



**UNIVERSIDADE ESTADUAL DA PARAÍBA
CAMPUS I – CAMPINA GRANDE
CENTRO DE CIÊNCIAS E TECNOLOGIA
CURSO DE QUÍMICA INDUSTRIAL**

PEDRO QUEIROZ GUIMARÃES

**COMPARAÇÃO ENTRE TÉCNICAS DE CALIBRAÇÃO MULTIVARIADA
PARA A QUANTIFICAÇÃO DE ZIDOVUDINA EM COMPRIMIDOS
UTILIZANDO ESPECTROSCOPIA NIR**

CAMPINA GRANDE – PB

2014

PEDRO QUEIROZ GUIMARÃES

**COMPARAÇÃO ENTRE TÉCNICAS DE CALIBRAÇÃO MULTIVARIADA
PARA A QUANTIFICAÇÃO DE ZIDOVUDINA EM COMPRIMIDOS
UTILIZANDO ESPECTROSCOPIA NIR**

Monografia apresentada a Universidade Estadual da Paraíba, como requisito para a obtenção do título de Bacharel em Química Industrial.

Orientadora: Prof.^a Dra. Simone da Silva Simões

CAMPINA GRANDE – PB
2014

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

G963c Guimarães, Pedro Queiroz.

Comparação entre técnicas de calibração multivariada para a quantificação de zidovudina em comprimidos utilizando espectroscopia NIR [manuscrito] / Pedro Queiroz Guimarães. - 2014.

53 p. : il. color.

Digitado.

Trabalho de Conclusão de Curso (Graduação em Química Industrial) - Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2014.

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

1. Espectroscopia NIR. 2. Fármacos anti-HIV. 3. Quimioterapia. 4. AZT. I. Título.

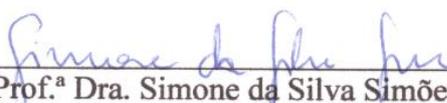
21. ed. CDD 616.979 2

PEDRO QUEIROZ GUIMARÃES

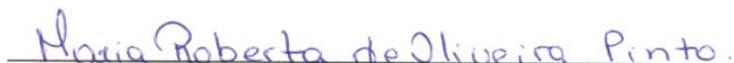
**COMPARAÇÃO ENTRE TÉCNICAS DE CALIBRAÇÃO MULTIVARIADA
PARA A QUANTIFICAÇÃO DE ZIDOVUDINA EM COMPRIMIDOS
UTILIZANDO ESPECTROSCOPIA NIR**

Monografia apresentada a Universidade Estadual da Paraíba, como requisito para a obtenção do título de Bacharel em Química Industrial.

Aprovada em 05 de junho de 2014.


Prof.ª Dra. Simone da Silva Simões/ UEPB
Orientadora


Prof. Dr. José Germano Vêras Neto / UEPB
Examinador


Prof.ª Dra. Maria Roberta Oliveira Pinto / UEPB
Examinadora

DEDICATÓRIA

*Dedico este trabalho aos meus queridos pais, irmãos e esposa.
Por acreditar que a família será sempre o meu porto seguro e também
Pela confiança e esperança depositados em mim.*

AGRADECIMENTOS

Ao Deus, Todo Poderoso, pelas maravilhas que tem feito em minha vida e por me fazer persistir, mesmo nos momentos mais difíceis.

A estimável pesquisadora e professora da Universidade Estadual da Paraíba, Simone da Silva Simões, pelo apoio, estímulo, atenção, confiança e paciência que sempre teve para comigo durante a orientação deste trabalho.

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

A minha adorável esposa Maria Geiza, pelo seu carinho, companheirismo e apoio incondicional.

Ao Laboratório Farmacêutico de Pernambuco, em especial ao pesquisador Severino Granjeiro Júnior, pela preparação em planta piloto dos comprimidos dose fixa combinada de AZT+3TC, cessão dos comprimidos comerciais e dados espectrais para o desenvolvimento deste trabalho.

A Universidade Estadual da Paraíba, em especial, a todos os professores que contribuíram ao longo do curso para a minha formação.

Aos caros colegas de turma, Bráulio, Amanda, Edijane e Giáconys, pela enorme colaboração recebida ao longo de todo o curso e, principalmente, pela amizade construída durante todo esse tempo.

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

Ver bem não é ver tudo: é ver o que os outros não veem.

José Américo de Almeida.

RESUMO

Neste trabalho foi estudada a potencialidade da espectroscopia no infravermelho próximo (NIR) como metodologia analítica para o controle de teor do princípio ativo zidovudina (AZT) em comprimidos dose fixa combinada (AZT+3TC). Juntamente à espectroscopia NIR, foi empregada a calibração multivariada através dos métodos (PLS e MLR) visando o desenvolvimento de modelos de regressão. Para a construção dos modelos foram utilizadas 33 amostras produzidas na planta piloto do Laboratório Farmacêutico de Pernambuco (LAFEPE), de modo a ampliar a faixa de concentração dos comprimidos. Estas amostras foram desenvolvidas no LAFEPE a partir de um planejamento fatorial experimental e cedidas para este trabalho. Cada uma das misturas do planejamento experimental possuía uma massa média de 300 mg e foi preparada através da mistura dos princípios ativos e dos excipientes. Deste modo, as massas de AZT e 3TC foram variadas dentro de uma faixa de $\pm 15\%$ do valor rotulado. Com isso, a concentração dos comprimidos fabricados na planta piloto variou de 255 a 345 mg. Também foram coletados comprimidos de 29 lotes distintos produzidos na linha de produção do LAFEPE. Totalizando 62 amostras. A determinação do teor de AZT pelo método de referência foi realizada no mesmo laboratório, por meio de análises cromatográficas e de acordo com o procedimento descrito na Farmacopeia Brasileira. Os espectros das amostras foram registrados um espectrômetro FTIR da marca BOMEM, na faixa de 14.000 a 4.000 cm^{-1} , utilizando 32 varreduras e resolução de 8 cm^{-1} . O programa UNSCRAMBLER[®]v9.8 da CAMO foi empregado para pré-processamento dos espectros e construção dos modelos PLS. Os algoritmos MLR/APS e SP-Xy foram executados empregando o ambiente MATLAB[®], versão 6.5. Os algoritmos APS e SP-Xy foram usados, respectivamente, para a seleção de variáveis espectrais nos modelos MLR e para seleção de amostras do conjunto de calibração. Para o pré-tratamento dos dados espectrais, foram testados os seguintes pré-tratamentos: SNV, MSC, derivada e suavização, e as combinações destes. A verificação da presença de amostras anômalas (*outliers*), antes da construção dos modelos multivariados, foi realizada através de um estudo do gráfico de influência versus resíduo na matriz X e no vetor y. Na sequência, com o auxílio do algoritmo SP-Xy foram selecionadas 42 amostras para o conjunto de calibração e 20 para o conjunto de predição. A espectroscopia NIR mostrou-se uma técnica adequada para o uso no controle de qualidade de medicamentos, no tocante a metodologia desenvolvida para a determinação do teor de AZT em comprimidos. A melhor técnica de pré-processamento espectral e foi a combinação da suavização utilizando o polinômio Savitzky-Golay com polinômio de 2ª ordem e janela de 11 pontos e MSC. Em relação aos métodos de regressão avaliados, podemos afirmar de acordo com o teste F ($F_{\text{cal}} = 1,07 < F_{0,05;20;20} = 2,12$) que não há diferença estatística entre os mesmos. Ambos os modelos apresentaram erros dentro dos limites estabelecidos para indústria farmacêutica, ou seja, 10% do teor. Deste modo, podem ser utilizados no controle de qualidade do teor de AZT em comprimidos dose fixa combinada (AZT+3TC).

PALAVRAS-CHAVE: Zidovudina. Espectroscopia NIR. Calibração multivariada. PLS. MLR.

ABSTRACT

In this work the potential of near infrared spectroscopy (NIR) was studied as an analytical method for controlling content of active ingredient zidovudine (AZT) in tablets of fixed-dose combination (AZT +3 TC). NIR spectroscopy was used with multivariate calibration methods (PLS and MLR) for the development of regression models. In development of the models, 33 samples were produced in the Pharmaceutical Laboratory of Pernambuco (LAFEPE) in order to expand the concentrations range of the tablets. These samples were developed according to an experimental factorial design and donated to this job. Each experimental mixture had an average weight of 300 mg and was prepared by mixing the active ingredients and the excipients. Thus, the masses of AZT and 3TC were varied within a range of $\pm 15\%$ of the labeled value. The concentration of the tablets manufactured in the pilot plant ranged 255-345 mg. 29 tablets of different batches produced on production line of LAFEPE were also collected, totaling 62 samples. The determination of AZT by the reference method was performed in the LAFEPE by chromatographic analysis and according to the procedure described in the Brazilian Pharmacopoeia. The spectra of the samples were recorded on FTIR spectrometer from Bomem brand, in the range 14000-4000 cm^{-1} , using 32 scans and a resolution of 8 cm^{-1} . The Unscrambler® v9.8 from CAMO was used for preprocessing the spectra and construction of the PLS models. MLR/APS and SP-Xy algorithms were run using the MATLAB® environment, version 6.5. The APS and SP-Xy algorithms were used, respectively, for the selection of spectral variables in the MLR models and selection of samples for the calibration and prediction set. For the pre-treatment of the spectral data, the following pre-treatments were tested: SNV, MSC, derived, smoothing and combinations of these were tested. Checking for the presence of anomalous samples (outliers) before the construction of the multivariate models was performed through a study of the plot of influence versus residual in the matrix X and in the vector y. In the following, with the aid of SP-Xy algorithm 42 samples were selected for the calibration set and 20 for the set of prediction. NIR spectroscopy proved to be adequate for use in the quality control of medicines, regarding the methodology developed for the determination of AZT tablets technical content. The best technique of spectral pre-processing was a combination of smoothing using the polynomial Savitzky-Golay with 2nd order polynomial, 11 points window with MSC. Regarding regression methods evaluated, we can say according to the F test ($F_{\text{cal}} = 1,07 < F_{0,05;20;20} = 2,12$) that there is no statistical difference between them. Both models had errors within the limits established for the pharmaceutical industry, ie 10% of the content. Thus, they can be used in the quality control of the content of AZT in fixed-dose combination tablets (AZT +3 TC).

KEY-WORD: Zidovudine. NIR Spectroscopy. Multivariate Calibration. PLS. MLR.

LISTA DE ILUSTRAÇÕES

Figura 1 - Estruturas moleculares planas da AZT (a) e 3TC (b)	13
Figura 2 - Funções de potenciais harmônicos para um oscilador diatômico	17
Figura 3 - Funções de potenciais harmônicos (---) e anarmônicos (—) para um oscilador diatômico	19
Figura 4 - Transição entre os níveis vibracionais	19
Figura 5 - Tipos de medidas realizados na região NIR.....	22
Figura 6 - Organização dos dados para calibração multivariada.....	31
Figura 7 - Representação dos <i>scores</i> (a) e <i>loadings</i> (b)	34
Figura 8 - Representação genérica do método PLS	36
Figura 9 - Espectros no infravermelho próximo, não tratados, obtidos para os comprimidos da dose fixa combinada de AZT+3TC	41
Figura 10 - Espectros pré-tratados, utilizando a suavização com o algoritmo Savitzky-Golay, polinômio de 2ª ordem e janela de 11 pontos combinado ao MSC	42
Figura 11 - Gráfico de influência versus resíduo.....	43
Figura 12a - Gráfico da influência versus resíduos para os dados estudados em X	43
Figura 12b - Gráfico da influência versus resíduos para os dados estudados em y.....	44
Figura 13 - Gráfico de variância residual versus número de variáveis latentes para os dados sem pré-tratamento.....	45
Figura 14 - Curva de calibração para os dados de refletância sem pré-tratamento	45
Figura 15 - Gráfico dos valores de referência versus valores previstos pelo modelo construído com os dados sem pré-tratamento.....	46
Figura 16 - Gráfico de variância residual versus número de variáveis latentes para os dados pré-tratados.....	47
Figura 17 - Curva de calibração para os dados de refletância pré-tratados.....	47
Figura 18 - Gráfico dos valores de referência versus valores previstos pelo modelo construído com os dados pré-tratados.....	48
Figura 19 - Gráfico de número de variáveis versus RMSE para os dados pré-tratados	49
Figura 20 - Variáveis selecionadas pelo algoritmo da projeções sucessivas (APS).....	50
Figura 21 - Gráfico dos valores de referência versus valores previstos pelo modelo construído com os dados pré-tratados.....	50

LISTA DE SIGLAS

AIDS	<i>Acquired Immunodeficiency Syndrome</i>
ANVISA	Agencia Nacional de Vigilância Sanitária
ASTM	<i>American Society for Testing and Materials</i>
BPF	Boas Práticas de Fabricação
CV	<i>Cross Validation</i>
DFC	Dose Fixa Combinada
EMA	<i>European Medicines Agency</i>
FDA	<i>Food and Drug Administration</i>
FTIR	<i>Fourier Transform Infrared</i>
HIV	<i>Human Immuno deficiency Virus</i>
GRAMS	<i>graphic relational - array management system</i>
HPLC	<i>High Performance Liquid Chromatography</i>
MLR	<i>Multiple Linear Regression</i>
MSC	<i>Multiplicative Scatter Correction</i>
NIPALS	<i>Non-linear Iterative Partial Least Squares</i>
NIRS	<i>Near Infrared Spectrometry</i>
OMS	Organização Mundial de Saúde
GASP	<i>Group of Analytical Sciences and Pharmaceutical</i>
PC	<i>Principal Components</i>
PCA	<i>Principal Components Analysis</i>
PCR	<i>Principal Component Regression</i>
PLS	<i>Partial Least Squares</i>
SIDA	Síndrome da Imunodeficiência Adquirida
SNV	<i>Standard Normal Deviation</i>

SUMÁRIO

1	INTRODUÇÃO	9
1.1	Objetivos	11
1.1.1	Objetivo geral.....	11
1.1.2	Objetivos específicos.....	11
2	FUNDAMENTAÇÃO TEÓRICA	12
2.1	Fármacos anti-HIV	12
2.2	Controle de qualidade de fármacos	13
2.3	Espectroscopia no infravermelho próximo	15
2.4	Instrumentação	21
2.5	Tipos de medidas na espectroscopia NIR	22
2.6	Transformada de Fourier	25
2.7	Espectroscopia NIR aplicada à indústria farmacêutica	26
2.8	Quimiometria	27
2.9	Análise multivariada	28
2.10	Calibração multivariada	28
2.10.1	Regressão por componentes principais (PCR).....	32
2.10.1.1	Análise de componentes principais (PCA).....	32
2.10.2	Regressão por mínimos quadrados parciais (PLS).....	35
2.10.3	Regressão linear múltipla (MLR).....	36
2.11	Cálculo dos erros	37
2.12	Coefficiente de correlação	38
3	PARTE EXPERIMENTAL	39
3.1	Amostras	39
3.2	Determinação do teor de AZT pelo método de referência	39
3.3	Aquisição dos espectros NIR	40
3.4	Softwares	40
4	RESULTADOS E DISCUSSÃO	41
4.1	Espectros	41
4.2	Construção dos modelos multivariados	42
4.2.1	Construção dos modelos PLS para a determinação do teor de AZT.....	44
4.2.2	Construção dos modelos MLR para a determinação do teor de AZT	49
5	CONCLUSÕES	52
	REFERÊNCIAS	53

1 INTRODUÇÃO

Em 1964, o pesquisador Horwitz da Fundação contra o Câncer de Michigan (EUA) sintetizou pela primeira vez a zidovudina (3'-azido-2',3'-didesoxitimidina), também conhecido por (AZT), que a princípio destacou-se por apresentar atividade anticancerígena. Não demorou muito e no início da década de 70 sua atividade antiviral foi descrita pela primeira vez. Em 1985, o AZT assumiu um papel de destaque sendo o primeiro composto a apresentar atividade anti-HIV. O que o levou, no ano seguinte, a ser aprovado pelo órgão norte-americano de controle sobre produtos farmacêuticos (do inglês: *Food and Drug Administration* - FDA) e comercializado pelos laboratórios Burroughs-Wellcome em 1987. Com a descoberta do AZT, inúmeras substâncias foram sintetizadas, obtendo-se novos e potentes fármacos no combate à replicação do vírus HIV (SOUZA, M.; ALMEIDA, 2003; SOUZA, M., 2005).

Estudos clínicos e imunológicos demonstram que a sua introdução no tratamento aumentou significativamente o tempo de sobrevivência de pacientes infectados pelo vírus HIV, quando comparado a outros. No entanto, o uso prolongado apresentou efeitos tóxicos, resistência a droga e falha terapêutica sendo necessária a associação com outros antirretrovirais (RANDAU et al., 2005).

