



*Universidade Estadual de Goiás*

*Unidade Universitária de Ciências Exatas e Tecnológicas (UnUCET)*

*PROGRAMA DE PÓS-GRADUAÇÃO STRICTO SENSU EM CIÊNCIAS*

*MOLECULARES*

**Desenvolvimento e Validação de um Método  
Quimiométrico para a Determinação de  
Hidroclorotiazida em Formulação Farmacêutica  
usando Medidas de Reflectância Difusa no  
Infravermelho Próximo**

**Aluno: Marcus Henrique Ferreira**

*Anápolis*

*Julho/2012*

**Marcus Henrique Ferreira**

**Desenvolvimento e Validação de um Método  
Quimiométrico para a Determinação de  
Hidroclorotiazida em Formulação Farmacêutica  
usando Medidas de Reflectância Difusa no  
Infravermelho Próximo**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ciências Moleculares da Universidade Estadual de Goiás para obtenção do título de Mestre em Ciências Moleculares. Área de Concentração: Físico–Química Molecular.

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

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

***Anápolis***

***Julho/2012***

DESENVOLVIMENTO E VALIDAÇÃO DE UM MÉTODO QUIMIOMÉTRICO PARA A  
DETERMINAÇÃO DE HIDROCLOROTIAZIDA EM FORMULAÇÃO  
FARMACÊUTICA USANDO MEDIDAS DE REFLECTÂNCIA DIFUSA NO  
INFRAVERMELHO PRÓXIMO

**MARCUS HENRIQUE FERREIRA**

Dissertação apresentada ao corpo docente do Programa de Pós-Graduação *Stricto Sensu* em Ciências Moleculares da Universidade Estadual de Goiás apresentada como parte dos requisitos necessários para obtenção do título de Mestre em Ciências Moleculares.

Aprovada por:



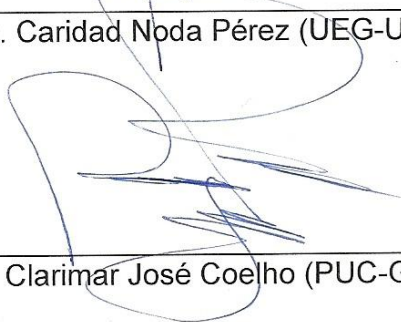
Prof. Dr. Marcelo Martins de Sena (UEG-UFMG) – Orientador



Prof. Dr. Jez Willian Batista Braga (UNB) – Co-Orientador



Profª. Dra. Caridad Noda Pérez (UEG-UFMG) - Membro



Prof. Dr. Clarimar José Coelho (PUC-GO) - Membro

ANÁPOLIS, GO - BRASIL  
JULHO 2012

Os investimentos em conhecimento são os que geram os melhores dividendos.

Benjamin Franklin

Dedico este trabalho aos meus pais, José Ferreira Ribeiro e Maria Roza Ribeiro, que já não se encontram em nosso convívio, mas os ensinamentos me deixados por eles tornaram meus sonhos mais próximos e facilitaram meu caminhar.

## Agradecimentos

- Primeiramente a Deus, que me permitiu tropeçar, cair e até me machucar, mas me curou e me deu forças para levantar e seguir mais forte para enfrentar novos desafios.
- Ao Prof. Dr. Marcelo Martins de Sena, pela amizade, paciência e pelos imensuráveis conhecimentos transmitidos.
- Ao Prof. Dr. Jez Willian Batista Braga, pela paciência e atenção ao responder os emails e aos longos telefonemas em diversas discussões.
- À minha querida esposa Jeaninny Ribeiro de Bastos, pela paciência, companheirismo e incentivo; e aos meus filhos Fabianno Otávio e Amanda Laurien por alegrar meus dias e motivar minha busca por dias ainda melhores.
- Aos meus doze irmãos que, cada um do seu modo, incentivou a minha busca por novos conhecimentos e desafios.
- Ao amigo Jorge Fernando Ferreira Gomes, pela confiança e disposição para enfrentar desafios, dando início à ideia de fazer este trabalho.
- À IQUEGO que forneceu toda a estrutura física e instrumental para a execução das etapas experimentais deste trabalho.
- Aos funcionários e estagiários do Laboratório de Controle da Qualidade da IQUEGO, que auxiliaram na preparação das amostras e execução de análise.
- À amiga e incentivadora, Alice Helena Rios e aos amigos Daniel Jesus, Eduardo Soares e Fritz Kasbaum, pelo incentivo e longas conversas e discussões técnicas.
- À Universidade Estadual de Goiás e a todos os professores do programa de mestrado que contribuíram grandiosamente para minha formação.

# CURRICULUM VITAE

## DADOS PESSOAIS

**Nome:** Marcus Henrique Ferreira

**Nascimento:** 10/09/1975 - Anicuns - GO

**E-mail:** marcushenriquef@yahoo.com.br

## FORMAÇÃO ACADÊMICA

Graduação em Farmácia-Bioquímica com Formação Generalista - 2007 - 2009  
Universidade Paulista, UNIP, Goiânia-GO, Brasil

Graduação em Ciências - Habilitação em Química - 2000 - 2003  
Universidade Estadual de Goiás, UEG, Anápolis-GO, Brasil

## ATUAÇÃO PROFISSIONAL

Indústria Química do Estado de Goiás – IQUEGO - 2005 - Atual  
Vínculo: Servidor público.

Enquadramento funcional: Supervisor - Controle de Qualidade Físico-Químico

Secretaria Estadual de Educação de Goiás – SEE - 2004 – 2005.  
Vínculo: Servidor público.

Enquadramento funcional: Professor de Química - nível médio

Laboratório Teuto Brasileiro S/A – TEUTO - 2004 - 2005  
Vínculo: Celetista formal.

Enquadramento funcional: Analista de Laboratório de Controle da Qualidade Físico-Químico

## ATIVIDADES ACADÊMICAS

Co-Orientação - Trabalhos de conclusão de curso de graduação

Gomes, J. F. F. “Análise de Figuras de Mérito na Calibração Multivariada, em Formulação Farmacêutica de Dipirona, por Espectroscopia de Infravermelho Próximo – NIRS”. 2008. Curso (Tecnologia Química Agroindustrial) - Centro Federal de Educação Tecnológica de Goiás

## **ARTIGOS PUBLICADOS**

Silva, M. A.M., FERREIRA, M. H., Braga, J. W.B., Sena, M.M. "Development and analytical validation of a multivariate calibration method for determination of amoxicillin in suspension formulations by near infrared spectroscopy". Talanta, v.89, p.342 - 351, 2012.

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". Journal of the Brazilian Chemical Society, v.20, p.1680 - 1686, 2009.

## **ARTIGO ACEITO PARA PUBLICAÇÃO**

FERREIRA, M. H., Braga, J. W. B., Sena, M. M. "Development and validation of a chemometric method for direct determination of hydrochlorothiazide in pharmaceutical samples by diffuse reflectance near infrared spectroscopy". Microchemical Journal, 2012

## **TRABALHOS RESUMIDOS PUBLICADOS EM ANAIS DE EVENTOS**

Ferreira, M. H., Sena, M. M., Braga, J. Desenvolvimento de um método quimiométrico para determinação de hidroclorotiazida por espectroscopia NIR, 16º ENQA - Campos do Jordão – SP, 2011.

Silva, M., Ferreira, M. H., Braga, J., Sena, M. M. Desenvolvimento e Validação Analítica de um Método Quimiométrico para a Determinação de Amoxicilina no Infravermelho Próximo, 2011- 34º Reunião Anual SBQ – Florianópolis-SC.

Ferreira, M. H., Gomes, J. F. F., Sena, M. M. Determinação de Dipirona em Formulação Farmacêutica por Espectroscopia no Infravermelho Próximo e Calibração Multivariada – 30º Reunião Anual SBQ, Águas de Lindóia – SP, 2007.

Ferreira, M. H., Gomes, J. F. F., Sena, M. M. Validação de um Modelo de Calibração Multivariada para a Determinação de Dipirona no Infravermelho Próximo - 14º ENQA, João Pessoa - Paraíba, 2007.



## RESUMO

### **Desenvolvimento e Validação de um Método Quimiométrico para a Determinação de Hidroclorotiazida em Formulação Farmacêutica usando Medidas de Reflectância Difusa no Infravermelho Próximo**

Autor: Marcus Henrique Ferreira

Orientador: Prof. Dr. Marcelo Martins de Sena

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

No presente trabalho se desenvolveu e validou um novo método analítico usando medidas de reflectância difusa no infravermelho próximo e calibração multivariada para a determinação direta do fármaco hidroclorotiazida em amostras de pó para produção de comprimidos. O melhor modelo por mínimos quadrados parciais (PLS) foi obtido na região espectral de 1640 a 1780 nm, com dados centrados na média e pré-processados usando primeira derivada e alisamento Savitzky-Golay, seguidos por normalização vetorial. Este modelo foi construído com 4 variáveis latentes, forneceu um erro quadrático médio de previsão (RMSEP) de 1,7% e erros relativos individuais menores em módulo que  $\pm 3\%$ . O método foi validado, de maneira a conciliar regulamentações específicas (ANVISA) e gerais (ICH, ASTM), na faixa 21,25 a 29,00 mg de hidroclorotiazida por 150 mg de pó, que equivale à massa média de um comprimido. A validação analítica foi realizada através da estimativa de figuras de mérito, tais como exatidão, precisão, linearidade, sensibilidade analítica, capacidade de detecção, teste de erro sistemático (*bias*) e relação de desempenho do desvio (RPD). O conceito de sinal do analito líquido (NAS) foi usado para estimar algumas figuras de mérito e para traçar uma curva de calibração pseudo-univariada. Os resultados para as determinações em pó para produção de comprimidos se mostraram de acordo com os resultados obtidos por cromatografia líquida de alta eficiência (CLAE). Finalmente, o método foi extrapolado para determinações de comprimidos intactos, proporcionando erros de previsão menores que  $\pm 9\%$ . O método desenvolvido apresentou a vantagem de ser quinze vezes mais rápido que o método CLAE de referência.

## ABSTRACT

In the present work it was developed and validated a new multivariate diffuse reflectance near infrared method for direct determination of hydrochlorothiazide in powder pharmaceutical samples (tablets). The best partial least squares (PLS) model was obtained in the spectral region from 1640 to 1780 nm, with mean centered data preprocessed by first derivative and Savitzky-Golay smoothing followed by vector normalization. This model was built with 4 latent variables and provided a root mean square error of prediction (RMSEP) of 1.7 %. The method was validated according to the appropriate regulations in the range from 21.25 to 29.00 mg of hydrochlorothiazide per 150 mg of powder (average mass tablet), by the estimate of figures of merit, such as accuracy, precision, linearity, analytical sensitivity, capability of detection, bias and residual prediction deviation (RPD). The concept of net analyte signal (NAS) was used to estimate some figures of merit and to plot a pseudo-univariate calibration curve. The results for determinations in powdered manufactured tablets were in agreement with those of the official high performance liquid chromatographic method (HPLC). Finally, the method was extrapolated for determinations in intact tablets, providing prediction errors smaller than  $\pm 9\%$ . The developed method presented the advantage of being about fifteen times faster than the reference HPLC method.

# SUMÁRIO

1. Introdução .....	1
2. Espectroscopia na região do infravermelho .....	6
2.1. Infravermelho .....	6
2.2. Aspectos teóricos .....	7
2.3. Modelo mecânico para uma molécula diatômica .....	9
2.4. Vibrações moleculares .....	11
2.5. Reflectância Difusa e Infravermelho Próximo .....	13
3. Métodos Quimiométricos .....	15
3.1. Calibração Multivariada .....	16
3.2. Disposição dos dados .....	17
3.3. Regressão linear múltipla (MLR, do inglês <i>Multiple Linear Regression</i> ) .....	18
3.4. Regressão por componentes principais (PCR, do inglês <i>Principal Component Regression</i> ) .....	19
3.5. Mínimos Quadrados Parciais (PLS, do inglês <i>Partial Least Squares</i> ) ..	20
3.6. Validação Analítica de Métodos Quimiométricos .....	22
3.7. Sinal analítico líquido - NAS .....	23
3.8. Figuras de Mérito .....	26
3.9. Detecção de amostras anômalas ( <i>outliers</i> ) .....	30
4. Parte experimental .....	31
4.1. Instrumentação e processamento dos dados .....	31
4.2. Reagentes e amostras .....	32
4.3. Metodologia .....	33
4.3.1. Planejamento experimental .....	33
4.3.2. Preparação das amostras .....	36
4.3.3. Obtenção dos dados .....	37
4.3.4. Análise de amostras reais .....	37

4.3.5. Análise cromatográfica.....	38
4.4. Validação analítica.....	38
5. Resultados e discussão.....	39
5.1. Espectro de HCTZ pura .....	39
5.2. Controle de pesos e teor de umidade .....	40
5.3. Modelo PLS.....	41
5.4. Detecção de amostras anômalas.....	43
5.5. Validação Analítica.....	46
5.6. Análises de Amostras Reais .....	51
6. Conclusões .....	55
7. Referências Bibliográficas .....	56

## LISTA DE TABELAS

<b>Tabela 1.</b> Composição das amostras do conjunto de calibração .....	34
<b>Tabela 2.</b> Composição das amostras do conjunto de validação .....	36
<b>Tabela 3.</b> Controle de massa final .....	40
<b>Tabela 4.</b> Controle de do teor de umidade (TU).....	40
<b>Tabela 5.</b> Otimização do modelo PLS pela detecção de amostras anômalas .....	44
<b>Tabela 6.</b> Parâmetros estimados para avaliar as principais figuras de mérito .....	48
<b>Tabela 7.</b> Resultados de previsão das amostras de validação, após a exclusão das amostras anômalas. ....	48
<b>Tabela 8.</b> Amostras lidas em diferentes recipientes.....	51
<b>Tabela 9.</b> Determinação de HCTZ em comprimidos triturados a pó fino pelo método proposto e pelo método oficial CLAE .....	52
<b>Tabela 10.</b> Determinação de HCTZ em dez comprimidos intactos pelo método desenvolvido e pelo método oficial CLAE .....	53

## LISTA DE FIGURAS

<b>Figura 1.</b> Fórmula estrutural da Hidroclorotiazida. ....	2
<b>Figura 2.</b> (a) Faixas espectrais e subdivisões da região do infravermelho e (b) tipos de transições. ....	8
<b>Figura 3.</b> Representação esquemática do comportamento harmônico (A) e anarmônico (B) para o modelo mecânico clássico de vibração [12].....	10
<b>Figura 4.</b> Tipos de vibração molecular.....	13
<b>Figura 5.</b> Reflectância difusa. Nota: “ $R_p$ ” é a reflexão do padrão e “ $R_a$ ” reflexão da amostra, como mostrado na equação 7. ....	14
<b>Figura 6.</b> Organização dos dados usados em calibração multivariada:.....	17
<b>Figura 7.</b> Formação de uma VL por rotação no eixo da componente principal. ....	22
<b>Figura 8.</b> Representação geométrica da propriedade de ortogonalidade do NAS.....	23
<b>Figura 9.</b> Representação esquemática da estrutura de um espectrofotômetro da marca Foss Modelo 4500. ....	31
<b>Figura 10.</b> Foto da parte interna de um espectrofotômetro da marca Foss modelo 4500.....	32
<b>Figura 11.</b> Planejamento experimental utilizado no trabalho, no qual a escala de cada componente é apresentada em termos de pseudocomponentes. ....	34
<b>Figura 12.</b> Espectro NIR obtido por reflectância difusa de HCTZ pura. A região espectral usada no melhor modelo PLS está destacada com um retângulo. ....	39
<b>Figura 13.</b> Espectros das 93 amostras, correspondentes aos conjuntos de calibração e validação. A região espectral usada no melhor modelo PLS está destacada por um retângulo.....	42
<b>Figura 14.</b> Espectros pré-processados com primeira derivada, alisamento e normalização vetorial, na faixa de 1640 a 1780nm. ....	43
<b>Figura 15.</b> Detecção de amostras anômalas no conjunto de calibração para o primeiro modelo. (a) Histogramas dos valores de influência ( <i>leverage</i> ). (b) Gráfico dos resíduos espectrais (X) <i>versus</i> resíduos de concentração (Y). As linhas contínuas indicam os limites aceitáveis para a detecção de amostras anômalas.....	44
<b>Figura 16.</b> Detecção de amostras anômalas no conjunto de calibração para o segundo modelo. (a) Histogramas dos valores de influência ( <i>leverage</i> ). (b)	

Gráfico dos resíduos espectrais (X) versus resíduos de concentração (Y). As linhas contínuas indicam os limites aceitáveis para a detecção de amostras anômalas.....	45
<b>Figura 17.</b> Detecção de amostras anômalas no conjunto de validação para o terceiro modelo. (a) Histogramas dos valores de influência ( <i>leverage</i> ). (b) Gráfico dos resíduos espectrais (X) versus resíduos de concentração (Y). As linhas contínuas indicam os limites aceitáveis para a detecção de amostras anômalas.....	45
<b>Figura 18.</b> Resíduos do modelo PLS para as amostras de calibração (círculos) e validação (asteriscos).....	50
<b>Figura 19.</b> Curva de calibração pseudo-univariada. Norma do NAS versus valores de referência das concentrações para as amostras de calibração (círculos) e validação (asteriscos).....	50
<b>Figura 20.</b> Cromatograma obtido para uma amostra de HCTZ pelo método oficial (CLAE). .....	52
<b>Figura 21.</b> Resultados (médios e desvios padrão de triplicatas) de determinações de HCTZ em pó de seis amostras coletados após mistura e antes da compressão. As linhas pontilhadas indicam os limites para liberação durante o controle em processo. Resultados da análise por CLAE (x) e por NIR (♦).....	54

## LISTA DE ABREVIATURAS

- ANVISA** - Agência Nacional de Vigilância Sanitária
- ASTM** - *American Society for Testing and Materials*
- CC $\beta$**  - Capacidade de detecção
- CLAE** - Cromatografia Líquida de Alta Eficiência
- DP** - Desvio padrão
- DPR** - Desvio padrão relativo
- DPV** - Desvio padrão de erros de validação
- Drifts** - *Diffuse Reflectance Infrared with Fourier Transform Spectroscopy*
- HCTZ** - Hidroclorotiazida
- IQUEGO** - Indústria Química do Estado de Goiás
- IUPAC** - *International Union of Pure and Applied Chemistry*
- IV** - Infravermelho
- LD** - Limite de detecção
- LQ** - Limite de quantificação
- MLR** - *Multiple Linear Regression (Regressão Linear Múltipla)*
- MSC** - *Multiplicative Scatter Correction (Correção do Espalhamento Multiplicativo)*
- NAS** - *Net Analyte Signal (Sinal Analítico Líquido)*
- NIR** - *Near Infrared (Infravermelho Próximo)*
- NIRS** – *Near Infrared spectroscopy*
- PCA** - *Principal Component Analysis (Análise de Componentes Principais)*
- PCR** - *Principal Component Regression (Regressão em Componentes Principais)*
- PLS** - *Partial Least Squares (Mínimos Quadrados Parciais)*
- RMSEC** - *Root Mean Squares Errors of Calibration*
- RMSECV** - *Root Mean Squares Errors of Cross Validation*
- RMSEP** - *Root Mean Squares Errors of Prediction*
- RPD** - *Residual Prediction Deviation ou Relative Predictive Determinant*
- SEN** - Sensibilidade
- SNV** - *Standard Normal Variate*
- TF** - Transformada de Fourier
- UV** - Ultra Violeta
- VIS** - Visível
- VL** - Variável Latente



## 1. Introdução

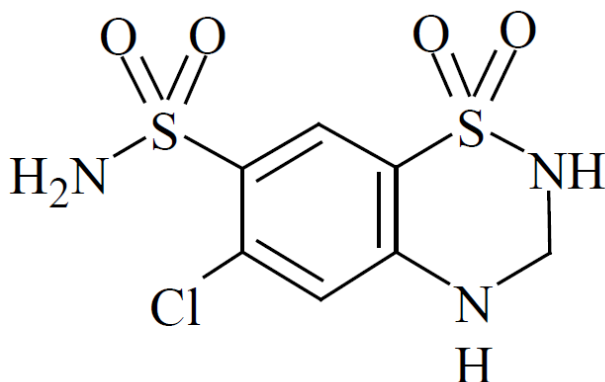
A hipertensão arterial sistêmica é o agravo mais comum na população adulta em todo o mundo e um fator de risco muito importante para as doenças cardiovasculares, dentre as quais se destaca o acidente vascular cerebral (AVC). A hipertensão é definida como pressão arterial sistólica superior a 140 mmHg e diastólica superior a 90 mmHg em pelo menos duas tomadas em momentos distintos. O uso clínico de medicamentos para o controle dos níveis pressóricos na hipertensão arterial surgiu por volta de 1947, portanto há pouco mais de 60 anos, o que torna essa prática bem recente do ponto de vista histórico [1].

