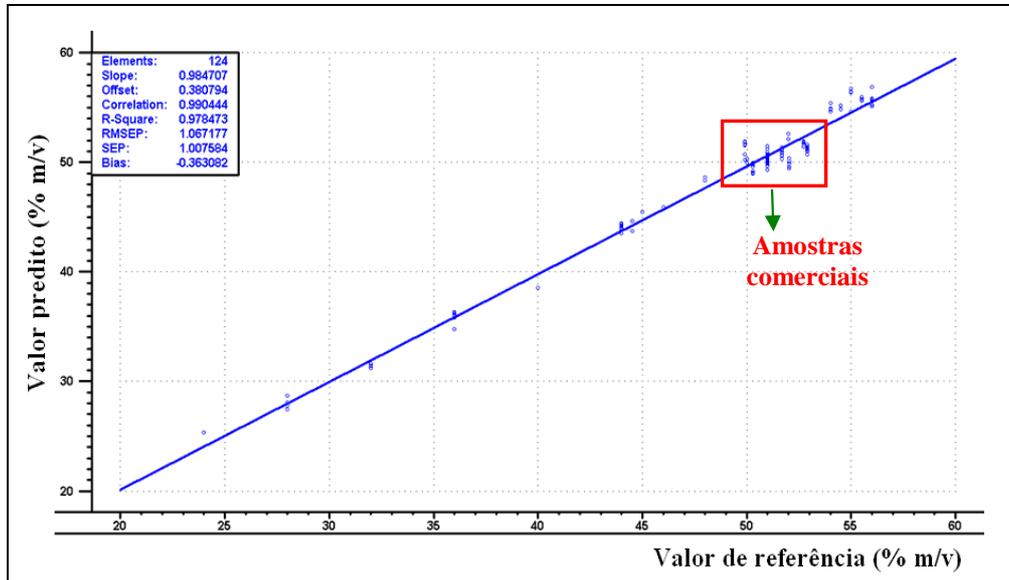


Figura 22 - Gráfico do valor predito versus o valor de referência para o conjunto de predição do modelo *PLS1*.

### 3.6.2 Modelo *MLR-SPA*

A partir dos conjuntos de amostras selecionados pelo *SPXY*, o algoritmo das projeções sucessivas (*SPA*) selecionou 9 variáveis espectrais mais informativas e não redundantes. A escolha do número ótimo de variáveis determinadas pelo *SPA* foi realizada a partir dos valores dos *RMSEV* obtidos. A **Figura 23** mostra a variação do *RMSEV* versus o número de variáveis selecionadas.

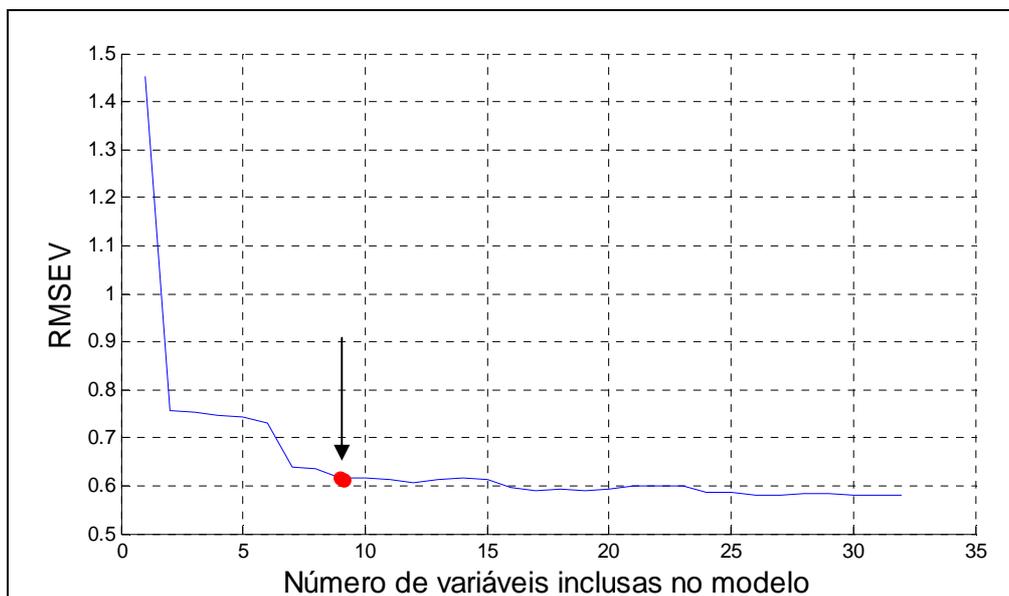
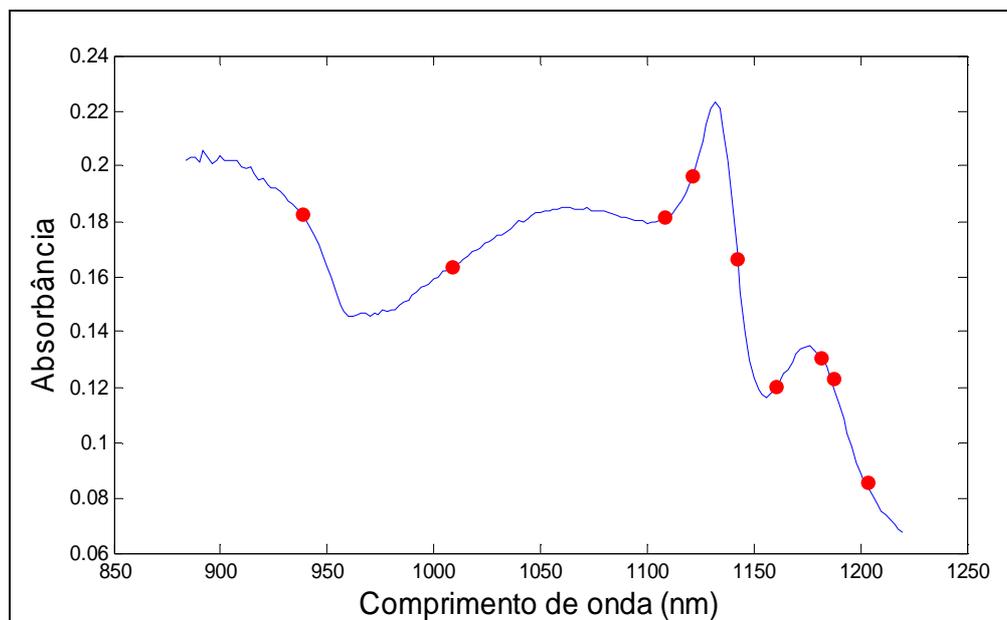


Figura 23 – Variação do *RMSEV* (% m/v) em função do número de variáveis incluídas no modelo *MLR-SPA*.

A **Figura 24** apresenta o gráfico do espectro *NIR* de uma amostra de dipirona sódica, indicando aproximadamente a localização das variáveis selecionadas pelo algoritmo *SPA*. Nesta figura, pode-se observar que as variáveis selecionadas estão localizadas em regiões espectrais informativas.



**Figura 24** - Gráfico do espectro *NIR* de uma amostra de dipirona sódica, indicando aproximadamente a localização das variáveis selecionadas pelo algoritmo *SPA*.

As variáveis selecionadas pelo algoritmo *SPA* foram escolhidas no procedimento de validação do modelo de regressão *MLR*. A **Figura 25** apresenta o gráfico do valor predito versus o valor de referência para o conjunto de validação do modelo *MLR-SPA*, onde se verifica um coeficiente de correlação linear de 0,9974 e um *RMSEV* de 0,62% (m/v). Neste gráfico, pode-se observar que as amostras do conjunto de validação apresentaram-se distribuídas de forma aleatória em torno da bissetriz, sendo um indicativo da ausência de erro sistemático no modelo *SPA-MLR*.

Da mesma forma como utilizado no *PLS1*, o modelo *MLR-SPA* elaborado (calibrado e validado) foi, então, aplicado nas amostras comerciais de medicamentos injetáveis e nos padrões de concentrações de 20,0 a 56,0% (m/v), para a predição das concentrações de dipirona sódica.

O gráfico da **Figura 26** apresenta o gráfico do valor predito versus o valor de referência das concentrações de dipirona sódica para o conjunto de predição do modelo *MLR-SPA*. O valor da correlação obtido e o *RMSEP* foram respectivamente iguais a 0,9938 e 0,81% (m/v). As amostras do conjunto do conjunto de predição localizadas na **Figura 26** apresentaram-se distribuídas de forma aleatória em torno da bissetriz, indicando ausência de erro sistemático no modelo *SPA-MLR*.

