

UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE QUÍMICA

DEPARTAMENTO DE QUÍMICA ANALÍTICA
LAQA - LABORATÓRIO DE AUTOMAÇÃO EM QUÍMICA ANALÍTICA
LQTA - LABORATÓRIO DE QUIMIOMETRIA TEÓRICA E APLICADA



Dissertação de mestrado

Espectroscopia NIR e Calibração Multivariada Aplicadas ao
Controle de Qualidade de Gases Combustíveis Naturais e
Derivados do Petróleo

Autor: Yuri Guimarães Dias
Orientadora: Profa. Dra. Márcia Miguel Castro Ferreira
Co-orientador: Prof. Dr. Jarbas José Rodrigues Rohwedder

Campinas, 2011

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA DO INSTITUTO DE
QUÍMICA DA UNICAMP

D543e Dias, Yuri Guimarães.
Espectroscopia NIR e calibração multivariada aplicadas ao controle de qualidade de gases combustíveis naturais e derivados do petróleo / Yuri Guimarães Dias. -- Campinas, SP: [s.n], 2011.

Orientador: Profa. Dra. Márcia Miguel Castro Ferreira.

Coorientador: Prof. Dr. Jarbas José Rodrigues Rohwedder.

Mestrado - Universidade Estadual de Campinas, Instituto de Química.

1. Espectroscopia no infravermelho próximo. 2. Calibração multivariada. 3. Figuras de mérito. 4. Gás natural. II. Ferreira, Márcia Miguel Castro. II. Rohwedder, Jarbas José Rodrigues. III. Universidade Estadual de Campinas. Instituto de Química. IV. Título.

Título em inglês: NIR spectroscopy and multivariate calibration applied to quality control of natural gas and other fuel gases derived from petroleum

Palavras-chaves em inglês: NIR spectroscopy, Multivariate calibration, Figures of merit, Natural gas

Área de concentração: Química Analítica

Titulação: Mestre em Química na área de Química Analítica

Banca examinadora: Profa. Dra. Márcia Miguel Castro Ferreira (orientador), Profa. Dra. Adriana Vitorino Rossi (IQ-UNICAMP), Prof. Dr. Jez Willian Batista Braga (IQ-UnB)

Data de defesa: 25/07/2011

Dedicatória

Este trabalho foi concluído com a ajuda impagável de algumas pessoas que, pela companhia e apoio prestado, não cairão no esquecimento.

Dedico esta conquista ao Hailhi, um tio que desempenhou papel de pai em toda minha trajetória.

AGRADECIMENTOS

- Agradeço imensamente aos meus irmãos que me ajudaram a manter o foco e seguir em frente, Igor e Ingrid; à Alyne pelo companheirismo e ao colega Adriano pela descontração proporcionada durante as conversas no MSN aos finais de semana.

- Aos colegas do grupo LQTA e do IQ, Euzébio, Clécio, Juliano, Eduardo, Samuel, Kerly, Flávia, Mário, Francisco e Sabrina, em especial ao João Paulo e Reinaldo pelos conhecimentos transmitidos.

- Ao professor Jarbas pela compreensão, ensinamentos e a amizade.

- À professora Márcia pela orientação.

- Aos professores com quem tive contato “pelos corredores” do IQ ou em disciplinas, professores Bruns, Ronei, Ivo, Célio, Wilson, Baccan, Solange e Adriana.

- À Unicamp e ao IQ pela excelente estrutura e apoio durante todo o trabalho.

- Ao CNPq pela bolsa concedida.

CURRICULUM VITAE

1 - DADOS PESSOAIS

Yuri Guimarães Dias

Brasileiro, solteiro

Natural de Maceió – AL

Nascimento: 20/03/1984

E-mail: diasyuri@iqm.unicamp.br, diasyuri@hotmail.com

2 - FORMAÇÃO ACADÊMICA

Ensino superior

- Bacharelado em química, Universidade de Brasília – **UnB**, DF, 2007

3 - CURSOS DE APERFEIÇOAMENTO

Cursos de curta duração:

- Norma ISO – 17025 para auditores internos
Universidade de Brasília, UnB, Brasil (16h) – ministrado por consultor Anvisa
- Validação em Análises Químicas
Universidade Católica de Brasília, UCB, Brasil. (16h)
- Estatística Multivariada Aplicada aos Dados Químicos
Sociedade Brasileira de Química, SBQ, Poços de Caldas, MG, Brasil. (6h).
- Espectroscopia de Absorção Atômica
Universidade de Brasília, UnB, Brasil, (6h).
- Espectroscopia Vibracional – Raman e Infravermelho
Universidade de São Paulo, USP, Brasil, (20h).
- Gestão Ambiental de Resíduos
Universidade de Brasília, UnB, Brasil, (6h).
- Tratamento de Resíduos Químicos em Rejeitos Industriais
Universidade de Brasília, UnB, Brasil. (6h)

4 – ATIVIDADES ACADÊMICAS

- BOLSISTA DO CNPQ NA REDE DE PESQUISA PROSAB 4 – 2005/2006
(rede nacional de projetos de pesquisa e desenvolvimento na área de saneamento básico).
INICIAÇÃO CIENTÍFICA

Durante esta oportunidade trabalhamos em cooperação com a equipe da faculdade de engenharia civil da Universidade de Brasília em um estudo que avaliava a possibilidade do uso de águas residuárias para irrigação ornamental em Brasília, bem como a espécie paisagística mais indicada para tal uso.

Neste período realizei a determinação e quantificação de diversas espécies químicas orgânicas e inorgânicas através da análise química por via úmida e instrumental. As amostras constituíam-se em matrizes ambientais como solo, tecido vegetal, água e efluentes, os quais eram analisados afim de avaliar-se o impacto de compostos contaminantes (considerando-se os limites de potabilidade atuais, resoluções CONAMA

etc) e a presença de substâncias recalcitrantes. Parâmetros físico-químicos como condutividade elétrica e umidade das amostras de solo também foram analisados. Além da avaliação ambiental, algumas rotinas foram modificadas e aperfeiçoadas.

- LABORATÓRIO DE QUÍMICA ANALÍTICA E AMBIENTAL - UNIVERSIDADE DE BRASÍLIA, 2004/2005
DISCIPLINA TÉCNICA DE PESQUISA II - QUÍMICA ANALÍTICA
Responsável pela análise de íons metálicos em matrizes ambientais como solo, água, efluentes e tecido vegetal utilizando-se análises químicas por via úmida e instrumentais como: absorção atômica (AAS) e plasma indutivamente acoplado (ICP-OES), além da análise estatística dos dados.
- LABORATÓRIO DE QUÍMICA INORGÂNICA PREPARATIVA - UNIVERSIDADE DE BRASÍLIA, 2003/2004
DISCIPLINA TÉCNICA DE PESQUISA I - QUÍMICA INORGÂNICA
Durante esta disciplina participei da pesquisa do grupo do prof. Doutor Sebastião de Souza Lemos e pude trabalhar na síntese e caracterização de complexos com potencial biológico de zinco e estanho com ligantes como a dimetilpirimidinona.
- UNIVERSIDADE DE BRASÍLIA – UnB, DF, 2005-2007
MONITOR DE DISCIPLINAS EXPERIMENTAIS DO CURSO DE BACHARELADO DE QUÍMICA
(Área de concentração: química analítica, carga horária: acima de 200 horas)

5 - TRABALHO EM EVENTO NACIONAL

- Dias, Y. G.; Ferreira, M. M. C.; Rohwedder, J. J. R.; Figuras de Mérito e Espectroscopia NIR para Determinação de Parâmetros de Espécies Presentes em Gás Natural, 14^a Encontro Nacional de Química Analítica, Salvador – BA, 2009.

6 - PARTICIPAÇÃO EM EVENTO NACIONAL

- Apresentação Oral do Trabalho Intitulado: Figuras de Mérito e Espectroscopia NIR para Determinação de Parâmetros de Espécies Presentes em Gás Natural, na Sessão Coordenada de Quimiometria, 14^a Encontro Nacional de Química Analítica, Salvador - BA, 2009.

7 - EXPERIÊNCIA PROFISSIONAL

- FLORA – PRODUTOS DE HIGIENE E LIMPEZA, LUZIÂNIA-GO – 2010/
(Indústria de produtos domésticos e cosméticos, detentora das marcas Albany e Minuano)
ANALISTA QUÍMICO SÊNIOR
Responsável pela gestão do controle de qualidade de alguns laboratórios durante o 2º turno da fábrica, desempenhando atividades de supervisão da equipe técnica, atualização das metodologias de análise, pesquisa e levantamento do custo de novos equipamentos para substituição de análises volumétricas e implementação de rotinas analíticas automatizadas.
Encarregado da montagem e estabelecimento do laboratório de cromatografia da empresa.

RESUMO

ESPECTROSCOPIA NIR, CALIBRAÇÃO E VALIDAÇÃO MULTIVARIADA PARA CONTROLE DOS PARÂMETROS DE QUALIDADE DE GASES COMBUSTÍVEIS NATURAIS E DERIVADOS DO PETRÓLEO

Autor: Yuri Guimarães Dias

Orientadora: Márcia Miguel Castro Ferreira

Co-orientador: Jarbas José Rodrigues Rohwedder

Este estudo avalia a capacidade do uso da espectroscopia NIR para a determinação do teor de 6 hidrocarbonetos e dióxido de carbono em 40 misturas gasosas com composição semelhante à do gás natural. Os espectros NIR foram obtidos a partir de um equipamento experimental, construído no próprio laboratório, baseado em um monocromador do tipo AOTF e simultaneamente em um espectrofotômetro comercial NIR-FT. Modelos de calibração multivariada PLS-1 foram construídos, para os 6 gases, com os dois conjuntos de dados. A validação do método foi realizada calculando-se as figuras de mérito multivariadas, envolvendo em alguns casos o conceito do sinal analítico líquido (NAS). Os valores de RMSECV variam de 0,1858% para o 1,3-butadieno, cuja faixa de porcentagem volumétrica é de: 0,8% a 4,9%; a 0,9732% para o metano cuja faixa de porcentagem volumétrica é de: 61,7% a 89,9%. Estes valores confirmam a possibilidade de determinação do teor das espécies constituintes do gás natural e uma comparação das variâncias, por meio do teste F, dos modelos obtidos com os espectrofotômetros NIR-AOTF e NIR-FT indicam a adequação do equipamento experimental utilizado.

ABSTRACT

NIR SPECTROSCOPY, MULTIVARIATE CALIBRATION AND VALIDATION IN QUALITY CONTROL OF NATURAL GAS AND OTHER FUEL GASES DERIVED FROM PETROLEUM

Author: Yuri Guimarães Dias

Adviser: Márcia Miguel Castro Ferreira

Co-adviser: Jarbas José Rodrigues Rohwedder

The present study evaluates the potential of NIR spectroscopy in determining the content of 6 hydrocarbons and carbon dioxide in 40 gas mixtures very similar in composition to the Brazilian natural fuel gas. The NIR spectra were obtained from an equipment built in our laboratory and based on a AOTF wavelength selector device. All the spectra were simultaneously recorded in a commercial NIR-FT spectrophotometer (Bomem MB160D). Multivariate calibration models (PLS-1) were built for the six gases with both data sets. Method validation was performed calculating the figures of merit based on the Net Analyte Signal (NAS). RMSECV values ranged from 0.1858%, for 1,3-butadiene which has a volumetric ratio of 0.8% to 4.9%; to 0.9732 for methane which has a volumetric ratio of 65.8% to 89.9%. These values confirm the possibility of measuring the content of all the constituents evaluated in this study. A model variance comparison regarding the calibrations obtained with the NIR-AOTF and NIR-FT spectrophotometers showed the adequacy of the experimental equipment employed.

SUMÁRIO

LISTA DE ABREVIATURAS	xix
LISTA DE SÍMBOLOS	xxi
ÍNDICE DE TABELAS	xxiii
ÍNDICE DE FIGURAS	xxv
PREFÁCIO	01
CAPÍTULO 1 – Calibração Multivariada	07
1.0 – Introdução.....	19
1.1 – PLS.....	12
1.2 – Ferramentas para Avaliação do Desempenho de Modelos de Calibração.....	15
1.2.1 – RMSEC.....	15
1.2.2 – RMSEP.....	16
1.2.3 – RMSECV.....	17
1.2.4 – Escolha do melhor modelo de calibração.....	17
1.3 – Técnicas de Pré-tratamento de Dados.....	19
1.3.1 – Suavização por média-móvel.....	19
1.3.2 – Suavização Savitz-Golay.....	19
1.4 – Ferramentas para Detecção de Amostras Anômalas (Outliers).....	20
1.4.1 – Alavancagem (<i>Leverage</i>).....	20
1.4.2 – Resíduo de Student.....	21
CAPÍTULO 2 – Figuras analíticas de mérito	23
2.0 – Introdução.....	25
2.1 – Figuras de Mérito em Calibração Univariada - breve histórico.....	26
2.2 – Notações e Convenções.....	32
2.3 – Validação em Calibração Multivariada.....	32
2.4 – Surgimento do Sinal Analítico Líquido.....	33
2.5 – Figuras de Mérito em Calibração Multivariada.....	43
2.5.1 – Limite de detecção.....	43
2.5.2 – Limite de quantificação.....	44

2.5.3 – Sensibilidade.....	45
2.5.4 – Sensibilidade Analítica.....	45
2.5.5 – Seletividade.....	46
2.5.6 – Relação Sinal/Ruído.....	46
2.5.7 – Exatidão.....	47
2.5.8 – Precisão.....	47
2.5.8.1 – Repetibilidade.....	48
2.5.8.2 – Precisão intermediária.....	48
2.5.8.3 – Reprodutibilidade.....	48
2.5.9 – Faixa linear de trabalho.....	49
2.5.10 – Robustez.....	51
CAPÍTULO 3 – Espectroscopia no infravermelho próximo.....	53
3.0 – Introdução.....	55
3.1 – Fundamentos.....	56
3.2 – Vibrações Moleculares.....	59
3.2.1 – Modos vibracionais.....	62
3.3 – Medidas de Absorbância.....	65
3.4 – Instrumentação.....	66
3.4.1 – Espectrofotômetros Dispersivos.....	67
3.4.2 – Espectrofotômetros por Transformada de Fourier.....	67
3.4.3 – Espectrofotômetros com filtros ópticos.....	68
3.4.3.1 – Filtro óptico-acústico sintonizável.....	68
3.5 – Medidas no Infravermelho Próximo de Amostras Gasosas.....	69
CAPÍTULO 4 – Parte Experimental.....	73
4.0 – Equipamento Experimental NIR-AOTF.....	75
4.1 – Sistema de Preparação das Amostras Gasosas.....	76
4.2 – Planejamento de Misturas.....	80
CAPÍTULO 5 – Calibração de espécies presentes em amostras sintéticas de gás natural.....	83
5.0 – Resultados e Discussão.....	85
5.1 – Validação Externa.....	98
5.2 – Figuras de Mérito.....	101

CONCLUSÃO	105
APÊNDICE	109
REFERÊNCIAS BIBLIOGRÁFICAS	119

LISTA DE ABREVIATURAS

ANVISA	-	Agência Nacional de Vigilância Sanitária
ANP	-	Agência Nacional de Petróleo
AOAC	-	Association of Official Analytical Chemists
AOTF	-	Acousto Optical Tunable Filter
ASTM	-	American Society for Testing and Materials
ATR	-	Attenuated Total Reflectance
CLS	-	Classical Least Squares
FDA	-	Food and Drug Administration
FIR	-	Far Infrared
GLP	-	Gás Liquefeito de Petróleo
HLA	-	Hybrid Linear Analysis
INMETRO	-	Instituto Nacional de Metrologia, Normalização e Qualidade Industrial
ISO	-	International Organization for Standardization
IUPAC	-	International Union of Pure and Applied Chemistry
LD	-	Limit of Detection
LQ	-	Limit of Quantification
MIR	-	Mid Infrared
MLR	-	Multiple Linear Regression
NAS	-	Net Analyte Signal
NIPALS	-	Nonlinear Iterative Partial Least Squares
NIR	-	Near Infrared
NIR-AOTF	-	Near Infrared – Acousto Optical Tunable Filter
NIR-FT	-	Fourier Transform Near Infrared
PCA	-	Principal Component Analysis
PCR	-	Principal Component Regression
PLS	-	Partial Least Squares
PTFE	-	Politetrafluoretileno
RMSEC	-	Root Mean Squared Error of Validation

RMSECV	-	Root Mean Squared Error of Cross Validation
RMSEP	-	Root Mean Squared Error of Prediction
SVD	-	Singular Value Decomposition
USP	-	United States Pharmacopeia
UV-VIS	-	Ultravioleta - Visível
TC	-	Technical Comittee

LISTA DE SÍMBOLOS

α	-	Probabilidade de ocorrência de erros do tipo I
β	-	Probabilidade de ocorrência de erros do tipo II
γ	-	Sensibilidade analítica
δ	-	Deformação angular no plano
δ_s	-	Deformação angular do tipo <i>scissoring</i> , simétrica.
δ_{as}	-	Deformação angular do tipo assimétrica.
Δ	-	Parâmetro de não-centralidade
Δv_q	-	Diferença entre dois números quânticos vibracionais
ϵ	-	Coefficiente de absorvidade molar ($L \cdot mol^{-1} \cdot L^{-1}$)
λ	-	Comprimento de onda
μ	-	Momento de dipolo
v_i	-	Número de graus de liberdade do sistema “i” (i=1,2,3...)
v_q	-	Número quântico vibracional
v_m	-	Frequência de vibração do sistema “m”.
v_{as}	-	Estiramento assimétrico
v_s	-	Estiramento simétrico
ζ	-	Termo normalizador de concentração.
ξ	-	Seletividade
π	-	Constante matemática pi (= 3,141...)
ρ	-	Deformação angular no plano do tipo <i>rocking</i> .
σ	-	Desvio-padrão populacional
τ	-	Deformação angular fora do plano do tipo <i>twisting</i> .
\emptyset	-	Massa reduzida
ψ	-	Vetor indicativo da direção de maior correlação
ω	-	Deformação angular fora do plano do tipo <i>wagging</i> .
e	-	Constante matemática euler (=2,718...)
∂	-	Derivada parcial
Σ	-	Função somatório
H_A	-	Hipótese alternativa

- H_0 - Hipótese nula
- \perp - Ortogonal

ÍNDICE DE TABELAS

Tabela 01	-	Regiões do espectro infravermelho.....	55
Tabela 02	-	Principais modos vibracionais das compostos utilizados neste trabalho.....	64
Tabela 03	-	Faixas de valores das razões volumétricas empregadas nos conjuntos de calibração.....	80
Tabela 04	-	Conjunto de calibração final, C3.....	82
Tabela 05	-	Tabela comparativa dos modelos PLS-1 para as 6 espécies químicas.....	86
Tabela 06	-	Tabela comparativa dos modelos PLS-1 para as 7 espécies químicas (DADOS OBTIDOS COM O EQUIPAMENTO EXPERIMENTAL BASEADO NO FILTRO AOTF – 40.....	89
Tabela 07	-	Tabela comparativa dos modelos PLS-1 para as 7 espécies químicas (DADOS OBTIDOS COM O EQUIPAMENTO COMERCIAL BOMEM® - 40 AMOSTRAS).....	96
Tabela 08	-	Resultado da validação externa para o metano CONJUNTO AOTF - 30 amostras de calibração, 10 de previsão.....	99
Tabela 09	-	Resultado da validação externa para o etano CONJUNTO AOTF - 30 amostras de calibração, 10 de previsão.....	99
Tabela 10	-	Resultado da validação externa para o propano CONJUNTO AOTF - 30 amostras de calibração, 10 de previsão.....	99
Tabela 11	-	Resultado da validação externa para o butano CONJUNTO AOTF - 30 amostras de calibração, 10 de previsão.....	99
Tabela 12	-	Resultado da validação externa para o dióxido de carbono CONJUNTO AOTF - 30 amostras de calibração, 10 de previsão.....	100
Tabela 13	-	Resultado da validação externa para o 1-buteno CONJUNTO AOTF - 30 amostras de calibração, 10 de previsão.....	100
Tabela 14	-	Resultado da validação externa para o 1,3-butadieno CONJUNTO AOTF - 30 amostras de calibração, 10 de previsão.....	100
Tabela 15	-	Figuras de mérito para os modelos PLS-1 apresentados nas Tabelas 8 a 14.....	102

ÍNDICE DE FIGURAS

Figura 01	-	Gráfico de avaliação do número ótimo de variáveis latentes de um modelo de calibração do tipo PLS18	18
Figura 02	-	<p>a) distribuição probabilística para os resultados dos ensaios em branco e o valor inicialmente definido como limite de detecção, tendo a probabilidade de ocorrência do erro do tipo I, $\alpha/2$, ilustrada em cinza;</p> <p>b) representação em conjunto da distribuição de probabilidade do ensaio em branco e de uma amostra contendo o analito em concentração próxima ao limite de detecção anteriormente definido, assim como as probabilidades de erros do tipo I ($\alpha/2$) e do tipo II ($\beta/2$), além do novo limite de detecção ($\bar{y}_{br} + 6\sigma$).29</p>	29
Figura 03	-	Correspondência através da curva de calibração univariada dos valores do ensaio em branco e do limite de detecção no domínio da concentração.....30	30
Figura 04	-	<p>a) representação geométrica tradicional do conceito do vetor NAS.</p> <p>b) representação de um caso hipotético onde apenas 3 comprimentos de onda são registrados por espectro.....34</p>	34
Figura 05	-	Sistema massa-mola fundamental.....56	56
Figura 06	-	Diagrama de energia potencial para um oscilador harmônico.59	59
Figura 07	-	<p>Diagramas de energia potencial</p> <p>a) oscilador harmônico quantizado.</p> <p>b) oscilador anarmônico, de = distância de equilíbrio, Ed = energia de dissociação.....60</p>	60

Figura 08	-	Atenuação da radiação monocromática ao atravessar um material absorvente de espessura b (caminho óptico). Neste caso, $P_0 > P$, onde P_0 e P são respectivamente a intensidade inicial e final do feixe de radiação.....	65
Figura 09	-	Componentes comumente encontrados em equipamentos utilizados para medidas espectrofotométricas.....	66
Figura 10	-	Monocromador do tipo AOTF.....	69
Figura 11	-	Equipamento experimental NIR-AOTF.....	75
Figura 12	-	Sistema de preparo das misturas gasosas.....	78
Figura 13	-	Peça evaporadora de acrílico.....	79
Figura 14	-	Configuração para utilização da célula de evaporação. O símbolo " E " representa a célula de evaporação e " B " representa uma bomba utilizada para o refluxo da amostra gasosa. " X ", Indica que as válvulas de entrada e saída da célula de medida permanecem fechadas durante a etapa de evaporação.....	79
Figura 15	-	Configuração para utilização a obtenção simultânea no equipamento NIR-AOTF e no NIR-FT (Bomem®, Camo®).....	79
Figura 16	-	Espectros para o conjunto de calibração com 85 amostras, obtidos com espectrofotômetro experimental baseado em AOTF.....	86
Figura 17	-	Espectros para o conjunto de calibração com 40 amostras, obtidos com espectrofotômetro experimental baseado em AOTF.....	87
Figura 18	-	Variação do valor de RMSECV com a aplicação progressiva do pré-tratamento alisamento por médio móvel.....	88
Figura 19	-	Erro relativo na validação cruzada para o metano CONJUNTO AOTF - 40 amostras de calibração.....	91
Figura 20	-	Erro relativo na validação cruzada para o etano CONJUNTO AOTF - 40 amostras de calibração.....	91

Figura 21	-	Erro relativo na validação cruzada para o propano CONJUNTO AOTF - 40 amostras de calibração	91
Figura 22	-	Erro relativo na validação cruzada para o butano CONJUNTO AOTF - 40 amostras de calibração	91
Figura 23	-	Erro relativo na validação cruzada para o dióxido de carbono CONJUNTO AOTF - 40 amostras de calibração	92
Figura 24	-	Erro relativo na validação cruzada para o 1-buteno CONJUNTO AOTF - 40 amostras de calibração	92
Figura 25	-	Erro relativo na validação cruzada para o 1,3-butadieno CONJUNTO AOTF - 40 amostras de calibração	92
Figura 26	-	Espectros para o conjunto de calibração com 40 amostras, obtidos com espectrofotômetro comercial Bomem®.....	94
Figura 27	-	Variação do valor de RMSECV com a aplicação progressiva do pré-tratamento alisamento por médio móvel.....	95
Figura 28	-	À esquerda tem-se os espectros obtidos com o espectrofotômetro experimental baseado em AOTF (Figura 28A); à direita tem-se os espectros obtidos com equipamento comercial da marca Bomem®, após suavização por janela móvel de 9 pontos (Figura 28B).....	97
Figura 29	-	Projeção do vetor de y no espaço-coluna (plano) da matriz X	112

PREFÁCIO

O crescente desenvolvimento dos métodos matemáticos, fundamentados em axiomas estatísticos e nos princípios da álgebra linear, de amplo uso no universo multivariado têm se mostrados extremamente adequados, versáteis e robustos em diversos campos do conhecimento. As técnicas de análise multivariada apesar de utilizadas há tempos na matemática somente nas últimas décadas tiveram uma expansão considerável para as demais áreas do conhecimento. Nesse sentido, observam-se estudos em sistemas econômicos, em segmentos das ciências humanas como a psicologia e diversos avanços em problemas de domínio químico^[1]. E devido ao seu enorme potencial vêm a cada momento se mostrando mais presentes na comunidade científica e no setor produtivo.

Métodos de compressão de dados e análise exploratória, como a análise de componentes principais (*PCA*, do inglês *Principal Component Analysis*), têm sido extensivamente empregados e determinações de espécies químicas em amostras complexas e sujeitas a interferentes

são constantemente realizadas mediante a aplicação da calibração multivariada. Métodos como a regressão por componentes principais (*PCR*, do inglês *Principal Component Regression*) ou a regressão por mínimos quadrados parciais (*PLS*, do inglês *Partial Least Squares*) estão entre os mais difundidos. Metodologias de cálculo como o PLS permitem desde a calibração de inúmeras espécies químicas com informações sobrepostas à possibilidade do uso da espectroscopia no infravermelho próximo para quantificação de diversos compostos sem a tradicional atribuição de bandas no espectro.

Todavia, assim como nos métodos de determinação univariados, os resultados obtidos com as ferramentas quimiométricas precisam, de alguma forma, expressar o quão confiáveis estes podem ser. Análises clínicas, biomédicas ou farmacêuticas estão todas sujeitas a validações e devem conter informações cruciais como o grau de confiança do resultado, a exatidão e precisão do procedimento, além dos limites de detecção aos quais o processo utilizado está sujeito. Toda nova metodologia analítica necessita ser validada para ser passível de comparação global e assim ser classificada de acordo com seu "desempenho". Órgãos de fiscalização e instituições metrológicas nacionais e internacionais como *Food and Drug Administration (FDA)*, *United States Pharmacopeia (USP)*, *American Society for Testing and Materials (ASTM)*, Instituto Nacional de Metrologia, Normalização e Qualidade Industrial (INMETRO) ou a Agência Nacional de Vigilância Sanitária (ANVISA) impõem normas de aceitação referentes aos mais diversos resultados, sejam eles químicos, biológicos, médicos, sanitários ou de qualquer natureza e por sua vez acabam discriminando ou selecionando o procedimento que mais atende às suas exigências com base em seus parâmetros de desempenho. Desta forma mesmo com resultados notáveis, os procedimentos analíticos baseados nas ferramentas quimiométricas devem ter seus parâmetros de desempenho, usualmente conhecidos por figuras de mérito, determinados.

Observa-se durante as 2 últimas décadas que uma maior atenção vem sendo direcionada ao computo das figuras de mérito multivariadas^[2-16]. Na literatura encontram-se propostas para o cálculo de intervalos de confiança, sinal

analítico líquido, limite de detecção, grau de superposição espectral entre outras; não estando, portanto, estabelecidas e sendo alvo de diversos estudos.

O crescente uso das ferramentas quimiométricas, o qual está fundamentalmente associado à evolução computacional, também tem possibilitado um grande avanço nos métodos instrumentais de análise, bem como um expressivo desenvolvimento da automação em laboratórios. A indústria, em especial, vem se favorecendo em grande parte com a implementação de sistemas de monitoramento in-line e on-line permitindo a determinação de inúmeros parâmetros simultaneamente. Equipamentos dedicados que possibilitam o acompanhamento dos processos diretamente na linha de produção, em tempo real e dispensando a necessidade de um analista de laboratório são o grande desejo do setor produtivo. Neste contexto encontramos também analisadores portáteis, tecnologia bastante presente e em constante desenvolvimento em indústrias farmacêuticas, nas forças armadas e em órgãos de segurança pública. Atualmente podem ser encontrados espectrofotômetros NIR-FT^[17, 18], Raman^[17, 19] e UV-VIS^[20] que realizam as medidas de forma não invasiva, não destrutiva e extremamente rápida, de amostras líquidas, sólidas, pastosas, resinas e até mesmo gasosas, sejam elas materiais inertes, contaminantes ou explosivos. As metodologias de análise modernas utilizadas pela indústria, as quais estão em grande parte associadas ao desenvolvimento de equipamentos dedicados e que por sua vez estão estruturados majoritariamente em técnicas espectroscópicas e métodos quimiométricos constituem um campo de pesquisa de grande potencial para os próximos anos.

Neste cenário, o gás natural se apresenta como um dos sistemas químicos carentes de implementação de metodologias de monitoramento mais ágeis e de alto processamento, devido à sua posição atual de destaque entre as principais fontes de energia da civilização moderna.

GÁS NATURAL

Trata-se de um combustível fóssil constituído em sua maior parte por hidrocarbonetos de baixa massa molar como o metano, etano, propano e butano; bastante voláteis e estruturalmente muito semelhantes. Em menores concentrações estão presentes dióxido de carbono, nitrogênio, vapor de água e compostos sulfurados, além de raras ocorrências de gases nobres como hélio e argônio. O gás natural encontra-se acumulado em rochas porosas no subsolo, isoladas do exterior por rochas impermeáveis, pode estar associado ou não a depósitos petrolíferos, quando é chamado de gás associado. É o resultado da degradação da matéria orgânica de forma anaeróbica.

Em vista das matrizes energéticas às quais a civilização moderna está estruturada pode-se dizer que o gás natural é um combustível moderno e seu uso em um primeiro momento predominantemente industrial concede espaço, há algum tempo, ao abastecimento residencial, por exemplo. Este combustível apresenta diversas aplicações, podendo substituir com eficiência qualquer combustível sólido, líquido ou gasoso^[21].

O gás natural encontra aplicações na indústria petroquímica, sendo usado como matéria-prima para a produção de metanol, fabricação de fertilizantes, para a síntese de amônia, uréia entre outros; no setor energético permite a geração e co-geração de energia elétrica através de motores de combustão interna. Também encontra emprego como combustível industrial sendo utilizado como substituinte de uma variedade de combustíveis alternativos como a madeira, carvão, óleos combustíveis, diesel e GLP. Em fornos pode ser utilizado para a fabricação de cerâmicas, vidro e cimento, pois proporciona uma combustão limpa e completa. Para o setor comercial e de prestação de serviços possui aplicações nos mais diversos negócios, como restaurantes, churrascarias, pizzarias, lanchonetes, cozinhas industriais, padarias, supermercados, academias, clubes, hotéis, hospitais, lavanderias, shoppings ou qualquer outro estabelecimento que tenha necessidade de cocção, aquecimento ou refrigeração^[21], finalmente como

combustível automotivo o gás natural já é a opção de milhares de brasileiros devido a benefícios como a economia e o rendimento proporcionados.