As combinações de nucleosídeos análogos que mais demonstraram benefícios clínicos são zidovudina/didanosina e zidovudina/zalcitabina. Porém, a combinação que apresentou maior grau de tolerância, bem como, potência antirretroviral (HIV-1 e HIV-2) comparável a essas duas foi a combinação zidovudina/lamivudina (BECK et al., 2007; GRANGEIRO JÚNIOR, 2011).

A lamivudina (3TC) foi sintetizada na perspectiva de aumentar a segurança durante o longo prazo da terapia, modificando-se a estrutura da molécula 2'3'-didesoxicitidina (análogo sintético da citidina) com a substituição do metileno por um átomo de enxofre no carbono 3' do anel da ribose (BECK et al., 2007; GRANGEIRO JÚNIOR, 2011).

A associação entre dois ou mais fármacos é um recurso terapêutico bastante comum, utilizado na maioria das vezes com o objetivo de potencializar a ação esperada do medicamento. Vários exemplos de associações de fármacos, hormônios, vitaminas, antibióticos, anti-hipertensivo entre outros estão descritos na literatura. Nesse contexto, combinações de medicamentos, como é o caso da dose fixa combinada (AZT+3TC), são cada vez mais empregados nos processos de quimioterapia que combatem a infecção do vírus HIV. O termo dose fixa combinada (DFC) refere-se à associação de classes variadas de fármacos de

diversas combinações, sendo o regime terapêutico indicado pela Organização Mundial de Saúde (OMS) (GRANGEIRO JÚNIOR, 2011; SOUZA, M., 2005).

O emprego terapêutico e a comercialização da associação da lamivudina e zidovudina estão bem estabelecidos, porém não há monografia oficial para a determinação dessa associação de fármacos em comprimidos ou outras formas farmacêuticas (SOUZA, M., 2005). Isso requer um grande esforço por parte dos laboratórios farmacêuticos produtores no sentido de adaptar ou até mesmo desenvolver metodologias capazes de acompanhar as etapas de produção, garantido o fornecimento de um produto de qualidade aos seus consumidores (GRANGEIRO JÚNIOR, 2011).

A metodologia analítica para quantificação do AZT existente em compêndios oficiais fundamenta-se em cromatografia líquida de alta eficiência (do inglês: *High Performance Liquid Chromatography* - HPLC), o que representa alto custo e tempo para a indústria farmacêutica (RANDAU et al., 2005).

Esse método, geralmente é lento e bastante trabalhoso, pois envolve um grande número de manipulações analíticas como: extração e dissolução da amostra, preparação constante de reagentes, padronização de soluções destinadas à calibração, necessita de pessoal qualificado, requer o uso de solventes e reagentes caros, entre outros. Além disso, os métodos baseados HPLC, devido à grande quantidade de resíduos que costumam produzir, podem ser considerados bastante prejudiciais ao meio ambiente (SIMÕES, 2008).

Uma boa alternativa para contornar os problemas apresentados pelas técnicas convencionais de HPLC tem sido a utilização da espectrofotometria de absorção na região do infravermelho próximo (do inglês: *Near Infrared Spectrometry* - NIRS). Uma característica fundamental da NIRS, comumente conhecida apenas como NIR, é que ela é uma técnica não destrutiva, o que a torna uma das técnicas de monitoramento e controle de qualidade em processo com maior potencial na indústria farmacêutica (GRANGEIRO JÚNIOR, 2011; SIMÕES, 2008).

No entanto, a espectrometria NIR apresenta desvantagens como bandas largas e sobrepostas, sendo necessária a utilização de métodos quimiométricos como os de calibração multivariada para eliminação desses problemas. Desse modo, a associação da NIR a técnicas de calibração multivariada vem permitindo a realização análises de forma rápida, não destrutiva, de baixo custo e com excelente confiabilidade (SIMÕES, 2008).

Nesse contexto, o presente trabalho pretende desenvolver metodologias baseadas em espectroscopia NIR e quimiometria para prever a concentração do fármaco zidovudina em

comprimidos. Para tanto, uma comparação entre dois tipos de modelos de regressão (PLS e MLR) será realizada.

1.1 Objetivos

1.1.1 Objetivo geral

Comparar a capacidade preditiva de duas técnicas de calibração multivariada, a regressão em mínimos quadrados parciais (PLS) e a regressão linear múltipla (MLR), para o controle de teor do princípio ativo zidovudina (AZT) em comprimidos dose fixa combinada (AZT+3TC).

1.1.2 Objetivos específicos

- Verificar qual dos métodos de pré-tratamento espectral é mais adequado para remoção de informação espúria dos espectros dos comprimidos dose fixa combinada (AZT+3TC).
- Avaliar a capacidade preditiva dos modelos construídos.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 Fármacos anti-HIV

Antirretrovirais são os medicamentos utilizados no combate ao Vírus da Imunodeficiência Humana (do inglês: *Human Immuno deficiency Vírus* - HIV). A função de tais medicamentos é inibir a replicação desse vírus, mantendo-o sob controle pelo maior tempo possível (SOUZA, M., 2003).

Desde 1981, quando foi descoberta da Síndrome da Imunodeficiência Adquirida (do inglês: *Acquired Immuno deficiency Syndrome* - AIDS), tem-se alcançado considerável progresso no desenvolvimento de agentes anti-HIV (BECK et al., 2007; SOUZA, M., 2005). A quimioterapia atual utiliza a combinação de medicamentos, visando o combate à infecção do HIV pela inibição das enzimas (transcriptase reversa e protease), necessárias à replicação viral, após a invasão da célula hospedeira. Com base nessa informação existem várias drogas nas classes: inibidores análogo e não análogo de nucleotídeo e inibidores de protease que compõem o coquetel utilizado no tratamento da AIDS. Dentre estas drogas, o AZT assume um papel de destaque podendo ser associada tanto ao 3TC quanto a didanosina (RANDAU et al., 2005; SOUZA, M., 2005).

O 3TC, que também é um análogo dos dideoxynucleosídeos, sofre fosforilação intracelular, dando origem ao metabólito ativo denominado trifosfato de lamivudina, que assim como o AZT, age inibindo a atividade da transcriptase reversa. Desse modo, o 3TC atua induzindo a resistência viral a ele próprio, restituindo a sensibilidade em pacientes que apresentam processo de resistência associada à zidovudina por meio de mutação, enfraquecendo a virulência e a capacidade de replicação viral do HIV-1, com a diminuição da diversificação genética de sua enzima (GRANGEIRO JÚNIOR, 2011; NASCIMENTO, et al., 2004).

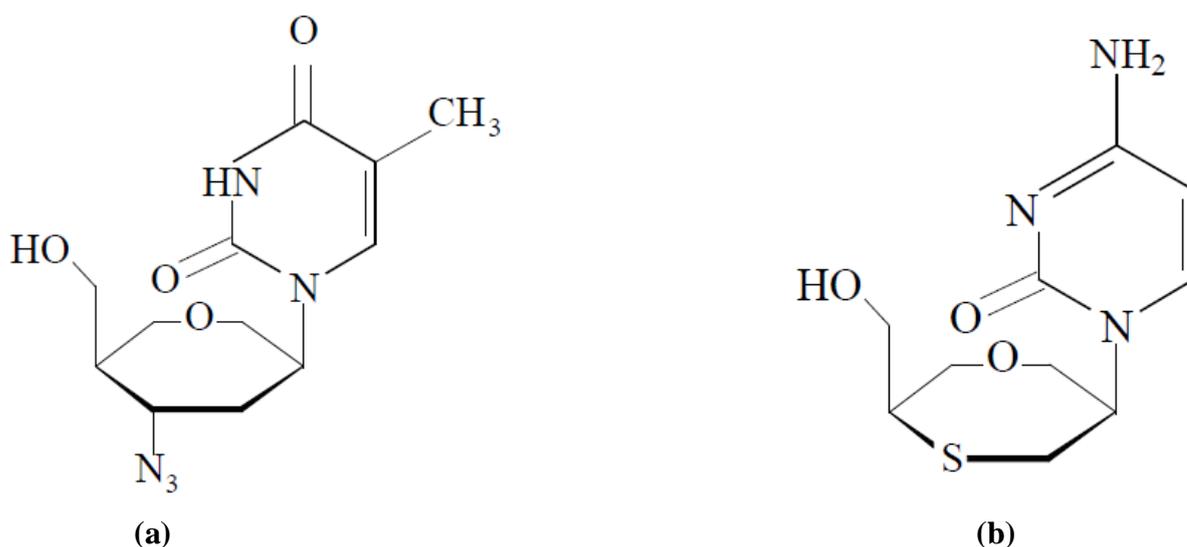
A terapia combinada (AZT+3TC) tem demonstrado maior eficácia em relação à monoterapia que emprega apenas o AZT. Estudos comprovam que a associação entre esses fármacos mostrou sinergismo *in vitro* através da inibição a replicação do HIV em culturas celulares (GRANGEIRO JÚNIOR, 2011; NASCIMENTO, et al., 2004).

Além disso, essa DFC assim como todas as outras recomendadas pela OMS apresenta uma série de vantagens como: o baixo custo, eficácia e tolerabilidade semelhante a outros esquemas terapêuticos, melhor adesão ao tratamento e menor número de comprimidos e/ou cápsulas ingeridos. Onde essa última, é uma das características mais importantes, pois é

considerada como responsável pela ineficácia de grande parte dos tratamentos (GRANGEIRO JÚNIOR, 2011).

O AZT (Figura 1a) se apresenta como pó branco ou acastanhado e tem fórmula química ($C_{10}H_{13}N_5O_4$) e massa molar de 267,2 g/mol. Apresenta solubilidade em etanol, água e tem um ponto de fusão de aproximadamente 124 °C. Já o 3TC (Figura 1b) se apresenta sob a forma de pó cristalino de cor branca ou branco-amarelada, possui fórmula molecular ($C_8H_{11}N_3O_3S$), massa molar de 229,26 g/mol e ponto de fusão na faixa de 176 a 178 °C, sendo solúvel também em etanol e água (GRANGEIRO JÚNIOR, 2011).

Figura 1 - Estruturas moleculares planas do AZT (a) e 3TC (b)



Fonte: GRANGEIRO JÚNIOR (2011)

2.2 Controle de qualidade de fármacos

A consciência da importância do controle de qualidade é histórica. Pois, em 1352, João o Bom (Rei da França), legislou sobre a necessidade de serem controladas as preparações farmacêuticas (SOUZA, J.; FERRÃO, 2006). Atualmente a qualidade é utilizada como uma espécie de selo de garantia para qualquer produto industrial. Nas indústrias farmacêuticas esse controle é essencial tendo em vista a finalidade dos produtos por ela produzidos. Já que alterações na qualidade desses produtos podem causar efeitos prejudiciais à saúde dos consumidores, podendo em alguns casos levar à morte (SIMÕES, 2008).

Segundo a Norma NBR ISO 9000 (2000 apud SIMÕES, 2008), qualidade é a totalidade de características de uma entidade que lhe confere a capacidade de satisfazer as

necessidades explícitas e implícitas. De acordo com essa definição quando se usa a expressão “boa qualidade” pretende-se exprimir a excelência de um produto ou serviço e o cumprimento das especificações, quantitativas ou qualitativas, definidas pelas autoridades competentes. Desse modo, o controle de qualidade consiste em realizar medidas de alguns parâmetros do produto considerado e verificar se os valores obtidos estão de acordo com as especificações pré-estabelecidas para o mesmo (SIMÕES, 2008).

O controle de qualidade na indústria farmacêutica faz parte de um conjunto de ações denominadas de Boas Práticas de Fabricação (BPF), direcionadas no sentido de diminuir os riscos inerentes a cada fase do processo. Elas podem ser divididas em duas etapas bem definidas: o controle de qualidade da formulação durante o processo de produção e o controle de qualidade do produto final. Para essas duas etapas, há exigência, por parte de órgãos reguladores como a Agência Nacional de Vigilância Sanitária (ANVISA), de métodos de análises confiáveis e robustos e que possam ser utilizados rotineiramente para assegurar a qualidade do produto final (GRANGEIRO JÚNIOR, 2011). Além da observação quanto às BPF, é necessário que os medicamentos produzidos sejam submetidos a testes de biodisponibilidade relativa e de bioequivalência farmacêutica antes de serem disponibilizados aos consumidores (SIMÕES, 2008).

As metodologias utilizadas com maior frequência no controle de qualidade de fármacos são: a espectrometria UV-Visível e a cromatografia líquida de alta eficiência (do inglês: *High Performance Liquid Chromatography* - HPLC). A HPLC apesar de ser um método capaz de atingir baixos limites de detecção, fornecer resultados precisos e exatos, apresenta algumas desvantagens como: é dispendiosa, laboriosa, consome grandes quantidades de reagentes, necessita de muitas manipulações analíticas e gera resíduos químicos potencialmente nocivos tanto para o homem quanto para o meio ambiente. A espectrometria UV-Visível, por sua vez, é uma técnica destrutiva. Ambos os métodos, demandam bastante tempo na linha de produção, necessitando também de equipamentos e mão-de-obra especializados (GRANGEIRO JÚNIOR, 2011; SIMÕES, 2008).

Sendo assim, a NIR em conjunto com a técnica de calibração multivariada, tem se apresentado como uma boa alternativa para contornar os problemas apresentados pelas técnicas convencionais.

2.3 Espectroscopia no infravermelho próximo

A espectroscopia é uma técnica instrumental analítica que se baseia nas propriedades de absorção e emissão de energia eletromagnética das moléculas em regiões do espectro eletromagnético. A região do espectro eletromagnético perceptível aos nossos olhos (região visível) está compreendida entre comprimentos de onda de 380 e 780 nm. Acima desse limite até cerca de 50.000 nm, situa-se a região do infravermelho (CALDEIRA; SANTOS; SACON, 201 - ?).

Quando a faixa de comprimento de onda se situa entre (780 e 2.500 nm), tem-se então a região denominada de infravermelho próximo (CABRAL, 2011). Nessa região, as vibrações moleculares que resultam em transições harmônicas (*overtones*) são responsáveis pela absorção (CALDEIRA; SANTOS; SACON, 201 - ?). Essas vibrações moleculares podem ser estiramentos, deformações ou torções. As vibrações de estiramento envolvem uma mudança contínua na distância interatômica ao longo do eixo da ligação entre dois átomos; as vibrações de deformação envolvem mudanças nos ângulos planos entre duas ligações e as torções envolvem mudanças no ângulo diédrico entre dois planos (GRANGEIRO JÚNIOR, 2011).

Os comprimentos de onda nos quais estas vibrações ocorrem para um composto qualquer são funções de sua estrutura e composição. Portanto, o espectro de NIR pode ser utilizado para identificar espécies moleculares em complexas misturas químicas e avaliar sua concentração (CALDEIRA; SANTOS; SACON, 201 - ?).

Na região NIR as bandas de absorção apresentam, de forma geral, intensidades bastante inferiores àquelas observadas no infravermelho convencional (cerca de 10 a 100 vezes). Essa característica, por um lado, pode dificultar até certo ponto a obtenção de medidas. Mas por outro, permite o uso em amostras que possuam alta umidade, como também, proporciona alta penetração em medidas de refletância de amostras sólidas. Essa região espectral dispõe ainda de fontes de alta intensidade e eficientes detectores como forma de compensação do baixo índice de absorção das bandas (CABRAL, 2011).

Apesar das bandas de absorção observadas nessa região não apresentarem fácil compreensão, a utilização de seus espectros de absorbância e refletância difusa vem se mostrando muito eficientes. Diversas metodologias analíticas baseadas nesses espectros demonstram o potencial que esta técnica possui na resolução de um grande número problemas referentes à análise de alimentos, produtos da indústria de polímeros e petróleo, produtos farmacêuticos e na agricultura (CABRAL, 2011; MORGANO et al., 2005).

A espectroscopia NIR surgiu como uma técnica de avaliação não invasiva para materiais orgânicos, sendo amplamente empregada no controle de qualidade e em processos industriais. Essa técnica possui vantagens frente às convencionais, já que é rápida, não invasiva, adequada para o uso em linha de produção e exige pouca manipulação no preparo da amostra. Além disso, pode ser aplicada na análise de qualquer material que possua em suas moléculas, principalmente ligações do tipo (C-H, O-H, N-H, S-H ou C=O) (CABRAL; MORGANO et al., 2011).

Os espectros nessa região podem ser obtidos a partir de amostras de diferentes naturezas: sólidas, líquidas e gasosas. A qualidade de seus espectros depende do método de preparação da amostra e do tipo de acessórios utilizado para a sua obtenção. Os métodos mais empregados para amostras sólidas envolvem reflectância e transmitância (GRANGEIRO JÚNIOR, 2011).