Os diuréticos têm sido as drogas mais usadas no tratamento da hipertensão por mais de 30 anos e apresentam um papel vital no controle dessa condição [1]. Por definição, os diuréticos são fármacos que aumentam a taxa de fluxo de urina. Entretanto, os diuréticos clinicamente úteis também aumentam a taxa de excreção de  $\text{Na}^+$  (natriurese) e de um ânion associado, habitualmente  $\text{Cl}^-$ . O  $\text{NaCl}$  no organismo é o principal determinante do volume de líquido extracelular, e a maioria das aplicações clínicas dos diuréticos visa à redução do volume de líquido extracelular ao diminuir o conteúdo corporal total de  $\text{NaCl}$  [2].

Os benzotiazídicos, uma classe de diuréticos inibidores do simporte (transporte através da membrana de duas substâncias no mesmo sentido) de  $\text{Na}^+\text{Cl}^-$  foram sintetizados com o propósito de aumentar a potência dos inibidores da anidrase carbônica (outra classe de diuréticos). Entretanto, ao contrário dos inibidores da anidrase carbônica, que aumentam primariamente a excreção de  $\text{NaHCO}_3$ , foi constatado que os benzotiazídicos aumentam predominantemente a excreção de  $\text{NaCl}$ , um efeito independente da inibição da anidrase carbônica. A expressão *diuréticos tiazídicos* é frequentemente utilizada em referência a todos os membros da classe de inibidores do simporte de  $\text{Na}^+\text{Cl}^-$  [2].

A hidroclorotiazida (HCTZ), denominada quimicamente 1,1-dióxido de 6-cloro-3,4-diidro-2H-1,2,4-benzotiazina-7-sulfonamida e cuja estrutura é mostrada na Figura 1, é um diurético tiazídico, amplamente utilizado no tratamento de edemas, doenças cardíacas congestivas e diversas formas de disfunção renal e hepática. Geralmente, ela é administrada em conjunto com anti-hipertensivos, resultando em uma redução efetiva da pressão por mecanismos associados, como a redução da resistência periférica [3]. Como outros diuréticos tiazídicos, a HCTZ reduz a atividade da anidrase carbônica, aumentando a excreção de bicarbonato. Contudo,

este efeito é geralmente de pequena intensidade, em comparação com seu efeito sobre a excreção de cloreto, e não altera consideravelmente o equilíbrio ácido-base, nem o pH urinário. A HCTZ possui efeito anti-hipertensivo por ação direta sobre os vasos sanguíneos [1].



**Figura 1.** Fórmula estrutural da Hidroclorotiazida.

A HCTZ se apresenta na forma de pó cristalino branco ou quase branco, inodoro, muito pouco solúvel em água, pouco solúvel em etanol, solúvel em acetona e em soluções diluídas de hidróxidos alcalinos [4].

Os métodos oficiais para determinação de HCTZ em formulações farmacêuticas são baseados em CLAE e UV [4,5], mas muitos outros métodos têm sido publicados, com base em espectrofotometria de absorção no UV/visível [6,7], quimioluminescência [8], eletroforese capilar [9] e técnicas eletroanalíticas [10,11]. No entanto, a maioria desses métodos não têm sido capazes de determinar diretamente HCTZ na presença de interferentes, tais como excipientes, impurezas e outros princípios ativos, exigindo etapas de separação e uso de reagentes ou solventes. O método proposto neste trabalho é direto e não utiliza solventes ou reagentes.

A Farmacopeia Brasileira, V edição [4], preconiza para quantificação de HCTZ metodologias baseadas em espectrofotometria de absorção no ultravioleta e em cromatografia líquida de alta eficiência (CLAE), usando como fase móvel uma mistura de água, metanol e acetonitrila. Em ambos os métodos é necessário que a amostra seja dissolvida em solventes com o uso de ultrassom ou agitadores

mecânicos no intuito de fazer a extração e/ou separação físico-química dos componentes e só após estas etapas, que demandam cerca de vinte a trinta minutos por amostra, devido à necessidade de se dissolver o analito, a amostra é submetida à análise instrumental para se obter a resposta relacionada à concentração da HCTZ.

A espectroscopia de absorção no infravermelho próximo é uma técnica analítica de importância reconhecida, principalmente pelo custo relativamente baixo da instrumentação, rapidez nas análises e sensibilidade, que satisfazem muitas exigências analíticas. Além disso, ela permite a realização de medidas não destrutivas, o que é de grande vantagem para o controle de qualidade na indústria farmacêutica [12].

A partir da década passada, o emprego conjunto de técnicas de espectroscopia molecular e de métodos quimiométricos de calibração multivariada passou a representar uma alternativa para a determinação direta de fármacos [13]. Em situações onde a determinação de um analito é dificultada ou impedida devido à presença de um ou mais interferentes, ao invés da eliminação dessas espécies (separação cromatográfica, extração, etc.), o uso de calibração multivariada permite a modelagem desses interferentes junto com o analito de interesse. A ideia central consiste em substituir a separação física das espécies pela separação quimiométrica de seus sinais. As vantagens potenciais da determinação direta são maior rapidez, menor custo e simplificação ao máximo das etapas de tratamento da amostra, diminuindo as possibilidades de erro causadas pela sua manipulação. Nos últimos anos, esta estratégia vem sendo aplicada em análises clínicas e no controle de qualidade de produtos farmacêuticos. Em análises clínicas, tornou-se possível a determinação de fármacos em amostras complexas, tais como, plasma, soro e urina, sem a necessidade de separação prévia dos constituintes [14]. Em aplicações voltadas para o controle de qualidade de produtos farmacêuticos, tornaram-se possíveis a determinação simultânea direta de vários princípios ativos em um mesmo produto farmacêutico usando espectrofotometria no ultravioleta (UV) [15] e a determinação não destrutiva de fármacos na presença de interferentes através da técnica de refletância difusa no infravermelho [16].

O infravermelho próximo (NIR, do inglês *Near Infrared*) corresponde à região espectral de comprimento de onda entre 780 e 2500 nm (números de onda de 12800 a 4000  $\text{cm}^{-1}$ ), na qual os sinais observados consistem principalmente de sobretons e bandas de combinação, provenientes de vibrações moleculares.

Diferentemente da espectroscopia no infravermelho médio, os espectros de absorção no NIR são pouco úteis para identificação e atribuição de bandas de absorção a grupos específicos da molécula (análise qualitativa), pois apresentam bandas fracas e sobrepostas. Porém, quando usada em conjunto com métodos quimiométricos, a espectroscopia NIR é uma ferramenta útil para análises quantitativas de compostos contendo grupos funcionais constituídos de hidrogênio ligado a carbono, nitrogênio e oxigênio. Tais compostos podem frequentemente ser determinados com exatidão e precisão equivalentes à espectroscopia UV/Vis [17]. A complexa natureza da região NIR raramente permite o uso de um único comprimento de onda para construir um modelo de calibração. Dessa forma, não foi por acaso que a difusão do uso de metodologias baseadas em espectroscopia NIR ganhou um grande impulso paralelamente ao desenvolvimento de métodos quimiométricos, os quais usam matemática e estatística para extrair informação de dados químicos multivariados. Se no começo dos anos oitenta, ela ainda era considerada uma “adormecida entre as técnicas espectrais” [18], atualmente, pode-se afirmar que a espectroscopia NIR é a técnica analítica que mais faz uso de métodos quimiométricos [12].

Métodos baseados em NIRS (do inglês *Near-infrared spectroscopy*) são rápidos, não destrutivos e de larga aplicação, permitindo o registro direto de espectros para sólidos com pequeno ou nenhum pré-tratamento da amostra. Estas vantagens tornaram a NIRS uma opção atraente para a determinação de componentes farmacêuticos ativos nos últimos anos [16,19-22].

Os componentes básicos dos equipamentos para espectroscopia NIR, a princípio, não diferem muito dos equipamentos para medidas no UV/Vis. No entanto, notórias diferenças são encontradas, principalmente na fase de seleção de comprimento de onda, dando origem a equipamentos dispersivos ou interferométricos. Os equipamentos interferométricos baseados na Transformada de Fourier (TF) são os mais usados atualmente e possuem as vantagens de apresentar maiores resolução, reprodutibilidade e razão sinal/ruído [17]. O uso de equipamentos baseados na TF combinado com reflectância difusa dá origem à técnica DRIFTS (do inglês, *Diffuse Reflectance Infrared with Fourier Transform Spectroscopy*) [12]. A espectroscopia de reflexão no infravermelho (IV) apresenta especial interesse para amostras sólidas. Os espectros de reflexão, embora não idênticos aos espectros de absorção correspondentes, são semelhantes na aparência geral e fornecem a mesma informação. A maioria dos fabricantes de

instrumentos oferece acessórios facilmente adaptáveis para obtenção de medidas de reflexão. Dentre os tipos de reflexão, a que encontra maior aplicação na região do NIR é a reflectância difusa. Esta ocorre quando um feixe de radiação atinge a superfície irregular de um pó finamente dividido e, após penetrar alguns milímetros, é refletido em ângulos diferentes do ângulo de incidência. Estritamente, não se trata de um fenômeno de superfície, pois ocorre absorção pela interação entre a amostra e a radiação IV incidente. A DRIFTS possui as vantagens de ser uma técnica não destrutiva, relativamente rápida e, principalmente, requerer um mínimo de preparo de amostra. Contudo, algumas precauções devem ser consideradas, tais como manter o tamanho das partículas e o grau de compactação uniforme, e a superfície das amostras plana. A espectroscopia na região do NIR, juntamente com métodos quimiométricos de calibração multivariada, tem sido usada para o controle de qualidade de fármacos, tais como dipirona sódica [19], genfibrozil [20], carbamazepina [16], nistatina e metronidazol [21], paracetamol [22] e amoxicilina [23].

Outro aspecto importante a ser considerado neste tipo de aplicação é a validação analítica. Os métodos analíticos baseados em calibração multivariada são relativamente recentes e a sua aceitação e implementação é ainda restrita, principalmente em áreas como a farmacêutica e a médica, devido às exigências de validação pelas agências e organismos reguladores [24-26]. Essa validação pode ser realizada pela determinação de parâmetros que são conhecidos como figuras de mérito, tais como exatidão, precisão, sensibilidade e limite de detecção. No entanto, até bem pouco tempo atrás, a maioria dos artigos de calibração multivariada não abordava a questão da validação analítica, que no caso multivariado não é tão simples como na calibração univariada. Na estimativa de figuras de mérito em calibração multivariada [16,27], a parte do sinal que se relaciona unicamente ao analito de interesse é mais importante que o sinal total e é denominada sinal analítico líquido (NAS, do inglês *Net Analyte Signal*). O NAS é definido como a parte do sinal analítico que é ortogonal ao sinal dos interferentes presentes na amostra [28] e pode ser usado na estimativa das figuras de mérito: sensibilidade, seletividade e limites de detecção e quantificação.

Diante da necessidade de se realizar rigorosos testes de controle de qualidade na indústria farmacêutica, atendendo às normas e exigências legais, e ao mesmo tempo reduzir custos com mão de obra e insumos, a técnica NIRS se apresenta como alternativa em potencial que, embora inicialmente demande mão

de obra especializada e uma quantidade de padrões relativamente alta, propicia um retorno bastante satisfatório em termos de custos quando aplicada à rotina.

O objetivo do presente trabalho foi desenvolver e validar um método alternativo para determinar HCTZ, podendo ser aplicado no controle em processo durante a etapa de mistura conforme o procedimento farmacotécnico da IQUEGO, evitando assim a realização de análises por métodos dispendiosos e demorados.

## 2. Espectroscopia na região do infravermelho

### 2.1. Infravermelho

Quase todos os compostos que tenham ligações covalentes, sejam orgânicos ou inorgânicos, absorvem radiação em várias frequências na região do infravermelho do espectro eletromagnético. Essa região envolve comprimentos de onda maiores do que aqueles associados à luz visível, que vão de aproximadamente de 400 a 800 nm, mas menores que aqueles associados a microondas, que são maiores que 1 mm. Assim como ocorre com outros tipos de absorção de energia, as moléculas, quando absorvem radiação no infravermelho, são excitadas para atingir um estado de maior energia [29].

A espectrometria infravermelha moderna é uma ferramenta versátil, que é aplicada à determinação qualitativa e quantitativa de espécies moleculares de todos os tipos. A região mais amplamente utilizada é o infravermelho médio, que vai de 670 a 4000  $\text{cm}^{-1}$  [17]. Essa faixa de comprimento de onda no infravermelho é amplamente usada na indústria farmacêutica para identificação de matérias primas, seja princípio ativo ou excipiente, em atendimento às exigências legais [4]. A região do infravermelho próximo, que é objeto de estudo deste trabalho, vai de 4.000 a 14.000  $\text{cm}^{-1}$  (780 nm a 2500 nm), e também encontra uso considerável para determinações quantitativas de rotina [17]. O infravermelho próximo vem ganhando mercado em relação ao infravermelho médio nas indústrias farmacêuticas devido ao menor custo do equipamento, sua robustez e pouca sensibilidade à umidade no ar, além da maior rapidez na obtenção de espectros e a facilidade do tratamento quimiométrico via *software* para diferenciar espécies químicas sem tratamento prévio da amostra, levando assim a uma identificação rápida e segura por meio de comparação do espectro da amostra e um padrão contido em banco de dados, o

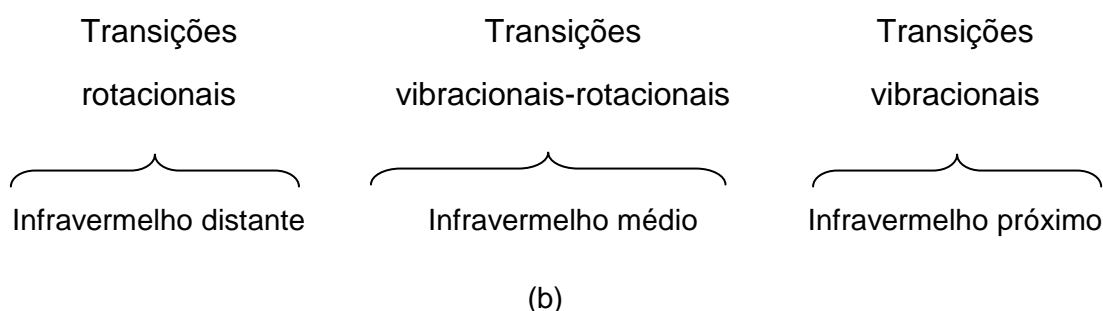
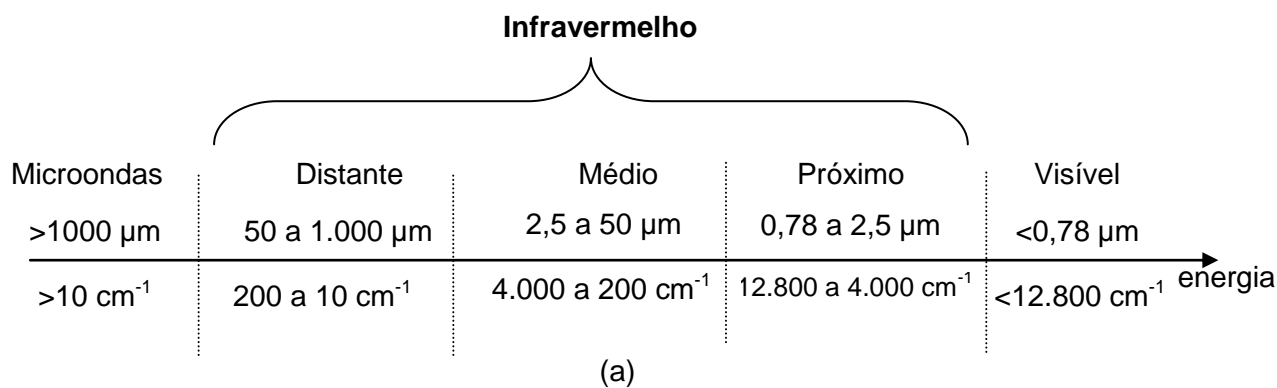
que é possível e já aplicado na rotina de indústrias farmacêuticas devido ao alto grau de pureza de suas matérias-primas, pureza essa que é garantida por análises específicas.

As bandas de absorção na região do infravermelho próximo são harmônicas (sobretons) ou combinações de bandas de estiramento fundamentais. As ligações envolvidas são normalmente C-H, O-H, N-H e S-H [17].

## **2.2. Aspectos teóricos**

A radiação infravermelha não é energética o suficiente para causar transições eletrônicas, como no ultravioleta e visível. A absorção deste tipo de radiação está muito restrita a espécies moleculares que têm diferenças de energia pequenas entre seus vários estados vibracionais ou rotacionais. Para absorver radiação infravermelha, uma molécula precisa sofrer uma variação no momento de dipolo como consequência do movimento vibracional ou rotacional. Apenas nessas circunstâncias o campo elétrico alternado da radiação pode interagir com a molécula e causar variações na amplitude de um de seus movimentos. A energia necessária para causar uma mudança de nível rotacional é diminuta e, como os níveis rotacionais são quantizados, a absorção por gases na região do infravermelho distante (50 a 1000  $\mu\text{m}$ ) é caracterizada por linhas discretas e bem definidas. Por outro lado, a rotação está rigorosamente restrita em líquidos e sólidos, nos quais as linhas discretas vibracionais/rotacionais desaparecem, deixando picos vibracionais alargados [17].

A Figura 2 mostra as faixas de radiação eletromagnéticas, entre a região visível e microondas e as subdivisões na região do infravermelho, com os tipos de transições em cada faixa desta região do espectro eletromagnético.



**Figura 2.** (a) Faixas espectrais e subdivisões da região do infravermelho e (b) tipos de transições.

A equação 1 mostra a relação para conversão das principais unidades usadas para expressar comprimento de onda ( $\lambda$ ) na região do infravermelho.

$$\text{Número de onda}(cm^{-1}) = \frac{10^4}{\lambda(\mu m)} = \frac{10^7}{\lambda(nm)} \quad (1)$$

Em temperatura ambiente, a maioria das moléculas está no estado fundamental de energia vibracional. Átomos ou grupos de átomos que participam de ligações são deslocados uns em relação aos outros em frequências definidas pela força da ligação e da massa individual de cada átomo ou grupos de átomos. A amplitude dessas vibrações é da ordem de alguns nanômetros e aumenta se alguma energia é transferida à molécula [12]. As vibrações podem ser classificadas como de estiramento e de deformação angular. Uma vibração de estiramento envolve uma variação contínua na distância interatômica ao longo do eixo de ligação entre dois átomos. As vibrações de deformação angular são caracterizadas pela variação do



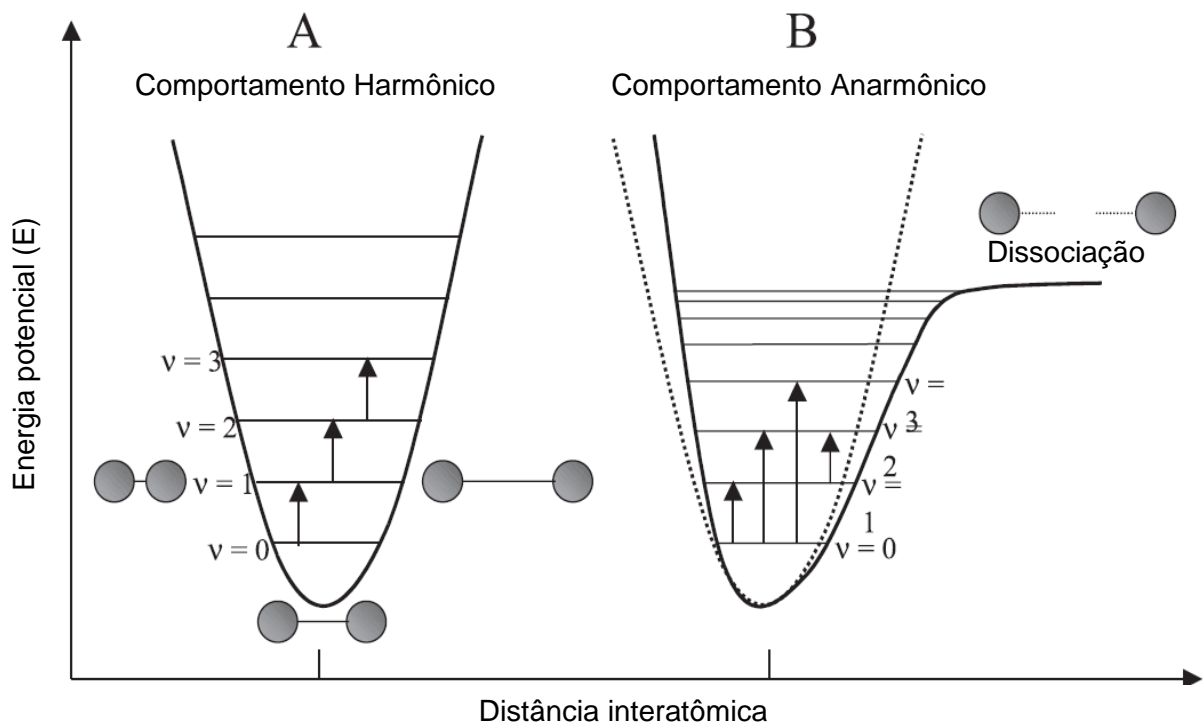
ângulo entre duas ligações. Além disso, interações ou acoplamento de vibrações podem ocorrer se as vibrações envolvem ligações de um mesmo átomo central, provocando uma variação das suas características [17].