A legislação acerca da comercialização do gás natural^[22] estabelece diversos parâmetros de qualidade com relação ao produto final, tais como poder calorífico, índice de Wobb, número de metano, porcentagem de espécies como metano, etano, propano, butano, oxigênio, gases inertes entre outros. Para a verificação destes parâmetros de qualidade é necessário o emprego diferentes métodos e vários equipamentos para as medidas físico-químicas^[23, 24]. No caso da porcentagem das espécies gasosas a maioria das análises é realizada por cromatografia gasosa^[25-28], técnica com alto tempo de análise e baixa capacidade de processamento de amostras. Este fato tem levado algumas instituições de pesquisa a buscar novos métodos^[29] e a desenvolver novos equipamentos^[30] que possam substituir as atuais análises, uma vez que são relativamente morosas e necessitam de diferentes instrumentos.

Neste sentido, os métodos espectroscópicos surgem como uma alternativa bastante viável, existindo atualmente inúmeras pesquisas e aplicações contemplando a região espectral do infravermelho em equipamentos dedicados e de análise rápida. Esta faixa espectral tem se mostrado bastante adequada para o desenvolvimento de equipamentos tão versáteis que podem ser operados sobre uma simples bancada de laboratório ou em condições hostis, tais como em altas temperaturas e elevadas pressões^[31, 32].

CONTRIBUIÇÃO DESTE TRABALHO

Este estudo dá continuidade ao projeto que vem sendo desenvolvido pelo laboratório de automação em química analítica da Universidade Estadual de Campinas, cujo objetivo constitui no desenvolvimento de um espectrofotômetro NIR-AOTF dedicado ao monitoramento dos parâmetros de qualidade de gases combustíveis naturais e derivados do petróleo. O presente trabalho tem como foco avaliar o uso da espectroscopia NIR para a determinação do teor de 6 hidrocarbonetos e dióxido de carbono em misturas gasosas com composição

semelhante à do gás natural, bem como realizar a validação dos métodos de calibração. A validação foi realizada através do cálculo dos parâmetros de desempenho, também conhecidos como figuras de mérito, os quais foram obtidos a partir das propostas presentes na literatura. Este sistema ilustra uma necessidade real da indústria nacional de petróleo com relação à otimização e automatização dos métodos de controle dos parâmetros de qualidade do gás natural.

Esta dissertação sintetiza todo o trabalho realizado ao longo de 2 anos e está estruturada em 5 capítulos, conclusão, apêndices e bibliografia. O primeiro capítulo apresenta os cálculos e os fundamentos dos métodos de calibração multivariada utilizados neste estudo, além das ferramentas utilizadas na construção dos modelos de calibração. O embasamento utilizado para realizar a validação dos modelos de calibração é apresentado no segundo capítulo, o qual contém a maioria das figuras de mérito atualmente discutidas na literatura. O terceiro capítulo é reservado para introdução da espectroscopia na região do infravermelho próximo. Neste capítulo é feita uma descrição a respeito da técnica, mostrando um panorama da espectroscopia NIR atual, apresentando os fundamentos da técnica e ilustrando em maiores detalhes o equipamento previamente construído e utilizado neste trabalho. Destaque maior é dado ao dispositivo seletor de comprimento de onda do tipo AOTF (do inglês, *Acousto Optical Tunable Filter*). No quarto capítulo, os métodos e procedimentos experimentais, além do sistema de preparação de amostras são descritos e no quinto é último capítulo os resultados e avanços obtidos durante todo o estudo são apresentados e discutidos. Por fim, as conclusões desta etapa são registradas e apresentadas ao leitor. Algumas informações relevantes e cálculos mais longos são deixados como informação complementar e desta forma são apresentados nos apêndices.

Capítulo

1

CALIBRAÇÃO MULTIVARIADA

1.0 - INTRODUÇÃO

O crescente uso de métodos multivariados na química e em particular na química analítica culminou no nascimento de uma área de pesquisa denominada quimiometria. Segundo Gemperline^[86], o termo foi criado em 1971 para descrever este novo campo do conhecimento de caráter interdisciplinar que envolve a estatística multivariada, a aplicação de modelos matemáticos, ferramentas da ciência da computação e a química analítica.

Os segmentos químicos com as principais aplicações são a calibração, a validação de métodos experimentais, os teste de hipóteses, planejamentos experimentais e a otimização de experimentos. Visando na maioria dos casos a extração da maior quantidade de informação útil dos dados obtidos e, sendo assim, fazem uso de diversas variáveis simultaneamente.

Neste universo multivariado a calibração possui grande destaque. O processo de calibração consiste na obtenção e validação de um modelo matemático que relaciona a informação fornecida por um equipamento com alguma propriedade da amostra sob investigação.

Para a construção de um modelo de calibração em espectroscopia, por exemplo, amostras com valores de concentração conhecidos são analisadas por um espectrofotômetro. A informação obtida através do instrumento, neste caso valores de absorbância, junto com os valores de concentração são utilizados para o cálculo dos parâmetros de calibração do modelo matemático. Em seguida o modelo é validado e finalmente utilizado na etapa de previsão, onde os valores de concentração de amostras desconhecidas são estimados após medirem-se os valores de absorbância das mesmas.

Quando, apenas o valor de absorbância em um único comprimento de onda é utilizado para estabelecer-se a relação entre a concentração e a resposta instrumental, tem-se a univariada ou de ordem zero. Todavia esta abordagem é falha quando há espécies interferentes na amostra, pois o sinal instrumental deve ser específico para a propriedade de interesse. A calibração de ordem zero também possui desvantagens como o fato de ser necessário conhecer-se previamente a relação matemática entre a resposta instrumental e a concentração.

Ou seja, ao variar-se a concentração a absorvância deve modificar-se em um padrão conhecido (de forma linear, exponencial, logarítmica...). Além disso, a medida que o ruído instrumental aumenta as previsões perdem qualidade, o que no caso multivariado é atenuado pela vantagem multicanal. Tais restrições fazem com que a calibração univariada tenha seu uso limitado a sistemas mais simples.

Nestes casos, apesar da sofisticação matemática dos métodos quimiométricos, o que resulta em ferramentas mais complexas, os métodos multivariados se mostram superiores exatamente onde a calibração univariada é falha.

Historicamente, observa-se primeiramente o emprego da regressão linear múltipla (MLR, do inglês *Multiple Linear Regression*), em vista da maior robustez e maior exatidão dos resultados devido à vantagem multicanal. Apesar dos benefícios, esta abordagem possui sua utilização limitada devido à condição matemática relativa à quantidade de variáveis utilizadas, para o emprego do MLR o número de variáveis deve ser menor ou igual ao número de amostras. Além disso, as variáveis devem ter baixa correlação caso contrário o problema de calibração pode não ter solução. Matematicamente diz-se que a matriz $\mathbf{X}^T\mathbf{X}$, derivada da equação matricial 1.0, é singular e não pode ser invertida.

$$\text{Sistema linear de equações} \quad \mathbf{y} = \mathbf{Xb} \quad (1.0)$$

$$\begin{array}{l} \text{Quando há baixa correlação e o } n^\circ \text{ de variáveis } \leq n^\circ \text{ amostras} \\ \text{Solução pelo método MLR} \quad \mathbf{b} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \end{array} \quad (1.1)$$

$$\begin{array}{l} \text{Quando há correlação significativa ou } n^\circ \text{ variáveis } > n^\circ \text{ amostras} \\ \text{Sistema indeterminado} \quad (\mathbf{X}^T\mathbf{X})^{-1} = \text{não existe} \end{array} \quad (1.2)$$

Nesse contexto, as metodologias de compressão de dados surgem para contornar esse problema. Nestes casos, a matriz \mathbf{X} sofre um rearranjo. Novos eixos, chamados de *componentes principais*, são definidos de forma a compilar a maior parcela possível da variância dos dados nas primeiras componentes

principais. Desta forma, a informação relevante é condensada nestes novos eixos, fazendo com as componentes principais expressem a informação anteriormente dispersa em um número bem maior de variáveis.

Geralmente após a etapa de compressão de dados, o número de variáveis no novo sistema de componentes principais é menor que o número de amostras, além de serem variáveis independentes. A nova matriz gerada é, portanto passível de inversão, havendo assim solução para o problema de calibração. Este método é conhecido como Regressão por Componentes Principais (PCR, do inglês *Principal Components Regression*).

$$\text{Sistema} \quad \mathbf{y} = \mathbf{Xb} \quad (1.3)$$

$$\text{Após a compressão de dados} \quad \mathbf{X} = \mathbf{TP}^T \quad (1.4)$$

$$\text{Novo sistema} \quad \mathbf{y} = \mathbf{TP}^T\mathbf{b} \quad (1.5)$$

$$\text{Solução pelo PCR} \quad \hat{\mathbf{b}} = \mathbf{P}(\mathbf{T}^T\mathbf{T})^{-1}\mathbf{T}^T\mathbf{y} \quad (1.6)$$

Onde, $\hat{\mathbf{b}}$ é a estimativa do vetor de regressão.

Existem diferentes algoritmos matemáticos para a realização da etapa representada na equação 1.4, o mais popular é o NIPALS (do inglês, *Non-linear Iterative Partial Least Squares*).

A compressão de dados é realizada visando reorientar o sistema de forma a criar novos eixos, ao longo dos quais a variância dos dados é a máxima possível.

O cálculo gera duas matrizes \mathbf{T} e \mathbf{P} , a primeira é a matriz de *scores* e representa os antigos vetores no novo sistema de eixos; e a segunda é chamada de matriz de *loadings* (ou pesos), pois atua de forma ponderada na matriz \mathbf{T} combinando as colunas de forma linear para gerar os antigos vetores. A matriz de *loadings* pode ser entendida como uma matriz de transformação.

A calibração pelo método PCR funciona considerando-se que a variância do conjunto de dados é devida unicamente à propriedade de interesse. No âmbito da espectroscopia, poderia-se dizer que a variação nos valores de absorvância é devida unicamente à variação na concentração do analito. Desde que este postulado seja cumprido o método funciona como esperado.

A leitura dos tópicos sobre os métodos PCR e PLS nas referências números 86^[86] e 87^[87] sugere que, dependendo do nível de ruído instrumental e de outras fontes de variação presentes no conjunto de dados, os resultados do método PCR não serão satisfatórios ou mesmo válidos, pois a variância evidenciada na etapa de obtenção das componentes principais pode não estar relacionada à propriedade de interesse. Nesse sentido, a abordagem dos Mínimos Quadrados Parciais (PLS, do inglês *Partial Least Squares*), possui avanços que garantem que a informação modelada esteja diretamente associada à propriedade de interesse.

Ao contrário do PCR, onde o critério para o cálculo da nova matriz é a maximização da variância ao longo dos novos eixos, no método PLS a nova matriz é construída de forma que a correlação entre os *scores* e o vetor com os valores da propriedade de interesse (**y**), seja maximizada. Ou seja, a informação presente nos novos eixos, chamados neste caso de *variáveis latentes*, está correlacionada unicamente à propriedade de interesse. Todavia, outros autores parecem não concordar com as conclusões acima o que demonstra que não há um consenso a respeito de qual método realmente propicia os melhores resultados.

1.1 – Método PLS

Esta metodologia foi desenvolvida por Herman Wold no campo da economia no final dos anos 1960. Posteriormente, Harold Martens e Swante Wold difundiram o método com aplicações em sistemas químicos^[86].

O método é extremamente eficaz em situações onde a resposta instrumental não é específica para a espécie química de interesse, pois ainda assim consegue extrair a informação necessária permitindo estabelecer a relação entre a resposta instrumental e a propriedade de interesse.

O método faz uso da vantagem multicanal que surge quando utilizam-se duas ou mais variáveis correlacionadas, desta forma mesmo quando as variáveis sofrem interferência a informação pode ser extraída. Isso ocorre, pois a informação redundante presente nas diversas variáveis utilizadas tornam o cálculo robusto. Nesse sentido, a informação redundante atua de forma a minimizar as

possíveis distorções causadas por interferentes, uma vez que todas as variáveis contribuem com a informação para o cálculo do modelo, já no caso univariado o modelo é construído com base em uma única variável deixando-o bem mais vulnerável quando o ruído instrumental é considerável.

A etapa de compressão de dados presente neste método também funciona como um filtro, eliminando parte do ruído instrumental. Algebricamente a nova matriz contém o mesmo número de variáveis da matriz original, mas o cálculo realizado permite condensar toda a informação do sistema em algumas poucas variáveis latentes. A matriz original pode então ser reconstruída apenas com as novas variáveis latentes, fazendo com que outras fontes de variação, geralmente ruído, sejam minimizadas.

Para a aplicação da metodologia PLS existem diferentes algoritmos como o SVD, SIMPLS e o NIPALS. Para o sucesso do cálculo, os dados também devem ser submetidos a algum tipo de pré-processamento. Neste trabalho os dados originais foram “centrados na média”. Este processo consiste em determinar o valor médio da 1ª variável da matriz de dados e em seguida subtraí-lo de cada um dos valores individuais daquela variável. O procedimento é repetido com todas as demais variáveis. De forma análoga ao PCR, no método PLS a matriz de dados também é decomposta em duas matrizes, de *scores* e *loadings* respectivamente (equação 1.4).

Considerando que os dados foram centrados na média, a forma geral^[88] dos algoritmos não iterativos, como o SVD e o SIMPLS, é descrita abaixo, o cálculo inicializa com $j = 1$, $\mathbf{X}_1 = \mathbf{X}$ e $\mathbf{y}_1 = \mathbf{y}$ e prossegue até o número de variáveis latentes de interesse “ g ”.

$$1. \mathbf{w}_j = \frac{\mathbf{X}_j^T \mathbf{y}_j}{\|\mathbf{X}_j^T \mathbf{y}_j\|} \quad (1.7)$$

$$2. \mathbf{t}_j = \mathbf{X}_j \mathbf{w}_j \quad (1.8)$$

$$3. q_j = \frac{\mathbf{t}_j^T \mathbf{y}_j}{\mathbf{t}_j^T \mathbf{t}_j} \quad (1.9)$$

$$4. \boldsymbol{\rho}_j = \frac{\mathbf{X}_j^T \mathbf{t}_j}{\mathbf{t}_j^T \mathbf{t}_j} \quad (1.10)$$

$$5. \mathbf{X}_{j+1} = \mathbf{X}_j - \mathbf{t}_j \boldsymbol{\rho}_j^T \quad \text{e} \quad \mathbf{y}_{j+1} = \mathbf{y}_j - \mathbf{t}_j \hat{c}_j \quad (1.11)$$

O processo finaliza quando $j = g$, caso contrário repetem-se as etapas acima.

A etapa 01 (equação 1.7) consiste na determinação da direção de maior co- variação entre a propriedade de interesse y e os dados experimentais. Esta direção é dada pelo vetor \mathbf{w} , o qual é obtido pelo cálculo da covariância ($\mathbf{X}^T \mathbf{y}$) entre o vetor \mathbf{y} e as todas as variáveis em \mathbf{X} . O denominador, $\|\mathbf{X}_j^T \mathbf{y}_j\|$, serve unicamente para normalizar \mathbf{w} transformando-o em um vetor unitário, visto que a informação desejada é apenas a direção do vetor.

O vetor de *scores* \mathbf{t}_1 é obtido de partindo-se da equação 1.4, partindo-se da premissa que a matriz de *loadings* é (na realidade está sendo condicionada a ser) ortonormal.

$$\mathbf{X} = \mathbf{TP}^T \quad (1.4)$$

$$\mathbf{X} = \mathbf{t}_i \mathbf{p}_i^T \quad (1.12)$$

$$\mathbf{X} \mathbf{p}_i = \mathbf{t}_i \mathbf{p}_i^T \mathbf{p}_i \quad (1.13)$$

$$\mathbf{X} \mathbf{p}_i = \mathbf{t}_i \quad (1.14)$$

O vetor \mathbf{w} atua como um vetor de pesos, ponderando a participação das variáveis (as quais estão dispostas em colunas) de forma a orientar o vetor \mathbf{t}_1 na direção de maior correlação. O vetor \mathbf{t}_1 obtido é desta forma uma combinação linear das colunas da matriz de dados. Na etapa 3 (equação 1.9), é realizada uma regressão linear ($\mathbf{y} = \mathbf{t}_1 q$) entre \mathbf{y} e \mathbf{t}_1 e o coeficiente de regressão q obtido.

De posse do vetor \mathbf{t}_1 , o vetor de *loadings*, \mathbf{p}_1 , é obtido manipulando a equação 1.4 de forma semelhante à sequência iniciada na equação 1.12. Todavia, sem assumir que os *scores* são normalizados.

$$\mathbf{X} = \mathbf{t}_i \mathbf{p}_i^T \quad (3.12)$$

$$\mathbf{X}^T = \mathbf{p}_i \mathbf{t}_i^T \quad (1.15)$$

$$\mathbf{X}^T \mathbf{t}_i (\mathbf{t}_i^T \mathbf{t}_i)^{-1} = \mathbf{p}_i \mathbf{t}_i^T \mathbf{t}_i (\mathbf{t}_i^T \mathbf{t}_i)^{-1} \quad (1.16)$$

$$\mathbf{X}^T \mathbf{t}_i (\mathbf{t}_i^T \mathbf{t}_i)^{-1} = \mathbf{p}_i \quad (1.17)$$

$$\frac{\mathbf{X}^T \mathbf{t}_i}{\mathbf{t}_i^T \mathbf{t}_i} = \mathbf{p}_i \quad (1.18)$$

No final da seqüência, reconstrói-se a matriz \mathbf{X} (denominada de \mathbf{X}_1) com a informação da 1ª variável latente ($\mathbf{X}_1 = \mathbf{t}_1 \mathbf{p}_1$) e subtrai-se da matriz \mathbf{X} original. Desta forma, extrai-se da matriz \mathbf{X} uma parcela (em geral a maior parte) da correlação entre \mathbf{X} e \mathbf{y} . A correlação residual é extraída nas etapas seguintes, utilizando-se as matrizes $\mathbf{X}_2, \mathbf{X}_3, \dots, \mathbf{X}_g$ nas etapas de 1 a 5.

Ao contrário do algoritmo NIPALS, onde o cálculo é iterativo, no caso do método PLS o algoritmo é recursivo. Ou seja, em cada etapa do PLS uma variável latente é extraída, essa nova variável é então utilizada para o cômputo da segunda variável latente e assim por diante. O processo termina após as “g” etapas. O mesmo não ocorre com o NIPALS, uma vez que uma variável latente é obtida mediante “n” iterações, não sabendo-se antecipadamente quantas iterações serão necessárias. Neste sentido, os algoritmos recursivos são computacionalmente mais atraentes devido ao menor número de etapas para atingir o mesmo objetivo.

1.2 - Ferramentas para avaliação do desempenho de modelos de calibração multivariada

1.2.1 – RMSEC

O valor do erro quadrático médio de calibração (do inglês *Root Mean Square Error of Calibration* ou RMSEC) descreve o grau de concordância entre os valores previstos pelo modelo de calibração e os valores verdadeiros utilizados no cômputo dos parâmetros do modelo.

O RMSEC é calculado de acordo com a equação 1.19.

$$\text{RMSEC} = \sqrt{\frac{1}{l-p} \sum_{i=1}^l (y_i - \hat{y}_i)^2} \quad (1.19)$$

Onde,

l : número total de amostras no conjunto de calibração.

p : número de parâmetros estimados no modelo de calibração.

y_i : concentração da amostra i .

\hat{y}_i : concentração da amostra i estimada pelo modelo de calibração.

No caso de modelos de calibração cujos dados sejam centrados na média, o conjunto de dados perde 1 grau de liberdade. Deste modo o número de graus de liberdade ($l - p$) presente na equação 1.19 passa a ser ($l - p - 1$).

Entretanto os valores de RMSEC em geral superestimam a capacidade de previsão do modelo de calibração, uma vez que as amostras utilizadas na etapa de previsão são as mesmas utilizadas para a calibração.

1.2.2 – RMSEP (ou RMSEV)

Para avaliar de forma autêntica os modelos de calibração, os mesmos devem ser submetidos a amostras externas, ou seja, a um conjunto de validação. As amostras deste conjunto são preparadas exclusivamente para a etapa de validação, são amostras independentes com a função de simular amostras reais as quais não foram utilizadas no cálculo do modelo de calibração.

A concordância entre as previsões do modelo de calibração para as amostras do conjunto de validação e os valores reais é mensurada através do cálculo do RMSEP (do inglês *Root Mean Square Error os Prediction*, em português erro quadrático médio de previsão)

$$\text{RMSEP} = \sqrt{\frac{1}{A} \sum_{a=1}^A (y_i - \hat{y}_i)^2} \quad (1.20)$$

Onde,

A: número de amostras no conjunto de validação.

1.2.3 - RMSECV

Esta abordagem é utilizada para estimar a capacidade de previsão de um modelo de calibração. Para um conjunto com i amostras, o processo consiste no cálculo de i modelos de calibração com $i - 1$ amostras em cada modelo, ou seja, para cada novo modelo calculado 1 das “ i ” amostras é deixada de fora, funcionando como uma amostra independente.

Na validação cruzada a amostra remanescente é usada como uma amostra de validação. Neste caso aplica-se a equação 1.20 para testar a capacidade de previsão do modelo em cada um dos “ i ” modelos gerados.

1.2.4 - Escolha do melhor modelo de calibração

O cálculo de um modelo de calibração do tipo PLS envolve a escolha do número de variáveis latentes que gera o modelo mais eficaz e robusto. O melhor modelo deve necessariamente gerar boas previsões tanto para as amostras de calibração quanto para as amostras do conjunto de validação, ou seja, os valores dos erros quadráticos médios de calibração e validação (RMSEC, RMSEV e RMSECV) devem estar em uma margem aceitável, além de serem num caso ideal estatisticamente iguais. A igualdade estatística pode ser avaliada aplicando-se um teste F, uma vez que ambas expressões podem ser consideradas como medidas de variância.

A escolha do melhor modelo pode ser alcançada calculando-se os valores de RMSEC e RMSECV para uma seqüência de modelos, os quais são calculados com uma quantidade crescente de variáveis latentes. Desta forma, o primeiro modelo é construído com apenas uma variável latente, em seguida o modelo é

recalculado com duas variáveis latentes, posteriormente com três variáveis latentes e o procedimento continua até a quantidade desejada.

Modelos com um elevado número de variáveis latentes em geral tendem a apresentar bons valores de RMSEC, todavia elevados valores de RMSECV. Estas constatações indicam o caso conhecido como *overfitting*. Nesta situação obtém-se um modelo de calibração super ajustado, o qual apresenta boas previsões na etapa de treinamento, mas elevados erros na validação externa.

Ao analisar-se graficamente o número de variáveis latentes versus os valores de RMSEC e RMSECV, em geral observa-se uma diminuição inicial acentuada, seguida de um decréscimo menos pronunciado, uma constância ou até mesmo um crescimento. O número de variáveis latentes correspondente ao melhor modelo é escolhido com base nesta espécie de ponto de inflexão, onde os valores dos erros deixam de cair de forma brusca e começam a apresentar valores mais estáveis. Pela informação apresentada na Figura 1, por exemplo, o melhor modelo provavelmente teria 6 variáveis latentes.

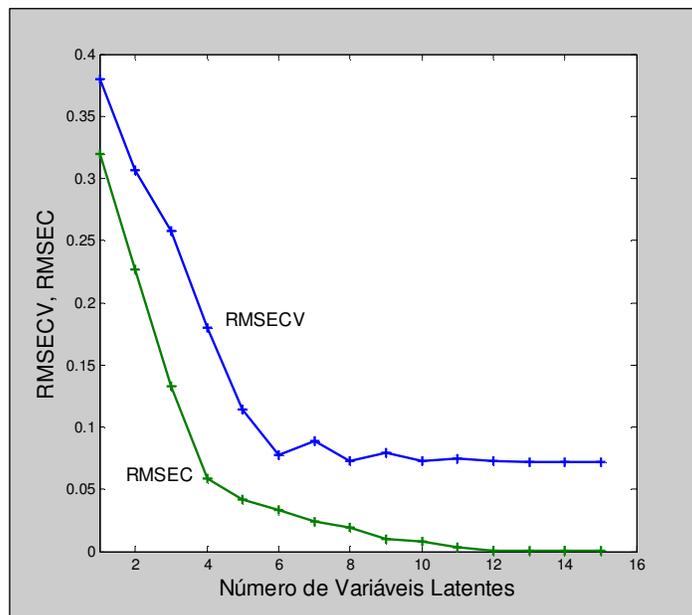


Figura 01 – Gráfico de avaliação do número ótimo de variáveis latentes de um modelo de calibração do tipo PLS.

1.3 – TÉCNICAS DE PRÉ-TRATAMENTO DE DADOS

Os resultados obtidos com métodos quimiométricos também podem ser melhorados controlando-se o nível de ruído no conjunto de dados antes do cálculo do modelo de calibração.

Nesse sentido o uso de ferramentas atenuadoras de ruído como a suavização de dados, filtro Savitz-Golay, Transformada de Fourier, Transformada de *Wavelet* entre outros, é de grande utilidade como em um pré-tratamento dos dados. Uma remoção de ruído eficiente é uma etapa crucial e presente em quase todos os métodos quimiométricos.

1.3.1 – Suavização por média móvel

A suavização por média móvel age de forma a atenuar as oscilações abruptas ou os valores isolados que sejam fortemente distoantes do conjunto de dados.

Em um espectro, por exemplo, a aplicação desta ferramenta proporciona uma considerável diminuição na oscilação da linha de base, evitando que valores mais elevados sejam considerados picos de absorção de baixa intensidade. Este procedimento também permite uma melhor definição dos picos ou bandas de absorção, permitindo uma visualização mais nítida no caso de picos sobrepostos.

Matematicamente esta técnica funciona calculando-se a média aritmética de um intervalo de dados, como os 5 primeiros valores de absorbância em um espectro, substituindo em seguida o valor central do intervalo pela média calculada. Este intervalo, tecnicamente chamado de janela de 5 pontos, percorre todo o espectro recalculando a média para a janela de 5 pontos subsequente e atribuindo o valor médio ao ponto central do intervalo. Desta forma a tendência do conjunto de dados é mantida, minimizando ou eliminando fortes oscilações ou saltos repentinos.

1.3.2 – Suavização Savitz-Golay

De forma análoga à suavização por média móvel, esta ferramenta também possibilita a atenuação do ruído, entretanto de forma mais elaborada.

A suavização Savitz-Golay realiza uma interpolação polinomial em um intervalo de valores escolhido pelo usuário. Neste caso tanto o grau do polinômio, quanto a dimensão do intervalo podem ser previamente selecionados.

1.4 - FERRAMENTAS PARA DETECÇÃO DE AMOSTRAS ANÔMALAS (*OUTLIERS*)

1.4.1 – Alavancagem (*Leverage*)

Este parâmetro indica o quão influente uma determinada amostra é sobre o conjunto. O valor de *leverage* é uma medida do distanciamento da amostra i em relação às $n-1$ amostras restantes.

Matematicamente o *leverage* para uma amostra i , h_i , é definido como o i -ésimo elemento da diagonal da matrix \mathbf{H} (equação 1.21), os quais assumem valores de 0 a 1.

$$\mathbf{H} = \mathbf{T}(\mathbf{T}^T\mathbf{T})^{-1}\mathbf{T}^T \quad (1.21)$$

As amostras mais distantes dos valores centrais da variável x em geral exercem maior influência sobre o conjunto e, portanto, possuem maiores valores de *leverage*.

O conceito é melhor compreendido visualizando-se a reta de obtida em um modelo de calibração univariada. De forma análoga ao *leverage*, os valores de concentração próximos ao valor de médio de concentração possuem pouca influência na inclinação da reta e desta forma são pouco eficazes para causar grandes distorções nos resultados do modelo. Já os valores localizados nos extremos da reta possuem maior poder de alavancagem e podem distorcer de forma apreciável as previsões realizadas pelo modelo de calibração.

Alguns autores recomendam que os pontos com valores de *leverage* maior que $2p/l$ ou $3p/l$ (onde “ p ” é o n° de variáveis latentes do modelo de calibração e “ l ” é o número total de amostras) devam ser tratados como amostras anômalas, *outliers*.

1.4.2 - Resíduo de *Student*

Uma das formas de identificação de amostras anômalas consiste em examinar os gráficos dos resíduos, $\hat{e}_i = y_i - \hat{y}_i$, em conjunto com os valores de *leverage*. Ao analisar tais gráficos, procura-se por amostras com elevados resíduos as quais são avaliadas para a confirmar-se se são realmente *outliers*.

Um problema comum nesta avaliação é que os resíduos das amostras próximas ao valor médio da variável x possuem variância mais elevada em relação às amostras nos valores extremos. Uma forma de resolver esta situação consiste na normalização dos resíduos, originando os chamados resíduos de *student* (ou *studentized residuals*), r_i .

$$r_i = \frac{\hat{e}_i}{\sigma_y \sqrt{1 - h_i}} \quad (1.22)$$

Onde,

r_i : resíduo de *student* da amostras i .

\hat{e}_i : resíduo da amostra i .

σ_y^2 : variância dos resíduos.

h_i : leverage da amostras i .

Capítulo

2

**FIGURAS
ANALÍTICAS
DE MÉRITO**

2.0 - INTRODUÇÃO

A necessidade de avaliar o desempenho de um novo método de estudo ou verificar a sua compatibilidade com a pesquisa e trabalho de rotina, além da comparação entre abordagens ou metodologias sistemáticas têm sido sempre uma preocupação e real imposição do estudo científico e controle de qualidade. A decisão a respeito de qual técnica a ser utilizada para a quantificação de um determinado composto químico, por exemplo, constitui-se a priori de uma análise comparativa entre os diversos parâmetros de desempenho dos métodos existentes. Questões como a magnitude da concentração esperada ou no caso de várias determinações, a faixa de concentração a ser empregada definem, fatores como limite de detecção, limite de quantificação e limite de linearidade os quais por sua vez definem a faixa dinâmica de trabalho. No intuito de estabelecer medidas comparativas entre procedimentos distintos, o analista deverá recorrer a informações como precisão, exatidão, reprodutibilidade, repetibilidade e robustez do método, medidas estas que viabilizam a decisão principal de escolha do processo a ser utilizado, envolvendo questões indispensáveis e sempre presentes como o custo agregado e benefício proporcionado.

Pode-se dizer que os parâmetros de desempenho dos métodos analíticos, formalmente referidos como figuras de mérito, têm sua existência justificada pela necessidade das agências reguladoras e órgãos de fiscalização, além da indústria, pela validação das metodologias empregadas, para que possam apresentar de forma criteriosa os argumentos que comprovam o desempenho e a aplicabilidade dos métodos em uso. Já no meio acadêmico as mesmas são utilizadas, em sua maioria, como uma informação de caráter complementar.