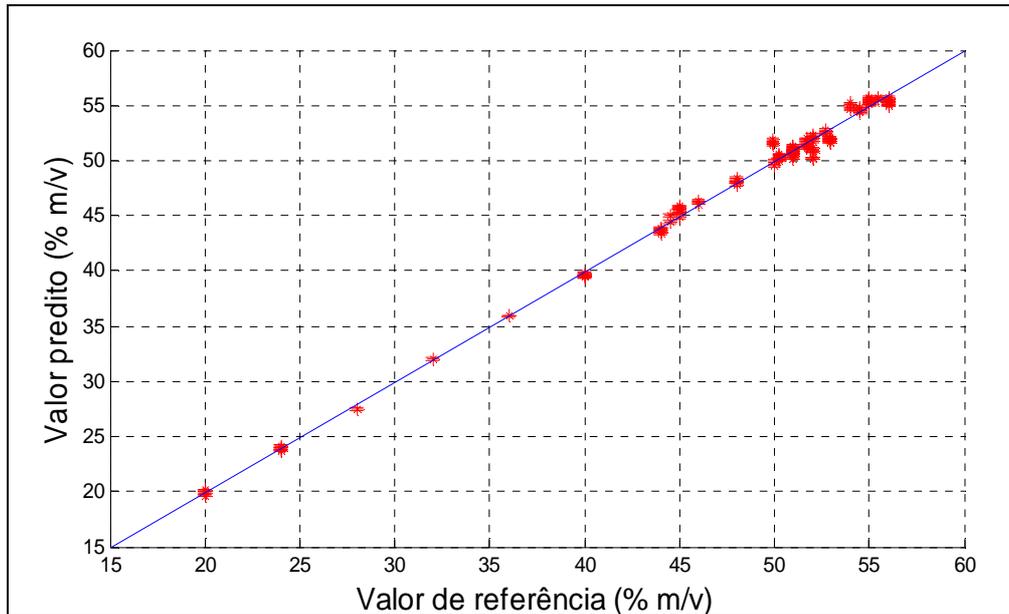


Figura 25 - Gráfico do valor predito versus o valor de referência para o conjunto de validação do modelo *MLR-SPA*.

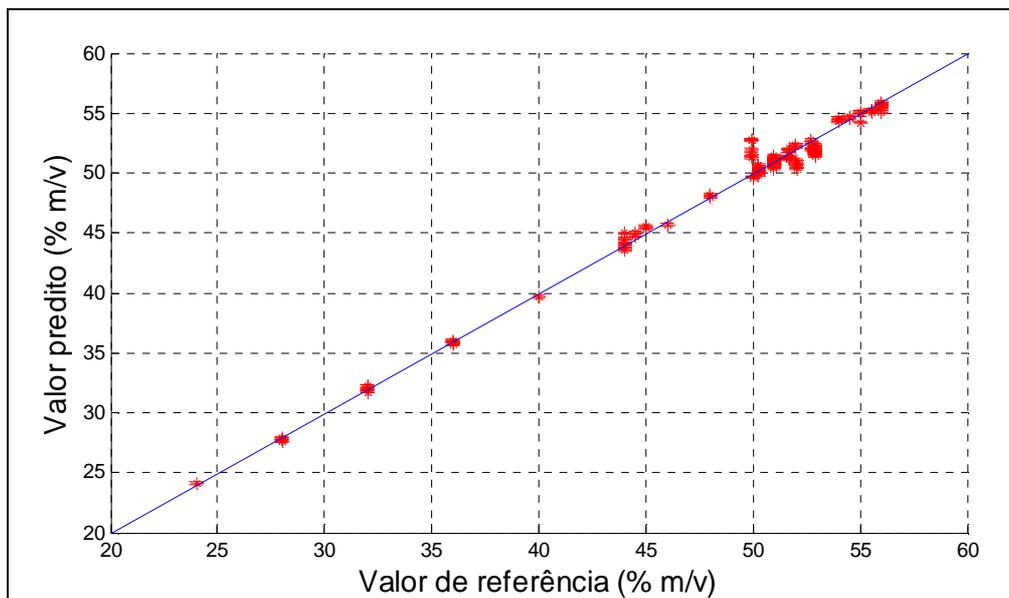
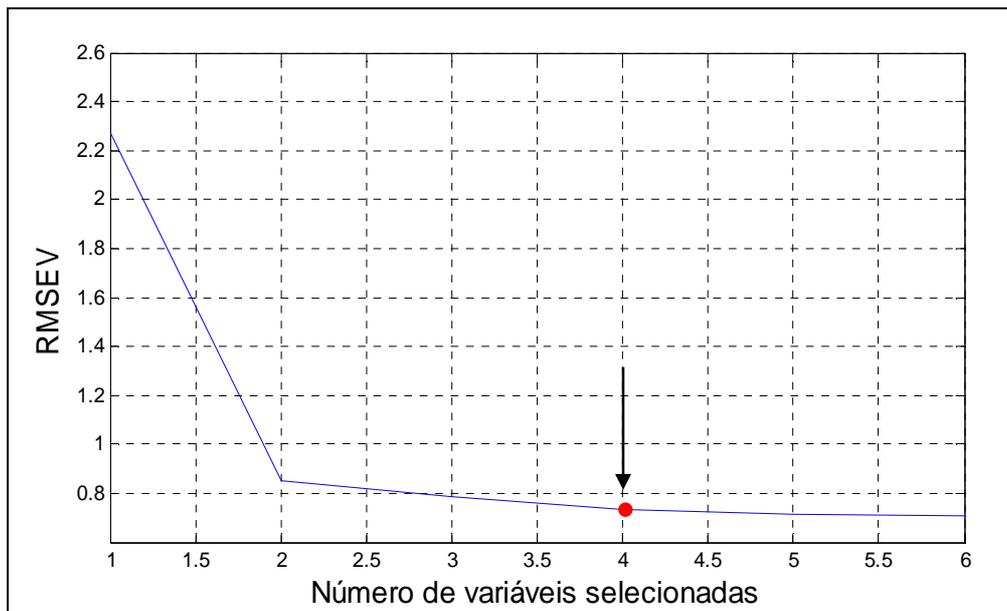


Figura 26 - Gráfico do valor predito versus o valor de referência para o conjunto de predição do modelo *MLR-SPA*.

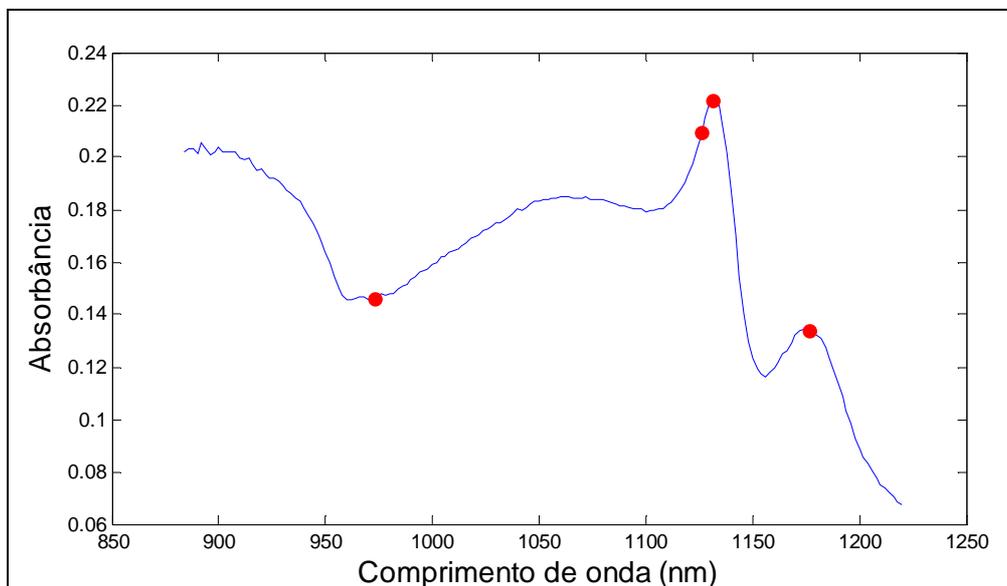
### 3.6.3 Modelo *MLR-ASA*

Na a construção do modelo *MLR-ASA* o valor adotado para o *VIF* foi 5. O algoritmo *ASA* selecionou, a partir do *RMSEV*, apenas 4 variáveis espectrais mais informativas e não redundantes, para serem usadas para a construção do modelo de regressão *MLR*, conforme mostrado na **Figura 27**.



**Figura 27** - Variação do *RMSEV* (% m/v) em função do número de variáveis espectrais selecionadas pelo *ASA*.

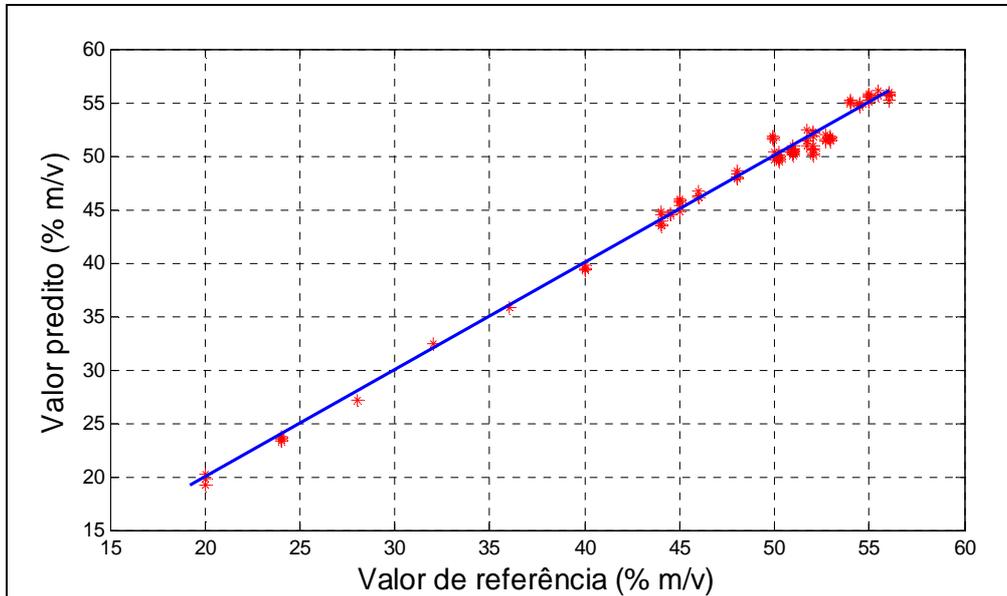
A **Figura 28** apresenta o gráfico do espectro *NIR* de uma amostra de dipirona sódica, indicando aproximadamente a localização das variáveis selecionadas pelo algoritmo *ASA*. Nesta figura, pode-se observar que as variáveis selecionadas estão localizadas em regiões espectrais bastante informativas.