### 2.3. Modelo mecânico para uma molécula diatômica

Pode-se fazer uma comparação, aproximada, das características de uma vibração de estiramento com um modelo mecânico constituído por duas massas ligadas por uma mola. Um deslocamento em uma das massas ao longo do eixo da mola provoca uma vibração denominada movimento harmônico simples. Dessa forma, considerando dois átomos como sendo massas esféricas ligadas por uma mola de constante de força  $k$ , de acordo com a lei de Hooke a força de restauração é proporcional ao deslocamento “ $x$ ” e à energia potencial  $E$ , que é dada pela equação 2.

$$E = \frac{1}{2} kx^2 \quad (2)$$

A curva de energia potencial deduzida a partir da equação 2 é uma parábola como mostrada numa representação esquemática na Figura 3A. Observe que a energia potencial é máxima quando os átomos estão muito distantes ou muito próximos um do outro, e decresce para zero na posição de equilíbrio, representativa do estado fundamental.



**Figura 3.** Representação esquemática do comportamento harmônico (A) e anarmônico (B) para o modelo mecânico clássico de vibração [12].

Segundo o modelo do oscilador harmônico, a energia vibracional de um determinado sistema é uma função quadrática da distância entre os átomos e a oscilação entre eles é simétrica. Esse modelo considera que a ligação entre os átomos é feita por uma mola, cuja frequência de vibração ( $\nu_m$ ) é descrita pela lei de Hooke e representada pela equação 3 [12].

$$\nu_m = \frac{1}{2\pi} \sqrt{\frac{k}{\mu}} \quad (3)$$

Onde “ $k$ ” representa a constante de força e  $\mu$  é a massa reduzida:

$$\mu = \frac{m_1 m_2}{m_1 + m_2} \quad (4)$$

Esta aproximação é útil para se entender o conceito de energia vibratória, mas não descreve completamente o comportamento de partículas de dimensões

atômicas. A falha surge pelo fato de que sistemas moleculares e atômicos não podem assumir qualquer valor de energia de maneira contínua, como duas massas presas por uma mola. Um sistema molecular pode assumir níveis de energia discretos definidos pela mecânica quântica, de acordo com a equação 5, onde  $h$  é a constante de Plank e  $\nu_m$  é o número quântico vibracional definido na equação 3.

$$E = \left( \nu + \frac{1}{2} \right) h \nu_m \quad (5)$$

Outra restrição do modelo clássico é que as transições ocorrem apenas entre níveis adjacentes com  $\Delta\nu = \pm 1$ , já que as distâncias entre os níveis são iguais.

A Figura 3B mostra o comportamento de um oscilador anarmônico, uma representação mais realística para uma molécula diatômica. Pode-se observar que à medida que a energia aumenta, ocorre uma diminuição na distância entre níveis energéticos adjacentes; note também que, devido à repulsão coulombiana, conforme os átomos vão se aproximando a energia tende a se elevar mais rapidamente que no modelo harmônico. No outro extremo da oscilação, um decréscimo na força de restauração e, portanto, da energia potencial, ocorre quando a distância interatômica se aproxima daquela em que ocorre a dissociação dos átomos [12,17].

## 2.4. Vibrações moleculares

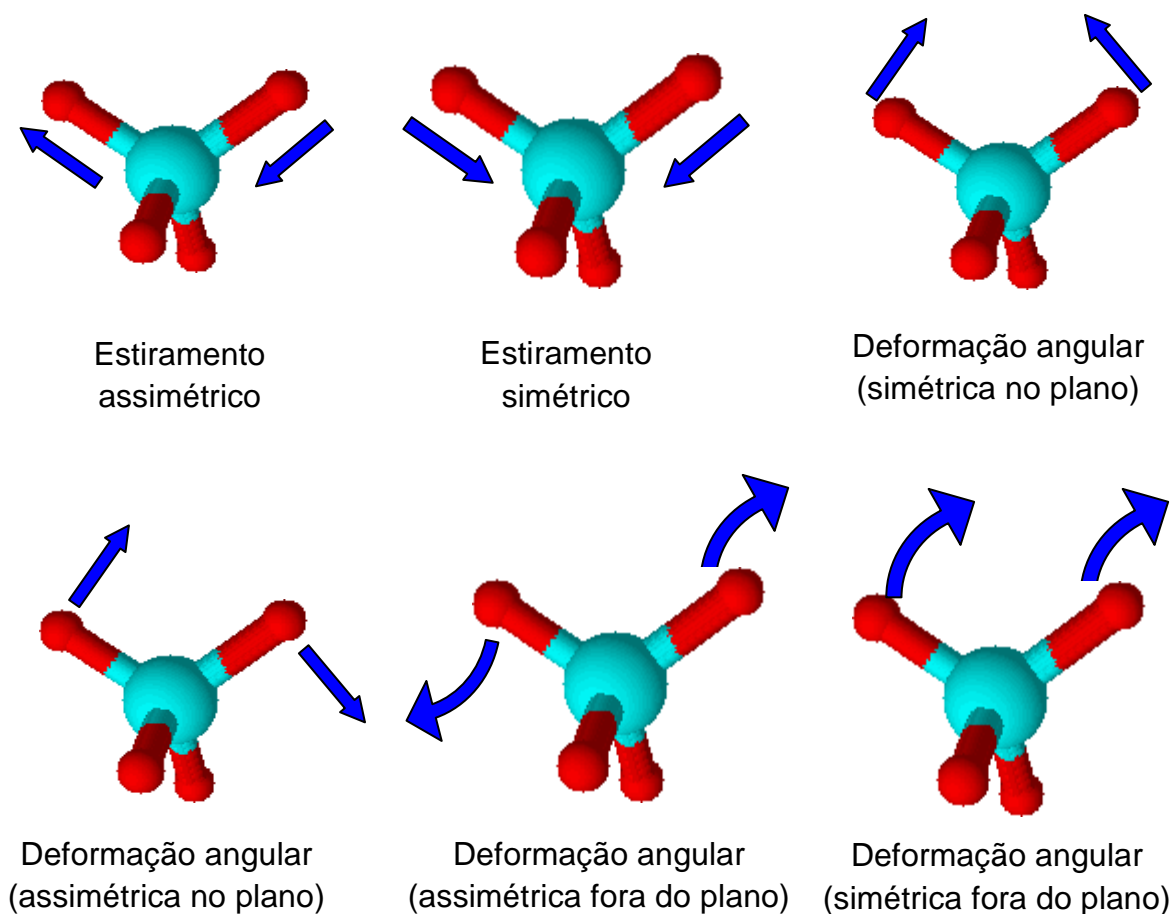
Cada ligação distinta em uma molécula absorve em um comprimento de onda específico do espectro. Mesmo ligações iguais em ambientes químicos levemente diferentes, com átomos “vizinhos” diferentes, irão apresentar diferentes padrões de absorção. Em uma molécula os átomos estão sempre em movimento nas três dimensões do espaço, não só com movimento de translação em que a molécula muda as coordenadas de todos os seus átomos no espaço, mas também movimentos rotacionais e vibracionais. Se, em um sistema, há “N” átomos livres para se movimentarem nas três dimensões, o sistema terá “3N” graus de liberdade. Se, no entanto, esses átomos estiverem ligados entre si, formando uma molécula, continuarão ainda existindo 3N graus de liberdade, sendo três graus para a translação do centro de massa da molécula e, para uma molécula não linear, três graus para a rotação da mesma em torno dos três eixos, restando, assim, 3N-6

graus de liberdade para as vibrações. Para moléculas lineares, como não há rotação em torno do eixo internuclear, restam  $3N-5$  graus de liberdade para as vibrações. Esses graus de liberdade correspondem aos diferentes modos normais de vibração de uma molécula. Um modo normal de vibração é aquele em que cada núcleo realiza uma oscilação harmônica simples em torno de sua posição de equilíbrio, todos os núcleos se movem com a mesma frequência e em fase, e o centro de gravidade da molécula permanece inalterado [30].

Para cada modo normal de vibração, existe uma energia potencial relacionada. Assim, cada modo normal de vibração deveria apresentar uma banda de absorção específica, porém nem sempre isso acontece. Vários são os fatores que podem interferir e tornar diferentes o número de picos obtidos experimentalmente e o número previsto pela teoria. Por exemplo, um número menor de picos é encontrado quando a simetria da molécula impede uma variação de dipolo, quando as energias de duas ou mais vibrações são iguais ou muito parecidas, sobrepondo os picos. Ocorre um número menor de picos também quando a intensidade de absorção é muito baixa e não pode ser detectada pelo equipamento ou a energia vibracional está em uma região de comprimento de onda além da faixa do instrumento. Algumas vezes são encontrados mais picos que o previsto com base no número de modos normais de vibração. Isso ocorre, por exemplo, quando se observa bandas em frequências de valores que são múltiplos ( $2\nu$ ,  $3\nu$ , etc.) de valores de frequências normais de vibração ( $\nu$ ), formando assim os chamados sobretons. Também é observada maior quantidade de bandas quando ocorrem combinações lineares de frequências normais de vibração. Por exemplo, sendo  $\nu_1$  e  $\nu_2$  bandas normais de vibração que, ao se combinarem, geram as bandas de combinação em valores de frequência de aproximadamente  $(\nu_1+\nu_2)$ ,  $(\nu_1-\nu_2)$ ,  $(2\nu_1+\nu_2)$ , etc. Em todos esses casos, as bandas se apresentam com intensidades menores que as bandas dos modos normais de vibração, já que elas ocorrem devido à excitação de dois ou mais modos vibracionais pelo mesmo fóton [12,17].

As vibrações moleculares podem ser divididas nas categorias de deformações axiais ou estiramentos e de deformações angulares. As vibrações de estiramento envolvem uma variação contínua na distância interatômica ao longo do eixo da ligação entre dois átomos. Nas vibrações de deformação angular ocorrem variações nos ângulos entre dois átomos ligados, que podem ser simétricas ou

assimétricas e ocorrer no plano ou fora do plano da ligação. A Figura 4 ilustra os diferentes tipos de vibrações moleculares.



**Figura 4.** Tipos de vibração molecular.

## 2.5. Reflectância Difusa e Infravermelho Próximo

A região do NIR foi a primeira faixa do espectro não-visível a ser descoberta, porém sua utilização como ferramenta analítica só ocorreu muito tempo depois. As primeiras aplicações analíticas apareceram na década de 50, com o aparecimento dos primeiros espectrofotômetros comerciais baseados em detectores fotoelétricos [31].

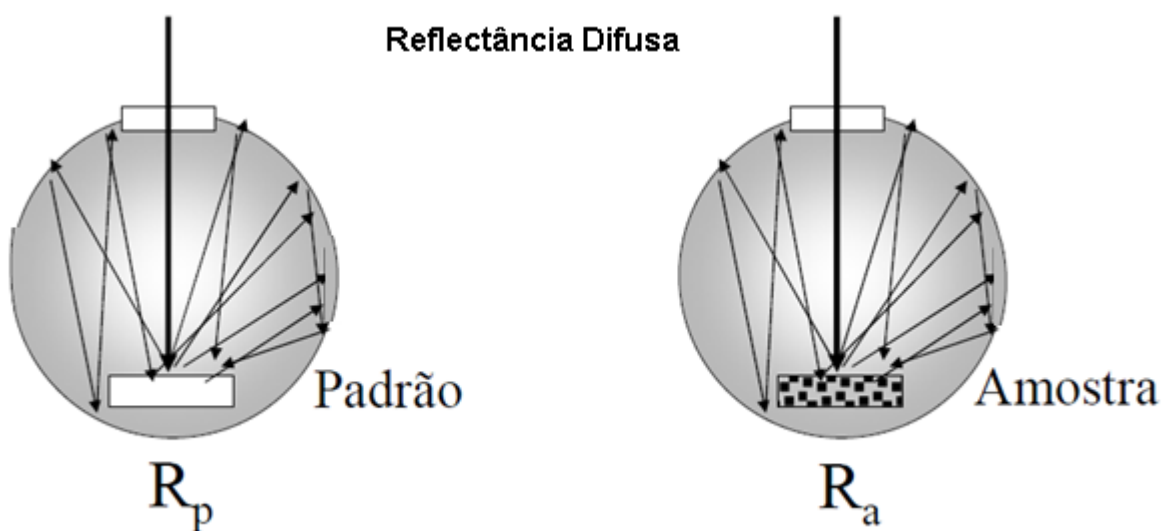
A espectrometria de reflexão difusa no infravermelho pode ser um modo efetivo de se obter espectros na região do infravermelho diretamente na amostra com o mínimo ou sem nenhum pré-tratamento, o que leva a uma grande redução no tempo de análise e minimiza os erros inerentes à intervenção humana.

A reflectância difusa é um processo complexo que ocorre quando um feixe de luz incide na superfície de uma amostra e é refletido aleatoriamente sem ângulo preferencial. Considera-se que ocorrem reflexões especulares em cada micro partícula que compõe a superfície da amostra ou do pó no caso de amostras pulverizadas. No entanto, como existem muitas destas partículas e elas estão orientadas aleatoriamente, a radiação é refletida difusamente em todas as direções. Apesar de ocorrer a reflexão, este não é um fenômeno de superfície, pois ocorre a interação entre a amostra e a radiação, e a luz não absorvida é refletida após penetrar alguns milímetros na amostra [17].

Vários modelos foram desenvolvidos para descrever em termos quantitativos a intensidade da radiação refletida difusamente. Um dos modelos mais usado foi o desenvolvido por Kubelka e Munk, através do qual foi demonstrado que a intensidade relativa de reflectância de um pó,  $f(R_d)$ , é dada pela equação 6.

$$R_d = \frac{(1 - R^2)}{2R} = \frac{k}{s} \quad (6)$$

Onde  $s$  é o coeficiente de espalhamento,  $k$  é o coeficiente de absorção molar do analito e  $R$  é a razão entre as intensidades refletidas pela amostra e por um padrão não-absorvente, por exemplo, o brometo de potássio, teflon, sulfato de bário, óxido de magnésio, placas cerâmicas de alumina de alta pureza. A reflexão difusa é ilustrada na Figura 5.



**Figura 5.** Reflectância difusa. Nota: “ $R_p$ ” é a reflexão do padrão e “ $R_a$ ” reflexão da amostra, como mostrado na equação 7.

$$R = \frac{R_a}{R_p} \quad (7)$$

No entanto, os espectros na região do NIR são normalmente obtidos como gráficos de  $\log(1/R)$  em função do comprimento de onda.

### 3. Métodos Quimiométricos

A química analítica quantitativa faz uso de inúmeras ferramentas e conceitos matemático-estatísticos, além da fundamentação no conhecimento das propriedades físicas e químicas de uma matriz complexa, com o objetivo de fornecer informações sobre a quantidade de um ou mais compostos contidos nesta matriz.

A quimiometria pode ser definida como a área da química que utiliza métodos matemáticos e estatísticos para definir ou selecionar as condições ótimas de medidas ou experimentos e extrair o máximo de informações possíveis, a partir da análise de dados químicos [32]. 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 [32].

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. 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 [33].

### **3.1. Calibração Multivariada**

A calibração é definida como o processo que permite estabelecer a relação entre a resposta instrumental (sinal analítico) e uma determinada propriedade da amostra (concentração do analito, por exemplo). A equação matemática que descreve esta relação é denominada modelo de calibração e a representação gráfica é denominada curva analítica ou de calibração [31].

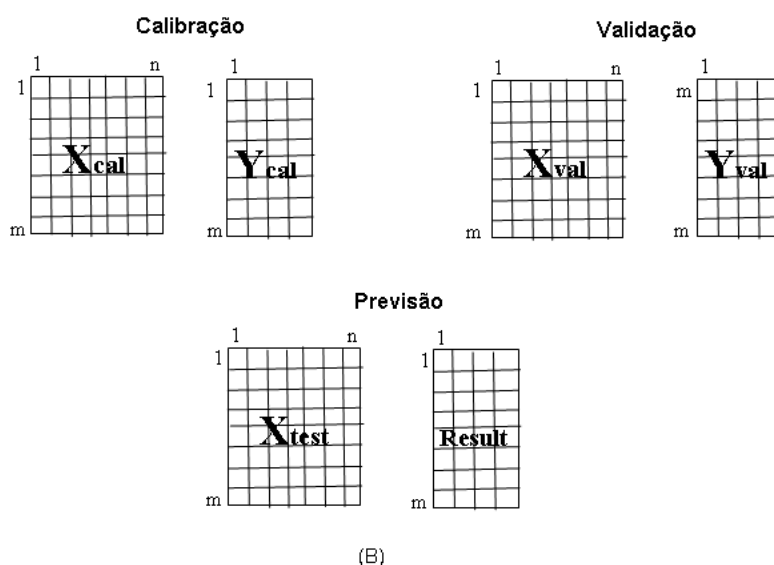
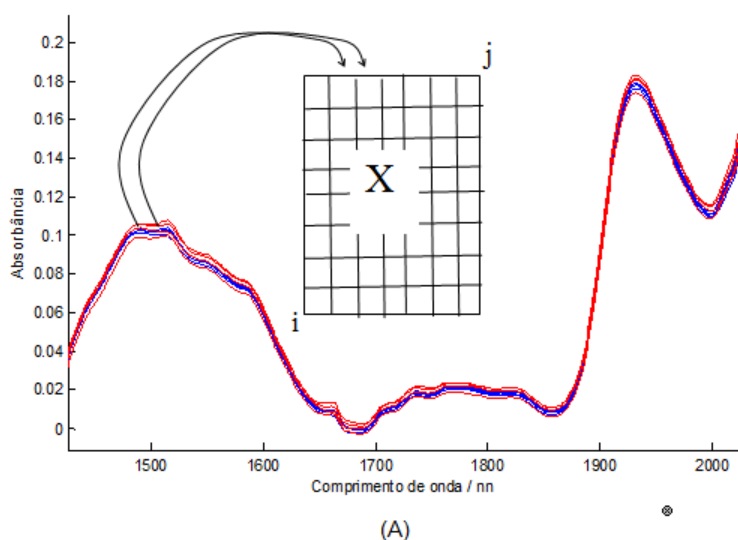
Nas últimas décadas, a calibração multivariada vem ganhando espaço e se tornando uma ferramenta de grande potencial na manipulação e extração de informações de grandes quantidades de dados químicos. Com uma gama de possibilidades de aplicação, a calibração multivariada tem sido utilizada na determinação tanto de propriedades físicas quanto químicas, nos setores de alimentos, agricultura, pecuária, ambiental e farmacêutico [19,20,23].

Em métodos de calibração multivariada duas ou mais respostas instrumentais são relacionadas com uma ou mais propriedades de interesse. Esses métodos possibilitam análises mesmo na presença de interferentes, desde que estes estejam presentes nas amostras de calibração, determinações simultâneas e análises sem resolução do sinal analítico. Isso faz com que, modelos de calibração multivariada sejam uma alternativa quando métodos univariados não podem ser aplicados. Diversos modelos de calibração multivariada vêm sendo utilizados, tais como Regressão Linear Múltipla (RLM), Regressão por Componentes Principais (PCR) e Regressão por Mínimos Quadrados Parciais (PLS) [27], sendo este último, o método utilizado neste trabalho.