Apesar de sempre terem sido enfatizadas e discutidas, mesmo que apenas de forma qualitativa, o real movimento demonstrando a preocupação a respeito do cômputo das figuras é recente, uma vez que acompanhou o desenvolvimento tecnológico dos equipamentos usualmente presentes nos centros de pesquisa. Publicações por parte de organizações e institutos metrológicos como a International Union of Pure and Applied Chemistry (IUPAC) e a International

Organization for Standardization (ISO) datam do início da década 90^[33-35], sendo interessante notar a crescente preocupação durante este processo evolucionário que pode ser observada na citação do presidente, sr. Wilrich, P. Th., do comitê técnico TC 69 da *International Organization for Standardization - ISO (ISO/TC69/SC6 - Application of Statistical Methods)*, obtida na publicação de L.A. Currie^[36]. "*There is an urgent need for answers ... a growing number of regulations and recommendations of the European Community concerning limits... for trace constituents in food, water, air and soil... almost all lack of a reliable background (on) 'detection limits' "*.

2.1 - FIGURAS DE MÉRITO EM CALIBRAÇÃO UNIVARIADA, BREVE HISTÓRICO

A necessidade das figuras de mérito e o cálculo de alguns destes parâmetros pode ser reconhecido em livros de química analítica datados de 1947^a. Na obra *Análise Qualitativa*^[37] de Alekseev, a percepção da sensibilidade de um determinado analito é demonstrada através do estudo do que hoje se conhece por análise sistemática de cátions e ânions, onde a espécie de interesse é identificada por meio de reações com formação de soluções ou precipitados coloridos. A estes resultados associava-se o fato de que uma maior intensidade de cor para reações distintas, mas com compostos em concentrações equivalentes, correspondia a uma maior sensibilidade. A pesquisa destas reações qualitativas, tipicamente visuais, fez com que a linha de pensamento vigente apresentasse a sensibilidade como sendo caracterizada quantitativamente por dois índices, chamados de limite de sensibilidade e diluição limite.

O limite de sensibilidade era definido como a quantidade mínima de uma substância que podia ser detectada por meio de uma dada reação em condições pré-estabelecidas e era usualmente expresso em microgramas. Já a diluição limite consistia na mínima concentração da substância para qual os testes visuais de identificação ainda podiam ser realizados, sendo definida como 1/G, onde G representava a massa do solvente correspondente a 1 unidade de massa da substância pesquisada. Assim em uma reação como a formação do precipitado de

a: apesar das traduções relativamente recentes para a língua inglesa e portuguesa (de Portugal), a obra original em russo foi publicada na década de 40.

cromato de prata, estes índices eram obtidos por meio da repetição do experimento em concentrações progressivamente menores até que a coloração indicativa da formação do composto pudesse ou não ser observada. Ou seja, resultados positivos e negativos eram obtidos localizando-se desta forma região de dúvida do teste analítico.

Na época convencionou-se considerar como diluição limite a concentração da espécie de interesse para qual metade das reações realizadas dessem resultado positivos. Sendo assim partindo-se de uma solução que contém 1g/L de Ag^+ (1,57g de AgNO_3) pode-se verificar que a diluição máxima é de 25 vezes, logo a diluição limite é apresentada como 1:25000 (g/g) e o limite de sensibilidade, calculado por meio desta proporção, é de $0,8\mu\text{g}$ (considerando o volume de estudo de 1 gota como 0,02 mL).

Outras figuras de mérito como a seletividade são ainda puramente qualitativas e tratadas de forma apenas descritiva, utilizando-se termos como *relação limite* entre a concentração de um interferente e o analito na tentativa de expressar a influência do contaminante na resposta da espécie de interesse.

É interessante notar que a concentração mínima detectável, conhecida atualmente como limite de detecção, está inclusa na caracterização da sensibilidade sendo atualmente um parâmetro distinto. O limite de detecção, citado como diluição limite, para metodologias qualitativas como a análise sistemáticas de cátions e ânions se mantém o mesmo desde então^[38].

Com a introdução dos métodos estatísticos na química analítica, o rigor matemático passou a reger o cômputo das figuras de mérito. Surgem então parâmetros como precisão, exatidão e variações como a reprodutibilidade e a repetibilidade. Os resultados analíticos são apresentados com intervalos de confiança baseados na distribuição normal de probabilidade, na prática emprega-se a distribuição “*t*” de *student* inspirada na distribuição normal de probabilidade de Gauss. Neste cenário, os primeiros artigos sobre limites de detecção^[39, 40] sugerem então a *definição de limite de detecção* (representado aqui por LD) como sendo a concentração equivalente ao sinal de fundo, ou ruído, adicionado de 3

vezes o desvio-padrão^b do ensaio em branco (representado como σ_{br}). Desta forma o sinal correspondente ao limite de detecção era equivalente ao valor mais elevado de um intervalo de confiança, obtido com base na distribuição Gauss, e calculado para englobar 99,85% dos possíveis resultados para o ensaio em branco. No caso de uma relação linear entre a concentração e o sinal analítico, o limite de detecção seria expresso como:

$$\text{Sinal correspondente ao LD: } y_{LD} = (y_{br} + 3\sigma_{br}) \quad (2.0)$$

$$\text{Equação de calibração : } y = bx + a \quad (2.1)$$

$$\text{Logo, } y_{LD} = bx + a \quad (2.2)$$

$$(y_{br} + 3\sigma_{br}) = bx_{LD} + a \quad (2.3)$$

$$\text{LD: } x_{LD} = \frac{(y_{br} + 3\sigma_{br} - a)}{b} \quad (2.4)$$

Onde,

y_{LD} : Sinal correspondente ao limite de detecção.

y_{br} : sinal correspondente ao ensaio em branco.

x : valor de concentração.

σ_{br} : desvio-padrão de réplicas do ensaio em branco.

b : coeficiente angular da equação de calibração.

a : coeficiente linear da equação de calibração.

Posteriormente, analisou-se o fato de que apesar de existir somente 0,15% de chance dos resultados para o ensaio em branco se apresentarem na faixa de detecção do analito, constituindo o que convencionou-se chamar de um resultado *falso positivo*, havia também a chance de uma amostra contendo de fato o analito ter um sinal na faixa dos resultados possíveis para o ensaio em branco, ou seja um resultado *falso negativo*. Mais tarde estes conceitos foram aprimorados com a aplicação dos testes de hipótese, onde definiu-se um resultado falso positivo como um *erro do tipo I* e o resultado falso negativo como erro do tipo II. Estas novas

b: Neste caso deve ser obtido a partir de um elevado número de réplicas do ensaio em branco, geralmente acima de 30, de forma a englobar todas as fontes de variação possíveis.

definições derivam da formulação da hipótese nula (representada por H_0) e da hipótese alternativa (representada por H_A). No primeiro caso, H_0 , tem-se a comparação estatística entre 2 resultados supostamente idênticos de um mesmo parâmetro. Quando verifica-se que para um determinado nível de confiança (99,85%, por exemplo) a diferença é realmente zero, diz-se que a hipótese nula é verdadeira com 0,15% de probabilidade de ocorrência do erro do tipo I. Caso contrário, a hipótese alternativa é verdadeira, ou seja, os 2 valores são estatisticamente diferentes.

Sendo assim, ficou claro que a proposta do limite de detecção devia considerar um segundo intervalo de confiança, aquele referente a uma amostra contendo a espécie de interesse. Esta idéia está ilustrada na Figura-02.

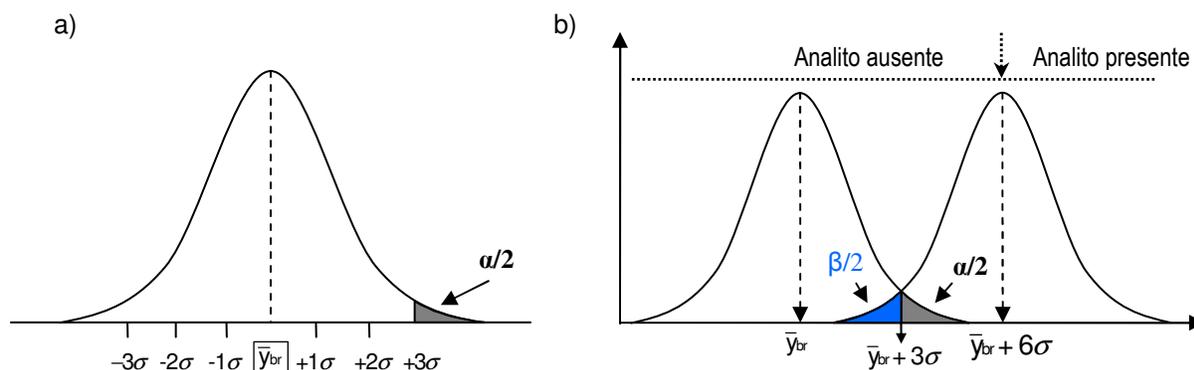


Figura 02 - a) distribuição probabilística para os resultados dos ensaios em branco e o valor inicialmente definido como limite de detecção, tendo a probabilidade de ocorrência do erro do tipo I, $\alpha/2$, ilustrada em cinza; b) representação em conjunto da distribuição de probabilidade do ensaio em branco e de uma amostra contendo o analito em concentração próxima ao limite de detecção anteriormente definido, assim como as probabilidades de erros do tipo I ($\alpha/2$) e do tipo II ($\beta/2$), além do novo limite de detecção ($\bar{y}_{br} + 6\sigma$).

Com a probabilidade de ocorrência de um erro do tipo II (resultado falso negativo) agora em foco, o limite de detecção proposto torna-se então uma função das probabilidades α e β , exemplificadas na Figura 02b. Em uma situação homocedástica, ou seja, quando a variância dos dados é constante ao longo da

faixa de trabalho, o sinal correspondente ao limite de detecção pode se calculado como: $\bar{y}_{br} + 6\sigma_{br}$, considerando $\alpha = \beta = 0,15\%$.

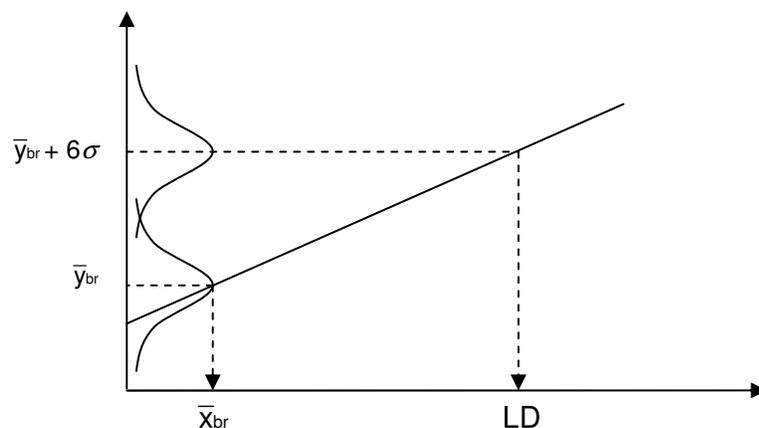


Figura 03 - correspondência através da curva de calibração univariada dos valores do ensaio em branco e do limite de detecção no domínio da concentração.

Mais tarde^[41] sugeriu-se a modificação do cálculo do limite de detecção com a substituição do índice precedente ao erro associado ao sinal do branco, por uma função delta ou parâmetro de não centralidade (Δ).

O valor da função delta também pode ser calculado a partir dos valores pré-definidos de α e β , bem como obtido a partir de tabelas^[42]. Desta forma o limite de detecção no domínio do sinal analítico passa a ser:

$$y_{br} + \Delta(\alpha, \beta) \cdot \sigma_{br}. \quad (2.5)$$

O emprego deste parâmetro é justificado por Clayton pelo fato da hipótese nula, H_0 , não ser verdadeira neste caso e sim a segunda condição H_A , sendo assim a distribuição de probabilidade adequada é a distribuição normal não-centralizada. A modificação introduzida por Clayton tem sido aplicada^[43, 44] e difundida no entanto é alvo de críticas em um artigo de revisão recente sobre figuras de mérito^[45].

É interessante ressaltar que geralmente as definições contemplam o domínio da variável independente, mas para fins práticos o limite de detecção é

calculado como sendo a quantidade correspondente a este valor mínimo detectável da variável experimentalmente medida. No campo da espectroscopia de absorção, por exemplo, o conceito e o tratamento estatístico referente ao limite de detecção é aplicado à absorvância, que é a variável dependente, entretanto o limite de detecção de fato é a concentração equivalente ao sinal mínimo detectável calculado.

Outras figuras de mérito como a sensibilidade, seletividade e a relação sinal ruído já não possuem uma trajetória comparável ao do limite de detecção. Seus cálculos não sofreram grandes mudanças e permanecem basicamente os mesmos.

A sensibilidade é comumente expressa como o coeficiente angular, b , da equação de calibração, equação 2.1^[46]. De forma descritiva equivale à mudança no sinal correspondente à variação de 1 unidade de concentração da espécie de interesse. Todavia, posteriormente a influência da precisão foi considerada^[47] e o conceito adaptado para incluir esta informação passou a ser conhecido então como sensibilidade analítica:

$$S = b / \sigma_{br} \quad (2.6)$$

A seletividade é tida como a parcela do sinal disponível para quantificação. Quantitativamente pode ser expressa por uma razão de sensibilidades, imaginando um sistema onde a absorção em um determinado comprimento de onda é devida ao analito “k” e a um interferente “l”, a seletividade do analito mediante a presença do interferente “l” é calculada como^[46, 48]:

$$\text{Seletividade}_{l,a} = S_k / S_l \quad (2.7)$$

Por fim, a relação sinal ruído tem sido consensualmente calculada como a razão da resposta instrumental pela incerteza do sinal. Em espectroscopia corresponde, por exemplo, à absorção da amostra i no comprimento de onda j , dividida pelo desvio-padrão dos valores de absorvância no comprimento de onda j

para um conjunto de espectros onde somente a linha-base é registrada.

$$S/R = S_{i,j}/\sigma_j \quad (2.8)$$

2.2 – NOTAÇÕES E CONVENÇÕES

Durante o restante deste capítulo a seguinte convenção será adotada: todo valor escalar será representado por uma letra grafada em minúsculo e no formato itálico (como em “escalar x ”), todo vetor será representado por uma letra grafada em minúsculo e no formato negrito (como em “vetor \mathbf{x} ”), as matrizes serão representadas por letras em maiúsculo e no formato negrito (como em “matriz \mathbf{X} ”) e finalmente, um índice numérico será representado por letras em itálico e em posição de subscrito (por exemplo, vetor \mathbf{x} com índice i : \mathbf{x}_i). Nos demais capítulos, esta convenção não será necessariamente seguida, todavia os vetores e matrizes serão devidamente indicados no texto.

Nos cálculos envolvendo matrizes, eventualmente os termos espaço-linha e espaço-coluna serão utilizados, os mesmos se referem aos espaços geométricos gerados pela combinação linear dos vetores-linha ou vetores coluna, respectivamente, de uma determinada matriz \mathbf{X} .

Também serão utilizados símbolos como o da norma euclidiana de um vetor, $\|\mathbf{x}\|$; a transposta de um vetor, \mathbf{x}^T ; a pseudo-inversa de uma matriz, \mathbf{X}^+ ; e a parte ou complemento ortogonal de um vetor, \mathbf{x}^\perp .

2.3 – VALIDAÇÃO EM CALIBRAÇÃO MULTIVARIADA

Com o surgimento dos modelos inversos de calibração multivariada^c, as definições das figuras de mérito do universo univariado naturalmente necessitaram de adaptações. Entretanto devido à natureza inversa e multicanal destes modelos, as novas definições não foram imediatas, uma vez que o raciocínio anteriormente empregado não era válido para métodos multivariados atuais. No caso do limite de

^c: métodos nos quais o papel das variáveis independente e dependente são invertidos visando a obtenção de melhores resultados. Em espectroscopia, por exemplo, os valores de concentração (tipicamente a variável independente) estariam no eixo y e as absorbâncias no(s) eixo(s) x .

detecção, por exemplo, a idéia constituía em utilizar a equação do modelo univariado como relação de conversão, uma vez que a detecção era definida por meio do menor sinal que poderia ser atribuído ao analito com posterior conversão deste na concentração equivalente ao limite de detecção. Ocorre que nos métodos de calibração como o PCR e PLS existem infinitas combinações de sinais que geram uma determinada concentração, não havendo desta forma uma relação unívoca entre a concentração, variável y_i , e o vetor de variáveis independentes^[49], sendo falha portanto a abordagem previamente utilizada no âmbito univariado.

2.4 - SURGIMENTO DO SINAL ANALÍTICO LÍQUIDO

Tendo em vista a problemática dos sistemas multicomponentes, onde o sinal bruto obtido passa a ser uma soma do sinal das diversas espécies presentes na amostra, ou seja a resposta instrumental não é mais específica para o analito, existindo então uma sobreposição das informações, Avraham Lorber introduziu em 1986 o conceito de **sinal analítico líquido**^[50], com o intuito de obter uma medida cuja informação fosse proveniente unicamente da espécie de interesse.

Inicialmente o conceito do NAS (do Inglês *Net Analyte Signal*) foi aplicado em um conjunto de dados proveniente do estudo de quatro compostos orgânicos utilizando-se espectroscopia na região do ultravioleta. Uma vez que havia sobreposição no sinal observado e a absorbância registrada era então a composição da informação da espécie de interesse e das demais espécies interferentes, Lorber definiu o sinal analítico líquido como a parcela do sinal que é ortogonal ao espaço gerado pelos interferentes. Algebricamente, a parte do espectro que não é ortogonal à contribuição das demais espécies é basicamente uma combinação linear dos respectivos espectros puros.

Graficamente, o NAS pode ser visto como o vetor ortogonal ao subespaço da matriz de dados gerado apenas pelos interferentes, Figura - 4a.

Uma vez que o espectro de cada amostra é representado por um vetor em um conjunto de eixos que são os comprimentos de onda da região em estudo, os vetores obtidos a partir dos espectros dos compostos puros, exceto o analito, geram o espaço dos interferentes. Portanto qualquer amostra contendo apenas as

espécies interferentes origina um vetor presente neste subespaço, mais precisamente cada uma destas amostras pode ser representada por um vetor que é uma combinação linear dos espectros puros dos interferentes. Desta forma, os vetores que não podem ser descritos apenas como uma combinação linear, dos vetores que geram subespaço dos interferentes, contêm informação adicional, sendo esta a informação que está associada ao analito. Esta informação pode ser visualizada como os valores de absorbâncias nos comprimentos de onda onde os interferentes não apresentam sinal de absorção.

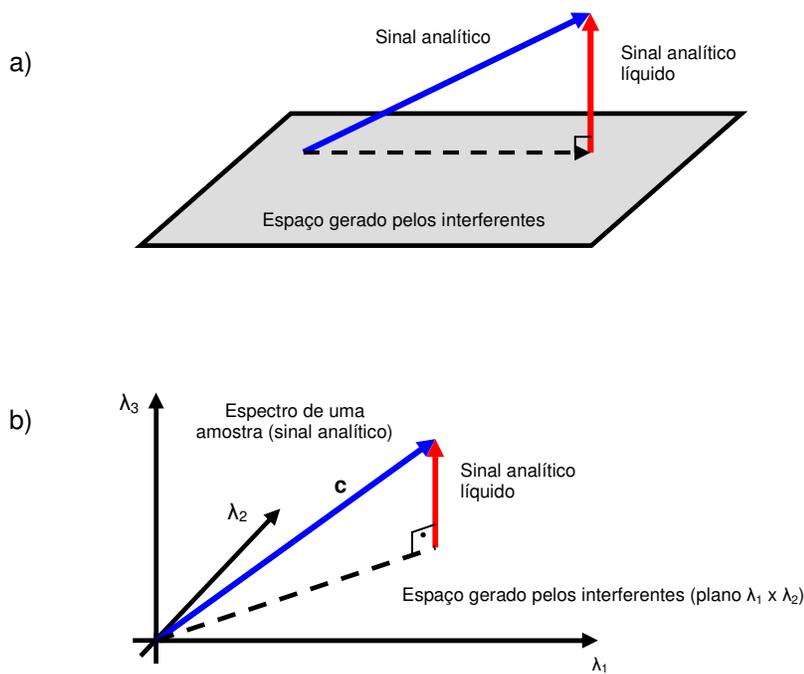


Figura 04 – (a) representação geométrica tradicional do conceito do vetor NAS, (b) representação de um caso hipotético onde apenas 3 comprimentos de onda são registrados por espectro.

Quimicamente o vetor NAS pode ser compreendido considerando um sistema hipotético, representado na parte b da figura 3, com apenas duas espécies químicas interferentes A e B, além do analito K. Assume-se que o interferente A absorve somente em λ_1 e B somente em λ_2 , o analito absorve em λ_1 , λ_2 e λ_3 . Tem-se então os vetores **a** e **b** representando respectivamente os espectros puros dos interferentes de mesma denominação e um vetor **c**

representando uma amostra contendo as espécies A, B e K. Os vetores **a** e **b** coincidem com os eixos das variáveis λ_1 e λ_2 e geram o plano (λ_1, λ_2) . Para facilitar a visualização, apenas o espaço bidimensional gerado por estes vetores é identificado. O sinal analítico líquido é o vetor que contém a informação unicamente relacionada à espécie de interesse, neste caso a espécie K. Devido às circunstâncias impostas para este sistema, a absorção em λ_3 naturalmente assume o papel da informação única do analito, uma vez que as absorções em λ_1 e λ_2 são a composição das informações das 3 espécies devido à sobreposição espectral. Geometricamente a absorção em λ_3 pode então ser vista como a componente do vetor **c** que é ortogonal ao espaço gerado pelos vetores **a** e **b**, ou seja, é o vetor NAS.

Matematicamente, o vetor NAS para uma amostra, doravante representado por \mathbf{r}_{nas} , pode ser obtido através da aplicação do operador $[\mathbf{I} - \mathbf{X}_{-k}^T \cdot (\mathbf{X}_{-k}^T)^+]$ no vetor \mathbf{v}_i , de acordo com a equação (2.9). Onde k representa o analito; \mathbf{X}_{-k} , a matriz de calibração sem a informação do analito (contemplando apenas a informação dos interferentes); $(\mathbf{X}_{-k}^T)^+$, a pseudo-inversa da transposta da matriz \mathbf{X}_{-k} , \mathbf{I} uma matriz identidade de dimensões adequadas e \mathbf{v}_i um vetor representando o espectro de uma das amostras do conjunto de calibração. O operador $\mathbf{X}_{-k}^T \cdot (\mathbf{X}_{-k}^T)^+$ tem por função projetar o vetor \mathbf{v}_i no espaço-coluna da matriz \mathbf{X}_{-k}^T , ou seja no espaço dos interferentes que é gerado pelas linhas da matriz \mathbf{X}_{-k} (ver apêndice A1). Desta forma, se a amostra for uma combinação linear dos vetores-base do espaço-linha associado à matriz \mathbf{X}_{-k} , a aplicação deste retorna o próprio vetor \mathbf{v}_i como resultado, e ao aplicar-se o operador $[\mathbf{I} - \mathbf{X}_{-k}^T \cdot (\mathbf{X}_{-k}^T)^+]$ tem-se como resposta o vetor nulo. Portanto, esta operação atua no vetor em questão anulando apenas a contribuição proveniente dos interferentes, explicitando a informação ortogonal referente à espécie k.

$$\mathbf{r}_{\text{nas}} = [\mathbf{I} - \mathbf{X}_{-k}^T \cdot (\mathbf{X}_{-k}^T)^+] \cdot \mathbf{v}_i \quad (2.9)$$

Historicamente, o NAS foi proposto para sistemas estudados sob o ponto de vista espectroscópico e onde métodos de calibração multivariada direta como o *CLS* (do inglês *Classical Least Squares*) podiam ser aplicados. Ou seja, em

situações onde todas as espécies presentes eram conhecidas e a matriz \mathbf{X}_k podia ser facilmente estabelecida, uma vez que consistia basicamente dos espectros puros dos interferentes dispostos como vetores-linha.

$$\mathbf{X}_k = \begin{pmatrix} \text{Interferente 1} \\ \text{Interferente 2} \\ \cdot \\ \cdot \\ \cdot \\ \text{Interferente Z} \end{pmatrix} \quad (2.10)$$

A introdução do sinal analítico líquido permitiu o cômputo de figuras de mérito cruciais na validação de metodologias analíticas, como limite de detecção, sensibilidade, seletividade e relação sinal-ruído, parâmetros estes que ainda eram desconhecidos aos métodos de calibração multivariada do tipo inversa. Ainda hoje, apesar das diversas propostas e trabalhos publicados, estas figuras de mérito não estão formalmente definidas e, portanto, seu uso não é totalmente aceito. Considerando a definição apresentada por Lorber e a natureza dos métodos inversos onde na maioria dos casos os interferentes não são conhecidos e em determinadas circunstâncias apenas as concentrações do analito são conhecidas, o foco no cálculo do NAS foi direcionado para a determinação da matriz \mathbf{X}_k , uma vez que a equação (2.9) não poderia ser diretamente aplicada, devido a ausência dos espectros puros de todos interferentes ou do espectro puro da espécie k.

Lorber et al.^[3] apresentaram uma forma de contornar o problema e propuseram obter a matriz \mathbf{X}_k por meio de uma técnica matemática de redução do posto (*rank annihilation* em inglês).

De posse da matriz de calibração constituída de j ($j = 1, \dots, J$) comprimentos de onda (variáveis) e i ($i = 1, \dots, I$) amostras de calibração, originando a matriz \mathbf{X} de dimensões ($I \times J$) e das concentrações do analito k para todas as amostras, a matriz \mathbf{X}_k pode ser obtida através da equação (1.11).

$$\mathbf{X}_{A,k} = \mathbf{X}_A - \zeta \mathbf{y}_k \mathbf{v}_{cl}^T \quad (2.11)$$

Primeiramente, \mathbf{X} é reconstruída utilizando somente os primeiros A fatores significativos (\mathbf{X}_A), resultantes da decomposição da matriz \mathbf{X} por meio da aplicação da análise de componentes principais (*PCA*) ou mínimos quadrados parciais (*PLS*), por exemplo. Essa reconstrução é necessária para evitar a inversão da matriz singular $\mathbf{X}^T\mathbf{X}$ posteriormente no cálculo dos coeficientes de regressão, além de minimizar o ruído presente na matriz original. O valor ζ é um escalar definido como:

$$\zeta = 1/\mathbf{v}_{cl}^T\mathbf{X}_A^+\mathbf{y}_{k^*} \quad (2.12)$$

O vetor de concentrações do analito, \mathbf{y}_k , é então projetado no espaço de dimensão A , gerando \mathbf{y}_{k^*} , ($\mathbf{c}_{k^*} = \mathbf{X}_A\mathbf{X}_A^+\mathbf{y}_k$), \mathbf{v}_{cl} é uma combinação linear das linhas de \mathbf{X}_A . Apesar de rigorosamente o vetor \mathbf{v}_{cl} , utilizado na técnica de redução de posto, ser o espectro puro da espécie de interesse, o mesmo resultado pode ser obtido através da combinação linear das linhas ou amostras da matriz \mathbf{X}_A ou \mathbf{X} , influenciando somente no valor de ζ .

O vetor NAS para cada amostra i é então obtido com o auxílio da matriz $\mathbf{X}_{A,-k}$:

$$\mathbf{r}_{nas} = (\mathbf{I} - \mathbf{X}_{A,-k}^T(\mathbf{X}_{A,-k}^T)^+)\mathbf{v}_i \quad (2.13)$$

A idéia central nesta definição da matriz dos interferentes se baseia na substituição do espectro puro do analito com concentração unitária por uma combinação linear das linhas da matriz \mathbf{X} reconstruída, fato este que pode ser facilmente compreendido ao considerar o ajuste feito posteriormente pela introdução do índice α . Ao analisar-se a equação (1.12) conclui-se que o vetor \mathbf{v}_{cl} é apenas uma nova amostra que não está presente no conjunto de calibração e naturalmente contém a informação do analito, entretanto em uma concentração não unitária. O escalar ζ tem por função normalizar o vetor \mathbf{v}_{cl} referente à concentração da espécie de interesse k , ou seja ζ é basicamente o inverso da concentração do analito na amostra representada pelo vetor \mathbf{v}_{cl} ($\mathbf{X}_A^+\mathbf{y}_{k^*} = \mathbf{b}$, vetor de regressão). Portanto, a multiplicação do vetor \mathbf{y}_{k^*} pelo produto $\alpha.\mathbf{v}_{cl}^T$ gera uma matriz onde todos os sinais unicamente gerados a partir do analito são os mesmos

presentes na matriz \mathbf{X}_A , e ao efetuar-se a diferença $\mathbf{X}_A - \zeta \mathbf{y}_k \cdot \mathbf{v}_{cl}^T$ tem-se então a subtração da contribuição do analito na matriz de calibração, dando origem ao espaço dos interferentes.

Uma vez que as espécies químicas, ou mais precisamente as mudanças nas suas respectivas concentrações em cada uma das amostras, constituem as fontes de variação no conjunto de dados, é natural esperar que o posto químico da matriz \mathbf{X}_A seja igual ao número de espécies presentes no sistema. Sendo assim ao excluir-se a informação do analito da matriz \mathbf{X}_A , matematicamente diminui-se o posto da matriz de uma unidade e portando uma etapa de redução de posto (do inglês *rank annihilation step*) foi realizada, o que explica o termo utilizado para esta abordagem.

Xu e Schechter ^[51] sugeriram um forma de obtenção da matriz \mathbf{X}_{-k} a qual dispensa a inserção do escalar α introduzido anteriormente.

Primeiramente a matriz de calibração $\mathbf{X}(I,J)$ é normalizada com referência à concentração do analito, o que é realizado através da divisão de cada amostra ou linha da matriz \mathbf{X} pela concentração correspondente do analito:

$$\mathbf{X}_n = \mathbf{Y}_k^{-1} \mathbf{X} \quad (2.14)$$

onde \mathbf{X}_n representa a matriz normalizada; \mathbf{Y}_k^{-1} é a inversa da matriz diagonal \mathbf{Y} ($I \times I$) cujos elementos são as concentrações do analito em cada amostra. Em seguida um vetor de absorbâncias médio é calculado somando-se todas os vetores-linha da matriz \mathbf{X} e dividindo-se cada elemento do vetor resultante pelo número total de linhas. Em seguida este vetor, \mathbf{x}_{av} , é subtraído de cada linha da matriz normalizada, \mathbf{X}_n :

$$\mathbf{X}_{-k} = \mathbf{X}_n - \mathbf{X}_{av} \quad (2.15)$$

Sendo \mathbf{X}_{-k} a representação do espaço dos interferentes e \mathbf{X}_{av} uma matriz cujas linhas se repetem sendo todas iguais ao vetor de absorbâncias médio \mathbf{x}_{av} , o qual é calculado através da equação:

$$\mathbf{x}_{av} = \frac{1}{l} \sum_{i=1}^l (\mathbf{x}_{n,i}) \quad (2.16)$$

Sendo $\mathbf{x}_{n,i}$ o i -ésimo vetor-linha da matriz \mathbf{X}_n e l o número de linhas da matriz.

Diferentemente da definição presente na equação (2.11), esta abordagem normaliza a própria matriz de calibração para que os valores de absorbância nos comprimentos de onda unicamente relacionados ao analito sejam correspondentes à concentração unitária da espécie k . Em seguida é calculada a média de cada variável J ($j = 1, \dots, J$) em todas as l amostras de modo que o vetor denominado \mathbf{x}_{av} contenha o valor médio de cada variável. Teoricamente os valores individuais das absorbâncias, após a normalização, e o valor médio devem ser iguais, logo ao subtrair-se o espectro médio de cada amostra as variáveis que representam a informação única da espécie k são automaticamente zeradas e a contribuição do analito é anulada, resultando na matriz \mathbf{X}_{-k} .

