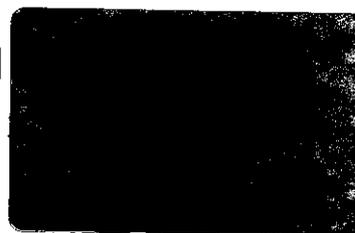


Universidade Estadual de Campinas

Instituto de Química



**Aplicação de Técnicas Espectroscópicas de
Reflexão no Infravermelho no Controle de
Qualidade de Farinha de Trigo**

Tese de Doutorado

Marco Flôres Ferrão

Orientador: Prof. Dr. Celso Ulysses Davanzo

Campinas, julho de 2000.



BANCA EXAMINADORA

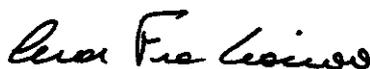
Prof. Dr. Celso Ulysses Davanzo (Orientador)



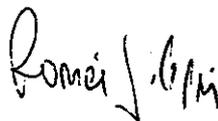
Profa. Dra. Ieda Spacino Scarmínio (DQ-UEL)



Prof. Dr. Cesar Francisco Ciacco (FEA-UNICAMP)



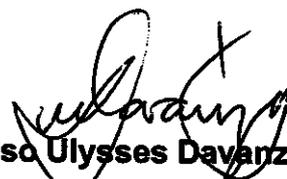
Prof. Dr. Ronei Jesus Poppi (IQ-UNICAMP)



Prof. Dr. Jarbas José Rodrigues Rohwedder (IQ-UNICAMP)



Este exemplar corresponde à redação final da Tese de Doutorado defendida pelo aluno **MARCO FLÔRES FERRÃO**, aprovada pela Comissão Julgadora em 12 de julho de 2000.



Prof. Dr. Celso Ulysses Davanzo

(Presidente da Banca)

Família:

Conjunto que nenhum
método multivariado
consegue discriminar.

Dedico esta tese as mulheres da minha vida, símbolos de
dedicação e vitalidade:

Vó Irene (in memorium)

Vó Floriana

Irmã Fernanda

Sobrinha Sofia

E principalmente a minha Mãe Teresa.

Agradecimentos

- Ao professor Celso pela dedicação ao trabalho e amizade nestes quatro anos de convívio;
- A Empresa Filler pelo financiamento do projeto, fornecimento das amostras e colaboração;
- Aos professores da UNICAMP pela sua amizade e exemplo de conduta científica;
- Aos meus novos amigos: Sílvia, Marcelo Hanada, Nancy, Luiz, Josué, Nick, Pablo, Ângela, Afrânio, Iara, Cláudio e Ritinha, pelos bons momentos que marcaram minha passagem por Campinas;
- Aos colegas César, Marcelo Morgano, Ricardo, Márcia, Melissa, Dalmo, Carla, Rita, Neife, Carol, Débora, Marilene e Kesley pelas experiências trocadas na UNICAMP;
- Aos colegas e funcionários da UNISC;
- Aos funcionários do Instituto de Química – UNICAMP;
- À todas as pessoas e instituições envolvidas neste trabalho;
- E a Universidade de Santa Cruz do Sul e a CAPES pela bolsa concedida.

Resumo

Neste trabalho são combinadas as potencialidades das técnicas quimiométricas e as facilidades na obtenção de espectros apresentadas pelas técnicas de espectroscopia por reflexão difusa no infravermelho com transformada de Fourier (DRIFTS), espectroscopia por reflexão difusa no infravermelho próximo (NIRRS) e espectroscopia por reflexão total atenuada horizontal (HATR) visando a previsão simultânea dos teores de cinzas, de proteína e no número de queda em amostras de farinha de trigo.

Foram utilizadas cem amostras de farinhas de trigo fornecidas pela Filler S.A., indústria localizada na BR 471 km 58, Santa Cruz do Sul. As amostras de farinha foram analisadas em duplicata pelas técnicas espectroscópicas de reflexão no infravermelho. A determinação do teor de cinza tomada como referência foi realizada através do método gravimétrico 08-03 da *American Association of Cereal Chemists* (1984); para a determinação do teor de proteína o método 2055 da *Association of Official Analytical Chemists* (1984); e para a determinação do número de queda foi utilizado o método 56-81B da *American Association of Cereal Chemists* (1982).

Na construção dos modelos foram empregados os dados brutos em forma de $\text{Log}(1/R)$ ou as suas derivadas de 1ª e 2ª ordem para as regiões espectrais selecionadas. A calibração foi efetuada com o método de regressão dos mínimos quadrados parciais (PLS), utilizando os dados centrados na média e/ou escalados pela variância, como ferramentas de pré-processamento, para um conjunto de 55 amostras.

Para validar os modelos foi empregado o SEV (*standard error of validation*), sendo os valores de previsão estimados para as demais 45 amostras, não empregadas na modelagem.

Também foram desenvolvidos modelos empregando dois métodos para a seleção das variáveis. O primeiro foi baseado na subtração de espectros de amostras que continham valores extremos nas propriedades a serem determinadas; o segundo método utilizou as informações dos coeficientes de regressão, dos espectros e dos *loadings* de um modelo PLS referencial.

Os resultados refletem a importância na escolha do tratamento dos dados, bem como do tipo de pré-processamento empregado. Melhores resultados foram obtidos quando da utilização simultânea dos dados centrados na média e escalados pela variância. Os conjuntos de dados modelados (espectros) também apresentaram melhor desempenho para as modelagens nas quais foram empregados os dados com derivadas de 1ª ordem ou 2ª ordem.

Com relação aos métodos espectroscópicos de reflexão, os métodos que empregam espectros no infravermelho médio (DRIFTS e HATR) apresentaram um desempenho superior ao que utiliza espectros no infravermelho próximo (NIRRS), quando é levado em consideração o desempenho destes frente ao conjunto de validação (SEV).

Estes resultados nos permitiram concluir que bons modelos visando a previsão simultânea dos teores de cinzas, de proteína e número de queda podem ser obtidos, apesar de não se ter conhecimento de bandas (números de onda) correlacionados diretamente com o parâmetro teor de cinzas, e que as técnicas de reflexão são adequadas por permitir uma obtenção rápida do espectro e por não gerar resíduos que sejam nocivos ao ambiente.

Abstract

In this study are combined the potentialities of chemometrics techniques and the facilities in the acquisition of spectra presented by the techniques Diffuse Reflectance Infrared Fourier Transform Spectroscopy (DRIFTS), Near Infrared Diffuse Spectroscopy (NIRRS) and Horizontal Attenuated Total Reflection Spectroscopy (HATR) bearing in mind the simultaneous prediction of ash and protein levels, as well as the falling number in wheat flour samples.

One hundred wheat flour samples, which were supplied by Filler S. A. - industry located on BR 471 km 58, Santa Cruz do Sul, were used during the research. The samples were analyzed in duplicate by infrared reflection spectroscopy techniques. For the determination of ash level reference, the gravimetric method 08-03 from the American Association of Cereal Chemists (1984) was used, for the determination of protein level, the 2055 method from the Association of Official Analytical Chemists (1984) was used and for the determination of falling number, the 56-81B method from the American Association of Cereal Chemists (1982) was used.