### 3.2. Disposição dos dados

Em calibração multivariada, frequentemente trabalha-se com grande quantidade de dados; esses podem ser organizados em uma matriz onde cada linha representa uma amostra, com diferentes concentrações ou outra propriedade de interesse e cada coluna representa uma variável, como a absorbância em dados comprimento de onda ou outra variável relacionada com a propriedade de interesse, como representado na Figura 6 A.



**Figura 6.** Organização dos dados usados em calibração multivariada:

- (A) Cada coluna da matriz contém absorbâncias em um dado comprimento de onda e cada linha representa uma amostra;
- (B) Cada conjunto, de calibração e de validação, possui uma matriz de dados  $X$  e uma matriz  $Y$  ou vetor  $y$  (no caso de um único analito) contendo as concentrações.

### 3.3. Regressão linear múltipla (MLR, do inglês *Multiple Linear Regression*)

A RLM busca estabelecer uma relação linear entre sinal analítico e concentração (ou outra propriedade) aplicando o método dos mínimos quadrados. Na construção de um modelo RLM cada variável dependente, ou seja, cada vetor  $\mathbf{y}_{cal}$  que contém o valor da propriedade de interesse da matriz  $\mathbf{Y}_{cal}$ , é expressa como uma combinação linear das variáveis independentes contidas na matriz  $\mathbf{X}_{cal}$  e um vetor,  $\mathbf{b}$ , que contém os coeficientes de regressão, e é dada pela equação 8.

$$\mathbf{y} = \mathbf{X} \mathbf{b} + \mathbf{e} \quad \text{ou} \quad \mathbf{y} = \mathbf{b}_1 \mathbf{x}_1 + \mathbf{b}_2 \mathbf{x}_2 + \dots + \mathbf{b}_n \mathbf{x}_n \quad (8)$$

A solução por mínimos quadrados para se determinar os coeficientes do vetor  $\mathbf{b}$  é dada pela equação 9.

$$\mathbf{b} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad (9)$$

Amostras de previsão ou de concentração desconhecida podem ser estimadas por:

$$\hat{\mathbf{y}} = \mathbf{X} \mathbf{b} + \mathbf{e} \quad (10)$$

Apesar do grande potencial da RLM, este método apresenta alguns problemas, restringido sua aplicação. Um destes problemas é o fato do modelo ser baseado na resolução de um sistema de equações lineares simultâneas, assim o número de amostras deve ser pelo menos igual ao número de variáveis, caso contrário o sistema se torna indeterminado. Outra limitação da RLM é que toda a informação contida na matriz  $\mathbf{X}$  é incorporada ao modelo, independente se relevante ou não, tornando o modelo instável e pouco robusto. Outro problema está na etapa de inversão da matriz  $(\mathbf{X}^T \mathbf{X})$ ; se a matriz  $\mathbf{X}$  contiver menos linhas que colunas (menos amostras que variáveis), ela não poderá ser invertida e não existirá solução única [33].

Os problemas da RLM apresentados podem ser contornados pelo método PCR, no qual a regressão é feita nos escores, que são as projeções das amostras no novo sistema de coordenadas obtido na PCA, ao invés de ser feita nas variáveis originais.

### 3.4. Regressão por componentes principais (PCR, do inglês Principal Component Regression)

A PCR decompõe os dados originais da matriz  $\mathbf{X}$  em componentes principais, nas quais é realizada a regressão, relacionando os escores desta matriz com a propriedade de interesse.

A concentração (ou outra propriedade) de cada analito  $p$  a ser determinado na amostra é representada por um vetor na matriz  $\mathbf{Y}$  e o espectro de cada amostra é formado pelas absorvâncias nos diferentes comprimentos de onda, definindo cada amostra por um vetor na matriz de dados  $\mathbf{X}$ . Ao se construir um modelo com  $m$  amostras no conjunto de calibração e registrados os espectros em  $n$  comprimentos de onda, os dados podem ser descritos por uma matriz de dados  $\mathbf{X}$  com dimensões  $m$  linhas por  $n$  colunas e uma matriz  $\mathbf{Y}$  com as concentrações com dimensões  $m$  por  $p$ .

O processo de calibração inicia com a decomposição da matriz  $\mathbf{X}$  em componentes principais.

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E} \quad (11)$$

Onde  $\mathbf{T}$  é a matriz de escores com as componentes principais do modelo e  $\mathbf{P}$  a matriz de pesos, que contém informações sobre a contribuição de cada componente principal para explicar a variância dos dados.

Um número de componentes principais é escolhido para descrever a matriz  $\mathbf{X}$ , conciliando o menor número de componentes principais e maior variância explicada; então a matriz  $\mathbf{X}$  pode ser representada por sua matriz de escores  $\mathbf{T}$ :

$$\mathbf{T} = \mathbf{XP} \quad (12)$$

A matriz com as concentrações  $\mathbf{Y}$  é então relacionada com os escores de  $\mathbf{X}$ :

$$\mathbf{Y} = \mathbf{TB} + \mathbf{E} \quad (13)$$

Sendo  $\mathbf{B}$  uma matriz com os coeficientes de regressão obtidos, relacionados com os valores de concentrações conhecidas do conjunto de calibração:

$$\mathbf{B} = (\mathbf{T}^T \mathbf{T})^{-1} \mathbf{T}^T \mathbf{Y} \quad (14)$$

Após estar definido o modelo de calibração, pode se prever a concentração de amostras desconhecidas a partir dos espectros registrados.

A partir da matriz de pesos  $\mathbf{P}$ , obtida na etapa de calibração, calcula-se os escores ( $\mathbf{T}^*$ ) das amostras de previsão ( $\mathbf{X}^*$ ).

$$\mathbf{T}^* = \mathbf{X}^* \mathbf{P} \quad (15)$$

Então se calcula a concentração das novas amostras usando os coeficientes de regressão obtidos na etapa de calibração.

$$\mathbf{Y}_{\text{previsão}} = \mathbf{T}^* \mathbf{B} \quad (16)$$

Uma deficiência da PCR é que a decomposição dos dados é baseada apenas nos dados da matriz  $\mathbf{X}$ , não considerando a variância relacionada às concentrações  $\mathbf{Y}$ , a qual pode ter uma influencia significativa na variância global e assim comprometer a capacidade de previsão do modelo.

### 3.5. Mínimos Quadrados Parciais (PLS, do inglês *Partial Least Squares*)

O método PLS é o mais utilizado para a construção de modelos de calibração multivariada. Esse método relaciona a matriz  $\mathbf{X}$ , que contém os dados espectrais, e o vetor  $\mathbf{y}$  (ou matriz  $\mathbf{Y}$  quando mais de uma variável é prevista) contendo os valores das concentrações. Essas duas matrizes são decompostas simultaneamente em uma soma de  $A$  variáveis latentes (VL), de acordo com o seguinte modelo estrutural, dado pelas equações 17 e 18.

$$\mathbf{X} = \mathbf{T} \mathbf{P}^T + \mathbf{E} = \sum \mathbf{t}_A \mathbf{p}_A^T + \mathbf{E} \quad (17)$$

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.

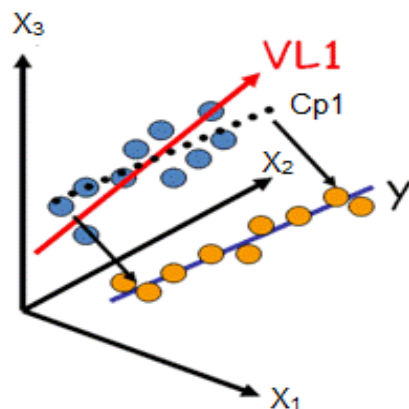
$$\mathbf{y} = \mathbf{T}\mathbf{q}^T + \mathbf{f} = \sum \mathbf{t}_A \mathbf{q}_A^T + \mathbf{f} \quad (18)$$

Onde  $\mathbf{q}$  é o vetor de pesos para a decomposição de  $\mathbf{y}$ ,  $\mathbf{f}$  é o vetor de resíduos do vetor de concentrações  $\mathbf{y}$ . O produto de  $\mathbf{T}$  e  $\mathbf{P}^T$  é uma aproximação das variáveis independentes (ex: espectros) e o produto de  $\mathbf{T}$  e  $\mathbf{q}^T$  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 ( $\mathbf{y}$ ) e às medidas ( $\mathbf{x}$ ). A concentração das novas amostras pode ser estimada a partir dos novos escores,  $\hat{T}$ , e dos pesos do modelo,  $\mathbf{q}$ , de acordo com a equação:

$$y_n = \hat{T} \mathbf{q}^T \quad (19)$$

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. Quando o número de amostras de calibração for superior a 20, outras alternativas de validação cruzada são recomendadas, tais como divisão em blocos contínuos, subconjuntos aleatórios ou *venetians blinds* [34].

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 VLs não são ortogonais entre si. Como pode ser notado, há um compromisso entre a habilidade das VLs em descrever as amostras nos espaços individuais (modelagem dos blocos  $\mathbf{X}$  e  $\mathbf{y}$ ) simultaneamente, levando à perda de ortogonalidade [35].



**Figura 7.** Formação de uma VL por rotação no eixo da componente principal.

Deve-se ainda mencionar a diferença entre PLS1, quando a regressão é feita individualmente para cada variável dependente ( $y$  é um vetor), e PLS2, quando todas as variáveis dependentes são previstas simultaneamente ( $Y$  é uma matriz). O número após o PLS indica, portanto, a dimensão do arranjo de dados a ser previsto [35].

### 3.6. Validação Analítica de Métodos Quimiométricos

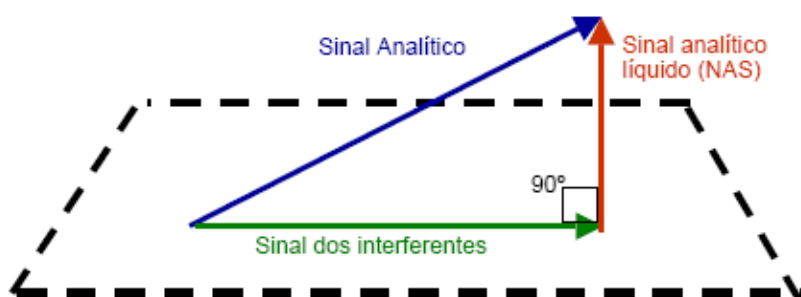
A validação de um método analítico é realizada para garantir, por meio de estudos experimentais, que o método atenda às exigências das aplicações a que se propõe, assegurando a confiabilidade dos resultados para um determinado intervalo. Para tanto, é necessário realizar um estudo quantitativo dos resultados da aplicação do modelo em amostras que não fizeram parte do conjunto de calibração. Assim, utiliza-se um conjunto de amostras autênticas e independentes das utilizadas na calibração, mas que sejam representativas das mesmas e das futuras amostras a analisar, denominado conjunto de validação. As concentrações das amostras deste conjunto de validação devem ser conhecidas e comparadas com aquelas previstas pelo modelo construído [30,31].

Para isto, são estimadas figuras de mérito, tais como linearidade, exatidão, precisão, capacidade de detecção e limite de quantificação. O modo de estimar essas figuras de mérito e os critérios de aceitação é estabelecido em normas e resoluções nacionais, regulamentadas pela Agência Nacional de Vigilância Sanitária (ANVISA) [24], e internacionais, como a Sociedade Americana de Testes e Materiais (ASTM, do inglês *American Society for Testing and Materials*) [36], 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” [23]. A seguir, será apresentada a maneira como são calculadas figuras de mérito, para as quais, no caso de métodos multivariados, é muito útil o conceito de Sinal Analítico Líquido (NAS, do inglês *Net Analyte Signal*).

### 3.7. Sinal analítico líquido - NAS

Em calibração multivariada, o conceito de NAS exerce uma importante função na determinação de figuras de mérito. Esse conceito foi proposto por Lorber em 1986 [37], e é definido como a parte do sinal analítico que é ortogonal às contribuições de possíveis interferentes presentes na amostra, ou seja, representa a parte do sinal analítico que pode ser atribuído exclusivamente ao analito. Essa propriedade de ortogonalidade do NAS está representada geometricamente na Figura 8. De posse do módulo (ou norma) do NAS, é possível calcular um valor escalar relacionado unicamente com a propriedade a ser determinada no analito e assim construir uma nova representação dos dados multivariados, que agora pode ser representados de forma pseudo-univariada.



**Figura 8.** Representação geométrica da propriedade de ortogonalidade do NAS.

Para o cálculo do NAS, as matrizes de dados  $\mathbf{X}$  e  $\mathbf{y}$  são reconstruídos com as  $A$  variáveis latentes utilizadas no modelo, gerando  $\mathbf{X}_A$  e  $\mathbf{y}_A$ , conforme as equações 20 e 21, respectivamente [30,38,39].

$$\mathbf{X}_A = \mathbf{T}_A \mathbf{P}_A + \mathbf{E} \quad (20)$$

Onde  $\mathbf{X}$  representa os espectros,  $\mathbf{T}$  os escores,  $\mathbf{P}$  os pesos, e  $\mathbf{E}$  contém os erros de decomposição da matriz  $\mathbf{X}$ .

$$\mathbf{y}_A = \mathbf{T}_A \mathbf{q}_A + \mathbf{f} \quad (21)$$

Onde  $\mathbf{T}$  é a matriz contendo os escores,  $\mathbf{q}$  é o vetor contendo os pesos e  $\mathbf{f}$  é o vetor com os erros de decomposição do vetor  $\mathbf{y}$ . O próximo passo é determinação da matriz  $\mathbf{X}_{A,-k}$ , que é uma 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 [39,40]:

$$\mathbf{X} = (\mathbf{X}\mathbf{X}^+)\mathbf{X} \quad (22)$$

$$\mathbf{X}^+ = \mathbf{X}^+ (\mathbf{X}\mathbf{X}^+) \quad (23)$$

Tem-se então, através das informações acima, que se um vetor  $\mathbf{z}$  for uma combinação linear de  $\mathbf{X}$ , a multiplicação deste vetor pela matriz  $\mathbf{X}\mathbf{X}^+$  fornecerá como resultado o próprio vetor  $\mathbf{z}$ . 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  $\mathbf{X}$ . Assim, a matriz  $\mathbf{X}_{A,-k}$  é calculada como:

$$\mathbf{X}_{A,-k} = [\mathbf{I} - \mathbf{y}_{A,k} \mathbf{y}_{A,k}^+] \mathbf{X}_A \quad (24)$$

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



$$\mathbf{y}_{A,k} = \mathbf{X}_A \mathbf{X}_A^+ \mathbf{y}_k. \quad (25)$$

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  $\mathbf{x}_{A,k,i}^{\text{NAS}}$ , que contém as informações da espécie k, livre de interferentes, para cada amostra i, de acordo com a seguinte equação:

$$\mathbf{x}_{A,k,i}^{\text{nas}} = (\mathbf{I} - \mathbf{X}_{A,-k}^T (\mathbf{X}_{A,-k}^T)^+ ) \mathbf{x}_{A,i} \quad (26)$$

Assim, o módulo de cada vetor fornece um valor escalar de nas ( $\text{nâs}_i$ ), que pode ser usado como um sinal analítico univariado seletivo/puro.

$$\text{nâs}_i = \|\mathbf{x}_{A,k,i}^{\text{nas}}\| \quad (27)$$

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 pode ser expresso de maneira pseudo-univariada [41,42]. Essa representação requer: (i) o cálculo do  $\text{nâs}$  para todas as amostras de calibração; (ii) determinação do coeficiente  $b_{\text{nas}}$  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  $\mathbf{y}$ , conforme a equação a seguir.

$$b_{\text{nas}} = (\mathbf{nâs}^T \mathbf{nâs})^{-1} \mathbf{nâs}^T \mathbf{y} \quad (28)$$

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

$$\mathbf{y} = b_{\text{nas}} \mathbf{nâs} + \mathbf{e} \quad (29)$$

O modelo com base na regressão dos valores de referência contra o NAS, obtido por meio das equações 28 e 29, é chamado de modelo pseudo-univariado e fornece resultados equivalentes aos obtidos pelo modelo multivariado (PLS).

Quando os dados são centrados na média, antes da determinação do coeficiente de regressão  $b_{nas}$ , é necessário fazer uma correção no vetor  $\mathbf{n\hat{a}s}$ , para 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  $(y-\bar{y})$ , onde  $\bar{y}$  é a média do vetor  $\mathbf{y}$ , que contém os valores de referência [43].

### 3.8. Figuras de Mérito

A regulamentação tradicional [24-26] preconiza que na validação de métodos analíticos para determinação de princípios ativos sejam estimadas as seguintes figuras de mérito: especificidade / seletividade (SEL), linearidade, exatidão, precisão e intervalo. O limite de detecção não é exigido para este tipo de método, mas apenas para a quantificação de impurezas e em métodos para a validação de limpeza. Visando uma completa validação analítica, este trabalho também estimou as figuras de mérito: sensibilidade (SEN), sensibilidade analítica ( $\gamma$ ), viés (*bias* ou teste para detecção de erros sistemáticos) e capacidade de detecção ( $CC\beta$ ). A SEL para métodos multivariados é calculada como a razão entre a norma do vetor **NAS** e a norma de cada espectro. Uma vez que cada amostra apresenta uma SEL diferente, um valor médio é usado para caracterizar o método [44,45]. Como métodos multivariados NIRS não necessitam de resolução do sinal, não há sentido em estabelecer um valor limite para SEL. Uma alternativa para avaliar a especificidade é a demonstração de que o método é capaz de distinguir a espécie de interesse de outros componentes, tais como impurezas, produtos de degradação ou de outros princípios ativos.

A precisão e a exatidão podem ser estimadas de forma semelhante para os métodos univariados e multivariados. Neste trabalho, a precisão foi avaliada em dois níveis, repetibilidade e precisão intermediária. A exatidão pode ser avaliada através dos erros de previsão relativa das amostras individuais. No entanto, a exatidão para

métodos multivariados costuma ser também avaliada através de parâmetros, tais como RMSEC (raiz quadrada do erro quadrático médio de calibração), equação 30, RMSECV (raiz quadrada do erro quadrático médio de validação cruzada) e RMSEP (raiz quadrada do erro quadrático médio de previsão). Dentre estes, o RMSEP, equação 31 é o parâmetro mais robusto, porque é estimado a partir de amostras de validação externa.

$$RMSEC = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{v}} \quad (30)$$

onde  $y_i$  é o valor de referência da concentração do analito de interesse,  $\hat{y}_i$  é o valor estimado pelo modelo, e  $v$  é o número de graus de liberdade, que é igual ao número de amostras usadas na calibração menos o número de variáveis latentes do modelo. Para dados centrados na média, usual em métodos espectroscópicos, perde-se ainda mais um grau de liberdade.

$$RMSEP = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (31)$$

em que  $n$  é o número de amostras de validação.

A linearidade pode ser avaliada a partir do gráfico dos valores de referência *versus* valores previstos, através do ajuste e do coeficiente de correlação ( $r$ ). Porém, antes é necessária a verificação do comportamento aleatório dos resíduos deste ajuste. O intervalo é estabelecido pela confirmação de que o método proporciona um grau aceitável de linearidade, exatidão e precisão quando aplicado a amostras no intervalo de concentração a que o método se propõe. A SEN de métodos multivariados é estimada como o NAS em concentração unitária, que é equivalente à seguinte equação [44,46].

$$SEN = \frac{1}{\|b\|} \quad (32)$$

onde " $\| \cdot \|$ " indica a norma euclidiana de um vetor. Uma figura de mérito mais útil é a sensibilidade analítica ( $\gamma$ ), que é definida, por analogia com a calibração univariada [47], como a relação entre SEN e o ruído instrumental ( $\varepsilon$ ). O ruído instrumental pode ser estimado através da norma do vetor de resíduos do modelo PLS [48] ou através do desvio padrão combinado de um vetor contendo replicatas de dez espectros do branco [44].

$$\gamma = \frac{SEN}{\varepsilon} \quad (33)$$

O inverso de  $\gamma$  ( $\gamma^{-1}$ ) fornece uma estimativa da diferença de concentração mínima que é discernível pelo método analítico, considerando o ruído aleatório experimental como a única fonte de erro, independentemente da técnica específica empregada.

Segundo a IUPAC [49], o termo viés (teste para erro sistemático ou *bias*, em inglês) expressa erros sistemáticos, sendo definido como a diferença entre a média e o valor de referência, correspondendo a todas as componentes de erro que não são aleatórias. Ele é calculado apenas a partir do conjunto de validação, conforme norma da ASTM [36], de acordo com a equação 34:

$$\text{bias} = \sum_{i=1}^{n_v} \frac{(y_i^{\text{ref}} - \hat{y}_i)}{n_v} \quad (34)$$

Onde  $y_i^{\text{ref}}$  e  $\hat{y}_i$  são os valores de concentração de analito de referência e previstos pelo modelo PLS, respectivamente, e  $n_v$  é o número de amostras no conjunto de validação. O desvio padrão de erros de validação (DPV) também é calculado e utilizado em um teste t para determinar se as estimativas de validação mostram uma tendência estatisticamente significativa.

$$DPV = \sqrt{\frac{\sum [(y_i^{\text{ref}} - \hat{y}_i) - \text{bias}]^2}{n_v - 1}} \quad (35)$$

O valor de t calculado é então comparado com o valor t crítico com o número adequado de graus de liberdade, que é igual a  $n_v$ .