Berger e colaboradores^[52], em seu artigo de introdução da metodologia de análise linear híbrida (em inglês *Hybrid Linear Analysis, HLA*), que trata de um novo método de calibração multivariada, apresentaram e formalizaram uma maneira lógica e intuitiva de obtenção de \mathbf{X}_{-k} , onde o produto do vetor de concentrações de calibração da espécie k pela transposta de um espectro representativo do analito k , medido na concentração unitária é subtraído da matriz de calibração.

O subespaço da matriz \mathbf{X} contendo apenas a contribuição dos interferentes pode ser representado pela equação genérica:

$$\mathbf{X}_{-k} = \mathbf{X} - \mathbf{y}_k \cdot \mathbf{v}^T \quad (2.17)$$

Onde \mathbf{X} corresponde à matriz de calibração; \mathbf{y}_k , é o vetor com as concentrações de calibração do analito e \mathbf{v} é um espectro puro do analito correspondente à concentração unitária.

Tendo em mente as 3 propostas previamente discutidas, em todas elas o subespaço \mathbf{X}_k é genericamente obtido de acordo com a equação (2.17), modificando-se apenas a compreensão do vetor \mathbf{v}^T . Este vetor pode ser tanto o espectro puro do analito na concentração unitária ou uma combinação linear apropriada das linhas da matriz de calibração gerando um novo espectro (equivalente a uma nova amostra) que naturalmente também contém a informação do analito. Neste segundo caso, basta posteriormente normalizar a combinação linear para a concentração unitária da espécie de interesse, k .

Entretanto Goicoechea e colaboradores^[53], sugeriram uma forma totalmente distinta para a obtenção da matriz \mathbf{X}_k e propuseram que o espaço dos interferentes poderia ser calculado como:

$$\mathbf{X}_k = (\mathbf{I} - \mathbf{y}_k(\mathbf{y}_k)^+) \mathbf{X} \quad (2.18)$$

Os termos da equação 2.18 possuem as mesmas definições discutidas anteriormente.

A idéia por traz da equação (2.18) pode ser compreendida fazendo-se uma comparação através do raciocínio inverso de obtenção do vetor NAS. Assim como no método dos mínimos quadrados onde o vetor de regressão é obtido por meio da projeção do vetor de concentração no espaço-coluna da matriz \mathbf{X} (ver apêndice 1), o procedimento de cálculo do vetor NAS utiliza o operador $\mathbf{X}_{-k}^T(\mathbf{X}_{-k}^T)^+$ para projetar o vetor \mathbf{v}_i no espaço-coluna de \mathbf{X}_{-k}^T e obter desta forma a componente de \mathbf{v}_i que contém apenas a contribuição dos interferentes. Em seguida esta componente é subtraída do vetor original resultando então no vetor NAS, subtração essa possibilitada por meio da matriz identidade que atua simplesmente para reproduzir o vetor \mathbf{v}_i . Ou seja, a aplicação do operador $(\mathbf{I} - \mathbf{X}_{-k}^T(\mathbf{X}_{-k}^T)^+)$ no vetor \mathbf{v}_i , elimina a informação das espécies interferentes e evidencia a informação do analito; de forma semelhante, a aplicação do operador $(\mathbf{I} - \mathbf{y}_k(\mathbf{y}_k)^+)$ sobre a matriz \mathbf{X} elimina a informação do analito e evidencia a informação das espécies interferentes.

O vetor NAS calculado empregando-se qualquer uma das alternativas citadas pode ser então utilizado tanto para a determinação das figuras de mérito como para um meio de calibração independente. Os métodos de calibração baseados no NAS (conhecidos em inglês como *NAS Based Methods*, *NBM*) como o *HLA*, fazem uso tanto do vetor NAS como da norma euclidiana deste vetor. Uma vez que o vetor é livre de interferentes sua proporcionalidade é mantida ao substituí-lo pelo escalar $\|\mathbf{r}_{nas}\|$ transformando um problema inicialmente multivariado em uma calibração de ordem zero ou pseudounivariada.

É importante mencionar que Lorber introduziu o conceito do sinal analítico líquido como uma ferramenta para o cálculo de algumas figuras de mérito, valores estes que descrevem a habilidade da abordagem matemática em extrair a informação do conjunto de dados. Entretanto o NAS calculado através de qualquer uma das alternativas previamente descritas é uma característica específica do conjunto de dados obtidos, independente da abordagem utilizada para calibração. Bro ^[54] prudentemente alerta para este fato e enfatiza que as figuras de mérito devem estar intimamente relacionadas ao modelo para que seja possível por exemplo comparar calibrações obtidas por modelos distintos. Entretanto esta premissa não é satisfeita ao calcularem-se os parâmetros de desempenho utilizando-se o sinal analítico líquido obtido através das perspectivas anteriores.

Neste cenário inicia-se então uma recente discussão a respeito do que denomina-se no momento de sinal analítico líquido verdadeiro (*true net analyte signal*)^[54]. O termo encontrado em poucos trabalhos é atribuído ao vetor NAS ideal, concebendo-o como uma propriedade do conjunto de calibração e assumindo-se desta forma que os diferentes métodos apresentados até aqui são formas de obtenção de uma estimativa do *true net analyte signal*. Logo em uma situação ideal todos estes cálculos resultariam em um mesmo vetor e conseqüentemente as figuras de mérito seriam por sua vez todas idênticas. Esta nova interpretação do sinal analítico líquido reforça os argumentos apresentados por Bro e reitera a necessidade de um caminho alternativo para o cômputo do sinal analítico líquido de forma a ser representativo da capacidade do modelo em extrair a informação do analito presente em um determinado conjunto de dados.

Devido ao crescente interesse e às aplicações do conceito do sinal analítico líquido, Faber^[55] derivou uma forma mais eficiente em termos computacionais para o cálculo do NAS, baseando-se na publicação de Sanchez e Kowalski^[56], os quais apresentam por meio da álgebra tensorial que o vetor de regressão constitui-se da parcela do sinal do analito ortogonal ao sinal dos interferentes, Faber propõe que o NAS seja calculado através do vetor de regressão uma vez que ambos estão no espaço ortogonal ao espaço dos interferentes, o que nos faz concluir que são vetores colineares.

$$\mathbf{y} = \mathbf{X} \cdot \mathbf{b} \quad (2.19)$$

$$y = \mathbf{b}^T \cdot \mathbf{x} \quad (2.20)$$

$$y = \mathbf{b}^T \cdot (\mathbf{x}_{nas} + \mathbf{x}^\perp) \quad (2.21)$$

$$y = \mathbf{b}^T \cdot \mathbf{x}_{nas} + \mathbf{b}^T \cdot \mathbf{x}^\perp \quad (2.22)$$

$$y = \mathbf{b}^T \cdot \mathbf{x}_{nas} \quad (2.23)$$

$$\mathbf{x}_{nas} = (\mathbf{b} \cdot \mathbf{b}^T)^{-1} \cdot \mathbf{b} \cdot y \quad (2.24)$$

$$\mathbf{r}_{nas\ k,i} = \mathbf{b}(\mathbf{b}^T \mathbf{b})^{-1} y_i \quad (2.25)$$

Onde,

y: vetor de concentrações do analito.

y: concentração de uma amostra genérica.

X: matriz de dados (absorbâncias).

b: vetor de regressão.

Na prática a concentração prevista para a amostra i (\hat{y}_i) substitui o valor esperado (y_i). Diferentemente dos demais métodos a equação acima traz uma complicação para os modelos onde a aplicação do pré-processamento "centrar os dados pela média" é necessária. Nestes casos a equação (2.25) já não é mais válida, uma vez que produz resultados não equivalentes para a norma do vetor NAS. Dessa forma é necessário realizar uma correção onde a norma do vetor NAS para a espécie k e amostra i , $\|\mathbf{r}_{nas\ k,i}\|$, deve ser multiplicada pelo respectivo sinal algébrico obtido pela diferença entre a concentração da espécie k naquela amostra e a concentração média do conjunto, $(y_i - \bar{y})$ ^[57].

Nesta linha de pensamento Bro^[58] também propõe uma nova maneira de obter-se o sinal analítico líquido baseada no vetor de regressão, mas que gerencia automaticamente a questão da utilização do pré-processamento. O autor categoricamente recomenda a utilização da correção, devido à etapa de pré-processamento, mesmo se os dados não foram submetidos a esse ajuste, argumentando que a inadequação dos valores calculados através do vetor de regressão não está presente somente nos casos onde a opção de pré-processamento é necessária, mas também em situações onde haja valores de concentração previstos abaixo de zero. Neste trabalho utilizou-se esta última abordagem.

2.5 - FIGURAS ANALÍTICAS DE MÉRITO EM CALIBRAÇÃO MULTIVARIADA

2.5.1 - Limite de detecção

Consensualmente é considerado como o menor sinal obtido que pode ser atribuído à propriedade de interesse sem que o mesmo se confunda com o ruído, ou seja, o erro associado à medida.

No universo multivariado uma das formas de cálculo do limite de detecção (LD) propostas^[15] é expressa como:

$$LD = 3 \|\mathbf{b}\| \sigma \quad (2.26)$$

Sendo σ a incerteza associada à medida, é uma forma de estimar o ruído instrumental. $\|\mathbf{b}\|$ é a norma euclidiana do vetor de regressão.

Segundo Ortiz e colaboradores^[49], o limite de detecção pode ser calculado realizando um ajuste por mínimos quadrados com as concentrações experimentais da espécie de interesse (y) como variável independente e as concentrações previstas pelo modelo como variável dependente. Desta maneira, o LD assume um perfil univariado sendo definido de acordo com a equação 2.27.

$$LD_{\text{Ortiz}} = \Delta_{(\alpha, \beta, \nu)} \frac{\sigma}{b} \sqrt{\frac{1}{L} + \frac{1}{n} + \frac{\bar{y}^2}{\sum_{i=1}^L (y_i - \bar{y})^2}} \quad (2.27)$$

Onde o parâmetro delta (Δ), anteriormente discutido, é o valor que associa as probabilidades de ocorrência de erro do tipo I (α), do tipo II (β) e o número de graus de liberdade do sistema (ν); b é o coeficiente angular da reta ajustada, L é o número de amostras, n é a quantidade de amostras de mesma concentração, σ equivale ao valor de RMSEP, definido na equação 2.36, calculado para $L-2$ graus de liberdade e \bar{y} é o valor médio global das concentrações experimentais.

Neste trabalho ambas definições foram utilizadas para estimar o limite de detecção.

2.5.2 - Limite de quantificação

Considera-se que somente a partir do limite de quantificação a incerteza associada a um resultado analítico é tolerável para fins de quantificação. Sendo assim, é recomendado como o valor mínimo para que medidas quantitativas possam ser realizadas.

De forma análoga ao limite de detecção, o limite de quantificação foi estabelecido como sendo 10 vezes a incerteza multiplicada pela norma do vetor de regressão^[15].

$$LQ = 10 \|\mathbf{b}\| \sigma \quad (2.28)$$

Nesta dissertação esta figura de mérito foi calculada por meio da equação 2.28 e complementarmente com a proposta apresentada na equação 2.27, modificando-se apenas o valor do parâmetro delta.

2.5.3 - Sensibilidade

É definida como a fração do sinal responsável pelo acréscimo de uma unidade de concentração. É um parâmetro que indica quão sensível o método se apresenta para distinguir pequenas diferenças de concentração nas amostras. Assim como no caso univariado, as considerações a respeito da razão sinal/concentração e a precisão da medida, discutidas na seção 2.1, são o ponto de partida para as definições em calibração de 1ª ordem.

Para modelos de calibração multivariada, como o PLS, pode ser determinada como^[2]:

$$s_k = \frac{1}{\|\mathbf{b}\|} \quad (2.29)$$

Quando o NAS é determinado, pode-se calcular o vetor de sensibilidade líquida para cada amostra i .

$$\mathbf{s}_k = \frac{\mathbf{r}_{\text{NAS}, i}}{c_i} \quad (2.30)$$

sendo $\mathbf{r}_{\text{NAS}, i}$ o vetor sinal analítico líquido da espécie de interesse na amostra i e c_i a concentração do analito na respectiva amostra. Em uma situação ideal os vetores de sensibilidade calculados para cada uma das amostras devem ser iguais, todavia na realidade as estimativas individuais são combinadas em uma média aritmética para obter-se um vetor médio representativo do conjunto de calibração. E finalmente o escalar s_k pode ser determinado por:

$$s_k = \|\mathbf{s}_k\| \quad (2.31)$$

2.5.4 – Sensibilidade analítica

A sensibilidade analítica é tipicamente calculada com o intuito de estabelecer-se a menor diferença de concentração entre amostras que pode ser distinguida pelo método. Neste sentido, esta figura de mérito é definida^[9] como a

razão entre a sensibilidade e a flutuação do sinal analítico, σ_{NAS} . A flutuação do sinal analítico foi estimada por meio do desvio-padrão da norma do vetor NAS de 40 espectros do ensaio em branco.

$$SEN_{A_k} = \frac{S_k}{\sigma_{NAS}} \quad (2.32)$$

Finalmente o inverso deste valor é calculado, γ^{-1} .

$$\gamma^{-1} = \frac{1}{SEN_{A_k}} \quad (2.33)$$

2.5.5 - Seletividade

De forma ampla a seletividade expressa o quanto do sinal registrado está disponível para quantificação do analito, consiste em uma grandeza que permite avaliar a especificidade da medida.

Tendo a natureza escalar da sensibilidade univariada em mente, Lorber e colaboradores^[3] postulam que este parâmetro pode ser calculado como:

$$SEL_i = \frac{\|\mathbf{r}_{NAS,i}\|}{\|\mathbf{x}_i\|} \quad (2.34)$$

Sendo $\|\mathbf{r}_{NAS,i}\|$ a norma do vetor NAS para o analito k na amostra i e \mathbf{x}_i o espectro da amostra i .

2.5.6 - Relação sinal ruído

Este valor informa quão superior a resposta instrumental se apresenta em relação ao erro instrumental. É obtido através do cálculo da razão do sinal analítico pela incerteza.

A proposta de Lorber e outros^[3] segue este perfil, apenas com adaptações referentes ao sinal e sem modificações conceituais. Sugerindo então que a norma do vetor NAS seja dividida por uma estimativa do desvio padrão do erro associado à resposta instrumental.

$$S/N = \frac{\| \mathbf{r}_{NAS, i} \|}{\sigma_{NAS}} \quad (2.35)$$

2.5.7 - Exatidão

Normalmente avaliada através de testes de recuperação de amostras certificadas, a exatidão confronta o resultado final disponibilizado pelo processo analítico com o valor real ou verdadeiro. Nas análises que fazem uso da regressão a exatidão se refere à diferença entre as concentrações nominais dos padrões e os valores previstos pelo modelo de regressão. Geralmente é verificada por meio do cálculo do RMSEP (do inglês *Root Mean Square Error of Prediction*, em português: raiz quadrada do erro quadrático médio de previsão).

$$RMSEP = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \quad (2.36)$$

Onde \hat{x}_i e x_i são respectivamente o valor previsto pelo modelo e o valor de referência da amostra i , n é o número de amostras no conjunto de previsão.

2.5.8 - Precisão

Estabelece o grau de concordância entre diversas medidas independentes de uma mesma amostra^[90]. Tem como objetivo indicar a amplitude do erro associado ao resultado final. Pode ser avaliada em diversos níveis, sendo matematicamente computada por meio do desvio padrão do conjunto de observações. Medidas típicas de precisão são a repetibilidade, precisão intermediária e a reprodutibilidade.

$$\delta_x = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}} \quad (2.37)$$

2.5.8.1 - Repetibilidade

A repetibilidade^[90] consiste em, logo após ter-se implementado o método de análise, realizar-se novas medidas ou réplicas do procedimento analítico, geralmente em um curto intervalo de tempo. Essa medida tem por função contabilizar, por exemplo, a influência de pequenas mudanças, como variações nas condições ambientais. Neste caso vidrarias, equipamentos e o analista continuam os mesmos.

Pode ser calculada através da realização de diversas análises em uma mesma amostra, comumente no mesmo dia.

2.5.8.2 - Precisão intermediária

Consiste no cálculo do desvio-padrão dos resultados após a variação de uma determinada condição experimental. Por exemplo, análises realizadas em dias diferentes, ou por analistas diferentes ou com equipamentos diferentes. É importante destacar que caso todas estas variações ocorram, a medida não representará a precisão intermediária^[90], mas a reprodutibilidade.

2.5.8.3 - Reprodutibilidade

Em geral está erroneamente associada como uma medida de precisão temporal do método. A reprodutibilidade permite avaliar a distorção sofrida pela metodologia quando um determinado analista decide fazer uso de um procedimento já publicado. Na ação de reproduzir ou empregar o método previamente aceito o mesmo acaba por submeter a técnica às mais variadas

condições, como grandes intervalos de tempo, condições ambientais diferentes, analistas diversos, equipamentos e laboratórios distintos etc., sendo estas alterações os alvos a serem contabilizados por esta figura de método.

Uma vez que engloba a maior gama possível de variações ao qual o procedimento está sujeito, esta variante da precisão é obtida através de ensaios interlaboratoriais^[90].

2.5.9 - Faixa linear de trabalho

Uma grandeza que confirma se a faixa de valores da variável independente apresenta uma relação estritamente linear com o sinal instrumental. A faixa linear de trabalho pode ser avaliada através de um teste de falta de ajuste dos dados experimentais ao modelo escolhido. Para esta avaliação é necessário que sejam preparadas réplicas das amostras para comparar-se o erro experimental com a possível falta de ajuste do modelo. Neste trabalho, esta figura de mérito não foi calculada devido à insuficiência de dados para uma avaliação consistente.

O método fundamenta-se na soma quadrática das diferenças $(SQ_r)^{[59]}$ entre os valores observados das réplicas para um determinado nível “j” da variável de interesse, por exemplo a concentração, e o valor previsto pelo modelo.

$$(SQ_r)_j = \sum_j^J (y_{jh} - \hat{y}_j)^2 \quad (2.38)$$

Onde,

$(SQ_r)_j$: soma quadrática das diferenças para o nível j.

y_{jh} : valor de concentração da h-ésima replicata do nível “j”.

\hat{y}_j : valor da concentração previsto pelo modelo de calibração para o nível “j”.

Estas diferenças, ou resíduos, deixadas pelo modelo podem ser decompostas em dois termos.

$$(y_{jh} - \hat{y}_j) = (y_{jh} - \bar{y}_j) - (\hat{y}_j - \bar{y}_j) \quad (2.39)$$

Onde,

\bar{y}_j : média dos valores observados no nível “j”.

Elevando-se a equação 2.39 ao quadrado e realizando o somatório sobre todas as observações, tem-se a forma final da soma quadrática residual:

$$\sum_j^J \sum_h^H (y_{jh} - \hat{y}_j)^2 = \sum_j^J \sum_h^H (y_{jh} - \bar{y}_j)^2 + \sum_j^J \sum_h^H (\hat{y}_j - \bar{y}_j)^2 \quad (2.40)$$

Onde, $j = 1 \dots J$ e $h = 1 \dots H$.

O primeiro termo do lado direito da equação 2.40 não depende do modelo, estando unicamente relacionado à dispersão dos valores em relação ao valor médio, ou seja, trata-se do erro aleatório de natureza experimental (chamado de erro puro). Já o segundo termo reflete o quanto as estimativas do modelo (\hat{y}), para cada nível “j”, se desviam dos respectivos valores médios (\bar{y}_j). Desta forma, fornece uma medida da falta de ajuste do modelo às respostas observadas.

Quando dividimos essas somas quadráticas pelos seus respectivos números de graus de liberdade, temos médias quadráticas que *podem* representar *estimativas da variância das respostas*. A média quadrática devido ao erro puro (MQ_{ep}, equação 2.41) é uma estimativa da variância, esteja o modelo bem ajustado ou não, mas a média quadrática devido à falta de ajuste (MQ_{faj}, equação 2.42) só representará uma estimativa da variância caso o modelo esteja bem ajustado, senão este valor estimará a variância mais a contribuição da falta de ajuste. Sendo assim um teste F pode ser empregado para investigar se os valores MQ_{faj} e MQ_{ep} são estatisticamente iguais, neste caso significando que o modelo está bem ajustado.

$$MQ_{ep} = \frac{\sum_j^J \sum_h^H (y_{jh} - \bar{y}_j)^2}{h - j} \quad (2.41)$$

$$MQ_{faj} = \frac{\sum_j^J \sum_h^H (\hat{y}_j - \bar{y}_j)^2}{j - p} \quad (2.42)$$

Onde,

h : número total de observações (réplicas em todos os níveis).

j : número de níveis.

p : número de parâmetros calculados para o modelo.

Matematicamente compara-se o valor da razão MQ_{faj} / MQ_{ep} com o valor tabelado do parâmetro F para os respectivos graus de liberdade (F_{n_1, n_2} ; onde n_1 : graus de liberdade de MQ_{faj} e n_2 : graus de liberdade de MQ_{ep}).

Outras ferramentas como um gráfico dos resíduos versus as concentrações, que deve possuir caráter aleatório; ou no caso de modelos de calibração multivariada, um gráfico dos *scores* versus as concentrações^[60], que deve possuir comportamento linear, também auxiliam de forma qualitativa no julgamento a respeito do comportamento linear.

2.5.10 - Robustez

Um critério quase sempre discutido, todavia geralmente não mensurado. Assim como a exatidão, a precisão e suas variantes, a robustez de um método visa garantir a qualidade da informação provida pelo procedimento, este parâmetro confirma que os resultados são confiáveis, pois são independentes, dentro dos limites usuais, de pequenas variações durante a determinação.

Quando diferentes laboratórios usam o mesmo método, eles inevitavelmente introduzem pequenas variações no procedimento. Além destas, variáveis dificilmente controladas como a umidade, a temperatura, a pressão, a pureza dos reagentes ou ainda ligeiras flutuações nas quantidades dos componentes utilizados podem ter uma influência significativa no resultado final.

Tradicionalmente é avaliada por meio de um estudo interlaboratorial^[61, 62], sendo desta forma uma informação onerosa tanto financeiramente quanto em termos do tempo necessário para o estudo, e portanto determinada geralmente apenas para metodologias presentes em compêndios, protocolos oficiais, normas de garantia de qualidade e afins. Apesar de ser normalmente uma figura de mérito de difícil acesso, a robustez pode e deve ser testada sempre que for possível

introduzir pequenas alterações ou variar de forma deliberada estas fontes externas de erro para posterior análise das conseqüências. Nesta linha pensamento a AOAC (do inglês *Association of Official Analytical Chemists*) fazendo uso das técnicas de planejamento experimental propõe que as variáveis passíveis de manipulação sejam estudadas por meio de um planejamento fatorial de 2 níveis como apresentado na referência ^[63].

Em quimiometria também costuma-se utilizar o termo em situações onde a capacidade de previsão do modelo de calibração é avaliada. Nesse contexto, amostras são retiradas do conjunto de treinamento e um novo modelo é calculado, caso os coeficientes de regressão sejam pouco afetados, diz-se que o modelo é robusto. Neste caso, o termo é usado no sentido de robustez matemática, fazendo-se referência unicamente ao modelo de calibração empregado.

Capítulo

3

**ESPECTROSCOPIA
NO INFRAVERMELHO
PRÓXIMO**

3.0 - INTRODUÇÃO

A espectroscopia no infravermelho está fundamentada no fenômeno de absorção da radiação eletromagnética, de comprimento de onda de 780nm a 1.000.000 nm, por um sistema molecular promovendo o aumento de energia vibracional e ou rotacional das espécies químicas.

As técnicas de estudo espectroscópicas que fazem uso da radiação na região do infravermelho são melhor compreendidas subdividindo-se o espectro em regiões denominadas de infravermelho próximo (NIR – do inglês *near infrared*), médio (MIR –do inglês *mid infrared*) e distante (FIR – *far infrared*).

Tabela 1.0 – regiões do espectro infravermelho.

REGIÃO	COMPRIMENTO DE ONDA (nm)	NÚMERO DE ONDA (cm ⁻¹)	FREQUÊNCIA (Hz)
NIR	780 a 2.500	12.800 a 4.000	3,8 x 10 ¹⁴ a 1,2 x 10 ¹⁴
MIR	2500 a 50.000	4.000 a 200	1,2 x 10 ¹⁴ a 6,0 x 10 ¹²
FIR	50.000 a 1.000.000	200 a 10	6,0 x 10 ¹² a 3,0 x 10 ¹¹

As ferramentas baseadas em espectroscopia no infravermelho possuem diversas aplicações: na indústria alimentícia^[64-67], na indústria têxtil^[68, 69], no segmento farmacêutico^[70-72], no ramo agrícola^[65, 73, 74], no setor petroquímico^[75-77] e em diversos outros. Talvez seja a técnica espectroscópica mais utilizada nas últimas décadas, sendo alvo frequente de revisões bibliográficas que tentam compilar todas as suas aplicações.

Trata-se de uma técnica onde as medidas podem ser realizadas sem a necessidade de pré-tratamento da amostra, utilizando-se, por exemplo, acessórios de reflectância total atenuada – ATR. Desta forma, a espectroscopia no infravermelho pode ser considerada como uma ferramenta de análise não destrutiva. O desenvolvimento de novas fontes de radiação de elevada intensidade e diversos acessórios permitiu seu uso em campo, sendo possível realizar medidas *in loco* e *in natura* utilizando-se sondas conectadas ao equipamento portátil por meio de fibras ópticas.

3.1 - FUNDAMENTOS

Existem diversas fontes de informação que trazem todo os fundamentos e embasamento teórico a respeito dos diversos métodos espectroscópicos. Neste tópico, tem-se como referência o livro *Principles of Instrumental Analysis*^[46].

O princípio da espectroscopia na região do infravermelho consiste na absorção de energia, na forma de radiação eletromagnética, pelas moléculas presentes em uma amostra. A radiação correspondente a esta região possui energia apenas para promover as moléculas para níveis de energia vibracionais e rotacionais mais elevados. Desta forma, todo o estudo e o formalismo decorrentes da interpretação física deste fenômeno consideram o sistema molecular como um agregado de osciladores harmônicos, onde assume-se que uma molécula se comporta como um sistema massa-mola, o qual pode ser descrito através da lei de Hook. No caso de uma molécula diatômica, por exemplo, os núcleos são considerados como duas massas pontuais e a ligação química é tida como uma mola de massa desprezível.

Utilizando-se o caso mais simples de um oscilador harmônico da mecânica clássica, no qual tem-se uma massa que se movimenta na ausência de qualquer força de amortecimento, como o atrito com uma superfície, e que está conectada a uma mola cuja extremidade está fixa, os parâmetros que descrevem o movimento de vibração podem ser calculados.

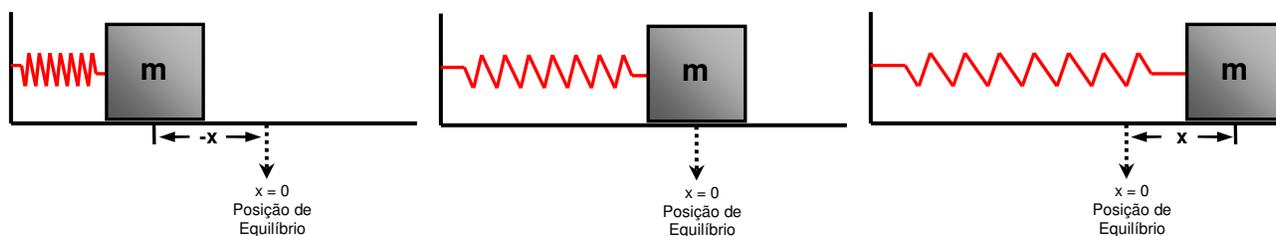


Figura 05 – Sistema massa-mola fundamental.

Ao promover-se o deslocamento de uma massa m a partir de sua posição de equilíbrio, surge sobre a mesma uma força que tende a restaurar a posição

inicial. A intensidade desta força pode ser obtida através da equação que representa a lei de Hook.

$$F = -k.x \quad \text{Lei de Hook} \quad (3.0)$$

Onde F é igual à força de restauração atuando na massa m.

x = deslocamento da massa em relação à posição e equilíbrio.

k = constante de força da mola, o sinal (-) indica que a força é contrária ao sentido do deslocamento.

Combinando-se a 2ª lei de Newton com a lei de Hook, obtém-se uma equação diferencial (EDO) de 2ª ordem:

$$F = m.a \quad 2^{\text{a}} \text{ lei de Newton} \quad (3.1)$$

$$a = \frac{d^2x}{dt} \quad (3.2)$$

$$m \frac{d^2x}{dt} = -kx \quad \text{EDO de } 2^{\text{a}} \text{ ordem} \quad (3.3)$$

Onde,

m = massa,

a = aceleração, derivada segunda do deslocamento com relação ao tempo.

Uma solução particular da equação 3.3 é a função periódica 3.4, a qual descreve o movimento da massa acoplada à mola.

$$x = A.\cos 2\pi v_m t \quad (3.4)$$

$$\frac{d^2x}{dt} = -4.\pi^2.v_m^2.A.\cos 2\pi v_m t \quad (3.5)$$

A amplitude do movimento é representada pela constante A.

Substituindo-se as equações 3.4 e 3.5 na equação 3.3, obtém-se a frequência de vibração do sistema, sendo o inverso de ν_m igual ao período.

$$\nu_m = (1/2\pi) \cdot (k/m)^{1/2} \quad (3.6)$$

Entretanto as equações acima descrevem o movimento de uma massa acoplada a uma mola cuja outra extremidade está fixa em um objeto de massa infinita. Para representar o movimento de uma molécula diatômica a massa m é substituída pela massa reduzida ϕ (ver Apêndice 2).

$$\phi = \frac{m_1 \cdot m_2}{(m_1 + m_2)} \quad (3.7)$$

$$\nu_m = \frac{1}{2\pi} \cdot \sqrt{\left(\frac{k}{\phi}\right)} \quad (3.8)$$

$$\nu_m = \frac{1}{2\pi} \sqrt{\frac{k(m_1 + m_2)}{(m_1 m_2)}} \quad (3.9)$$

A energia potencial é calculada definindo-se como zero a energia do sistema na posição de repouso ou posição de equilíbrio. Ao mover-se a massa m em um deslocamento dx tem-se pela 1ª lei de Newton e aplicando o cálculo integral/diferencial as seguintes conclusões:

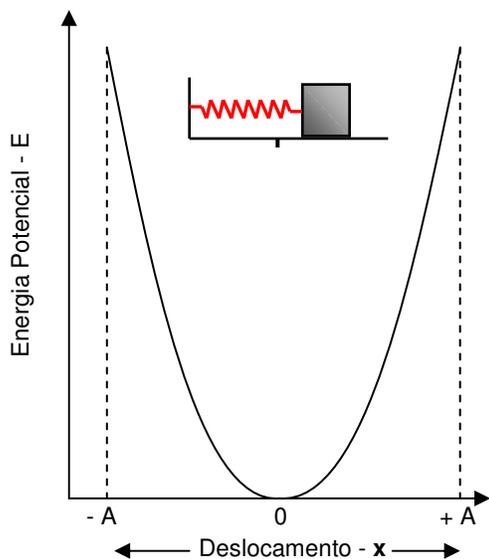
$$dE = -Fdx \quad (3.10)$$

$$dE = kxdx \quad (3.11)$$

$$\int_0^E dE = k \int_0^x x \cdot dx \quad (3.12)$$

$$E = \frac{1}{2} kx^2 \quad \text{Energia potencial} \quad (3.13)$$

A equação 3.13 representa a curva de energia potencial de um oscilador harmônico, a qual está ilustrada na Figura 06. O gráfico da energia potencial



possui o perfil de uma parábola cuja concavidade toca o eixo das abcissas em um único ponto, o zero de energia. Ao permitir-se que a massa oscile livremente, o primeiro máximo de energia é obtido quando a mola atinge a maior amplitude de deslocamento à esquerda ($- A$), em seguida a mesma sofre uma elongação gradual e a

Figura 06 – Diagrama de energia potencial para um oscilador harmônico.

energia potencial decresce continuamente até assumir o valor zero quando a massa passa pela posição de equilíbrio, a partir desse ponto a mola estende-se até a amplitude máxima à direita do ponto zero ($+ A$), atingindo novamente o valor máximo de energia potencial.

