



*Universidade Estadual de Goiás*

*Unidade Universitária de Ciências Exatas e Tecnológicas*

**PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS MOLECULARES**

## ***DISSERTAÇÃO DE MESTRADO***

**Desenvolvimento e Validação de Método  
Quimiométrico para Determinação de Amoxicilina  
em Formulação Farmacêutica**

**Aluno: Maurício Alves de Melo Silva**

**Orientador: Prof. Dr. Marcelo Martins de Sena**

**Co-Orientador: Prof. Dr. Jez Willian B. Braga (IQ/UnB)**

***Anápolis***

***Março/2011***



# **Desenvolvimento e Validação de Método Quimiométrico para Determinação de Amoxicilina em Formulação Farmacêutica**

Dissertação apresentada ao Programa de Pós Graduação em Ciências Moleculares da Universidade Estadual de Goiás como parte dos requisitos para a obtenção do título de Mestre em Ciências Moleculares.

**Aluno: Maurício Alves de Melo Silva**

**Orientador: Prof. Dr. Marcelo Martins de Sena**

**Co-Orientador: Prof. Dr. Jez Willian B. Braga (IQ/UnB)**

***Anápolis***

***Março/2011***



*"Tudo o que existe no Universo é fruto do Acaso e da Necessidade".*

Demócrito

*“ O amor é grande e cabe nesta janela sobre o  
mar. O mar é grande e cabe na cama e no  
colchão de amar. O amor é grande e cabe no  
breve espaço de beijar. ”*

Carlos Drummond de Andrade

*Rosi Graciele,*

*Obrigado, mas muito obrigado, por ser  
esta pessoa maravilhosa, que sempre está ao  
meu lado me falando uma palavra de apoio, me  
incentivando e cuidando de mim que Deus nos  
faça cada dia mais amigos e apaixonados.*

*Obrigado por tudo!*

*Te amo*

## Agradecimento

- Agradeço primeiramente a Deus por me dar a oportunidade de estar findando mais esta etapa tão importante da minha vida;
- Ao meu orientador Marcelo Martins de Sena meus sinceros agradecimentos pelos conhecimentos transmitidos que contribuíram e muito para minha formação;
- Ao meu co-orientador Jez Willian B. Braga por toda experiência passada me ajudando sempre que possível;
- Ao amigo Marcus Henrique que me ajudou em toda parte experimental junto à IQUEGO e também pelas boas discussões e aos estagiários da IQUEGO que me ajudaram nos experimentos;
- À IQUEGO por ceder seu laboratório, matéria prima e equipamentos para realização deste trabalho, enfim a todos que contribuíram direta e indiretamente para conclusão deste trabalho;
- A minha querida esposa, Rosi Graciele, que contribuiu de todas as formas para conclusão deste trabalho;
- A meus pais pela confiança em mim depositada e incentivo durante todo o processo.
- Aos meus amigos da graduação Francisco Nunes, Tarcísio Carvalho e Alex Wilkerson pelas conversas e discussões principalmente na disciplina de termodinâmica;
- À Cristina D. Bernardes que me acolheu tão bem em sua casa em Belo Horizonte e pelas conversas;
- À Unidade Universitária de Ciências Exatas e Tecnológicas (UnUCET) pela estrutura, pois sem a mesma não teria concluído o mestrado;
- Enfim, a todos que de alguma forma contribuíram para a conclusão deste trabalho.

## RESUMO

Este trabalho propôs um novo método para determinação de amoxicilina numa formulação farmacêutica (suspensão), baseado em medidas de transfectância por espectroscopia no Infravermelho Próximo e calibração multivariada usando o método dos Mínimos Quadrados Parciais. Uma metodologia completa foi implementada para o desenvolvimento do método proposto, incluindo um planejamento experimental com dois fatores independentes (amoxicilina e placebo), o processamento de dados usando Correção do Espalhamento Multiplicativo, a seleção de amostras pelo algoritmo de Kennard-Stone, e a detecção de amostras anômalas baseada em valores extremos de influência ou de resíduos nos dados espectrais (X) e nos valores de concentração previstos (Y). O melhor modelo foi obtido com sete variáveis latentes na faixa de 40,0 a 65,0 mg mL<sup>-1</sup> de amoxicilina, fornecendo uma Raiz Quadrada Média dos Erros de Previsão de 1,6 mg mL<sup>-1</sup>. O método proposto foi validado de acordo com normas Brasileiras e internacionais, através da estimativa de figuras de mérito, tais como linearidade, precisão, exatidão, robustez, seletividade, sensibilidade analítica, limites de detecção e quantificação e detecção de erros sistemáticos (viés). Uma curva de calibração pseudo-univariada também foi obtida baseada no Sinal Analítico Líquido. O método quimiométrico proposto apresenta as vantagens de rapidez, simplicidade, baixo custo, não uso de solventes e não geração de resíduos químicos, se comparado aos principais métodos propostos para a determinação de amoxicilina, os quais são baseados em Cromatografia Líquida de Alta Eficiência (CLAE).

## ABSTRACT

This work proposed a new method for determination of amoxicillin in suspension pharmaceutical formulations, based on transmittance NIR measurements and partial least squares multivariate calibration. A complete methodology was implemented for developing the proposed method, including experimental design with two independent factors (amoxicillin and placebo), data preprocessing by using multiple scatter correction, sample selection by Kennard-Stone algorithm, outlier detection based on high values of leverage, and X and Y residuals. The best model was obtained with seven latent variables in the range from 40.0 to 65.0 mg ml<sup>-1</sup> of amoxicillin, providing a root mean square error of prediction of 1.6 mg ml<sup>-1</sup>. The method was validated in accordance with Brazilian and international guidelines, through the estimative of figures of merit, such as linearity, precision, accuracy, robustness, selectivity, analytical sensitivity, limits of detection and quantitation, and bias. A pseudo-univariate calibration curve was also obtained based on the net analyte signal. The proposed chemometric method presented the advantages of rapidity, simplicity, low cost, no use of solvents and no chemical waste generation over the main alternative methods based on high performance liquid chromatography.

## SUMÁRIO

1. Introdução.....	1
1.1. Calibração Multivariada.....	2
1.1.1. Mínimos Quadrados Parciais .....	3
1.2. Validação Analítica de Métodos Quimiométricos.....	4
1.2.1. Sinal Analítico Líquido.....	5
1.2.2. Figuras de Mérito.....	8
1.3. Detecção de Amostras anômalas.....	13
1.4. Espectroscopia no Infravermelho Próximo.....	15
1.5. Amoxicilina.....	17
2. Procedimento.....	20
2.1. Equipamentos e Processamento dos Dados.....	20
2.2. Materiais, Reagentes e Suspensões.....	21
2.3. Metodologia.....	21
2.3.1. Estudos preliminares: medidas de reflectância difusa.....	21
2.3.2. Planejamento Experimental.....	22
2.3.3. Preparo das Amostras.....	26
2.3.4. Análise Cromatográfica.....	27
3. Resultados e Discussão.....	28
3.1. Estudos Preliminares: Medidas de Reflectância Difusa.....	28
3.2. Modelo de Calibração Multivariada com Medidas de Transreflectância	30
3.3. Detecção de Amostras Anômalas e Previsão das Amostras de	34
Validação.....	
3.4. Validação Analítica.....	37
3.4.1. Especificidade/Seletividade.....	38
3.4.2. Linearidade.....	39

3.4.3. Precisão.....	41
3.4.4. Exatidão.....	41
3.4.5. Intervalo.....	43
3.4.6. Robustez.....	43
3.4.7. Sensibilidade e Sensibilidade Analítica.....	43
3.4.8. Limites de Detecção e Quantificação.....	44
3.4.9. Teste para a Presença de Erro Sistemático (Viés).....	44
3.5. Modelo de Calibração Pseudo-Univariado Baseado no NAS.....	45
4. Conclusões.....	47
5. Referências Bibliográficas.....	48

## LISTA DE TABELAS

<b>Tabela 1.</b>	Composição das amostras de calibração	23
<b>Tabela 1.</b>	Composição das amostras de calibração (Continuação).	24
<b>Tabela 2.</b>	Composição das amostras de validação	25
<b>Tabela 3.</b>	Evolução dos parâmetros para as três etapas na detecção de outliers no conjunto de calibração.	35
<b>Tabela 4.</b>	Resultados da previsão das amostras do conjunto de validação.	36
<b>Tabela 5.</b>	Figuras de mérito para o modelo PLS.	38
<b>Tabela 6.</b>	Avaliação da exatidão através da determinação de três amostras pelo método quimiométrico proposto e pelo método oficial (CLAE).	42
<b>Tabela 7.</b>	Avaliação da robustez do método proposto na determinação de uma amostra de 100,00 % de AMX em três níveis diferentes de temperatura.	43

## LISTA DE FIGURAS

<b>Figura 1.</b>	Reapresentação geométrica da propriedade de ortogonalidade do NAS.	6
<b>Figura 2.</b>	Fórmula Estrutural da amoxicilina.	17
<b>Figura 3.</b>	Fórmula Estrutural de três polimorfos da amoxicilina	19
<b>Figura 4.</b>	Espectrômetro Foss NIR systems, com acessório para obtenção de espectros	20
<b>Figura 5.</b>	Espectros das amostras de formulações de AMX obtidos por reflectância difusa no NIR.	29
<b>Figura 6.</b>	Espectro de uma amostra de AMX pura obtido por reflectância difusa no NIR.	30
<b>Figura 7.</b>	Espectros das amostras de calibração e validação obtidos por transflectância.	32
<b>Figura 8.</b>	Espectros das amostras de calibração e validação obtidos por transflectância, tratados com MSC.	33
<b>Figura 9.</b>	Valores do RMSECV (círculos) e do RMSEC (quadrados) em função do número de variáveis latentes.	33
<b>Figura 10.</b>	Detecção de outliers na primeira etapa para o conjunto de calibração. Gráficos a) de influência e b) dos resíduos de Y em função de X, com os respectivos limites de aceitação.	35
<b>Figura 11.</b>	Detecção de outliers nas amostras contendo ampicilina ao invés de amoxicilina. Gráficos a) de influência e b) dos resíduos de X, com os respectivos limites de aceitação.	39
<b>Figura 12.</b>	Valores de referência (linha cheia) contra os previstos para as amostras de calibração (círculos) e de validação (triângulos). A linha mais fina indica o ajuste ideal.	40
<b>Figura 13.</b>	Resíduos do modelo para as amostras de calibração (círculos) e de validação (triângulos).	41
<b>Figura 14.</b>	Cromatograma obtido para uma amostra no nível médio de teor AMX pelo método oficial (CLAE).	42
<b>Figura 15.</b>	Curva de calibração pseudo-univariada baseada na  NAS . (●) amostras de calibração, (▼) amostras de validação, (—) reta ajustada para as amostras de calibração.	46

## **Lista de Siglas e Símbolos**

**AMX-** Amoxicilina

**ANVISA-** Agência Nacional de Vigilância Sanitária

**ASTM-** *American Society for Testing and Materials*

**CLAE-** Cromatografia Líquida de Alta Eficiência

**DPR-** Desvio padrão relativo

**IQUEGO-** Indústria Química do Estado de Goiás

**IUPAC-** *International Union of Pure and Applied Chemistry*

**LD-** Limite de detecção

**LQ-** Limite de quantificação

**MLR-** *Multiple Linear Regression*

**MSC-** *Multiple Scatter Correction*

**NAS-** Sinal Analítico Líquido

**NIR-** *Near Infrared*

**OMS-** Organização Mundial da Saúde

**PCA-** *Principal Component Analysis*

**PCR-** *Principal Component Regression*

**PLS-** *Partial Least Squares*

**RMSEC-** *Root Mean Squares Errors of Calibration*

**RMSECV-** *Root Mean Squares Errors of Cross Validation*

**RMSEP-** *Root Mean Squares Errors of Prediction*

**SEN-** Sensibilidade

**SNV-** *Standard Normal Variate*

**UEG-** Universidade Estadual de Goiás

**UNUCET**- Unidade Universitária de Ciências Exatas e Tecnológicas

**UV**- Ultra-Violeta

**VL**- Variável Latente

## 1. Introdução

A quimiometria pode ser definida como a área da química que utiliza métodos matemáticos e estatísticos para [1]: definir ou selecionar as condições ótimas de medidas ou experimentos; e extrair o máximo de informações possíveis, a partir da análise de dados químicos.

Muita ênfase é dada aos sistemas multivariados, nos quais é possível medir muitas variáveis simultaneamente (ou de forma sequencial) com grande eficiência, ao se estudar diversos sistemas químicos. Nestes sistemas, a conversão da resposta instrumental na resposta ou propriedade de interesse requer a utilização de técnicas de estatística multivariada, álgebra linear e análise numérica. Atualmente, essas técnicas se constituem na melhor alternativa para a interpretação de dados e para a aquisição do máximo de informações sobre o sistema [2].

A quimiometria surgiu da necessidade de se tratar a grande e crescente quantidade de dados gerados nos laboratórios químicos, através da aplicação de métodos matemáticos e estatísticos. Dentre seus principais objetivos estão o planejamento e a otimização de experimentos, a análise exploratória e classificatória, a resolução de curvas e a calibração multivariada [3]. Em particular, a calibração multivariada será o foco do presente trabalho.

Na Química Analítica, a aplicação de métodos quimiométricos permitiu o desenvolvimento de metodologias para determinações diretas, sem a necessidade de resolução do sinal analítico e na presença de interferentes, desde que estes estejam contidos no conjunto de calibração, além da identificação de amostras anômalas. Estas possibilidades caracterizam a chamada vantagem de primeira ordem, possível em métodos de calibração multivariada que utilizam como dados de entrada um vetor de dados por amostra [4]. Desta maneira, adota-se a estratégia analítica de substituir a separação física e/ou química dos analitos (ex: cromatografia, extrações, precipitações, etc.) pela separação dos seus sinais analíticos. O uso desta estratégia traz potenciais vantagens para o desenvolvimento de métodos mais rápidos, baratos, não destrutivos, que envolvam menor manipulação das amostras e sem gasto de solventes ou geração de resíduos.

## 1.1. Calibração Multivariada

O princípio básico da calibração multivariada é utilizar simultaneamente várias variáveis  $X_1, X_2, \dots, X_N$  (ex: valores de absorvância em diferentes comprimentos de onda) para quantificar alguma outra variável de interesse,  $y$  (ex: concentração). O método mais simples de calibração multivariada é a Regressão Linear Múltipla (MLR, do inglês *Multiple Linear Regression*). Na MLR a variável a ser prevista, contida no vetor  $\mathbf{y}$ , é estimada por uma combinação linear das variáveis independentes contidas na matriz  $\mathbf{X}$ ,

$$\mathbf{y} = \mathbf{X}\mathbf{b} \text{ ou } y = b_1x_1 + b_2x_2 + \dots + b_nx_n \quad \text{Equação (1)}$$

Onde  $\mathbf{y}$  é um vetor contendo os valores da propriedade de interesse, que na maioria das vezes trata-se da concentração,  $\mathbf{X}$  é uma matriz contendo as variáveis independentes, neste trabalho os espectros, e  $\mathbf{b}$  é o vetor com os coeficientes de regressão. A solução por regressão de mínimos quadrados para encontrar  $\mathbf{b}$  é dada por:

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y} \quad \text{Equação (2)}$$

A MLR é o melhor método para a previsão em sistemas bem comportados, que apresentam respostas lineares, sem interações entre analitos, com baixo ruído e nenhuma colinearidade. A limitação da MLR está no fato de ela usar toda a informação contida na matriz  $\mathbf{X}$ , não importando se é relevante ou não para construir o modelo. Com isso, quantidade significativa de informação irrelevante é incorporada levando a modelos instáveis e pouco robustos. Outro problema está na etapa de inversão da matriz  $(\mathbf{X}'\mathbf{X})$ ; se a matriz  $\mathbf{X}$  contiver menos linhas que colunas (menos amostras que variáveis), ela não poderá ser invertida e não existirá solução única [3,5].

A Regressão em Componentes Principais (PCR, do inglês *Principal Component Regression*) é um método eficiente para superar as limitações apresentadas pela RLM, sendo mais robusto em relação ao ruído e à colinearidade. Nele, a regressão, ao invés de ser feita nas variáveis originais, é feita nos escores obtidos através da Análise de Componentes Principais (PCA, do inglês *Principal Component Analysis*),

a qual é o método quimiométrico de análise exploratória e classificação não supervisionada mais conhecido e utilizado [3,5].