**Figura 28** - Gráfico do espectro *NIR* de uma amostra de dipirona sódica, indicando aproximadamente a localização das variáveis selecionadas pelo algoritmo *ASA*.

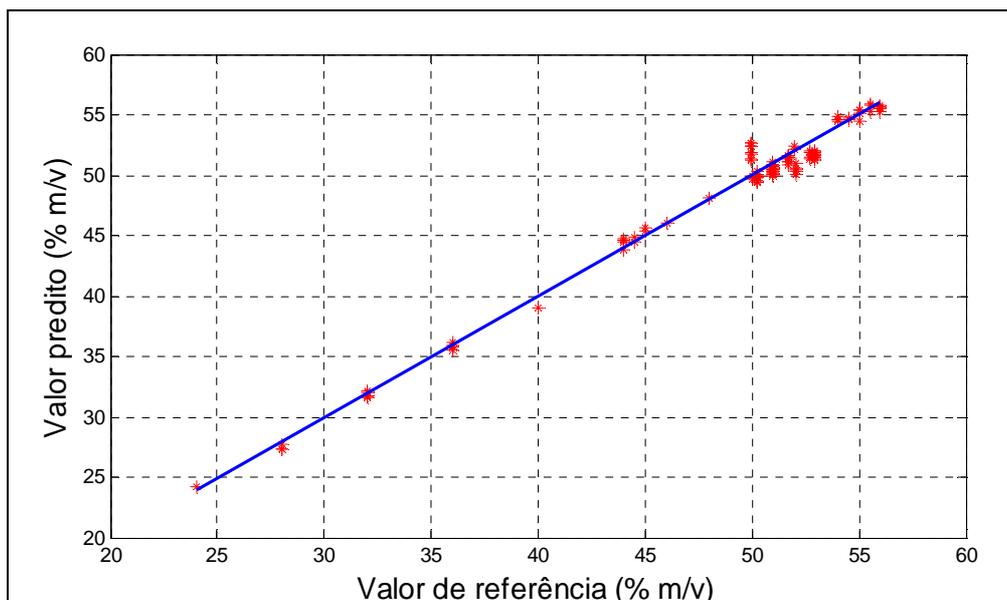
As variáveis selecionadas pelo *ASA* foram utilizadas para a construção e validação do modelo de regressão *MLR*. O gráfico da **Figura 29** apresenta o valor predito versus o valor de

referência das concentrações de dipirona para o conjunto de validação do modelo *MLR-ASA*. O valor do coeficiente de correlação linear e do *RMSEV* são respectivamente 0,9964 e 0,73% (m/v).



**Figura 29** - Gráfico do valor predito versus o valor de referência para o conjunto de validação do modelo *MLR-ASA*.

O gráfico da **Figura 30** apresenta o valor predito versus o valor de referência das concentrações de dipirona para o conjunto de predição do modelo *MLR-ASA*. O valor da correlação obtido e o *RMSEP* foram respectivamente iguais a 0,9930 e 0,89% (m/v).

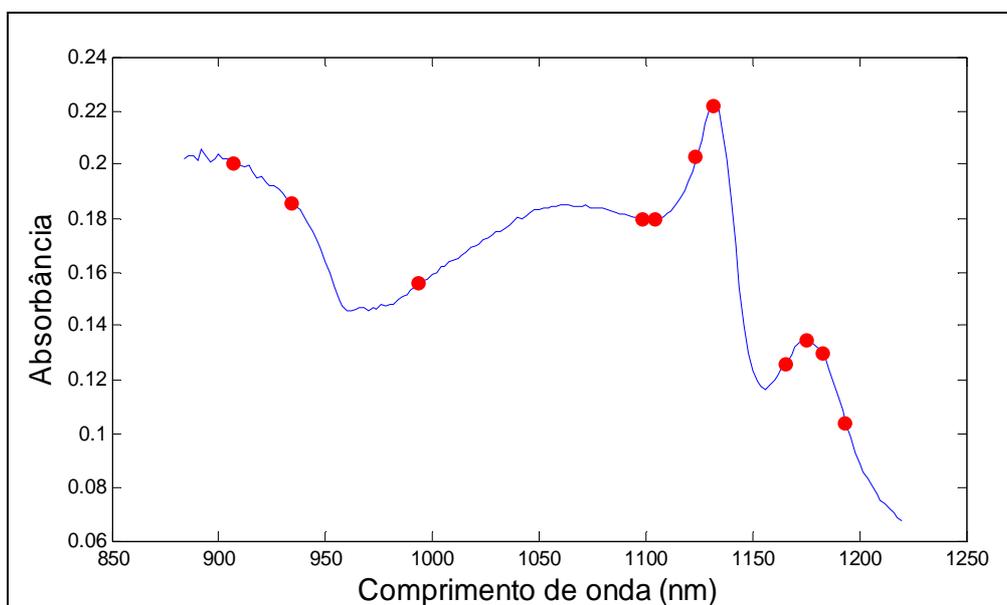


**Figura 30** - Gráfico do valor predito versus o valor de referência para o conjunto de predição do modelo *MLR-ASA*.

As amostras do conjunto de validação localizadas na **Figura 29**, e as amostras do conjunto de predição localizadas na **Figura 30**, apresentaram-se distribuídas de forma aleatória em torno da bissetriz, indicando, em ambos os casos, ausência de erro sistemático no modelo *MLR-ASA*.

### 3.6.4 Modelo *MLR-SW*

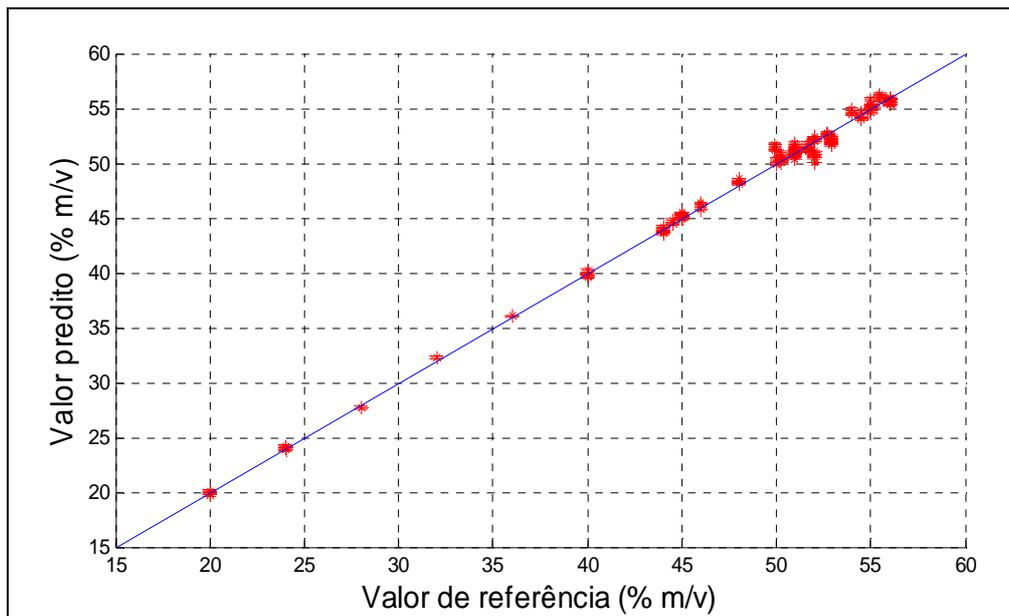
Na a construção do modelo *MLR-SW* o valor adotado para o alfa foi 0.05. O algoritmo *Stepwise (SW)* selecionou 11 variáveis espectrais, às quais foram usadas para a construção do modelo de regressão *MLR*. A **Figura 31** apresenta o gráfico do espectro *NIR* de uma amostra de dipirona sódica, indicando aproximadamente a localização das variáveis selecionadas pelo algoritmo *SW*. Além de ter selecionado poucas variáveis, estas também estão localizadas em regiões espectrais informativas (**Figura 31**).