To build the models the Partial Least-squares regression method (PLS) was used with raw spectra in $\log(1/R)$ form and its first and second order derivatives for select spectral range. The data were the mean-centered data and/or variance scaling, as pre-processing tools, for a 55-sample set.

To validate the models, Standard Error Validation (SEV) was used, and the prevision values were estimated for the other 45 samples not used in the modeling.

Models were also developed, in which two methods for variable selection were applied. One of them was based in the spectra subtraction of sample which included extreme values in the properties to be determined; the other method applied the information of regression coefficients, spectra and loading of a PLS referential model simultaneously.

The results reflect the importance to choose the data treatment, as well as the kind of pre-processing applied. Best results were obtained when there was simultaneous application of mean-centered data and variance scaling. The modeled data sets (spectra) also presented the best performance for the models in which data with first or second order derivatives were applied.

In relation to the spectroscopy methods applied, the ones which use mid-infrared spectra (DRIFTS and HATR) presented a better performance than the ones which apply near infrared spectra (NIRRS), when their performance is considered in relation to the validation set (SEV).

These results let us conclude that good models aiming the simultaneous prediction of ash and protein levels and falling number can be obtained, in spite of the fact that there is not any knowledge about bands (wave numbers) directly correlated to the ash level parameter, and that reflection techniques are adequate to allow a fast obtention of spectra and not to create wastes which are harmful to the environment.

Curriculum Vitae

➤ Formação Acadêmica:

Graduação: # Curso de Licenciatura em Química da UFRGS
Colação de Grau em 23 de dezembro de 1986

Curso de Química Industrial da UFRGS
Colação de Grau em 20 de julho de 1990

Pós-Graduação: # Mestrado em Química da UFRGS
Orientador: Prof. Dr. Dimitrios Samios
Defesa em 15 de maio de 1992

➤ Atividade Profissional: Professor do Departamento de Química e Física da Universidade de Santa Cruz do Sul – RS (antiga FISC) desde o segundo semestre de 1990.

➤ Publicações Científicas:

A Química da Erva-Mate – Erva-Mate: diagnóstico e perspectiva de desenvolvimento. Venâncio Aires: Prefeitura Municipal, p. 23-33, 1996.

Determinação da composição qualitativa e quantitativa dos ácidos graxos presentes em óleos de linhaça e de soja por cromatografia gasosa – Tecno-Lógica, v.1, n.1, p. 43-48, dez. 1997, Santa Cruz do Sul: UNISC.

Aplicação de métodos de regressão multivariados na determinação quantitativa de tñeres a partir de dados de espectroscopia na região do infravermelho com transformada de Fourier – Tecno-Lógica, v.2, n.1, p. 9-31, jan./jun 1998, Santa Cruz do Sul: UNISC.

➤ Principais Trabalhos Apresentados nos Últimos 12 meses:

Determinação Simultânea dos Teores de Cinzas e de Proteína Empregando Regressão por Mínimos Quadrados Parciais em Dados de Espectroscopia por Reflexão no Infravermelho com Transformada de Fourier (DRIFTS) em farinha de Trigo – Durante a 23ª Reunião Anual da SBQ, Poços de Caldas, MG, 2000.

Técnicas de Reflexão no Infravermelho Aplicadas na Previsão do Número de Queda de Farinhas de Trigo Empregando Regressão por Mínimos Quadrados – Durante a 23ª Reunião Anual da SBQ, Poços de Caldas, MG, 2000.

Avaliação de Métodos de Derivatização e Extração de Ácidos Graxos em Óleos Vegetais para Análise por Cromatografia Gasosa - Durante a 23ª Reunião Anual da SBQ, Poços de Caldas, MG, 2000.

Avaliação do Desempenho de Critérios de Seleção de Variáveis Visando a Determinação de Proteína em Farinha de Trigo Empregando Regressão por Mínimos Quadrados Parciais em Dados Obtidos por DRIFTS - Durante o 10º ENQA, Santa Maria, RS, 1999.

Quantificação de Resíduo Mineral em Farinha de Trigo Empregando Regressão Multivariada em Dados de Espectroscopia por Reflexão difusa no Infravermelho com Transformada de Fourier (DRIFTS) - Durante o 10º ENQA, Santa Maria, RS, 1999.

Construção de Modelos de Regressão Empregando Seleção de Variáveis com Base nas Componentes Principais Visando a Determinação de Carbonatos de Cálcio e Magnésio e de Óxido de Magnésio por DRIFTS - Durante o 10º ENQA, Santa Maria, RS, 1999.

Desempenho de Modelos de Regressão por Mínimos Quadrados Parciais e por Componentes Principais Visando a Determinação de Carbonatos de Cálcio e de Magnésio e de Óxido de Magnésio por DRIFTS - Durante o 10º ENQA, Santa Maria, RS, 1999.

Comparação de Modelos de Regressão por Mínimos Quadrados Parciais e por Componentes Principais Visando a Determinação de Matéria Orgânica em Solos Empregando Dados Obtidos por DRIFTS - Durante o 10º ENQA, Santa Maria, RS, 1999.

Estudo da Influência do Número de Varreduras Empregadas em Espectros de Solos Obtidos por DRIFT na Modelagem via Regressão por Mínimos Quadrados Parciais Visando a Determinação de Matéria Orgânica - Durante o 10º ENQA, Santa Maria, RS, 1999.

Desempenho de Modelos de Análise Discriminante Frente a Adulterantes Visando o Controle Qualitativo de Essências Alimentícias Através de Dados de FT-IR/ATR - Durante o 10º ENQA, Santa Maria, RS, 1999.

Previsão do Número de Queda para Farinhas de Trigo Empregando Regressão por Mínimos Quadrados Parciais em Dados obtidos por Espectroscopia HATR no Infravermelho Médio com Transformada de Fourier - Durante o 10º ENQA, Santa Maria, RS, 1999.

Abreviações

AACC	American Association of Cereal Chemists
AOAC	American Association of Official Analytical Chemists
ATR	Attenuated total reflectance
CLS	Classical least-square
CSR	Cyclic subspace regression
CV	Cross validation
DRIFTS	Diffuse reflectance infrared Fourier transform spectroscopy
FTIR	Fourier transform infrared spectroscopy
HATR	Horizontal attenuated total reflectance
HCA	Hierarchical cluster analysis
HLA	Hybrid linear analysis
LDA	Linear discriminant analysis
LVMR	Latent variable multivariate regression
LWR	Locally weighted regression
MLR	Multiple linear regression
MSC	Multiplicative scatter correction
NIRR	Near-infrared reflectance
PC	Principal component
PCA	Principal component analysis
PCR	Principal component regression

PLS	Partial least-square
PRESS	Prediction residual error sum of squares
RCE	Relevant component extraction
RMSECV	Root-mean square error cross validation
RR	Ridge regression
SEC	Standard error of calibration
SEP	Standard error of prediction
SEV	Standard error of validation
SIMCA	Soft independent modeling of class analogies
SVD	Singular value decomposition
UVE	Uninformative variables elimination
VL	Variáveis latentes