3.2 - VIBRAÇÕES MOLECULARES

No caso das vibrações moleculares, o modelo do oscilador harmônico já não é inteiramente válido e modificações devem ser feitas. Dessa forma as vibrações moleculares são melhor explicadas através da mecânica quântica.

Tendo-se em mente um oscilador macroscópico, é natural imaginar que uma maior amplitude no deslocamento da massa pode ser obtida ao fornecer-se um pouco mais de energia inicial ao sistema, podendo-se repetir esse processo para quaisquer valores de energia de interesse. Nesse sentido, o oscilador harmônico absorve energia de forma contínua. Entretanto estudos indicam que um sistema molecular que desenvolve um movimento de vibração não varia sua energia potencial de forma contínua, para que haja o aumento da amplitude de

vibração a molécula deve receber uma quantidade definida de energia, possuindo assim estados vibracionais associados a níveis de energia definidos, que a rigor são referidos como *níveis permitidos*.

Ao incidir-se um feixe de radiação infravermelha, também é constatado que apenas moléculas que possuem uma separação efetiva de cargas elétricas podem absorver a energia. Desta forma a separação de cargas, definida como o momento de dipolo permanente (μ), pode variar no decorrer da vibração da molécula. Sendo assim o campo elétrico oscilante da radiação pode acoplar com o campo elétrico oscilante da molécula resultante da variação periódica do momento de dipolo, transferindo energia e fazendo com que os átomos vibrem com maior amplitude.

$$\frac{\partial \mu}{\partial q} \neq 0 \tag{3.14}$$

Aos níveis de energia permitidos em um sistema molecular são atribuídos números inteiros e positivos, incluindo o zero e são denominados números quânticos vibracionais (v_q). Em contrapartida ao sistema massa-mola convencional, para $v = 0$ a energia correspondente não é nula e não pode ser exaurida por menor que seja a temperatura do sistema.

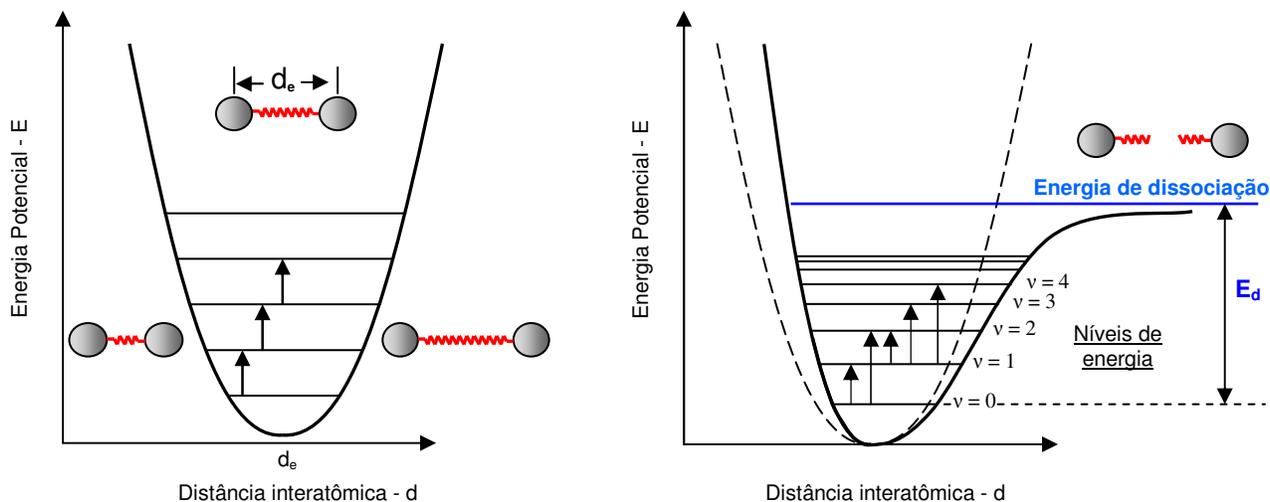


Figura 07 – Diagramas de energia potencial. (a) oscilador harmônico quantizado, (b) oscilador anarmônico, d_e = distância de equilíbrio, E_d = energia de dissociação.

Devido às características da ligação química, a curva de energia potencial perde seu perfil simétrico e assume um comportamento assintótico no sentido de alongação da "mola", sendo limitada pela energia de dissociação da molécula. Desta forma, os átomos não podem ser afastados indefinidamente, uma vez que ocorrerá a quebra da ligação e o sistema oscilador deixará de existir. Essas particularidades introduzem no sistema um caráter anarmônico que também é responsável pela distorção do espaçamento entre os níveis de energia esperados para a molécula. Idealmente a diferença de energia entre níveis adjacentes é idêntica e as mudanças de estado vibracional são permitidas apenas para o nível vizinho superior ou inferior ($\Delta v_q = \pm 1$), todavia, a anarmonicidade diminui o espaçamento energético entre os estados vibracionais a medida que o número quântico vibracional cresce, além de tornar possível saltos quânticos ou sobretons para estados vibracionais mais distantes do estado fundamental ou do estado excitado em que a molécula se encontra ($\Delta v_q = \pm i, i = 1, 2, 3, 4, \dots$).

Experimentalmente verificam-se transições apenas para os primeiros 4 números quânticos vibracionais, apesar das demais serem previstas teoricamente. Diferenças significativas entre as massas dos dois átomos, bem como uma elevada eletronegatividade de um dos átomos também são responsáveis pelo aumento do caráter anarmônico do oscilador.

A curva de energia potencial apresentada na Figura 07 pode ser razoavelmente aproximada pela função de Morse:

$$V = E_d \cdot [1 - e^{-\sqrt{k_m/2} \cdot E_d \cdot d}]^2 \quad (3.15)$$

Onde E_d = energia de dissociação,

k_m = constante de força,

d = comprimento da ligação.

Os valores discretos de energia podem ser obtidos resolvendo-se equações de onda com o auxílio da mecânica quântica.

$$E = \left(v_q + \frac{1}{2} \right) \left(\frac{h}{2\pi} \right) \sqrt{\frac{k}{\mu}} \quad (3.16)$$

Onde,

h = constante de Planck.

v = número quântico vibracional, o qual assume apenas valores positivos inteiros.

Substituindo-se a equação 3.8 na equação 3.16, tem-se:

$$E = \left(v_q + \frac{1}{2} \right) v_m \quad (3.17)$$

Expressão que relaciona a energia dos níveis vibracionais com a frequência da radiação eletromagnética necessária para a excitação molecular.

3.2.1 - Modos vibracionais

O número de modos vibracionais em uma molécula pode ser calculado considerando-se que cada átomo possui 3 possíveis direções para movimentar-se, chamadas de graus de liberdade. Desta forma uma molécula diatômica possui $3 \times 2 = 6$ graus de liberdade, entretanto quando ambos os átomos se deslocam para uma mesma direção tem-se um movimento de translação que pode ser na direção x , y ou z , ou seja dos 6 graus de liberdade 3 são utilizados para descrever o movimento translacional do centro de massa. Analogamente 3 graus de liberdade são necessários para descrever os movimentos rotacionais da molécula. Generalizando o raciocínio acima, para uma molécula não linear o número máximo de vibrações deve ser igual $3N-6$, onde N é o número de átomos; no caso de uma molécula linear não é considerada a rotação em torno do eixo da molécula, neste caso existem apenas 2 graus de liberdade e espera-se então, um máximo de $3N-5$ formas de vibração. Todos os $3N-6$ ou $3N-5$ modos vibracionais de uma molécula são denominadas modos normais ou fundamentais de vibração.

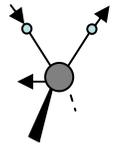
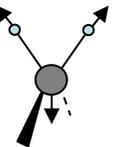
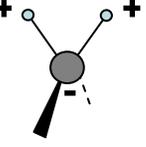
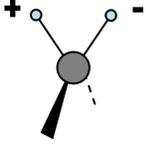
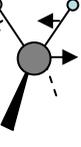
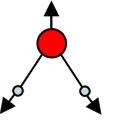
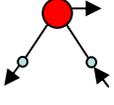
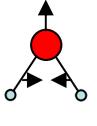
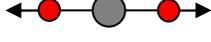
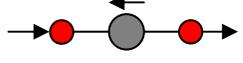
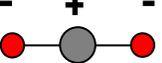
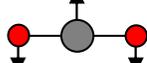
No caso específico dos compostos utilizados neste trabalho, os principais modos de vibrações esperados são apresentados na Tabela 02.

Ao incidir-se um feixe de radiação na região do infravermelho sobre uma determinada amostra, em princípio todos estes modos normais de vibração deveriam absorver uma parcela da radiação e, portanto, apresentariam um sinal correspondente no espectro de absorção. Entretanto, experimentalmente o número de picos de absorção pode ser menor ou maior que os previstos, isto se deve à presença de modos degenerados (quando as energias de 2 modos vibracionais são idênticas), ou quando uma determinada vibração não produz uma variação periódica no momento de dipolo permanente, ao fenômeno de ressonância, à combinação de determinados modos vibracionais ou ainda à presença de sobretons que são as freqüências cujos valores são múltiplos inteiros das freqüências de vibração fundamentais (tons). Para os casos de combinação de vibrações e ressonância, podem ocorrer, por exemplo, combinações entre dois modos fundamentais; ou em moléculas com muitas ligações é possível que haja diferentes vibrações cuja diferença de energia seja mínima, o que torna possível a interação entre um modo normal e um sobreton (ressonância de Fermi) ou ainda a interação entre dois sobretons ou entre sobretons e uma banda de combinação (ressonância de Darling-Dennison). As freqüências de vibração também podem ser afetadas pelo acoplamento entre vibrações vizinhas.

É importante destacar que a maioria dos modos fundamentais de vibração é observada na região do infravermelho médio. Sendo possível correlacionar os picos de absorbância presentes em um espectro com as ligações químicas. Entretanto o mesmo não ocorre na região do infravermelho próximo, uma vez que nesta região tem-se uma elevada sobreposição dos picos devido à grande quantidade de informação e também pelo fato de que a informação em grande parte corresponde a interações entre os modos vibracionais, fenômenos de ressonância e à presença de sobretons.

Neste caso, somente através de métodos quimiométricos torna-se viável a utilização desta região espectral para a quantificação de compostos químicos.

Tabela 02 – Principais modos vibracionais das moléculas utilizadas neste trabalho.

MOLÉCULAS	MODOS VIBRACIONAIS		
Metano Etano Propano Butano 1-buteno 1,3-butadieno			
	$\nu_{as} \text{CH}_2$ (2926 cm^{-1})	$\nu_s \text{CH}_2$ (2853 cm^{-1})	$\delta_s \text{CH}_2$ (1465 cm^{-1})
			
ωCH_2	τCH_2	ρCH_2 (722 cm^{-1})	
Água			
	$\nu_s \text{OH}$ (3652 cm^{-1})	$\nu_{as} \text{OH}$ (3756 cm^{-1})	$\delta_s \text{HOH}$ (1596 cm^{-1})
	Dióxido de carbono		
$\nu_s \text{CO}$ (1340 cm^{-1})		$\nu_{as} \text{CO}$ (2350 cm^{-1})	
			
$\delta_s \text{CO}_2$ (666 cm^{-1})		$\delta_s \text{CO}_2$ (666 cm^{-1})	

Simbologia: ν : estiramento (ν_{as} : assimétrico, ν_s : simétrico); δ : deformação angular no plano (δ_s : simétrico “scissoring”, δ_{as} : assimétrico); ω : deformação angular fora do plano (wagging), τ : deformação angular fora do plano (twisting), ρ : deformação angular no plano (rocking).

3.3 - MEDIDAS DE ABSORBÂNCIA

Para medidas quantitativas em espectroscopia na região do infravermelho, utilizam-se as leis de Bourguer/Lambert/Beer e Kubelka-Munk. A primeira relaciona a capacidade das moléculas de absorver a energia de um feixe de radiação que atravessa a amostra, e a segunda relaciona a absorção de radiação que sofre espalhamento em uma amostra, geralmente sólida. Neste trabalho utilizou-se unicamente a Lei de Bouguer/Lambert/Beer^[78] (ver apêndice 4).

Inicialmente Bouguer concluiu que a razão entre a intensidade inicial, P_0 , e a final, P , de um feixe de radiação que atravessa uma porção de um material de espessura definida, b , é sempre constante e denominou esta constante de transmitância, equação 3.18. Posteriormente, descobriu que a transmitância diminui em progressão geométrica com o aumento da espessura do material (caminho óptico).

Já Beer aplicou estas conclusões para casos onde o material é na realidade uma solução e, além da possibilidade de caminho óptico variável, a capacidade de absorção do “material” também pode ser modificada alterando-se, por exemplo, sua concentração.

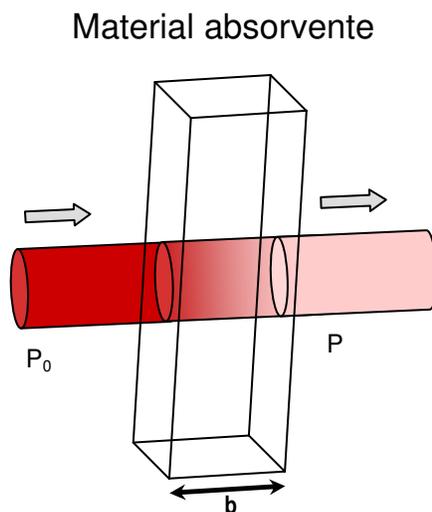


Figura 08 – Atenuação da radiação monocromática ao atravessar um material absorvente de espessura b (caminho óptico). Neste caso, $P_0 > P$, onde P_0 e P são respectivamente a intensidade inicial e final do feixe de radiação.

A forma final da lei de Bouguer/Lamber/Beer, expressa na equação 3.19, correlaciona a absorbância (não confundir com absorção), o caminho óptico e a concentração da solução.

$$T = \frac{P}{P_0} \quad (3.18)$$

$$A = \varepsilon \cdot b \cdot c \quad \text{Lei de Bouguert/Lambert/Beer} \quad (3.19)$$

Onde,

T: transmitância

P: intensidade final do feixe de radiação.

P₀: intensidade inicial do feixe de radiação.

A: absorbância.

ε: coeficiente de absorvidade molar em L.mol⁻¹.L⁻¹.

b: caminho óptico em cm.

c: concentração da solução em mol.L⁻¹.

3.4 - INSTRUMENTAÇÃO

Os equipamentos utilizados em espectroscopia contêm tipicamente os seguintes componentes^[46].

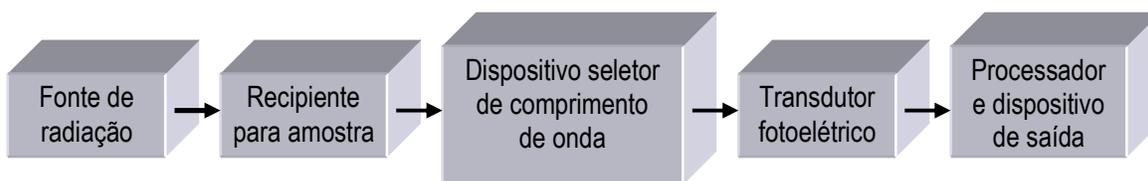


Figura 09 – Componentes comumente encontrados me equipamentos utilizados para medidas espectrofotométricas.

Atualmente, os instrumentos disponíveis no mercado diferem com relação ao sistema utilizado para realizar a seleção dos comprimentos de onda. Neste sentido existem 3 classificações: espectrofotômetros dispersivos,

espectrofotômetros por Transformada de Fourier e espectrofotômetros que utilizam filtros ópticos.

3.4.1 - Espectrofotômetros Dispersivos

São equipamentos mais antigos e que utilizam redes de difração para separação dos comprimentos de onda da luz. Em geral possuem muitos espelhos e partes móveis e desta forma são mais lentos na aquisição dos dados. Entretanto muitos instrumentos são providos de um sistema de feixe duplo, o que compensa as variações na fonte de radiações e outros tipos de flutuações.

3.4.2 - Espectrofotômetros por Transformada de Fourier

Funcionam com base em uma operação matemática conhecida como transformada de Fourier. Estes instrumentos não fazem uso de um dispositivo mecânico para separar os comprimentos de onda da luz, neste caso toda a radiação emitida pela fonte e que interage com a amostra é posteriormente direcionada ao detector. Todavia, entre a amostra e o detector é posicionado um dispositivo conhecido como interferômetro, o qual possui um divisor de feixes (*beam splitter*) responsável por refletir, para o detector, cerca de 50% da radiação incidente e permitir a passagem da outra metade a qual atinge um espelho móvel. O espelho se movimenta de forma a fazer com que a radiação percorra em determinados momentos uma distância igual a um múltiplo inteiro de algum dos comprimentos de onda da radiação incidente, promovendo desta maneira momentos de interferência construtiva ou destrutiva quando a radiação atinge o detector. Esta é uma condição matemática para que a transformada possa ser calculada, o software do equipamento obtém então o interferograma o qual representa a intensidade registrada no detector em função do tempo, o qual está indiretamente relacionado ao deslocamento do espelho móvel. A transformada de Fourier permite associar a intensidade luminosa registrada à frequência da radiação, passando a informação do domínio do tempo para a frequência (ou comprimento de onda), ou seja, transformando o interferograma em um espectro.

Os espectrofotômetros baseados em Transformada de Fourier possuem vantagens como uma alta velocidade de aquisição de dados, devido à detecção simultânea de todos os comprimentos de onda (conhecida como vantagem de *Felgett*). Além da ausência de fendas para aumento da resolução, neste caso toda a radiação é recebida pelo detector gerando uma melhor relação sinal-ruído (conhecida como vantagem de *Jaquinot*), além do fato de possuírem apenas um único espelho móvel.

3.4.3 - Espectrofotômetros com filtros ópticos

Tais instrumentos utilizam filtros de interferência (também conhecidos como filtros *Fabryt-Perot*) ou filtros óptico-acústicos para promover a separação dos comprimentos de onda da radiação. Existem ainda os filtros de absorção, mas estes são restritos a medidas na região visível do espectro.

Os filtros de interferência funcionam com base no fenômeno de interferência construtiva e destrutiva da radiação. Estes dispositivos consistem em um dielétrico transparente, em geral de CaF_2 ou MgF_2 , entre dois filmes metálicos semitransparentes, sendo a espessura do material dielétrico determinante no controle da resolução do comprimento de onda da radiação transmitida. Este arranjo é então prensado entre 2 placas de vidro ou outro material transparente.

3.4.3.1 - Filtro óptico-acústico sintonizável (AOTF)

Já o filtro AOTF^[79] é um dispositivo que utiliza um cristal birefringente de TeO_2 em contato com um material piezoelétrico (geralmente LiNiO_4), o qual produz uma onda mecânica quando excitado por uma fonte de rádio frequência. Ao ser excitado o material piezoelétrico produz uma série de ondas mecânicas que se propagam pelo cristal e através da compressão e expansão periódica do material produzem uma variação controlada da densidade e conseqüentemente do índice de refração do meio. A interação da onda eletromagnética e da onda acústica faz com que o cristal refrate seletivamente uma estreita faixa de comprimentos de onda da radiação que o atravessa.

O fenômeno da difração ocorrendo em um cristal birefringente faz com que o feixe de luz incidente origine dois novos feixes de radiação (por simplificação é mostrado apenas um dos feixes difratados na Figura 10), neste caso monocromáticos e com separação angular entre 8 a 10 graus, além do feixe policromático residual (não difratado) transmitido pelo cristal.

O filtro AOTF empregado neste equipamento é semelhante ao da Figura - 10. O dispositivo opera em uma configuração não-colinear, indicando que a onda acústica e o feixe de radiação se propagam em direções quase perpendiculares. O conceito do monocromador do tipo AOTF, totalmente sem partes móveis, confere aos novos equipamentos características apropriadas para uso em campo ou diretamente associados a processos industriais.

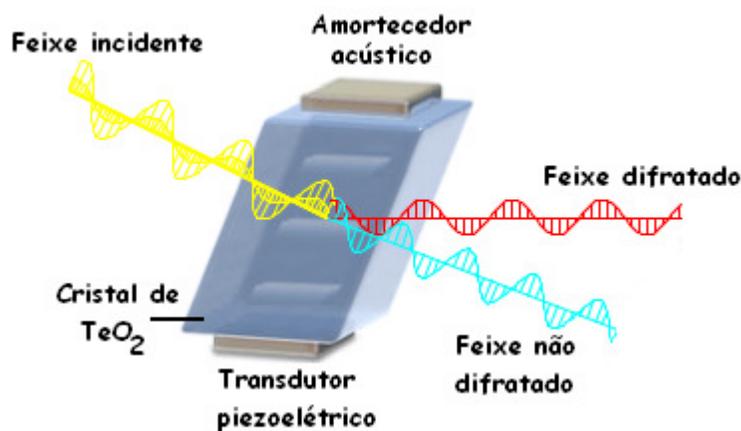


Figura 10 – Monocromador do tipo AOTF.

3.5 - MEDIDAS NO INFRAVERMELHO PRÓXIMO DE AMOSTRAS GASOSAS

Diferentemente dos estados sólido e líquido, no estado gasoso ou no vapor os átomos ou moléculas possuem uma elevada separação física. Desta forma as perturbações elétricas e magnéticas sofridas com maior intensidade quando existe a alta proximidade das moléculas, praticamente desaparecem ao considerar-se o estado gasoso. Este fato é facilmente compreendido quando consideramos que o raio de atuação da força elétrica decresce rapidamente com o inverso do

quadrado da distância. Ao contrário dos espectros NIR de amostras líquidas ou sólidas onde um sinal, de absorção ou transmissão, se apresenta geralmente como uma banda centrada no comprimento de onda de interesse e cuja informação principal é de natureza vibracional. No estado gasoso esperam-se espectros cujos sinais se apresentam como picos os quais carregam informação de origem vibracional e rotacional.

Talvez a maior desvantagem das medidas com amostras gasosas seja a intensidade do sinal registrado, que costuma ser várias vezes inferior ao de uma amostra nos demais estados físicos, podendo sofrer, neste caso, grande distorção devido à relação sinal/ruído desfavorável.

O aumento do caminho óptico quando se trabalha com amostras gasosas já não consiste em um artifício tão simples, pois geralmente as células de gás possuem vários centímetros, chegando a alguns metros em determinadas situações. Uma das soluções consiste na utilização aumento da pressão no interior da célula, entretanto esta opção insere algumas complicações práticas, além de materiais especiais, pois todo o sistema deve resistir à alta pressão. Entretanto existem soluções inovadoras que contornam a necessidade do aumento da pressão. Neste caso células de gás são equipadas com espelhos internos de maneira a permitir múltiplas reflexões do feixe de radiação, simulando um caminho óptico maior e possibilitando obter valores de absorbância na ordem de magnitude desejada. Esta tem sido uma alternativa bastante adequada, menos onerosa e com menor risco de acidentes para o estudo em laboratório de amostras gasosas sem a necessidade de montagem de células de fluxo ou sistemas de amostragem adaptados para trabalhar em altas pressões.

Makhoukhi et al^[80] avaliaram o uso da espectroscopia nas regiões do infravermelho próximo (NIR) e médio (MIR), aliado à calibração multivariada na determinação de metano, etano e propano em misturas sintéticas de composição semelhante ao gás natural. O estudo indica que o uso do modelo de calibração PLS e da região do infravermelho próximo proporciona melhores resultados quando comparados com a técnica do MIR.

Na publicação *Near-Infrared Spectral Analysis of Gas Mixtures* de Mullins et al^[31], encontramos um estudo semelhante. Mas neste caso, os espectros NIR são obtidos de misturas gasosas a elevadas pressões e temperatura, simulando as condições encontradas nas diferentes fases de produção e transporte do gás natural. Mesmo nestes casos, os resultados obtidos são comparáveis com aqueles encontrados por cromatografia gasosa, demonstrando a confiabilidade da técnica.

Para o acompanhamento de processos envolvendo sistemas gasosos, a espectroscopia NIR foi avaliada por Hans et al^[81] quanto à sua capacidade de monitoramento do rápido processo de separação de iso- e ciclo-alcanos de alcanos normais mediante o uso de colunas de peneiras moleculares. A região espectral do infravermelho próximo quando utilizada para a construção de uma biblioteca de espectros de espécies gasosas^[82] revelou-se bastante apropriada para identificação de misturas de gases. Indicando que trata-se de uma porção do espectro rica em informações. Adicionalmente a região do NIR associada à calibração multivariada também já foi utilizada para determinação do conteúdo energético do gás natural^[83-85].

Estes trabalhos confirmam que a espectroscopia no infravermelho próximo é altamente indicada como técnica analítica a ser empregada em um equipamento destinado à quantificação das espécies gasosas presentes em gases combustíveis. Podendo ter aplicações importantes como o controle de qualidade durante todo o processo de geração do gás ou no abastecimento em postos de gás natural veicular.

Capítulo

4

**PARTE
EXPERIMENTAL**

4.0 – EQUIPAMENTO EXPERIMENTAL NIR-AOTF

Neste trabalho utilizou-se um equipamento construído no próprio laboratório, Figura 11, baseado em um filtro óptico-acústico sintonizável ou AOTF (do inglês Acousto Optical Tunable Filter).

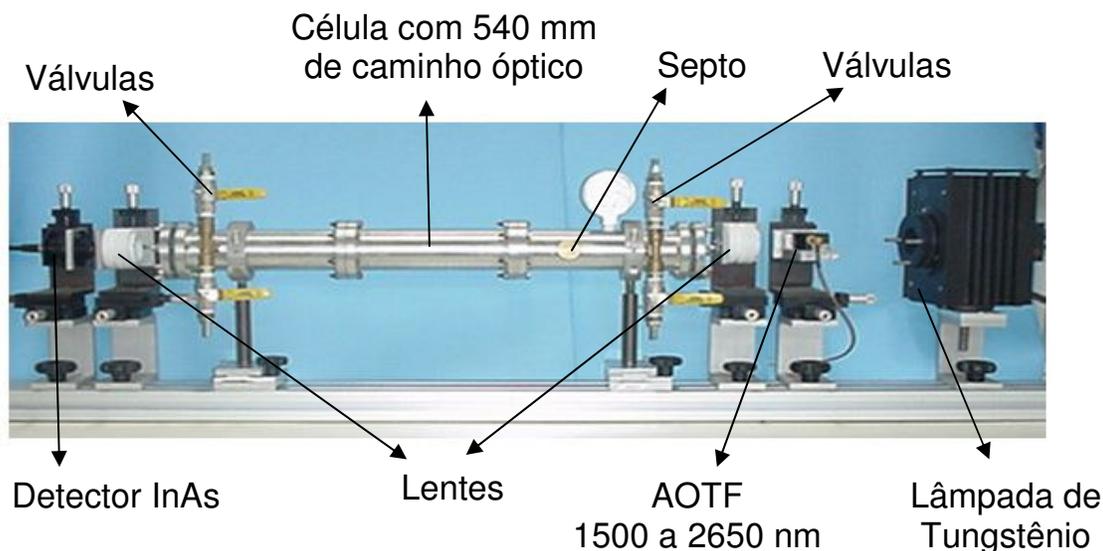


Figura 11 – Equipamento experimental NIR-AOTF.

O equipamento consiste em um filtro seletor de comprimento de onda do tipo AOTF (*Brimrose*®, modelo TEAF% - 1.5-2.65 UH), um gerador de rádio frequência (*Brimrose*®) com faixa de trabalho de 1500 a 2650 nm e resolução de 4nm a 14nm e um amplificador lock-in (*Stanford Research Systems / SR 830*). Uma fonte de alimentação elétrica *Iso-Tech*® foi utilizada para o funcionamento de uma lâmpada de tungstênio (*Osram* - 50 W); a radiação é colimada por meio de lentes de CaF₂ (fabricadas pela *Ealing Electro Optics*) e mensurada através de um detector de arseneto de índio (InAs, EG&G) provido de um resfriador baseado em um sistema peltier. Todo o controle do instrumento é feito através de um microcomputador compatível com um PC e um programa escrito em *Visual Basic 3.0 (Microsoft)*. A célula de gás empregada tem caminho óptico de 54 cm e volume de 490 mL, possuindo um medidor de pressão acoplado. Os espectros registrados com este equipamento são a média de 5 aquisições

Além deste equipamento também utilizou-se paralelamente um espectrofotômetro comercial NIR-FT da marca Bomem, modelo D160, operando na mesma faixa espectral do NIR-AOTF, no entanto com resolução de 4cm^{-1} . Os espectros obtidos neste caso foram a média de 100 *scans*.

4.1– SISTEMA DE PREPARAÇÃO DAS AMOSTRAS GASOSAS

Para a preparação das amostras foi empregado um sistema baseado em fluxímetros, da marca *Alborg®*, para controle da vazão dos gases e conseqüentemente de suas razões volumétricas nas misturas. Os gases; metano, etano, propano, butano, dióxido de carbono, 1-buteno e 1,3-butadieno; eram conduzidos dos cilindros de alta pressão por tubos de PTFE ou Poliamida de $\frac{1}{4}$ de polegada passando pelos respectivos controladores de fluxo, sendo direcionados em seguida a um dispositivo misturador. Este foi confeccionado em um cilindro de latão contendo entradas individuais para cada gás e uma única saída. Os canais internos do misturador apresentam diâmetro de 5 mm visando melhorar as misturas entre os gases. Após a etapa de homogeneização, a mistura gasosa era conduzida para a célula espectrofotométrica cujo fluxo permanecia por um período de 5 minutos garantindo a limpeza da célula. Após este período o fluxo as válvulas de entrada e saída da célula eram fechadas e o fluxo interrompido, sendo o espectro obtido logo em seguida.