**Figura 31** - Gráfico do espectro NIR de uma amostra de dipirona sódica, indicando aproximadamente a localização variáveis selecionadas pelo algoritmo *SW*.

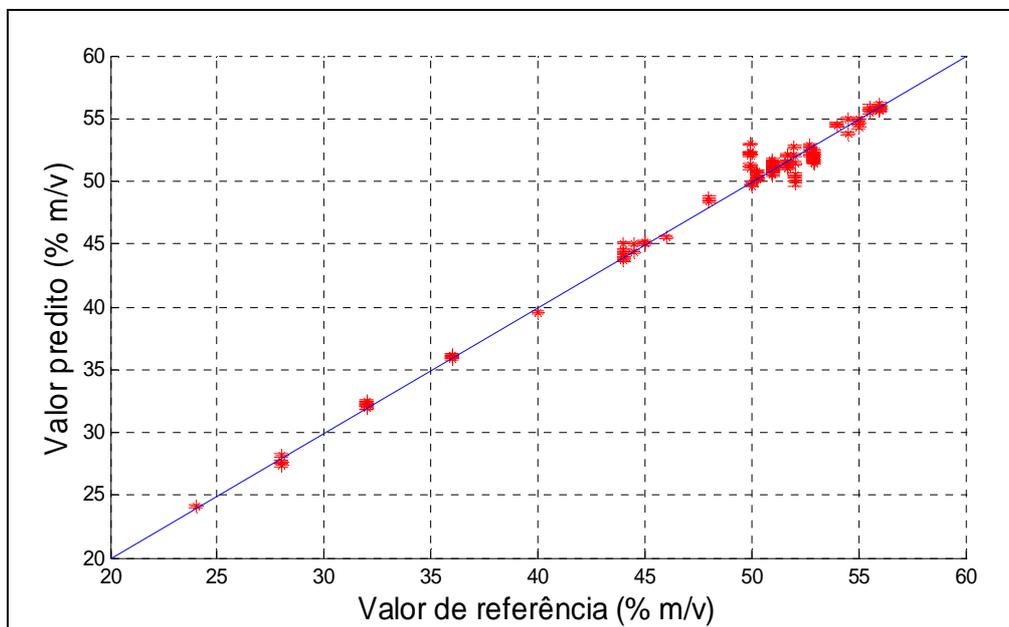
A partir das variáveis selecionadas, construiu-se e validou-se o modelo *MLR-SW*. A **Figura 32** ilustra o gráfico do valor predito versus o valor de referência das concentrações de dipirona para o conjunto de validação do modelo *MLR-SW*. O valor do coeficiente de correlação linear e do *RMSEV* obtido foram respectivamente 0,9978 e 0,56% (m/v).

As amostras do conjunto de validação (**Figura 32**) encontram-se distribuídas de forma aleatória em torno da bissetriz, sendo um indicativo da ausência de erro sistemático no modelo *MLR-SW*.



**Figura 32** - Gráfico do valor predito versus o valor de referência para o conjunto de validação do modelo *MLR-SW*.

O gráfico da **Figura 33** apresenta o valor predito versus o valor de referência das concentrações de dipirona para o conjunto de predição do modelo *MLR-SW*. O valor do coeficiente de correlação obtido e o *RMSEP* foram respectivamente iguais a 0,9940 e 0,80% (m/v).



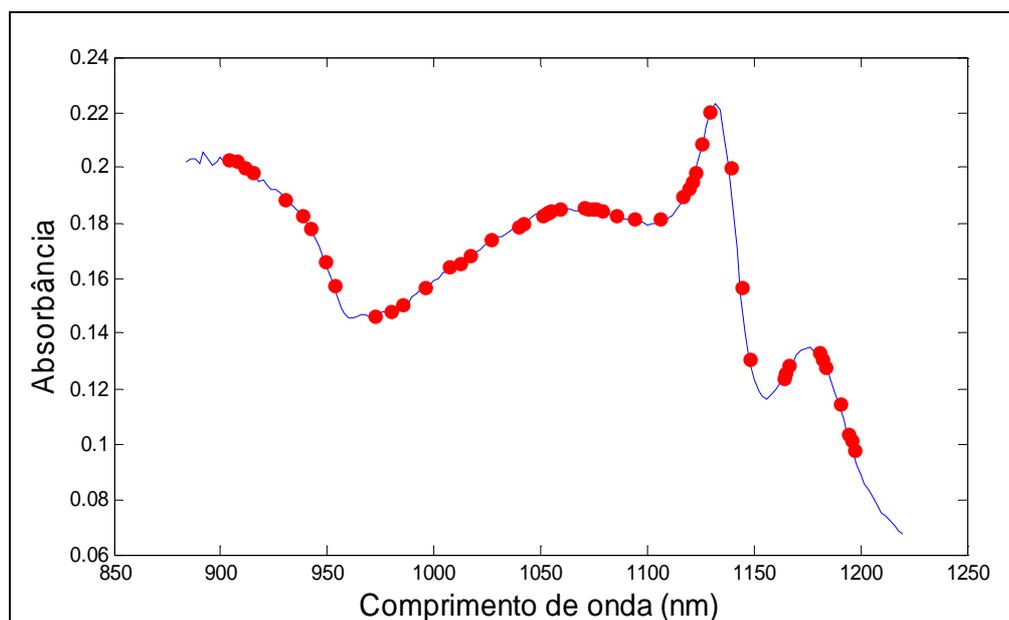
**Figura 33** - Gráfico do valor predito versus o valor de referência para o conjunto de predição do modelo *MLR-SW*.

As amostras do conjunto de predição localizadas no gráfico da **Figura 33** apresentaram-se distribuídas de forma aleatória em torno da bissetriz, sendo um indicativo ou tendência de que não ocorre erro sistemático no modelo *MLR-SW*.

### 3.6.5 Modelo MLR-GA

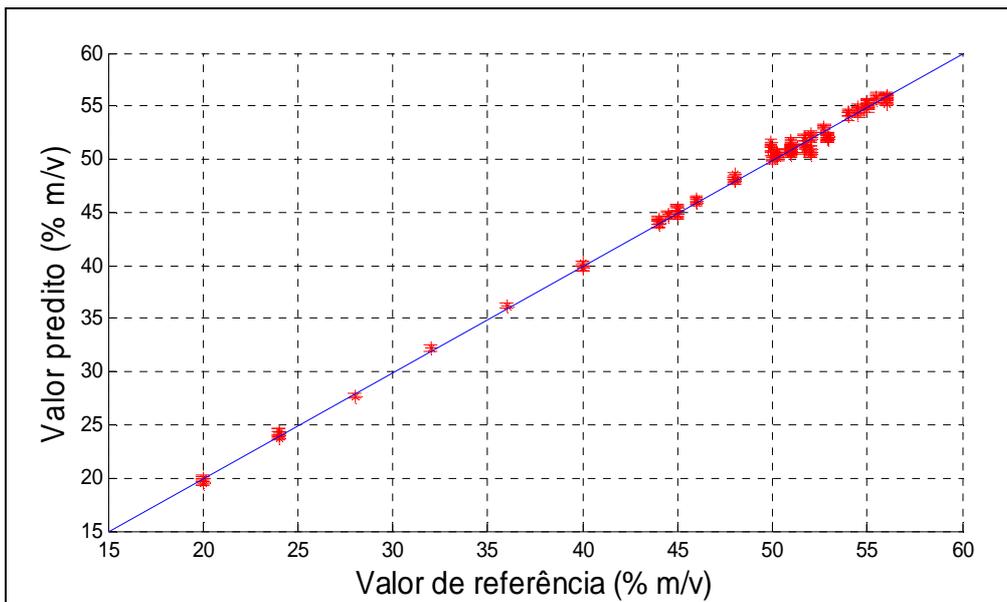
Na a construção do modelo *MLR-SW* o valor adotado para a população foi 100, para o número máximo de gerações foi 200, para a probabilidade de *crossover* foi 0.6 e para a probabilidade de mutação foi 0.05. O algoritmo genético (*GA*), foi executado (10) dez vezes e a partir de então, escolheu-se o melhor resultado para a construção do modelo *MLR*. Esta escolha se deu pelo menor *RMSEP* obtido pela regressão *MLR* efetuada sobre as variáveis selecionadas pelo algoritmo *GA*. No melhor resultado, 52 variáveis espectrais foram selecionadas pelo *GA*, sendo estas utilizadas para a elaboração do modelo de regressão *MLR*.

A **Figura 34** apresenta o gráfico do espectro *NIR* de uma amostra de dipirona sódica, indicando aproximadamente a localização das variáveis selecionadas pelo algoritmo *GA*. Neste gráfico, pode-se observar que o *GA* selecionou variáveis em regiões informativas e em regiões, aparentemente, pouco ou nada informativas. As variáveis selecionadas foram usadas na construção do modelo *MLR*.