Índice Analítico

CAPÍTULO 1	INTRODUÇÃO	1
CAPÍTULO 2	OBJETIVOS	3
CAPÍTULO 3	Farinha de Trigo	4
3.1	INTRODUÇÃO	4
3.2	CLASSIFICAÇÃO DO GRÃO DE TRIGO	5
3.3	ESTRUTURA DO GRÃO DE TRIGO	7
3.4	COMPONENTES BIOQUÍMICOS DO TRIGO	8
3.4.1	PROTEÍNAS	8
3.4.2	GLUCÍDIOS	10
3.4.3	LIPÍDIOS	11
3.4.4	ENZIMAS	12
3.4.5	VITAMINAS E MINERAIS	13
3.5	PROCESSAMENTO DO GRÃO DE TRIGO	15
3.6	CLASSIFICAÇÃO DA FARINHA DE TRIGO	16
3.7	CRITÉRIOS DE QUALIDADE DA FARINHA DE TRIGO	17
3.7.1	UMIDADE	17
3.7.2	TEOR DE PROTEÍNA	19
3.7.3	TEOR DE GLÚTEN	21
3.7.4	TEOR DE CINZA	22
3.7.5	CONCENTRAÇÃO DAS ENZIMAS AMIOLÓTICAS	23

Capítulo 4	TÉCNICAS DE REFLEXÃO NO INFRAVERMELHO	27
4.1	INTRODUÇÃO	27
4.2	REFLEXÃO ESPECULAR	31
4.3	REFLEXÃO DIFUSA	35
4.4	REFLEXÃO INTERNA ATENUADA	44
Capítulo 5	TÉCNICAS DE ANÁLISE MULTIVARIADA	52
5.1	INTRODUÇÃO	52
5.2	ANÁLISE POR COMPONENTES PRINCIPAIS	53
5.3	MÉTODOS DE REGRESSÃO MULTIVARIADA	59
5.3.1	REGRESSÃO LINEAR MÚLTIPLA (MLR)	62
5.3.2	REGRESSÃO POR COMPONENTE PRINCIPAIS (PCR)	65
5.3.3	REGRESSÃO POR MÍNIMOS QUADRADOS PARCIAIS (PLS)	69
5.4	CRITÉRIOS PARA AVALIAÇÃO DE MODELOS DE REGRESSÃO	75
5.5	SELEÇÃO DE VARIÁVEIS E OTIMIZAÇÃO DE MODELOS DE REGRESSÃO	78
CAPÍTULO 6	PARTE EXPERIMENTAL	84
6.1	AMOSTRAGEM	84
6.2	ANÁLISES ATRAVÉS DOS MÉTODOS DE REFERÊNCIA	84
6.3	AQUISIÇÃO DOS ESPECTROS NO INFRAVERMELHO	86
6.3.1	AQUISIÇÃO DOS ESPECTROS HATR	89
6.3.2	AQUISIÇÃO DOS ESPECTROS DRIFTS	91
6.3.3	AQUISIÇÃO DOS ESPECTROS NIRRS	95
6.4	TRATAMENTO DOS ESPECTROS	96
6.4.1	CORREÇÃO DO ESPALHAMENTO DE LUZ (MSC)	97
6.4.2	ALISAMENTO SAVITZKY-GOLAY	98

6.4.3. PRIMEIRA E SEGUNDA DERIVADAS _____	99
6.4.4 FLUXOGRAMAS DOS TRATAMENTOS EMPREGADOS _____	100
6.5 MODELAGEM DOS DADOS ESPECTRAIS _____	103
6.6 VALIDAÇÃO DOS ESPECTROS E DOS DADOS DE REFERÊNCIA _____	106
6.7 SELEÇÃO DAS AMOSTRAS _____	110
6.8 SELEÇÃO DAS VARIÁVEIS _____	113
6.8.1 SELEÇÃO DAS VARIÁVEIS E A ATRIBUIÇÃO DE SINAIS _____	122
CAPÍTULO 7 RESULTADOS E DISCUSSÕES _____	124
7.1 TÉCNICA DE HATR _____	124
7.2 TÉCNICA DE DRIFTS _____	142
7.3 TÉCNICA DE NIRR _____	150
7.4 MODELAGENS PARA ESTIMAR O TEOR DE PROTEÍNA _____	159
7.5 MODELAGENS PARA ESTIMAR O TEOR DE CINZA _____	163
7.6 MODELAGENS PARA ESTIMAR O NÚMERO DE QUEDA _____	168
CAPÍTULO 8 CONCLUSÕES _____	171
CAPÍTULO 9 SUGESTOES PARA TRABALHOS FUTUROS _____	174
REFERÊNCIAS BIBLIOGRÁFICAS _____	175
APÊNDICE A _____	200
APÊNDICE B _____	204
APÊNDICE C _____	206
APÊNDICE D _____	210
APÊNDICE E _____	213

Introdução

Dentre os insumos amplamente utilizados na indústria alimentícia encontramos a farinha de trigo, para qual são realizados diversos ensaios de qualidade destacando-se os teores de proteína e de cinzas.

Ao lado dos ensaios reológicos, ambos parâmetros contribuem na certificação do emprego da farinha na elaboração de determinado produto, sendo que a proteína é também associada às qualidades nutricionais e o teor de cinzas, geralmente, ao preço de mercado.

Além destes parâmetros, também podemos destacar as determinações de umidade, relacionada com a processabilidade dos grãos e da farinha de trigo, do número de queda (*falling number*), que estima a atividade enzimática fator importante na elaboração de produtos de panificação, ensaios de granulometria, de cor, do teor de glúten e os que indicam a presença de impurezas.

Praticamente para todos estes ensaios existem procedimentos que foram sendo definidos ao longo dos anos. Porém, muitos destes procedimentos apesar de fornecerem resultados bastante significativos, podem vir a consumir muito tempo de análise, bem como gerar resíduos.

Nos últimos anos, grande parte das empresas do setor alimentício, em particular àquelas que concentram sua produção na elaboração de massas e biscoitos, vem trabalhando com baixos estoques de suas diversas matérias-primas. Entre estas matérias-primas encontram-se, principalmente, as diferentes farinhas empregadas na linhas de produção, com as quais são realizados diversos ensaios de rotina.

Entretanto, muitos dos ensaios oficiais utilizados não possibilitam um controle imediato dos lotes de farinhas que são diariamente recebidos, e com isso as planilhas contendo os valores destes ensaios, que deveriam gerar uma espécie de diagnóstico de cada lote, acaba por ser apenas um mera conferência formal da matéria-prima, já que os produtos acabados (massas e biscoitos) costumam chegar as empacotadoras, bem antes dos resultados dos ensaios terem sido fechados.

Diante desse quadro empresas do ramo alimentício, tem buscado junto aos centros de pesquisa, alternativas que venham a suprir, em tempo real, o diagnóstico de suas matérias-primas, afim de garantir a qualidade de seus processos e produtos finais, diminuindo as perdas oriundas do emprego inadequado de determinado insumo em sua linha de produção.

Isto significa dizer, que as indústrias do ramo alimentício, além de buscarem novas tecnologias para a elaboração de seus produtos, também vem buscando controles de qualidade cada vez mais rápidos, de forma que estes ensaios possam vir a acompanhar a real velocidade dos processos industriais.