É importante lembrar que, o grande avanço na aplicação da NIR só foi possível devido a associação com métodos quimiométricos de análise. Os quais envolvem, em sua grande maioria, a utilização da regressão por mínimos quadrados parciais (do inglês: *Partial Least Squares* - PLS) e análise de componentes principais (do inglês: *Principal Component Analysis* - PCA) (CABRAL; GRANGEIRO JÚNIOR, 2011; SCAFI, 2000).

2.3.1 Fundamentos de espectroscopia

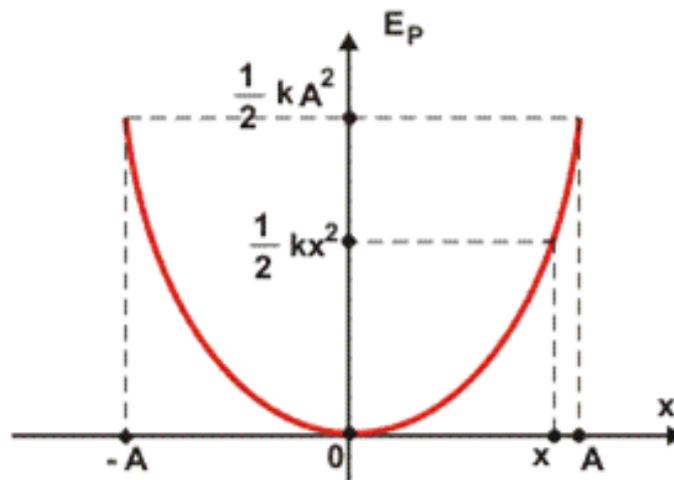
A mecânica quântica, assumindo o modelo do oscilador harmônico, prevê separações iguais entre os níveis vibracionais e impõe a regra de seleção ($\Delta v = \pm 1$). Para oscilações anarmônicas, a separação entre os níveis já não é igualmente espaçada e diminui para energias mais altas. Com isso, a regra de seleção ($\Delta v = \pm 1$) é relaxada e transições com ($\Delta v = \pm 2$) ou ($\Delta v = \pm 3$) dão origem aos sobretons, que são as transições observadas no NIR. As bandas de combinação aparecem entre (1.900 e 2.500 nm) e são resultantes de interferências construtivas de frequências, isto é, suas frequências são a soma dos múltiplos de cada frequência de interação (GRANGEIRO JÚNIOR, 2011; SIMÕES, 2008).

A radiação no infravermelho absorvida por uma molécula faz suas ligações individuais vibrarem com maior amplitude e frequência, de maneira similar a um oscilador harmônico. Nesse sentido, o conceito básico da espectroscopia vibracional se baseia no comportamento de uma molécula diatômica. Para um oscilador harmônico ideal, a energia potencial V é representada por uma equação quadrática simples (CABRAL, 2011; SCAFI, 2000).

$$V = \frac{1}{2}k(r - r_e)^2 = \frac{1}{2}kx^2 \quad \text{Equação (1)}$$

Onde: k é a constante de força de ligação, r é distância internuclear, (r_e) é a distância nuclear de equilíbrio e $(x = r - r_e)$ é à distância de deslocamento. A Figura 2 ilustra como a energia potencial de um oscilador harmônico varia com a amplitude da vibração (SCAFI, 2000).

Figura 2 - Funções de potenciais harmônicos para um oscilador diatômico



Fonte: GOOGLE IMAGENS (2014)

Onde: (E_p) é a energia potencial, (A) é a amplitude máxima quando a mola do oscilador harmônico está completamente distendida. Por outro lado, $(-A)$ é a amplitude máxima quando a mola está completamente comprimida.

A energia potencial assume a forma de uma parábola e tem o comportamento variando em torno da distância internuclear de equilíbrio (r_e) . A frequência de vibração de um oscilador harmônico depende da força de ligação e das massas dos átomos que estão ligados. A energia total da ligação é proporcional à frequência de vibração. O modelo de um oscilador harmônico simples se baseia na Lei de Hooke, onde a frequência vibracional (ν_0) em número de onda é definido da seguinte forma (CABRAL, 2011; SCAFI, 2000):

$$\nu_0 = \frac{1}{2\pi} \sqrt{\frac{k}{\mu}} \quad \text{Equação (2)}$$

Onde: (k) representa a força de ligação; (μ) é a massa reduzida do sistema, sendo definida por:

$$\mu = \frac{m_1 m_2}{m_1 + m_2} \quad \text{Equação (3)}$$

O tratamento da mecânica quântica sobre o modelo simples do oscilador harmônico mostra que o nível de energia vibracional entre os dois átomos de uma molécula é quantizado. Estes níveis de energia são dados por:

$$E_{VIB} = h\nu_0\left(\nu + \frac{1}{2}\right) \quad \text{Equação (4)}$$

Onde: (h) é a constante de Planck, (ν_0) é a frequência vibracional clássica e (ν) é o número quântico vibracional.

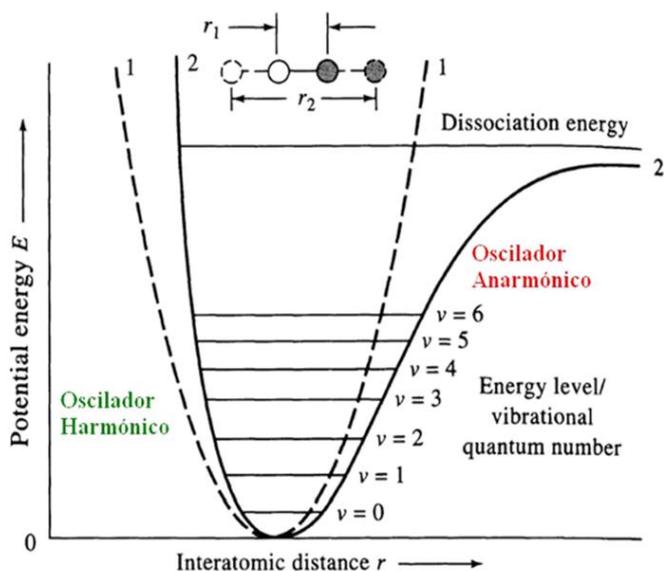
Os níveis de energia podem ser expressos em número de onda (cm^{-1}) de acordo com a expressão:

$$G(\nu)(\text{cm}^{-1}) = \frac{E_{VIB}}{hc} = \nu_0\left(\nu + \frac{1}{2}\right) \quad \text{Equação (5)}$$

Onde: (ν_0) é o número de onda da transição vibracional e (c) é a velocidade da luz no vácuo.

Na temperatura ambiente, ocorre a existência de mais moléculas no nível vibracional fundamental ($\nu = 0$), e conseqüentemente a transição permitida ($\nu = 0 \rightarrow \nu = 1$), denominada transição fundamental, domina o espectro de absorção do infravermelho. Se os níveis energéticos vibracionais se comportassem como previsto para um oscilador harmônico, estes seriam igualmente espaçados, mas na prática os níveis de energia quantizados não são igualmente espaçados, porque as moléculas se desviam da idealidade comportando-se como um oscilador anarmônico. A Figura 3 representa a forma do potencial anarmônico para um sistema diatômico (SCAFI, 2000).

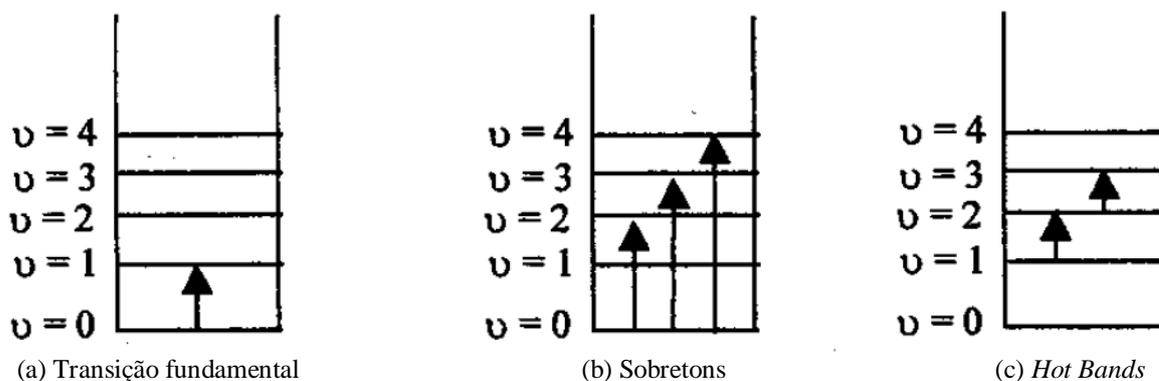
Figura 3 - Funções de potenciais harmônicos (---) e anarmônicos (—) para um oscilador diatômico



Fonte: GOOGLE IMAGENS (2014)

Evidências experimentais mostram que as moléculas se comportam como osciladores anarmônicos, basicamente por dois motivos: primeiro, porque as bandas de transição denominadas Bandas Quentes (do inglês: *Hot Bands*), que são transições originárias de níveis vibracionais excitados ($v \neq 0$), por exemplo ($v = 1 \rightarrow v = 2 \rightarrow v = 3, \dots$), não possuem a mesma frequência de banda fundamental. Isso se deve ao fato de que os níveis de energia vibracionais não se encontram igualmente espaçados. Em segundo lugar, as transições denominadas sobretons como ($v = 1 \rightarrow v = 2, 3, 4, \dots$), são permitidas. Na Figura 4 são descritas todas as transições entre os níveis vibracionais (SCAFI, 2000).

Figura 4 - Transição entre os níveis vibracionais



Fonte: SCAFI (2000)

Com isso, a Equação 5 pode ser reescrita com um termo de anarmonicidade:

$$G(\nu)(cm^{-1}) = \frac{E_{VIB}}{hc} = \nu_0 \left(\nu + \frac{1}{2} \right) - x_e \nu_0 \left(\nu + \frac{1}{2} \right)^2 \quad \text{Equação (6)}$$

Em que, o termo (x_e) é a constante de anarmonicidade.

A absorção de radiação na região NIR por um sistema se deve a variação periódica do seu momento dipolo elétrico (μ), sendo a frequência da radiação absorvida idêntica a oscilação do dipolo. Desse modo, a oscilação do momento dipolo provocada pela frequência da radiação incidente, fará com que a molécula absorva radiação (SCAFI, 2000). Esse momento dipolo pode ser expresso como:

$$\mu = \mu_0 + \left[\left(\frac{d\mu}{dx} \right)_e x \right] \quad \text{Equação (7)}$$

Onde: (μ_0) é vetor do momento dipolo à distância internuclear de equilíbrio e a derivada $\left[\left(\frac{d\mu}{dx} \right)_e x \right]$ é considerada em relação à posição de equilíbrio. A absorção no infravermelho implica $\left[\left(\frac{d\mu}{dx} \right)_e \neq \mathbf{0} \right]$ pelo menos para uma das componentes (μ_x , μ_y ou μ_z) (SCAFI, 2000).

Além disso, há também a chamada anarmonicidade elétrica, que é responsável pelo surgimento no espectro infravermelho, de bandas de sobretons (*overtones*) correspondentes a transições entre os níveis energéticos que diferem de duas ou três unidades para o nível quântico vibracional ($\Delta v = +2, +3, \dots$) (SCAFI, 2000).

$$\mu = \mu_0 + \left(\frac{d\mu}{dx} \right)_e x + \frac{1}{2!} \left(\frac{d^2\mu}{dx^2} \right)_e x^2 + \frac{1}{3!} \left(\frac{d^3\mu}{dx^3} \right)_e x^3 + \dots \quad \text{Equação (8)}$$

Para o surgimento do primeiro sobreton é suficiente se considerar apenas até o termo quadrático. Já para o segundo sobreton, será necessário considerar o termo cúbico. A partir do terceiro sobreton, será necessária a consideração de derivadas de ordem maiores no desenvolvimento da série (SCAFI, 2000).

2.4 Instrumentação

Atualmente, existem diversos tipos de instrumentos que operam na região NIR. No entanto, esses instrumentos apresentam características muito semelhantes àqueles empregados na região do UV-Visível, os quais em geral possuem a seguinte configuração (SKOOG, 2009):

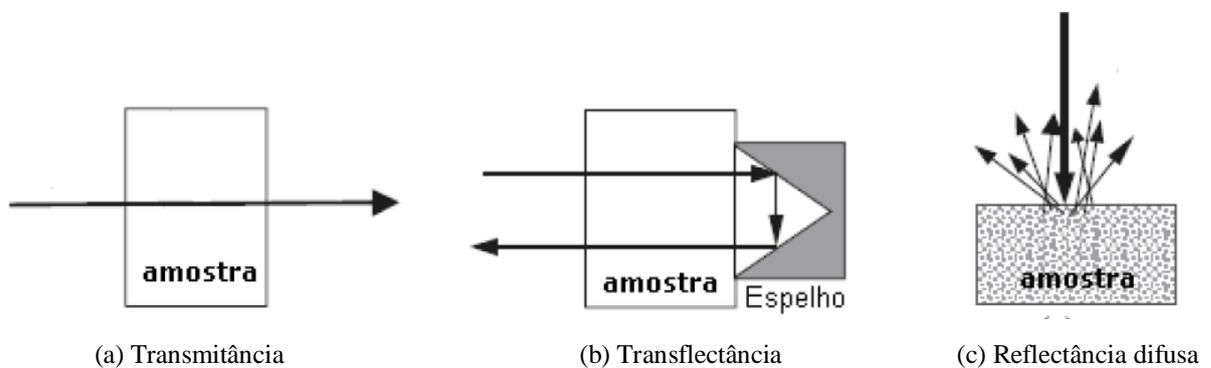
- **Fontes:** é o local de origem de um feixe com potência de radiação contínua, cuja sua função é de emitir radiação. Para a região NIR aplicam-se fontes contínuas como lâmpadas de filamento de tungstênio/halogênio com janelas de quartzo. Essas lâmpadas contêm uma pequena quantidade de iodo dentro do bulbo de quartzo que aloja o filamento. O quartzo permite que o filamento seja operado a temperaturas de cerca de (3.500 K), o que leva à alta intensidade e estende a faixa da lâmpada até a região do ultravioleta. O tempo de vida de uma lâmpada de tungstênio/halogênio é mais que o dobro daquele de uma lâmpada comum de tungstênio porque a vida dessa última é limitada pela sublimação do tungstênio do filamento. Desse modo, a lâmpada de tungstênio/halogênio é especificamente indicada para a região de comprimento de onda compreendida entre (350 e 2.500 nm) que cobre a região do UV-Visível e infravermelho próximo.
- **Seletor de comprimento de onda:** este dispositivo seleciona a radiação em uma faixa de comprimento de onda de interesse, a exemplo de (monocromadores ou filtros).
- **Detector:** é o dispositivo que tem a função de captar a radiação transmitida pelo material analisado e converter em sinal elétrico, através do qual são obtidos os espectros de absorção. Os detectores comumente empregados são os fotocondutores que se baseiam no processo de fotocondução. Neste processo, a radiação promove os elétrons para bandas de condução, ou seja, a condução é baseada no aumento da condutividade. Os fotocondutores mais utilizados no infravermelho próximo são os de sulfeto de chumbo (PbS) que é empregado para uma região de (3.000 a 1.000 nm) e pode ser operado a temperatura ambiente. E arsenato de gálio e índio (InGaAs): esse material semiconductor é depositado sobre uma superfície não condutora e selada em um invólucro sob vácuo com finalidade de impedir a interação entre o material e a atmosfera.

- **Software para NIR:** GRAMS do (inglês: *grafic relational - array management system*), é um conjunto de módulos de programas instalados centrados em torno do programa (*GRAMS/AI*) de processamentos de dados espectroscópicos.
- **Registrador:** é responsável por processar e registrar o sinal emitido pelo detector, sendo a informação obtida através de um espectro.

2.5 Tipos de medidas na espectroscopia NIR

Inicialmente, quando a técnica NIR foi descoberta, acreditava-se que as medidas espectrais podiam ser obtidas através da análise da informação contida em uma porção de radiação refletida de maneira difusa por amostras sólidas através de um fraco sinal de transmitância. Atualmente, a refletância difusa é apenas uma das várias maneiras para se realizar medidas analíticas na região NIR. A Figura 5 representa as formas mais comuns de medidas realizadas na espectroscopia NIR (SIMÕES, 2008).

Figura 5 - Tipos de medidas realizados na região NIR



Fonte: SIMÕES (2008)

Nas medidas de transmitância (Figura 5a), cubetas de vidro ou quartzo com caminhos ópticos variando entre 1 e 50 mm, são empregadas como suporte transparente para colocação das amostras no momento do registro dos espectros NIR. O caminho óptico é, em princípio, definido de acordo com a região espectral que está sendo estudada (SIMÕES, 2008).