### 1.1.1. Mínimos Quadrados Parciais

O PLS (do inglês *partial least squares*) pode ser considerado o método mais utilizado para a construção de modelos de calibração multivariada que relacionam uma matriz  $\mathbf{X}$  e um vetor  $\mathbf{y}$  ( $\mathbf{Y}$  é uma matriz quando mais de uma variável dependente/propriedade é prevista), os quais contêm as variáveis independentes  $x$ , e as dependentes  $y$ , respectivamente. Essas duas matrizes são decompostas simultaneamente em uma soma de  $f$  Variáveis Latentes (VL), de acordo com o seguinte modelo estrutural, dado pelas Equações (3) e (4).

$$\mathbf{X} = \mathbf{TP}' + \mathbf{E} = \sum_{f} \mathbf{t}_f \mathbf{p}'_f + \mathbf{E} \quad \text{Equação (3)}$$

Onde  $\mathbf{T}$  é a matriz de escores;  $\mathbf{P}$  é a matriz de pesos (*loadings*) para a decomposição de  $\mathbf{X}$ , e  $\mathbf{E}$  é a matriz de resíduos. O vetor ( $\mathbf{y}$ ) ou matriz ( $\mathbf{Y}$ ) contendo as variáveis dependentes é simultaneamente decomposto, de acordo com a seguinte equação.

$$\mathbf{y} = \mathbf{Tq}' + \mathbf{f} = \sum_{f} \mathbf{t}_f \mathbf{q}'_f + \mathbf{f} \quad \text{Equação (4)}$$

Onde  $\mathbf{q}$  é a matriz de pesos para a decomposição de  $\mathbf{y}$ ,  $\mathbf{f}$  é o vetor de resíduos. O produto de  $\mathbf{T}$  e  $\mathbf{P}'$  é uma aproximação das variáveis independentes (ex: espectros) e o produto de  $\mathbf{T}$  e  $\mathbf{q}'$  é uma aproximação das variáveis dependentes (ex: concentrações). Uma importante característica do PLS é a possibilidade de obter escores que estão simultaneamente relacionados às concentrações ( $y$ ) e às medidas ( $x$ ). A concentração das novas amostras pode ser estimada a partir dos novos escores,  $\hat{\mathbf{T}}$ , e dos pesos do modelo,  $\mathbf{Q}$ , de acordo com a equação:

$$\mathbf{y}_n = \hat{\mathbf{T}}\mathbf{q}' \quad \text{Equação (5)}$$

O número de VLs do modelo (equivalentes do PLS às componentes principais dos modelos PCA e PCR) é escolhido através de um procedimento de validação cruzada, baseado no menor erro de previsão. A maneira mais comum de realizar a validação cruzada é através do *leave-one-out*, em que uma amostra de calibração é retirada de cada vez e sua propriedade de interesse é estimada por um modelo

construído com as amostras restantes; a média dos erros de previsão é calculada para cada número de variáveis latentes e o número que apresentar menor erro de previsão é escolhido.

Uma Variável Latente (VL) é definida de maneira semelhante a uma componente principal, e é formada a partir de uma leve rotação no eixo das componentes principais para produzir uma maior correlação com o vetor de dados  $\mathbf{y}$ . Portanto, ao contrário das componentes principais, as variáveis latentes não são ortogonais entre si. Como pode ser notado, há um compromisso entre a habilidade das variáveis latentes em descrever as amostras nos espaços individuais (modelagem dos blocos  $\mathbf{X}$  e  $\mathbf{y}$ ) simultaneamente, levando a perda de ortogonalidade [3,6,7].

Deve-se ainda mencionar a diferença entre PLS1, quando a regressão é feita individualmente para cada variável dependente ( $\mathbf{y}$  é um vetor), e PLS2, quando todas as variáveis dependentes são previstas simultaneamente ( $\mathbf{Y}$  é uma matriz). O número após o PLS indica, portanto, a dimensão do arranjo de dados a ser previsto.

Uma relação entre os dois blocos ( $\mathbf{X}$  e  $\mathbf{y}$ ) pode ser realizada correlacionando-se os escores para cada componente de cada vez, utilizando-se um modelo linear. Esse modelo, entretanto, não é o melhor possível. Isto porque a PCA é realizada em cada matriz separadamente, podendo resultar numa relação não muito satisfatória (não linear) entre os escores dos dois blocos. No PLS, isto é realizado por uma leve mudança nos valores dos escores, de forma a produzir a melhor relação possível, ou seja, há uma leve rotação no eixo das componentes principais para obtenção de uma maior correlação com o vetor de dados  $\mathbf{y}$  [3,6,7].

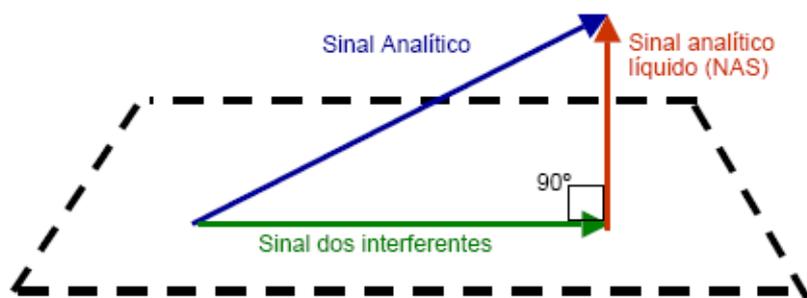
## **1.2. Validação Analítica de Métodos Quimiométricos**

A validação de um método analítico é feita de forma a garantir uma metodologia exata e que possa ser reproduzida sobre uma determinada faixa específica e flexível na análise de uma substância. Para avaliar se os parâmetros estão de acordo com as exigências legais, são feitos testes laboratoriais com o objetivo de demonstrar que o método é adequado ao propósito ao qual se destina. Para isto, são estimadas figuras de mérito, tais como linearidade, exatidão, precisão, limites de detecção e quantificação. O modo de estimar essas figuras e os critérios

de aceitação é reportado em normas e resoluções nacionais regulamentadas pela Agência Nacional de Vigilância Sanitária (ANVISA) [8] ou internacionais, como a Sociedade Americana de Testes e Materiais (ASTM, do inglês *American Society for Testing and Materials*) [9], para validação de métodos analíticos. A maioria desses documentos foi elaborada sob um “pensamento univariado”, com base nos métodos de análise tradicionais. O recente desenvolvimento de métodos quimiométricos gerou a necessidade de adaptação dessas normas ao “pensamento multivariado”. Com isso, a validação analítica passou a ser um desafio no desenvolvimento de novos métodos multivariados e se tornou objeto de estudo [7,10,11], considerando a necessidade de reconhecimento desses métodos pelos organismos reguladores. A seguir, será apresentada a maneira como são calculadas figuras de mérito para métodos multivariados, começando pelo importante conceito de Sinal Analítico Líquido (NAS, do inglês *Net Analyte Signal*).

### **1.2.1. Sinal analítico líquido**

A definição do NAS é de grande importância na determinação de figuras de mérito para calibrações multivariadas. A partir da necessidade de validação dos métodos analíticos multivariados, a estimativa do NAS se tornou uma ferramenta essencial, pois através dele são calculados vários parâmetros para mostrar a qualidade do modelo. O conceito de NAS foi proposto por Lorber em 1986 [12], e é definido como a parte do sinal analítico que é ortogonal às contribuições de possíveis interferentes presentes na amostra. Essa propriedade de ortogonalidade pode ser observada na representação geométrica da Figura 1.



**Figura 1.** Representação geométrica da propriedade de ortogonalidade do NAS

O modelo proposto por Lorber foi aprimorado por vários autores [11,12,13], de forma a possibilitar o cálculo preciso do NAS para modelos de calibração construídos a partir de PCR e PLS. No cálculo do NAS, inicialmente  $\mathbf{X}$  e  $\mathbf{y}$  são reconstruídos com as  $A$  variáveis latentes (VLs) utilizadas no modelo gerando  $\mathbf{X}_A$  e  $\mathbf{y}_A$ , segundo as equações (6) e (7) respectivamente [7,10,11].

$$\mathbf{X}_A = \mathbf{T}_A \mathbf{P}_A + \mathbf{E} \quad \text{Equação (6)}$$

Observa-se que na Equação (6)  $\mathbf{X}$  representa os espectros,  $\mathbf{T}$  os escores,  $\mathbf{P}$  os pesos, e  $\mathbf{E}$  contém os erros de decomposição da matriz  $\mathbf{X}$ .

$$\mathbf{y}_A = \mathbf{T}_A \mathbf{q}_A + \mathbf{f} \quad \text{Equação (7)}$$

Na Equação (7),  $\mathbf{T}$  é a matriz contendo os escores,  $\mathbf{q}$  é o vetor contendo os pesos  $\mathbf{f}$  é o vetor com os erros de decomposição do vetor  $\mathbf{y}$ . Em seguida, é determinada a matriz  $\mathbf{X}_{A,-k}$ , que é a matriz que contém a informação de todas as espécies presentes na amostra, exceto da espécie de interesse  $k$ . Isso é feito através de uma projeção ortogonal, que tem base na operação matricial que estabelece que para uma matriz  $\mathbf{X}$  qualquer,  $\mathbf{X}\mathbf{X}^+$  (onde o símbolo “+” sobrescrito indica a pseudoinversa de Moore–Penrose de uma matriz ou vetor) é uma matriz de projeção que contém as seguintes propriedades [10,11]:

$$\mathbf{X} = (\mathbf{X}\mathbf{X}^+)\mathbf{X} \quad \text{Equação (8)}$$

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

Tem-se então, através das informações acima, que a multiplicação de qualquer vetor  $\mathbf{z}$  pela matriz  $\mathbf{X}\mathbf{X}^+$ , fornecerá como resultado o próprio vetor  $\mathbf{z}$ , que é

uma combinação linear da matriz  $\mathbf{X}$ . No entanto, ao multiplicarmos  $\mathbf{z}$  por  $(\mathbf{I} - \mathbf{X}\mathbf{X}^+)$ , onde  $\mathbf{I}$  é uma matriz identidade de dimensões adequadas, tem-se como resultado um vetor de zeros, ou seja, um vetor ortogonal à matriz  $\mathbf{X}$ . Desta maneira, a matriz  $\mathbf{X}_{A,-k}$  é calculada como:

$$\mathbf{X}_{A,-k} = [\mathbf{I} - \mathbf{y}_{A,k}\mathbf{y}_{A,k}^+]\mathbf{X}_A \quad \text{Equação (10)}$$

onde  $\mathbf{y}_{A,k}$  é o vetor de concentração da espécie de interesse  $k$  estimado com  $A$  variáveis latentes segundo a equação:

$$\mathbf{y}_{A,k} = \mathbf{X}_A \mathbf{X}_A^+ \mathbf{y}_k. \quad \text{Equação (11)}$$

Desta forma, a matriz  $\mathbf{X}_{A,-k}$ , está livre de qualquer colaboração da espécie  $k$  e pode ser usada em uma projeção ortogonal para estimar um vetor  $\mathbf{x}_{A,k,i}^{\text{NAS}}$  para cada amostra  $i$  (cada espectro original  $\mathbf{x}_i$ , reconstruído com  $A$  VLs como  $\mathbf{x}_{A,i}$ , de acordo com a seguinte equação:

$$\mathbf{x}_{A,k,i}^{\text{NAS}} = (\mathbf{I} - \mathbf{X}_{A,-k}^T (\mathbf{X}_{A,-k}^T)^+ ) \mathbf{x}_{A,i} \quad \text{Equação (12)}$$

Dessa forma, a norma de cada vetor destes fornece um valor escalar de nas ( $\text{nâs}_i$ ), que pode ser usado como um sinal analítico univariado seletivo/puro.

$$\text{nâs}_i = \|\mathbf{x}_{A,k,i}^{\text{NAS}}\| \quad \text{Equação (13)}$$

onde  $\|\cdot\|$  representa o módulo do vetor.

Através do cálculo do escalar  $\text{nâs}_i$  de cada amostra, o qual é livre de interferentes, é possível uma nova forma de representar a calibração multivariada, em que o modelo possa ser expresso de maneira pseudo-univariada. Essa representação requer: (i) o cálculo do  $\text{nâs}$  para todas as amostras de calibração; (ii) em seguida, o coeficiente  $b_{\text{nas}}$  é determinado através da regressão por mínimos quadrados do vetor contendo os escalares do sinal analítico líquido das amostras de calibração ( $\text{nâs}$ ) contra o vetor de concentrações  $\mathbf{y}$ :

$$b_{\text{nas}} = (\text{nâs}^T \text{nâs})^{-1} \text{nâs}^T \mathbf{y} \quad \text{Equação (14)}$$

O modelo de regressão pode, então, ser representado por:

$$\mathbf{y} = b_{\text{nas}} \text{nâs} + \epsilon \quad \text{Equação (15)}$$

O modelo com base na regressão dos valores de referência contra o NAS, obtido por meio das equações (14) e (15), é chamado de modelo pseudo-univariado e deve fornecer resultados equivalentes aos obtidos pelo modelo multivariado.

Se os dados foram centrados na média, antes da determinação do coeficiente de regressão  $b_{nas}$ , é necessário fazer uma correção no vetor  $\mathbf{n\hat{a}s}$ , de forma a se evitar um erro de sinal, introduzido pelo uso da norma Euclidiana. Para essa correção se faz a multiplicação de cada elemento do vetor  $\mathbf{n\hat{a}s}$  pelo seu respectivo sinal no vetor  $(\mathbf{y}-\bar{y})$  onde  $\bar{y}$  é a média do vetor  $\mathbf{y}$ , que contém os valores de referência [14].

### 1.2.2. Figuras de Mérito

De acordo com a regulamentação brasileira [8], o tipo de aplicação da qual trata este trabalho é classificado na Categoria I: “Testes quantitativos para a determinação de princípio ativo em produtos farmacêuticos ou matérias primas”. Para esta categoria são exigidas as estimativas das figuras de mérito especificidade, linearidade, exatidão, precisão, intervalo, e robustez. Além disso, este trabalho também estimou as figuras de mérito sensibilidade, sensibilidade analítica, limite de detecção, limite de quantificação e viés (teste para a presença de erro sistemático). As quatro primeiras figuras de mérito mencionadas não são exigidas legalmente, mas são importantes para a comparação da qualidade do método proposto com outros métodos, sendo estimadas através do NAS.

Segundo a União Internacional de Química Pura e Aplicada (IUPAC, do inglês *International Union of Pure and Applied Chemistry*) é recomendado que seja utilizado o termo seletividade preferencialmente ao termo especificidade, como descrito nas normas brasileiras [17]. A SEL é um parâmetro que mede a pureza do sinal analítico em relação ao sinal total, sendo requerida a resolução total do sinal analítico por métodos univariados. Para métodos multivariados, a SEL pode ser expressa como a razão entre o  $n\hat{a}s_i$  e o módulo de cada espectro,  $\|\mathbf{x}_i\|$ , de acordo com a equação (16). Assim, um valor diferente é obtido para cada amostra e um valor médio de SEL pode ser usado para caracterizar o método de calibração

[16,17]. Considerando que os métodos de calibração multivariada são indicados somente quando não é possível se obter um sinal analítico seletivo, não há necessidade de se estabelecer um valor limite exigido de SEL.

Como uma das grandes vantagens dos modelos multivariados é justamente a capacidade de lidar com sinais não seletivos ou sobrepostos, a estimativa da SEL para métodos multivariados não tem sentido prático, porque a sua exigência pelos organismos reguladores é concebida de uma maneira univariada. Por um lado, a aplicação de métodos univariados só é possível na presença de total seletividade do sinal analítico (especificidade). Por outro lado, o uso de métodos multivariados somente é justificado quando não há resolução do sinal analítico. Desta forma, a mudança da exigência deste conceito pelos organismos reguladores permanece um dos principais desafios para o reconhecimento oficial de métodos quimiométricos [16,18]. Uma alternativa para avaliar a SEL/especificidade em métodos multivariados é demonstrar que o método é capaz de distinguir outros componentes, tais como impurezas, produtos de degradação ou outros princípios ativos semelhantes.

$$SEL_i = n\hat{s}_i / \|\mathbf{x}_i\| \quad \text{Equação(16)}$$

A exatidão mostra o grau de concordância do valor estimado ou medido com o valor de referência ou tido como verdadeiro. Normalmente, em aplicações de calibração multivariada, a exatidão média é expressa através da raiz quadrada do erro quadrático médio de previsão (RMSEP, do inglês *Root Mean Squares Errors of Prediction*)

$$RMSEP = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad \text{Equação (17)}$$

em que  $n$  é o número de amostras de validação,  $y_i$  é o valor de referência que contém o analito de interesse e  $\hat{y}_i$  é o valor estimado pelo modelo. Por ser estimado a partir de um conjunto independente de amostras de validação, o RMSEP é considerado um parâmetro “não viciado”. Mas, além dele pode-se estimar também o RMSEC (do inglês *Root Mean Squares Errors of Calibration*), de definição semelhante em relação às amostras de calibração, como mostrado na Equação (18).

$$RMSEC = \sqrt{\frac{\sum (y_r - \hat{y}_p)^2}{\nu}} \quad \text{Equação (18)}$$

onde  $\nu$  é o número de graus de liberdade, que é igual ao número de amostras usadas na calibração menos VL. Para dados centrados na média, usual em métodos espectroscópicos, perdem-se ainda mais um grau de liberdade.