O conceito de  $CC\beta$ , presente na ISO 11843-2 [50], foi estendido para a calibração multivariada por Ortiz *et al.* [51], de acordo com a equação 36, considerando um conjunto de calibração, sem amostras replicadas.

$$CC\beta = \delta_{\alpha,\beta,v} \frac{s(e)}{b_{cal}} \sqrt{\frac{1}{n_c} + 1 + \frac{\bar{y}^2}{\sum_{i=1}^{n_c} (y_i - \bar{y})^2}} \quad (36)$$

Nesta equação  $s(e)$  e  $b_{ca}$  são, respectivamente, o desvio padrão dos resíduos e a inclinação da curva dos valores de referência contra os valores previstos de concentração,  $n_c$  é o número de amostras no conjunto de calibração e  $\delta_{\alpha,\beta,v}$  é o parâmetro de uma distribuição t não-central com probabilidades  $\alpha$  e  $\beta$  e o número de graus de liberdade ( $v$ ) igual a  $n_c-2$ ; neste trabalho, foi usado  $\delta = 3,328$ , correspondendo às probabilidades  $\alpha$  e  $\beta$  de 0,05 (nível de confiança de 95%) [52].

Também foi estimada a relação de desempenho do desvio (RPD, do inglês *residual prediction deviation* ou *relative predictive determinant*). A RPD é a razão entre a raiz quadrada da variância dos valores de referência nas amostras de calibração ou validação e a magnitude dos erros que ocorrem na previsão. O RPD representa o quão bem o modelo de calibração prevê um conjunto de amostras específico e é mais adequado para avaliar o desempenho de um modelo em termos absolutos [53-55]. Especificamente, o RPD é calculado para o conjunto de calibração e de validação de acordo com as seguintes equações:

$$RPD_{cal} = \frac{DP_{cal}}{RMSECV} \quad (37)$$

$$RPD_{val} = \frac{DP_{val}}{RMSEP} \quad (38)$$

Onde DP é o desvio padrão dos valores de referência para os conjuntos de calibração,  $DP_{cal}$ , e validação,  $DP_{val}$ .

### 3.9. Detecção de amostras anômalas (*outliers*)

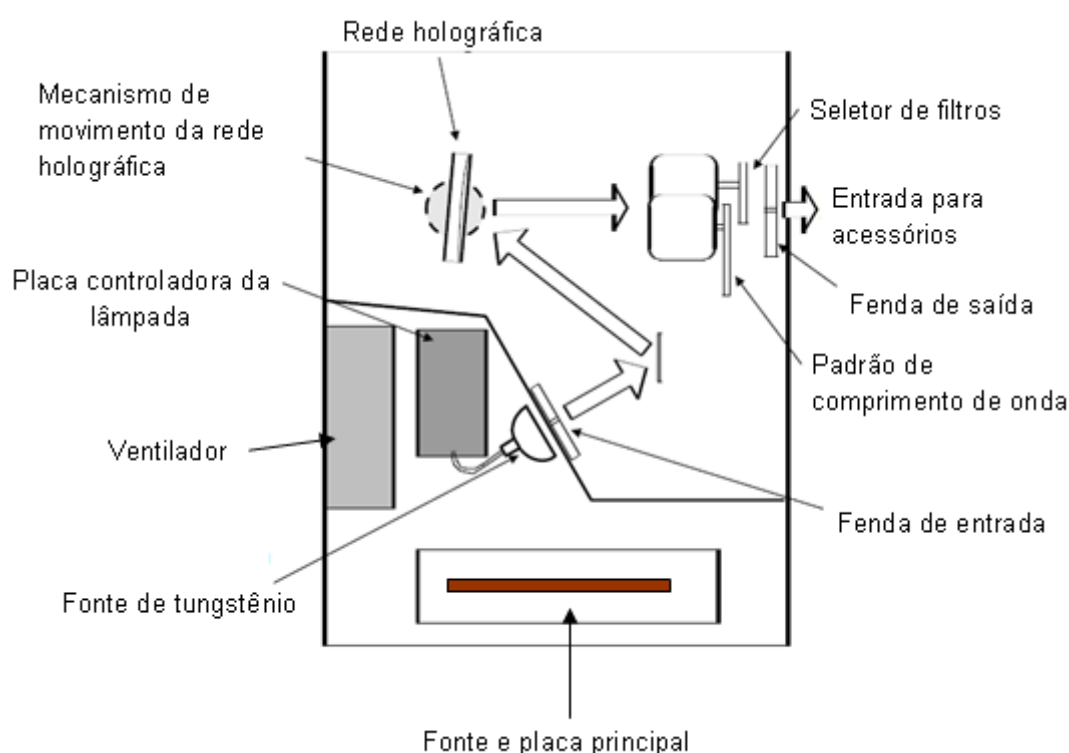
A detecção de amostras anômalas (*outliers*) é fundamental para o desenvolvimento de bons modelos NIRS. Existem várias metodologias que foram propostas para a detecção de amostras anômalas [56-58]. Este trabalho empregou para esse fim um procedimento robusto baseado na identificação de amostras com extrema influência, grandes resíduos nos dados espectrais ou nos valores de concentração analítica [38]. O *leverage*,  $h_i$ , é uma medida da influência de cada amostra no modelo. Amostras com  $h_i$  maior do que um valor limite (três vezes o número de LVs mais um, dividido pelo número de amostras de calibração) devem ser consideradas amostras anômalas. As amostras anômalas com altos resíduos espectrais são detectados por comparação do desvio padrão total,  $s(e)$ , com o desvio padrão de cada amostra,  $s(e_i)$ , relacionadas com os valores de absorbância ou  $\log(1/R)$ , medidos e previstos com A LVs. Se houver uma amostra com  $s(e_i) > 2s(e)$ , esta deve ser excluída, com cerca de 95% de confiança. As amostras anômalas com altos valores de resíduos nas concentrações podem ser detectadas através da comparação do RMSEC do modelo com os erros absolutos de amostras individuais. Se uma amostra tiver uma diferença entre seu valor de referência e sua estimativa maior do que três vezes o RMSEC, ela deve ser excluída.

## 4. Parte experimental

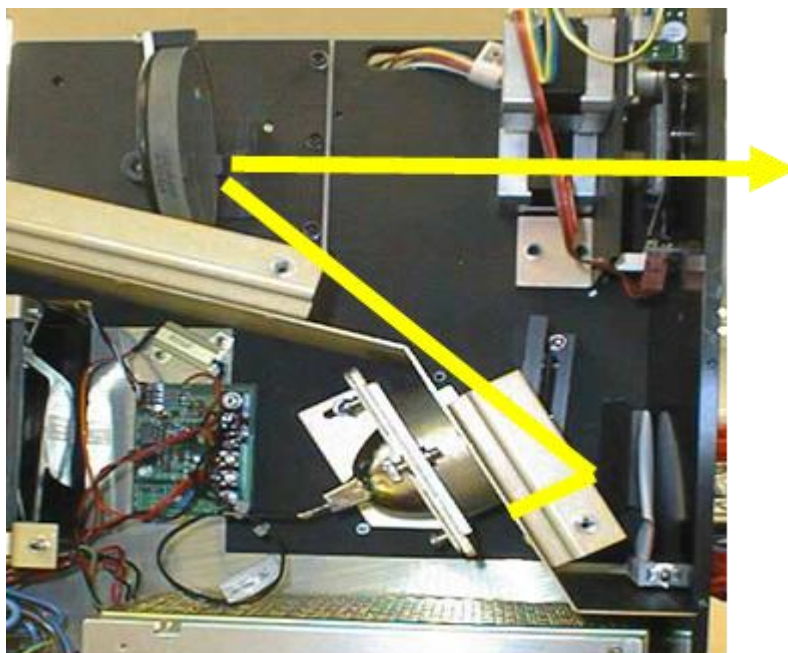
### 4.1. Instrumentação e processamento dos dados

O procedimento experimental deste trabalho foi conduzido nas dependências da Indústria Química do Estado de Goiás (IQUEGO), em Goiânia, Goiás, que disponibilizou a estrutura física e instrumental de seu Laboratório de Controle de Qualidade para a preparação das amostras e obtenção dos dados.

As medidas foram realizadas utilizando um espectrofotômetro específico para a região do infravermelho próximo da marca Foss, modelo 4500 *smart probe analyzer*. Como a maioria dos equipamentos dispersivos disponíveis no mercado, o espectrofotômetro possui fonte de tungstênio para gerar radiação infravermelha, monocromador constituído de rede de difração holográfica côncava e detector de sulfeto de chumbo (PbS). Para as leituras foi usado um acessório de reflectância difusa conectado ao espectrofotômetro por sonda de fibra ótica de 1,5 m. As Figuras 9 e 10 mostram a estrutura e alguns componentes do equipamento.



**Figura 9.** Representação esquemática da estrutura de um espectrofotômetro da marca Foss Modelo 4500.



**Figura 10.** Foto da parte interna de um espectrofotômetro da marca Foss modelo 4500.

O sinal de referência empregado para o branco foi obtido pela medida da cela de porcelana vazia. Os dados foram adquiridos usando o programa do próprio equipamento: Vision 3.3.0.0 (Foss). O processamento dos dados foi feito no programa MATLAB, versão 7.13 (*The MathWorks, Natick, EUA*), usando o pacote PLS Toolbox, versão 6.5 (*Eigenvector Technologies, Manson, EUA*), o qual contém as rotinas para métodos de calibração multivariada, além de uma rotina para a detecção de amostras anômalas elaborada por Jez W.B. Braga [38].

#### **4.2. Reagentes e amostras**

A formulação farmacêutica estudada contém 25 mg de HCTZ e a composição de excipiente é a seguinte: celulose microcristalina, croscamelose sódica, dióxido de silício coloidal e estearato de magnésio. Todas as matérias primas utilizadas na preparação das amostras foram obtidas do estoque da IQUEGO, portando com o mesmo grau de pureza e mesmas especificações utilizadas na produção e foram usadas sem purificação adicional. As amostras de calibração e validação foram preparadas usando para pesagem uma balança analítica (0,0001 g), de acordo com um planejamento experimental, em recipientes de vidro âmbar.



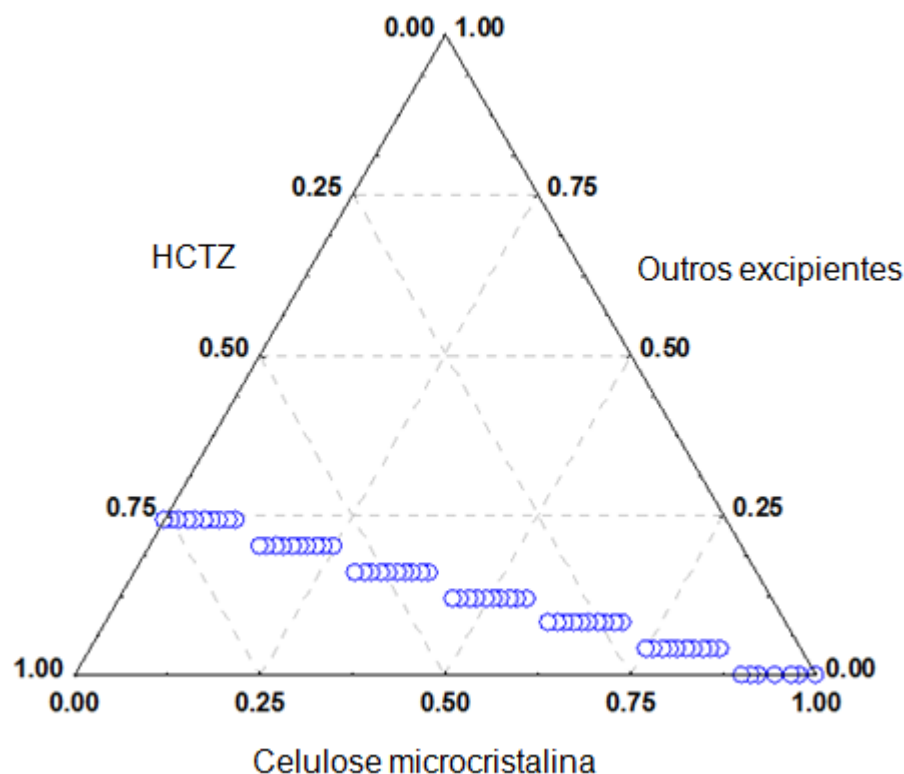
## 4.3. Metodologia

### 4.3.1. Planejamento experimental

Noventa e três amostras foram preparadas de acordo com um planejamento experimental com três fatores: HCTZ, que é o princípio ativo, celulose (principal excipiente), e outros excipientes (croscarmelose, dióxido de silício e estearato de magnésio). A faixa de conteúdo HCTZ abrangeu de 21,25 mg a 29,00 mg por peso médio (150mg) de amostra em intervalos aproximadamente equidistantes, correspondendo de 85,0 a 116,0 % do conteúdo de princípio ativo (25 mg). Esta faixa foi escolhida para cobrir a quantidade de 90 a 110% (22,50 a 27,50 mg de ativo por peso médio) do conteúdo de HCTZ, que são os limites aceitáveis estabelecidos pela Farmacopeia Brasileira [4] para esta formulação. A massa de 25 mg de princípio ativo, corresponde a cerca de 17% m/m do comprimido de HCTZ produzido. Os outros dois fatores (celulose e outros excipientes) também foram variados de acordo com o planejamento experimental apresentado na Figura 11 em termos de pseudocomponentes, que são uma forma prática de se representar as amostras em situações em que a variação de alguns componentes da mistura se encontra em um intervalo muito restrito. Nesse caso, a utilização das proporções reais dos componentes dificulta a visualização da dispersão das amostras no planejamento, e o uso de pseudocomponentes permite, como definido na equação 39, uma visualização mais clara.

$$x_i = \frac{(c_i - a_i)}{(1 - \sum_{i=1}^q a_i)} \quad (39)$$

Onde  $x_i$  é o valor codificado da concentração em termos de pseudocomponentes,  $a_i$  é o limite inferior para o componente  $i$ ,  $c_i$  é a concentração real do componente  $i$  e  $q$  é o número de componentes da mistura.



**Figura 11.** Planejamento experimental utilizado no trabalho, no qual a escala de cada componente é apresentada em termos de pseudocomponentes.

Do total de 93 amostras preparadas, 56 foram separadas para o conjunto de calibração, e 28 formaram o conjunto de validação, juntamente com as 9 amostras usadas para avaliar a precisão. A Tabela 1 mostra a composição das amostras utilizadas no conjunto de calibração e a Tabela 2 as amostras de validação. Por questões de confidencialidade da formulação e segredo industrial, nas tabelas 1 e 2, a coluna que se refere a placebo é constituída da soma das proporções de celulose e dos outros componentes.

**Tabela 1** Composição das amostras do conjunto de calibração

Amostra	Proporção de HCTZ na mistura (m/m)	Proporção de placebo na mistura (m/m)	Massa de HCTZ (g)	Massa de placebo (g)	Teor de HCTZ (%)
1	0,1424	0,8576	4,2726	25,7274	85,00
2	0,1432	0,8568	4,2953	25,7047	85,45
3	0,1447	0,8553	4,3405	25,6595	86,35
4	0,1454	0,8546	4,3631	25,6369	86,80
5	0,1469	0,8531	4,4084	25,5916	87,70

6	0,1477	0,8523	4,4310	25,5690	88,15
7	0,1492	0,8508	4,4762	25,5238	89,05
8	0,1500	0,8500	4,4988	25,5012	89,50
9	0,1522	0,8478	4,5667	25,4333	90,85
10	0,1530	0,8470	4,5893	25,4107	91,30
11	0,1545	0,8455	4,6346	25,3654	92,20
12	0,1552	0,8448	4,6572	25,3428	92,65
13	0,1560	0,8440	4,6798	25,3202	93,10
14	0,1567	0,8433	4,7024	25,2976	93,55
15	0,1575	0,8425	4,7250	25,2750	94,00
16	0,1583	0,8417	4,7477	25,2523	94,45
17	0,1590	0,8410	4,7703	25,2297	94,90
18	0,1598	0,8402	4,7929	25,2071	95,35
19	0,1605	0,8395	4,8155	25,1845	95,80
20	0,1613	0,8387	4,8381	25,1619	96,25
21	0,1620	0,8380	4,8608	25,1392	96,70
22	0,1628	0,8372	4,8834	25,1166	97,15
23	0,1641	0,8359	4,9221	25,0779	97,92
24	0,1643	0,8357	4,9286	25,0714	98,05
25	0,1654	0,8346	4,9633	25,0367	98,74
26	0,1658	0,8342	4,9739	25,0261	98,95
27	0,1668	0,8332	5,0045	24,9955	99,56
28	0,1673	0,8327	5,0191	24,9809	99,85
29	0,1688	0,8312	5,0643	24,9357	100,75
30	0,1696	0,8304	5,0870	24,9130	101,20
31	0,1703	0,8297	5,1096	24,8904	101,65
32	0,1709	0,8291	5,1282	24,8718	102,02
33	0,1718	0,8282	5,1548	24,8452	102,55
34	0,1723	0,8277	5,1694	24,8306	102,84
35	0,1733	0,8267	5,2001	24,7999	103,45
36	0,1737	0,8263	5,2106	24,7894	103,66
37	0,1748	0,8252	5,2453	24,7547	104,35
38	0,1751	0,8249	5,2518	24,7482	104,48
39	0,1764	0,8236	5,2905	24,7095	105,25
40	0,1764	0,8236	5,2931	24,7069	105,30
41	0,1778	0,8222	5,3343	24,6657	106,12
42	0,1779	0,8221	5,3358	24,6642	106,15
43	0,1792	0,8208	5,3755	24,6245	106,94
44	0,1794	0,8206	5,3810	24,6190	107,05
45	0,1809	0,8191	5,4263	24,5737	107,95
46	0,1816	0,8184	5,4489	24,5511	108,40
47	0,1831	0,8169	5,4941	24,5059	109,30
48	0,1839	0,8161	5,5167	24,4833	109,75
49	0,1862	0,8138	5,5846	24,4154	111,10
50	0,1869	0,8131	5,6072	24,3928	111,55
51	0,1884	0,8116	5,6525	24,3475	112,45
52	0,1892	0,8108	5,6751	24,3249	112,90
53	0,1907	0,8093	5,7203	24,2797	113,80
54	0,1914	0,8086	5,7429	24,2571	114,25
55	0,1929	0,8071	5,7882	24,2118	115,15
56	0,1937	0,8063	5,8108	24,1892	115,60

---

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

Amostra	Proporção de HCTZ na mistura (m/m)	Proporção de placebo na mistura (m/m)	Massa de HCTZ (g)	Massa de placebo (g)	Teor de HCTZ (%)
1	0,1439	0,8561	4,3179	25,6821	85,90
2	0,1462	0,8538	4,3857	25,6143	87,25
3	0,1485	0,8515	4,4536	25,5464	88,60
4	0,1507	0,8493	4,5215	25,4785	89,95
5	0,1537	0,8463	4,6119	25,3881	91,75
6	0,1558	0,8442	4,6748	25,3252	93,00
7	0,1572	0,8428	4,7160	25,2840	93,82
8	0,1586	0,8414	4,7572	25,2428	94,64
9	0,1599	0,8401	4,7984	25,2016	95,46
10	0,1613	0,8387	4,8397	25,1603	96,28
11	0,1635	0,8365	4,9060	25,0940	97,60
12	0,1650	0,8350	4,9512	25,0488	98,50
13	0,1665	0,8335	4,9965	25,0035	99,40
14	0,1682	0,8318	5,0457	24,9543	100,38
15	0,1696	0,8304	5,0870	24,9130	101,20
16	0,1711	0,8289	5,1322	24,8678	102,10
17	0,1726	0,8274	5,1774	24,8226	103,00
18	0,1741	0,8259	5,2227	24,7773	103,90
19	0,1756	0,8244	5,2679	24,7321	104,80
20	0,1771	0,8229	5,3132	24,6868	105,70
21	0,1786	0,8214	5,3584	24,6416	106,60
22	0,1801	0,8199	5,4036	24,5964	107,50
23	0,1824	0,8176	5,4715	24,5285	108,85
24	0,1854	0,8146	5,5620	24,4380	110,65
25	0,1877	0,8123	5,6298	24,3702	112,00
26	0,1899	0,8101	5,6977	24,3023	113,35
27	0,1922	0,8078	5,7656	24,2344	114,70
28	0,1944	0,8056	5,8334	24,1666	116,05

#### 4.3.2. Preparação das amostras

Foi preparado um placebo contendo os três excipientes em menor quantidade na formulação de Hidroclorotiazida 25mg produzida pela IQUEGO: croscarmelose sódica, dióxido de silício coloidal e estearato de magnésio. Essas substâncias foram misturadas em saco plástico e homogeneizadas manualmente. Em seguida, esse placebo foi misturado com HCTZ e celulose microcristalina em recipiente de vidro âmbar de 240 mL, adicionando pequenas porções de cada

componente e fazendo a homogeneização por agitação manual, sendo no final obtido uma massa total de 30,0000 g para cada amostra preparada.