A Figura 5b representa um modo especial para realizar medidas de transmitância, conhecida como transflectância. Este modo de medida é utilizado quando feixes de fibra óptica são empregados. A diferença em relação às medidas de transmitância é o caminho

óptico duplo pelo qual o feixe de radiação passa duas vezes através da amostra (SIMÕES, 2008).

As medidas de refletância difusa de amostras sólidas (Figura 5c), como mencionado anteriormente, é um modo de medida característico para análises empregando a espectrometria NIR. A refletância difusa ocorre em superfícies não totalmente planas, podendo o substrato ser contínuo ou fragmentado. Este modo de medida é apropriado quando a superfície do material reflectante é pouco absorvente no comprimento de onda incidente. Dessa forma, a energia radiante dos modos vibracionais da molécula a ser analisada, penetra na camada superficial das partículas, excitando-os e então, tem-se um espalhamento em todas as direções. A energia refletida emerge aleatoriamente da amostra em todas as direções. Esta energia refletida difusamente carrega informação suficiente para obtenção do espectro na região do infravermelho próximo (CABRAL; GRANGEIRO JÚNIOR, 2011; SIMÕES, 2008).

Devido à sua complexidade, não há teorias rigorosas sobre a refletância difusa. No entanto, uma série de teorias empíricas foram propostas, entre as quais a mais popular é a teoria de Kubelka-Munk descrita na Equação 9 (CABRAL; GRANGEIRO JÚNIOR, 2011; SIMÕES, 2008). Para amostras opacas, como é o caso de um pó (CABRAL, 2011), de espessura infinita a função de Kubelka-Munk é descrita como segue, onde (R_{∞}) é a refletância absoluta da amostra. Esta teoria assume que a radiação incidente sobre o meio dispersante sofre um processo de absorção e dispersão, simultaneamente. Deste modo, a radiação refletida pode ser descrita em função das constantes de absorção (k_a) e dispersão (s) (SIMÕES, 2008):

$$f(R_{\infty}) = \frac{(1-R_{\infty})^2}{2R_{\infty}} = \frac{k_a}{s} \quad \text{Equação (9)}$$

Em análises quantitativas, a Equação 9 pode ser descrita em função da concentração do analito absorvente (c) de acordo com a seguinte equação:

$$f(R_{\infty}) = \frac{(1-R_{\infty})^2}{2R_{\infty}} = \frac{k_a}{s} = \frac{ac}{s} \quad \text{Equação (10)}$$

Onde: (a) é a absorvidade molar.

Na prática, ao invés da refletância absoluta (R_{∞}) é utilizada a refletância relativa (R), que é a relação entre a intensidade de luz refletida pela amostra e a intensidade de luz refletida pelo padrão. Este padrão deve ser um material estável, com refletância absoluta elevada e

relativamente constante na região espectral do infravermelho próximo, por exemplo, brometo de potássio, teflon, sulfato de bário, óxido de magnésio.

Reescrevendo a Equação 10 em função da refletância relativa, obtém-se:

$$f(R_{\infty}) = \frac{(1-R)^2}{2R} = \frac{ac}{s} \quad \text{Equação (11)}$$

Para amostras que obedecem a Equação 11, o gráfico $f(R)$ em função da concentração é uma linha reta com coeficiente angular igual a (a/s) . Entretanto, se a matriz também absorve ou se o analito apresenta forte absorção, a refletância difusa não obedece a esta relação linear.

Com isso, a equação de Kubelka-Munk, assim como a lei de Beer, é limitada e só pode ser aplicada para bandas de absorção de baixa intensidade. Esta limitação ocorre quando se utiliza a espectrometria NIR, pois não é possível separar a absorção do analito da absorção da matriz, ocorrendo desvios de linearidade (GRANGEIRO JÚNIOR, 2011; SIMÕES, 2008).

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

$$\log \frac{R_{padr\tilde{a}o}}{R_{amostra}} = \log \frac{1}{R_{amostra}} + \log R_{padr\tilde{a}o} \cong \frac{ac}{s} \quad \text{Equação (12)}$$

Para a radiação monocromática o termo $(\log R_{padr\tilde{a}o})$ é constante, sendo assim, a Equação 12 pode ser escrita como:

$$A = \log \frac{1}{R} = ac \quad \text{Equação (13)}$$

Onde: (A) é a absorbância aparente e (a) uma constante que engloba os termos $(1/s)$ e $(\log R_{padr\tilde{a}o})$. Apesar de não ter as bases teóricas da equação de Kubelka-Munk, esta equação apresenta resultados muito satisfatórios nas aplicações da espectroscopia por refletância difusa (SIMÕES, 2008).

2.6 Transformada de Fourier

Em 1950, os astrônomos desenvolveram a espectroscopia com transformada de Fourier (do inglês: *Fourier Transform Infrared - FTIR*) com o objetivo de estudar o espectro de infravermelho de estrelas distantes. Eles utilizaram um dispositivo denominado interferômetro de Michelson, o qual foi desenvolvido pelo professor de física da Universidade de Chicago (Albert Abraham Michelson) para determinar o efeito da rotação da terra na velocidade da luz (CABRAL, 2011; SKOOG, 2009).

Quando os espectrômetros FTIR apareceram pela primeira vez no mercado, no início da década de 1970, eram enormes e muito caros (mais de 100 mil dólares) e requeriam ajustes mecânicos frequentes. Por essas razões, seu uso estava limitado a aplicações especiais nas quais as suas características únicas (alta velocidade, alta resolução, alta sensibilidade e excelente precisão e exatidão em relação ao comprimento de onda) eram essenciais. Contudo atualmente os espectrômetros FTIR tiveram seu tamanho reduzido podendo ser alocados em bancadas e têm se tornado muito confiáveis e de fácil manutenção. Além disso, os modelos mais simples atualmente apresentam um preço similar aos espectrômetros dispersivos simples. Dessa forma, os espectrômetros FTIR estão substituindo largamente os instrumentos dispersivos nos laboratórios (SKOOG, 2008).

A maioria dos espectrômetros FTIR de bancada são do tipo de feixe único. Para se obter o espectro da amostra, primeiro obtém-se um espectro do fundo, o qual é conhecido como *background* (solvente, água presente no ambiente e dióxido de carbono). Depois, consegue-se o espectro da amostra. Finalmente, a razão entre o espectro de feixe único da amostra e o espectro do fundo é calculada e a absorvância ou transmitância *versus* o comprimento de onda é registrada. Frequentemente, os instrumentos de bancada são purgados com um gás inerte ou ar seco, livre de CO₂, para reduzir a absorção de vapor de água e dióxido de carbono de fundo (SKOOG, 2008).

As maiores vantagens dos instrumentos FTIR sobre os espectrômetros dispersivos incluem uma maior velocidade e sensibilidade, melhor aproveitamento da potência luminosa, calibração do comprimento de onda mais exata, desenho mecânico simples e a eliminação virtual de problemas de radiação espúria e emissão infravermelha. Em virtude dessas vantagens, quase todos os novos instrumentos empregam sistemas FTIR (CABRAL, 2011; SKOOG, 2008).

2.7 Espectroscopia NIR aplicada à indústria farmacêutica

O infravermelho próximo não foi considerado inicialmente como uma técnica analítica com algum valor prático, sendo originalmente, uma extensão da região do visível que não foi explorada até 1970. Os primeiros trabalhos que despertaram interesse pelo estudo da espectrofotometria no infravermelho próximo como ferramenta de análise na indústria, foram desenvolvidas na década de setenta pelo professor Karl Norris, quando este era responsável por um grupo de pesquisa do Departamento de Agricultura dos Estados Unidos. Entretanto, as limitações tecnológicas da época não permitiram o uso mais abrangente desta técnica como atualmente vem sendo observado em diversos setores industriais (GRANGEIRO JÚNIOR, 2011).

Na literatura recente são apresentadas várias propostas de utilização da espectroscopia NIR no controle de qualidade de fármacos. Aplicações no controle de qualidade de insumos e do processo de homogeneização de misturas de constituintes sólidos de comprimidos têm demonstrado as principais vantagens desta técnica, em termos de possibilitar determinações em tempo real, sem que haja a necessidade de pré-tratamento da amostra. A identificação de insumos é uma aplicação padrão da espectrofotometria NIR, e em geral, a principal motivadora para a aquisição de um equipamento. O procedimento usual é a comparação dos espectros dos insumos com os espectros arquivados em uma biblioteca utilizando técnicas quimiométricas de reconhecimento de padrões. Além disso, são possíveis outras aplicações como determinação de dureza em comprimidos e acompanhamento do perfil de dissolução (GRANGEIRO JÚNIOR, 2011; SCAFI, 2000).

Recentemente, vários trabalhos científicos vêm sendo desenvolvidos visando o estudo e à aplicação da espectrometria NIR. Atualmente, esta técnica tem ganhado aceitação pela indústria farmacêutica como meio de identificação, quantificação e controle de parâmetros físicos e químicos, seja da matéria-prima, seja do produto final. Inclusive, várias agências reguladoras, como a Agência Europeia de Medicamentos (do inglês: *European Medicines Agency* – EMEA) através de publicações como (EMEA/CVMP/961/01), o Grupo de Ciências Analíticas e Farmacêuticas (do inglês: *Group of Analytical Sciences and Pharmaceutical* - GASP) e farmacopeias (como por exemplo: a Farmacopeia Europeia e a Farmacopeia dos Estados Unidos) vêm recomendando as metodologias baseadas na espectrometria NIR como uma boa alternativa analítica para controle de qualidade dos produtos farmacêuticos, devido às vantagens apresentadas em relação aos métodos convencionais. No entanto, sua aceitação e emprego na indústria farmacêutica, bem como, em

outros setores industriais, tem sofrido restrições devido às exigências para a sua validação, como aquelas expressas na norma (ASTM E 1655 05) e outras normas específicas. Nesta norma são estabelecidas as recomendações necessárias para a construção e validação de modelos de calibração multivariada usando espectrometria NIR (SIMÕES, 2008).

Validação de um método analítico é o processo pelo qual se estabelece evidência documentada, através de estudos laboratoriais, de que as características de desempenho ou parâmetros analíticos do método alcançam requisitos necessários para as aplicações analíticas pretendidas. Atualmente, a validação das metodologias analíticas é requisito fundamental no processo de registro de medicamentos genéricos pelo Ministério da Saúde. No processo de validação de uma metodologia analítica é necessário avaliar alguns parâmetros como: robustez, linearidade, faixa de variação, especificidade, precisão, exatidão, limites de detecção e quantificação. Esses parâmetros são conhecidos como figuras de mérito e têm como objetivo garantir que a metodologia proposta seja confiável para a estimativa das concentrações de amostras desconhecidas, podendo ser usada efetivamente para o controle de qualidade de parâmetros físicos e químicos (RANDAU et al., 2005; SIMÕES, 2008).

2.8 Quimiometria

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

A aplicação desta técnica tem se difundido nas mais diversas áreas da Química e ciências afins. O desenvolvimento de novos instrumentos e a aplicação dos mesmos em laboratórios de pesquisas passou a fornecer uma grande quantidade de informações, o que criou a necessidade do desenvolvimento de uma ferramenta estatística sofisticada que pudesse ser aplicada na análise desses dados. Esse grande número de informações e tal necessidade, aliados a uma facilidade crescente de acesso a microcomputadores, possibilitou o crescimento da quimiometria (CABRAL; GRANGEIRO JÚNIOR, 2011).

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

A quimiometria pode ser considerada como um dos mais recentes campos da Química Analítica (SIMÕES, 2008). Sendo subdividida em outras áreas principais como: processamento de sinais analíticos, planejamento e otimização de experimentos, reconhecimento e classificação de padrões, calibração multivariada, monitoramento e modelagem de processos multivariados e métodos de inteligência artificial. Nesse contexto, a principal linha de pesquisa da quimiometria aplicada à Química Analítica tem sido a construção de modelos de regressão a partir de dados de primeira ordem, ou seja, dados que podem ser representados através de um vetor para cada amostra, sendo a construção desses modelos denominada de calibração multivariada (GRANGEIRO JÚNIOR, 2011).

2.9 Análise multivariada

A necessidade de entender as relações entre muitas variáveis faz da análise multivariada um assunto de difícil compreensão. Além do mais, a matemática necessária para fazer inferências de técnicas é mais complexa do que em técnicas univariadas (FONSECA, 2011).

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

- a) Redução de dados ou Simplificação Estrutural;
- b) Triagem e Agrupamento;
- c) Investigação da dependência entre variáveis;
- d) Predição;
- e) Construção e teste de hipóteses.

2.10 Calibração multivariada

Dentre os métodos de calibração existentes, certamente os mais difundidos são ainda os métodos de calibração univariada. Os quais são conhecidos também como “calibração de

ordem zero”, ou seja, tem-se apenas uma medida instrumental para cada uma das amostras de calibração. Em outras palavras, para cada amostra tem-se apenas uma escalar. Contudo, a aplicação desse tipo de calibração é restrita, uma vez que, quando a amostra não é livre de interferentes e a medida é realizada diretamente por uma metodologia instrumental, isso provavelmente provocará desvios na determinação da propriedade de interesse e a aplicação deste método de calibração torna-se inviável (FONSECA, 2011).

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

Dentre os métodos quimiométricos, os que mais se aplicam a Química Analítica são os de calibração multivariada. Esse tipo de calibração consiste na construção de modelos matemáticos para relacionar dados de saída de um instrumento (medidas instrumentais multivariadas) às concentrações dos analitos ou a outras propriedades das amostras (SIMÕES, 2008).

As técnicas multivariadas são adequadas para solução de problemas analíticos porque permitem estudos com várias espécies presentes ao mesmo tempo, não importando a existência ou ausência de diferenças espectrais marcantes entre elas nem a existência de alta correlação nos dados. É possível também, a correção de problemas eventuais com linha base ou interferentes nas amostras usadas na calibração e nas novas amostras de predição (GRANGEIRO JÚNIOR, 2011).

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

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

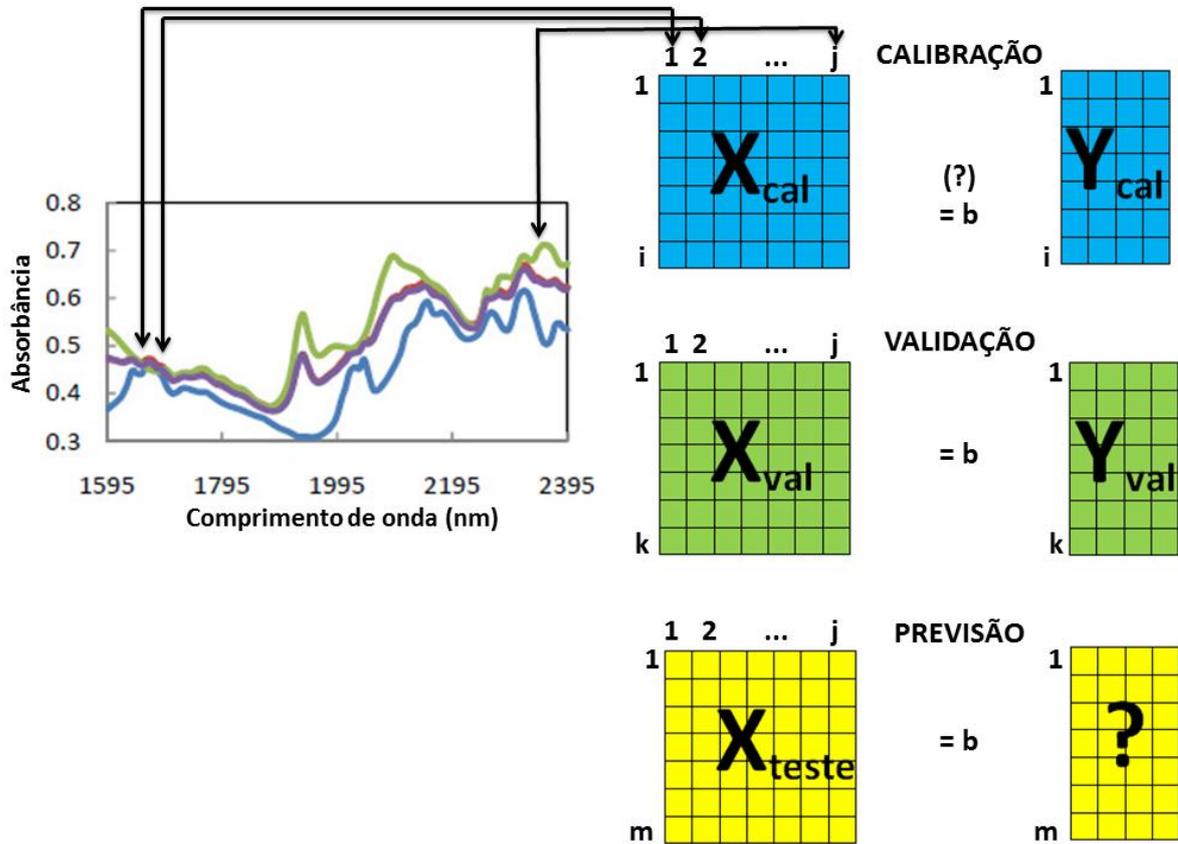
$$y = f(X) = Xb$$

Equação (14)

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

Os dados para a calibração multivariada podem ser organizados conforme ilustra a Figura 6. Os valores de absorvância dos espectros, a cada número de onda, correspondem às variáveis independentes (matriz X), e as concentrações das amostras, definida pelo método de referência, indicam as variáveis dependentes (vetor y) (GRANGEIRO JÚNIOR, 2011).