Como a regulamentação atual não é pensada em termos multivariados, a figura do RMSEP normalmente não é reconhecida. Contudo, cabe ressaltar que o RMSEP apenas expressa a exatidão média do modelo, não podendo ser utilizado para atestar sozinho a exatidão do método, por não ser capaz de avaliar a presença de erros sistemáticos. Dessa maneira, este trabalho também avaliou a exatidão de acordo com a ANVISA [8], que exige que o grau de recuperação seja estimado para um mínimo de nove determinações em pelo menos três níveis de concentração (um alto, um médio e um baixo) que cubram a faixa de aplicabilidade do modelo [18].

A linearidade é a capacidade de uma metodologia analítica de demonstrar que os resultados obtidos são diretamente proporcionais à concentração do analito na amostra, dentro de um intervalo especificado [8]. Para métodos univariados, a linearidade é normalmente avaliada através do coeficiente de correlação ( $r$ ) da reta ajustada entre o sinal analítico e a concentração de referência (curva de calibração). Para métodos multivariados não são ajustadas curvas de calibração, a não ser a chamada curva pseudo-univariada, estimada através do NAS, conforme descrito na seção anterior. Desta forma, a linearidade para métodos multivariados costuma ser acessada através do ajuste de uma curva dos valores de referência contra os valores previstos [16,18]. Contudo, é importante notar que ambas as curvas, pseudo-univariada e de valores de referência *versus* valores estimados, não podem ser utilizadas para atestar a linearidade do método multivariado, pois em ambos os casos, os valores de referência são correlacionados com um parâmetro ( $n$  ou  $\hat{y}_i$ ) que já é obtido por um método linear. Logo, apenas uma estimativa qualitativa da linearidade pode ser obtida para modelos multivariados, a qual apenas indicaria se a aplicação de um modelo linear nos dados é válida. Essa estimativa qualitativa consiste em observar os resíduos do modelo para as amostras e calibração e validação, que devem se distribuir de forma aleatória.

A precisão é a avaliação da proximidade dos resultados obtidos em uma série de replicatas autênticas, que são amostras diferentes com o mesmo teor do princípio ativo e placebo. Neste caso, a estimativa não difere do caso univariado para o caso multivariado. Neste trabalho, a precisão será avaliada em dois níveis, repetibilidade e precisão intermediária, a partir das mesmas nove amostras (triplicatas de três níveis de concentração) usadas para avaliar a exatidão. A precisão intermediária pode ser avaliada através do desvio padrão relativo dos resultados obtidos por diferentes analistas em diferentes dias [8].

O intervalo é a amplitude de valores para os quais o método analítico apresenta exatidão, precisão e linearidade adequadas quando aplicado a amostras contendo quantidades de substâncias dentro do intervalo especificado [8].

A robustez: é a medida da capacidade de um método analítico de suportar pequenas variações de seus parâmetros. Ela indica sua confiança durante o uso normal [8]. O parâmetro escolhido nesse trabalho para avaliar a robustez do método será a temperatura, conforme sugerido pela ANVISA [8].

A sensibilidade (SEN) é definida como a fração de sinal responsável pelo acréscimo de uma unidade de concentração da propriedade de interesse. Em modelos de calibração multivariada, como o PLS, a SEN pode ser determinada como [16,19],

$$SEN = 1 / \| \mathbf{b} \| \quad \text{Equação (19)}$$

onde  $\mathbf{b}$  é o vetor dos coeficientes de regressão com A variáveis latentes do PLS. Alternativamente, esta equação pode ser expressa também em função da norma do vetor  $\mathbf{NAS}$ , que deve ser igual ao inverso da norma de  $\mathbf{b}$ .

$$SEN = \| \mathbf{NAS} \| = 1 / \| \mathbf{b} \| \quad \text{Equação (20)}$$

A sensibilidade analítica ( $\gamma$ ) é uma figura de mérito que normalmente não é abordada em normas ou protocolos de validação. Entretanto, ela apresenta, de forma clara, a sensibilidade do método em função da unidade de concentração que é utilizada. Em analogia com a calibração univariada [20], a  $\gamma$  é definida como a

razão entre a SEN e o ruído instrumental ( $\varepsilon$ ), de acordo com a equação (21). O ruído instrumental  $\varepsilon$  pode ser estimado através do desvio padrão combinado de um vetor contendo certo número de replicatas do espectro do branco (placebo) [16] ou através da norma do vetor de resíduos do modelo PLS [21], como adotado neste trabalho e mostrado na equação (22).

$$\gamma = \text{SEN} / \varepsilon \quad \text{Equação (21)}$$

$$\varepsilon = \sqrt{\frac{\|\mathbf{E}_{cal}\|^2}{\nu}} \quad \text{Equação (22)}$$

Onde  $\mathbf{E}_{cal}$  é a matriz de resíduos que contem os vetores de resíduos para todas as amostras de calibração,  $\|\cdot\|$  representa o módulo ou norma de Frobenius da matriz, e  $\nu$  é o número de graus de liberdade. Graus de liberdade é o número de valores independentes que serão utilizados na estimativa de um valor estatístico. Em geral, o número de graus de liberdade de uma estimativa é igual ao número de valores utilizados no seu cálculo menos o número de parâmetros estimados no cálculo intermediário para sua obtenção.

O inverso da  $\gamma$  ( $\gamma^{-1}$ ) é uma estimativa da diferença de concentração mínima que é discernível pelo método analítico considerando-se o ruído instrumental aleatório como única fonte de erro.

Os limites de detecção (LD) e de quantificação (LQ) de um método analítico expressam as menores quantidades da espécie de interesse que podem ser detectadas e determinadas quantitativamente, respectivamente [8]. Conforme mencionado anteriormente, estas figuras não são exigidas para a validação deste tipo de método. No entanto, elas serão estimadas aqui para fins comparativos. Para métodos multivariados, o LD e o LQ são estimados de acordo com as equações a seguir, que são coerentes com a definição adotada pela ANVISA:

$$\text{LD} = 3,3 \varepsilon \|\mathbf{b}\| = 3,3 \varepsilon / \text{SEN} \quad \text{Equação (23)}$$

$$\text{LQ} = 10 \varepsilon \|\mathbf{b}\| = 10 \varepsilon / \text{SEN} \quad \text{Equação (24)}$$

Segundo a IUPAC [22], o termo viés (teste para erro sistemático ou *bias*, em inglês) é atribuído a erros sistemáticos que são calculados pela diferença entre a média da população e o valor verdadeiro, correspondendo a todas as componentes de erro que não são aleatórias. Neste trabalho esse teste será feito através de dois testes t para comparar se os intervalos de confiança calculados para os coeficientes linear e angular da reta ajustada entre os valores de referência *versus* valores estimados para as amostras do conjunto de validação. Os intervalos de confiança estimados para esses parâmetros devem ser conter os valores esperados, zero e um, respectivamente [23].

### 1.3. Detecção de amostras anômalas

Amostras anômalas (*outliers*) podem ser definidas como amostras que possuem um comportamento diferente das demais amostras do conjunto de dados. A ocorrência dessas amostras pode ser causada por diversas razões, tais como erros de procedimento, erros instrumentais, amostras provenientes de outras populações, etc., e sua identificação e exclusão são fundamentais para assegurar a qualidade preditiva dos modelos. Existem várias metodologias que foram propostas para a detecção de *outliers* [25-27]. As três formas mais simples de identificá-los são baseadas na detecção de amostras com extrema influência no modelo (*leverage*), altos resíduos não modelados nos dados espectrais e altos resíduos na variável dependente (concentração). Neste trabalho será adotada a metodologia proposta por Valderrama *et al.* [24], a qual se baseia nas três formas mencionadas.

O valor de *leverage* indica o quanto uma amostra está distante do centro do conjunto de dados e pode ser definido de acordo como:

$$h_i = \mathbf{t}_{A,i}^T (\mathbf{T}_A^T \mathbf{T}_A)^{-1} \mathbf{t}_{A,i} \quad \text{Equação (25)}$$

onde  $\mathbf{T}$  representa os escores das amostras de calibração,  $\mathbf{t}_i$  é o vetor de escores da amostra  $i$  e  $A$  é número de variáveis latentes. De acordo com a norma ASTM E1655-05 [9], amostras com  $h_i$  mais altos que o valor limite ( $h_{\text{limite}}$ ) dado pela equação (26), devem ser removidas do conjunto de calibração e o modelo reconstruído.

$$h_{\text{limite}} = 3 \frac{A+1}{I_c} \quad \text{Equação (26)}$$

onde  $I_c$  é o número de amostras de calibração. Para modelos não centrados na média o fator 1 da equação (26) é omitido.

A identificação de *outliers* baseada nos resíduos não modelados dos dados espectrais é feita por comparação do desvio padrão residual total ( $s(e)$ ) com o desvio padrão de uma amostra ( $s(e_i)$ ), definidos nas equações (27) e (28)

$$s(e) = \frac{1}{I_c J - J - A \text{Max}(C, J)} \sum_{c=1}^{I_c} \left( \sum_{j=1}^J (e_{ij} - \hat{x}_{ij})^2 \right) \quad \text{Equação (27)}$$

$$s(e_i) = \frac{1}{I_c J - J - A \text{Max}(C, J)} \sum_{j=1}^J (e_{ij} - \hat{x}_{ij})^2 \quad \text{Equação (28)}$$

onde  $J$  é o número de variáveis espectrais,  $x_{ij}$  é o valor de absorvância da amostra  $i$  no comprimento de onda  $j$  e  $\hat{x}_{ij}$  é o valor estimado com  $A$  variáveis latentes. Se uma amostra apresentar  $s(e_i) > n \cdot s(e)$ , onde  $n$  é uma constante que pode variar entre 2 e 3, de acordo com o nível de confiança desejado [25], ela deve ser removida do conjunto de calibração. Neste trabalho, foi adotado  $n=2$ , valor considerado satisfatório para identificar amostras com resíduos espectrais significativamente maiores dos que as outras [24].

Finalmente, *outliers* também podem ser identificados através dos resíduos nas variáveis dependentes, pela comparação do valor de RMSEC do modelo (Equação (18)) com o erro absoluto de previsão da amostra. Se uma amostra apresentar uma diferença entre o valor de referência e o valor estimado maior do que uma constante que pode variar de duas a três vezes o valor de RMSEC, ela será identificada como *outlier* [25]. Neste trabalho foi adotado um valor de 3 para essa constante [24].

#### 1.4. Espectroscopia no Infravermelho Próximo

A Espectroscopia no Infravermelho Próximo (NIR, do inglês *Near Infrared*) é uma técnica espectroscópica que se baseia na medida da intensidade da absorção de luz por uma determinada amostra na região do NIR em função do comprimento de onda. Essa região do espectro eletromagnético é compreendida entre 800 e 2500 nm ( $12500$  e  $4000$   $\text{cm}^{-1}$ ), possuindo energia suficiente para gerar sobretons e bandas de combinações de vibrações moleculares, principalmente das ligações O-H, C-H, S-H e N-H, com de intensidade fracas, quando comparadas com as intensidades das bandas observadas no infravermelho médio. Como esses sinais são de intensidade baixa e altamente sobrepostos, essa região possui baixa sensibilidade e fornece pouca informação qualitativa. Por isso, os espectros NIR não são facilmente interpretáveis, o que fez com que essa região do espectro fosse mais explorada para análises quantitativas, requerendo para isso em quase todas as aplicações a utilização de métodos de calibração multivariada para extrair o máximo das informações quantitativas contidas nos espectros [28-30].

A espectroscopia NIR começou a ser usada em análises quantitativas de uma maneira mais ampla quando Karl Norris propôs analisar o conteúdo das informações da medição espectral usando reflectância difusa, ao invés do sinal fraco de transmitância. Hoje, a reflectância difusa é uma das várias possibilidades de se obter o espectro na região do NIR [28-30]. Dentre os principais modos de se obter as medidas no NIR as mais usadas são: absorvância, refletância difusa e transflectância. Nas medidas de transflectância os espectros são obtidos com sondas de reflexão, nas quais a radiação interage com a amostra e é em seguida refletida por um espelho passando novamente pela amostra e voltando ao detector por meio de outras fibras óticas.

Quando se usa o modo de reflectância difusa para a aquisição dos espectros, a leitura é feita diretamente na amostra com um mínimo pré-tratamento e quase nenhuma intervenção humana. Considerando acessórios que empregam sondas e fibras óticas, a leitura é feita através da incidência da radiação na amostra, na qual penetra alguns milímetros além da superfície, sendo parcialmente absorvida e refletindo o restante de forma difusa. A radiação é então coletada por

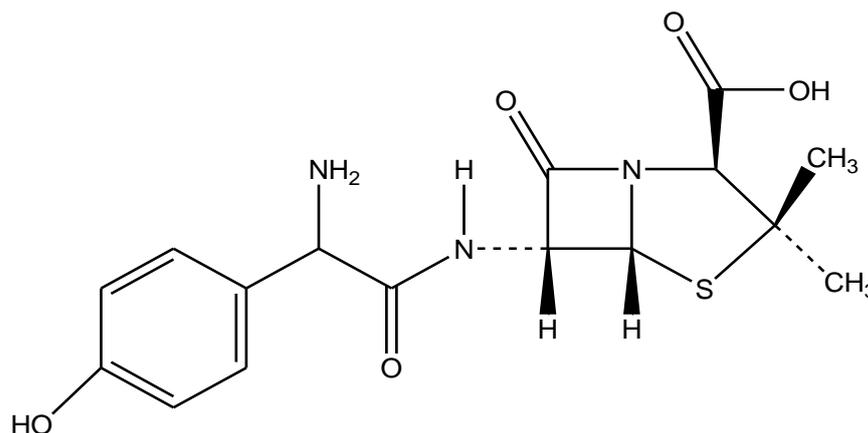
fibras ópticas e, após processamento matemático pela transformada de Fourier, gera um espectro que é relacionado com a concentração da amostra. Embora os espectros de reflectância difusa não sejam idênticos aos de transmitância, eles são bastante similares e contêm as mesmas informações químicas. A presença de espalhamento multiplicativo da luz, causado pela falta de homogeneidade no tamanho das partículas, é um aspecto crítico, que pode ser corrigido por métodos matemáticos, mas que em casos extremos inviabiliza a análise. Portanto, esse método de medição é ideal para amostras bem homogêneas [28-30].

No modo de transfectância a radiação interage com a amostra, é refletida por um espelho, passando novamente pela amostra e volta ao detector através de fibras ópticas, gerando um espectro de resposta. Esse modo de aquisição dos espectros pode ser usado para amostras líquidas, viscosas e em suspensão. As análises são rápidas, superando o problema de quantificar amostras que não podem ser analisadas na forma sólida (pó) por falta de homogeneidade e formação de cargas eletrostáticas na superfície das partículas. Neste trabalho esse será o modo utilizado para obtenção dos espectros.

Dentre as principais vantagens que a espectroscopia NIR pode apresentar, menciona-se o mínimo pré-tratamento da amostra, a rapidez e o conseqüente baixo custo, a possibilidade de análises não destrutivas, com potencial de aplicação universal, a simplicidade e a menor intervenção humana, gerando menos etapas de propagação de erros, a não necessidade de usar reagentes e/ou solventes e a não geração de resíduos químicos. A partir da década de 80, começou-se usar a aplicar a espectroscopia no NIR em análises agrícolas, poliméricas e clínicas. Esta técnica vem sendo muito utilizada nos últimos anos, sobretudo nas indústrias de alimentos, na determinação de moléculas orgânicas (gorduras e proteínas), petróleo, química, cosmética, têxtil, tintas [31,32], sendo ainda muito empregada para análise de matéria prima nas indústrias farmacêuticas.

## 1.5. Amoxicilina

A amoxicilina (AMX) é um antibiótico de amplo espectro, que foi introduzido no mercado mundial em 1972 e está na lista dos fármacos essenciais da Organização Mundial da Saúde (OMS) [33]. A fórmula da AMX, d-(-)-alfa-amino-p-hidroxibenzilpenicilina, é representada na Figura 2.



**Figura 2.** Fórmula estrutural da amoxicilina

A AMX é uma aminopenicilina semi-sintética do grupo betalactâmico e possui atividade antibacteriana contra muitos microrganismos gram-positivos e gram-negativos, agindo ao ligar-se a proteínas de ligação de penicilina, localizadas na membrana citoplasmática desses microrganismos. Ela possui atividade contra infecções respiratórias e otorrinolaringológicas (pneumonias, bronquite, amigdalite, faringite), urinárias (cistite, uretrite, gonorréia), de pele e do tecido subcutâneo (erisipela, celulite, abscessos, ferimentos infectados, furunculose, antraz), intestinais, cirúrgicas, febre tifóide, dentre outras [34]. Embora a AMX possa ser usada na forma de sal de sódio, seu uso em produtos farmacêuticos de administração oral se dá exclusivamente na forma de trihidrato [33].