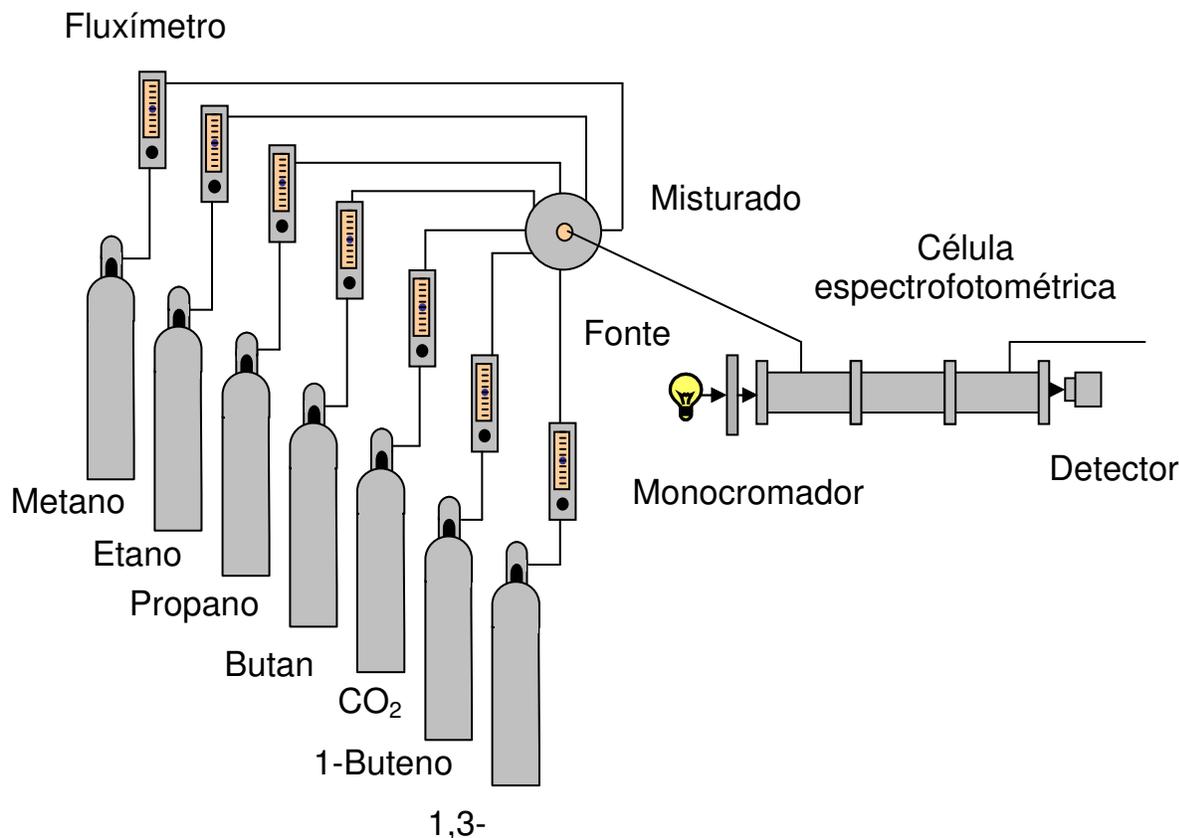


Figura 12 – Sistema de preparo das misturas gasosas.

À configuração principal, ilustrada na Figura 12, foram adaptados outros dispositivos de acordo com a composição das misturas a serem preparadas ou da necessidade de acoplar-se outro equipamento para obtenção simultânea dos espectros no instrumento experimental NIR-AOTF e um espectrofotômetro FT-NIR comercial.

Para o preparo de misturas com teores controlados de umidade, desenvolveu-se um sistema de inserção de umidade através da injeção de água líquida em uma célula de evaporação construída em acrílico dotada de um septo de teflon, Figura 13. Após a injeção do volume de água desejado, uma bomba promovia a circulação da mistura entre a célula de medida e a célula de evaporação até a completa evaporação da água, como ilustrado na Figura 14.

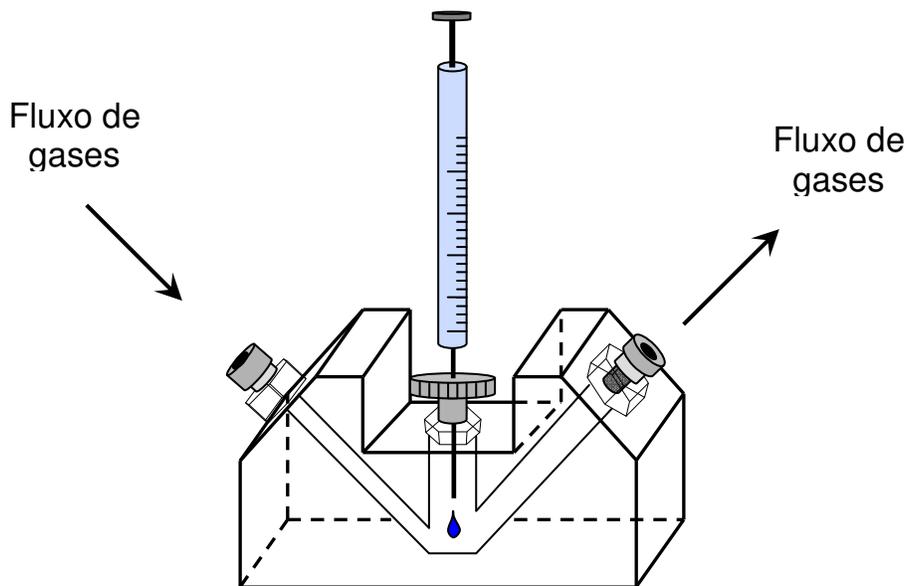


Figura 13 – Peça evaporadora de acrílico.

Nos experimentos em que os dois espectrofotômetros foram utilizados simultaneamente para a obtenção de espectros, um tubo foi adicionado ao sistema permitindo a transferência da mistura gasosa da célula de medida do equipamento NIR-AOTF diretamente para a célula de medida do espectrofotômetro NIR-FT, Figura 15. Desta forma, uma mesma mistura tinha seu espectro obtido em ambos equipamentos simultaneamente, evitando diferenças devidos ao preparo de amostra.

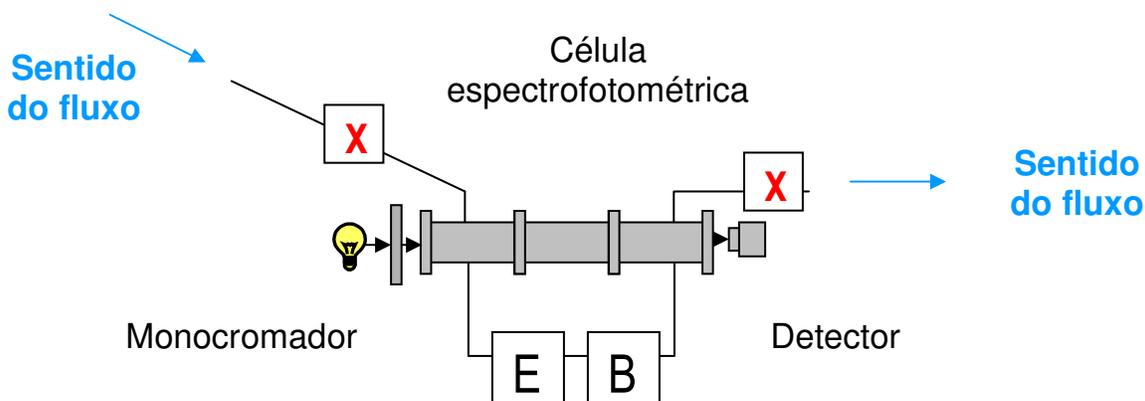


Figura 14 - Configuração para utilização da célula de evaporação. O símbolo "E" representa a célula de evaporação e "B" representa uma bomba utilizada para o refluxo da amostra gasosa. "X", Indica que as válvulas de entrada e saída da célula de medida permanecem fechadas durante a etapa de evaporação.

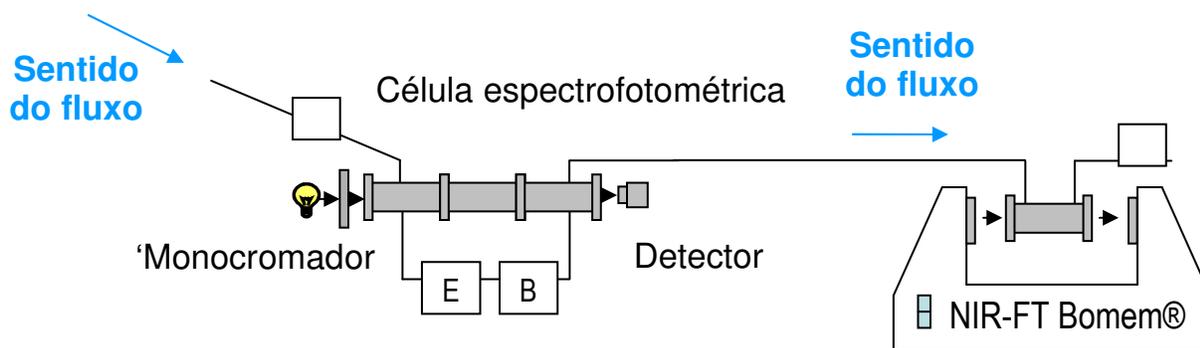


Figura 15 - Configuração para utilização a obtenção simultânea no equipamento NIR-AOTF e no NIR-FT (Bomem®, Camo®).

4.2 – PLANEJAMENTO DE MISTURAS

Os conjuntos de misturas foram preparados de acordo com as faixas de porcentagem volumétrica apresentadas na Tabela 03. Estas faixas foram escolhidas com base na portaria ANP n° 104 a qual regulamenta os parâmetros de qualidade dos gases naturais combustíveis.

Tabela 03 – faixas de valores das razões volumétricas empregadas nos conjuntos de calibração.

AMPLITUDE DE VALORES	RAZÃO VOLUMÉTRICA - % (V/V)								CONJUNTO DE AMOSTRAS
	METANO	ETANO	PROPANO	BUTANO	CO ₂	UMIDADE	1-BUTENO	1,3-BUTADIENO	
MÁXIMO	92,1	25,0	9,2	2,9	6,8	1,3	-	-	Conj. 1
MÍNIMO	63,0	3,0	0,7	0,7	1,7	0,3	-	-	
MÁXIMO	87,6	23,5	6,7	4,0	6,4	-	4,9	4,9	Conj. 2
MÍNIMO	61,7	3,1	1,0	0,5	2,1	-	0,5	0,8	

Utilizou-se o software Unscrambler® para o cálculo de um novo conjunto de calibração (referido como C1) por meio de um planejamento fatorial completo de 6 fatores em 2 níveis, o qual gerou um conjunto de 64 amostras. Este conjunto foi complementado com mais 21 amostras estrategicamente escolhidas.

O segundo conjunto (referido como C2) foi calculado por uma metodologia própria, na qual o número total de amostras é escolhido pelo usuário, além de possuir vantagens como um elevado número de níveis e baixa correlação entre as variáveis.

O conjunto C1 preparado a partir do software Unscrambler® era composto por 6 espécies: metano, etano, propano, butano, dióxido de carbono e vapor e água (umidade) e continha 85 amostras.

O conjunto com os melhores resultados, C2, continha os gases metano, etano, propano, butano, dióxido de carbono, 1-buteno e 1,3-butadieno e era composto por 40 amostras, Tabela 03.

Para a preparação das misturas, os valores da Tabela 03 eram convertidos em unidades de fluxo e posteriormente transformados nos valores presentes nas escalas dos fluxímetros. A correspondência entre os valores arbitrários das escalas dos fluxímetros e os valores reais de fluxo (mL/min) foi obtida por meio da calibração com um bolhometro, aplicando-se o processo de regressão polinomial para a obtenção as equações de calibração.

Tabela 04 – conjunto de calibração final, C2.

N° da amostra	RAZÃO VOLUMÉTRICA - % V/V						
	Metano	Etano	Propano	Butano	CO ₂	1-buteno	1,3-butadieno
1°	78,1	10,4	2,7	1,7	2,1	1,2	3,7
2°	70,5	6,4	5,0	4,0	5,6	4,5	3,9
3°	78,9	10,2	1,3	3,0	2,6	0,5	3,3
4°	70,2	12,0	3,3	2,6	5,2	3,0	3,7
5°	73,1	8,8	5,9	2,9	3,4	4,7	1,3
6°	64,3	20,4	5,3	2,0	2,5	2,8	2,8
7°	76,0	6,4	5,2	4,0	2,8	4,6	1,0
8°	79,2	3,1	5,9	3,9	5,2	1,7	1,0
9°	72,1	15,7	5,9	0,9	2,3	0,6	2,3
10°	70,4	16,2	3,0	2,1	4,3	3,0	1,0
11°	79,7	3,9	1,3	2,5	5,5	4,9	2,2
12°	70,9	15,7	3,0	3,0	4,9	1,5	1,0
13°	71,6	20,4	1,1	1,1	2,6	1,7	1,5
14°	78,4	5,2	3,1	0,9	5,6	4,9	1,8
15°	71,9	14,3	2,8	2,1	3,6	0,8	4,6
16°	87,6	5,0	1,3	1,2	2,6	1,1	1,3
17°	79,7	8,1	5,8	2,4	2,6	0,6	0,9
18°	83,7	4,8	1,6	0,7	2,8	2,4	4,0
10°	80,4	4,2	1,4	1,6	3,3	4,3	4,9
20°	66,1	17,5	1,4	3,4	6,4	0,8	4,5
21°	73,4	12,9	5,7	2,8	2,1	1,2	1,9
22°	77,8	5,6	6,7	1,2	5,7	1,2	1,9
23°	63,2	23,5	1,7	3,7	2,4	2,3	3,2
24°	85,7	4,0	1,2	1,2	3,1	0,8	4,0
25°	61,7	22,5	1,9	1,5	4,3	4,5	3,6
26°	73,5	15,0	4,5	0,9	2,3	1,5	2,3
27°	70,8	18,6	3,2	3,0	2,8	0,6	1,0
28°	74,6	8,3	1,7	2,6	5,2	3,2	4,5
29°	83,6	10,0	1,2	0,8	2,3	1,0	1,1
30°	74,9	13,2	1,7	1,1	4,7	3,2	1,1
31°	70,8	18,7	2,1	1,3	3,6	1,4	2,0
32°	81,7	7,5	1,4	1,0	3,0	3,8	1,6
33°	73,0	11,8	4,0	1,4	4,3	1,5	4,1
34°	78,5	4,7	1,0	2,8	5,9	4,3	2,8
35°	75,5	9,6	5,1	1,5	3,5	2,1	2,7
36°	64,5	23,5	2,6	0,6	3,1	3,0	2,7
37°	83,7	4,1	1,3	0,5	6,1	2,6	1,7
38°	76,0	7,4	4,1	0,8	5,9	3,4	2,5
39°	68,7	22,1	2,6	0,6	3,6	1,6	0,8
40°	64,0	20,4	2,1	2,9	4,1	4,3	2,2

Capítulo

5

**CALIBRAÇÃO DE
ESPÉCIES GASOSAS
PRESENTES EM
AMOSTRAS SINTÉTICAS
DE GÁS NATURAL**

5.0 – RESULTADOS E DISCUSSÃO

A Tabela 05 apresenta os resultados para os modelos de calibração envolvendo o primeiro conjunto de amostras. Para a construção dos modelos de calibração foi realizada a correção de linha base dos espectros e eliminação das primeiras 49 variáveis (presentes na faixa de 1524nm a 1627nm) que representavam unicamente ruído de difícil correção.

Os melhores modelos foram escolhidos com base nos valores de RMSECV, sendo necessárias de 6 a 7 variáveis latentes para extrair a informação necessária da matriz de dados. Estes números estão coerentes com a quantidade de espécies químicas presente nas misturas.

Os valores dos erros de calibração de cada espécie são próximos dos valores dos erros de validação indicando que o modelo está bem ajustado. Este fato também pode ser confirmado pelos valores de correlação e do coeficiente angular da reta ajustada entre os valores reais, de porcentagem volumétrica, e os previstos pelos modelos de calibração. Exceção é observada para o butano que apresenta valores de correlação e inclinação inferiores a 0,95. Os valores dos coeficientes lineares são próximos a zero indicando que não há erro sistemático nos modelos construídos. Os valores de RMSECV representam erros percentuais de 0,6% para o metano; 2,7% para o etano; 5,5% para o propano; 11,4% para o butano; 5,8% para o CO₂ e 8,9% para o teor de umidade, considerando nestes casos os valores médios das faixas de porcentagem volumétrica para cada uma das espécies.

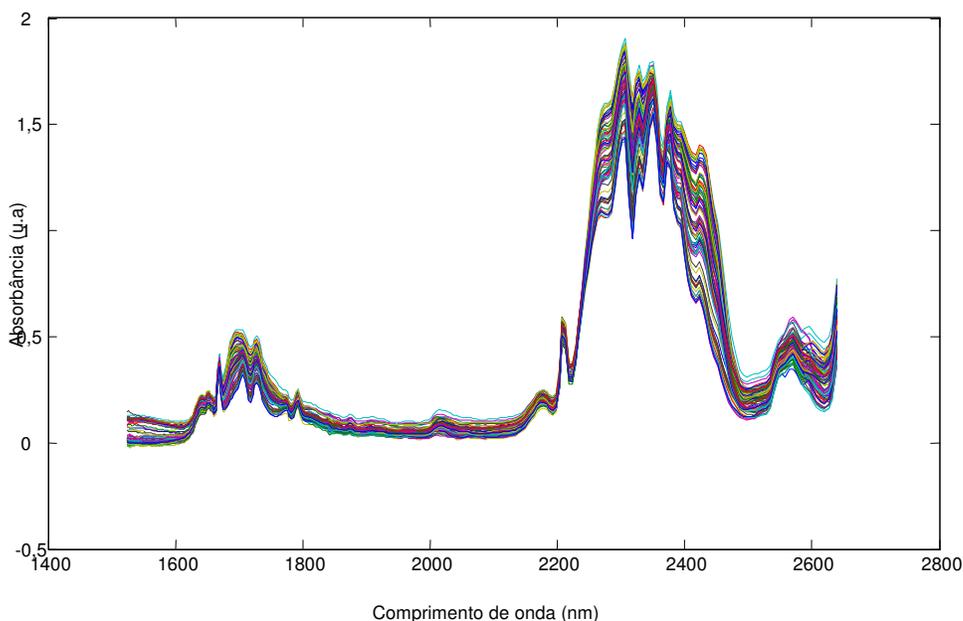


Figura 16 - espectros para o conjunto de calibração com 85 amostras, obtidos com espectrofotômetro experimental baseado em AOTF.

Tabela 05 – Resultados comparativos dos modelos PLS-1 para as 6 espécies químicas.

Espécie química	VL	RMSEC	RMSECV	Coefficiente de Correlação (real x previsto ^a)	Coefficiente angular	Coefficiente linear (% v/v)
METANO	6	0,4115	0,4680	0,9983	0,9969	0,2413
ETANO	6	0,3294	0,3768	0,9990	0,9978	0,0251
PROPANO	7	0,2152	0,2654	0,9961	0,9936	0,0307
BUTANO	7	0,1752	0,2125	0,9467	0,9107	0,1334
CO₂	6	0,2278	0,2611	0,9851	0,9686	0,1133
UMIDADE	6	2,3304	2,6735	0,9727	0,9519	1,1710

(DADOS OBTIDOS COM O EQUIPAMENTO EXPERIMENTAL BASEADO EM AOTF – 85 amostras)

^a: valores previstos na validação cruzada (*leave one out*), **Obs.:** em todos os casos foi realizado ajuste de linha base e eliminação das primeiras 49 variáveis (1524nm – 1627nm) que representavam unicamente ruído de difícil correção. Para o cálculo dos modelos todos os dados foram previamente centrados na média, VL: variáveis latentes.

Apesar de apresentar erros de duas unidades, a determinação do teor de umidade nas misturas gasosas não apresenta repetibilidade. Este fato está relacionado a problemas de adsorção do vapor de água nas paredes internas da

cela espectrofotométrica, o que gera resultados com baixa repetibilidade, afetando assim os erros de calibração e previsão.

Devido à dificuldade no preparo de amostras com resultados coerentes referentes ao teor de água, um novo conjunto foi preparado substituindo-se este parâmetro por dois de hidrocarbonetos gasosos, 1-buteno e 1,3-butadieno. Estas misturas foram empregadas para determinação das principais figuras de mérito do procedimento analítico. A Figura 17 apresenta os espectros do segundo conjunto de calibração obtido com o espectrofotômetro baseado em AOTF'

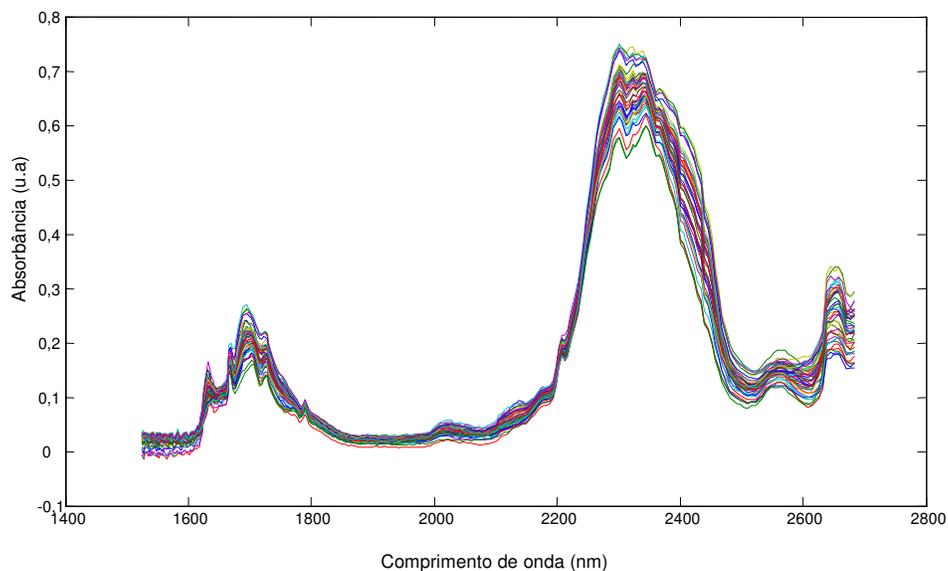


Figura 17 - espectros para o conjunto de calibração com 40 amostras, obtidos com espectrofotômetro experimental baseado em AOTF.

Comparando-se os espectros do primeiro conjunto de amostras (Figura 16) com os apresentados na Figura 17, observa-se claramente uma perda de resolução onde o conjunto de 40 amostras possui espectros com resolução inferior ao primeiro grupo. Isto é consequência de uma falha ocorrida com o dispositivo monocromador de AOTF que teve provavelmente a falha no funcionamento de alguns piezoelétricos responsáveis pela vibração do cristal de Te_2O , e que deve ter modificado a resolução do monocromador.

Uma segunda observação importante com relação aos dois conjuntos de espectros, refere-se à região final próximo a 2650 nm na Figura 16, a qual destaca-se pelo fato de haver um aumento na absorção inexistente na Figura 17. Esta observação é decorrente do aparecimento do início da banda fundamental da molécula de água, a qual ocorre em 2778 nm.

Visando otimizar os novos modelos a serem construídos, uma vez que os espectros apresentam ruído de até 20 miliabsorbância (0,020 u.a.), um estudo utilizando o pré-tratamento alisamento por janela móvel foi realizado para descobrir a quantidade ótima de pontos da janela a ser aplicada. Para este estudo todos os espectros sofreram ajuste de linha base e eliminação das primeiras 44 variáveis, representando unicamente ruído de difícil correção.

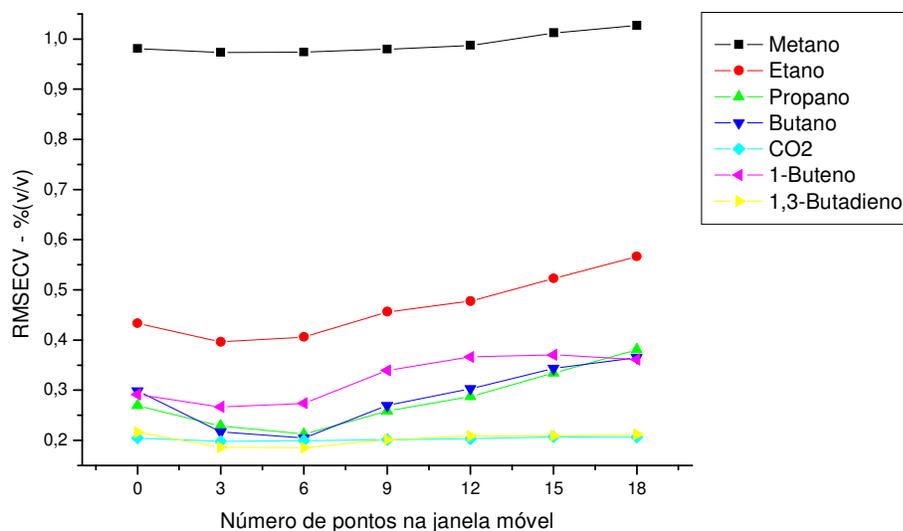


Figura 18 – Variação do valor de RMSECV com a aplicação progressiva do pré-tratamento alisamento por médio móvel.

(DADOS OBTIDOS COM ESPECTROFÔTOMETRO EXPERIMENTAL BASEADO EM AOTF - 40 amostras)

Obs.: Os conjuntos de calibração utilizados nesta análise correspondem aos conjuntos de dados dos modelos apresentados na Tabela 02, o número de variáveis latentes dos referidos modelos também se aplica neste caso. Para o cálculo dos modelos todos os dados foram previamente centrados na média.

Em seguida, uma seqüência de modelos foi construída para uma mesma espécie química, onde apenas o nível de alisamento dos espectros na matriz de calibração era modificado de forma crescente, neste caso foram aplicadas janelas de 3, 6, 9, 12, 15 e 18 pontos o que possibilitou observar uma melhora nos modelos que utilizavam o pré-tratamento com janela móvel de até 3 pontos. A Figura 18 indica que acima deste valor os modelos apresentam decréscimo na capacidade de previsão, exceto para 1-buteno e o dióxido de carbono os quais apresentaram uma ligeira tendência ao aumento da habilidade de previsão para o uso de janela móvel a partir de 15 pontos. Durante este processo constatou-se que para o metano houve o deslocamento do número ótimo de variáveis latentes de 5 para 6, entretanto a queda no valor de RMSECV associado a esta mudança era irrisória.

De acordo com os resultados da Figura 18, os modelos presentes na Tabela 06 foram calculados utilizando-se os espectros com ajuste de linha base; eliminação das primeiras 44 variáveis, representando unicamente ruído de difícil correção e suavização por média móvel com janela de 3 pontos.

Tabela 06 - tabela comparativa dos modelos PLS-1 para as 7 espécies químicas.

Espécie química	VL	RMSEC	RMSECV	Coefficiente de Correlação (real x previsto ^e)	Coefficiente angular	Coefficiente linear (% v/v)
METANO	5	0,7712	0,9732	0,9883	0,9832	1,2664
ETANO^a	7	0,2475	0,3959	0,9981	0,9970	0,0285
PROPANO	8	0,1363	0,2290	0,9914	0,9661	0,0997
BUTANO^b	8	0,1194	0,2169	0,9784	0,8756	0,2165
CO₂^c	7	0,1383	0,1979	0,9878	0,9817	0,0677
1-BUTENO	7	0,1589	0,2667	0,9825	0,9633	0,0897
1,3-BUTADIENO^d	8	0,1045	0,1858	0,9879	0,9775	0,0476

(DADOS OBTIDOS COM O EQUIPAMENTO EXPERIMENTAL BASEADO EM AOTF – 40 amostras)

^a: exclusão da 12^a amostra, ^b: exclusão da 2^a amostra, ^c: exclusão da 20^a amostra, ^d: exclusão da 1^a e 15^a amostra,

^e: valores previstos na validação cruzada (*leave one out*). **Obs.:** em todos os casos foi aplicado respectivamente: alisamento por média móvel, janela de 3 pontos, ajuste de linha base e eliminação das primeiras 44 variáveis (1524nm-1612nm) que representavam unicamente ruído de difícil correção. Para o cálculo dos modelos todos os dados foram previamente centrados na média, **VL:** variáveis latentes.

De forma similar ao caso anterior, o novo conjunto de amostras possibilitou a obtenção de modelos nos quais a informação química estava presente nas 7 ou 8 primeiras variáveis latentes, números condizentes com a quantidade de espécies presentes nas misturas gasosas. Os coeficientes angular e linear também estão próximos dos valores ideais de 1 e 0 respectivamente, indicando um bom ajuste.

A aplicação do teste $F(v_1: 80, v_2: 40, 95\%$ de confiança, $F: 1,545$)^d para comparação dos modelos de calibração apresentados nas Tabelas 05 e 06, revela que os resultados são semelhantes com 95% de confiança, com exceção para o metano e dióxido de carbono. O aumento do erro da validação cruzada em cerca de 2 vezes para o conjunto atual indica uma perda da capacidade de previsão para esta espécie. Esta perda poderia estar relacionada à menor resolução dos espectros, entretanto esta suposição ainda precisa ser comprovada. Os gráficos das Figuras 19 a 25 trazem os erros relativos individuais para o teor de cada um das espécies que compõem as amostras do segundo conjunto de calibração. Uma análise individual dos resultados revela que os erros de previsão mais elevados se encontram na faixa inferior de porcentagem volumétrica independente da espécie estudada, excetuando-se o metano. Estes resultados revelam que possivelmente a razão sinal ruído deve ser a responsável pelos elevados erros, indicando com grande probabilidade que os limites inferiores de “concentração” da maioria dos gases estão próximos do ruído instrumental, que por sua vez afeta a capacidade de previsão dos modelos. Todavia possíveis erros de preparação das amostras não podem ser descartados, vista as baixas porcentagens volumétricas de algumas espécies nas amostras.

^d: devido ao difícil acesso a tabelas mais completas do parâmetro F, utilizou-se o valor de F(crítico) para a razão mais próxima: $v_1:80/v_2:40$

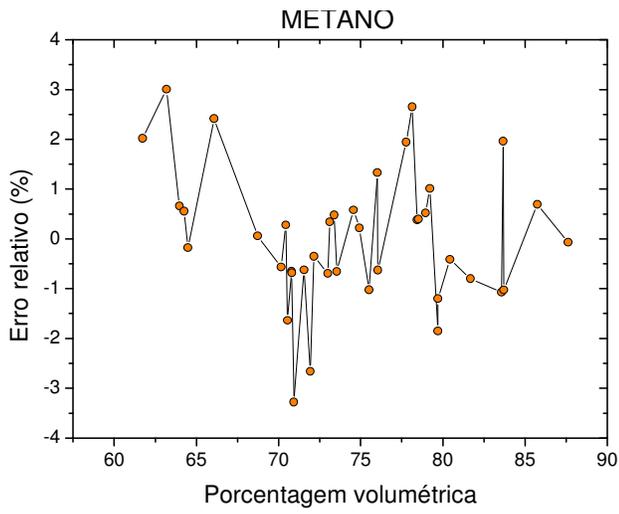


Figura 19 - Erro relativo na validação cruzada para o metano.