Figura 6 - Organização dos dados para calibração multivariada



Fonte: adaptado de SIMÕES (2008)

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

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

Tem sido possível através de métodos quimiométricos, especialmente os métodos de calibração multivariada, a determinação de propriedades químicas e físicas de amostras ativas na região do Infravermelho e Raman. Isso tem sido feito por métodos com modelagem local, como regressão linear múltipla (MLR, do inglês: *Multiple Linear Regression*), que utiliza a absorbância em comprimentos de onda selecionados, ou métodos de modelagem global, tais

como a regressão por componentes principais (PCR, do inglês: *Principal Component Regression*) ou mínimos quadrados parciais (PLS, do inglês: *Partial Least Squares*) (FONSECA, 2011; SIMÕES, 2008).

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

2.10.1 Regressão por componentes principais (PCR)

O método da regressão dos componentes principais emprega as variações espectrais para calcular as equações de calibração. Como as variações mais significativas no conjunto de calibração são as mudanças no espectro relacionadas às diferentes concentrações das amostras, é possível calcular um conjunto que represente as mudanças nas absorvâncias em todo o espectro. Essas variações são conhecidas como autovetores (componentes principais ou fatores) e as constantes utilizadas para multiplicar os espectros são conhecidas como *scores* (pesos) (OLIVEIRA, 2006; SCAFI, 2000).

2.10.1.1 Análise de componentes principais (PCA)

Uma das formas de calcular os autovetores do espectro é através do método de Análise de Componentes Principais (do inglês: *Principal Components Analysis*). Este modelo foi introduzido na Química por Malinowski na década de 1960, inicialmente com o nome de Análise de Fatores, e consiste em um processo de simplificação dos dados que torna possível encontrar relações entre objetos (FONSECA, 2011; SCAFI, 2000).

A PCA tem por objetivo a redução do número de variáveis, por meio da substituição das variáveis originais por outras chamadas Componentes Principais (do inglês: *Principal Components - PC*). Além disso, esse método apresenta vantagens como: detecção de *outliers* (amostras que não correspondem bem ao modelo de calibração) e a possibilidade de classificação e predição de resultados (BEEB; RANDY; SEASHOLTZ, 1998).

Assim sendo, a PCA é capaz de decompor uma determinada matriz original, dada por (m x n) em uma soma de matrizes, que são produtos de vetores denominados de *scores* e *loadings*, os quais representam o peso das variáveis. Deste modo, a primeira componente

principal (PC1), é definida em uma direção com máxima variância de um conjunto de dados. Enquanto que, a segunda componente principal (PC2) consiste de uma descrição da máxima variância no espaço da primeira componente principal, e assim sucessivamente. Cada componente principal é responsável por uma parcela de contribuição em uma fração de variância dos dados. Sendo descritas como (PC1, PC2, PC3, PC4, PC5, ...), essas componentes estão num sistema de coordenadas ortogonais entre si, não existindo correlação entre elas. É importante ressaltar que a maior parte da variância total é explicada pelas primeiras componentes principais (CABRAL, 2011; TAYLOR & FRANCIS GROUP, 2006).

Como mencionado anteriormente, a PCA decompõe a (matiz X) no produto de duas matrizes. Logo:

$$X = M_1 + M_2 + M_3 + M_4 + M_5 \dots + M_n \quad \text{Equação (15)}$$

$$X = T_1P_1 + T_2P_2 + T_3P_3 + T_4P_4 + T_5P_5 \dots + T_nP_n \quad \text{Equação (16)}$$

Sabendo-se que a matriz X é determinada pela Equação 17, realiza-se a decomposição de valores singulares (SVD) por meio de um algoritmo. Obtendo-se em seguida, o cálculo dos *scores* e *loadings* (CABRAL, 2011).

$$X = USP^t \quad \text{Equação (17)}$$

Onde: (P) é a matriz dos *loadings* e (US) a matriz dos *scores*. (U e P) são colunas ortogonais, enquanto que, (S) é uma matriz diagonal no qual seus elementos são valores singulares e fornece informações sobre a variância que cada componente descreve. Sendo esta variância descrita pela próxima equação (CABRAL, 2011):

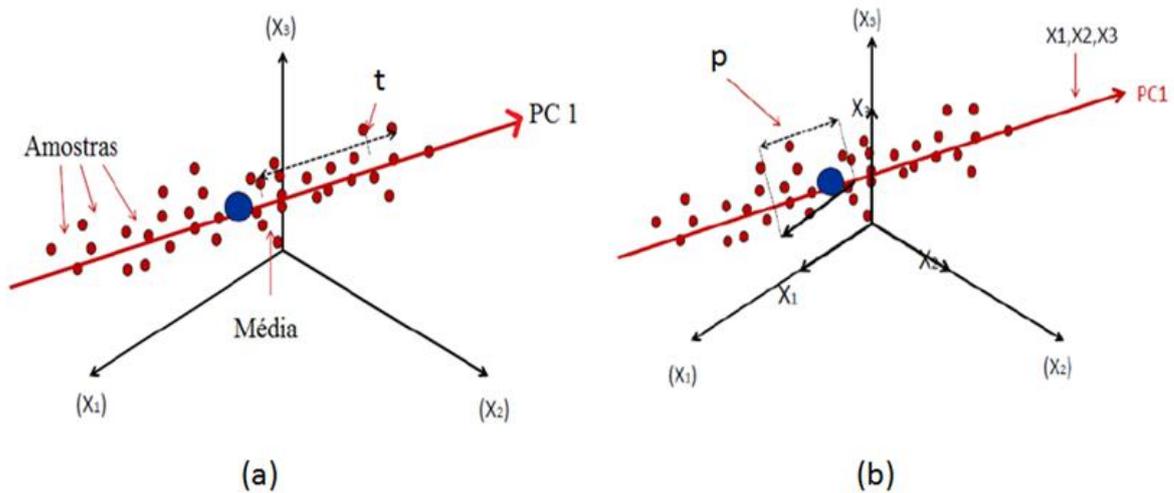
$$\%VAR(PC_j) = \frac{s_{jj}^2}{\sum_{j=1}^p s_{jj}^2} \quad \text{Equação (18)}$$

A Equação 18 mostra que o denominador refere-se à variância total, enquanto que, (P) indica o número total de valores singulares do conjunto de dados, com o numerador se referindo à variância de cada componente principal (CABRAL, 2011).

Pode-se evidenciar logo mais abaixo, que no gráfico da figura (7a), as distâncias entre a média e a componente principal são os *scores* (t) para a amostra e os *loadings* são

exatamente os pesos que as variáveis exercem. Já no gráfico da figura (7b), “p” representa a distância do (vetor X_1) ao ponto de coloração azul. Também é possível observar que a seta em vermelho simboliza a descrição dos *loadings* (X_1, X_2, X_3) (CABRAL, 2011).

Figura 7 - Representação dos *scores* (a) e *loadings* (b)



Fonte: CABRAL (2011)

Para se construir um gráfico que represente as componentes principais, deve-se tomar por base alguns critérios. Inicialmente, é necessário o auxílio de um software, seguido de um tratamento da matriz de dados. Desse modo, geram-se algumas variáveis de categoria para cada classe de amostra individual, selecionam-se também os números de variáveis e amostras, fazendo uma sugestão do número de componentes principais a serem utilizadas (CABRAL, 2011).

Para a determinação do número correto de componentes principais o método mais utilizado é o da Validação Cruzada (do inglês: *Cross Validation* - CV), o qual se baseia na habilidade de predição de um modelo construído por parte de um conjunto de dados, seguido pela predição do restante desse próprio conjunto de dados, que é realizada pelo modelo construído. A validação cruzada pode ser realizada em blocos, ou seja, um número determinado de amostras é deixado de fora do processo de construção no modelo e a seguir essas amostras são preditas pelo modelo construído, ou ainda por um caso conhecido como *leave-one-out* (deixe uma fora por vez), em que uma amostra é deixada de fora no processo de construção do modelo, sendo também predita pelo modelo construído. Em ambos os casos, o

processo é repetido até que todas as amostras tenham sido preditas, e assim, o erro quadrático médio da validação cruzada (RMSECV) possa ser calculado (FONSECA, 2011).

2.10.2 Regressão por mínimos quadrados parciais (PLS)

O PLS ou regressão por PLS (PLSR) foi desenvolvido por Herman Wold em meados de 1964, sendo considerado o método de regressão mais utilizado para a construção de modelos de calibração multivariada a partir de dados de primeira ordem (FONSECA, 2011; SIMÕES, 2008). Para a construção de tais modelos, o método utiliza tanto as informações da matriz de dados espectroscópicos (matriz X), como as informações da matriz de concentrações (matriz Y) (SIMÕES, 2008).

A regressão por mínimos quadrados parciais estende o conceito do modelo inverso (propriedade como função da resposta instrumental) trocando as variáveis originais por um subconjunto truncado das variáveis latentes dos dados originais (BEEB; RANDY; SEASHOLTZ, 1998; OLIVEIRA, 2006). Considerando um caso geral para a determinação de mais de uma espécie de interesse, as matrizes X_{cal} e Y_{cal} (Figura 6) são decompostas em suas matrizes de pesos e escores, respectivamente, sendo esse processo representado pelas Equações 19 e 20, respectivamente (SIMÕES, 2008).

$$X = TP^T + E \quad \text{Equação (19)}$$

$$Y = UQ^T + F \quad \text{Equação (20)}$$

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

A decomposição em escores e pesos pode ser realizada empregando diferentes tipos de algoritmos. No entanto, eles produzem no final resultados semelhantes. Um exemplo desses algoritmos é o NIPALS (do inglês: *Non-linear Iterative Partial Least Squares*). O modelo final em PLS consiste em relacionar linearmente os escores da (matriz X) com os escores da (matriz Y), de acordo com as seguintes equações (SIMÕES, 2008):

$$U = BT + G \quad \text{Equação (21)}$$

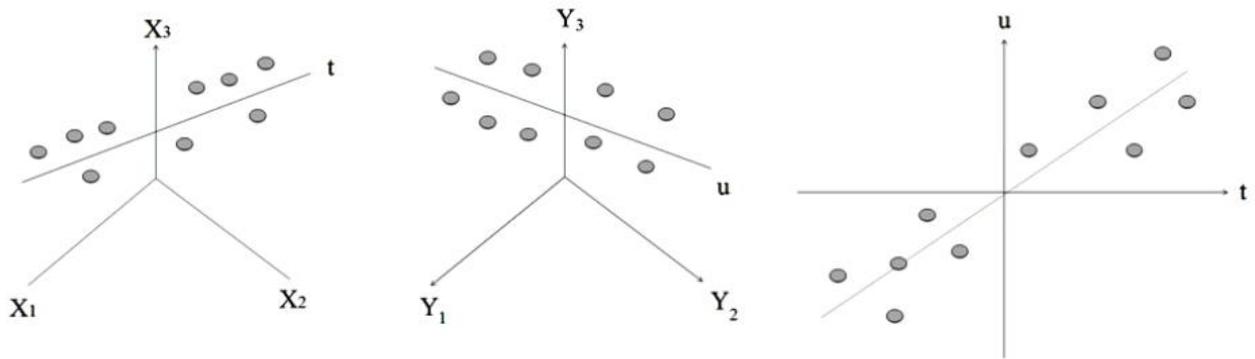
$$Y = BTQ^T + H$$

Equação (22)

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

A Figura 8 exemplifica, de forma genérica, como o método PLS processa os dados da matriz. Logo, para a construção desse modelo, a regressão dos mínimos quadrados retira informações de um conjunto de dados da (matriz X) correlacionando-os com as informações retiradas do conjunto de dados da (matriz Y) (CABRAL, 2011).

Figura 8 - Representação genérica do método PLS



Fonte: CABRAL (2011)

2.10.3 Regressão linear múltipla (MLR)

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

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

$$\mathbf{y}_{cal} = \mathbf{X}_{cal}\mathbf{b}_{MLR} + \mathbf{e} \quad \text{Equação (23)}$$

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

$$\mathbf{b}_{MLR} = (\mathbf{X}_{cal}^T \mathbf{X}_{cal})^{-1} \mathbf{X}_{cal}^T \mathbf{y}_{cal} \quad \text{Equação (24)}$$

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

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

$$\hat{\mathbf{y}}_{desc} = \mathbf{X}_{desc} \mathbf{b}_{MLR} + \mathbf{e} \quad \text{Equação (25)}$$

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

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

2.11 Cálculo dos erros

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

$$RMSE = \frac{\sqrt{\sum_{i=1}^n (y_{predição} - y_{real})^2}}{n} \quad \text{Equação (26)}$$

Em que, (n) é número de amostras.

$$E = \sqrt{\frac{(y_{predição} - y_{real})^2}{(y_{real})^2}} \cdot 100 \quad \text{Equação (27)}$$

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

2.12 Coeficiente de correlação

Uma simples associação entre as variáveis (X e Y) produz um par de valores ou, graficamente, aponta esta dispersão em uma linha reta. Uma pequena dispersão muito próxima a linha indica forte associação (dispersão). A medida numérica desta relação é chamada Coeficiente de Correlação da Amostra. Também conhecido como Coeficiente de Correlação de Pearson, em homenagem ao seu criador, sendo simbolizado por (R) (FONSECA, 2011). Observando-se a Equação 28, infere-se que o valor de (R) varia no intervalo $(-1,0 \leq R \leq 1,0)$.

$$R = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n [(x_i - \bar{x})^2] [(y_i - \bar{y})^2]}} \quad \text{Equação (28)}$$

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

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

3 PARTE EXPERIMENTAL

3.1 Amostras

Para a construção dos modelos foram utilizadas 33 amostras fabricadas na planta piloto do Laboratório Farmacêutico de Pernambuco (LAFEPE), de modo a ampliar a faixa de concentração dos comprimidos. Estas amostras foram desenvolvidas no LAFEPE a partir de um planejamento fatorial experimental e cedidas para este trabalho. Cada uma das misturas do planejamento experimental possuía uma massa média de 300 mg e foi preparada através da mistura dos dois princípios ativos e dos excipientes. Deste modo, as massas de AZT e 3TC foram variadas dentro de uma faixa de $\pm 15\%$ do valor rotulado. Com isso, a concentração dos comprimidos fabricados na planta piloto variou de 255 a 345 mg. Os diluentes (celulose e amido) foram usados para completar a massa total do comprimido (800 mg) e a massa dos demais excipientes (aerosil, polivinilpirrolidona, estearato de magnésio) foi mantida constante para preservar as características físico-químicas dos comprimidos. Além destes, também foram coletados comprimidos de 29 lotes distintos produzidos na linha de produção do LAFEPE. Totalizando 62 amostras.

3.2 Determinação do teor de AZT pelo método de referência

As análises cromatográficas foram realizadas no LAFEPE, de acordo com o procedimento descrito na Farmacopeia Brasileira (Farmacopeia Brasileira, 2010), empregando-se um sistema cromatográfico Shimadzu[®], constituído por bombas binárias LC-10ADvp, detector UV/VIS SPD 10 Avp em 270 nm, forno CTO-10ASvp (30°C), controlador do sistema SCL-10ADvp e auto-injetor SIL-10ADvp, fase estacionária (coluna C18 125 x 4,6 cm, porosidade 5 μm) e fase móvel (tampão de acetato de amônio pH 5,4 e metanol 65:35), fluxo de 1,0 mL/min e volume de injeção de 20 μL . Os cromatogramas foram obtidos e analisados através do programa Class-VP. O metanol e o acetato de amônio utilizados na fase móvel foram das marcas J.T.Baker e Merck, respectivamente. A água deionizada utilizada em todas as etapas foi obtida por sistema Milli-Q[®]-Gradient. Todas as vidrarias utilizadas foram da marca Pyrex[®] e certificadas por lote. Além disso, todos os equipamentos foram previamente qualificados e certificados.