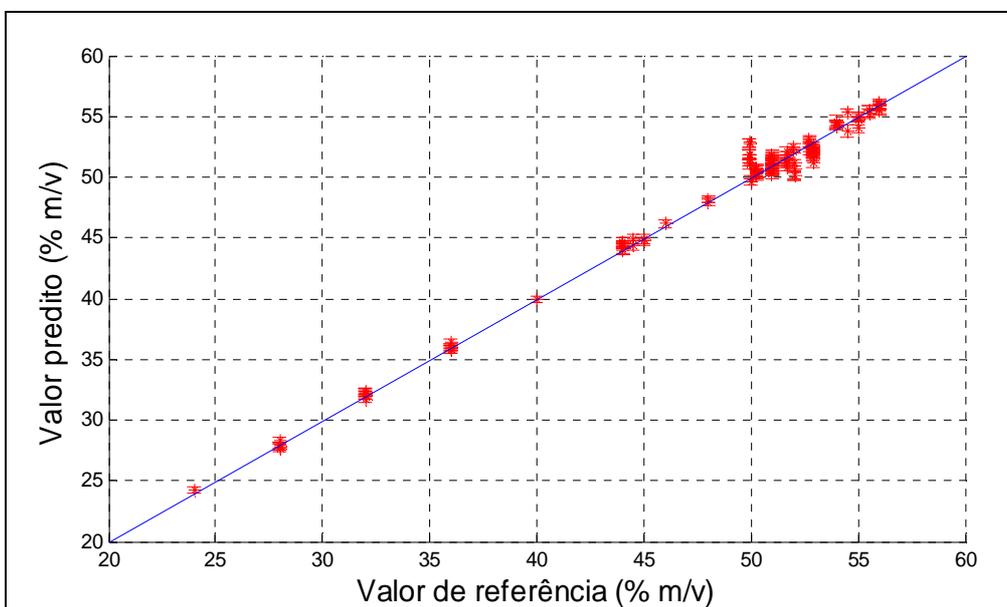
**Figura 34** - Gráfico do espectro *NIR* de uma amostra de dipirona sódica, indicando aproximadamente a localização variáveis selecionadas pelo algoritmo *GA*.

O gráfico do valor predito versus o valor de referência das concentrações de dipirona para o conjunto de validação do modelo *MLR-GA*, ilustrado na **Figura 35**, apresenta um *RMSEV* de 0,49% (m/v) e um coeficiente de correlação linear de 0.9983.



**Figura 35** - Gráfico do valor predito versus o valor de referência para o conjunto de validação do modelo *MLR-GA*.

O gráfico do valor predito versus o valor de referência das concentrações de dipirona para o conjunto de predição do modelo *MLR-GA* encontra-se na **Figura 36**. O valor da correlação obtido e o *RMSEP* foram respectivamente iguais a 0,9944 e 0,77% (m/v).



**Figura 36** - Gráfico do valor predito versus o valor de referência para o conjunto de predição do modelo *MLR-GA*.

Tanto as amostras do conjunto de validação localizadas na **Figura 35**, como as amostras do conjunto de predição localizadas na **Figura 36**, apresentaram-se distribuídas de forma aleatória em torno da bissetriz, sendo um indicativo, em ambos os casos, de ausência de erro sistemático no modelo *MLR-GA*.

Para facilitar a comparação dos resultados das previsões dos valores das concentrações de dipirona para os modelos desenvolvidos foram dispostos resumidamente na **Tabela 5**.

**Tabela 5.** Valores de *RMSEV*, *RMSEP* e correlação linear de validação e de previsão dos modelos.

Modelo	<i>PLS1</i>	<i>MLR-SPA</i>	<i>MLR-ASA</i>	<i>MLR-SW</i>	<i>MLR-GA</i>
<b><i>RMSEV</i></b> (% m/v)	0,98 (1)	0,62 (9)	0,73 (4)	0,56 (11)	0,49 (52)
<b>Correlação validação</b>	0,9938	0,9974	0,9964	0,9978	0,9983
<b><i>RMSEP</i></b> (% m/v)	1,07 (1)	0,81 (9)	0,89 (4)	0,80 (11)	0,77 (52)
<b>Correlação predição</b>	0,9904	0,9934	0,9930	0,9940	0,9944

Os parênteses (“( )”) representam o n° de variáveis selecionadas ou de variáveis latentes de cada modelo.

Todos os modelos gerados apresentaram valores de *RMSEV* e *RMSEP* muito baixos, com relação ao método de referência (50% m/v  $\pm$ 10%), além de apresentarem bons coeficientes de correlação linear na validação e na predição de seus modelos. Neste sentido, todos os modelos de calibração multivariada construídos mostraram-se adequados para a predição das concentrações de amostras de dipirona sódica injetável.

Dentre os modelos com seleção de variáveis, o *MLR-ASA* apresentou o maior erro de predição (maior *RMSEP*), entretanto, esse erro é menor que o *RMSEP* do modelo *PLS1* que utiliza todas as variáveis da faixa espectral. O modelo *MLR-ASA* necessita apenas de 4 variáveis para sua construção, enquanto que o *MLR-SW*, o *MLR-SPA* e o *MLR-GA* necessitam de muito mais variáveis, ou seja, o modelo *MLR-ASA* é mais parcimonioso.

O modelo *MLR-SW* apresentou um erro menor que o *MLR-APS*, o *MLR-ASA* e o *PLS1*, entretanto, precisou de 11 variáveis para sua construção, enquanto que o *MLR-APS* precisou de 9 e o *MLR-ASA* precisou de apenas 4 variáveis.

O modelo *MLR-GA* foi o que apresentou o menor erro de predição, entretanto foram necessárias 52 variáveis para a elaboração deste modelo. Outro fator importante a se levar em consideração, é que o *GA* é um algoritmo estocástico, ou seja, seus resultados não são reprodutíveis.

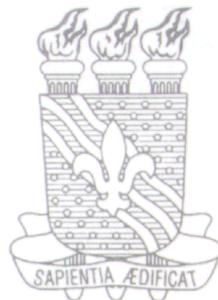
Os modelos *MLR-SPA* e *MLR-SW* apresentaram valores equiparáveis de *RMSEV*, *RMSEP* e correlação, mas vale salientar que o algoritmo *APS* é determinístico, ou seja, sempre apresenta os mesmos resultados para o mesmo conjunto de dados, enquanto que as variáveis selecionadas e o valor do *RMSEP* do método *SW* dependem do valor do alfa adotado.

Embora o modelo *ASA-MLR* tenha sido mais parcimonioso em relação aos demais modelos, o algoritmo *ASA* ainda está em fase de teste.

# **CAPÍTULO IV**

## **CONCLUSÕES**

---



## 4. CONCLUSÕES

Uma nova metodologia, não destrutiva, para a análise de medicamentos injetáveis diretamente na ampola foi desenvolvida. Ela se fundamenta no uso da espectroscopia *NIR* aliada a técnicas quimiométricas de análises para a determinação do teor de dipirona sódica injetável, com vistas a ser aplicada em métodos de controle de qualidade deste medicamento. A metodologia proposta permite determinar a concentração do princípio ativo do medicamento injetável, diretamente em sua embalagem, sem a violação da mesma. Esta metodologia apresentou as seguintes características: análise não-destrutiva e não-invasiva, ou seja, mantém a integridade das amostras comerciais. Essas características dispensam o tratamento prévio das amostras e uso de reagentes caros, tóxicos e danosos ao meio ambiente. Além disso, essa metodologia apresentou algumas vantagens, tais como: maior simplicidade e rapidez das análises.

Os modelos *SIMCA* gerados, empregando um nível de confiança de 95%, demonstraram através da análise *Screening*, que não há diferença entre os espectros *NIR* das amostras preparadas no laboratório e as amostras comerciais, sendo todas amostras apropriadas para a construção e validação dos modelos de calibração para a predição do teor de dipirona sódica em medicamentos injetáveis.

Os resultados obtidos com a construção/validação e predição dos modelos de calibração multivariada mostraram que todos os modelos gerados apresentaram erros de predição muito baixos, sendo bastante promissores para a predição das concentrações de amostras de dipirona sódica injetável.

Os modelos *MLR-SPA* e *MLR-SW* apresentaram valores equiparáveis de *RMSEV*, *RMSEP* e correlação, entretanto, sabe-se que as variáveis selecionadas e o valor do *RMSEP* obtidos pelo método *SW* podem mudar dependendo do valor do alfa adotado.

O modelo *MLR-GA* apresentou o menor erro de predição em relação aos outros modelos, porém ele não foi parcimonioso. O algoritmo *GA* selecionou 52 variáveis para a construção do modelo *MLR*, enquanto que os outros algoritmos de seleção de variáveis selecionaram uma quantidade muito inferior a esta para a construção de seus modelos de calibração (o *SPA* selecionou 9 variáveis, o *SW* selecionou 11 variáveis e o *ASA* selecionou apenas 4 variáveis).

Dentre os modelos com seleção de variáveis, o *MLR-ASA* apresentou o maior erro de predição (maior *RMSEP*), porém, o valor de *RMSEP* obtido ainda foi menor que o do modelo *PLS1* (que utiliza todas as variáveis da faixa espectral).

Como o algoritmo *ASA* selecionou apenas de 4 variáveis para a construção do modelo *MLR*, e os demais modelos utilizaram muito mais variáveis, o modelo *MLR-ASA* foi o mais parcimonioso.