Neste sentido, o emprego de técnicas de reflexão no infravermelho para estimar parâmetros de qualidade em farinhas de trigo tem sido intensamente estudado nos últimos anos, merecendo destaque aqueles que comparam o desempenho dos métodos de regressão multivariada em relação as técnicas de reflexão no infravermelho próximo (NIRR) e no infravermelho médio (DRIFTS).

Cabe ainda ressaltar, que estas técnicas de reflexão que vem sendo empregadas, atendem às necessidades anteriormente citadas, uma vez que tratam-se de técnicas de fácil execução e que depois de modeladas, permitem fornecer estimativas de parâmetros de qualidade em poucos minutos.

Objetivos

- Desenvolver uma rotina de análise simultânea dos teores de cinza e proteína e do número de queda de amostras de farinha de trigo, empregando espectroscopia no infravermelho;
- Comparar o desempenho das técnicas de reflexão difusa e de reflexão total atenuada nas regiões do infravermelho médio e do infravermelho próximo frente a previsão de parâmetros de qualidade de farinhas de trigo;
- Buscar uma metodologia, empregando regressão multivariada, que permita a obtenção de modelos de regressão robustos e de fácil manipulação, que possam ser facilmente transferidos para as Empresas do setor alimentício que têm na farinha de trigo seu principal insumo;
- Otimizar os modelos de regressão multivariada, fazendo uso de ferramentas de pré-processamento de dados espectroscópicos e de tratamento de dados espectrais, afim de prever os parâmetros teor de cinza, teor de proteína e número que queda em amostras de farinha de trigo, com os menores valores de erro possíveis.

Farinha de Trigo

3.1 Introdução

As primeiras sementes de trigo foram trazidas ao Brasil por Martin Affonso de Souza, em 1534, que as plantou na Capitania de São Vicente, e a partir da qual, se estenderam pelo planalto na direção da Região Sul, onde as condições climáticas eram mais favoráveis [1].

Nas duas primeiras décadas do século XIX, o Brasil figurou entre os grandes produtores de trigo, porém o aumento do consumo interno, e do aumento no custo da mão de obra e a substituição por culturas mais lucrativas como o arroz e a soja, tornaram o Brasil um dos grandes importadores.

Em virtude das crescentes dificuldades enfrentadas pelo setor agrícola, a produção nacional da última década tem mostrado uma queda acentuada na produção de trigo, tendo como seus principais produtores os estados do Paraná e Rio Grande do Sul, que somados detêm mais de 90% da produção nacional deste grão. Mesmo assim, a produção brasileira de trigo não supri a metade da demanda nacional, elevando a importação deste produto.

O trigo consumido no Brasil é importado principalmente da Argentina, Canadá, Estados Unidos e França, tendo sido importados a cada ano da década de noventa mais de 5 milhões de toneladas deste produto.

A farinha de trigo, principal derivado do grão de trigo, apresenta diversas aplicações destacando-se seu emprego na panificação, produção de massas e biscoitos, além da comercialização na forma de farinha para a elaboração de diversificados produtos alimentícios.

3.2 Classificação do Grão de Trigo

O trigo é uma gramínea, pertencente à família *Gramineae* e ao gênero *Triticum*, possuindo diversas espécies.

Os tipo de maior interesse comercial são [2]:

- *Triticum aestivum* L. – trigo comum, utilizado no elaboração de pães, bolos, biscoitos e produtos de confeitaria;
- *Triticum compactum* Host. – cultivado nos Estados Unidos da América com o nome de “Club”, também utilizado em produtos de confeitaria;
- *Triticum durum* Desf. – trigo durum, utilizado no preparo de massas alimentícias.

Outro sistema importante de classificação do trigo, que visa estabelecer um valor de mercado para este cereal, é baseado em padrões comerciais, refletindo as políticas de produção e marketing dos países exportadores.

No caso do Brasil, a classificação aceita no mercado nacional é normatizada segundo a Portaria nº 167 de 29 de julho de 1994, que descreve as normas de identidade, qualidade, embalagem e apresentação do trigo [3].

Segundo esta norma, o trigo é classificado em 4 classes e 3 tipos, de acordo com os seguintes critérios:

- Classes: Melhorador, Superior, intermediário e comum, em função dos parâmetros de Farinografia, de Alveografia (W) e do Número de Queda, conforme tabela 3.1.
- Tipos: Expresso por números, é definido em função do limite mínimo do peso hectolitro e dos limites máximos dos percentuais de umidade, grãos danificados, matérias estranhas e impurezas, conforme tabela 3.2.

Tabela 3.1 – Parâmetros para a classificação do trigo.

Classes	Farinografia Estabilidade (min.)	Alveografia (W) (10⁻⁴ J)	Número de Queda (seg.)
Melhorador	14 (mínimo)	280 (mínimo)	200 (mínimo)
Superior	5 (mínimo)	200 (mínimo)	200 (mínimo)
Intermediário	3 (mínimo)	140 (mínimo)	200 (mínimo)
Comum	Quando não se enquadrar em nenhuma das classes acima		

Tabela 3.2 – Parâmetros para a tipificação do trigo.

Tipos	Umidade (%)	Peso do Hectolitro (Kg/hl)	Matérias Estranhas e Impurezas (%)
01	13,00 (máximo)	78 (mínimo)	1,00 (máximo)
02	13,00 (máximo)	75 (mínimo)	1,50 (máximo)
03	13,00 (máximo)	72 (mínimo)	2,00 (máximo)
	Grãos danificados pelo calor, mofados e ardidos (%)	Grãos chochos, triguilhos e quebrados (%)	Grãos danificados por insetos e/ou outras pragas, germinados e esverdeados (%)
01	0,50 (máximo)	1,50 (máximo)	1,0 (máximo)
02	1,00 (máximo)	2,50 (máximo)	1,5 (máximo)
03	2,00 (máximo)	5,00 (máximo)	2,0 (máximo)

3.3 Estrutura do Grão de Trigo

O grão de trigo divide-se, basicamente, em duas partes: o pericarpo (parte mais externa) e a semente, formada pelo endosperma e o germe. Na parte externa da semente, recobrendo o endosperma e o germe, encontram-se a testa (camada onde concentram-se os pigmentos que originam a cor), a hialina e a aleurona.