3.3 Aquisição dos espectros NIR

Os espectros de todas as amostras foram registrados no Departamento de Engenharia Química da Universidade Federal de Pernambuco (UFPE). Os mesmos foram registrados utilizando um espectrômetro FTIR da marca BOMEM (modelo FTLA 2.000), na faixa de 14.000 a 4.000 cm^{-1} , utilizando 32 varreduras e resolução de 8 cm^{-1} . O branco utilizado foi Teflon (99% reflective Spectralon, Labsphere me., North Sutton, NH, USA).

3.4 Softwares

O programa UNSCRAMBLER[®]v9.8 da CAMO foi empregado para pré-processamento dos espectros e construção dos modelos PLS. Os algoritmos MLR/APS (ARAÚJO, 2001) e SPXY (GALVÃO, 2005) foram executados empregando o ambiente MATLAB[®], versão 6.5. Esses programas foram usados, respectivamente, para a seleção de variáveis nos modelos MLR e para seleção de amostras do conjunto de calibração.

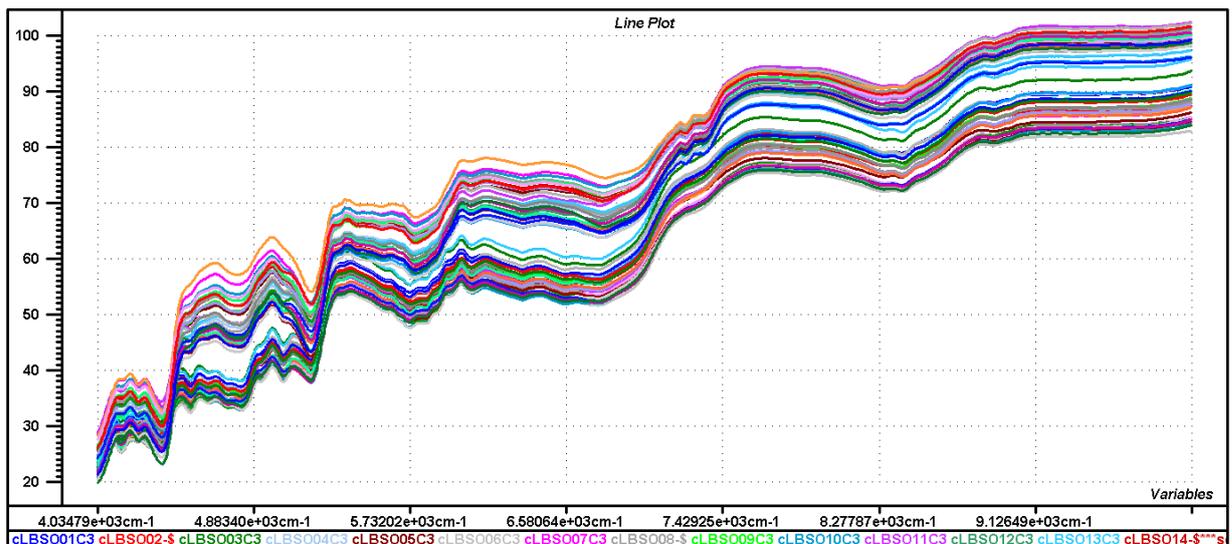
4 RESULTADOS E DISCUSSÃO

4.1 Espectros

Toda medida espectral é composta por uma porção de sinal analítico relacionado à amostra estudada e uma porção de ruído, que é informação espúria não relacionada ao problema analítico estudado. Para retirar ou minimizar a presença destes ruídos podem ser utilizados pré-tratamentos matemáticos, como por exemplo: algoritmos. Neste trabalho serão testados os seguintes pré-tratamentos: SNV (do inglês: *Standard Normal Deviation*), MSC (do inglês: *Multiplicative Scatter Correction*), derivada e suavização, e as combinações destes.

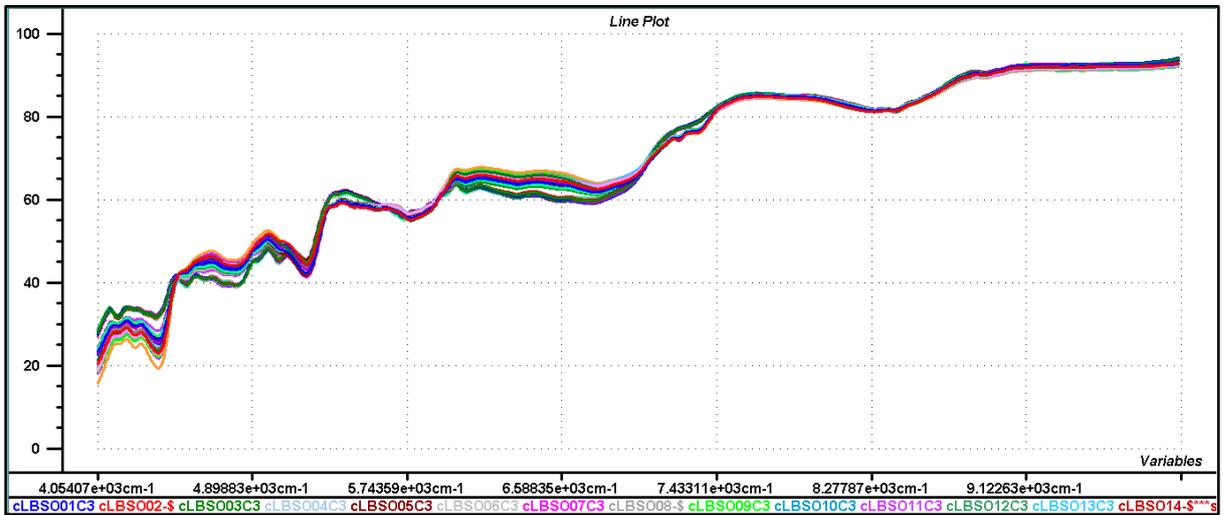
A Figura 9 mostra os dados sem pré-tratamento para os espectros obtidos. Pode-se observar nestes espectros a presença de ruído espectral e deslocamento de linha de base que podem vir a prejudicar a capacidade preditiva dos modelos construídos. Deste modo, vários pré-tratamentos espectrais foram testados e o que apresentou o melhor resultado foi a combinação de suavização utilizando o polinômio Savitzky-Golay com janela de 11 pontos e polinômio de 2ª ordem combinado ao MSC (Figura 10).

Figura 9 - Espectros no infravermelho próximo, não tratados, obtidos para os comprimidos da dose fixa combinada de AZT+3TC



Fonte: autoria própria

Figura 10 - Espectros pré-tratados, utilizando a suavização com o algoritmo Savitzky-Golay, polinômio de 2ª ordem e janela de 11 pontos combinado ao MSC

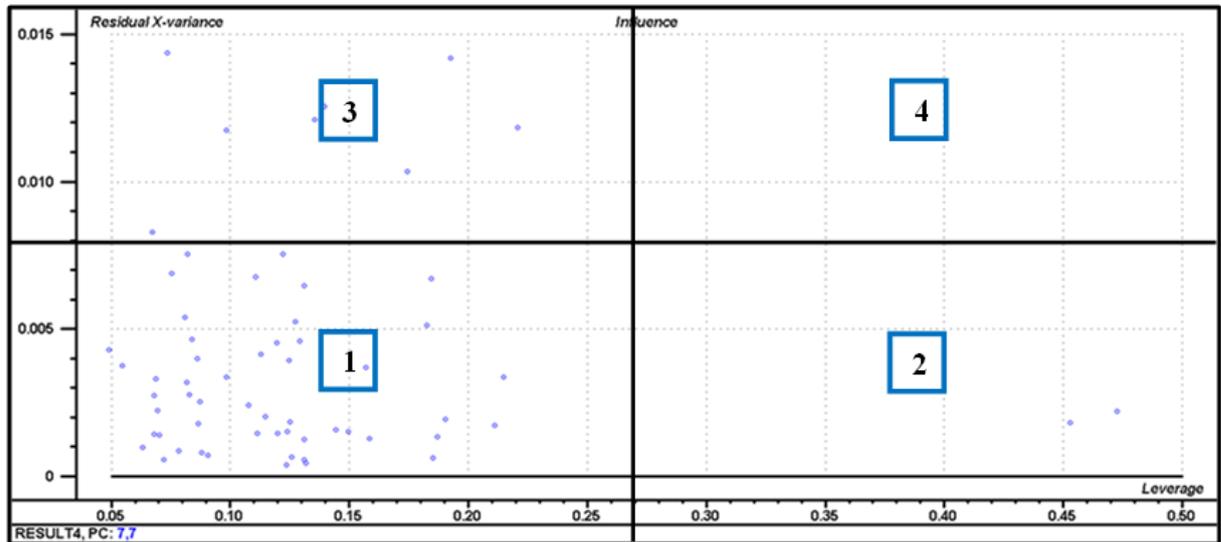


Fonte: autoria própria

4.2 Construção dos modelos multivariados

A construção de qualquer modelo multivariado deve ser precedida pela verificação da presença de amostras anômalas (*outliers*). Isto pode ser feito a partir do estudo do gráfico de influencia versus resíduo (Figura 11). Amostras com baixa influência e baixo resíduo (1), baixa influência e alto resíduo (2), alta influência e baixo resíduo (3), mostram um comportamento normal em relação aos espectros (X) e aos teores (y). Já amostras com alto resíduo e alta influência (4), apresentam provavelmente comportamento anômalo em relação a X e y, devendo ser observadas e, se adequado, excluídas do conjunto de dados.

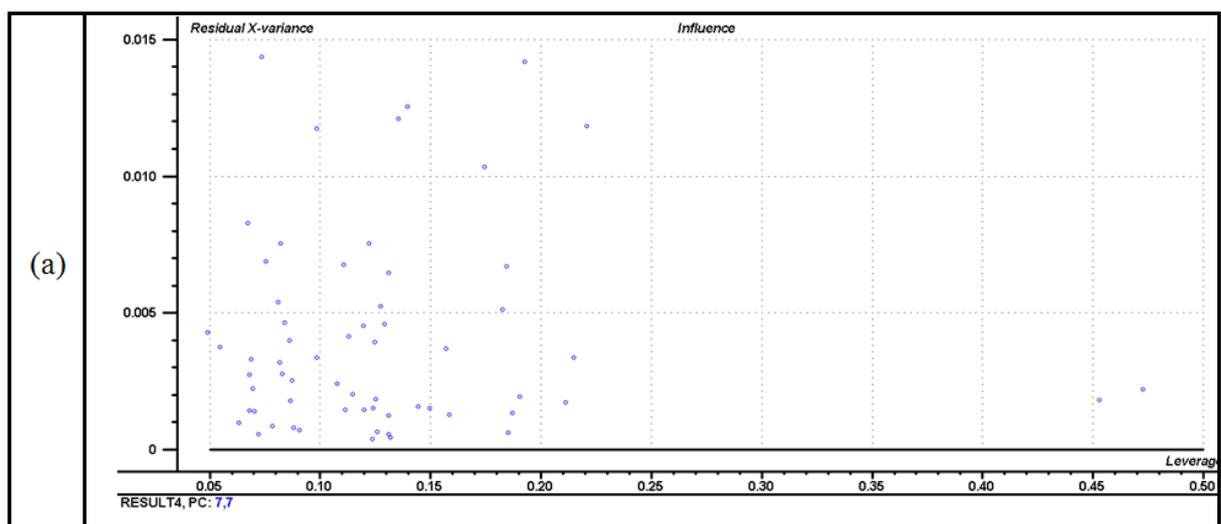
Figura 11 - Gráfico de influência versus resíduo



Fonte: autoria própria

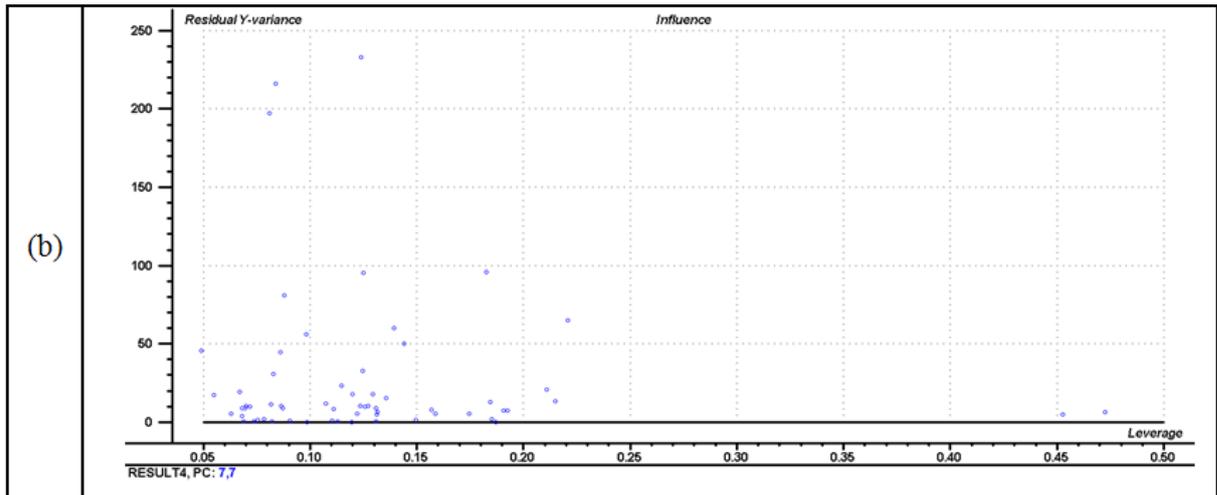
As figuras 12a e 12b mostram os gráficos de influência versus resíduo em X e y, respectivamente, para os dados brutos obtidos por reflectância. De acordo com a análise destes gráficos pode-se afirmar que não há a presença de amostras anômalas em nenhum dos conjuntos de dados. Já que não foi possível verificar a presença de amostras com alto resíduo, seja de X ou de y, e alta influência.

Figura 12a - Gráfico da influência versus resíduos para os dados estudados em X



Fonte: autoria própria

Figura 12b - Gráfico da influência versus resíduos para os dados estudados em y



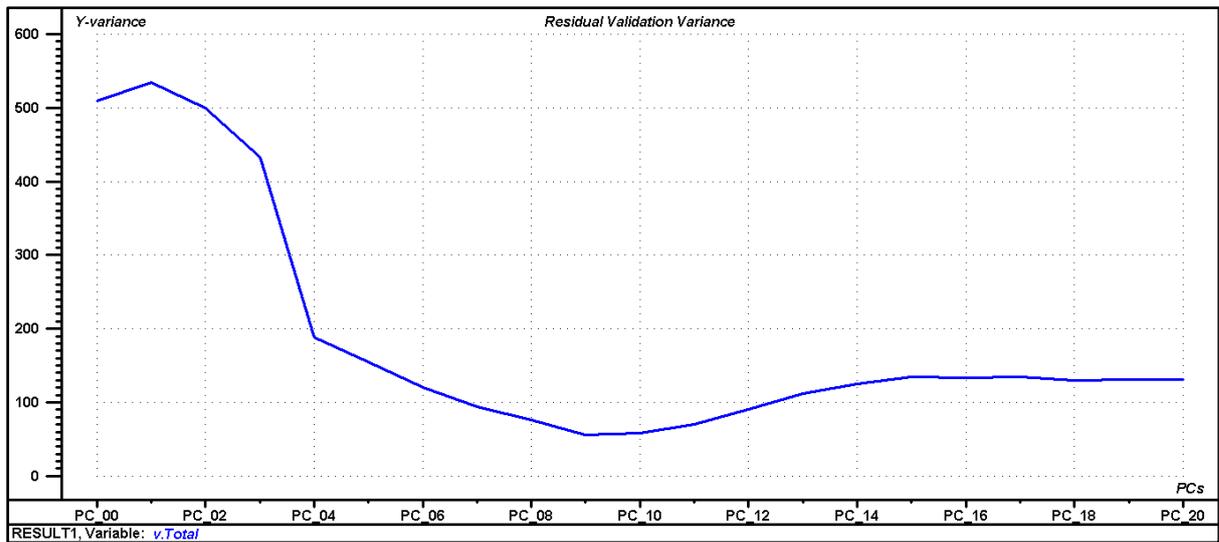
Fonte: autoria própria

A segunda etapa da construção do modelo multivariado é a seleção dos conjuntos de calibração e predição. Neste trabalho foi utilizado o algoritmo SP-Xy (GALVÃO, 2005) para este fim. Deste modo, foram selecionadas 42 amostras para o conjunto de calibração e 20 para o conjunto de predição.