#### 4.1 Propostas futuras

Pretende-se, como continuação desse trabalho:

- *Melhorar a capacidade de predição dos modelos elaborados utilizando-se outros pré-tratamentos que não chegaram a ser testados e/ou outras ferramentas de seleção de amostras, como por exemplo, o SPA para seleção de amostras;*
- *Verificar a capacidade preditiva de outros modelos quimiométricos, como o uso da regressão iPLS, por exemplo;*
- *Ampliar o número lotes e de amostras, a fim de se construir e validar o método de acordo com a norma ASTM E 1655 00, e deste modo poder sugerir esta metodologia como método oficial para controle de qualidade do teor de dipirona sódica em medicamentos injetáveis fabricados pelas Indústrias Farmacêuticas, desde que este método seja acreditado pelos órgãos responsáveis;*
- *Construir novos modelos para a determinação de outros parâmetros de qualidade, tais como: densidade, pH, viscosidade, índice de refração, etc.;*
- *Desenvolver novas metodologias para o controle de qualidade de outros medicamentos injetáveis, sem a violação da ampola, usando as mesmas estratégias;*
- *Desenvolvimento de fotômetros a base de LED's-NIR microcontrolados para análise dedicada desses medicamentos, de forma rápida, não dispendiosa e acessível às indústrias farmacêuticas, aos hospitais, clínicas médicas e farmácias e aos órgãos responsáveis pelo o controle de qualidade desses medicamentos.*

# **CAPÍTULO V**

## **REFERÊNCIAS BIBLIOGRÁFICAS**

---



---

---

## 5. REFERÊNCIAS BIBLIOGRÁFICAS

1. Trissel, L. A. *Handbook on injectable drugs*. 10ed. Houston, TX-USA: 1998.
2. Brasil. Portaria n.º 802, de 08 de outubro de 1998. Institui o sistema de controle e fiscalização em toda a cadeia dos produtos farmacêuticos. *Diário Oficial da República Federativa do Brasil*. Brasília, v.1, n.24-E, p. 036, 4 de fevereiro de 1999. Seção 1, pt 1.
3. Conselho Federal de Farmácia, *A Organização Jurídica da Profissão Farmacêutica*, 2ª edição, Editora Cidade Gráfica LTDA, 2000, 1396p.
4. Organização Mundial de Saúde. *The importance of Pharmacovigilance. Safety monitoring of medicinal products*. 2002.
5. Jornal Diário de São Paulo, publicado em 7 de abril de 2009. São Paulo.
6. Jornal O Dia, publicado em 15 de março de 2009. Rio de Janeiro.
7. Jornal O Estado de São Paulo, publicado em 8 de setembro de 2009. São Paulo.
8. ANVISA, Informes Técnicos Institucionais, Rev. Saúde Pública 2004, 38(5):748-9.
9. Site: <http://www.redenoticia.com.br/noticia/?p=17928>, publicado em 19 novembro de 2009, acessado em 05 de dezembro de 2009.
10. <http://www.anvisa.gov.br/DIVULGA/noticias/2009/160409.htm>, publicado em 14 de abril de 2009, acessado em 05 de dezembro de 2009.
11. United States Pharmacopoeia 25, NF20, Rock Ville, 2002.
12. Gilman A. G; Hardman J. G.; Limbird L. E., *As Bases Farmacológicas da Terapêutica*, Rio de Janeiro, McGraw-Hill, 2003.
13. Ciurczak, E. W., *Appl. Spectrosc. Rev.* 23 (1987) 147.
14. Burns, D. A.; Ciurczak, E.W., *Hand-Book of Near-infrared analysis - Second Edition, Revised and Expanded*, Marcel Dekker, Inc, N. York, 2001.
15. Pasquini, C. , *J. of Braz. Chem. Soc.*, 14 (2003) 198.
16. Workman, J. J.; Mobvley, P. R.; Kowalski, B. R. , *Appl. Spectrosc. Reviews*, 31 (1996) 73.
17. Roggo, Y.; Chalus, P.; Maurer, L.; Martinez, C. L.; Edmond, A.; Jent, N., *J. Pharm. Biomed. Analysis* 44 (2007) 683.
18. Beebe, K.R.; Pell, R.J.; Seasholtz, B., *Chemometrics - A Practical Guide*, Wiley, Nova Iorque, 1998.
19. Forina, M.; Lanteri, S.; Casale, M., *Multivariate calibration- Review. J. Chromatography A*, 1158 (2007) 61.

20. Prista, L. N.; Alves, A. C.; Morgado, R. M. R., *Técnica Farmacêutica e Farmácia Galênica*, volume III, 3ª edição, Fundação Calouste Gulbenkian/Lisboa, 1990.
21. Prista, L. N.; Alves, A. C.; Morgado, R. M. R., *Tecnologia Farmacêutica*, volume III, 4ª edição, Fundação Calouste Gulbenkian/Lisboa, 1995.
22. Schwidetzky O. History of needles and syringes. *Anesth Analg* 1944; 23:34-38.
23. Howard-Jones N. A critical study of the origins and early development of hypodermic medication. *J Hist Med* 1947; 2:201-249.
24. Howard-Jones N. The origins of hypodermic medication. *Scientific American* 1971; 224:96-102.
25. Haller JS. Hypodermic medication. *NYS J Med* 1981; 1671-1679.
26. Brunton D. A question of priority: Alexander Wood, Charles Hunter and the hypodermic method. *Proc R Coll Physicians Edinb* 2000; 30:349-351.
27. DAVISON MH, The evolution of anaesthesia. *Br J Anaesth.* 1958 Mar;30(3):142-7 contd.
28. History of injections. Pictures from the history of otorhinolaryngology highlighted by exhibits of the German History of Medicine Museum in Ingolstadt by Feldmann H. *HNO-Klinik, Universität Munster. Laryngorhinootologie.* 2000 Apr;79(4):239-46.
29. Lachman, L.; Lierman, H. A.; Kaning, J. L., *Teoria e prática na indústria farmacêutica*. Volume II, Fundação Calouste Gulbenkian/Lisboa, 2001.
30. Prista, L. N.; Alves, A. C.; Morgado, R. M. R., *Tecnologia Farmacêutica*, volume II, 4ª edição, Fundação Calouste Gulbenkian/Lisboa, 1995.
31. Thompson, J. E., *A prática farmacêutica na manipulação de medicamentos*, Artmed, 2006.
32. Burke, A.; Smyth, E., Fitzgerald G.A. - Analgesic-Antipyretic and Antiinflammatory Agents; Pharmacotherapy of Gout, em: Brunton, L.L., Lazo, J.S., Parker, K.L. - *Goodman & Gilman's The Pharmacological Basis of Therapeutics*, McGrawHill 11<sup>th</sup>, N.Y., 2005, p 671.
33. Brack, A.; Rittner, H. L.; Schafer, M., *Anaesthesist.* 53 (2004) 263.
34. Arellano, F.; Sacristán, J. A., *Eur. J. Clin. Pharmacol.* 38 (1990) 617.
35. Doretto, M.; Garcia-Cairasco, N.; Pimenta, N.J. et al., *Neuroreport.* 9 (1998) 2415.
36. Sabatowski, R.; Schafer, D.; Kasper, S. M. et al., *Pharm. Design.* 10 (2004) 701.
37. Auer, A., *Ther D Gegenw.* 72 (1931) 154.
38. Brogden, R.N., *Drugs.* 32 (1986) 60. Suppl 4.
39. International Agranulocytosis and Aplastic Anemia Study (The Boston Study). *J Am Med Assoc.* 256 (1986) 1749.

40. Hamerschlak, N. ; Maluf, E., Paquini, R. et al., *Med. J.*, 123 (2005) 101.
41. Cavalcanti, I. L.; Cantinho, F. A. F; Assad, A., *Medicina Perioperatória*, SAERJ- Sociedade de Anestesiologia do Estado do Rio de Janeiro, Rio de Janeiro, 2006.
42. Benseñor, M. I., *Med. J.*, 119 (200 1) 190.
43. Kehlet, H., *Br. J. Anaesth* . 63 (1989) 189.
44. Farmacopéia Brasileira. 4ª Edição, janeiro de 2000.
45. British Pharmacopoeia, Volume I, December 2000.
46. Vassileva-Alexandrova, P.; Shishmanov, P. S., *Anal. Chem.*, 47 (1975) 1432.
47. Sakiara, K. A.; Pezza, L.; Melios, C. B.; Pezza, H. R.; De Moraes, M.; *Il Fármaco*. 54 (1999) 629.
48. Pereira, A. V.; Penckowski, L.; Vosgerau, M.; Sassá, M. F.; Fatibello Filho, O., *Quim. Nova*. 25 (2002) 553.
49. Senyuva, H. Z.; Aksahin, I.; Ozcan, S.; Kabasakal, B. V., *Anal. Chim. Acta*. 547 (2005) 73.
50. Perez-Ruiz, T.; Martinez-Lozano, C.; Tomas, V.; Carpena, J.; *Microchem. J.*, 47( 1993) 296.
51. Qureshi, S. Z.; Saeed, A.; Haque, S.; *Microchem. J.*, 41 (1990) 362.
52. Qureshi, S. Z.; Saeed, A.; Hassan, T.; *Talanta*, 36 (1989) 869.
53. Marcolino-Jr., L. H.; Bonifácio, V. G.; Fatibello-Filho, O., *Quim. Nova*, 28 (2005) 783.
54. Pereira, A. V.; Penckowski, L.; Vosgerau, M.; Sassa, M. F.; Fatibello-Filho, O., *Quim. Nova*, 25 (2002) 553.
55. Buhl, F.; Hachula, U., *Chem. Anal.*, 29 (1984) 483.
56. Buhl, F. Hachula, U., *Chem. Anal.*, 26 (1981) 395.
57. Vassileva-Alexandrova, P., *Mikrochim. Acta*, 2 (1976) 247.
58. Abdine, H.; Soliman, S. A.; Morcos, M. G., *J. Pharm. Science*, 62 (1973) 1834.
59. Perez-Ruiz, T.; Martinez-Lozano, C.; Tomas, V., *J. Pharm. Biomed. Anal.*, 12(1994) 1109.
60. Matos, R. C.; Angnes, L.; Araujo, M. C. U.; Saldanha, T. C. B., *Analyst*, 125 (2000) 2011.
61. Munoz, R. A. A.; Matos, R. C.; Angnes, L.; *J. Pharm. Sci.*, 90 (2001) 1972.
62. Huang, Y. M.; Zhang, C.; Zhang, X. R.; Zhang, Z. J.; *J. Pharm. Biomed. Anal.*, 21 (1999) 817.
63. Huang, Y. M.; Zhang, C.; Zhang, X. R.; Zhang, Z. J. F., *J. Anal. Chem.*, 365 (1999) 381.