Uma revisão da literatura indica a presença de vários métodos para a determinação de AMX em formulações farmacêuticas, dentre os quais se podem citar os baseados em técnicas titulométricas [35], cromatográficas [36-41], espectrofotométricas [42-45], de luminescência molecular [44-47], eletroforéticas [48] e eletroanalíticas [49-51], sendo a mais utilizada a cromatografia. No entanto, a

grande maioria desses métodos não é capaz de determinar diretamente a AMX na presença de interferentes, tais como excipientes, impurezas e outros princípios ativos associados.

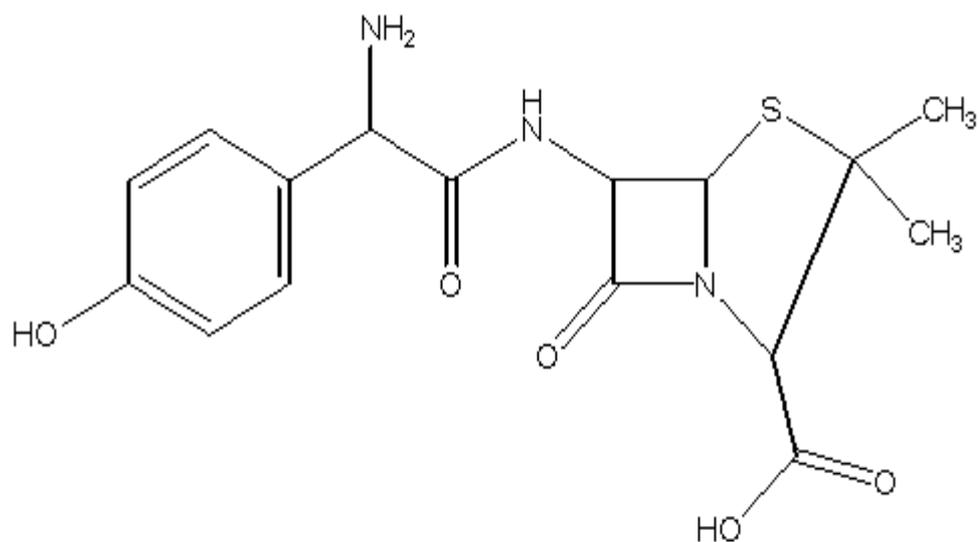
Alguns métodos envolvem reações de derivatização [42,43], tornando-se mais trabalhosos e sujeitos a erros. Já os métodos cromatográficos [36-41], embora estejam bem estabelecidos e sejam reconhecidos por organismos reguladores [36], podem apresentar desvantagens potenciais em certas situações, tais como alto custo, gerar resíduos, tempo de análise relativamente alta, entre outros. Um único trabalho foi publicado recentemente na literatura [52] propondo o uso de calibração multivariada para a determinação de AMX em uma formulação farmacêutica (pó) através de medidas de reflectância difusa no NIR.

Os melhores resultados desse trabalho foram obtidos com um modelo de redes neurais com função de base radial e PCA, uma metodologia bastante complexa, embora modelos PLS e PCR tenham apresentado resultados considerados satisfatórios. Cabe ressaltar que a formulação analisada por Qu *et al.* é consideravelmente mais simples, tendo amido como único excipiente, ao contrário do método proposto aqui, que analisou uma formulação com 7 substâncias excipientes, uma situação bem mais próxima da realidade. Além disso, o último trabalho mencionado [52] não realizou validação analítica.

No presente trabalho, tentou-se inicialmente elaborar um método para determinar AMX diretamente no pó para suspensão por medidas de reflectância difusa, numa situação similar à mencionada anteriormente [52]. No entanto, não foram obtidos resultados satisfatórios, por motivos que serão discutidos posteriormente (Seção 3.1). Optou-se então por preparar suspensões do pó em água e realizar a determinação usando medidas de transflectância, tornando a análise destrutiva. Desta maneira, o objetivo deste trabalho foi elaborar um método para a determinação de AMX em uma formulação farmacêutica baseado em medidas de transflectância no NIR e calibração multivariada (PLS). Além disso, foi empregada uma metodologia quimiométrica robusta que incluiu seleção de amostras, detecção de *outliers* e, principalmente, o processo de validação analítica completo, fundamental para a futura aceitação oficial de um método analítico como

este. Cabe ressaltar que a metodologia completa proposta aqui serve de base para ser aplicada na elaboração de outros métodos de controle de qualidade de formulações farmacêuticas.

Finalmente, cabe ressaltar que a molécula de AMX pode existir em mais de uma forma cristalina no estado sólido. No desenvolvimento de métodos analíticos baseados em espectroscopia no infravermelho é importante levar em conta a existência de espécies polimórficas, que podem apresentar diferentes sinais. No entanto, este aspecto só é considerado quando as amostras são analisadas no estado sólido, como no modo de aquisição dos espectros por reflectância difusa. Como o presente método foi desenvolvido para a análise de suspensões no modo de transflectância, não existiu interesse em analisar o comportamento das espécies polimórficas de AMX. A seguir é mostrado um polimorfo da AMX.



**Figura 3.** Polimorfo da amoxicilina

## 2. Procedimento

### 2.1. Equipamentos e processamento dos dados

Toda a parte experimental deste trabalho foi conduzida no Laboratório de Controle de Qualidade, nas dependências da Indústria Química do Estado de Goiás(IQUEGO), em Goiânia. As medidas foram feitas em um Espectrofotômetro de NIR da marca *Foss*, modelo 4500, o qual possui acessório para a obtenção de espectros de transflectância, modelo *smart probe analyzer*. Para as medidas preliminares foi usado também um acessório de reflectância difusa. O sinal de referência empregado para o branco das medidas de transflectância foi obtido pela medida da cela vazia na ausência de luz.

Para o presente trabalho o espectrofotômetro NIR foi controlado e os dados adquiridos usando o programa do próprio equipamento: Vision 3.3.0.0 (Foss). Toda a análise dos dados foi feita no programa MATLAB, versão 7.9 (R2009b) (*The MathWorks, Natick, EUA*), usando o pacote PLS Toolbox, versão 5.2 (*Eigenvector Technologies, Manson, EUA*), o qual contém as rotinas para métodos de calibração multivariada, além uma rotina para a detecção de *outliers* elaborada por J. W. B. Braga [24].

A seguir é mostrado o equipamentos utilizado na obtenção dos espectros.



**Figura 4.** Espectrofotômetro FOSS NIR Systems com acessório para obtenção de espectros de transflectância

## **2.2. Materiais, reagentes e suspensões**

Todos os reagentes utilizados foram de grau analítico, obtidos pela IQUEGO de fornecedores certificados. A amoxicilina triidratada foi adquirida com teor de 86,6%). O produto a ser analisado, pó para suspensão oral, Amoxicilina similar, produzido pela IQUEGO, é apresentado na forma de frascos de vidro âmbar de 80, 150 e 200 mL (contendo 60, 120 e 150 mL após reconstituição, respectivamente). Cada 5 mL de suspensão, a ser preparada pelo paciente, deve conter 250 mg de AMX. A composição de excipientes é a seguinte: carboximetilcelulose sódica, benzoato de sódio, dióxido de silício coloidal, aroma de cereja, corante vermelho eritrosina, citrato de sódio anidro e sacarose refinada [34]. As amostras em pó foram preparadas por pesagem em balança analítica (Sartorius) de precisão 0,0001g, de acordo com a composição da formulação, contendo em torno de 10% em massa do princípio ativo. Todas as suspensões foram preparadas com água deionizada (Ultrapurificador Maxima, Elga). Os modelos foram construídos em uma faixa de 80,00 a 130,00% (40,00 a 65,00 mg mL<sup>-1</sup>) da massa do princípio ativo, para abranger a faixa de aceitação para a liberação do fármaco, que está entre 90,00 e 120,00 % (45,00 a 60,00 mg mL<sup>-1</sup>) [53]. Portanto, o teor alvo de 100% do princípio ativo nas amostras (bula) corresponde a 50 mg mL<sup>-1</sup> de AMX.

## **2.3. Metodologia**

### **2.3.1. Estudos preliminares: medidas de reflectância difusa**

Inicialmente, tentou-se elaborar um modelo baseado em medidas de reflectância difusa, feitas diretamente nas misturas em pó. Para isso, foram preparadas 39 amostras, cujos espectros foram registrados na faixa de 1100 a 2500nm. Após a constatação da inviabilidade deste modelo, optou-se por construir outro modelo baseado em medidas de transflectância feitas nas suspensões preparadas a partir da dissolução das amostras de pó em água. Para uma discussão qualitativa, foi obtido ainda um espectro de AMX pura por reflectância difusa.

### **2.3.2. Planejamento experimental**

As amostras foram preparadas na faixa de concentração de AMX desejada, segundo um planejamento experimental. Originalmente, o objetivo foi utilizar um planejamento com três fatores: o princípio ativo AMX, o excipiente em maior proporção (sacarose refinada), e a mistura dos outros seis excipientes. No entanto, devido a dificuldades práticas na pesagem e na homogeneização enfrentadas na produção das amostras, e apontadas pelos técnicos de produção da IQUEGO, optou-se por um planejamento com apenas dois fatores: AMX e a mistura total de excipientes (placebo).

Desta forma, foi feito um planejamento com dois fatores para a preparação de 82 amostras de calibração e 50 amostras de validação, num total de 132 amostras. O detalhamento deste planejamento é mostrado nas Tabelas 1 e 2.

**Tabela 1.** Composição das amostras de calibração.

teor de AMX na mistura sólida (m/m)	teor do placebo na mistura sólida (m/m)	massa de AMX (g)	massa de placebo (g)	teor de AMX em relação à formulação (%)
0,0800	0,920	2,3202	26,683	80,55
0,0805	0,919	2,3349	26,668	81,05
0,0810	0,919	2,3495	26,653	81,56
0,0820	0,918	2,3788	26,624	82,58
0,0825	0,917	2,3935	26,609	83,09
0,0830	0,917	2,4081	26,595	83,60
0,0835	0,916	2,4228	26,580	84,10
0,0840	0,916	2,4374	26,566	84,61
0,0845	0,915	2,4521	26,551	85,12
0,0851	0,915	2,4667	26,536	85,63
0,0856	0,914	2,4813	26,522	86,14
0,0861	0,914	2,4960	26,507	86,65
0,0871	0,913	2,5253	26,478	87,66
0,0876	0,912	2,5399	26,463	88,17
0,0881	0,912	2,5546	26,448	88,68
0,0886	0,911	2,5692	26,434	89,19
0,0891	0,911	2,5839	26,419	89,70
0,0896	0,910	2,5985	26,404	90,21
0,0901	0,910	2,6132	26,390	90,71
0,0911	0,909	2,6425	26,360	91,73
0,0916	0,908	2,6571	26,346	92,24
0,0921	0,908	2,6718	26,331	92,75
0,0926	0,907	2,6864	26,317	93,26
0,0936	0,906	2,7157	26,287	94,27
0,0941	0,906	2,7303	26,273	94,78
0,0946	0,905	2,7450	26,258	95,29
0,0952	0,905	2,7596	26,243	95,80
0,0957	0,904	2,7743	26,229	96,31
0,0962	0,904	2,7889	26,214	96,82
0,0967	0,903	2,8036	26,199	97,32
0,0972	0,903	2,8182	26,185	97,83
0,0977	0,902	2,8329	26,170	98,34
0,0982	0,902	2,8475	26,155	98,85
0,0987	0,901	2,8622	26,141	99,36
0,0992	0,901	2,8768	26,126	99,87
0,0997	0,900	2,8914	26,112	100,38
0,1002	0,900	2,9061	26,097	100,88
0,1007	0,899	2,9207	26,082	101,39
0,1012	0,899	2,9354	26,068	101,90
0,1017	0,898	2,9500	26,053	102,41
0,1022	0,898	2,9647	26,038	102,92

**Tabela 1.** Composição das amostras de calibração (Continuação).

<b>teor de AMX na mistura sólida (m/m)</b>	<b>teor do placebo na mistura sólida (m/m)</b>	<b>massa de AMX (g)</b>	<b>massa de placebo (g)</b>	<b>teor de AMX em relação à formulação (%)</b>
0,1027	0,897	2,9793	26,024	103,43
0,1032	0,897	2,9940	26,009	103,93
0,1042	0,896	3,0233	25,980	104,95
0,1053	0,895	3,0526	25,950	105,97
0,1058	0,894	3,0672	25,936	106,48
0,1063	0,894	3,0819	25,921	106,98
0,1073	0,893	3,1111	25,892	108,00
0,1078	0,892	3,1258	25,877	108,51
0,1083	0,892	3,1404	25,863	109,02
0,1093	0,891	3,1697	25,833	110,04
0,1098	0,890	3,1844	25,819	110,54
0,1113	0,889	3,2283	25,775	112,07
0,1123	0,888	3,2576	25,745	113,09
0,1133	0,887	3,2869	25,716	114,10
0,1143	0,886	3,3162	25,687	115,12
0,1148	0,885	3,3308	25,672	115,63
0,1154	0,885	3,3455	25,657	116,14
0,1164	0,884	3,3748	25,628	117,15
0,1169	0,883	3,3894	25,614	117,66
0,1174	0,883	3,4041	25,599	118,17
0,1184	0,882	3,4334	25,570	119,19
0,1189	0,881	3,4480	25,555	119,70
0,1194	0,881	3,4627	25,540	120,20
0,1199	0,880	3,4773	25,526	120,71
0,1204	0,880	3,4920	25,511	121,22
0,1209	0,879	3,5066	25,496	121,73
0,1214	0,879	3,5212	25,482	122,24
0,1224	0,878	3,5505	25,452	123,26
0,1229	0,877	3,5652	25,438	123,76
0,1234	0,877	3,5798	25,423	124,27
0,1239	0,876	3,5945	25,408	124,78
0,1244	0,876	3,6091	25,394	125,29
0,1249	0,875	3,6238	25,379	125,80
0,1255	0,875	3,6384	25,365	126,31
0,1260	0,874	3,6531	25,350	126,81
0,1265	0,874	3,6677	25,335	127,32
0,1270	0,873	3,6824	25,321	127,83
0,1275	0,873	3,6970	25,306	128,34
0,1280	0,872	3,7117	25,291	128,85
0,1285	0,872	3,7263	25,277	129,86
0,1295	0,871	3,7556	25,247	130,37

**Tabela 2.** Composição das amostras de validação.

<b>teor de AMX na mistura sólida (m/m)</b>	<b>teor do placebo na mistura sólida (m/m)</b>	<b>massa de AMX (g)</b>	<b>massa de placebo (g)</b>	<b>teor de AMX em relação à formulação (%)</b>
0,080	0,920	2,3202	26,683	80,55
0,081	0,919	2,3500	26,653	81,58
0,082	0,918	2,3797	26,623	82,61
0,083	0,917	2,4094	26,594	83,64
0,084	0,916	2,4391	26,564	84,68
0,085	0,915	2,4689	26,534	85,71
0,086	0,914	2,4986	26,504	86,74
0,087	0,913	2,5283	26,475	87,77
0,088	0,912	2,5581	26,445	88,80
0,089	0,911	2,5878	26,415	89,84
0,090	0,910	2,6175	26,385	90,87
0,091	0,909	2,6472	26,356	91,90
0,092	0,908	2,6770	26,326	92,93
0,093	0,907	2,7067	26,296	93,96
0,094	0,906	2,7364	26,267	95,00
0,095	0,905	2,7662	26,237	96,03
0,096	0,904	2,7959	26,207	97,06
0,097	0,903	2,8256	26,177	98,09
0,098	0,902	2,8553	26,148	99,12
0,099	0,901	2,8851	26,118	100,16
0,101	0,900	2,9148	26,088	101,19
0,102	0,898	2,9445	26,058	102,22
0,103	0,897	2,9743	26,029	103,25
0,104	0,896	3,0040	25,999	104,28
0,105	0,895	3,0337	25,969	105,32
0,106	0,894	3,0634	25,940	106,35
0,107	0,893	3,0932	25,910	107,38
0,108	0,892	3,1229	25,880	108,41
0,109	0,891	3,1526	25,850	109,44
0,110	0,890	3,1823	25,821	110,48
0,111	0,889	3,2121	25,791	111,51
0,112	0,888	3,2418	25,761	112,54
0,113	0,887	3,2715	25,731	113,57
0,114	0,886	3,3013	25,702	114,60
0,115	0,885	3,3310	25,672	115,64
0,116	0,884	3,3607	25,642	116,67
0,117	0,883	3,3904	25,613	117,70
0,118	0,882	3,4202	25,583	118,73
0,119	0,881	3,4499	25,553	119,76
0,120	0,880	3,4796	25,523	120,80
0,121	0,879	3,5094	25,494	121,83
0,122	0,878	3,5391	25,464	122,86
0,123	0,877	3,5688	25,434	123,89
0,124	0,876	3,5985	25,404	124,92
0,125	0,875	3,6283	25,375	125,96
0,126	0,874	3,6580	25,345	126,99
0,127	0,873	3,6877	25,315	128,02
0,128	0,872	3,7175	25,286	129,05
0,129	0,071	3,7472	25,256	130,08
0,130	0,870	3,7769	25,226	131,12

### 2.3.3. Preparo das amostras

O preparo de todas essas amostras foi feito durante um período de aproximadamente 80 dias. Devido ao longo intervalo de preparo, notou-se que o placebo, mesmo sendo lacrado todo dia, estava apresentando mudança de coloração perceptível, passando de branco para levemente rosa. Isto é um indicativo do aumento da umidade. Na tentativa de elaborar os primeiros modelos quimiométricos, notou-se a presença de tendência sistemática nos dados em função do período de preparo das amostras. Esta constatação levou a decisão de modificar o critério de seleção das amostras para os conjuntos de calibração e validação, adotando-se o algoritmo de Kennard-Stone [54] que é um algoritmo que seleciona amostras de forma homogênea em um espaço amostral, sem levar em consideração o valor de referência [27], muito usado em métodos de calibração multivariada para selecionar um subconjunto de amostras de acordo com sua distribuição homogênea em toda a faixa de variância do espaço amostral. As quantidades de amostras nos conjuntos de calibração e validação foram mantidas as mesmas.