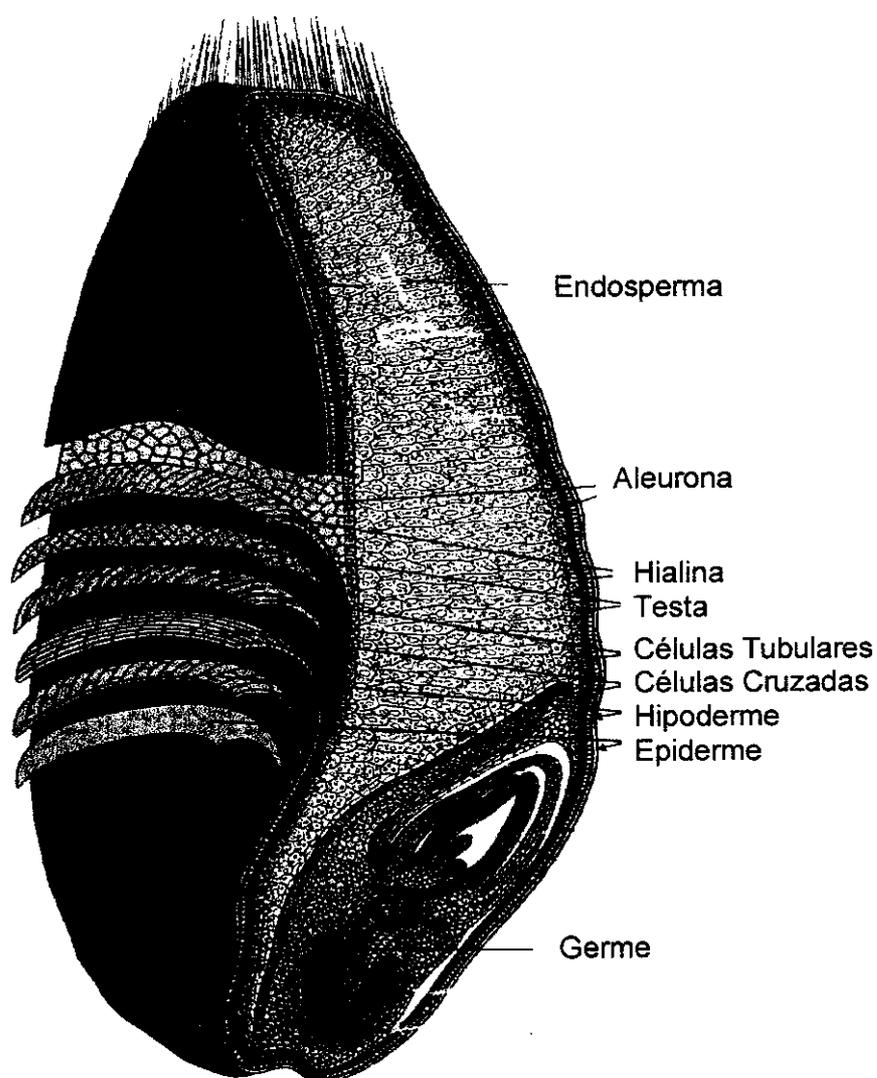


Figura 3.1 – Ilustração da seção longitudinal de um grão de trigo[2].

3.4 Componentes Bioquímicos do Trigo

Dentre as principais classes de compostos que estão presentes no trigo e que apresentam importância nutricional ou tecnológica (de processamento) encontramos as proteínas, os carboidratos, os lipídios, as enzimas, as vitaminas e os minerais.

3.4.1 Proteínas

As proteínas são compostos heteropoliméricos constituídos geralmente por 18 aminoácidos, podendo este número, ser superior a 20. Mediante a hidrólise ácida, as proteínas liberam seus aminoácidos constituintes e mais o íon amônio [4].

Os aminoácidos são compostos orgânicos que contêm um grupo amina (-NH₂) e um grupo carboxílico (-COOH). A estrutura básica de um aminoácido é apresentada na figura 3.2.

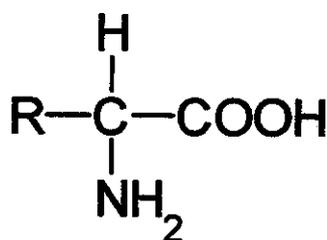


Figura 3.2 – Fórmula estrutural básica de um amino-ácido.

Existem diversas formas de classificar as proteínas, todavia a classificação de maior aplicação pelos pesquisadores e técnicos que trabalham na área de alimentos é baseada na solubilidade [4]:

- albuminas: são proteínas solúveis em água e em soluções salinas diluídas;
- globulinas: são proteínas simples, insolúveis em água pura, mas solúveis em soluções diluídas de sais neutros;
- gluteninas: são proteínas simples, insolúveis em soluções de sais neutros, mas solúveis em soluções ácidas ou alcalinas muito diluídas;
- prolaminas: são proteínas geralmente de origem vegetal, solúveis em soluções alcólicas (70-80%), mas insolúveis em água ou álcool puros;
- albuminóides: são proteínas insolúveis em solventes neutros clássicos, soluções diluídas de ácidos, álcalis ou sais;
- histonas: são proteínas com caráter básico, solúveis em água e soluções ácidas diluídas;
- protaminas: são proteínas com caráter básico acentuado, solúveis em água, em soluções de amônia e não precipitam com o calor.

A quantidade de proteína no trigo é determinada por fatores ambientais e genéticos. Além da quantidade, devemos levar em conta a qualidade das proteínas presentes. Do ponto de vista nutricional, uma proteína é dita balanceada, quando contém a mesma quantidade de cada amino-ácido.

Entre as proteínas balanceadas que compõem o trigo, encontramos as albuminas e as globulinas, que se concentram na aleurona, farelo e germe. Já entre as de importância tecnológica, destacam-se as prolaminas e as gluteninas, presentes principalmente no endosperma, e responsáveis pelas características de viscoelasticidade, adequadas na produção do pão.

3.4.2 Glicídios

Ao grupo dos glicídios pertencem os compostos de estruturas químicas diversas, sendo constituídos sempre de carbono, hidrogênio e oxigênio; às vezes, ainda, de nitrogênio e fósforo. São poli-hidróxi-cetonas ou poli-hidróxi-aldeídos ou compostos que na hidrólise os fornecem [5].

A distribuição dos glicídios pelos diferentes ciclos de moagem pode ser facilmente entendida se considerarmos a estrutura do grão. O farelo, que constitui a parte externa, é formado principalmente por fibras, tendo a função de proteção da semente. O germe, por sua vez, necessita para o crescimento e germinação farta presença de açúcares, sendo a principal fonte de açúcares livres. Na germinação, o sistema enzimático completa o seu desenvolvimento e isto possibilita a utilização do amido, como fonte de energia, sendo o amido necessário retirado do endosperma.

Na tabela 3.3 são apresentadas as composições médias dos glicídios do trigo distribuídos entre as diferentes partes do grão [1].

Tabela 3.3 – Composição média em glicídios distribuídos no grão de trigo.

Componentes	Endosperma	Germe	Farelo
Açúcares	1,5%	36,4%	7,6%
Amido	95,8%	31,5%	14,1%
Hemicelulose	2,4%	15,3%	43,1%
Celulose	0,3%	16,8%	35,2%
Glicídios totais	86,0%	50,5%	70,0%

3.4.3 Lipídios

Os lipídios são substâncias de natureza química e fisiológica bem distintas, que possuem em comum, a insolubilidade em água e a solubilidade em solventes orgânicos.

Os lipídios contêm em sua cadeia carbônica átomos de carbono, hidrogênio e oxigênio, às vezes estão presentes nitrogênio e fósforo e raramente enxofre. São geralmente ésteres de ácidos graxos com álcoois fornecendo por hidrólise ácidos graxos livres.

Os ácidos graxos são ácidos orgânicos que possuem uma cadeia carbônica longa, com um grupo carboxílico terminal.