64. Huang, Y. M.; Zhang, C.; Zhang, X. R.; Zhang, Z. J., *Anal. Lett.*, 32 (1999) 933.
65. Agundez, J. A. G.; Benitez, J.; *Ther. Drug Monit.*, 18 (1996) 104.
66. Tulaganov, A. A.; Mikhalev, A. V.; Osokina, L. M.; Arzamastsev, A. P., *Khimiko-Farm. Zh.*, 26 (1992) 76.
67. Agundez, J. A. G.; Martinez, C.; Martin, R.; Benitez, J., *Ther. Drug Monit.*, 16 (1994) 316.
68. Katz, E. Z.; Granit, L.; Drayer, D. E.; Levy, M., *J. Chromatogr.*, 305 (1984) 477.
69. Eddine, N. H.; Bressolle, F.; Mandrou, B., *Analyst*, 107 (1982) 67.
70. Skoog, D.A.; Holler, F.J; Nieman, T.A.. *Análise Instrumental*, 5ª Edição, Bookman, 2006.
71. Willard, H.; Merritt, L.; Dean, Jr. J., *Análise Instrumental*. 2ª Edição, Fundação Calouste Gulbenkian/Lisboa, 1984.
72. Burns, D. A.; Ciurczak, E. W., *Handbook of Near-Infrared Analysis*, 3rd ed,(practical spectroscopy, 35), CRC Press Taylor & Francis Group, 2008.
73. Silva, E. A.; Araújo, M. C. U.; Santos, S. R. B.; Vêras, G., *Quim. Nova*, 20 (1997), 2, 37.
74. Gaião, E. N.; Santos, S. R. B.; Santos, V. B.; Nascimento, E. C. L.; Lima, R. S.; Araújo, M. C. U., *Talanta*, 2007.
75. Scaffi, S.H.F.; Pasquini, C., *Analyst*, 126 (2001) 2218.
76. Blanco, M.; Eustaquio, A.; González, J. M.; Serrano, D., *J. Pharm. Biom. Analysis*, 22 (2000) 139.
77. Trafford, A. D.; Lee, R. D.; Moffat, A. C.; Graham, P., *Analyst*, 124 (1999) 163.
78. Blanco, M.; Coello, J.; Iturriaga, H.; Maspoch, S.; Pezuela, C., *Analyst*, 122 (1997) 761.
79. Wetzel, D. L., *Anal. Chem.*, 55 (1983) 1165A.
80. FEEMA, *Nitrogênio Kjeldahl total (micro Kjeldahl)*. Manual do Meio Ambiente Volume II – Métodos FEEMA, R.J., MF 421.R2, p.45. 1983.
81. Baiulescu, G. E.; Stoica, A. I., *Analytical Letters*, 37 (2004) 3105.
82. Mansfield, C. T.; Barman, B. N., *Anal. Chem.*, 71 (12) (1999) 81R.
83. Corti, P.; Savini, L.; Dreassi, E.; Ceramelli, G.; Montecchi L.; Lonardi, S., *Pharm. Acta Helv.*, 67 (1992) 57.
84. Dreassi, E.; Celesti, L.; Ceramelli, G.; Savini, L.; Corti, P., *Pharm. Acta Helv.*, 67 (1992) 341.
85. Ciurczak, E.W; Maldacker, T.A., *Spectroscopy*, 1(1986) 36.
86. Candolfi, A.; Massart, D.L.; Heuerding, S., *Anal. Chim. Acta*, 345 (1997) 185.

- 
- 
87. Dempster, M.A. et al., *J. Pharm. Biomed. Anal.*, 11(1993) 1087.
  88. Moffat, A.C. et al., *The Analyst*, 125 (2000) 1341.
  89. Harmonisation, I.C.o., ICH harmonised tripartite guideline-text on validation of analytical procedures. *Federal Register*, 60 (1995) 11260.
  90. Harmonisation, I.C.o., ICH harmonised tripartite guideline-validation of analytical procedures: Methodology. *Federal Register*, 60 (1995) 27463.
  91. Workman, J. J.; Mobvley, P. R.; Kowalski, B. R.; Bro R., *Appl. Spectrosc. Reviews*, 31 (1996) 73.
  92. Ferreira, M. M. C.; Antunes, A. M.; Melgo, M. S.; Volpe, P. L. O. Quimiometria I: calibração multivariada, um tutorial, *Quím. Nova* vol.22 n.5. São Paulo, 1999.
  93. <http://pt.wikipedia.org/wiki/Quimiometria>, acessado em 25 de novembro de 2009.
  94. Massart, D.L.; Vandeginste, B.G.M.; Buydens, S.J.; Lewi, P.J.; Smeyers-Verbeke, J., *Handbook of Chemometrics and Qualimetrics: Parte B*, Elsevier, Amsterdam, 1997
  95. Brereton, R.G., *Chemometrics: data analysis for the laboratory and chemical plant*. Ed. Wiley. University of Bristol, UK, 2003.
  96. Urabano, M.; Luque de Castro, M.D.; Pérez, P.M.; García-Olmo, J.; Gómez-Nieto, M.A., *Food Chemistry*, 97 (2006) 166.
  97. Ni, Y.; Qi, Z.; Kokot, S., *Chemometrics and Intelligent Laboratory Systems* 82 (2006) 241.
  98. Stuart, A. D.; Trotman S. M.; Doolan, K. J.; Frederiks, P. M., Spectroscopy measurement of used lubricating oil quality, *appl. Spectrosc.* 43 (1995) 55.
  99. CAMO S.A. *Manual do Usuário. UNSCRAMBLER 7.5*. Noruega. 1998.
  100. Pantoja, P. A., Dissertação de mestrado intitulada: *Aplicação da espectroscopia de infravermelho próximo na caracterização da carga de petróleo para o processo de destilação atmosférica*. USP, São Paulo/SP, novembro de 2006.
  101. Savitzky, A.; Golay, M. J. E., *Analytical Chemistry*, 8 (1964) 1627.
  102. Pearson, K., On lines and planes of closest fit to systems of points in space. *Philosophical Magazine* 2 (1901) 559.
  103. Jolliffe, I.T., *Principal Component Analysis*. Springer Verlag – New York Inc, 2002.
  104. Hotelling, H., *Analysis of a complex of statistical variable into principal components*; J. Educ. Psych., vol. 24 (1933) 417.
  105. Brown, S.D., *Appl. Spectrosc.*, Baltimore, v.49, n.12, p.14A, 1995.
  106. Sabin, J. A.; Ferrão, M. F.; Furtado, J. C., *Brazilian Journal of Pharmaceutical Sciences*, vol. 40, n. 3, 2004.