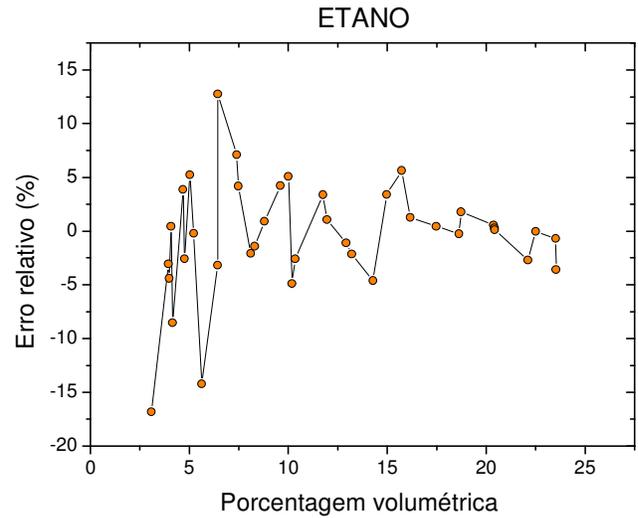


Figura 20 - Erro relativo na validação cruzada para o etano.

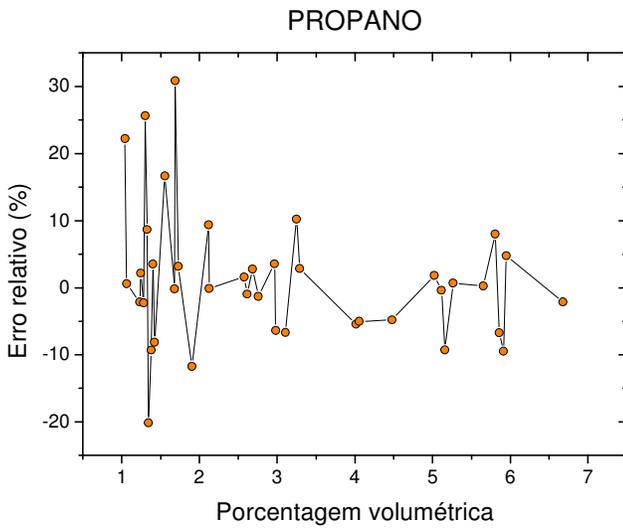


Figura 21 - Erro relativo na validação cruzada para o propano.

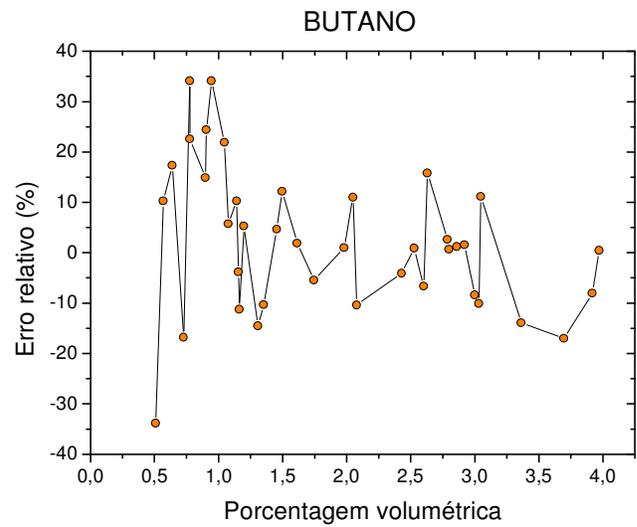


Figura 22 - Erro relativo na validação cruzada para o butano.

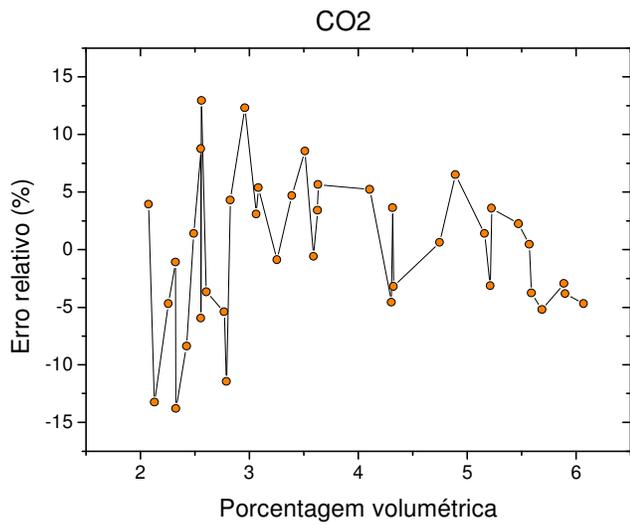


Figura 23 - Erro relativo na validação cruzada para o dióxido de carbono.

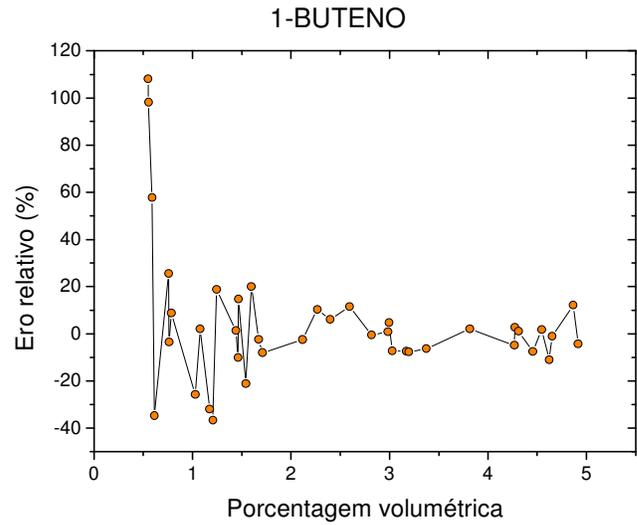


Figura 24 - Erro relativo na validação cruzada para o 1-buteno.

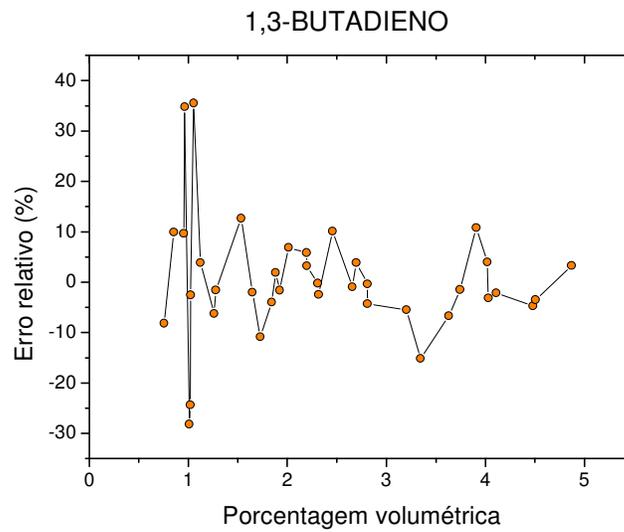


Figura 25 - Erro relativo na validação cruzada para o 1,3-butadieno.

Uma visão geral indica que os erros relativos estão localizados entre $\pm 3\%$ para o metano, além de estarem distribuídos de forma aleatória confirmando que não foi introduzido erro sistemático no preparo das amostras. O mesmo comportamento é exibido pelos demais gases.

Para o etano os dados revelam que apesar dos altos valores dos erros relativos, os valores absolutos são sempre inferiores a 1% (v/v), sendo este um valor bastante apropriado para a indústria, segundo normas que estabelecem os padrões de qualidade. O propano apresenta um erro absoluto máximo de 0,6% (v/v); no caso do butano o erro máximo é inferior a 0,75% (v/v); o dióxido de carbono possui 0,5% (v/v) como maior erro; a espécie 1-buteno apresenta uma amostra isolada com erro de aproximadamente 0,77% (v/v), entretanto todos as demais possuem erro abaixo de 0,5% (v/v) e para o 1,3-butadieno o maior erro da faixa é de 0,6% (v/v).

Espectros para o mesmo conjunto de 40 amostras também foram obtidos em um espectrofotômetro com transformada de Fourier da marca Bomem® (Figura 26). Este estudo visou comparar os dados fornecidos pelo equipamento comercial com os obtidos com o espectrofotômetro experimental, em vista da maior resolução oferecida pelo primeiro 4 cm^{-1} , equivalente a 0,9 nm (para o início da faixa - 1500 nm) e 2,8 nm (ao final da faixa - 2650 nm).

Para o equipamento experimental a resolução aumenta de 4nm (no início da faixa – 1500 nm) para 14 nm (ao final da faixa – 2650 nm). Adicionalmente, o espectrofotômetro FT apresenta níveis de ruído bem inferiores ao equipamento experimental baseado em AOTF; cerca de 0,4 mA e 20 mA respectivamente, ou seja 50 vezes menor.

Em contrapartida, o espectrofotômetro Bomem® só pode operar com célula cujo caminho óptico máximo é de 13 cm, enquanto o equipamento experimental utiliza uma célula para gases de 54 cm de caminho óptico, cerca de 4,2 vezes maior.

A Figura 26 apresenta os espectros obtidos com o espectrofotômetro NIR-FT. Observa-se neste caso uma riqueza de detalhes tanto na região do primeiro sobreton (1600 nm a 1800 nm) quanto na região das bandas de combinação (2000 nm a 2500 nm). Vêem-se claramente pontos de absorção bem definidos, apresentando picos bem agudos relacionados às absorções rotacionais as quais estão bastante nítidas diferentemente dos espectros obtidos com o espectrofotômetro NIR-AOTF, onde as mesmas estão consideravelmente envelopadas.

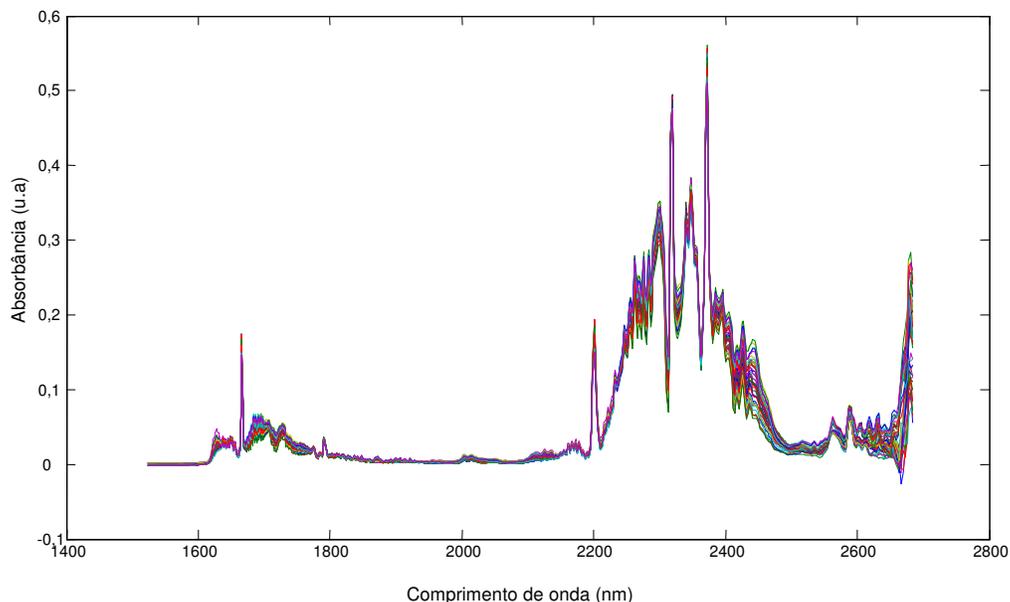


Figura 26 – Espectros para o conjunto de calibração com 40 amostras, obtidos com espectrofotômetro comercial Bomem®.

O conjunto de espectros obtidos por meio do espectrofotômetro Bomem® também foi alvo de um estudo com a aplicação do pré-tratamento alisamento por janela móvel. Neste caso, um novo modelo de calibração PLS-1 era calculado para cada etapa de suavização, iniciando-se com uma janela de 3 pontos e prosseguindo-se com incrementos também de 3 pontos sobre o valor inicial. O

processo foi repetido para todas espécies acompanhando-se sempre a evolução no valor do parâmetro RMSECV.

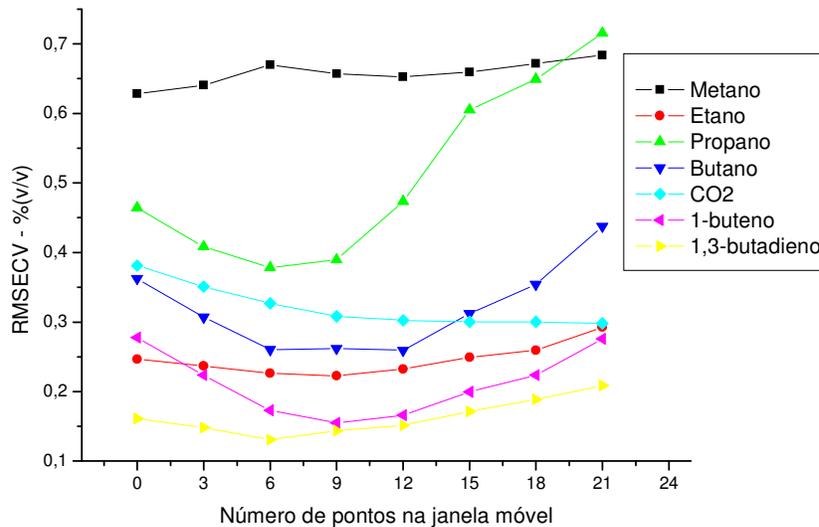


Figura 27 – variação do valor de RMSECV com a aplicação progressiva do pré-tratamento alisamento por médio móvel

(DADOS OBTIDOS COM ESPECTROFÔTOMETRO COMERCIAL BOMEM® - 40 amostras)

Obs.: Os conjuntos de calibração utilizados nesta análise correspondem aos conjuntos de dados dos modelos apresentados na Tabela 07, o número de variáveis latentes dos referidos modelos também se aplica neste caso. Para o cálculo dos modelos todos os dados foram previamente centrados na média.

A Figura 27 apresenta a progressão no valor do erro médio da validação cruzada com o aumento da janela utilizada no processo de suavização dos espectros. É possível observar, com exceção do metano, que todas as demais espécies apresentam uma queda inicial no valor do RMSECV com um posterior aumento quando o comprimento da janela móvel se encontra entre 9 e 12 pontos. É um consenso que espectros com ruído moderado a alto apresentam um melhora quando os mesmos são submetidos a pré-tratamentos deste gênero visando apenas a correção do erro aleatório. Entretanto o conjunto de espectros em questão possui um nível de ruído bastante aceitável (0,4 mA) para a magnitude das absorções registradas o que descarta a atribuição da melhora dos modelos de calibração simplesmente à correção do ruído espectral. Este fato põe em

evidência a alta resolução do espectrofotômetro, quando comparado ao equipamento experimental, indicando-a como responsável pelos valores iniciais de RMSECV superiores, ou seja calibrações menos eficientes.

Tabela 07 - tabela comparativa dos modelos PLS-1 para as 7 espécies químicas

Espécie química	VL	RMSEC	RMSECV	Coefficiente de Correlação (medido x previsto ^d)	Coefficiente angular	Coefficiente linear (% v/v)
METANO	8	0,3598	0,6285	0,9951	0,9894	0,7627
ETANO^a	8	0,1511	0,2226	0,9994	1,0004	-0,0114
PROPANO	8	0,2492	0,3895	0,9744	0,9461	0,1814
BUTANO	9	0,1633	0,2621	0,9685	0,9469	0,0791
CO₂^b	8	0,1909	0,3083	0,9722	0,9552	0,1722
1-BUTENO	9	0,0952	0,1552	0,9941	0,9802	0,0541
1,3-BUTADIENO^c	8	0,0865	0,1438	0,9927	0,9878	0,0187

(DADOS OBTIDOS COM ESPECTROFÔTOMETRO COMERCIAL BOMEM® - 40 amostras)

^a: exclusão da 12ª amostra, ^b: exclusão da 29ª amostra, ^c: exclusão da 19ª amostra, ^d: valores previstos na validação cruzada (*leave one out*). **Obs.:** em todos os casos foi utilizada a região (6561cm^{-1} - 3726cm^{-1}) equivalente à faixa (1524nm – 2683nm) analisada com o equipamento experimental baseado em AOTF. Em todos os casos foi aplicado alisamento por janela móvel de 9 pontos, exceto para os espectros do gás metano onde não foi realizada suavização, e para o cálculo dos modelos todos os dados foram previamente centrados na média, **VL**: variáveis latentes.

Para a construção dos modelos referentes aos dados provenientes do espectrofotômetro NIR-FT, foi realizada apenas a suavização dos espectros por média móvel de acordo com os resultados da Figura 27. Excetuando-se o metano cujos espectros não foram suavizados, em todos os demais casos foi aplicada uma janela de 9 pontos. Para este conjunto de dados, os menores valores do erro de previsão para a validação cruzada indicam, em geral, 8 variáveis latentes como o número ótimo para a calibração. Considerando o número de espécies presentes e a contribuição da variância espectral, a quantidade é coerente com o sistema químico em questão.

Analisando-se por meio de um teste $F(v_1: 40, v_2: 40, 95\%$ de confiança, $F: 1,693$) os modelos atuais em conjunto com aqueles apresentados na Tabela 06, verifica-se que apenas as calibrações para os gases butano e 1,3-butadieno são

semelhantes com 95% de confiança. Para as espécies metano, etano, propano, dióxido de carbono e 1-buteno as calibrações são estatisticamente distintas para 95% de confiança. Sendo que no caso do propano e dióxido de carbono os modelos obtidos a partir dos espectros registrados com o equipamento baseado em AOTF apresentam um melhor (menor) erro de previsão.

Uma investigação um pouco mais minuciosa indica que a aplicação do alisamento por janela móvel distorce gradativamente os espectros de forma a assemelharem-se aos espectros brutos obtidos com o equipamento experimental (Figura 28).

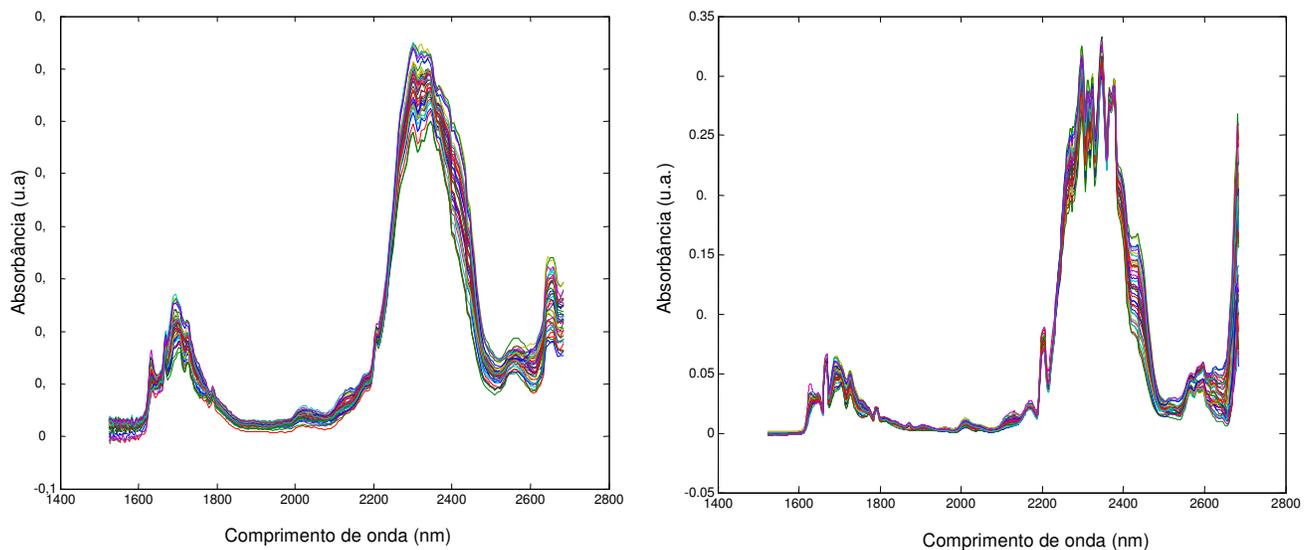


Figura 28 – à esquerda tem-se os espectros obtidos com o espectrofotômetro experimental baseado em AOTF (Figura 28A); à direita tem-se os espectros obtidos com equipamento comercial da marca Bomem®, após suavização por janela móvel de 9 pontos (Figura 28B).

Estes resultados também indicam a possibilidade de obtenção de modelos mais eficientes que os atuais quando faz-se uso deste pré-tratamento em espectros obtidos nas condições de resolução iniciais do equipamento experimental.

5.1 - VALIDAÇÃO EXTERNA

Até o momento o procedimento analítico foi avaliado exclusivamente por modelos construídos com base apenas em um conjunto de treinamento. Isto se deve ao fato de que estatisticamente o melhor modelo está intrinsecamente ligado ao maior espaço amostral, desta forma as potencialidades do sistema podem ser melhor avaliadas a partir deste panorama. Para a avaliação em termos práticos foi realizada uma validação externa. Para este fim o grupo maior de 40 amostras foi dividido em um conjunto de calibração com 30 misturas e um subconjunto de 10 amostras para validação externa.

VALIDAÇÃO EXTERNA – CONJUNTO AOTF
(30 amostras de calibração, 10 de previsão)

Tabela 08 – resultado da validação externa para o metano

	MODELO PLS-1 (5PCs)	
METANO	RMSECV: 1,0036	
	RMSEP: 1.0293	
Porcentagem volumétrica		
	Valor real	Valor previsto
1°	73,1	73,7
2°	78,4	78,6
3°	71,9	69,9
4°	83,7	82,5
5°	63,2	64,9
6°	74,6	74,6
7°	83,6	82,6
8°	70,8	70,3
9°	73,0	72,2
10°	68,7	68,8

Tabela 09 – resultado da validação externa para o etano

	MODELO PLS-1 (8PCs) ^a	
ETANO	RMSECV: 0,4060	
	RMSEP: 0,3680	
Porcentagem volumétrica		
	Valor real	Valor previsto
1°	8,8	8,9
2°	5,2	5,2
3°	14,3	13,7
4°	4,8	4,7
5°	23,5	22,8
6°	8,3	8,1
7°	10,0	10,3
8°	18,7	19,0
9°	11,8	11,9
10°	22,1	21,6

^a: exclusão da 11ª amostra.

Tabela 10 – resultado da validação externa para o propano.

	MODELO PLS-1 (8PCs)	
PROPANO	RMSECV: 0,2348	
	RMSEP: 0,3133	
Porcentagem volumétrica		
	Valor real	Valor previsto
1°	5,9	5,3
2°	3,1	2,7
3°	2,8	2,7
4°	1,6	1,6
5°	1,7	2,3
6°	1,7	1,8
7°	1,2	1,4
8°	2,1	2,4
9°	4,0	4,0
10°	2,6	2,6

Tabela 11 – resultado da validação externa para o butano.

	MODELO PLS-1 (9PCs) ^b	
BUTANO	RMSECV: 0,2637	
	RMSEP: 0,2098	
Porcentagem volumétrica		
	Valor real	Valor previsto
1°	2,9	2,8
2°	0,9	0,9
3°	2,1	2,1
4°	0,7	0,6
5°	3,7	3,2
6°	2,6	2,4
7°	0,8	1,0
8°	1,3	1,1
9°	1,4	1,2
10°	0,8	0,7

^b: exclusão da 11ª amostra.

VALIDAÇÃO EXTERNA – CONJUNTO AOTF
(30 amostras calibração, 10 previsão)

Tabela 12 – resultado da validação externa para o dióxido de carbono.

		MODELO PLS-1 (6PCs) ^c
CO₂		RMSECV: 0,2384
		RMSEP: 0,2304
Porcentagem volumétrica		
	Valor real	Valor previsto
1°	3,4	3,1
2°	5,6	5,4
3°	3,6	4,3
4°	2,8	3,2
5°	2,4	2,1
6°	5,2	5,2
7°	2,3	2,2
8°	3,6	3,7
9°	4,3	4,7
10°	3,6	3,8

^c: exclusão da 6ª amostra.

Tabela 13 – resultado da validação externa para o 1-buteno.

		MODELO PLS-1 (9PCs) ^d
1-BUTENO		RMSECV: 0,2328
		RMSEP: 0,2223
Porcentagem volumétrica		
	Valor real	Valor previsto
1°	4,7	4,6
2°	4,9	5,1
3°	0,8	1,1
4°	2,4	2,8
5°	2,3	2,3
6°	3,2	3,0
7°	1,0	0,7
8°	1,4	1,6
9°	1,5	1,7
10°	1,6	1,9

^d: exclusão da 10ª amostra.

Tabela 14 – resultado da validação externa para o 1,3-butadieno.

		MODELO PLS-1 (8PCs) ^e
1,3-BUTADIENO		RMSECV: 0,2318
		RMSEP: 0,3990
Porcentagem volumétrica		
	Valor real	Valor previsto
1°	1,3	1,2
2°	1,8	1,8
3°	4,6	5,8
4°	4,0	4,2
5°	3,2	3,0
6°	4,5	4,2
7°	1,1	1,3
8°	2,0	2,1
9°	4,1	4,0
10°	0,8	0,7

^e: exclusão da 1ª amostra.

Obs.: em todos os conjuntos foi aplicada a correção de linha de base, alisamento por janela móvel de 3 pontos e eliminação das primeiras 44 variáveis, representando unicamente ruído de difícil correção. Para o cálculo dos modelos todos os dados foram previamente centrados na média.

Os resultados presentes nas tabelas de 8 a 14 tendem a acompanhar o desempenho dos conjuntos únicos de 40 amostras, não havendo portanto divergências expressivas.

Excetuando-se a amostra n° 3 do conjunto de validação externa para o 1,3-butadieno, a qual consiste na amostra n° 15 do grupo de 40 que por sua vez trata-se de uma amostra anômala, em todos os demais casos os valores de RMSEP são bastante favoráveis. Considerando-se o valor médio das faixas de porcentagem volumétrica das espécies, o RMSEP representa um erro de 1,4% para o metano; 2,8% para o etano; 8,4% para o propano; 0,9% para o butano; 5,4% para o CO₂; 8,1% para o 1-butadieno e 13,3% para o 1,3-butadieno.

Assim como no caso anterior, é possível observar que as curvas de calibração fornecem resultados cujos erros de previsão (para a validação cruzada) estão de acordo com os valores estabelecidos pelas normas oficiais, onde emprega-se a cromatografia gasosa como método padrão.

5.2 - FIGURAS DE MÉRITO

As figuras de mérito têm a função de garantir que um novo método de análise seja confiável, assegurando a validade dos resultados e tornando possível a comparação com outros métodos já em uso. Um método não validado, não é confiável, podendo levar a resultados e interpretações equivocadas.

Os conceitos dos parâmetros de desempenho surgiram com a calibração de ordem zero e como reflexo, os cálculos das figuras de mérito são em grande parte específicos para os métodos de calibração univariados. No caso das calibrações de ordem superior, as abordagens matemáticas precisam ser adaptadas considerando-se a natureza multivariada dos métodos.

Devido à complexidade da calibração multivariada, muitas figuras de mérito ainda não estão estabelecidas, não havendo, portanto, um consenso entre os autores das propostas presentes na literatura. Pode-se dizer que o cálculo destes

parâmetros ainda é um desafio, cujo objetivo é desvendar as abordagens matemáticas que melhor traduzem as definições anteriormente formuladas.

Neste trabalho, a validação do procedimento analítico foi realizada calculando-se as figuras de mérito multivariadas de acordo com as propostas mais difundidas na literatura. Devido à presença de espécies presentes em baixas porcentagens, os limites de detecção e quantificação tiveram uma maior ênfase e foram estimados por mais de um método.

Tabela 15 – Figuras de mérito para os modelos PLS-1 apresentados nas Tabelas 8 a 14.

Figura de mérito		METANO	ETANO	PROPANO	BUTANO	CO ₂	1-BUTENO	1,3-BUTADIENO
Limite de detecção ¹ LOD % (v/v)	Ortiz ^a	5,9	1,4	0,5	0,5	0,7	0,8	0,8
	Lorber ^b	0,2 ₉₆₄	0,7 ₇₁₇	0,9 ₃₅₄	0,8 ₈₉₉	0,1 ₈₄₇	0,7 ₅₈₅	0,4 ₀₁₀
Limite de quantificação LOQ % (v/v)	Ortiz ^a	19,8	4,7	1,6	1,7	2,4	2,6	2,9
	Lorber ^b	0,9 ₈₈₀	2,5 ₇₂₃	3,1 ₁₈₀	2,9 ₆₆₃	0,6 ₁₅₇	2,5 ₂₈₃	1,3 ₃₆₇
Sensibilidade		0,0088	0,0096	0,0075	0,0079	0,0501	0,0094	0,0186
γ^{-1} % (v/v)		0,3 ₁	0,2 ₈	0,3 ₆	0,3 ₄	0,0 ₅	0,2 ₉	0,1 ₅
Seletividade ξ	Máxima	0,2458	0,0544	0,0144	0,0088	0,0901	0,0120	0,0267
	Média	0,1807	0,0290	0,0076	0,0051	0,0543	0,0067	0,0122
	Mínima	0,1322	0,0081	0,0037	0,0021	0,0236	0,0025	0,0049
Relação sinal ruído S/N	Máxima	282,4	82,3	18,6	12,7	120,9	17,4	34,4
	Média	242,8	40,5	10,4	6,9	73,3	9,2	16,5
	Mínima	202,8	10,4	4,8	3,2	33,2	3,4	6,4
Precisão ^c % (v/v)		0,2 ₇	0,0 ₉	0,1 ₃	0,0 ₉	0,2 ₀	0,0 ₈	0,1 ₁
Precisão instrumental ^d % (v/v)		0,2 ₄	0,5 ₉	0,2 ₄	0,1 ₁	0,1 ₈	0,1 ₈	0,6 ₉
Exatidão % (v/v)		1,0	0,4	0,3	0,0 ₂	0,2	0,2	0,4

(DADOS OBTIDOS COM ESPECTROFÔTOMETRO EXPERIMENTAL BASEADO EM AOTF)

^{a, b} : referem-se às propostas apresentadas pelos pesquisadores cujos nomes são mencionados, no caso "c" o ruído foi estimado com base no desvio-padrão da norma do vetor NAS para 40 espectros do ensaio em branco, ^{1(a)} : LOD calculado para $\alpha = \beta = 0,05$ (95,0%) e 38 graus de liberdade. ^c: valor estimado com base em 3 réplicas de uma amostra cuja porcentagem volumétrica era de 70,2%; 12,0%; 3,3%; 2,6%; 5,2%; 3,0%; 3,7%; para as espécies químicas apresentadas na tabela acima, iniciando-se no metano. ^d: valor estimado com base em 4 réplicas de uma amostra cuja porcentagem volumétrica era de 71,9%; 14,3%; 2,8%; 2,1%; 3,6%; 0,8%; 4,6%; para as espécies químicas apresentadas na tabela acima, iniciando-se no metano.

Conceitualmente o limite de quantificação deve expressar o ponto a partir do qual medidas quantitativas possam ser realizadas, ou seja, o erro associado ao resultado apresenta-se em um nível aceitável. Analisando-se os valores dos limites de quantificação da Tabela 15 com o auxílio dos gráficos das Figuras 19 a 25, observa-se a coerência dos resultados para as espécies 1,3-butadieno; 1-buteno, butano e propano com as seguintes ressalvas: para o butano o método prevê uma condição favorável de quantificação a partir de 2,9%(v/v). Mas observando-se o gráfico do erro relativo versus porcentagem volumétrica para este gás, percebe-se que erro relativo sugere uma condição favorável para a quantificação a partir de 1,5%(v/v), uma vez que os valores estão contidos entre $\pm 10\%$ (erro). O mesmo ocorre para o propano cujo valor calculado; 3,1%; está 50% acima do ponto sugerido pela avaliação do erro relativo, considerando-se a estabilização do erro em $\pm 10\%$ (erro), a partir de 2%(v/v).