Os lipídios encontrados nos cereais consistem em glicerídeos de ácidos graxos, principalmente os ácidos palmítico, oléico e linoléico. Contém ainda cerca de 4% de fosfolipídios, como por exemplo a lecitina. O conteúdo de lipídios no trigo representa cerca de 2 a 3% em massa, distribuídos ao longo do grão, concentrando-se principalmente no germe. Na tabela 3.4 são apresentados a distribuição dos lipídios nas diferentes frações de trigo [1].

Tabela 3.4 – Distribuição dos lipídios nas frações do trigo.

Componentes	Endosperma	Germe	Farelo
Esteril ésteres	traços	3,7%	0,5%
Triglicerídeos	29,4%	57,0%	56,1%
Ácidos graxos, esteróis, mono e diglicerídeos	17,1%	17,8%	25,1%
Fosfo e glicolipídios	52,4%	16,5%	22,5%

3.4.4 Enzimas

A denominação de enzima é derivada do grego, significando "em levedura", foi atribuída por Khune em 1878, época em que se acreditava que as enzimas só eram ativas nas células vivas. Esta idéia se manteve até 1897, quando Buchner observou que o extrato obtido por prensagem de células de leveduras ainda possuía a propriedade de fermentar sacarose [6].

As enzimas são proteínas com propriedades catalíticas, que participam de reações bioquímicas nos sistemas vivos e que continuam atuando mesmo nas fases pós-colheita ou pós-abate. Algumas dessas reações são indesejáveis, pois provocam o escurecimento, alterações de sabor e da textura; outras reações são altamente desejáveis e são propositadamente provocadas pela adição de enzimas em determinados processos tecnológicos.

O trigo contém um sistema enzimático próprio, normalmente inativo durante o armazenamento, desde que o grão esteja seco e livre de contaminações por insetos e fungos.

Para a panificação, as enzimas mais importantes são as amilolíticas, que agem sobre o amido em suas ligações α -1,4. Existem também as enzimas desramificantes que agem sobre as ligações α -1,6.

No caso das α -amilases, apresentam atuação mais acentuada sobre o amido danificado ou gelatinizado, resultando na formação de unidades de maltose e glicose. Já as β -amilases atuam nas extremidades não redutoras, liberando unidades de maltose e deixando como resíduo dextrinas de alto peso molecular, conhecidas como dextrinas limite.

Ainda são de importância tecnológica as enzimas proteolíticas, sendo significativa sua ação na panificação e produção de biscoitos doces; e as enzimas lipolíticas que atuam principalmente nos grãos e na farinha integral.

3.4.5 Vitaminas e Minerais

A maior parte dos cereais apresentam-se como importantes fontes de vitaminas do grupo B, como a tiamina (B1), niacina, riboflavina (B2), piridoxina (B6) e ácido pantotênico concentrados no germe e na camada de aleurona. Também contêm tocoferol, precursor da vitamina E, que tem ação antioxidante.

Já os minerais estão principalmente presentes no farelo e consistem basicamente de fosfatos e sulfatos de potássio, magnésio e cálcio. Outros elementos que são encontrados em pequenas concentrações são o ferro, o manganês, o zinco e o cobre.

Na tabela 3.5 são apresentadas as concentrações médias dos principais micronutrientes presentes nos grãos de trigo [1].

Tabela 3.5 – Conteúdo médio de nutrientes do grão de trigo.

Micronutrientes (mg/100g)			
Vitaminas		Minerais	
Tiamina	0,55	Fósforo	410
Riboflavina	0,13	Potássio	580
Niacina	6,4	Cálcio	60
Ác. Pantotênico	1,36	Magnésio	180
Piridoxina	0,53	Ferro	6
Biotina	0,01	Cobre	0,8
Ác. fólico	0,05	Manganês	5,5
Tocoferol	2,7	Zinco	3,3

Os constituintes químicos presentes no trigo não se distribuem uniformemente pelo grão. O pericarpo é rico em celulose e nos componentes precursores das cinzas. As camadas testa e hialina são ricas em celulose, precursores de cinzas e proteína. Já a aleurona é rica em precursores de cinzas, proteínas, lipídios, vitaminas e enzimas. Por outro lado, o endosperma, que representa cerca de 80% em peso do grão, é basicamente constituído de amido. A tabela 3.6 apresenta a distribuição dos principais constituintes químicos presentes no trigo [1].

Tabela 3.6 – Composição química das partes do grão de trigo.

Parte do grão	Proteína (%)	Lipídios (%)	Teor de Cinza (%)	Açúcares redutores (%)	Pentosanas (%)	Celulose (%)	Amido (%)
Farelo							
Pericarpo	7,6	0	5,1	0	34,9	38,4	0
Testa e Hialina	15,7	0	8,1	0	51,1	11,1	0
Aleurona	24,3	8,1	11,1	0	39,0	3,5	0
Endosperma							
Parte externa	16,2	2,2	0,8	1,6	1,4	0,3	63,4
Parte interna	8,0	1,6	0,5	1,6	1,4	0,3	72,6
Germe	26,3	10,1	4,6	26,3	6,6	2,0	0
Grão inteiro	12,1	1,8	1,8	2,0	6,7	2,3	59,2

Valores calculados para 14% de umidade [1]

3.5 Processamento do Grão de Trigo

Para a obtenção da farinha de trigo, os grãos de trigo após terem sido condicionados (acerto da umidade) são submetidos ao processo de moagem. A moagem é sempre realizada de modo gradual para minimizar os gastos de energia e evitar o aquecimento do material.

A farinha obtida em cada etapa da moagem é denominada de fração. Para se alcançar 100% de conversão dos grãos em farinha pode-se gerar 20 ou mais frações, dependendo das características do grão (variedade, formato e umidade), bem como das condições operacionais do moinho.

O endosperma e o pericarpo não se encontram eqüitativamente distribuídos em cada fração, estando a maior parte do endosperma concentrada nas primeiras frações, enquanto que o farelo se concentra nas últimas frações, conforme apresentado na tabela 3.7.

Desta forma, cada tipo de farinha, disponível em determinado mercado, é composta pela mistura de diversas frações, podendo apresentar preço e processabilidade diferenciados.

Tabela 3.7 – Proporção de farelo, germe e endosperma em farinhas de trigo com diferentes taxas de extração.

Taxa de Extração	Farelo	Germe	Endosperma
80%	1,4%	1,6%	77,0%
85%	3,4%	1,9%	79,7%
100%	12,0%	2,5%	85,5%

3.6 Classificação da Farinha de Trigo

A classificação da farinha de trigo no Brasil segue a Norma Técnica conforme Portaria nº 354, de 18 de julho de 1996 [7], que de acordo com seu uso divide-se em farinha de uso doméstico e farinha de uso industrial.

A farinha de uso doméstico divide-se em três categorias:

1 – Farinha de trigo integral: obtida a partir do cereal limpo e com teor máximo de cinzas de 2.0 % na base seca.

2 – Farinha de trigo especial ou de primeira: obtida a partir do cereal limpo, desgerminado com teor máximo de cinzas de 0.65 % na base seca; 98 % do produto deverá passar através de peneira com abertura de malha de 250 μm .