A mistura sólida estocada do placebo/excipientes foi preparada em um saco plástico, que foi agitado manualmente, misturando-se os diversos componentes nas devidas proporções utilizadas na formulação, por um tempo de aproximadamente 2 horas, produzindo um total de aproximadamente 5 kg de placebo. As amostras sólidas foram preparadas de acordo com o planejamento experimental pela pesagem das massas de AMX em papel manteiga e das massas do placebo em béqueres de plástico de 100 mL. As massas do princípio ativo foram transferidas individualmente para cada béquer contendo o placebo e foram homogeneizadas usando um bastão de vidro.

Para a preparação das suspensões, acrescentou-se cerca de 20 mL de água deionizada a cada béquer e transferiu-se o seu conteúdo quantitativamente para balões volumétricos de 50,0 mL, os quais foram agitados até obtenção das suspensões e, em seguida, completados com água até o menisco. Para as medidas, as amostras de suspensão foram transferidas para frascos de polietileno, nos quais foi imersa a sonda de transflectância. Os espectros foram registrados na faixa de 1100 a 2500 nm (passo 2 nm) com um caminho ótico de 2 mm e 32 varreduras.

Entre cada leitura, a sonda era lavada e secada. Para cada amostra foram feitas três leituras e calculou-se o espectro médio. Durante todas as medidas, a temperatura ambiente foi mantida em 25 °C.

Para a validação analítica do método proposto foram feitas mais algumas medidas, visando a estimativa de algumas figuras de mérito. Foram preparadas seis replicatas de uma amostra com teor de AMX de 100 % (50 mg mL<sup>-1</sup>) para estimar a repetibilidade. Estas mesmas replicatas foram analisadas por outro analista em outro dia para estimar a precisão intermediária. Para avaliação da robustez, foram analisada triplicatas dessa mesma amostra em três temperaturas diferentes: 15, 25 e 35 °C. As medidas na temperatura inferior foram realizadas imergindo-se os frascos de polietileno contendo a amostra em um banho de água fria e monitorando-se a temperatura com um termômetro. De maneira semelhante, as medidas na temperatura superior foram feitas imergindo o recipiente em um banho de água aquecido. Para a estimativa da exatidão foram obtidas triplicatas de amostras em três diferentes níveis de concentração: um baixo, um médio e um alto (80,00; 100,00 e 130,00 % de AMX). Estas mesmas amostras foram determinadas usando o método cromatográfico oficial, baseado em CLAE e recomendado pela Farmacopéia Americana [36]. Finalmente, para avaliar a especificidade, dez amostra contendo ampicilina como princípio ativo, ao invés de AMX, na faixa de 40,0 a 65,0 mg mL<sup>-1</sup>, foram preparadas e analisadas pelo método proposto.

#### **2.3.4. Análise cromatográfica**

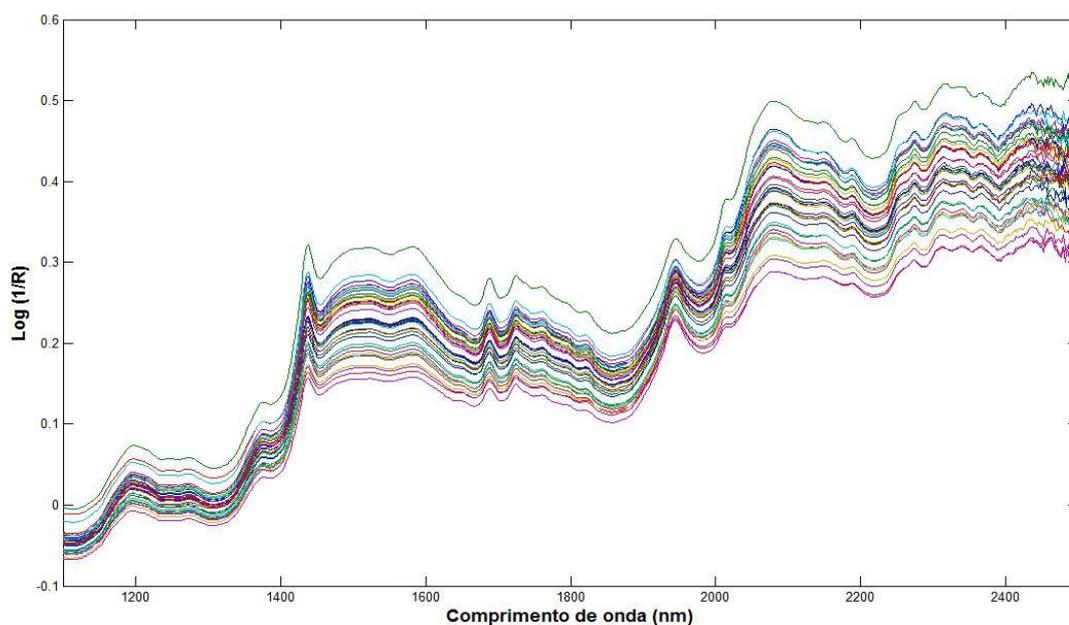
A validação da exatidão do método foi feita pela comparação dos resultados com o método oficial, baseado em Cromatografia Líquida de Alta Eficiência (CLAE) [36]. Foi usado um cromatógrafo Perkin-Elmer, com módulos da série 200 (bomba, válvula de injeção e detector). A detecção foi feita no ultravioleta, em 230 nm. Foi usada uma coluna analítica C-18 (250 x 4,6 mm), empacotada com partículas de sílica de 5 µm. A fase móvel foi tampão fostato/acetoneitrila (96:4, v/v), ajustada em pH 5,0(0,1, com um fluxo de 1,5 mL/min. Todas as injeções foram repetidas três vezes e as determinações foram feitas em triplicata.

### 3. Resultados e discussão

#### 3.1. Estudos preliminares: medidas de reflectância difusa

Conforme mencionado anteriormente, o objetivo inicial deste trabalho era elaborar um modelo de calibração multivariada com base em medidas de reflectância difusa no NIR feitas diretamente nas amostras em pó. Para isto, foram preparadas 39 amostras da formulação farmacêutica contendo AMX numa faixa de teor de 80 a 130%. Estas amostras foram divididas em 27 para o conjunto de calibração e 12 para o conjunto de validação. Os espectros obtidos são mostrados na Figura 5 e foram pré-processados usando correção de espalhamento multiplicativo (MSC, do inglês *Multiple Scatter Correction*). O melhor modelo preditivo usando PLS foi obtido com 8 variáveis latentes e forneceu um valor de RMSEP de 9,10%. As amostras de validação apresentaram erros individuais de previsão de até 23%, valores considerados inaceitáveis pela ANVISA [8].

Na seqüência, essas mesmas amostras foram usadas no preparo de suspensões em água e medidas por transflectância no NIR. O modelo PLS elaborado a partir destes dados apresentou RMSEP de 2,0% e erros de previsão de no máximo  $\pm 2,5\%$ , valores muito mais próximos dos limites de aceitação normalmente aceitos, erros máximos de  $\pm 2\%$ . Desta forma, decidiu-se direcionar este trabalho para a elaboração de um modelo mais completo e robusto, usando medidas de transflectância, o qual deve incluir um número maior de amostras e também a variação de mais um fator, a concentração do placebo/excipientes, que havia sido mantida fixa no modelo anterior.

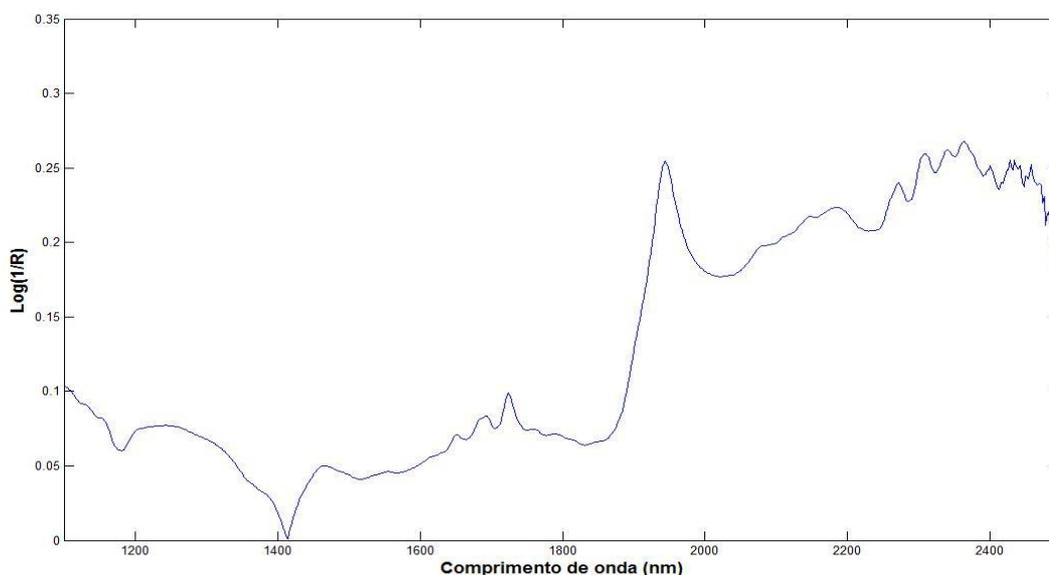


**Figura 5.** Espectros das amostras de formulações de AMX obtidos por reflectância difusa no NIR

Considerando que a partir das mesmas amostras sólidas o modelo com base nas medidas de transfectância apresentou resultados significativamente melhores, concluiu-se pela inviabilidade da elaboração de um modelo não destrutivo baseado nas medidas de reflectância difusa. A justificativa para isto pode ser dada a partir da dificuldade observada na pulverização e homogeneização das amostras. Em medidas de reflectância difusa no NIR, um fator fundamental para a qualidade dos modelos quantitativos é o controle do tamanho das partículas da amostra, que deve ser o mais homogêneo possível [55]. A falta de homogeneidade no tamanho das partículas leva a um aumento do espalhamento de luz, que em casos extremos não pode ser corrigido por métodos matemáticos, tais como o MSC. Durante a pulverização e mistura dos componentes para o preparo das amostras, feita com almofariz de ágata e pistilo, observou-se a formação constante de aglomerados que não se desfaziam. Isto deve estar associado à formação de cargas eletrostáticas na superfície das partículas, que dificultam a homogeneização da mistura. Desta forma, avalia-se que este fator foi o responsável pela inviabilização do modelo usando dados de reflectância difusa.

### 3.2. Modelo de calibração multivariada com medidas de transfectância

Os espectros de transfectância no NIR foram originalmente registrados entre 1100 e 2500 nm. Acima de 2000 nm foi observada uma banda larga de forte absorção, atribuída à ligação O-H da água. Devido à quase saturação do sinal analítico, decidiu-se eliminar essa região e restringir os modelos de 1100 a 2000 nm. Antes de discutir os modelos quantitativos, é interessante uma análise qualitativa, através da atribuição de bandas no espectro da AMX. Como não é viável obter espectros de transfectância da AMX pura, registrou-se um espectro de reflectância difusa de uma amostra sólida de AMX, mostrado na Figura 6. Espectros de reflectância, embora não idênticos aos correspondentes de absorção, são similares em forma e fornecem essencialmente as mesmas informações químicas [28]. Observando a Figura 6, é possível atribuir à banda de maior intensidade, centrada em 1945 nm, às combinações das vibrações das ligações O-H e N-H da molécula de AMX. Além disso, atribuiu-se a banda entre 1200 e 1300 nm ao segundo sobretom e combinação de ligações C-H; as bandas entre 1640 e 1780 ao primeiro sobretom de ligações C-H; e as bandas entre 2050 e 2220 às combinações de ligações C-H [55].