4.2.1 Construção dos modelos PLS para a determinação do teor de AZT

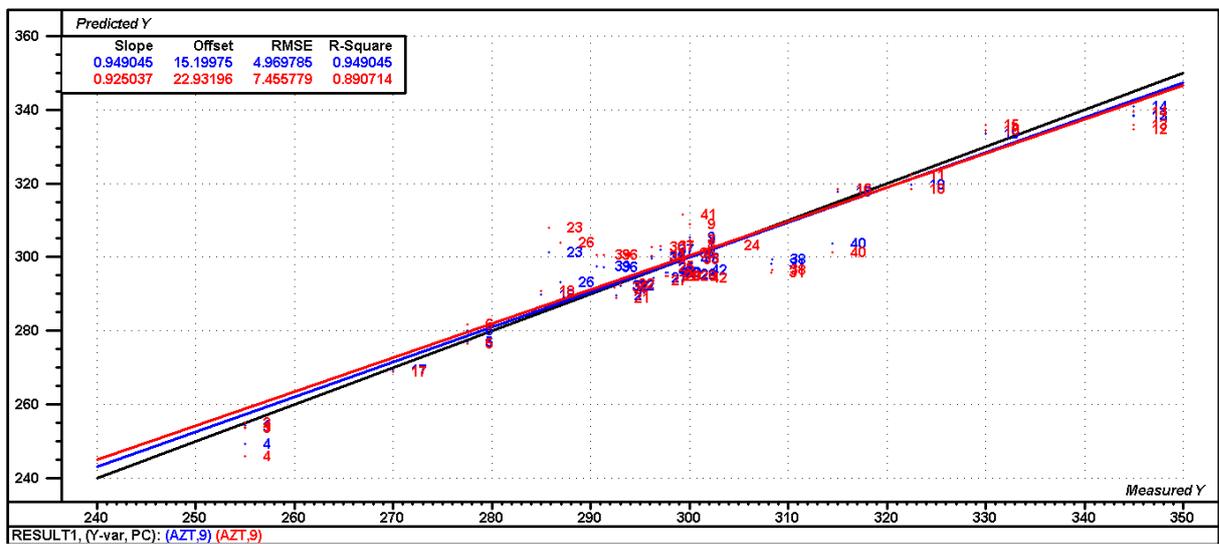
A construção do modelo PLS utilizando os espectros de reflectância, sem pré-tratamento, utilizou 9 variáveis latentes, como pode ser observado no gráfico de variância residual versus número de variáveis latentes apresentado na Figura 13. O modelo mencionado apresentou um RMSE da calibração e validação cruzada de 4,97 e 7,45 mg, respectivamente. O coeficiente de correlação para as etapas de calibração e validação cruzada foram 0,9490 e 0,8907, respectivamente. A Figura 14 mostra a curva de calibração para os dados de reflectância. A reta azul representa a calibração, a vermelha representa a predição e a preta ideal de calibração. Podemos notar que não há uma concordância perfeita entre estas três retas. O que indica que os dados possuem informações espúrias que comprometem a etapa de calibração. Isto é comprovado pela grande diferença entre os valores de RMSE das etapas de calibração e validação cruzada. Na etapa de predição (Figura 15) foi obtido um erro de 6,44 e o coeficiente de correlação desta etapa foi de 0,9668.

Figura 13 - Gráfico de variância residual versus número de variáveis latentes para os dados sem pré-tratamento



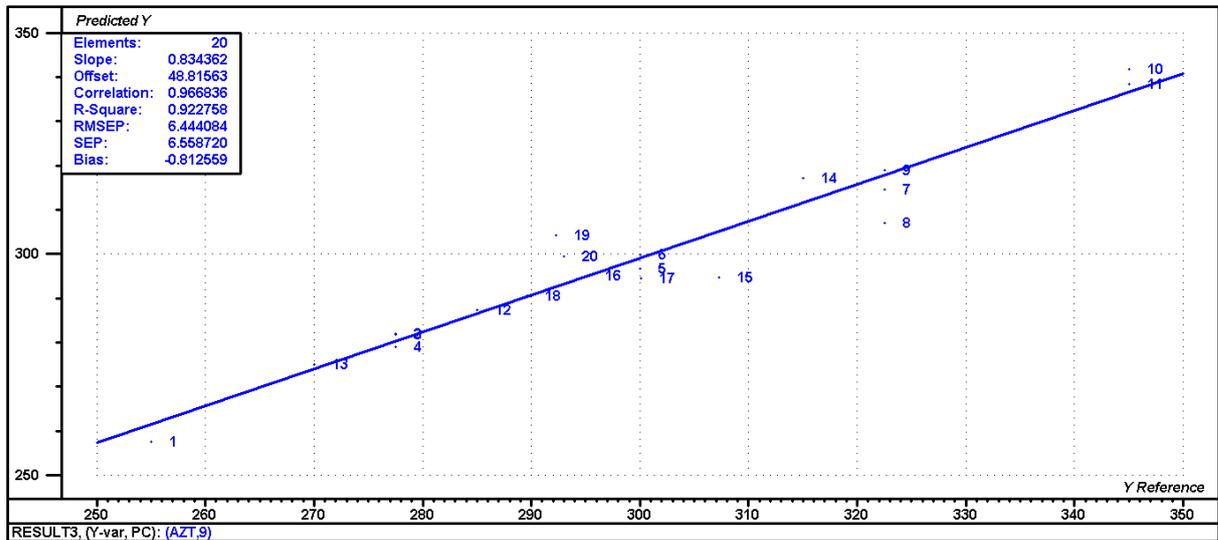
Fonte: autoria própria

Figura 14 - Curva de calibração para os dados de refletância sem pré-tratamento



Fonte: autoria própria

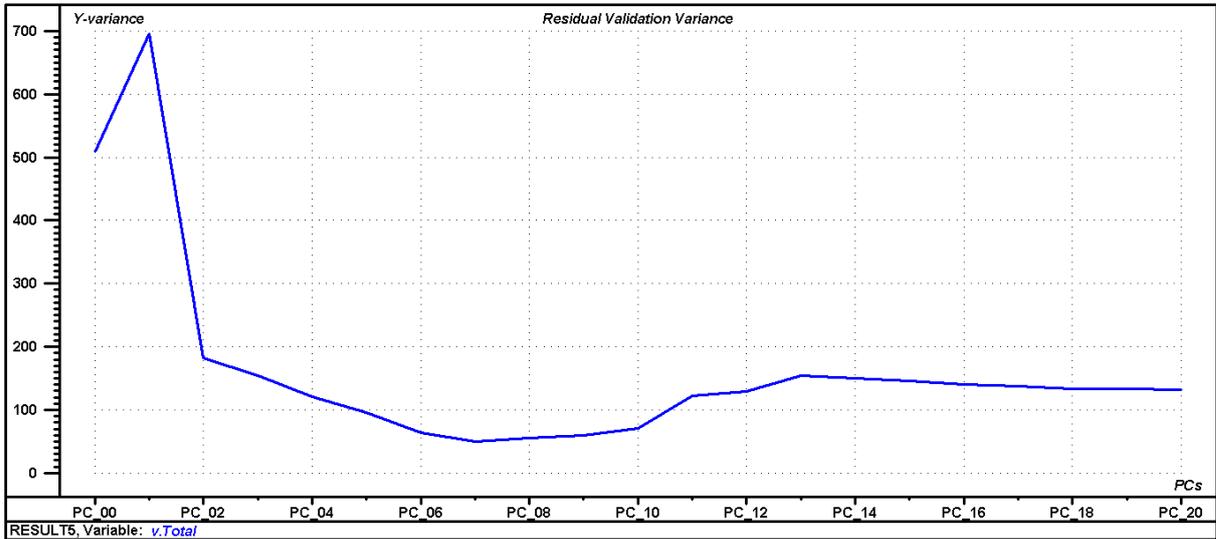
Figura 15 - Gráfico dos valores de referência versus valores previstos pelo modelo construído com os dados sem pré-tratamento



Fonte: próprio autor

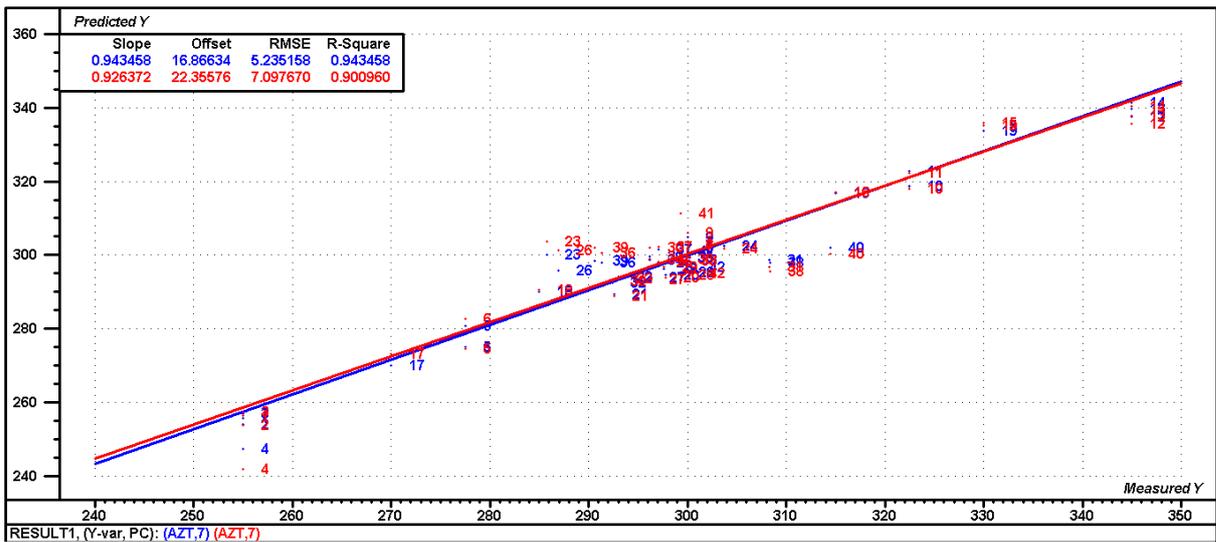
Desta forma, foram testados diversos pré-tratamentos espectrais e o que apresentou os melhores resultados foi a combinação da suavização (utilizando o algoritmo Savitzky-Golay, com polinômio de 2ª ordem e janela de 11 pontos) com a correção multiplicativa do sinal (MSC). Utilizando este pré-tratamento houve uma diminuição do número de variáveis latentes utilizadas para a construção do modelo (7 para os dados pré-tratados contra 9 para os dados brutos), como pode ser observado no gráfico de variância residual versus número de variáveis latentes, apresentado na Figura 16. O RMSE obtido nas etapas de calibração e predição foi de 5,23 e 7,09 mg, respectivamente. O coeficiente de correlação para as etapas de calibração e validação cruzada foram 0,9434 e 0,9009, respectivamente. A Figura 17 mostra a curva de calibração para os dados de refletância. A reta azul representa a calibração e a vermelha representa a predição. Podemos notar que não há uma concordância perfeita entre estas duas retas. O que indica que os dados possuem informações espúrias que comprometem a etapa de calibração. Isto é comprovado pela grande diferença entre os valores de RMSE das etapas de calibração e validação cruzada. Na etapa de predição (Figura 18) foi obtido um erro de 5,49 e o coeficiente de correlação desta etapa foi de 0,9758.

Figura 16 - Gráfico de variância residual versus número de variáveis latentes para os dados pré-tratados



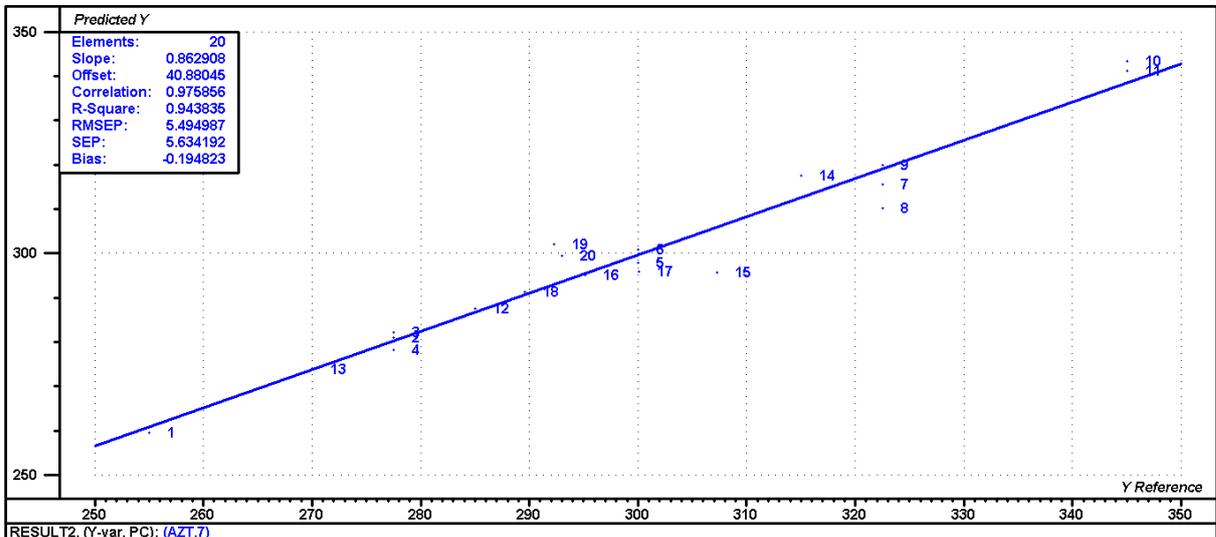
Fonte: autoria própria

Figura 17 - Curva de calibração para os dados de refletância pré-tratados



Fonte: autoria própria

Figura 18 - Gráfico dos valores de referência versus valores previstos pelo modelo construído com os dados pré-tratados



Fonte: autoria própria

Na Tabela 1, são observados os dados de performance do modelo construído utilizando os dados brutos e os dados pré-tratados.

Tabela 1 - Parâmetros para avaliação dos modelos construídos com os dados sem e com pré-tratamento

	RMSEC	Correlação	RMSECV	Correlação	Variável Latente	RMSEP	Correlação
Brutos	4,97	0,9490	7,45	0,8907	9	6,44	0,9668
Pré-tratados	5,23	0,9434	7,09	0,9009	7	5,49	0,9758

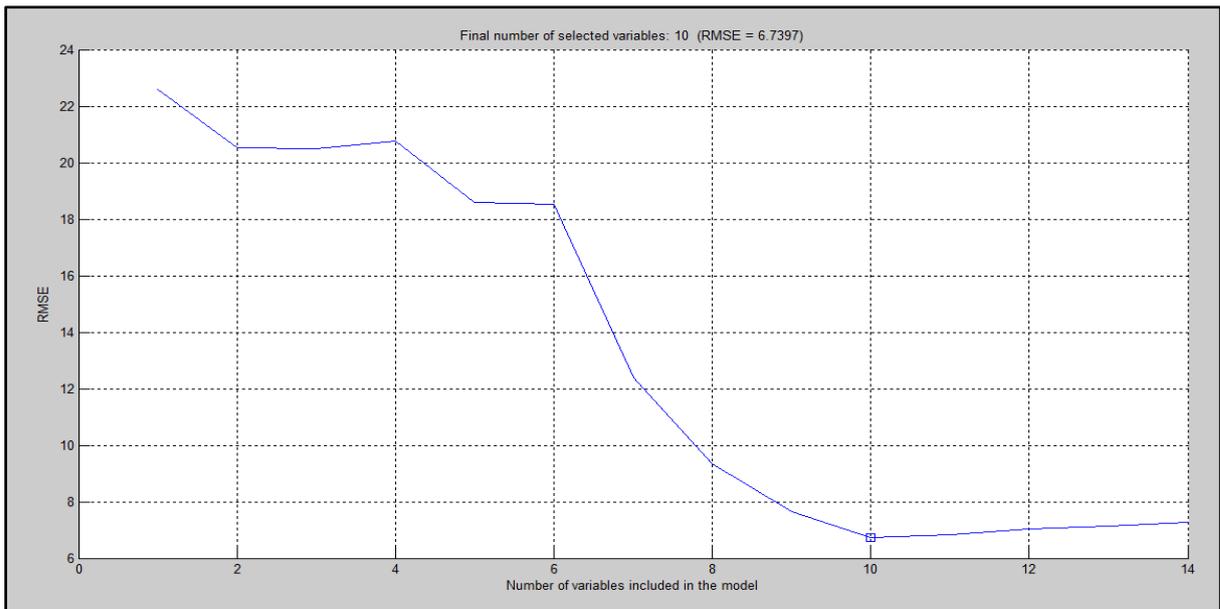
Fonte: autoria própria

De acordo com o teste t pareado bicaudal, a um nível de 0,05 de significância, não há diferença significativa entre os valores previstos pelo modelo PLS e aqueles obtidos pelo método de referencia ($t_{\text{calc}} = 0,8787 < t_{0,05;20} = 2,086$). Tais valores encontram-se listados no APENDICE A.