Com relação aos limites de detecção, percebe-se pela Tabela 15 que ambos métodos fornecem estimativas coerentes. Todavia, para a abordagem proposta por Lorber et al.^[2] o valor obtido para o dióxido de carbono está claramente muito abaixo de um limite de detecção possível de ser encontrado utilizando o método analítico proposto, mesmo que esta afirmação seja apenas uma conjectura necessitando, portanto, de novas comprovações.

Exclusivamente para o etano, propano e butano os limites de detecção univariados calculados^[89] (etano: 1,4%(v/v); propano 0,9%(v/v); butano: 0,6%(v/v)) mostram que a aplicação das definições multivariadas em foco retorna valores comparáveis em ordem de grandeza. O trabalho realizado por Franco^[89] também traz os valores de LQ para os 3 hidrocarbonetos citados (etano: 4,5%(v/v); propano 3,1%(v/v); butano: 2,1%(v/v)), sendo que no caso do gás propano o método multivariado proposto por Lorber prevê o mesmo resultado obtido pela abordagem univariada.

É importante mencionar, no entanto, que os valores univariados são utilizados apenas como um breve guia, auxiliando em simples observações como a delimitação das faixas onde possivelmente se encontrariam os limites de detecção multivariados. Os valores de LD e LQ para o metano são apresentados

apenas em caráter informativo. Os mesmos não podem ser considerados estimativas válidas, uma vez que os limites de detecção e quantificação também são funções das concentrações dos interferentes na matriz química, que no caso do metano pode sofrer mudanças significativas uma vez que o mesmo pode ser considerado como uma espécie de solvente da mistura.

CONCLUSÃO

Dentro da atual proposta do trabalho, acredita-se que as metas tenham sido cumpridas. Os estudos com o conjunto de calibração envolvendo o vapor de água demonstram ser possível a calibração de todas as espécies, apesar da falta de repetibilidade para os valores de umidade. A análise do segundo conjunto de calibração confirma que todos os hidrocarbonetos gasosos e o dióxido de carbono nas amostras sintéticas de gás natural puderam ser previstos com exatidão dentro das normas oficiais. Esta análise revelou ainda que os espectros obtidos com o espectrofotômetro comercial de alta resolução geram calibrações menos eficientes, indicando a necessidade de uma menor resolução espectral, destacando também o fato de que um caminho óptico maior deve ser um requisito importante para melhorar o desempenho dos modelos de calibração. O estudo também mostrou a possibilidade de obtenção de modelos de calibração multivariada baseados em uma quantidade bastante acessível de amostras, diminuindo consideravelmente o tempo de preparação do conjunto de misturas e a aquisição dos dados, bem como uma redução significativa no volume dos gases utilizados.

Provendo bons resultados, a escolha da metodologia de calibração PLS também se mostrou apropriada e por fim a avaliação das figuras de mérito fornece uma visão do desempenho do processo analítico e o torna passível de comparação com os demais métodos empregados para a determinação de espécies majoritárias e minoritárias em misturas de gases naturais.

APÊNDICE

OPERADOR DE PROJEÇÃO

Analisando-se algebricamente o processo de regressão por mínimos quadrados é possível identificar um dos operadores mais utilizados em quimiometria.

Admitindo-se uma relação linear entre duas ou mais variáveis independentes (x_1, x_2, \dots) e uma variável dependente y , é possível escrever uma equação de reta a qual representa a dependência entre as várias incógnitas (considere o intercepto igual a zero). Para uma série de dados experimentais afetados unicamente pelo erro aleatório inerente a toda medida, tem-se um conjunto de pares ordenados (x, y) . Para obter-se os valores dos parâmetros do modelo, admitindo anteriormente, pode-se aplicar por exemplo uma regressão por mínimos quadrados.

Na forma matricial, a equação de reta é escrita como a seguir:

$$\mathbf{y} = \mathbf{X} \cdot \mathbf{b} \quad (\text{A 1.0})$$

Onde, \mathbf{y} representa o vetor-coluna contendo os valores y_i , \mathbf{X} uma matriz cujas linhas representam as amostras e as colunas as variáveis em estudo e \mathbf{b} é o vetor de regressão o qual contém a taxa de variação de cada incógnita x_i com y , ou seja, os coeficientes angulares individuais.

Algebricamente, o vetor \mathbf{y} consiste em uma combinação linear dos vetores-coluna da matriz \mathbf{X} , ponderados pelos elementos do vetor de regressão. Ou seja, para que haja solução o vetor de respostas \mathbf{y} deve estar no espaço gerado pelos vetores-coluna da matriz \mathbf{X} .

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_i \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ x_{31} & x_{32} \\ \vdots & \vdots \\ x_{i1} & x_{i2} \end{pmatrix} \cdot \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \Leftrightarrow \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_i \end{pmatrix} = \begin{pmatrix} x_{11} \\ x_{21} \\ x_{31} \\ \vdots \\ x_{i1} \end{pmatrix} \cdot b_1 + \begin{pmatrix} x_{12} \\ x_{22} \\ x_{32} \\ \vdots \\ x_{i2} \end{pmatrix} \cdot b_2 \quad (\text{A 1.1})$$

Devido à presença dos erros experimentais o vetor \mathbf{y} não pode ser descrito exatamente como uma combinação linear dos vetores-coluna da matriz \mathbf{X} , o que equivale a dizer que o vetor \mathbf{y} não se encontra exatamente no espaço gerado pelas colunas de \mathbf{X} .

A abordagem dos mínimos quadrados consiste em encontrar o vetor de regressão que ajusta os valores experimentais da variável y com as variáveis independentes. Partindo-se da equação A 1.0, tem-se:

$$\hat{\mathbf{b}} = (\mathbf{X}^T \cdot \mathbf{X})^{-1} \cdot \mathbf{X}^T \cdot \mathbf{y} \quad (\text{A 1.2})$$

Do ponto de vista geométrico, encontrar a solução para um problema de mínimos quadrados equivale a projetar o vetor \mathbf{y} no espaço gerado pelas colunas de \mathbf{X} de forma que a projeção e o vetor original sejam o mais similar possível. A projeção de \mathbf{y} (\mathbf{y}_{proj}) é obtida a partir do vetor de regressão calculado:

$$\mathbf{y}_{\text{proj}} = \mathbf{X} \cdot \hat{\mathbf{b}} \quad \leftrightarrow \quad \mathbf{y}_{\text{proj}} = \mathbf{X} \cdot (\mathbf{X}^T \cdot \mathbf{X})^{-1} \cdot \mathbf{X}^T \cdot \mathbf{y} \quad \leftrightarrow \quad \mathbf{y}_{\text{proj}} = \mathbf{X} \cdot \mathbf{X}^+ \cdot \mathbf{y}$$

Onde \mathbf{X}^+ é igual à pseudoinversa da matriz \mathbf{X} e a matriz $\mathbf{X} \cdot (\mathbf{X}^T \cdot \mathbf{X})^{-1} \cdot \mathbf{X}^T$ representa o operador responsável pela projeção do vetor \mathbf{y} no espaço-coluna da matriz \mathbf{X} .

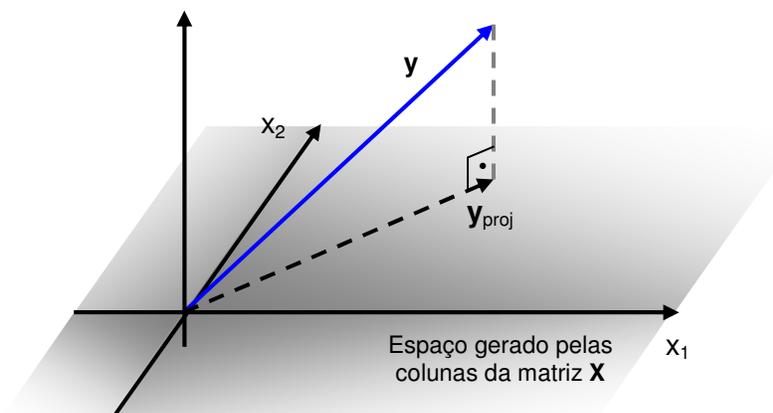


Figura 29 – Projeção do vetor de \mathbf{y} no espaço-coluna (plano) da matriz \mathbf{X} .

Analogamente, o conceito do sinal analítico líquido utiliza um operador semelhante para projetar um vetor-amostra no espaço gerado pelas linhas da matriz dos interferentes (teoricamente, uma matriz com amostras compostas unicamente com as espécies interferentes). A projeção do vetor-amostra neste espaço resulta em um vetor que pode ser descrito por uma combinação linear exata dos espectros dos interferentes. Desta forma, \mathbf{y}_{proj} é a parcela do vetor \mathbf{y} que está unicamente relacionada com os interferentes, sendo assim, ao subtrair-se

esta componente do vetor original obtém-se o vetor que representa o sinal puro do analito ou o **sinal analítico líquido**.

MASSA REDUZIDA

Ao aplicar-se a os conceitos da cinemática para análise de um oscilador harmônico, composto de duas massas acopladas por uma mola, o desenvolvimento matemático reestrutura o problema possibilitando reinterpretá-lo como o sistema massa-mola fundamental. Desta forma, um sistema molecular diatômico pode ser simplificado em um conjunto massa-mola convencional, onde analisamos o movimento de uma única massa acoplada a uma mola cuja outra extremidade está fixa em um objeto que se convencionou chamar de massa infinita, como uma parede por exemplo.

Considera-se um objeto de massa m_1 , cuja posição é descrita pelo vetor r_1 e um segundo objeto de massa m_2 , conectado ao primeiro por meio de uma mola, o qual pode ser localizado pelo vetor r_2 . Supõe-se então que o primeiro objeto exerce uma força f_{21} sobre o segundo objeto, logo de acordo com a 3ª lei de Newton, o segundo objeto exercerá uma força de igual magnitude mas em sentido oposto $f_{12} = - f_{21}$. Matematicamente:

$$m_2 \cdot \frac{d^2 r_2}{dt^2} = f \quad (\text{A 2.0})$$

$$m_1 \cdot \frac{d^2 r_1}{dt^2} = - f \quad (\text{A 2.1})$$

Onde $f_{21} = f$

O centro de massa do sistema está é representado por:

$$r_{cm} = \frac{m_1 \cdot r_1 + m_2 \cdot r_2}{m_1 + m_2} \quad (\text{A 2.3})$$

Reorganizando a equação e explicitando r_1 e r_2 , tem-se:

$$r_1 = r_{cm} - \frac{m_2}{m_1 + m_2} \cdot r \quad (\text{A 2.4})$$

$$r_2 = r_{cm} + \frac{m_1}{m_1 + m_2} \cdot r \quad (\text{A 2.5})$$

Onde $r = r_2 - r_1$

Substituindo as equações 4 e 5 nas equações 1 e 2 e lembrando-se que o centro de massa de um sistema isolado permanece estático e portanto não sofre aceleração, concluí-se:

$$\frac{m_1 \cdot m_2}{m_1 + m_2} \cdot \frac{d^2 r}{dt^2} = f \quad (\text{A 2.6})$$

$$\phi \cdot \frac{d^2 r}{dt^2} = f \quad (\text{A 2.7})$$

Onde ϕ = massa reduzida do sistema.

Neste caso, as posições individuais dos objetos, bem como as massas são substituídas por uma única distância e pela massa reduzida, assim como no sistema massa-mola fundamental

LEI DE BOUGUERT/LAMBERT/BEER

A primeira parte da lei de Bouguer enuncia que a energia radiante monocromática transmitida por um meio homogêneo e isotrópico^e é proporcional à energia radiante incidente, ou seja, a razão da potência de um feixe de radiação após atravessar um determinado material com relação à potência inicial é uma constante, denominada transmitância.

$$\frac{P}{P_0} = T \quad (\text{A 3.0})$$

Onde P = Potência do feixe transmitido,

P_0 = Potência do feixe incidente,

T = Transmitância.

A segunda parte da lei enuncia que a energia da radiação transmitida decresce em progressão geométrica quando o caminho óptico aumenta em progressão aritmética. Em outras palavras, se dividíssemos um cubo em várias fatias de igual caminho óptico, todas as fatias do material absorveriam individualmente uma mesma porcentagem da energia, matematicamente esse conceito é expresso pela equação A 3.1. A lei de Beer expressa a mesma relação, porém entre a transmitância e a concentração do material absorvente. De forma equivalente, para um caminho óptico fixo a transmitância diminui em progressão geométrica quando a concentração aumenta em progressão aritmética, equação A 3.2.

$$T = \left[\left(\frac{P}{P_0} \right)^* \right]^b \quad \text{Lei de Bouguert (A 3.1)} \quad T = \left[\left(\frac{P}{P_0} \right)^* \right]^c \quad \text{Lei de Beer (A 3.2)}$$

Onde,

T = transmitância.

^e: Cujas propriedades eletromagnéticas, como o índice de refração, são as mesmas em todas direções.

$(P/P_0)^*$ = Transmitância por unidade de concentração e caminho óptico.

b = caminho óptico e c = concentração.

Uma correlação linear entre a concentração e o caminho óptico é obtida definindo um novo parâmetro, a absorvância - A. Aplicando o logaritmo em ambas equações e realizando a combinação das duas leis, obtém-se a lei de Bouguer/Lambert-Beer.

$$- \text{Log } T = \epsilon . b . c \quad (\text{A 3.3})$$

$$A = - \text{Log } T \quad (\text{A 3.4})$$

$$\boxed{A = \epsilon . b . c} \quad (\text{A 3.5})$$

Lei de Bouguer/Lambert-Beer

Onde, $\epsilon = \text{Log } (P/P_0)^*$ = absorvância molar



REFERÊNCIAS BIBLIOGRÁFICAS

- [1] S. Ramos, *O Desenvolvimento da Quimiometria, Revista de Química Industrial*, **2006**, p. 11.
- [2] K. S. Booksh and B. R. Kowalski, *Theory of Analytical Chemistry, Analytical Chemistry* **1994**, 66, 782A-791A.
- [3] A. Lorber, K. Faber and B. R. Kowalski, *Net analyte signal calculation in multivariate calibration, Analytical Chemistry* **1997**, 69, 1620-1626.
- [4] A. L. B. R. K. Klaas Faber, *Analytical figures of merit for tensorial calibration, Journal of Chemometrics* **1997**, 11, 419-461.
- [5] N. M. Faber, *Multivariate sensitivity for the interpretation of the effect of spectral pretreatment methods on near-infrared calibration model predictions, Analytical Chemistry* **1999**, 71, 557-565.
- [6] H. C. Goicoechea and A. C. Olivieri, *Enhanced Synchronous Spectrofluorometric Determination of Tetracycline in Blood Serum by Chemometric Analysis. Comparison of Partial Least-Squares and Hybrid Linear Analysis Calibrations, Analytical Chemistry* **1999**, 71, 4361-4368.
- [7] H. C. Goicoechea and A. C. Olivieri, *MULTIVAR. A program for multivariate calibration incorporating net analyte signal calculations, Trac-Trends in Analytical Chemistry* **2000**, 19, 599-605.
- [8] A. Espinosa-Mansilla, M. I. A. Valenzuela, A. M. de la Pena, F. Salinas and F. C. Canada, *Comparative study of partial least squares and a modification of hybrid linear analysis calibration in the simultaneous spectrophotometric determination of rifampicin, pyrazinamide and isoniazid, Analytica Chimica Acta* **2001**, 427, 129-136.
- [9] A. Muñoz de la Peña, A. Espinosa-Mansilla, M. I. Acedo Valenzuela, H. C. Goicoechea and A. C. Olivieri, *Comparative study of net analyte signal-based methods and partial least squares for the simultaneous determination of amoxicillin and clavulanic acid by stopped-flow kinetic analysis, Analytica Chimica Acta* **2002**, 463, 75-88.
- [10] M. S. Collado, V. E. Mantovani, H. C. Goicoechea and A. C. Olivieri, *Simultaneous spectrophotometric-multivariate calibration determination of several components of ophthalmic solutions: phenylephrine, chloramphenicol, antipyrine, methylparaben and thimerosal, Talanta* **2000**, 52, 909-920.
- [11] G. Bergmann, B. Von Oepen and P. Zinn, *Improvement in the definitions of sensitivity and selectivity, Analytical Chemistry* **1987**, 59, 2522-2526.
- [12] A. C. Olivieri, N. K. M. Faber, J. Ferre, R. Boque, J. H. Kalivas and H. Mark, *Uncertainty estimation and figures of merit for multivariate calibration, Pure and Applied Chemistry* **2006**, 78, 633-661.
- [13] R. Liu, L. N. Lu, W. L. Chen and K. X. Xu, *Study of net analyte signal with near-infrared spectra for quantitative analysis, Spectroscopy and Spectral Analysis* **2004**, 24, 1042-1046.
- [14] G. Bauer, W. Wegscheider and H. M. Ortner, *Selectivity and Limits of Detection in Inductively Coupled Plasma Optical-Emission Spectrometry Using Multivariate Calibration, Spectrochimica Acta Part B-Atomic Spectroscopy* **1992**, 47, 179-188.
- [15] R. Boqué and F. X. Rius, *Multivariate detection limits estimators, Chemometrics and Intelligent Laboratory Systems* **1996**, 32, 11-23.
- [16] R. Boque, N. K. M. Faber and F. X. Rius, *Detection limits in classical multivariate calibration models, Analytica Chimica Acta* **2000**, 423, 41-49.

- [17] Thermo Scientific/Ahura Technologies, <http://www.ahurascientific.com/chemical-explosives-id/applications/index.php>, acessado em setembro de 2010.
- [18] A2 Technologies, http://www.a2technologies.com/handheld_exoscan.html, acessado em setembro de 2010.
- [19] BAYSPEC, http://www.bayspec.com/product_detail.php?p_id=57&a_id=4, acessado em outubro de 2010.
- [20] TECNAL - produtos para laboratório, http://www.tecnallab.com.br/padrao.aspx?produtos.aspx?id=produtos_tipo, acessado em setembro de 2010.
- [21] SCGÁS - Companhia de Gás de Santa Catarina, <http://www.scgas.com.br/info/ondeaplicarcomum/idse/309>, acessado em setembro de 2010.
- [22] Agência Nacional de Petróleo, Gás Natural e Biocombustíveis - ANP, *Estabelece a especificação do gás natural, de origem nacional ou importado, a ser comercializado em todo o território nacional*, Portaria ANP n° 104/2002 DOU 09/07/2002, **2002**.
- [23] American Society for Testing and Materials - ASTM. *Water Vapor Content of Gaseous Fuels Using Electronic Moisture Analyzers*, ASTM D-5454, ASTM International, **2004**.
- [24] American Society for Testing and Materials - ASTM. *Standard Practice for Calculating Heat Value, Compressibility Factor, and Relative Density of Gaseous Fuels*, ASTM D-3588, ASTM International, **2003**.
- [25] American Society for Testing and Materials - ASTM. *Standard Test Method for Analysis of Natural Gas by Gas Chromatography*, ASTM D-1945, ASTM International, **2010**.
- [26] American Society for Testing and Materials - ASTM. *Standard Test Method for Determination of Sulfur Compounds in Natural Gas and Gaseous Fuels by Gas Chromatography and Chemiluminescence*, ASTM D-5504, ASTM International, **2008**.
- [27] American Society for Testing and Materials - ASTM. *Standard Test Method for Determination of Sulfur Compounds in Natural Gas and Gaseous Fuels by Gas Chromatography and Flame Photometric Detection*, ASTM D-6228, ASTM International, **2003**.
- [28] Associação Brasileira de Normas Técnicas - ABNT/NBR. *Gás natural - Determinação da composição química por cromatografia em fase gasosa*, ABNT NBR 14903:2008, **2008**.
- [29] H. M. Hao, J. A. Cao, Z. Q. Yu, J. Ken and J. H. Liu, *Spectroscopy and Spectral Analysis* **2009**, 29, 2087-2091.
- [30] A. J. Rest, R. Warren and S. C. Murray, Near-infrared study of the light liquid alkanes, *Applied Spectroscopy* **1996**, 50, 517-520.
- [31] M. A. van Aghoven, G. Fujisawa, P. Rabbito and O. C. Mullins, Near-infrared spectral analysis of gas mixtures, *Applied Spectroscopy* **2002**, 56, 593-598.
- [32] G. Fujisawa, M. A. van Aghoven, F. Jenet, P. A. Rabbito and O. C. Mullins, Near-infrared compositional analysis of gas and condensate reservoir fluids at elevated pressures and temperatures, *Applied Spectroscopy* **2002**, 56, 1615-1620.
- [33] IUPAC, J.; Lengyel, A. M.; Ure, A.; Gelencsér, A.; Hulanicki, A. (Eds.) in *Compendium of analytical Nomenclature*, Vol. Blackwell, Oxford, **1998**.
- [34] Wilrich, P. *Capability of detection*, ISO/TC69/SC6, *International Standardization Organization - ISO Standard*, 11843-1, **1995**.

- [35] Commission on Analytical Nomenclature - IUPAC, Recommendations in Evaluation of Analytical Methods including Detection and Quantification Capabilities, *Pure and Applied Chemistry* **1995**, *67*, 1699-1723.
- [36] L. A. Currie, Detection: International update, and some emerging di-lemmas involving calibration, the blank, and multiple detection decisions, *Chemometrics and Intelligent Laboratory Systems* **1997**, *37*, 151-181.
- [37] Alekseev, V. N., *Análise Qualitativa*, Lopes da Silva Editora, **1982**, p. 563.
- [38] CITAC & Eurachem, *Guide to Quality in Analytical Chemistry - An Aid to Accreditation*, **2002**.
- [39] H. Kaiser, Quantitation in elemental analysis, *Analytical Chemistry* **1970**, *42*, 24A-41a.
- [40] L. A. Currie, Limits for qualitative detection and quantitative determination. Application to radiochemistry, *Analytical Chemistry* **1968**, *40*, 586-593.
- [41] C. A. Clayton, J. W. Hines and P. D. Elkins, Detection limits with specified assurance probabilities, *Analytical Chemistry* **1987**, *59*, 2506-2514.
- [42] D. B. Owen, The Power of Students T-Test, *Journal of the American Statistical Association* **1965**, *60*, 320-333.
- [43] M. C. Ortiz, L. A. Sarabia and M. S. Sanchez, *Analytica Chimica Acta* **2010**, *674*, 123-142.
- [44] I. Garcia, M. C. Ortiz, L. Sarabia and J. M. Aldama, Validation of an analytical method to determine sulfamides in kidney by HPLC-DAD and PARAFAC2 with first-order derivative chromatograms, *Analytica Chimica Acta* **2007**, *587*, 222-234.
- [45] E. Voigtman, Limits of detection and decision. Part 2, *Spectrochimica Acta Part B: Atomic Spectroscopy* **2008**, *63*, 129-141.
- [46] D. A. Skoog, F. J. Holler and T. A. Nieman, *Principles of instrumental analysis*, Thomson Learning Academic Resource Center, **1998**, p. 849.
- [47] J. Mandel and R. D. Stiehler, *Journal of Research of the National Bureau of Standards* **1964**, *68A*, 155.
- [48] M. Valcárcel, *Principles of Analytical Chemistry - A textbook*, Springer, **2000**, p. 371.
- [49] M. C. Ortiz, L. A. Sarabia, A. Herrero, M. S. Sánchez, M. B. Sanz, M. E. Rueda, D. Giménez and M. E. Meléndez, Capability of detection of an analytical method evaluating false positive and false negative (ISO 11843) with partial least squares, *Chemometrics and Intelligent Laboratory Systems* **2003**, *69*, 21-33.
- [50] A. Lorber, Error propagation and figures of merit for quantification by solving matrix equations, *Analytical Chemistry* **1986**, *58*, 1167-1172.
- [51] L. Xu and I. Schechter, A Calibration Method Free of Optimum Factor Number Selection for Automated Multivariate Analysis. Experimental and Theoretical Study, *Analytical Chemistry* **1997**, *69*, 3722-3730.
- [52] A. J. Berger, T.-W. Koo, I. Itzkan and M. S. Feld, An Enhanced Algorithm for linear Multivariate Calibration, *Analytical Chemistry* **1998**, *70*, 623-627.
- [53] H. C. Goicoechea and A. C. Olivieri, *Chemometrics and Intelligent Laboratory Systems* **2001**, *56*, 73-81.
- [54] C. M. A. Rasmus Bro, Theory of net analyte signal vectors in inverse regression, *Journal of Chemometrics* **2003**, *17*, 646-652.

- [55] N. M. Faber, Efficient Computation of Net Analyte Signal Vector in Inverse Multivariate Calibration Models, *Analytical Chemistry* **1998**, *70*, 5108-5110.
- [56] B. R. K. Eugenio Sanchez, Tensorial calibration: I. First-order calibration, *Journal of Chemometrics* **1988**, *2*, 247-263.
- [57] M. F. Nicolaas, Mean centering and computation of scalar net analyte signal in multivariate calibration, *Journal of Chemometrics* **1998**, *12*, 405-409.
- [58] R. Bro and C. M. Andersen, Theory of net analyte signal vectors in inverse regression, *Journal of Chemometrics* **2003**, *17*, 646-652.
- [59] B. B. Neto, I. S. Scarminio and E. R. Bruns, *Como Fazer Experimentos - Pesquisa e Desenvolvimento na Ciência e na Indústria*, Editora Unicamp, Campinas, **2003**, p. 401.
- [60] H. Martens and T. Naes, *Multivariate Calibration*, Wiley, **1989**.
- [61] Agência Nacional de Vigilância Sanitária - ANVISA, *Guia para Qualidade em Química Analítica - Laboratório*, Editora Nacional de Vigilância Sanitária, **2004**, pp. 41 - 44.
- [62] H. Günzler, *Accreditation and quality assurance in analytical chemistry*, Springer New York, **1994**, p. 260.
- [63] G. T. Wernimont, *Use of statistics to develop an evaluate analytical methods*, AOAC international, Washington, **1985**, p. 183.
- [64] D. Gerard, Review: Authentication of food and food ingredients by near infrared spectroscopy, *Journal of Near Infrared Spectroscopy* **1996**, *4*, 47-61.
- [65] P. Williams and K. Norris, *Near Infrared Technology in the Agricultural and Food Industries*, American Association of Cereal Chemists, Inc - AACC, **2001**, p. 296.
- [66] H. B. Huang, H. Y. Yu, H. R. Xu and Y. B. Ying, *Journal of Food Engineering* **2008**, *87*, 303-313.
- [67] N. S. Sahni, T. Isaksson and T. Naes, *Journal of near Infrared Spectroscopy* **2004**, *12*, 77-83.
- [68] G. F. Wu and Y. He, Identification of Varieties of Textile Fibers by Using Vis/NIR Infrared Spectroscopy Technique, *Spectroscopy and Spectral Analysis* **2010**, *30*, 331-335.
- [69] E. Cleve, E. Bach and E. Schollmeyer, Using chemometric methods and NIR spectrophotometry in the textile industry, *Analytica Chimica Acta* **2000**, *420*, 163-167.
- [70] K. Krämer and S. Ebel, Application of NIR reflectance spectroscopy for the identification of pharmaceutical excipients, *Analytica Chimica Acta* **2000**, *420*, 155-161.
- [71] M. Dyrby, S. B. Engelsen, L. Norgaard, M. Bruhn and L. Lundsberg-Nielsen, Chemometric quantitation of the active substance (containing C N) in a pharmaceutical tablet using near-infrared (NIR) transmittance and NIR FT-Raman spectra, *Applied Spectroscopy* **2002**, *56*, 579-585.
- [72] P. K. Aldridge, R. F. Mushinsky, M. M. Andino and C. L. Evans, Identification of Tablet Formulations inside Blister Packages by near-Infrared Spectroscopy, *Applied Spectroscopy* **1994**, *48*, 1272-1276.
- [73] J. Stuth, A. Jama and D. Tolleson, Direct and indirect means of predicting forage quality through near infrared reflectance spectroscopy, *Field Crops Research* **2003**, *84*, 45-56.
- [74] A. Alishahi, H. Farahmand, N. Prieto and D. Cozzolino, Identification of transgenic foods using NIR spectroscopy: A review, *Spectrochimica Acta Part a-Molecular and Biomolecular Spectroscopy* **2010**, *75*, 1-7.

- [75] Jerome Workman, Jr., Review: A brief review of near infrared in petroleum product analysis, *Journal of Near Infrared Spectroscopy* **1996**, 4, 69-74.
- [76] A. Espinosa, D. Lambert and M. Valleur, Use Nir Technology to Optimize Plant-Operations, *Hydrocarbon Processing* **1995**, 74, 86-&.
- [77] M. V. Reboucas and B. D. Neto, Near infrared spectroscopic prediction of physical properties of aromatics-rich hydrocarbon mixtures, *Journal of near Infrared Spectroscopy* **2001**, 9, 263-273.
- [78] G. H. Ayres, *Analisis Químico Cuantitativo*, New York, **1970**.
- [79] C. D. Tran, Acousto-Optic Devices: Optical Elements for Spectroscopy, *Analytical Chemistry* **1992**, 64, 971A-981A.
- [80] N. Makhoukhi, E. Péré, R. Creff and C. Pouchan, Determination of the composition of a mixture of gases by infrared analysis and chemometric methods, *Journal of Molecular Structure* **2005**, 744-747, 855-859.
- [81] H. F. M. Boelens, W. T. Kok, O. E. de Noord and A. K. Smilde, Fast on-line analysis of process alkane gas mixtures by NIR spectroscopy, *Applied Spectroscopy* **2000**, 54, 406-412.
- [82] C. W. Brown and J. Zhou, Near-infrared spectra of gases: Identification of mixtures, *Applied Spectroscopy* **2001**, 55, 44-49.
- [83] S. M. Donahue, C. W. Brown, B. Caputo and M. D. Modell, Near-infrared multicomponent analysis in the spectral and Fourier domains: energy content of high-pressure natural gas, *Analytical Chemistry* **1988**, 60, 1873-1878.
- [84] C. W. Brown and S. C. Lo, Near-infrared spectra of gases: Identification of mixtures, *Applied Spectroscopy* **1993**, 47, 812-815.
- [85] C. W. Brown, A. Mangal and S. M. Barnett, Novel Sampling Device for Obtaining Process Near-Infrared Spectra of High-Pressure Gases, *Applied Spectroscopy* **1998**, 52, 746-749.
- [86] P. Gemperline, *Practical Guide To Chemometrics*, CRC Press **2006**, p.
- [87] B. G. Jorgensen, Y. in *Master of Applied Statistics* <http://statmaster.sdu.dk/courses/ST02/module07/index.html>, **acesado em fevereiro/2011**.
- [88] J. A. Munoz and R. G. Brereton, Partial factorial designs for multivariate calibration: extension to seven levels and comparison of strategy, *Chemometrics and Intelligent Laboratory Systems* **1998**, 43, 89-105.
- [89] C. M. Franco in *Determinação de hidrocarbonetos majoritários presentes no gás natural utilizando espectroscopia no infravermelho próximo e calibração multivariada*, Dissertação (mestrado em química), Universidade Estadual de Campinas, Campinas, **2008**.
- [90] International Conference on Harmonization, ICH Harmonized Tripartite Guideline – Validation of Analytical Procedures: Metodology, Fed Regist. 1997, 62, 27463.