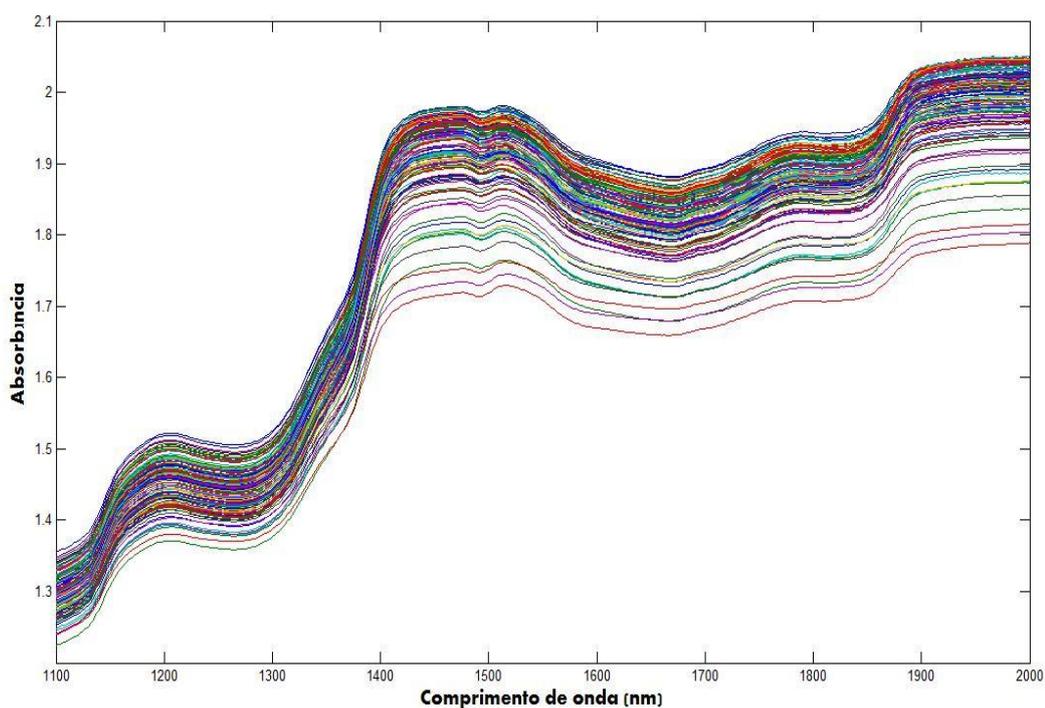


**Figura 6.** Espectro de uma amostra de AMX pura obtido por reflectância difusa no NIR

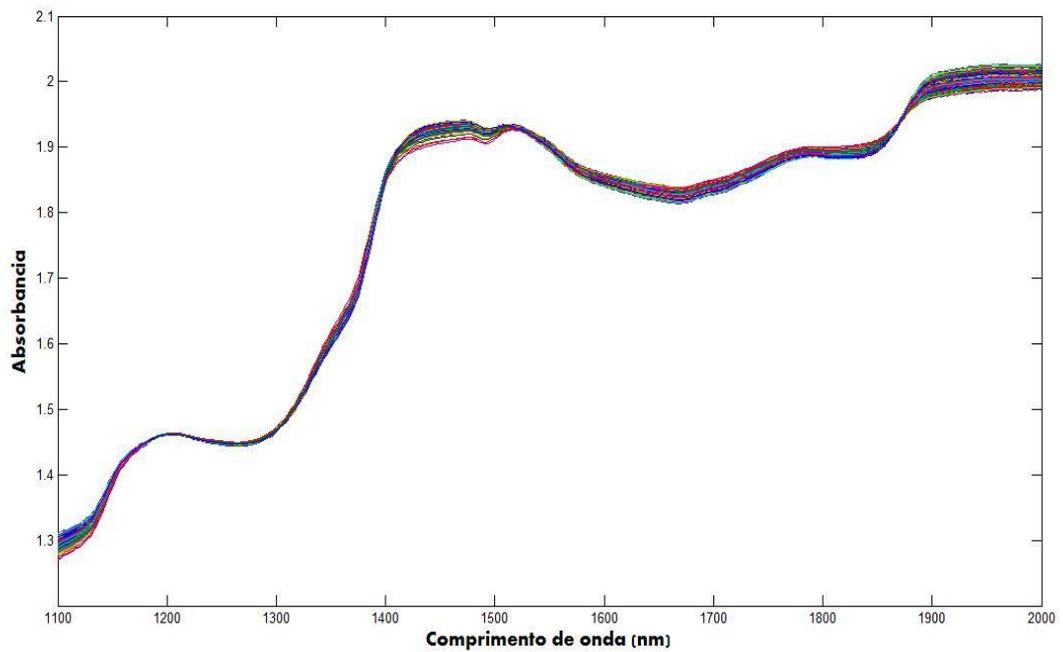
De acordo com o planejamento experimental descrito na seção 2.3.2, apresentado nas Tabelas 1 e 2, foram preparadas 132 amostras de formulações na faixa de 80 a 130 % de AMX. Os espectros de transfectância de todas estas amostras são apresentados na Figura 7. Essas amostras foram divididas em um conjunto de calibração (82) e um de validação (50) usando o algoritmo de Kennard-Stone [54]. Foram testados modelos locais de calibração multivariada usando regiões selecionadas do espectro, mas o melhor modelo foi obtido usando toda a região espectral. Os espectros mostrados na Figura 7 podem apresentar desvios de linha-base tanto lineares, quanto não lineares (multiplicativos), o que normalmente exige o uso de métodos de correção como uma etapa de pré-processamento dos dados. Neste trabalho foram testados modelos PLS usando os seguintes métodos de pré-processamento dos dados: MSC [57], correção por Variação Normal Padrão (SNV, do inglês *Standard Normal Variate*) [58], e primeira e segunda derivadas conjuntamente com alisamento Savitsky-Golay [59]. Esses desvios de linha-base são variações espectrais causadas por fatores físicos instrumentais, que não estão relacionadas com a composição química das amostras e, portanto, com a concentração do analito. Por isso, essas variações devem ser eliminadas. O uso de MSC ou SNV visa corrigir o deslocamento não linear da linha-base dos espectros causados pela dispersão de luz das partículas sólidas de diferentes tamanhos presentes na suspensão. A MSC consiste em fazer uma regressão dos espectros medidos contra um espectro de referência, que normalmente é o espectro médio, eliminando assim o deslocamento. A SNV consiste em diminuir de cada espectro seu valor médio e dividi-lo pelo seu respectivo desvio padrão (é o equivalente a autoescalar uma matriz transposta). Por outro lado, o uso de derivadas visa eliminar desvios de linha-base lineares. Um efeito colateral do uso de derivadas é o aumento do ruído, o que leva ao uso conjunto de métodos de alisamento que visam aumentar a razão sinal/ruído.

O melhor modelo foi obtido com os dados pré-processados com MSC e centrados na média. A Figura 8 mostra os espectros da Figura 7 depois de corrigidos com MSC. Os espectros do conjunto de calibração foram correlacionados com os valores de referência do teor de AMX, considerada a variável dependente, através de um modelo PLS. O número de variáveis latentes selecionadas através de

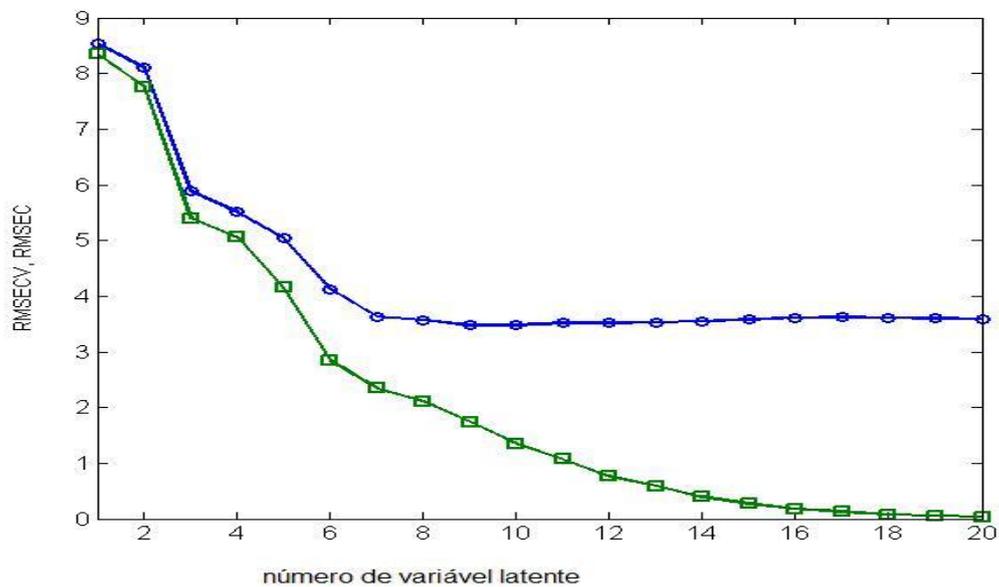
validação cruzada no modo “deixe uma fora” (do inglês *leave-one-out cross validation*) foi sete, correspondendo a um valor de RMSECV de 3,6 %. Esta escolha se baseou no gráfico apresentado na Figura 9, que mostra a variação do RMSECV e do RMSEC em função do número de variáveis latentes do modelo PLS. Os valores de RMSEC e RMSEP para este modelo foram 2,3 e 3,2 %, respectivamente.



**Figura 7.** Espectros das amostras de calibração e validação obtidos por transfectância



**Figura 8.** Espectros das amostras de calibração e validação obtidos por transfectância e pré-processados com MSC



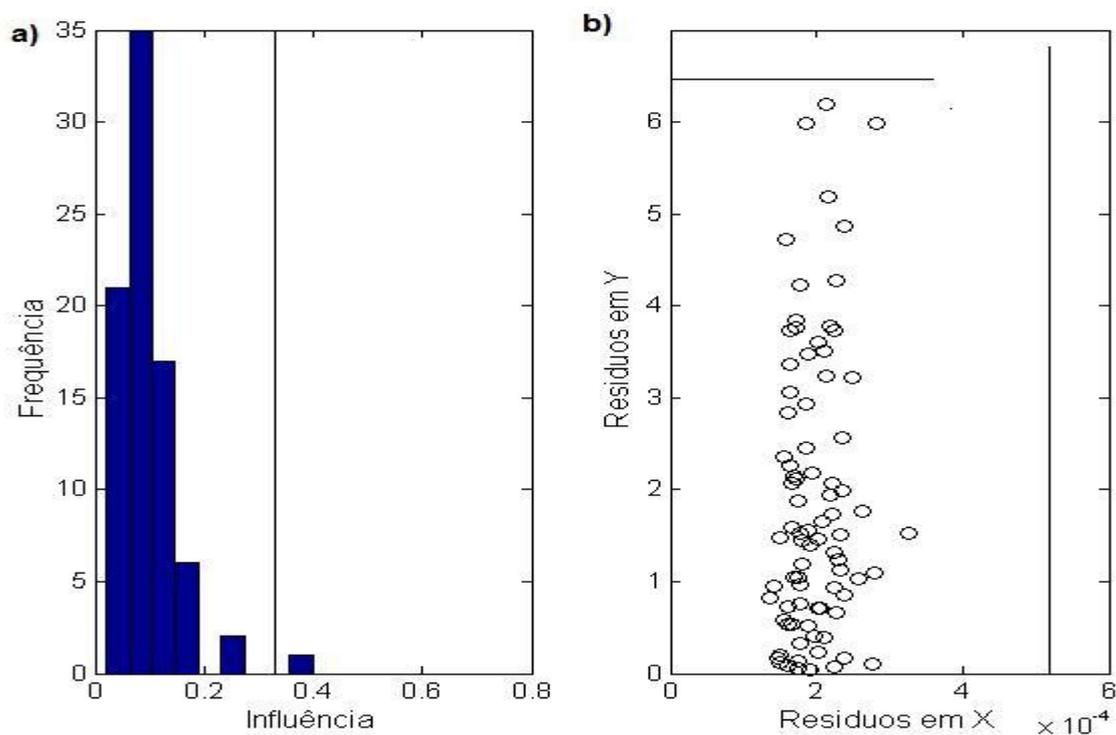
**Figura 9.** Valores do RMSECV (círculos) e do RMSEC (quadrados) em função do número de variáveis latentes

Considerando que o número de Variáveis Latentes (VLs) usadas no modelo foi igual a sete, constatou-se que os números de amostras usadas nos conjuntos de calibração e validação atenderam à norma da ASTM E1655-05 [9], a qual exige um

mínimo de  $6x(n^{\circ} \text{ de VLs} + 1)$  e  $4x(n^{\circ} \text{ de VLs})$  amostras para esses conjuntos, respectivamente. Neste trabalho, esses valores correspondem a 48 e 28 amostras, números inferiores aos utilizados.

### 3.3. Detecção de amostras anômalas e previsão das amostras de validação

Antes da elaboração do modelo final, o conjunto de calibração foi otimizado pela eliminação de amostras que apresentaram extrema influência (*leverage*), as quais são amostras que levam a erros muito altos, e/ou extremos valores de resíduos nos blocos X e/ou Y, de acordo com o procedimento relatado na seção 1.3 para a identificação de amostras anômalas. Uma vez identificado um *outlier*, este era retirado do conjunto e o modelo reconstruído. O processo foi repetido por três etapas, até que nenhum novo *outlier* foi detectado. Nas duas primeiras etapas foram eliminados dois *outliers* que apresentaram valores extremos de influência. Nenhum *outlier* foi observado com base nos resíduos. Após a otimização do conjunto de calibração, avaliou-se de modo semelhante o conjunto de validação, levando a eliminação de seis *outliers* que apresentaram resíduos nos valores de concentração (bloco Y) significativamente acima dos encontrados nas amostras de calibração. A detecção de *outliers* no conjunto de calibração está resumida na Tabela 3, que mostra a modificação dos parâmetros RMSEC, RMSECV e RMSEP em cada etapa. A Figura 10 ilustra a primeira etapa do processo de detecção de *outliers*, mostrando os gráficos de influência (Fig. 8a) e dos resíduos em Y em função dos resíduos em X (Fig. 8b), com os respectivos limites de confiança. Em cada uma das duas primeiras etapas para identificação de *outliers* foram detectados um *outlier* pelo critério de alta influência. Portanto, o modelo PLS final foi elaborado a partir de 80 amostras de calibração e 44 amostras de validação.



**Figura 10.** Detecção de *outliers* na primeira etapa para o conjunto de calibração. Gráficos a) de influência e b) dos resíduos de Y em função de X, com os respectivos limites de aceitação

**Tabela 3.** Evolução dos parâmetros para as três etapas na detecção de *outliers* no conjunto de calibração.

ETAPA	1	2	3	FINAL
N° Amostras Calibração	82	81	80	80
N° Amostras Validação	50	50	50	44
N° Variáveis Latentes	8	7	7	7
RMSEC (%)	2,3	2,5	2,3	2,3
RMSECV (%)	4,4	3,9	3,6	3,6
RMSEP (%)	3,9	4,8	4,9	3,2
h (influência)	1	1	0	---
X	0	0	0	---
Y	0	0	6	---

**Tabela 4.** Resultados da previsão das amostras do conjunto de validação.

<b>Valor de referência (%)</b>	<b>Valor previsto (%)</b>	<b>Erro relativo (%)</b>
96,82	96,12	-0,72
89,19	86,79	-2,70
120,71	121,30	0,49
91,73	86,84	-5,33
129,05	127,10	-1,51
122,86	120,25	-2,12
84,61	83,30	-1,54
84,68	79,40	-6,23
124,27	121,15	-2,51
120,80	119,58	-1,01
95,80	96,36	0,58
116,14	115,52	-0,53
111,05	117,40	5,72
81,58	79,22	-2,89
102,92	108,24	5,17
93,26	93,71	0,48
95,11	95,46	0,78
123,26	128,2	4,01
106,35	110,61	4,01
94,27	94,78	0,54
89,7	94,10	4,91
92,24	93,60	1,47
92,75	91,55	-1,29
94,78	95,42	0,68
98,85	98,97	0,12
100,88	103,89	2,98
101,39	105,74	4,29
102,41	103,78	1,34
103,43	108,24	4,65
103,93	110,15	5,98
106,48	108,17	1,59
91,21	85,66	-2,17
106,98	106,01	-0,91
108,51	108,30	-0,19
109,02	111,12	1,93
113,09	114,76	1,48
115,12	115,81	0,60
117,15	123,26	5,22
118,17	118,65	0,41
119,7	116,92	-2,32
122,24	121,16	-0,88
130,37	126,15	-3,24
88,80	91,10	2,59
97,06	93,05	-4,13

As amostras do conjunto de validação foram previstas com o modelo PLS otimizado. Os resultados para a previsão das 44 amostras são mostrados na Tabela 4. Os maiores erros relativos observados foram de  $\pm 6\%$ . Dessas 44 amostras, 23 (52%) apresentaram erros menores em módulo que 2,0%, que é o valor limite normalmente estabelecido para avaliação da exatidão de métodos analíticos e bioanalíticos na indústria farmacêutica [60].

### **3.4. Validação analítica**

A validação analítica do método proposto consistiu na estimativa das figuras de mérito linearidade, precisão, exatidão, intervalo, robustez, sensibilidade, sensibilidade analítica, limites de detecção e quantificação, além do teste para a presença de erro sistemático (viés). Conforme já mencionado, algumas dessas figuras foram estimadas da maneira tradicional, usada em calibração univariada, enquanto outras foram estimadas de maneira diferente da prescrita pela ANVISA [8]. A Tabela 5 resume a estimativa dessas figuras através do cálculo de vários parâmetros. A seguir, o resultado para cada figura de mérito será discutido individualmente.

**Tabela 5.** Figuras de mérito para o modelo PLS.

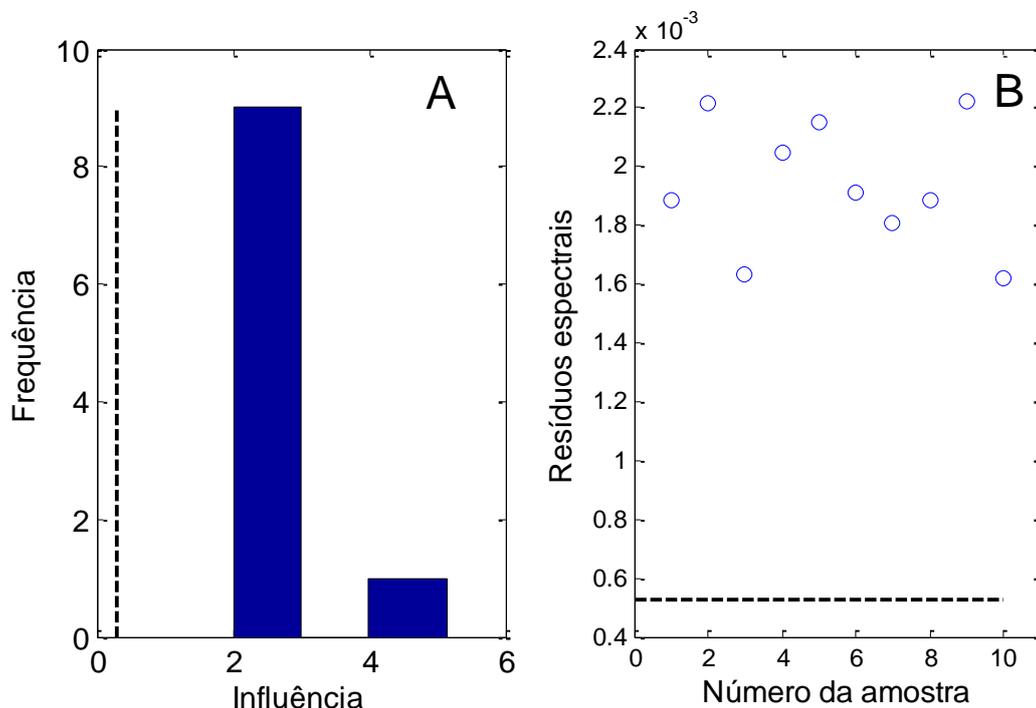
Figura de mérito	Parâmetro	Valor (mg/mL)
Exatidão	RMSECv (%)	1,8
	RMSEC (%)	1,2
	RMSEP (%)	1,6
Precisão	DPR repetibilidade	1,3 <sup>a</sup>
	DPR precisão intermediária	1,5 <sup>a</sup>
Linearidade/Ajuste	Inclinação	0,974 <sup>b</sup>
	Intercepto	2,42 <sup>b</sup>
	Coeficiente de correlação	0,989 <sup>b</sup>
Intervalo		80,0 a 130,0
SEL		1,4
SEN		1,8.10 <sup>-4</sup> <sup>c</sup>
$\gamma$		1,9 <sup>d</sup>
$\gamma^{-1}$		0,5
Limite de Detecção		1,8
Limite de Quantificação		5,4

<sup>a</sup> valores em %<sup>b</sup> valores para o ajuste da reta das amostras de calibração<sup>c</sup> valor expresso como a razão entre unidades de absorbância e %<sup>d</sup> mL/mg

### 3.4.1. Especificidade/Seletividade

Embora este trabalho considere que a SEL do método não tem nenhuma aplicação prática para o controle de qualidade para métodos de calibração multivariados, este valor foi estimado em 1,4%, indicando uma grande sobreposição dos interferentes no sinal da AMX. Além disso, a especificidade foi avaliada pela aplicação do método desenvolvido para a determinação de dez amostras contendo ampicilina ao invés de AMX, na faixa da 40,0 a 65,0 mg mL<sup>-1</sup>. Todas as dez amostras foram corretamente identificadas como *outliers*, com base em seus valores

de influência e nos altos resíduos espectrais (Figura 11). Assim o método foi capaz de distinguir outro princípio ativo.