3 – Farinha de trigo comum: obtida a partir do cereal limpo, desgerminado com teor de cinzas entre 0.66 e 1.35 % na base seca; 98 % do produto deverá passar através de peneira com abertura de malha de 250 μm .

A farinha de uso industrial divide-se em duas categorias:

1 – Farinha de trigo integral: obtida a partir do cereal limpo e com teor máximo de cinzas de 2.5 % na base seca, devendo obedecer aos requisitos específicos para cada segmento de aplicação.

2 – Farinha de trigo: obtida a partir do cereal limpo, desgerminado com teor máximo de cinzas de 1,35 % na base seca, devendo obedecer aos requisitos específicos para cada seguimento de aplicação; 98 % do produto deverá passar através de peneira com abertura de malha de 250 μm .

3.7 Critérios de Qualidade da Farinha de Trigo

Uma gama vasta de ensaios de qualidade para o trigo foram desenvolvidos afim de avaliar as características, tanto do grão como das farinhas obtidas em moinhos. Dentre os principais métodos, encontram-se aqueles que visam caracterizar a farinha quanto ao seu preço de mercado, bem como aqueles que definem características necessárias na produção de determinado produto. Esses critérios dão informações sobre as qualidades tecnológicas e de processamento da farinha.

3.7.1 Umidade

A quantidade de umidade é de suma importância, pois é diretamente responsável pela manifestação da maior parte das deteriorações. Muitas das reações bioquímicas que ocorrem no trigo dependem do teor de umidade, assim como os microrganismos, como por exemplo os fungos, precisam de água para se desenvolver.

Diversos métodos para a determinação da umidade são descritos na literatura oficial, adequando-se àquele que atenda às necessidades de cada linha de produção [1, 8]. O método gravimétrico comumente empregado, utiliza uma pequena quantidade de amostra (2-3 gramas), que é levada à estufa a 130°C, ficando por um período de 1 hora. A diferença de peso sofrida pela amostra, durante este processo, é referente a quantidade de água.

Uma das primeiras aplicações da espectroscopia no infravermelho próximo para a determinação da umidade em amostras de farinha de trigo foi proposta por Law e Tkachuk [9] em 1977, empregando regressão linear múltipla.

Em 1978, Williams e Thompson [10] discutem o efeito da moagem e da resultante granulometria das partículas da farinha de trigo no desempenho dos modelos de regressão visando a determinação da umidade. Eles empregam 3 tipos de moinho e estudam os valores médios de tamanho de partícula, bem como a distribuição do tamanho das partículas em diversas amostras.

No início dos anos 80 são disponibilizados comercialmente os primeiros equipamentos, que operavam na região do infravermelho próximo, com a função de determinar a porcentagem de umidade e o teor de proteína em amostras de farinha, visando o controle de qualidade [11-12].

Em 1982, Osborne *et alii* [13-14] apresentam as vantagens, principalmente no que se refere ao baixo tempo de análise, do emprego de técnicas de reflexão no infravermelho próximo no controle de qualidade da farinha de trigo.

Neste mesmo ano, Williams, Norris e Zarowski [15] publicam uma pesquisa onde é avaliada a influência da temperatura nas determinações espectrofotométricas de proteína e umidade em trigo.

Um extenso trabalho foi realizado por Osborne e Fearn em 1983 [16-17], onde são experimentados diversas calibrações para determinar proteína, umidade e dureza do grão de trigo, empregando diferentes espectrofotômetros no infravermelho próximo. Este estudo de abrangência interlaboratorial avaliou a reprodutibilidade e a precisão destas determinações espectrofotométricas.

Novos estudos foram posteriormente apresentados em 1987 por Osborne [18], em 1994 por Forina *et alii* [19], em 1996 por Hong *et alii* [20] e em 1998 por Gaines e Windham [21], todos empregando espectroscopia por reflexão no infravermelho próximo. Destaca-se o trabalho de Forina *et alii* [19] por ser um dos primeiros a empregar o método dos mínimos quadrados parciais e a estudar procedimentos de validação para estas metodologias.

3.7.2 Teor de proteína

Os teores de proteína são regidos por fatores genéticos e ambientais associados ao cultivo do trigo, que incluem o local de plantio, às condições climáticas, pelas práticas culturais (rotação de cultura e adubação nitrogenada) e ainda pelas doenças e pragas.

Do ponto de vista tecnológico, a principal característica correlacionada a quantidade de proteína, vem da capacidade desta em fornecer o glúten.

Para a determinação quantitativa de proteínas encontramos o método Kjeldahl, que dosa a porcentagem de nitrogênio, sendo descrito pela primeira vez em 1883, e ainda é considerado como um dos mais exatos [8, 22-24].

Outro método oficializado desde a década de 70, baseia-se na determinação espectrofotométrica por reflexão no infravermelho próximo (NIRR) [22]. Este método que tem sido utilizado como referencial nos mais recentes empregos de técnicas de regressão na quantificação da proteína em cereais.

Dentre os primeiros estudos empregando espectroscopia NIRR encontram-se os trabalhos publicados por Williams *et alii* [10, 25, 26], que apresentam o efeito do tamanho de partícula nas determinações espectroscópicas quantitativas.

Ainda no final dos anos 70 foram realizados estudos de quantificação de proteína em amostras de farinha de trigo de diferentes variedades e de diferentes origens, como no trabalho apresentado por Watson *et alii* [27], ou que abordam o efeito do moinho empregado para a geração da amostra no desempenho do modelo de calibração publicados por Williams *et alii* [15] e Hunt *et alii* [28].

Em 1978, Miller *et alii* [29] publicam os resultados do emprego da técnica NIRR na determinação de proteína em amostras de trigo visando avaliar a reprodutibilidade tanto do ponto de vista intralaboratorial, quanto do ponto de vista interlaboratorial.

Nos primeiros anos da década de 80, a espectroscopia NIRR foi amplamente explorada no controle de qualidade da farinha, sendo representativos os trabalhos que reportam a calibração de proteína, umidade, tamanho de partícula, cor e amido danificado, publicados por Diachuk *et alii* [11], Williams *et alii* [12], Osborne *et alii* [14], Tkachuk [30] e Czuchajowska e Pomeranz [31].

Em 1982, Osborne [32] investiga o desempenho da espectrofotometria no infravermelho próximo na determinação de proteína para diversos lotes de trigo, colhidos em diferentes anos, mostrando não ser necessário um ajuste do modelo com o passar dos anos.

Estudos empregando espectroscopia NIRR, que envolvem análises interlaboratoriais, foram também publicados nos anos 80 por Osborne e Fearn [16-17], por Williams e Krischenko [33], por Barton II e Windham [34], e nos anos 90, por Delwiche *et alii* [35].

Em meados dos anos 80, Williams *et alii* [36] desenvolvem modelos de regressão para determinar umidade e proteína em trigo, só que desta vez empregam espectroscopia no infravermelho próximo por transmissão.

Os anos 90, deram vez a aplicação da espectroscopia por reflexão difusa no infravermelho médio com transformada de Fourier (DRIFTS) na determinação quantitativa de fibra e proteína. A grande maioria dos trabalhos publicados tratam da comparação entre o desempenho da técnica consagrada NIRR com a recente aplicação da técnica DRIFTS para esta finalidade. Do ponto de vista quantitativo, destacam-se os trabalhos publicados por Reeves III *et alii* [37-40].