- 
- 
107. Wold, S. *Principal Component Analysis. Chemometrics and Intelligent Laboratory Systems*, 2 (1987) 37.
  108. Marques, G., *Análise em Componentes Principais, Apostila de Processamento Estatístico de Sinais*, 2006.
  109. Shlens, J., *A Tutorial on Principal Component Analysis*, Center for Neural Science, New York University New York City, NY 10003-6603 and Systems Neurobiology Laboratory, Salk Institute for Biological Studies La Jolla, CA 92037. Version 3.01, 2009.
  110. Smith, L.I., *A tutorial on Principal Components Analysis*, February 26, 2002, baseado em Howard Anton, Publisher John Wiley & Sons Inc, ISBN 0-471-85223-6.
  111. Wold, S., Pattern recognition by means of disjoint principal component models, *Pattern Recognition*, 8 (1976) 12.
  112. Sjöström e Kowalski, "A comparison of five pattern recognition methods based on the classification results from six real data bases". *Anal. Chim. 112* (1979) 11.
  113. Branden, K.V., Hubert, M. *Robust classification in high dimensions based on the SIMCA Method. Chemometrics and Intelligent Laboratory Systems* 79 (2005) 10.
  114. Candolfi, A.; Maesschalck, R. D.; Massart, D. L.; Hailey, P. A.; Harrington, A. C. E., *Journal of Pharmaceutical and Biomedical Analysis*, 19 (1999) 923.
  115. He, J.; Rodriguez-Saona, L. E.; Giusti, M. M., *Journal of Agricultural and Food Chemistry*, 55 (2007) 4443.
  116. Wang, L.; Lee, F. S. C.; Wang, X.; He, Y., *Food Chemistry*, 95 (2006) 529.
  117. Vogt, N. B.; Sjoegren, C. E., *Analytica Chimica Acta*, 222 (1989) 135.
  118. Trullols, E.; Ruisanchez, I.; Rius, F. X., *Trends in Anal. Chem.*, 24 (2004) 137.
  119. Valcarcel, M.; Cáardenas, S.; Gallego, M., *Trends in Anal. Chem.*, 18 (1999) 685.
  120. Song, R.; Schelcht, P. C.; Ashley, K., *J. Hazardous Materials*, 83 (2001) 29.
  121. Valcárcel, M. *Principios de química analítica*, Springer-Verlag Iberica, Madrid, 1999.
  122. Medeiros, V. M.; Araújo, M. C. U.; Galvão, R. K. H.; Silva, E. C.; Saldanha, T. C. B.; Toscano, I. A. S.; Oliveira, M. S. R.; Freitas, S. K. B.; Neto, M. M., *Water Research*, 39 (2005) 3089.
  123. Pulido, A.; Ruisánchez, I.; Bóque, R.; Rius, F. X. *Trends in Anal. Chem.*, 22 (2003) 647.
  124. Pulido, A.; Ruisánchez, I.; Bóque, R.; Rius, F. X., *Anal. Chim. Acta*, 455 (2002) 267.
  125. Rios, A.; Barcelo, D.; Buydens, L.; Cardenas, S.; Heydorn, K.; Karlber, B.; Klemm, K.; Lendl, B.; Milman, B.; Neidhart, B.; Stephany, R. W.; Townshend, A.; Zschunke, A.; Valcárcel, M., *Accreditation and Quality Assurance*, 8 (2003) 68.

- 
- 
126. Sabin, J.G *et al.*, “Análise exploratória aplicada na identificação de fármacos antidepressivos. Parte I: análise hierárquica de agrupamentos”. *Tecno-lógica*, 7 (2003) 69.
  127. Naes, T.; Isaksson, T.; Fern, T.; Davies, T., *Multivariate calibration and clasification*. NIR Publications, 2002.
  128. Daszykowski, M.; Walczak, B.; Massart, D. L., *Anal. Chim. Acta*, 468 (2002) 91.
  129. Tominaga, Y., *Chemometrics. Intelligent Laboratory System*, 43 (1998) 157.
  130. Sales, F.; Rius, A.; Callao, M.P.; Rius, F.X., *Talanta*, 52 (2000) 329.
  131. Wu, W.; Walczak, B.; Massart, D.L.; Heuerding, S.; Erni, F.; Last, I.R.; Prebble, K.A., *Chemometrics and Intelligent Laboratory System*, 33 (1996) 35.
  132. Kennard, R.W.; Stone, L. A., *Technometrics*, 11 (1969) 137.
  133. Galvão, R.K.H. *et al.*, *Talanta* 67 (2005) 736.
  134. Wold, H., *Soft Modeling by Latent Variables; the Non-linear Iterative Partial Least Squares Approach, in Perspectives in Probability and Statistics*, Ed. J. Gani, Academic Press, London, 1975.
  135. Geladi, P. Kowalski, B. R., Partial Least Squares Regression: A tutorial. *Anal. Chim. Acta*, 185 (1986) 1.
  136. Jackson, J. E., *User’s Guide to Principal Componentes*, Wiley, New York, 1991.
  137. Hair, J. F.; Tatham, R. L.; Anderson, R. E.; Black, W. *Análise Multivariada de Dados*. 5ª ed. Bookman, Porto Alegre, 2005.
  138. H. Martens, T. Naes, *Multivariate calibration*, Wiley, New York, 1996.
  139. M. Otto, *Chemometrics - Statistics and Computer application in Analytical Chemistry*, Willey-VCH, Berlin, 1999.
  140. Haaland, D. M.; Thomas, W. V., Partial least-squares methods for spectral analyses. Relation to other quantitative calibration methods and the extraction of qualitative information. *Anal. Chem.* 60 (1988) 11.
  141. Stuart, A. D.; Trotman S. M.; Doolan, K. J.; Frederiks, P. M., Spectroscopy measurement of used lubricating oil quality, *appl. Spectrosc.* 43 (1995) 55.
  142. Sternberg, J. C.; Stills, H. S.; Schwendeman, R. H., *Anal. Chem.*, 32 (1960) 84.
  143. Nunes, P. G. A., Tese de doutorado intitulada: *Uma nova técnica para seleção de variáveis em calibração multivariada aplicada às espectrometrias UV-VIS e NIR*. UFPB, João Pessoa/PB, Março de 2008.
  144. Nunes, P. G. A. ; Silva, E. C., Análise de umidade e proteína de uma amostra a partir de dados de espectrometria NIR usando o algoritmo ASA-MLR, *III JORNADA NACIONAL DA AGROINDÚSTRIA*. Bananeiras, agosto de 2008. ISSN 1980-1122.

- 
- 
145. Holland, J.H. *Adaptation in natural and artificial systems*. MIT Press, Ann Arbor, Michigan, 1975.
  146. Goldberg, David E.. *Genetic Algorithms in Search, Optimization, and Machine Learning*. EUA: Addison-Wesley, 1989.
  147. Koza, J.R.. *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. MIT Press, 1992.
  148. Poli, R., Langdon, W. B., McPhee, N. F.. *A Field Guide to Genetic Programming*. freely available via Lulu.com, 2008.
  149. Araújo, M. C. U.; Saldanha, T. C. B.; Galvão, R. K. H.; Yoneyma, T.; Chame, H. C.; Visani, V., *Chemometrics and Intelligent Laboratory Systems*, 57 (2001) 65.
  150. Dantas Filho, H. A.; Souza, E. S. O. N.; Visani, V.; Barros, S. R. R. C.; Saldanha, T. C. B.; Araújo, M. C. U.; Galvão, R.K.H., *J. Braz. Chem. Soc.*, 16 (2005) 58.
  151. Galvão, R. K. H.; Pimentel, M. F.; Araújo, M.C.U.; Yoneyama, T.; Visani, V., *Anal. Chim. Acta* 443 (2001) 107.
  152. Honorato, F. A.; Galvão, R. K. H.; Pimentel, M. F.; Neto, B. B.; Araújo, M. C. U.; Carvalho F. R., *Chemom. Intell. Lab. Syst.* 76 (2005) 65.
  153. Breitzkreitz, M. C.; Raimundo, I. M. Jr; Rohwedder, J. J. R.; Pasquini, C.; Dantas Filho, H. A.; José, G. E.; Araújo, M. C. U., *Analyst* 128 (2003) 1204.
  154. Dantas Filho, H. A. D.; Galvão, R. K. H.; Araújo, M. C. U.; Silva, E. C.; Saldanha, T. C. B.; José, G. E.; Pasquini, C.; Raimundo Jr, I. M.; Rohwedder, J. J. R., *Chemom. Intell. Lab. Syst.* 72 (2004) 83.
  155. Akhlaghi, Y.; Kompany-Zareh M.; *J. Chemometrics* 20 (2006) 1.
  156. Pontes, M. J. C.; Galvão, R. K. H.; Araújo, M. C. U.; Moreira, P. N. T.; Pessoa Neto, O. D.; José, G. E.; Saldanha, T. C. B., *Chemom. Intell. Lab. Syst.* 78 (2005) 11.
  157. Chong, I.; Jun, C., *Chemometrics and Intelligent Laboratory Systems*, 78 (2005) 103.
  158. Hocking, R. R., "The Analysis and Selection of Variables in Linear Regression," *Biometrics*, 32 (1976).
  159. Caneca, A. R.; Pimentel, M. F.; Galvão, R. K. H.; Matta, C. E.; Carvalho, F. R.; Raimundo JR, I. M.; Pasquini, C.; Rohwedder, J. J. R., *Talanta*, 70 (2006) 344.
  160. Draper, N. and Smith, H., *Applied Regression Analysis, 2d Edition*, New York: John Wiley & Sons, Inc, 1981.
  161. Duda, R. O.; Hart, P. E.; Stork, D. G. *Pattern Classification*, New York: John Wiley, 2001.
  162. Wilkinson, L.; Dallal, G.E., "Tests of significance in forward selection regression with an F-to enter stopping rule." *Technometrics*. 23(1981) 377.

163. Workman Jr., J.; Weyer, L., *Practical guide to interpretative near-infrared spectroscopy*. CRC-Press-Taylor & Francis Group. 1996.
164. Kelly, J. J.; Barlow, C. H.; Jinguji, T. M.; Callis, B., *Anal. Chem.*, 61 (1989) 313.