Foi realizado um controle de peso a cada 10 amostras preparadas, verificando-se a massa ao final da mistura, visando evitar possíveis erros de pesagem. Foi realizada análise de teor de umidade por método Karl Fischer em três amostras no momento de sua preparação e novamente no final da obtenção de todo o conjunto de espectros, visando avaliar a possível incorporação de água durante o período de armazenamento e aquisição dos dados.

#### **4.3.3. Obtenção dos dados**

Os espectros foram registrados na faixa de 1100 a 2500 nm (passo 2 nm) com 32 varreduras. Para cada amostra foram feitas três leituras e calculou-se o espectro médio. Entre cada leitura, a sonda foi colocada em diferentes posições na amostra para se evitar possíveis erros sistemáticos. Durante todas as medidas, a temperatura ambiente foi monitorada, permanecendo entre 22°C e 28°C e a umidade relativa do ar entre 35% e 65%.

#### **4.3.4. Análise de amostras reais**

O método desenvolvido foi aplicado a amostras reais de três maneiras diferentes. Primeiramente, triplicatas de amostras de comprimidos triturados a pó fino obtidas de três diferentes lotes reais foram analisadas. Cada amostra foi preparada pela pulverização de vinte comprimidos usando pistilo e almofariz de ágata. Em seguida, dez comprimidos intactos do mesmo lote foram amostrados e analisados diretamente com o acessório de reflectância difusa. Finalmente, triplicatas de amostras de pó consideradas sob controle foram coletados na linha de produção, após a etapa de mistura e antes da compressão. É nessa etapa que se avalia a homogeneidade do pó, na qual o método proposto poderá ser aplicado. Estas amostras foram obtidas a partir de seis diferentes lotes, durante um período de três meses, e analisadas. Todos estes resultados foram comparados com o método CLAE de referência.

#### **4.3.5. Análise cromatográfica**

A análise por CLAE foi baseada no método oficial [4]. Foi usado um cromatógrafo líquido da marca Perkin-Elmer, com módulos da série 200 (bomba, sistema de injeção e detector). A detecção foi feita no ultravioleta, em 254 nm. Foi usada uma coluna analítica C-18 (250 x 4,6 mm), empacotada com partículas de sílica de 5 µm e mantida à temperatura ambiente. A fase móvel foi tampão fosfato 0,1 mol/L e acetonitrila (9:1, v/v), ajustada em pH 3,0±0,1, com um fluxo de 2,0 mL/min. As amostras em pó foram dissolvidas em acetonitrila, sonicadas por 15 minutos e filtradas em membrana de 0,45 µm. Todas as injeções foram repetidas três vezes e as determinações foram feitas em triplicata.

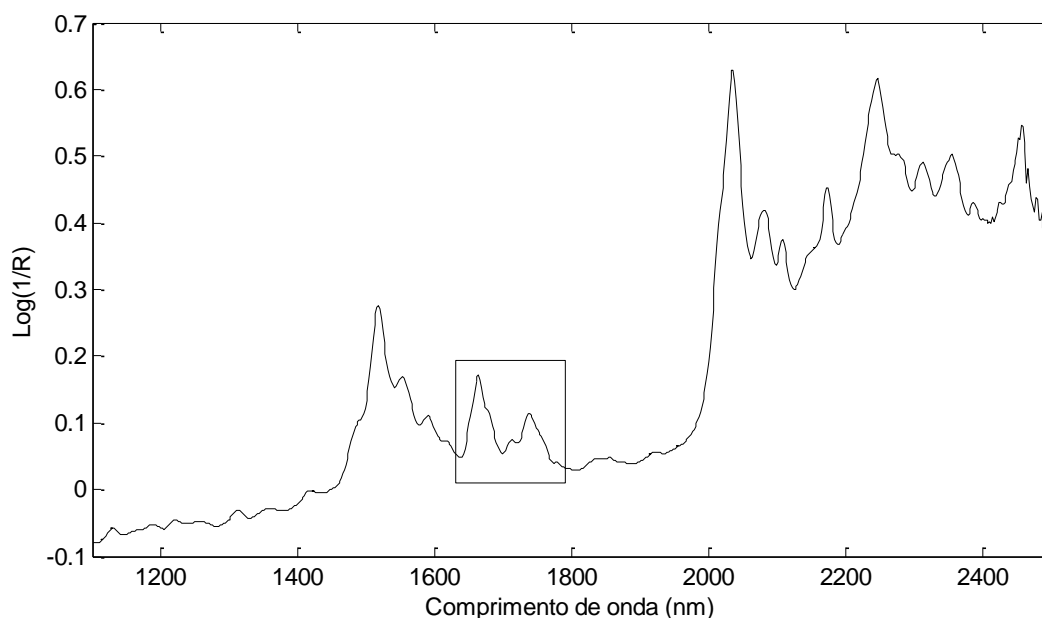
#### **4.4. Validação analítica**

Foram realizadas medidas visando à estimativa de algumas figuras de mérito para a validação do método proposto. Para se avaliar a repetibilidade, precisão intermediária e exatidão, foram analisadas triplicatas de amostras em três níveis de concentração, sendo um inferior com teor de 90,4 %, um médio com teor de 100,3 % e um superior com teor de 110,2 % de HCTZ. Estas mesmas amostras foram analisadas por outro analista em dia diferente para estimar a precisão intermediária. A robustez foi avaliada pela análise dessas mesmas triplicatas usando recipientes diferentes para leitura, sendo nomeados recipientes A, frasco de vidro âmbar, e recipiente B, frasco de polietileno usado para amostragem no setor de produção da indústria. Estas mesmas amostras foram determinadas usando o método cromatográfico oficial, baseado em CLAE e recomendado pela Farmacopeia Brasileira [4]. Finalmente, para avaliar a especificidade, foram analisados três produtos da linha de produção da IQUEGO que, apesar de terem estruturas químicas bem diferentes da HCTZ, contêm celulose microcristalina como principal excipiente e procedimento farmacotécnico semelhante ao da HCTZ. Os produtos selecionados foram Captopril 25 mg comprimido, Ranitidina comprimido revestido 150mg e Glibenclamida comprimido 40 mg, todos analisados na forma de pó. Foram realizadas também 15 leituras repetidas da cela vazia, visando estimar o ruído instrumental. A validação do método foi feita pela comparação dos resultados com o método oficial, baseado em Cromatografia Líquida de Alta Eficiência (CLAE) [4].

## 5. Resultados e discussão

### 5.1. Espectro de HCTZ pura

Uma análise qualitativa do espectro de reflectância difusa da HCTZ pura (Figura 12) ajuda a identificar as regiões espectrais que podem contribuir para a construção do melhor modelo de previsão. A região abaixo de 1400 nm, associada com vibrações de segundos sobretons, não apresenta bandas intensas. Os primeiros sobretons dos estiramentos de ligações N-H são observados na região entre 1480 e 1620 nm, na qual as bandas centradas em aproximadamente 1518, 1552 e 1590 nm podem ser atribuídas às ligações N-H dos três grupos amino da molécula de HCTZ. Os primeiros sobretons dos estiramentos de ligações C-H são observados na região de 1640 a 1780 nm, na qual as bandas centradas em 1664 e 1738 nm podem ser atribuídas às vibrações de C-H aromático e C-H alifático (CH<sub>2</sub> do anel não aromático), respectivamente. A banda mais intensa em 2034 nm é atribuída à combinação de vibrações de estiramento e deformação angular de N-H de amina primária e outra banda de combinação de N-H está presente em 2173 nm. As bandas de combinação de ligações de carbono são observadas acima de 2200 nm, com uma banda intensa em 2246 nm, que pode ser atribuída à combinação de estiramentos de ligações =CH e C=C [59].



**Figura 12.** Espectro NIR obtido por reflectância difusa de HCTZ pura. A região espectral usada no melhor modelo PLS está destacada com um retângulo.

## 5.2. Controle de pesos e teor de umidade

Para uma em cada dez amostras preparadas, foi realizada a determinação da massa de pó ao final da mistura, visando evitar possíveis erros de pesagem. A quantidade teórica de pó para todas as amostras é de 30,0000g. Os resultados das amostras pesadas estão listados na Tabela 3.

**Tabela 3.** Controle de massa final

Amostra	Massa recipiente vazio (g)	Massa recipiente com pó (g)	Massa do pó (g)	Erro (%)
1	145,2132	175,1810	29,9678	0,1073
2	145,5670	175,5470	29,9800	0,0667
3	147,5374	177,5479	30,0105	-0,0350
4	148,1510	178,0933	29,9423	0,1923
5	148,7618	178,7395	29,9777	0,0743
6	147,1168	177,0977	29,9809	0,0637
7	146,4162	176,4513	30,0351	-0,1170
8	147,0201	177,0051	29,9850	0,0500
9	147,1991	177,1758	29,9767	0,0777
10	146,7954	176,8091	30,0137	-0,0457

Das dez amostras avaliadas, obteve-se uma massa média de 29,9870g e um erro médio de 0,04% em relação à massa teórica de 30,0000g.

Durante a preparação das amostras, também foram realizadas análises do teor de umidade em três amostras, no momento de sua preparação e no final da obtenção de todos os dados. A especificação do teor de umidade para HCTZ comprimido é de no máximo 5,0 % e os resultados obtidos são mostrados na Tabela 4.

**Tabela 4.** Controle de do teor de umidade (TU)

TU inicial (%)	TU final (%)	Diferença (%)
4,65	4,78	0,13
4,88	5,05	0,17
4,76	4,88	0,12

Pode-se observar que houve um pequeno aumento de umidade, porém mantendo-se o seu teor abaixo ou próximo do limite especificado, não



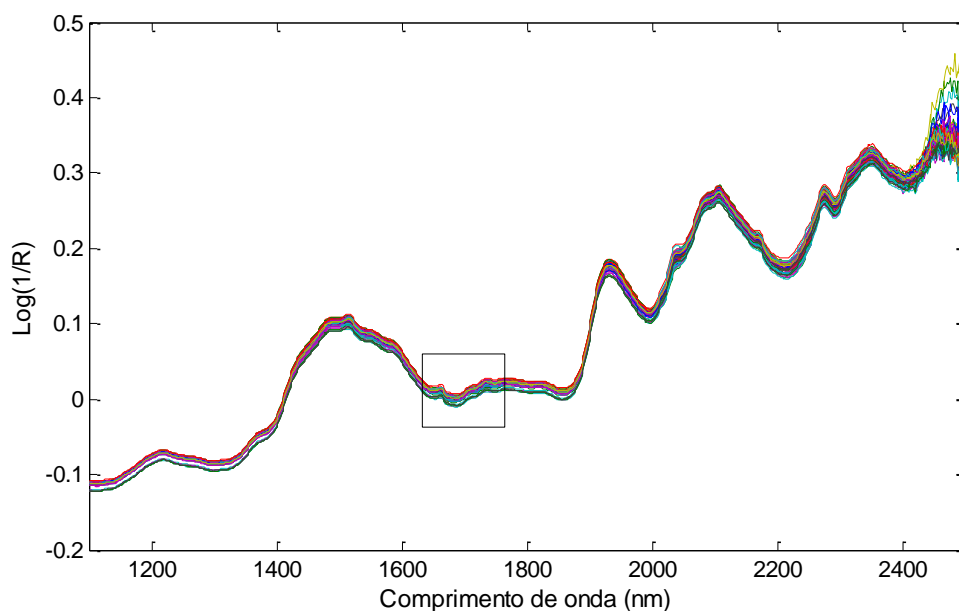
comprometendo a qualidade do modelo, já que as amostras a serem previstas são controladas conforme a especificação do teor de umidade.

### 5.3. Modelo PLS

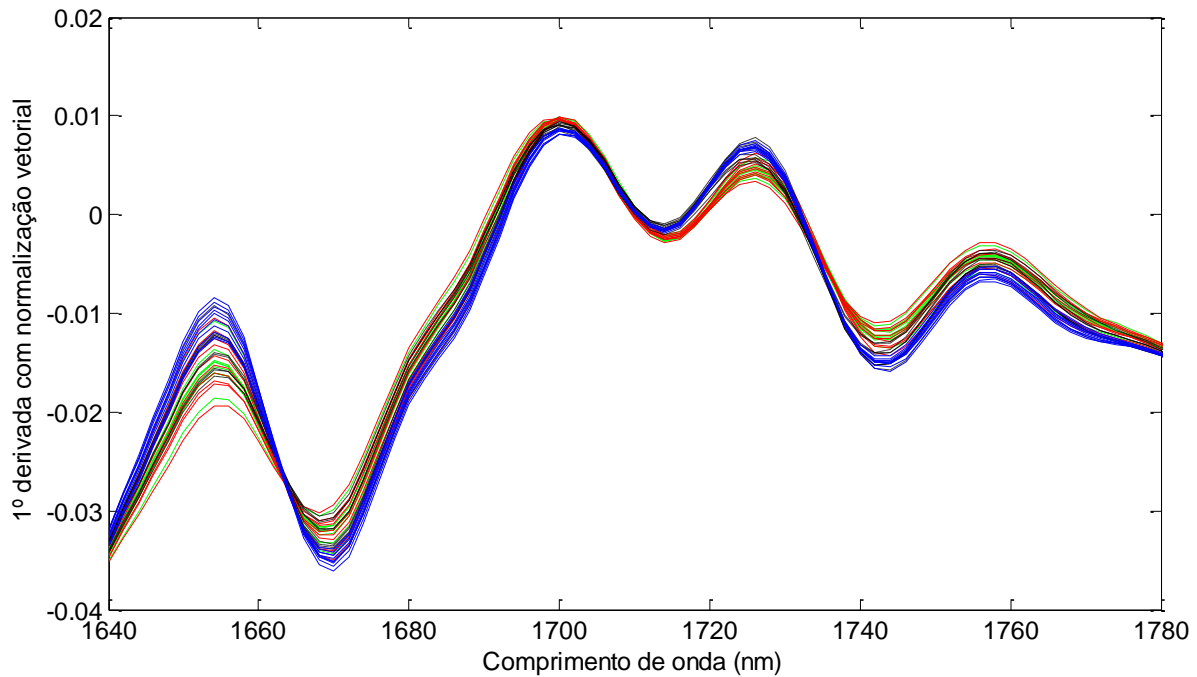
Os espectros das 93 amostras analisadas são mostrados na Figura 13. As amostras foram separadas em 56 para o conjunto de calibração e 37 para o conjunto de validação, distribuídas de forma homogênea e abrangendo toda a faixa de concentração de HCTZ, de acordo com o planejamento experimental. A região acima de 2400 nm apresentou grande ruído instrumental e por isso foi excluída dos modelos. Os espectros de reflectância difusa no NIR para amostras em pó podem apresentar desvios de linha de base relacionados ao espalhamento multiplicativo da luz, que é causado pelas irregularidades na distribuição do tamanho das partículas, um fenômeno estritamente físico. Estes desvios instrumentais não estão relacionados com a composição química da amostra e requerem o uso de métodos de pré-processamento. Foram testados os métodos de pré-processamento mais comuns para a correção de linha de base em espectrofotometria NIR [60]: correção de espalhamento multiplicativo (MSC, do inglês *multiplicative scatter correction*), correção por variação normal padrão (SNV do inglês *standard normal variate*) e alisamento Savitsky-Golay aplicando primeira derivada seguida de normalização vetorial. O número de variáveis latentes (VLs) do modelo PLS foi escolhido por validação cruzada em blocos contínuos e a combinação de primeira derivada, alisamento (filtro com janela de 11 pontos e ajuste polinomial de segunda ordem) e normalização vetorial forneceu o melhor modelo com 5 VLs e um RMSECV de 3,0 %. Modelos usando MSC e SNV forneceram valores de RMSECV maiores e com mais variáveis latentes, sete em ambos os casos. O pré-processamento escolhido teve por objetivo compensar possíveis variações nas condições experimentais e melhorar a relação sinal/ruído, estabilizando o modelo.

Os resultados foram otimizados através da construção de modelos usando diferentes regiões espectrais, que foram escolhidas com base na análise qualitativa do espectro da HCTZ pura, discutida na seção 5.1. Quatro regiões espectrais foram testadas, 1100-1400nm (A), 1480-1620 nm (B), 1640-1780 nm (C) e 2000-2400nm (D). Modelos PLS locais foram comparados com modelos PLS usando todo o espectro e o melhor resultado foi obtido para a região C, que forneceu um RMSECV

de 2,5% com 4 VLs. Outras regiões do espectro forneceram valores similares de RMSECV e com maior número de VLs, tais como 7 para a região B e 8 para a região D, indicando sobreajuste nesses modelos. Assim, o melhor modelo PLS foi construído com os espectros centrados na média e pré-processados aplicando primeira derivada, alisamento Svitsky-Golay e normalização vetorial na região de 1640 a 1780 nm (Figura 13), a qual está associada ao primeiro sobretom do estiramento da ligação C-H. Este modelo explicou 99,68% da variância total no bloco X e 97,31% no bloco Y. Considerando que este modelo utilizou 4 VLs, pode-se concluir que ele usou uma quantidade de amostras que atende às recomendações da ASTM [36], que prescreve  $6 \times (\text{número de VLs} + 1)$  e  $4 \times (\text{número de VLs})$  amostras para os conjuntos de calibração e de validação, respectivamente.



**Figura 13.** Espectros das 93 amostras, correspondentes aos conjuntos de calibração e validação. A região espectral usada no melhor modelo PLS está destacada por um retângulo.



**Figura 14.** Espectros pré-processados com primeira derivada, alisamento e normalização vetorial, na faixa de 1640 a 1780nm.