**Figura 11.** Detecção de *outliers* nas amostras contendo ampicilina, ao invés de amoxicilina. Gráficos a) de influência e b) dos resíduos espectrais, com os respectivos limites de aceitação

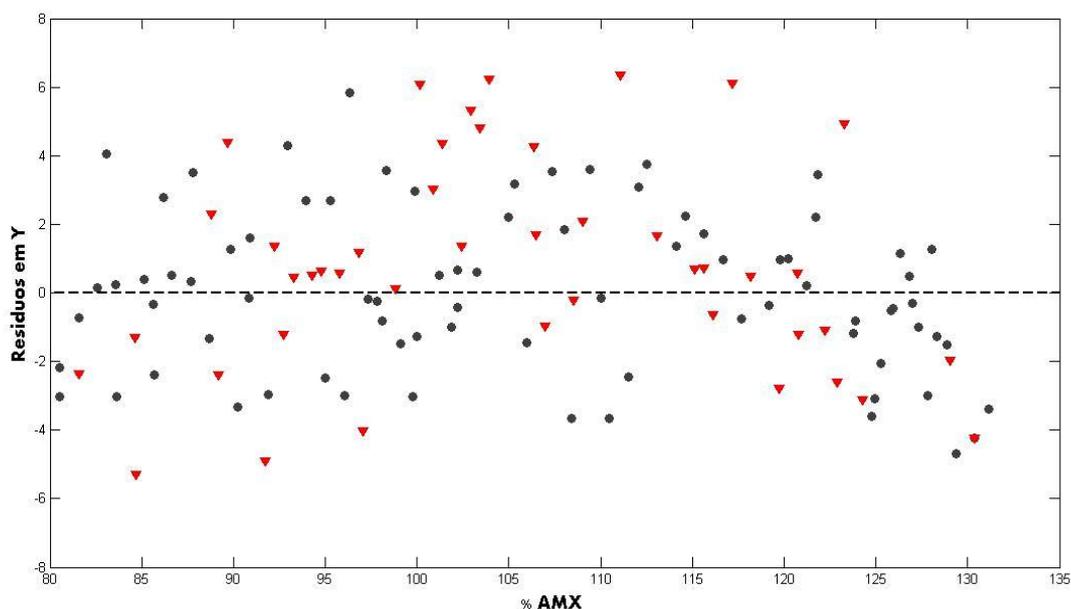
### 3.4.2. Linearidade

Conforme mencionado na seção 1.2.2, a linearidade do método apenas pode ser acessada qualitativamente para modelos de calibração multivariada, através do gráfico de resíduos. Após a constatação da aleatoriedade dos resíduos, indicando a validade do modelo linear, pode-se utilizar o coeficiente de correlação para inferir sobre a concordância entre os valores estimados e de referência pelo ajuste de uma reta entre eles.

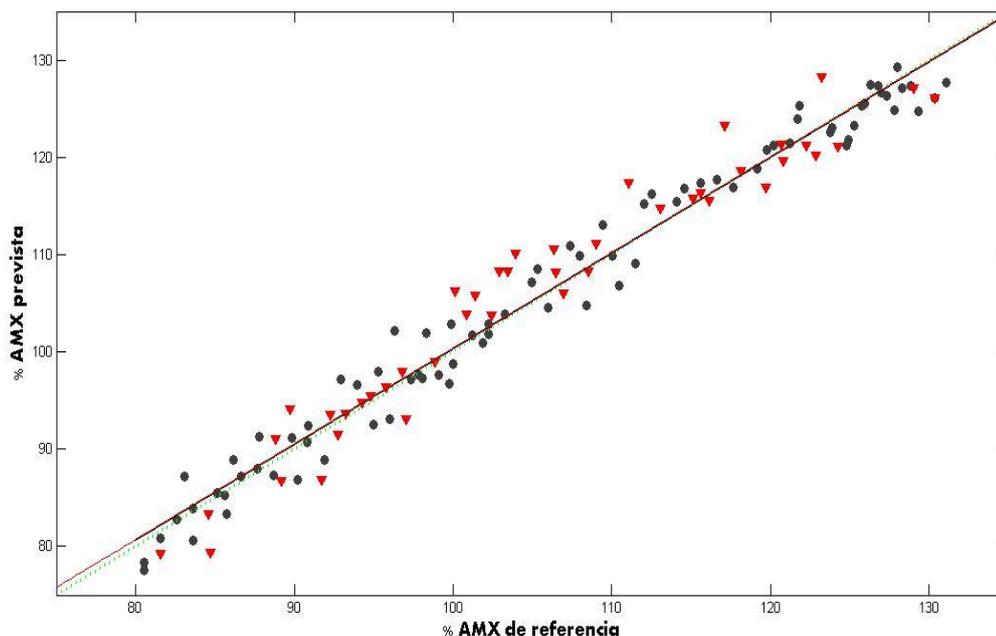
Avaliando-se a distribuição dos resíduos das amostras de calibração e validação, apresentados na Figura 12, nota-se a ausência de comportamento sistemático, indicando a validade do modelo linear.

O ajuste da reta entre os valores de referência e os valores estimados pode ser feito somente para as amostras de calibração, somente para as amostras de

validação, ou conjuntamente para ambas. De acordo com as normatizações existentes, considera-se correto avaliar o ajuste da reta para as amostras de calibração. Os valores dos coeficientes angular (inclinação) e linear (intercepto) são mostrados na Tabela 5, além do coeficiente de correlação ( $r$ ), que foi de 0,989. Este valor atende às exigências da ANVISA [8], que estabelece um valor mínimo de  $r$  aceitável de 0,99 (2 casas decimais). Os ajustes para as retas somente com as amostras de validação e contendo ambos os conjuntos forneceram valores de  $r$  de 0,972 e 0,983, respectivamente. A Figura 13 mostra a reta ajustada conjuntamente para as amostras de calibração e validação (linha cheia). A linha pontilhada indica o ajuste ideal.



**Figura 12.** Resíduos do modelo para as amostras de calibração (círculos) e de validação (triângulos)



**Figura 13.** Valores de referência (linha cheia) contra os previstos para as amostras de calibração (círculos) e de validação (triângulos). A linha pontilhada indica o ajuste ideal

### 3.4.3. Precisão

A precisão foi avaliada nos níveis de repetibilidade e precisão intermediária. A repetibilidade foi obtida a partir de seis replicatas autênticas obtidas por um mesmo analista em um mesmo dia para medidas de uma amostra com 100 % de AMX. O desvio-padrão relativo (DPR) estimado foi de 1,3%. A precisão intermediária foi obtida comparando-se outras seis replicatas obtidas por outro analista em outro dia. Conforme esperado, o DPR para o conjunto das amostras foi igual a 1,5%, um valor um pouco superior ao obtido no estudo de reprodutibilidade. Ambos os resultados atendem às exigências da ANVISA [8], que estabelece um DPR máximo aceitável de 5%.

### 3.4.4. Exatidão

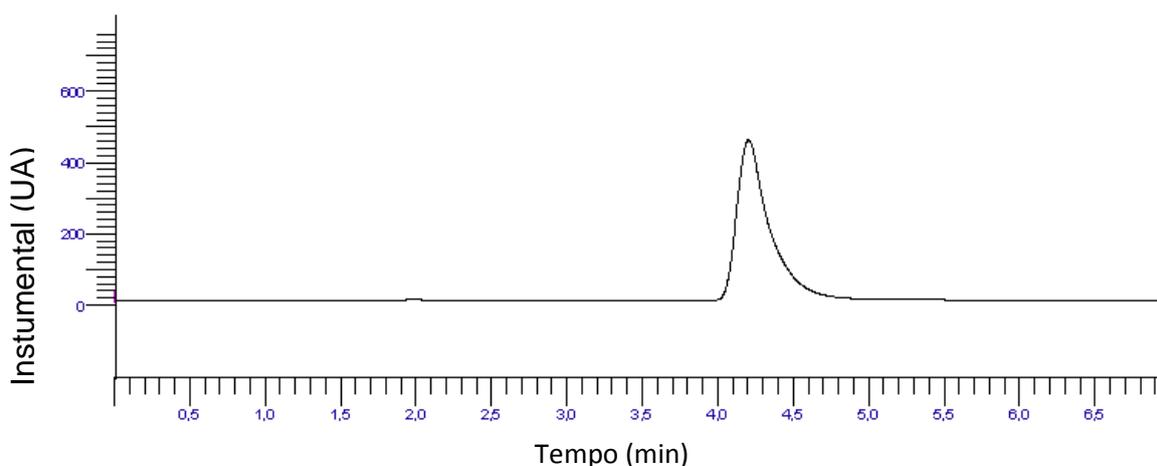
Conforme mencionado na seção 1.2.2, a exatidão de modelos de calibração multivariada pode ser avaliada a partir dos parâmetros RMSEC, RMSECV e RMSEP, cujos valores para o modelo estudado são mostrados na Tabela 5. O

parâmetro mais usado para este fim é o RMSEP. Contudo, conforme já mencionado na seção 1.1.2, esse parâmetro apenas representa uma medida do erro médio apresentado pelo modelo, não sendo suficiente para atestar a exatidão do método. Além disso, a ANVISA não reconhece este parâmetro e requer a avaliação da exatidão do método a partir de triplicatas de três pontos. Desta maneira, foram obtidas replicatas de três níveis, um inferior, um médio e um superior, as quais foram analisadas pelo método quimiométrico. Triplicatas de amostras independentes nos mesmos níveis foram também analisadas pelo método cromatográfico oficial [36], descrito na seção 2.3.4. Os resultados são mostrados na Tabela 6. Testes t não pareados com quatro graus de liberdade e 95% de confiança mostraram que os resultados dos dois métodos não apresentam diferença significativa. Na Figura 14 é mostrado um dos cromatogramas, obtido para uma amostra no nível médio de teor de AMX.

**Tabela 6.** Avaliação da exatidão através da determinação de três amostras pelo método quimiométrico proposto e pelo método oficial (CLAE).

Nível com Teor de AMX	Método Oficial (mg mL <sup>-1</sup> ) <sup>a</sup>	Método Proposto (mg mL <sup>-1</sup> ) <sup>a</sup>
Médio	39.5 ± 0.2	39.6 ± 0.5
Inferior	49.3 ± 0.3	49.4 ± 0.8
Superior	64.6 ± 0.5	63.1 ± 0.7

<sup>a</sup> Média dos valores e desvios padrão de três determinações



**Figura 14.** Cromatograma obtido para uma amostra no nível médio de teor AMX pelo método oficial (CLAE)

### 3.4.5. Intervalo

Considerando os estudos de linearidade, exatidão e precisão, o intervalo analítico do método proposto foi de 40,0 a 65,0 mg mL<sup>-1</sup> de AMX, correspondendo a uma faixa de 80,0 a 130,0 % de teor de AMX na formulação estudada.

### 3.4.6. Robustez

A robustez do método foi avaliada em relação à temperatura das medidas. Os resultados obtidos para triplicatas de amostras preparadas com 100,00 % de AMX (50,00 mg mL<sup>-1</sup>), analisadas em três diferentes temperaturas são mostrados na Tabela 7. Testes t com dois graus de liberdade cada e 95 % de confiança mostraram não haver diferenças significativas entre os resultados obtidos e o valor de referência. Portanto, o método foi considerado robusto em relação a variações de temperatura entre 15 e 35°C.

**Tabela 7.** Avaliação da robustez do método proposto na determinação de uma amostra de 100,00 % de AMX em três níveis diferentes de temperatura.

Temperatura (°C)	Teor de AMX (mg mL <sup>-1</sup> ) <sup>a</sup>	Recuperação em (%)
15	51,4 ± 0,7	102,8 ± 1,4
25	49,4 ± 0,8	98,8 ± 1,6
35	50,8 ± 0,6	101,6 ± 1,2

<sup>a</sup> Média dos valores e desvios padrão de três determinações

### 3.4.7. Sensibilidade e Sensibilidade analítica

A SEN foi calculada a partir dos coeficientes de regressão do modelo, conforme descrito na seção 1.2.2. O valor estimado foi  $1,8 \times 10^{-4}$  (Tabela 5). Este valor não é adequado para comparação entre diferentes métodos analíticos, pois é dependente da técnica empregada e da matriz analisada. Por isso, estimou-se também a sensibilidade analítica ( $\gamma$ ), dividindo-se o valor de SEN pela estimativa do ruído instrumental ( $\varepsilon$ ). O valor de  $\varepsilon$  foi estimado em  $1,9 \times 10^{-4}$  a partir do módulo de Frobenius da matriz de resíduos do modelo PLS equação (21) [21]. Dessa maneira, assume-se que o ruído instrumental é a única fonte de erro significativa. Portanto, foi

obtido um valor de  $\gamma$  igual a 0,95 (1/%). O inverso deste valor ( $\gamma^{-1}$ ) indica que o método é capaz de discriminar uma diferença de 1,06 % no teor de AMX, correspondendo a uma concentração de 0,53 mg mL<sup>-1</sup>, uma estimativa que considera apenas o ruído como fonte de erro. Esse resultado corrobora a boa qualidade do método para o fim ao qual ele se destina. Além disso, também concorda com a conhecida sensibilidade da técnica NIR, que é capaz de realizar determinações em concentrações da ordem de 0,1%, correspondendo a 1,0 mg mL<sup>-1</sup> [30].

#### **3.4.8. Limites de detecção e quantificação**

Os valores de LD e LQ foram estimados em 3,6 e 10,5 %, respectivamente, correspondendo a concentrações de 1,8 e 5,4 mg mL<sup>-1</sup> de AMX, respectivamente. Estes valores foram considerados adequados à finalidade do método proposto.

#### **3.4.9. Teste para a presença de erro sistemático (viés ou *bias*)**

Este teste foi realizado conforme descrito na seção 1.2.2, visando avaliar a existência de erros sistemáticos nas estimativas para as amostras de validação através do ajuste da reta dos valores de referência contra os estimados. O coeficiente angular estimado para essa reta foi  $0,9998 \pm 0,0372$  e o coeficiente linear foi  $0,3725 \pm 1,9759$ . Dois testes t, com 95% de confiança e 42 graus de liberdade, mostraram que não existiam diferenças significativas entre os valores estimados e de referência, uma vez que os coeficientes angular e linear da reta são estatisticamente iguais aos valores esperados ou ideais, um e zero, respectivamente. Dessa forma, foi constatada a ausência de tendência sistemática nos resultados e confirmada a exatidão do método proposto.