4.2.2 Construção dos modelos MLR para a determinação do teor de AZT

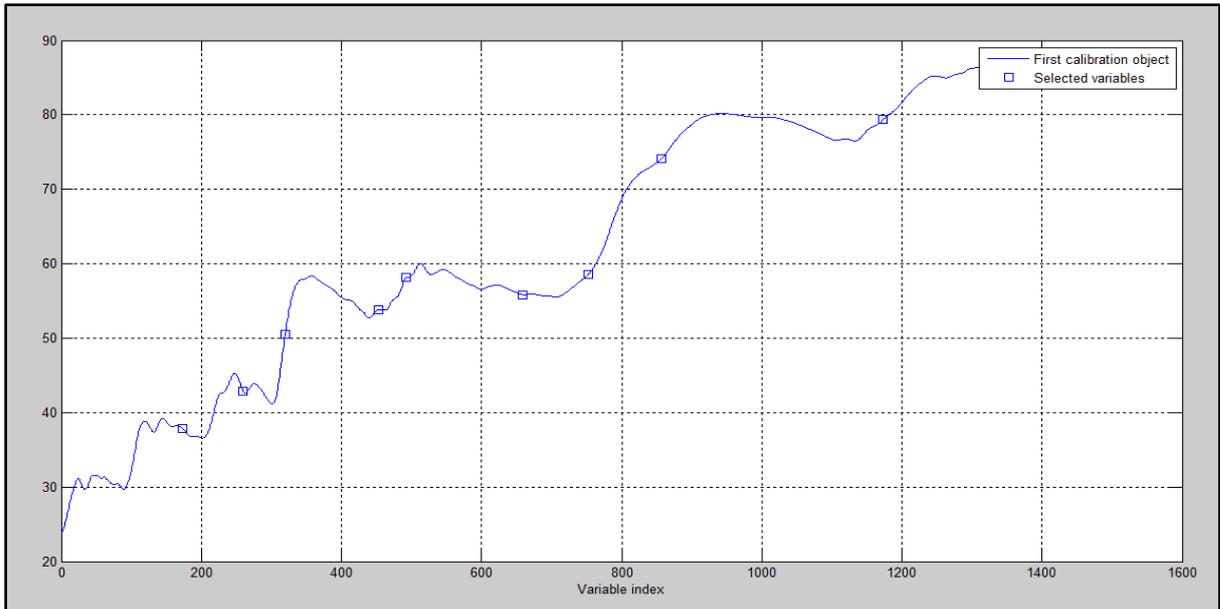
Para efeito de comparação entre os dois tipos de métodos de regressão (PLS E MLR), foi construído um modelo usando o método de regressão linear múltipla, tendo como base o mesmo conjunto de dados (sem a presença de amostras anômalas) e o mesmo pré-tratamento espectral. Foi obtido um erro na etapa de calibração de 6,74 mg e utilizou-se 10 variáveis espectrais na construção do modelo (Figura 19). As variáveis mais informativas, ou seja, mais relacionadas a variação de concentração do AZT nos comprimidos, selecionadas pelo algoritmo das projeções sucessivas (APS) são mostradas na Figura 20. O erro para a etapa de predição (RMSEP) foi de 5,67 mg (Figura 21). Na Tabela 2 estão listados os parâmetros para avaliação de performance dos dois modelos.

Figura 19 - Gráfico de número de variáveis versus RMSE para os dados pré-tratados



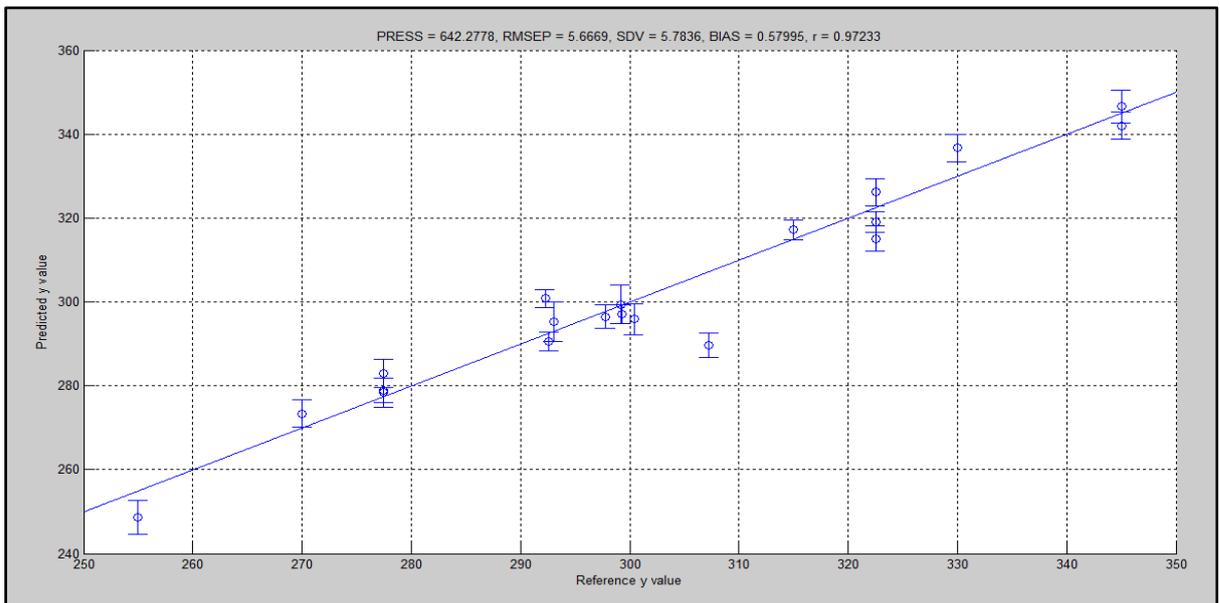
Fonte: autoria própria

Figura 20 - Variáveis selecionadas pelo algoritmo da projeções sucessivas (APS)



Fonte: autoria própria

Figura 21 - Gráfico dos valores de referência versus valores previstos pelo modelo construído com os dados pré-tratados



Fonte: autoria própria

Tabela 2 - Parâmetros para avaliação dos modelos construídos utilizando PLS e MLR

	RMSEC	RMSEP	Número de fatores	R
PLS	5,23	5,49	7	0,9758
MLR	6,74	5,67	10	0,9723

Fonte: autoria própria

De acordo com o teste F, para comparação de variâncias, não há diferença significativa entre os valores de RMSEP apresentados pelos dois métodos de regressão ($F_{\text{cal}} = 1,07 < F_{0,05;20;20} = 2,12$).

5 CONCLUSÕES

A espectroscopia no infravermelho próximo mostrou-se uma técnica adequada para o uso no controle de qualidade de medicamentos, no tocante a metodologia desenvolvida para a determinação do teor de AZT em comprimidos. Foram testados alguns métodos para a remoção de ruído espectral e aquele que mostrou o melhor desempenho foi a combinação da suavização utilizando o polinômio Savitzky-Golay com polinômio de 2ª ordem e janela de 11 pontos e MSC. Em relação aos métodos de regressão avaliados, podemos afirmar de acordo com o teste F ($F_{\text{cal}} = 1,07 < F_{0,05;20;20} = 2,12$) que não há diferença estatística entre os mesmos. Ambos os modelos apresentaram erros dentro dos limites estabelecidos para indústria farmacêutica, ou seja, 10% do teor. Deste modo, podem ser utilizados no controle de qualidade do teor de AZT em comprimidos dose fixa combinada (AZT+3TC).

REFERÊNCIAS

ANDRADE, C. R. et al. Calibrações por meio da técnica da espectroscopia no infravermelho próximo para teor de extrativos totais, de lignina total e holocelulose usando espectros obtidos na madeira sólida. **Ambiência Guarapuava (PR)**, v. 7, n. 1, p. 39-49, jan./abr. 2011. Disponível em: <revistas.unicentro.br/index.php/ambiencia/article/viewFile/678/1189>. Acesso em: 12 dez. 2013.

ARAÚJO, M. C. U. et al. "The successive projections algorithm for variable selection in spectroscopy multicomponent analysis". In: **Chemometrics and Intelligent Laboratory Systems**, v.57, p. 65-73, 2001.

BECK, R. C. R. et al. VALIDAÇÃO DE MÉTODO POR CROMATOGRAFIA LÍQUIDA DE ALTA EFICIÊNCIA PARA DETERMINAÇÃO DA LAMIVUDINA E ZIDOVUDINA EM COMPRIMIDOS. **Química Nova**, v. 30, n. 5, p. 1225-1228, 2007. Disponível em: <www.scielo.br/scielo.php?pid=S0100-40422007000500033&script=...>. Acesso em: 19 nov. 2013.

BEEB, K. R.; RANDY, J. P.; SEASHOLTZ, M. B. **Chemometrics: A practical guide**. New York: Wiley-Interscience, 1998. 348 p. ISBN 0-471-12451-6.

CABRAL, A. M. **Construção de modelos multivariados para a determinação de lipídios totais e umidade em leite em pó em leite comercial utilizando espectroscopia no infravermelho próximo**. Universidade Federal do Rio Grande do Norte. Natal, RN, p. 104. 2011. (RN/UFRN/BSE-Química CDU 637.147). disponível em: <<http://repositorio.ufrn.br:8080/jspui/handle/1/6959>>. Acesso em: 15 jan. 2014.

CALDEIRA, A. F.; SANTOS, S. L. D.; SACON, V. "HEXENURONIC ACID, KLASON LIGNIN AND VISCOSITY OF PULP PREDICTED BY NIR SPECTROSCOPY". Disponível em: <www.scielo.br/scielo.php?script=sci_arttext&pid=S0100>. Acesso em: 19 set. 2013.

Farmacopeia Brasileira; 5a ed. Volumes -1 e 2 ANVISA, 2010.

FONSECA, F. D. S. "**Espectrometria no Infravermelho Médio e Calibração Multivariada por PLS para quantificação de biodiesel metílico e etílico de soja, milho e girassol em misturas com diesel**". Universidade Federal de Uberlândia. Uberlândia, MG, p. 65. 2011. Disponível em: <https://www.google.com.br/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CD8QFjAA&url=http%3A%2F%2Frepositorio.ufu.br%2Fbitstream%2F123456789%2F788%2F1%2FEspectrometriaInfravermelhoMedio.pdf&ei=oyZEU_SnK8bisASV7IGoBw&usg=AFQjCNHqog-b_KcgLyHANYihb0f06bdvGA>. Acesso em: 14 abr. 2014.

GALVÃO, R.K.H. et al. A method for calibration and validation subset partitioning, *Talanta*, 67,736, 2005.

GOOGLE. **Google Imagens**, 2014. Disponível em: <[https://www.google.com.br/search?q=Funções%20de%20potenciais%20harmônicos%20\(---\)%20e%20anarmônicos%20\(---\)](https://www.google.com.br/search?q=Funções%20de%20potenciais%20harmônicos%20(---)%20e%20anarmônicos%20(---)>)>. Acesso em: 28jan. 2014.

GRANGEIRO JÚNIOR, S. **Aplicação de Métodos Quimiométricos em Análises de Formas Farmacêuticas Sólidas Lamivudina (3TC) + Zidovudina (AZT) Baseados na Espectrofotometria de absorção na região do Ultravioleta e do Infravermelho próximo.** Universidade Federal de Pernambuco. Recife, PE, p. 132. 2011. (UFPE 615.19 CDD (22.ed.) CCS2011-041). Disponível em:

<www.btdt.ufpe.br/btdt/tedeSimplificado/tde.../processaPesquisa.php?>. Acesso em: 15 out. 2013.

MORGANO, M. A. et al. DETERMINAÇÃO DE PROTEÍNA EM CAFÉ CRU POR ESPECTROSCOPIA NIR E REGRESSÃO PLS. **Ciência Tecnologia de Alimentos**, Campinas, SP, v. 25, n. 1, p. 25-31, jan.- mar. 2005. Disponível em:

<chipre.iqm.unicamp.br/~marcia/Pub78.pdf>. Acesso em: 25 out. 2013.

NASCIMENTO, J. E. D. et al. Avaliação farmacocinética de comprimidos contendo lamivudina e zidovudina em plasma humano. **Revista Brasileira de Ciências Farmacêuticas**, v. 40, n. 1, p. 59-66, jan./mar. 2004. Disponível em:

www.scielo.br/pdf/rbcf/v40n1/10.pdf. Acesso em: 28 dez. 2013.

OLIVEIRA, F. C. D. C. **MODELOS DE CALIBRAÇÃO MULTIVARIADA ASSOCIADOS À ESPECTROSCOPIA VIBRACIONAL PARA ANÁLISE DE MISTURAS DIESEL- ÓEOS VEGETAIS.** Universidade de Brasília. Brasília, DF, p. 108. 2006. Disponível em:

<https://www.google.com.br/url?sa=t&rct=j&q=&esrc=s&source=web&cd=7&ved=0CGwQFjAG&url=http%3A%2F%2Frepositorio.unb.br%2Fbitstream%2F10482%2F3106%2F1%2FFlavia%2520C%2520C%2520Oliveira.pdf&ei=oyZEU_SnK8bisASV7IGoBw&usg=AFQjCNFFnsgX89bVjzCwJ83eKsINdbkTcg>. Acesso em: 14 de abr. 2014.

RANDAU, K. P. et al. Desenvolvimento e Validação de Metodologia Analítica para Anti-Retroviral Zidovudina (AZT)-Matéria-Prima. **Acta Farm. Bonaerense**, v. 24, n. 1, p. 104-108, Dezembro 2005. ISSN 0326-2383. Disponível em: <sedici.unlp.edu.ar > ... > vol. 24, no. 01>. Acesso em: 7 jan. 2014.

SCAFI, S. H. F. **Espectroscopia no Infravermelho Próximo para a identificação de medicamentos falsificados.** Universidade de Campinas. Campinas, SP, p. 138. 2000. Disponível em: <<http://www.bibliotecadigital.unicamp.br>>. Acesso em: 23 maio 2013.

SIMÕES, S. D. S. **Desenvolvimento de métodos validados para a determinação de captopril usando espectrometria NIR e calibração multivariada.** Universidade Federal da Paraíba. João Pessoa, PB, p. 83. 2008. [UFPB/BC CDU:543(043)]. Disponível em: <www.ppgq.quimica.ufpb.br/teses/Tese_Simone_Simoes.pdf>. Acesso em: 20 ago. 2013.

SKOOG, D. A. et al. **Fundamentos de Química Analítica.** 8ª ed. São Paulo, SP: Thomson, 2008.

SKOOG, D.A.; HOLLER F.G.; CROUCH S.R. **Princípios de Análise Instrumental.** Porto Alegre, Rs: Bookman, 2009.

SOUZA, J. S. D.; FERRÃO, M. F. Aplicações da espectroscopia no infravermelho no controle de qualidade de medicamentos contendo diclofenaco de potássio. Parte I : Dosagem por regressão multivariada. **Revista Brasileira de Ciências Farmacêuticas**, v. 42, n. 3, p. 437-445, jul./set. 2006. Disponível em: <<http://www.scielo.br/pdf/rbcf/v42n3/a13v42n3.pdf>>. Acesso em: 29 nov. 2013.

SOUZA, M. V. N. D.; ALMEIDA, M. V. D. DROGAS ANTI-VIH: PASSADO, PRESENTE E PERSPECTIVAS FUTURAS. **Química Nova**, v. 26, p. 366-372, 2003. ISSN 3. Disponível em: <www.scielo.br/scielo.php?script=sci_arttext&pid=S0100...>. Acesso em: 05 jan. 2014.

SOUZA, M. V. N. D. Fármacos Inibidores de Fusão: uma Nova Estratégia no Combate à Replicação do Vírus VIH. **Acta Farm. Bonaerense**, v. 24, n. 2, p. 291-299, 4 Janeiro 2005. ISSN 0326-2383. Disponível em: <www.latamjpharm.org/trabajos/24/.../LAJOP_24_2_6_2_H4JJ04IZ7O.p...>. Acesso em: 5 de jan. de 2014.

TAYLOR & FRANCIS GROUP. **PRACTICAL GUIDE to CHEMOMETRICS**. 2^a. ed. Lodon - New York: [s.n.], 2006. 520 p. ISBN ISBN 1-57444-783-1 (alk. paper).

APÊNDICE A – Teores de AZT determinados pelo método de referência e previstos pelo modelo PLS utilizando os dados pré-tratados

Teor de AZT previsto pelo modelo PLS (mg)	Teor de AZT obtido pelo método de referencia (mg)
259,54	255,00
281,02	277,50
282,27	277,50
278,19	277,50
297,80	300,00
300,83	300,00
315,50	322,50
310,23	322,50
319,95	322,50
343,41	345,00
341,30	345,00
287,58	285,00
273,88	270,00
317,62	315,00
295,65	307,26
295,14	295,14
295,87	300,06
291,26	300,07
302,07	300,08
299,40	300,09