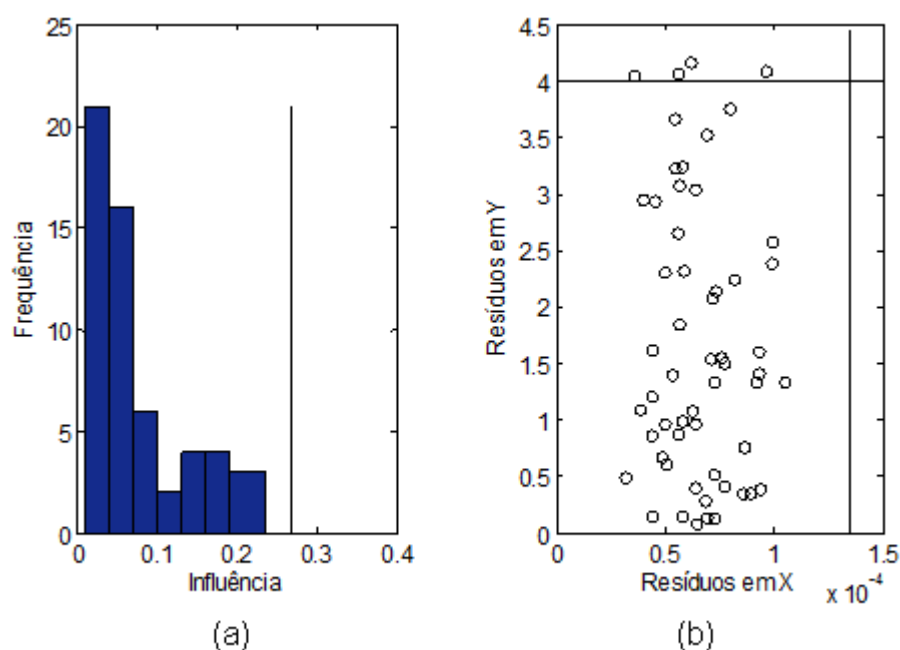
#### 5.4. Detecção de amostras anômalas

Depois de selecionada a região espectral, o modelo PLS foi otimizado através de um procedimento para detecção de amostras anômalas baseado na metodologia descrita na seção 3.9. Neste procedimento, sete amostras anômalas foram detectadas no conjunto de calibração e outras seis no conjunto de validação. A evolução do procedimento é detalhada na Tabela 5, que mostra as mudanças nos valores de RMSEC e RMSEP dos sucessivos modelos construídos. Todos esses modelos foram construídos usando quatro VLs. As Figura 15, 16 e 17 mostram a detecção de amostras anômalas para os modelos construídos, representadas pelo histograma de influência (*leverage*) das amostras e pelo gráfico de resíduos espectrais em função dos resíduos nos valores de concentrações, mostrando também os limites de aceitação. Pode ser observado que no primeiro modelo foram detectadas quatro amostras anômalas, baseadas nos resíduos dos valores das concentrações (Y). Nenhuma amostra anômala foi identificada com base na influência ou nos resíduos espectrais no primeiro modelo construído. Então, o

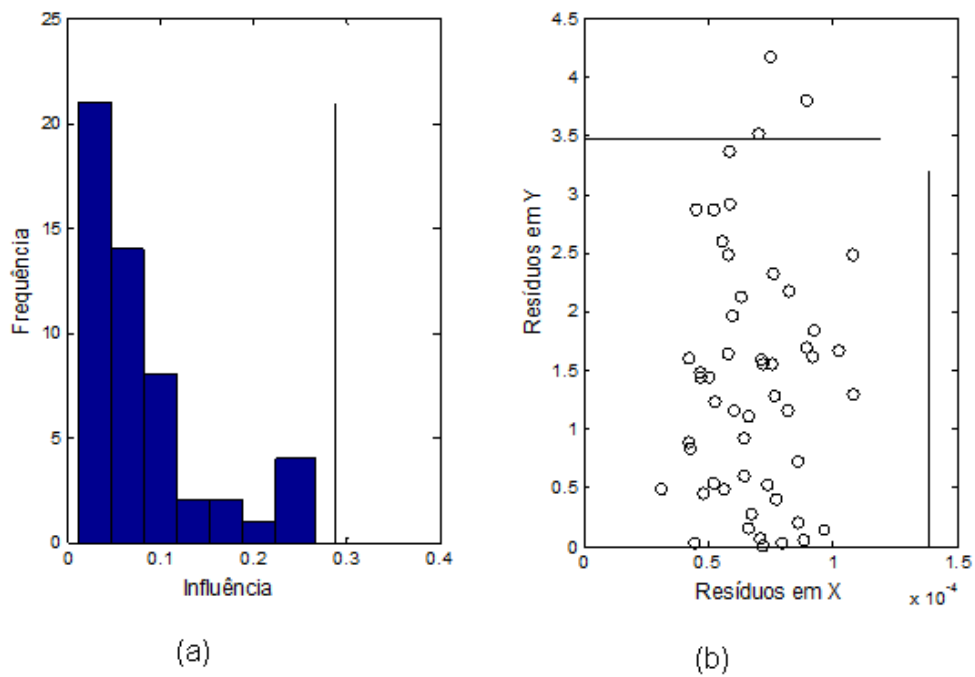
modelo PLS otimizado após detecção e eliminação de amostras anômalas foi construído com 49 amostras de calibração e 31 amostras de validação.

**Tabela 5.** Otimização do modelo PLS pela detecção de amostras anômalas

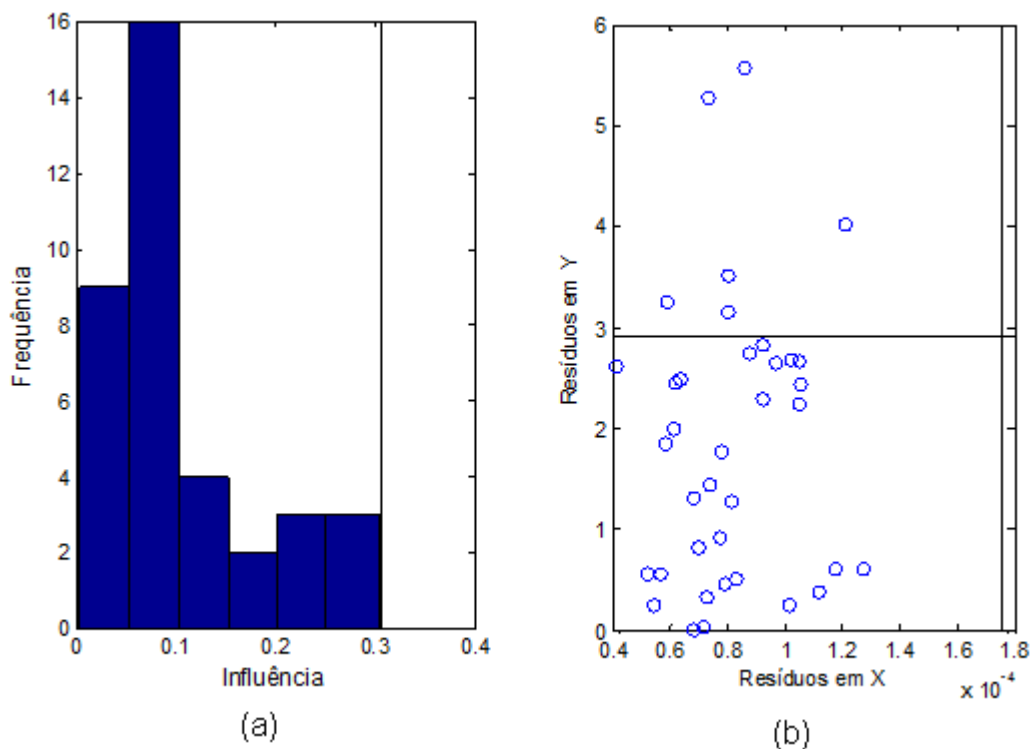
Modelo	1º	2º	3º	Final
Número de amostras de calibração	56	52	49	49
Número de amostras de validação	37	37	37	31
RMSEC (%)	2,0	1,9	1,4	1,4
RMSEP (%)	2,5	2,3	2,3	1,7



**Figura 15.** Detecção de amostras anômalas no conjunto de calibração para o primeiro modelo. (a) Histogramas dos valores de influência (*leverage*). (b) Gráfico dos resíduos espectrais (X) *versus* resíduos de concentração (Y). As linhas contínuas indicam os limites aceitáveis para a detecção de amostras anômalas.



**Figura 16.** Detecção de amostras anômalas no conjunto de calibração para o segundo modelo. (a) Histogramas dos valores de influência (*leverage*). (b) Gráfico dos resíduos espectrais (X) versus resíduos de concentração (Y). As linhas contínuas indicam os limites aceitáveis para a detecção de amostras anômalas.



**Figura 17.** Detecção de amostras anômalas no conjunto de validação para o terceiro modelo. (a) Histogramas dos valores de influência (*leverage*). (b) Gráfico dos resíduos espectrais (X) versus resíduos de concentração (Y). As linhas contínuas indicam os limites aceitáveis para a detecção de amostras anômalas.

## 5.5. Validação Analítica

A Tabela 6 mostra os parâmetros calculados para avaliar as principais figuras de mérito do método desenvolvido. A exatidão pode ser estimada pelos parâmetros RMSECV, RMSEC e, principalmente, RMSEP. O valor de RMSEP obtido foi 1,7%, o qual está abaixo do limite máximo de 2% que é normalmente aceito pela indústria farmacêutica [61]. Os erros de previsão para as amostras de validação ficaram entre -3,2 e +2,7% e são apresentados na Tabela 7. A precisão foi avaliada através do desvio padrão relativo (DPR) obtido para triplicatas em três níveis de concentração de HCTZ. Um DPR de 1,7% foi obtido para a repetibilidade e 2,2% para a precisão intermediária. Ambos os resultados estão dentro dos limites aceitáveis pela regulamentação brasileira [24], que estabelece um DPR máximo de 5%.

A linearidade do método foi avaliada qualitativamente por meio dos resíduos do modelo PLS, que são mostrados na Figura 18. É possível observar a ausência de tendência sistemática na distribuição desses resíduos, o que evidencia um comportamento aleatório dos dados. Uma vez que o gráfico dos resíduos indicou que um modelo linear é uma suposição válida, o ajuste de uma reta dos valores de referência *versus* os valores previstos pelo modelo pode ser usado para estimar um coeficiente de correlação, o qual expressa o grau de concordância entre os valores estimados e de referência. No entanto, este coeficiente de correlação por si só não pode ser usado como uma medida quantitativa da linearidade. O resultado para este ajuste para as amostras de calibração é mostrado na Tabela 6,  $r = 0,9865$ , e está de acordo com a regulamentação brasileira [24], que exige um valor mínimo de 0,99 (0,9865 arredondado para duas casas decimais é igual a 0,99). Considerando a linearidade, precisão e exatidão obtidas, o intervalo do método foi estabelecido de 85,0 a 116,0%, o que corresponde a faixa de concentração de 21,25 a 29,0 mg de HCTZ por 150,00 mg de pó, que equivale ao peso médio teórico do comprimido.

A estimativa da seletividade (SEL) para o método só é possível com a aplicação do conceito de NAS, já que sua norma fornece uma relação quantitativa do sinal instrumental, que é usado na construção do modelo de calibração para a determinação do analito. Embora a SEL seja um parâmetro fundamental para determinações em métodos como CLAE, em modelos multivariados usando espectroscopia NIR, ela não tem nenhum significado prático para o controle de qualidade, já que podem ser obtidos baixos valores de SEL para modelos com alta exatidão. A seletividade média estimada para o método foi de 33,0%, indicando uma

razoável sobreposição de interferentes (excipientes) no sinal da HCTZ. A especificidade foi avaliada, com a aplicação do método a três produtos da linha de produção da IQUEGO, com formulação e processo farmacotécnico similar à da HCTZ, os produtos analisados foram: Captopril 25 mg, Ranitidina 150 mg e Glibenclamida 40 mg, e todos foram identificados como amostras anômalas.

A sensibilidade (SEN) do método foi estimada em 0,00033. Este valor não é apropriado para comparação com outros métodos, já que depende da metodologia analítica empregada e da matriz analisada. Assim, a  $\gamma$  também foi calculada como  $1,9\%^{-1}$ , baseada na estimativa do ruído instrumental, que foi de 0,00038. O inverso de  $\gamma$ , 1,1%, indica a diferença mínima de concentração que o método pode diferenciar considerando o ruído instrumental aleatório como única fonte de erro. A  $CC\beta$  foi estimada como 12,3%, indicando que o método não é capaz de detectar concentrações abaixo deste teor. Porém, o método é adequado para o seu propósito, já que o limite inferior do intervalo analítico é cerca de sete vezes maior que a  $CC\beta$ . O viés (*bias*) foi estimado como  $-0,267 \pm 1,729$  e um teste de  $t$  com 30 graus de liberdade e nível de confiança de 95% demonstrou que o modelo não apresentou erro sistemático ( $t_{\text{calc}} = 0,860 < t_{\text{crit}} = 2,042$ ).

Os valores estimados do parâmetro RPD para a calibração e validação foram 5,6 e 5,1, respectivamente. Estes resultados são satisfatórios, pois modelos com valores de RPD superiores a 5,0 são considerados bons para controle de qualidade [54].

**Tabela 6.** Parâmetros estimados para avaliar as principais figuras de mérito

Figuras de mérito	Parâmetro	Valor
Exatidão	RMSECV	1,6%
	RMSEC	1,4%
	RMSEP	1,7%
Precisão	DPR repetibilidade	1,7%
	DPR intermediária	2,2%
Linearidade	Inclinação	1,02 ± 0,04 <sup>a</sup>
	Intercepto	-2,01 ± 3,6 <sup>a</sup>
	Coeficiente de correlação	0,9865 <sup>a</sup>
Intervalo		85,0 a 116,0%
Seletividade		0,33
Sensibilidade		3,3 x10 <sup>-4</sup> <sup>b</sup>
Sensibilidade analítica		0,9 <sup>c</sup>
$\gamma^{-1}$		1,1%
CC $\beta$	Capacidade de detecção	12,3%
Bias		-0,267±1,729%
RPD	Calibração	5,6
	Validação	5,1

<sup>a</sup> Valores para ajuste da curva para as amostras de calibração

<sup>b</sup> Valores expressos como a razão entre log(1/R) e o teor em %

<sup>c</sup> Valor expresso como %<sup>-1</sup>

**Tabela 7.** Resultados de previsão das amostras de validação, após a exclusão das amostras anômalas.

Valore de referência (%)	Valores estimados (%)	Erro relativo (%)
85,9	86,4	0,6
87,3	88,1	1,0
88,6	85,8	-3,2
90,0	92,4	2,7
91,8	89,5	-2,5
93,8	91,1	-2,9
95,5	94,5	-1,0
96,3	93,5	-2,9
97,6	97,6	0,1
98,5	101,0	2,5
99,4	102,0	2,6

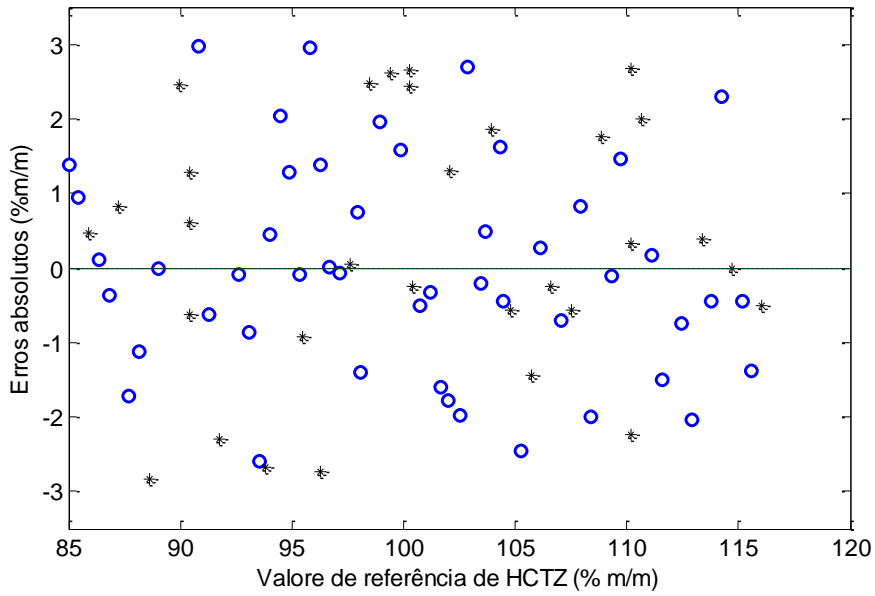


100,4	100,1	-0,2
102,1	103,4	1,3
103,9	105,8	1,8
104,8	104,2	-0,5
105,7	104,3	-1,4
106,6	106,3	-0,2
107,5	106,9	-0,5
108,9	110,6	1,6
110,7	112,6	1,8
113,4	113,7	0,3
114,7	114,7	0,0
116,1	115,5	-0,4
90,4	91,7	1,4
90,4	89,8	-0,7
90,4	91,0	0,7
100,3	102,7	2,4
100,3	102,9	2,6
110,2	110,5	0,3
110,2	112,9	2,4
110,2	108,0	-2,0

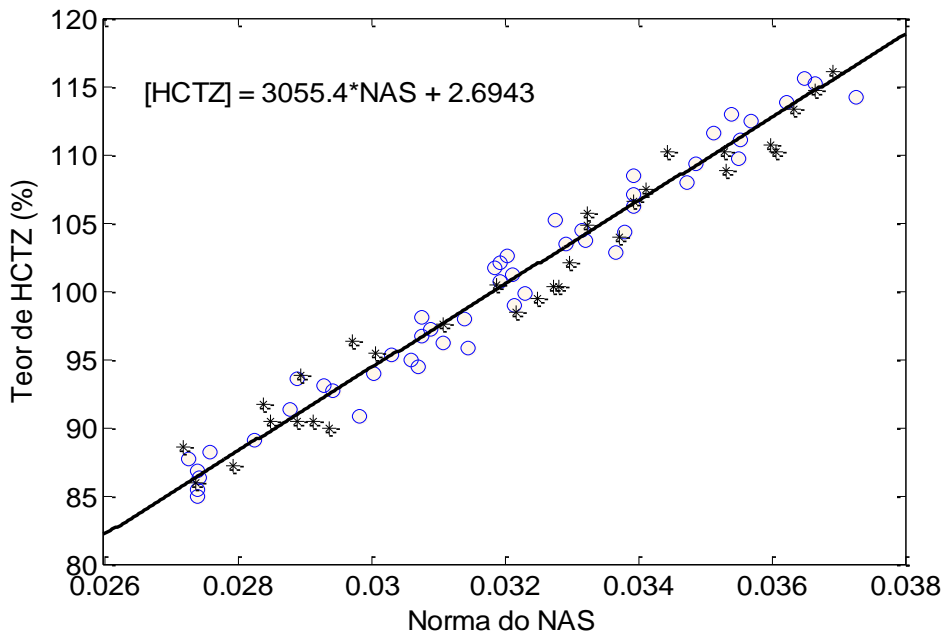
---

Finalmente, uma curva de calibração pseudo-univariada foi obtida (Figura 19), demonstrando uma maneira mais simples de expressar este modelo multivariado. Esta curva é particularmente útil para uma análise rotineira, possibilitando que o analista visualize o modelo PLS de modo univariado. Os valores da norma do NAS são equivalentes ao sinal seletivo obtido do espectro de cada amostra. O ajuste da curva pseudo univariada é dada pela seguinte equação:

$$[\text{HCTZ}] = 3055,4 \|\mathbf{n\hat{a}s}\| + 2,6943 \quad (40)$$



**Figura 18.** Resíduos do modelo PLS para as amostras de calibração (círculos) e validação (asteriscos).



**Figura 19.** Curva de calibração pseudo-univariada. Norma do NAS versus valores de referência das concentrações para as amostras de calibração (círculos) e validação (asteriscos).

A robustez foi avaliada pela análise de triplicatas em três níveis de concentração, um baixo, 90,4%, um médio, 100,3%, e um alto, 110,2%, usando recipientes diferentes para leitura, sendo nomeados recipientes A, frasco de vidro âmbar, e recipiente B, saco de polietileno usado para amostragem no setor de produção da indústria. Os resultados são apresentados na Tabela 8. Um teste F indicou que as variâncias foram estatisticamente iguais e um teste t não pareado com 4 graus de liberdade e 95% de confiança mostrou que os resultados usando o recipiente A foram estatisticamente equivalentes aos obtidos usando o recipiente B, nos três níveis, com  $t_{cal}=0,79$  para o nível inferior,  $t_{cal}=0,97$  para o nível médio e  $t_{cal}=0,14$  para o nível superior. Em todos os casos os valores estimados ficaram abaixo do valor crítico  $t_{tab}=2,776$ .

**Tabela 8.** Amostras lidas em diferentes recipientes

Nível	A (%) <sup>a</sup>	B (%) <sup>a</sup>	Média (%)	DPR (%)
90,4	90,9	91,4	91,2	0,4
100,3	102,2	101,5	101,9	0,5
110,2	110,2	110,4	110,3	0,2

<sup>a</sup>Média de três leituras

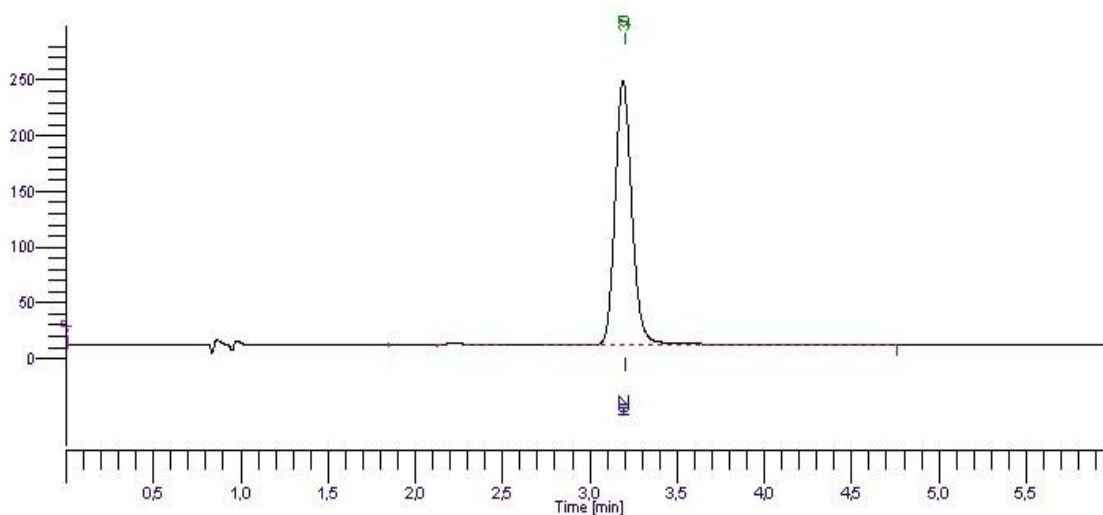
## 5.6. Análises de Amostras Reais

Inicialmente, o método foi aplicado à determinação de HCTZ em amostras na forma de pó, obtidas pela trituração de comprimidos prontos de três lotes diferentes. Os resultados foram comparados com os fornecidos pelo método CLAE de referência e são mostrados na Tabela 9; um cromatograma é mostrado na Figura 20. Um teste t não pareado com quatro graus de liberdade mostrou que não há diferenças significativas entre os métodos para os três lotes, no nível de confiança de 95%.