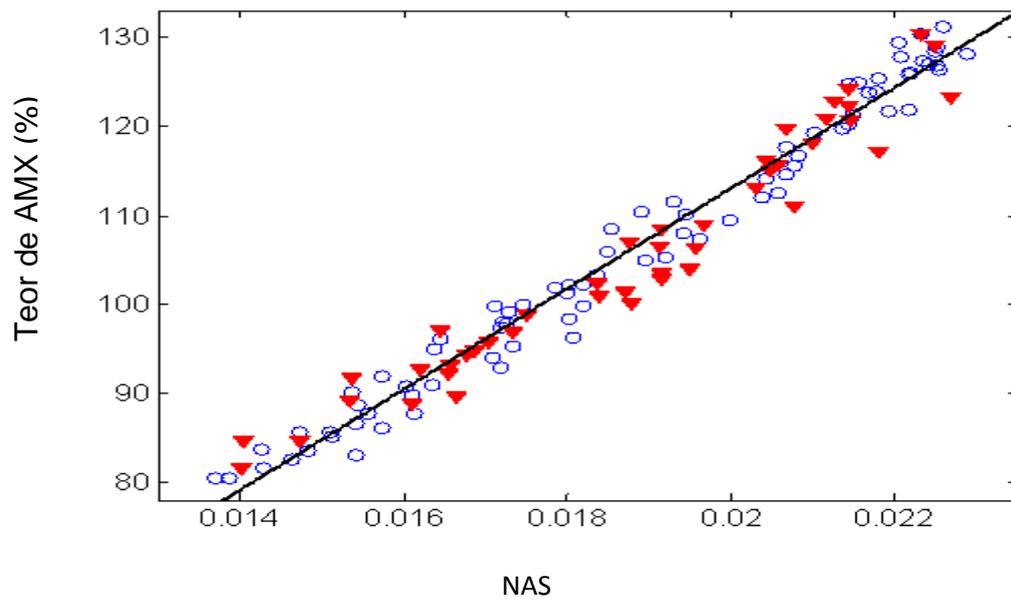
### 3.5. Modelo de calibração pseudo-univariado baseado no NAS

Quando se constrói modelos de calibração multivariada PCR ou PLS, não é possível representar uma curva de calibração de maneira tradicional. Por outro lado, conforme descrito no final da seção 1.2.1, é possível construir uma curva de calibração chamada de pseudo-univariada graficando os valores do módulo do vetor **nas** para cada amostra em função das concentrações de referência (**y**) [13]. Como o vetor **nas** para uma amostra indica uma direção que é afetada somente pelas mudanças na concentração do analito, ele pode ser usado em um procedimento totalmente seletivo para a sua determinação [61]. Isto equivale a obter um sinal analítico univariado “totalmente filtrado”, com o qual se pode construir uma relação linear com a concentração.

A curva de calibração pseudo-univariada obtida neste trabalho, construída tanto com as amostras de calibração, quanto de validação, é mostrada na Figura 15. A equação da reta ajustada para esta curva foi:

$$[AMX] = 5654,7||\mathbf{n\hat{a}s}|| + 0,0014682 \quad \text{Equação (29)}$$

O valor de *r* estimado foi 0,9885. Desta maneira, foi possível desenvolver um modelo similar ao de calibração univariada, tornando mais clara a interpretação dos resultados.



**Figura 15.** Curva de calibração pseudo-univariada baseada no **[NAS]**. (●) amostras de calibração, (▼) amostras de validação, (—) reta ajustada para as amostras de calibração

## 4. Conclusões

Um método baseado em espectroscopia no infravermelho próximo e calibração multivariada foi desenvolvido para a determinação do teor de AMX em uma formulação farmacêutica (pó para suspensão), através de medidas de transfectância de amostras em suspensão.

O método desenvolvido é cerca de 8 vezes mais rápido que os métodos propostos pelas Farmacopéias americana e brasileira, que se baseiam em CLAE [35], titulação iodimétrica e ensaios microbiológicos [53]. A estimativa de tempo gasto por análise é de 5 min para o método desenvolvido contra 40 minutos para o método cromatográfico. Além disso, este método apresentou outras vantagens, tais como um procedimento analítico simplificado, com menor intervenção humana, não uso de reagentes ou solventes orgânicos e não geração de resíduos químicos.

Outro aspecto importante foi a validação do método proposto, procurando atender as normas da ANVISA [8] e as mais recentes propostas de estimativa de figuras de mérito para métodos multivariados, num procedimento completo, que incluiu seleção de amostras e detecção de *outliers*. Mais do que a determinação específica de AMX, este tipo de procedimento possui potencial crescente de aplicação no desenvolvimento de novos métodos quimiométricos. O método desenvolvido aqui foi considerado exato, preciso, linear, robusto e sensível na faixa de 40,0 a 65,0 mg mL<sup>-1</sup> de AMX, correspondendo de 80,0 a 130,0 % de teor do princípio ativo na formulação estudada.

## 5. REFERÊNCIAS BIBLIOGRÁFICAS

- [1] Bruns, R.E.; Neto, B.B.; Scarminio, I.S.; “25 Anos de quimiometria no Brasil”, *Quím. Nova* **29**:1401 (2006).
- [2] Bruns, R.E.; Faigle, J.F.G.; “Quimiometria”, *Quím. Nova* **8**:84 (1985).
- [3] Sena, M.M., “Aplicação de métodos quimiométricos de primeira e segunda ordem na determinação direta de fármacos por espectroscopia molecular”, Tese de Doutorado, UNICAMP, Campinas, 2004.
- [4] Olivieri, C.A.; “Analytical advantages of multivariate data processing. One, two, three, infinity?”, *Anal. Chem.* **80**:5713 (2008).
- [5] Sena, M.M.; Freitas, C.B.; Silva, L.C.; Braga, J.W.B.; Poppi, R.J.; “Aplicação de Calibração Multivariada no Controle de Qualidade de Fármacos”, *Estudos* **32**:1953 (2005).
- [6] Ferreira, M.C.; Antunes, A.M.; Melgo, M.S.; Volpe, P.L.O.; “Quimiometria I: calibração multivariada, um tutorial”, *Quím. Nova* **22**:724 (1999).
- [7] Rocha, F.C.W.; “Utilização do sinal analítico líquido para validação de modelos de calibração multivariada através do cálculo de figuras de mérito e de cartas de controle”, Dissertação de Mestrado, UNICAMP, Campinas, 2007.
- [8] Agência Nacional de Vigilância Sanitária (ANVISA); “Guia para Validação de Métodos Analíticos e Bioanalíticos”, Resolução – RE nº 899, Brasil, 2003.
- [9] Annual Book of ASTM Standards, Standard Practices for Infrared Multivariate Quantitative Analysis-E1655-00, ASTM International, West Conshohocken: Pennsylvania, USA, 2000.
- [10] Braga, J.W.B.; “Aplicação e validação de modelos de calibração de segunda ordem em química analítica”, Tese de Doutorado, UNICAMP, Campinas, 2008.
- [11] Valderrama, P.; “Avaliação de figuras de mérito em calibração multivariada na determinação de parâmetros de controle de qualidade em indústria alcooleira por espectroscopia no infravermelho próximo”, Dissertação de Mestrado, UNICAMP, Campinas, 2005.
- [12] Lorber, A.; “Error propagation and figures of merit for quantification by solving matrix equation”, *Anal. Chem.* **58**:1167(1986).

- [13] Ferre, J.; Brown, S.D.; Rius, F.X.; "Improved calculation of the net analyte signal in inverse multivariate calibration", *J. Chemom.* **15**:537 (2001).
- [14] Faber, N.M.; "Exact presentation of multivariate calibration model as univariate calibration graph", *Chemom. Intell. Lab. Sys.* **50**:114. (2000).
- [15] Boqué, R.; Rius, F.X.; "Multivariate detection limits estimators", *Chemom. Intell. Lab. Sys.* **32**:11 (1996).
- [16] Valderrama, P.; Braga, J.W.B; Poppi, R.; "Estado da arte em calibração multivariada", *Quím. Nova* **32**:1278 (2009).
- [17] Olivieri, C.O.; Faber, N.K.M.; Ferré, J.; Boqué, R.; Kalivas, J. H.; Mark, H.; "Uncertainty estimation and figures of merit for multivariate calibration", *Pure Appl. Chem.* **78**: 633 (2006).
- [18] Ferreira, M.H.; Gomes, J.F.F.; Sena, M.M.; "Development and validation of a multivariate calibration model for determination of dipyrone in oral solutions by near infrared spectroscopy", *J. Braz. Chim. Soc.* **20**:1680 (2009).
- [19] Beebe, K. R.; Kowalski, B. R.; "An Introduction to multivariate calibration and analysis", *Anal. Chem.* **59**: 1007, (1987).
- [20] Rodriguez, L.C.; Campaña, A.M.G.; Jimenez-Linares C.; Román-Ceba, M.; "Estimation of performance characteristics of an analytical method using the data set of the calibration experiment", *Anal. Lett.* **26**:1243 (1993).
- [21] Braga, J.W.B; Carneiro, R.L.; Poppi, R.; "Evaluation of the number of factors needed for residual bilinearization in BLLS and UPLS models to achieve the second-order advantage", *Chemom. Intell. Lab. Syst.* **100**:99 (2010).
- [22] Currie, L.A.; "Nomenclature in evaluation of analytical methods including detection and quantification capabilities (IUPAC Recommendations 1995)", *Anal. Chim. Acta* **391**:105 (1999).
- [23] Miller, J.N.; Miller, J.C.; "Statistics and Chemometrics for Analytical Chemistry", 5<sup>th</sup> ed., Pearson-Prentice Hall: New York, 2005.
- [24] Valderrama, P.; Braga, J.W.B; Poppi, R.; "Variable selection, outlier detection, and figures of merit estimation in a partial least-squares regression multivariate calibration model. A case study for the determination of quality parameters in the alcohol industry by near-infrared spectroscopy", *J. Agric. Food Chem.* **55**:8331 (2007).

- [25] Martens, H.; Naes, T.; "Multivariate Calibration", John Wiley & Sons: New York, 1989.
- [26] Walczak, B.; Massart, D.L.; "Multiple outlier detection revisited", *Chemom. Intell. Lab. Syst.* **41**:1 (1998).
- [27] Fernández Pierna, J.A.; Wahl, F.; Noord, O.E.; Massart, D.L.; "Methods for outlier detection in prediction", *Chemom. Intell. Lab. Syst.* **63**:27 (2002).
- [28] Pasquini, C.; "Near Infrared Spectroscopy: Fundamentals, Practical Aspects and Analytical Applications", *J. Braz. Chim. Soc.*, **14**:198 (2003).
- [29] Hart, J.H.; Norris, K.H.; *J. Near Infrared Spectrosc.* **23**: 4 (1996).
- [30] Burns, D. A.; Ciurczak, E. W. (eds.); "Handbook of Near-Infrared Analysis", 2<sup>nd</sup> ed., Taylor & Francis: New York, 2001.
- [31] Neto, T.E.; Neto T.A.A.; "Modificação química de argilas: desafios científicos e tecnológicos para obtenção de novos produtos com maior valor agregado", *Quím. Nova* **32**:809 (2009).
- [32] Camolesi, V.J.; "Caracterização de querosene através da espectroscopia de infravermelho próximo", Dissertação de mestrado, USP, São Paulo, (2009).
- [33] Bird, A.E.; "Amoxicillin", em: Brittain, H.G. (ed.); *Analytical Profiles of Drug Substances* **23**:1 (1994).
- [34] IQUEGO; Amoxicilina, bula farmacêutica, 2006.
- [35] Belal, F., El-Brashy, A., Ibrahim, F.; "Titrimetric determination of some penicillins using 2-iodoxybenzoate", *J. AOAC* **73**:896 (1990).
- [36] The United States Pharmacopoeia, 25<sup>th</sup> rev., U.S.P. Convention: Rockville, 2002.
- [37] Tsou, T.L.; Wu, J.R.; Young, C.D.; Wang, T.M.; "Simultaneous determination of amoxicillin and clavulanic acid in pharmaceutical products by HPLC with beta-cyclodextrin stationary phase", *J. Pharm. Biomed. Anal.* **15**:1197 (1997).
- [38] Li, Y.M.; van Schepdael, A.; Zhu, Y.; Roets, E.; Hoogmartens, J.; "Development and validation of amoxicillin determination by micellar electrokinetic capillary chromatography", *J. Chromat.* **812**:227 (1998).
- [39] Wang, P.; Qi, M.; Sun, Y.; Yang, J.; "A liquid chromatographic method for simultaneous determination of amoxicillin sodium and sulbactam sodium in a combination formulation", *J. Pharm. Biomed. Anal.* **36**:565 (2004).

- [40] Nagaralli, B.S.; Seetharamappa, J.; Melwanki, M.B.; "Sensitive spectrophotometric methods for the determination of amoxicillin, ciprofloxacin and piroxicam in pure and pharmaceutical formulations", *J. Pharm. Biomed. Anal.* **29**:859 (2002).
- [41] Liu, H.; Wang, H.W., Sunderland, V.B.; "An isocratic ion exchange HPLC method for the simultaneous determination of flucloxacillin and amoxicillin in a pharmaceutical formulation for injection", *J. Pharm. Biomed. Anal.* **37**:395 (2005).
- [42] Salem, H.; "Selective spectrophotometric determination of phenolic beta-lactam antibiotics in pure forms and in their pharmaceutical formulations", *Anal. Chim. Acta* **515**:333 (2004).
- [43] Al-Abachi, M.Q.; Haddi, H.; Al-Abachi, A.M.; "Spectrophotometric determination of amoxicillin by reaction with N,N-dimethyl-p-phenylenediamine and potassium hexacyanoferrate(III)", *Anal. Chim. Acta* **554**:184 (2005).
- [44] Li, Y.H.; Lu, J.R.; "Chemiluminescence flow-injection analysis of beta-lactam antibiotics using the luminol-permanganate reaction", *Luminescence* **21**:251 (2006).
- [45] El Walily, A.F.M.; Gazy, A.A.K.; Belal, S.F.; Khamis, E.F.; "Selective spectrofluorimetric determination of phenolic beta-lactam antibiotics through the formation of their coumarin derivatives", *J. Pharm. Biomed. Anal.* **20**:643 (1999).
- [46] Li, Q.M.; Yang, Z.J.; "Study of spectrophotometric determination of amoxicillin using sodium 1,2-naphthoquinone-4-sulfonate as the chemical derivative chromogenic reagent", *Anal. Lett.* **39**:763 (2006).
- [47] Du, J.X.; Li, Y.H.; Lu, J.R.; "Chemiluminescence flow injection analysis of amoxicillin by a permanganate-based reaction", *Anal. Lett.* **35**:2295 (2002).
- [48] Pajchel, G.; Pawlowski, K.; Tyski, S.; "CE versus LC for simultaneous determination of amoxicillin/clavulanic acid and ampicillin/sulbactam in pharmaceutical formulations for injections", *J. Pharm. Biomed. Anal.* **29**:75 (2002).
- [49] Uslu, B.; Biryol, I.; "Voltammetric determination of amoxicillin using a poly (N-vinyl imidazole) modified carbon paste electrode", *J. Pharm. Biomed. Anal.* **20**:591 (1999).

- [50] Bergamini, M.F.; Teixeira, M.F.S.; Dockal, E.R.; Bocchi, N.; Cavalheiro, E.T.G; "Evaluation of different voltammetric techniques in the determination of amoxicillin using a carbon paste electrode modified with [N,N'-ethylenebis(salicylideneaminato)] oxovanadium(IV)", *J. Electr. Soc.* **153**:E94 (2006).
- [51] Rezaei, B.; Damiri, S.; "Electrochemistry and adsorptive stripping voltametric determination of amoxicillin on a multiwalled carbon nanotubes modified glassy carbon electrode", *Electroanalysis* **21**:1577 (2009).
- [52] Qu, N.; Zhub, M.; Mic, H.; Doud, Y.; Rena, Y.; " Nondestructive determination of compound amoxicillin powder by NIR spectroscopy with the aid of chemometrics" *Spect. Acta A* **70**:1146 (2008).
- [53] Farmacopéia Brasileira, 4<sup>a</sup> ed., Ateneu Editora: São Paulo, Brasil, 2002.
- [54] Kennard, R.W.; Stone, L.A.; "Computer aided design experiments" *Technometrics* **11**: 137 (1969).
- [55] Williams, P., "Sampling, Sample Preparation, and Sample Selection", em: Burns, D. A.; Ciurczak, E. W. (eds.); "Handbook of Near-Infrared Analysis", 2<sup>nd</sup> ed., Taylor & Francis: New York, 2001.
- [56] Workman Jr., J.J.; "Interpretive spectroscopy for near infrared", *Appl. Spectrosc. Rev.* **31**: 251 (1996).
- [57] Isaksson, T.; Naes, T.; "The effect of multiplicative scatter correction (MSC) and linearity improvement in NIR spectroscopy", *Appl. Spectrosc.* **42**:1273 (1988).
- [58] Dhanoa, M.S.; Lister, S.J.; Barners, R.J.; "On the scales associated with near-infrared reflectance difference spectroscopy", *Appl. Spectrosc.* **49**:765 (1995).
- [59] Savitzky, A.; Golay, M.J.E.; "Smoothing and differentiation of data by simplified least squares procedure", *Anal. Chem.* **36**:1627 (1967).
- [60] Huber L.; "Validation and Qualification in Analytical Laboratories", 2<sup>nd</sup> ed., LabCompliance: New York, 2007.
- [61] Rocha, W.F.C.; Rosa, A.L.; Martins, J.A.; Poppi, R.J.; "Determination and validation of nimesulide in pharmaceutical formulations by near infrared spectroscopy", *J. Braz. Chem. Soc.* **21**:1929 (2010).