**Tabela 9.** Determinação de HCTZ em comprimidos triturados a pó fino pelo método proposto e pelo método oficial CLAE

Lote	Método oficial (%) <sup>a</sup>	Método proposto (%) <sup>a</sup>
A	97,7 ± 2,3	99,0 ± 0,4
B	99,3 ± 1,3	96,8 ± 1,6
C	100,4 ± 2,3	98,9 ± 2,5

<sup>a</sup>Valores médios e desvios padrão de três determinações



**Figura 20.** Cromatograma obtido para uma amostra de HCTZ pelo método oficial (CLAE).

Embora o modelo obtido por espectroscopia NIR tenha sido desenvolvido para a determinação de HCTZ em amostras na forma de pó, ele também foi testado em 10 comprimidos intactos de um mesmo lote. Os resultados desta análise foram comparados com os obtidos pelo método CLAE de referência e são mostrados na Tabela 10. Um teste *t* pareado com 10 graus de liberdade e nível de confiança de 95% mostrou que não há diferença significativa entre os resultados ( $t_{\text{cal}} = 2,094 < t_{\text{tab}} = 2,228$ ). Os erros de previsão individual variaram entre -6,1 e + 8,6% e foi obtido um RMSEP de 4,6%. Estes resultados não são tão bons quanto os obtidos para as

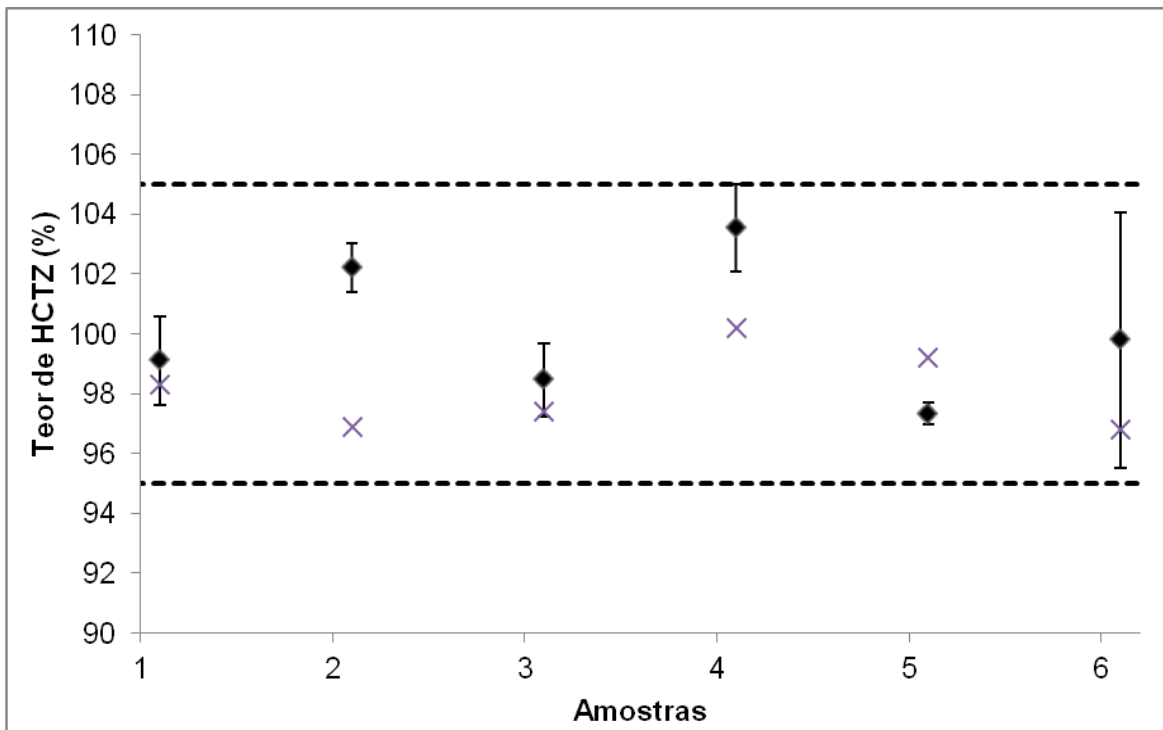
amostras em pó, mas levando em conta que o método foi desenvolvido para amostras em outra forma física, podemos considerá-los bastante razoáveis.

**Tabela 10.** Determinação de HCTZ em dez comprimidos intactos pelo método desenvolvido e pelo método oficial CLAE

Comprimido	Método oficial (%)	Método proposto (%)	Erros de previsão (%)
1	95,1	103,6	8,6
2	98,0	104,2	6,1
3	99,2	103,5	4,3
4	100,1	94,0	-6,1
5	100,3	101,2	0,8
6	98,2	98,0	-0,1
7	99,9	105,2	5,3
8	95,0	97,6	2,6
9	100,8	103,1	2,3
10	99,3	102,0	2,7

Finalmente, o método foi aplicado para a determinação de HCTZ em amostras reais na forma de pó, obtidas durante o processo de produção industrial, após a etapa de mistura e antes da compressão. Nesta situação, o método será de grande utilidade para se avaliar a eficiência do processo de mistura na homogeneização de cada lote. Estas amostras foram consideradas sob controle, com média de conteúdo de HCTZ entre 95,0 e 105,0% (23,75 e 26,25 mg de HCTZ por 150 mg de pó). Estes limites são considerados adequados para controle de qualidade em processo, já que os limites oficiais para liberação do produto acabado são de 90,0 a 110,0%. Esse controle também foi feito por CLAE e um teste de t não pareado mostra que não houve diferença significativa entre os dois métodos  $t_{cal} = 1,75$  e  $t_{tab} = 2,23$ . Portanto, o método desenvolvido foi aplicado na análise de amostras medidas em triplicatas coletadas de seis diferentes lotes durante o período de três meses. Os resultados foram dispostos, para os dois métodos, em um gráfico

similar ao usado para cartas de controle (Figura 21) e, como se pode observar, todos os valores estão dentro dos limites aceitáveis.



**Figura 21.** Resultados (médios e desvios padrão de triplicatas) de determinações de HCTZ em pó de seis amostras coletados após mistura e antes da compressão. As linhas pontilhadas indicam os limites para liberação durante o controle em processo. Resultados da análise por CLAE (x) e por NIR (◆).

## 6. Conclusões

Foi desenvolvido e validado um método por reflectância difusa no infravermelho próximo para determinação direta de HCTZ em formulação farmacêutica na forma de pó. O método desenvolvido possui muitas vantagens em relação ao método oficial (CLAE) e outros métodos alternativos, tais como baixo custo, procedimento simplificado, não utilização de reagentes ou solventes, portanto não gerando resíduos nocivos e, já que é um método não destrutivo, a amostra ainda pode ser reutilizada para outros testes de controle de qualidade.

A principal vantagem do método desenvolvido foi sua rapidez. A Farmacopeia Brasileira, quinta edição, prescreve dois métodos para quantificação de HCTZ, um baseado em espectrofotometria de absorção no ultravioleta (A), o qual para a execução de todas as etapas até se obter o resultado demanda cerca de 56 minutos, e um método por CLAE (B), o qual entre preparação da amostra, injeção no cromatógrafo e obtenção dos resultados leva cerca de 88 minutos. No método NIRS proposto, a amostra precisa apenas ser triturada e lida em um espectrofotômetro NIR, o que juntamente com o tratamento quimiométrico demanda cerca de 6 minutos. Sendo assim, o método proposto é cerca de 9 vezes mais rápido que o método A e 15 vezes mais rápido que o método B.

A região espectral entre 1640 nm e 1780 nm forneceu o melhor modelo PLS, região esta que corresponde ao primeiro sobretom de estiramento de ligações C-H de HCTZ. O método baseado em espectroscopia NIR foi validado de acordo com as normas brasileiras e internacionais [24-26] e foi considerado exato, preciso, linear e sensível no intervalo de 85,0 a 116,0% de HCTZ. Este modelo foi aplicado com sucesso em comprimidos prontos triturados a pó fino e em amostras coletadas na linha de produção após a etapa de mistura, antes da compressão. Além disso, o método também foi aplicado na determinação de HCTZ em comprimidos intactos, fornecendo erros de previsão que podem ser considerados razoáveis. Assim, foi desenvolvida uma metodologia multivariada robusta que empregou o conceito de NAS, e modelos quimiométricos que podem ser aplicados no desenvolvimento de outros métodos baseados em espectroscopia NIR e calibração multivariada com aplicação na indústria farmacêutica e em outras áreas, tais como na análise de alimentos, produtos agrícolas, combustíveis, e em tantas outras que podem ser beneficiadas com esta tecnologia.

## 7. Referências Bibliográficas

- [1] Silva, P.; "Farmacologia", 7<sup>a</sup> ed., Guanabara Koogan S.A., Rio de Janeiro, 2006.
- [2] Hardman, J.G., Limbird, L.E., Gilman A.G.; "Goodman & Gilman's The Pharmacological Basis of Therapeutics", 10<sup>th</sup> ed., McGraw-Hill, New York, 2001.
- [3] Rosa, T.C.C.; "Dissolução intrínseca de hidroclorotiazida de diferentes granulometrias e sua relação com a dissolução do ativo em comprimidos", Dissertação de Mestrado, UFRJ, Rio de Janeiro, 2005.
- [4] Farmacopeia Brasileira, V ed., Agência Nacional de Vigilância Sanitária - ANVISA, Brasília, 2011.
- [5] The United States Pharmacopoeia, 33<sup>rd</sup> rev., U.S.P. Convention, Rockville, 2011.
- [6] Satana, E., Altinay, S., Goger, N.G., Ozkan, S.A., Senturk, Z.; "Simultaneous determination of valsartan and hydrochlorothiazide in tablets by first-derivative ultraviolet spectrophotometry and LC", J. Pharm. Biomed. Anal. 25:1009 (2001).
- [7] Gotardo, M.A., Pezza, L., Pezza, H.R.; "Determination of hydrochlorothiazide in pharmaceutical formulations by diffuse reflectance spectroscopy", Eclet. Quim. 30:17 (2005).
- [8] Ji, J. Xi, X., Zhang, S., Ai, X., He, Z.; "Investigation of RuBPS-Ce(IV) chemiluminescence reaction and its application in determination of two diuretics", Anal. Chim. Acta 541:191 (2005).
- [9] Quaglia, M.G., Donati, E., Carlucci, G., Mazzeo, P., Fanali, S.; "Determination of losartan and hydrochlorothiazide in tablets by CE and CEC", J. Pharm. Biomed. Anal. 29:981 (2002).
- [10] Rezaei, B., Damiri, S.; "Multiwalled carbon nanotubes modified electrode as a sensor for adsorptive stripping voltammetric determination of hydrochlorothiazide", IEEE Sens. J. 8:1523 (2008).
- [11] Karimi-Maleh, H., Ensafi, A.A., Ensafi, H.R.; "Ferrocenedicarboxylic acid modified carbon paste electrode: A sensor for electrocatalytic determination of hydrochlorothiazide", J. Braz. Chem. Soc. 20:880 (2009).
- [12] Pasquini, C.; "Near infrared spectroscopy: Fundamentals, practical aspects and analytical applications", J. Braz. Chim. Soc. 14:198 (2003).



- [13] Sena, M.M., Freitas, C.B., Silva, L.C., Pérez, C.N., Paula, Y.O.; “Determinação espectrofotométrica simultânea de paracetamol e ibuprofeno em formulações farmacêuticas usando calibração multivariada”, *Quim. Nova* 30:75 (2007).
- [14] Arancibia, J.A., Escandar, G.M.; “Two different strategies for the fluorimetric determination of piroxicam in serum”, *Talanta* 60:1113 (2003).
- [15] Sena, M.M., Poppi, R.J.; “N-way PLS applied to simultaneous spectrophotometric determination of acetylsalicylic acid, paracetamol and caffeine”, *J. Pharm. Biomed. Anal.* 34:27 (2004).
- [16] Braga, J.W.B.; “Avaliação de figuras de mérito em calibração multivariada, aplicada na determinação de carbamazepina por espectroscopia no infravermelho próximo e médio”, *Dissertação de Mestrado, UNICAMP, Campinas, 2004.*
- [17] Skoog, D.A., Holler, F.J., Nieman, T.A.; “Princípios de Análise Instrumental”, 5ª ed., Bookman, Porto Alegre, 2002, pp. 342.
- [18] Wetzel D.L.; “Near-infrared reflectance analysis. Sleeper among spectroscopic techniques”, *Anal. Chem.* 55:1165A (1983).
- [19] Ferreira, M.H., Gomes, J.F.F., Sena, M.M.; “Development and validation of multivariate calibration model for determination of dipyron in oral solutions by near infrared spectroscopy”, *J. Braz. Chem. Soc.* 20:1680 (2009).
- [20] Blanco, M., Eustáquio, A., González, J.M., Serrano, D.; “Identification and quantitation assays for intact tablets of two related pharmaceutical preparations by reflectance near-infrared spectroscopy: validation of the procedure”, *J. Pharm. Biomed. Anal.* 22:139 (2000).
- [21] Baratieri, S.C., Barbosa, J.M., Freitas, M.P., Martins, J.A.; “Multivariate analysis of nystatin and metronidazole in a semi-solid matrix by means of diffuse reflectance NIR spectroscopy and PLS regression”, *J. Pharm. Biomed. Anal.* 40:51 (2006).
- [22] Li, W., Worosila, G.D.; “Quantitation of active pharmaceutical ingredients and excipients in powder blends using designed multivariate calibration models by near-infrared spectroscopy”, *Int. J. Pharm.* 295:213 (2005).
- [23] Silva, M.A.M.; “Desenvolvimento e Validação de Método Quimiométrico para determinação de Amoxicilina em formulação farmacêutica”, *Dissertação de Mestrado, UEG, Anápolis, 2011.*

- [24] Agência Nacional de Vigilância Sanitária (ANVISA); “Guia para Validação de Métodos Analíticos e Bioanalíticos”, Resolução – RE nº 899, Brasil, 2003.
- [25] International Conference on Harmonization, Tripartite Guideline – Q2A Text on Validation of Analytical Procedures, Fed. Regist., 60 FR 11260, USA, 1995.
- [26] International Conference on Harmonization, Tripartite Guideline – Q2B Validation of Analytical Procedures: Methodology, Fed. Regist., 62 FR 27464, USA, 1997.
- [27] Braga, J.W.B., Poppi, R.J.; “Validação de modelos de calibração multivariada: uma aplicação na determinação de pureza polimórfica de carbamazepina por espectroscopia no infravermelho próximo”, Quim. Nova 27:1004 (2004).
- [28] Lorber, A., Faber, K., Kowalski, B.R.; “Net Analyte Signal Calculation in Multivariate Calibration”, Anal. Chem. 69:1620 (1997).
- [29] Pavia, D.L., Lampman, G.M., Kriz, G.S., Vyvyan, J.R.; Introdução à Espectroscopia, 4ª ed., Cengage Learning, São Paulo, 2010.
- [30] Rocha, W.F.C.; “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.
- [31] Nunes, P.G.A.; “Uma nova técnica para seleção de variáveis em calibração multivariada aplicada às espectroscopias UV-Vis e NIR”, Tese de Doutorado, UFPB, João Pessoa, 2008.
- [32] Bruns, R.E., Faigle, J.F.G.; “Quimiometria”, Quím. Nova 8:84 (1985).
- [33] 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.
- [34] Wise, B.M., Gallagher, N.B., Bro, R., Shaver, J.M., Windig, W., Koch, R.S.; “Chemometrics Tutorial for PLS Toolbox and Solo”, Eigenvector Research, Wenatchee, 2006.
- [35] Ferreira, M.C., Antunes, A.M., Melgo, M.S., Volpe, P.L.O.; “Quimiometria I: calibração multivariada, um tutorial”, Quím. Nova 22:724 (1999).

- [36] Annual Book of ASTM Standards, Standard Practices for Infrared Multivariate Quantitative Analysis - E1655-05, ASTM International, West Conshohocken: Pennsylvania, USA, 2005.
- [37] Lorber, A.; "Error propagation and figures of merit for quantification by solving matrix equation", *Anal. Chem.* 58:1167(1986).
- [38] 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).
- [39] Braga, J.W.B.; "Aplicação e validação de modelos de calibração de segunda ordem em química analítica", Tese de Doutorado, UNICAMP, Campinas, 2008.
- [40] 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.
- [41] Ferre, J., Brown, S.D., Rius, F.X.; "Improved calculation of the net analyte signal in inverse multivariate calibration", *J. Chemom.* 15:537 (2001).
- [42] Faber, N.M.; "Exact presentation of multivariate calibration model as univariate calibration graph", *Chemom. Intell. Lab. Sys.* 50:114 (2000).
- [43] Bro, R., Andersen, C.M.; "Theory of net analyte signal vectors in inverse regression" *J Chemom.* 17:646 (2003)
- [44] Valderrama, P., Braga, J.W.B., Poppi, R.; "Estado da arte em calibração multivariada", *Quím. Nova* 32:1278 (2009).
- [45] 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).
- [46] Beebe, K.R., Kowalski, B.R.; "An Introduction to multivariate calibration and analysis", *Anal. Chem.* 59:1007 (1987).
- [47] 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).

- [48] 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).
- [49] Currie, L.A.; "Nomenclature in evaluation of analytical methods including detection and quantification capabilities (IUPAC Recommendations 1995)", *Anal. Chim. Acta* 391:105 (1999).
- [50] ISO 11843-2, Capability of detection, International Standards Organization, Geneva, Switzerland, 2000
- [51] Ortiz, M.C., Sarabia, L.A., Herrero, A., Sánchez, M.S., Sanz, M.B., Rueda, M.E., Giménez, D., Meléndez, M.E.; "Capability of detection of an analytical method evaluating false positive and false negative (ISO 11843) with partial least squares", *Chemom. Intell. Lab. Syst.* 69: 21 (2003).
- [52] Carneiro, H.S.P., Medeiros, A.R.B., Oliveira, F.C.C., Aguiar, G.H.M., Rubim, J.C., Suarez, P.A.Z.; "Determination of ethanol fuel adulteration by methanol using partial least-squares models based on Fourier transform techniques", *Energy Fuels* 22:2767 (2008).
- [53] Forouzangohar, M., Kookana, R.S., Forrester, S.T., Smernik, R.J., Chittleborough, D.J.; "Midinfrared spectroscopy and chemometrics to predict diuron sorption coefficients in soils", *Environ. Sci. Technol.* 42: 3283 (2008).
- [54] Igne, B., Hurburgh Jr., C.R.; "Local chemometrics for samples and variables: optimizing calibration and standardization processes", *J. Chemom.* 24: 75 (2010).
- [55] Williams, P.; "Near-infrared technology in the agricultural and food industries", 2<sup>nd</sup> ed., American Association of Cereal Chemists Inc., St. Paul, EUA, 2001.
- [56] Martens, H., Naes, T.; "Multivariate Calibration", John Wiley & Sons, New York, EUA, 1989.
- [57] Walczak, B., Massart, D.L.; "Multiple outlier detection revisited", *Chemom. Intell. Lab. Syst.* 41:1 (1998).
- [58] 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).
- [59] Workman Jr., J.J., Weyer, L.; "Practical Guide to Interpretive Near-Infrared Spectroscopy" CRC – Press, Boca Raton, EUA, 2008.

[60] Rinnan, A., van den Berg, F., Engelsen, S.B.; “Review of the most common pre-processing techniques for near-infrared spectra”, *TrAC – Trends Anal. Chem.* 28:1201 (2009).

[61] Huber, L.; “Validation and Qualification in Analytical Laboratories”, 2<sup>nd</sup> ed., *LabCompliance*, New York, EUA, 2007.