Recentemente, Sgrulletta e De Stefanis [41-42], revisitam o emprego da técnica de reflexão no infravermelho próximo visando a determinação simultânea do teor de proteína e da cor em amostras de farinha de trigo *Triticum durum*.

3.7.3 Teor de Glúten

Glúten é o nome genérico dado ao conjunto de proteínas insolúveis do trigo que possuem a capacidade de formar massa, ou seja, quando são misturadas farinha de trigo e água pode-se observar a formação de uma massa constituída da rede protéica do glúten ligado aos grânulos de amido [43].

Os métodos utilizados com a finalidade de avaliar a qualidade da proteína presente na farinha de trigo, baseiam-se na lavagem da farinha com soluções apropriadas, visando a eliminação principalmente do amido e demais proteínas não formadoras do glúten.

Os métodos de determinação do glúten para o controle de qualidade de farinhas encontram-se entre os mais recentes [8, 44].

Estes métodos são de difícil padronização, podendo os erros experimentais serem bastante elevados. De qualquer forma, eles permitem estimar a quantidade e a qualidade da proteína presente em uma determinada farinha [1].

O emprego das técnicas espectroscópicas ainda é restrito a análise qualitativa, a exemplo do trabalho de Robert *et alii* [45], que empregam espectros obtidos na região do infravermelho próximo e que são associados à análise discriminante.

3.7.4 Teor de Cinza

A cinza é o resíduo obtido após a incineração ou calcinação do grão e/ou da farinha proveniente da moagem. Uma vez que os minerais, principais constituintes da cinza, encontram-se mais concentrados nas camadas mais externas dos grão de trigo, as farinhas com altos teores de cinzas são consideradas de inferior qualidade tecnológica, seja por apresentarem maior quantidade de farelo incorporado, ou sujeira incorporada ao grão, antes da moagem.

No Brasil, o teor de cinzas é utilizado como critério para diferenciar os principais tipos de farinha de trigo existentes no mercado [7], estando diretamente relacionado ao preço praticado dentro do mercado nacional.

Os métodos gravimétricos são empregados na determinação das cinzas, podendo ser utilizadas outras variantes oficializadas na literatura [8]. Também, apesar de bastante antigos, os métodos gravimétricos não caem no esquecimento, sendo constantemente checados e avaliados como apresentado no estudo interlaboratorial publicado em 1989 por Mortensen e Wallin [46].

Alguns autores também empregam a medida de cinzas para estimar o grau de moagem e de extração de uma determinada farinha de trigo [1,47].

Do ponto de vista do uso da espectroscopia como uma ferramenta quantitativa, o principal emprego tem sido concentrado nas técnicas que utilizam a região do infravermelho próximo [11].

Em 1986, Iwamoto *et alii* [48] empregaram espectroscopia por reflexão difusa no infravermelho próximo (NIRR) para determinar o teor de cinza de farinhas obtidas de trigos cultivados no Japão.

A análise quantitativa do teor de cinza e outros parâmetros, foi estudada por Szalánczy em 1987 [49] empregando espectroscopia por reflexão no infravermelho próximo.

As determinações do teores de cinzas e umidade também foram estudadas empregando a técnica de termogravimetria, que é descrita no trabalho publicado por Acquistucci *et alii* em 1991 [50].

No início dos anos 90, Symons e Dexter [51-52] empregam a técnica de espectrofotometria por fluorescência para determinar o teor de cinzas e a cor de farinhas investigando o pericarpo e a camada aleurona.

A retomada da espectroscopia no infravermelho próximo ocorre a partir de 1996 com a publicação dos estudos de De Aldana *et alii* [53] que determinam cinza em pastagens; e dos estudos de Bruno-Soares *et alii* [54], que buscam estimar diversos parâmetros empregando espectroscopia NIRR.

3.7.5 Concentração das Enzimas Amilolíticas

A enzima do trigo de maior relevância tecnológica é a α -amilase, pois sua alta atividade pode dar indicativos de degradação, além de ser importante no êxito da panificação. Sendo assim, torna-se relevante a identificação da atividade enzimática, principalmente nas farinhas comercializadas.

Os testes mais freqüentes são o viscoamilógrafo, o emprego do método nefelométrico e a determinação do número de queda (falling number) [8], que apesar de não quantificarem a enzima, permitem comparar diferentes farinhas e discriminar aquelas que apresentam níveis adequados para a produção de determinado produto.

Na tabela 3.8 é apresentada a classificação da qualidade do grão de trigo de acordo com os valores de número de queda (falling number) sugerida por Perten [55].

Tabela 3.8 – Classificação da qualidade de grãos de acordo com os valores de número de queda.

Classificação	número de queda (s)
alta atividade enzimática	≤ 200
atividade enzimática ideal	201-350
baixa atividade enzimática	≥ 351

Uma excelente revisão dos métodos laboratoriais empregados na determinação da α -amilase é apresentada por Asp [56]. Neste artigo são destacados os métodos colorimétricos, viscométricos, nefelométrico, fluorométrico, cromatográfico e imunológico.

Dentre os métodos colorimétricos destaca-se o método cromogênico apresentado por Mathewson e Pomeranz [57-58], e em 1981 pelo estudo interlaboratorial publicado por Mathewson *et alii* [59].

Uma significativa divulgação dos métodos viscométricos foi realizada no início dos anos 80 que iniciou com o lançamento do método turbidimétrico ainda no final dos anos 70 [60-62]. Neste período o método do número de queda (descrito no apêndice A deste trabalho), também figurou como um dos mais empregados, principalmente pela facilidade de execução [63-66].

Paralelamente, o método nefelométrico para a determinação da α -amilase em cereais foi introduzido por Prasad *et alii* [67] em 1979, sendo nos anos seguintes abordado por O'connell *et alii* [68] e por Osborne *et alii* [69].

Neste mesmo período, Mathewson *et alii* [70], publicam novos resultados comparativos interlaboratoriais empregando o método colorimétrico para a determinação da α -amilase em trigo.

Os métodos viscométricos utilizados na determinação da α -amilase em trigo, foram então inicialmente comparados em 1982 através do trabalho publicado por D'appolonia *et alii* [71], no qual os resultados obtidos através do analisador de amylase (*grain amylase analyzer*) são correlacionados com valores obtidos de número de queda e via amilógrafo.

Em 1987, Ross *et alii* [72] apresentam o método do *visco-analyzer* para estimar o amido danificado, que posteriormente foi otimizado por Batey *et alii* [73].

Outras determinações incluem a quantificação das impurezas no grão e na farinha, como por exemplo os trabalhos que aplicam a espectrofotometria por reflexão no infravermelho próximo na detecção de infestações internas e externas por insetos publicados por Ridgway e Chambers [74] e por Dowell *et alii* [75].

A espectrofotometria por reflexão no infravermelho próximo também foi empregada na determinação do peso e da dureza de grãos de trigo. Os trabalhos de Osborne e Fearn [17], Gaines e Windham [21], Williams [25], Williams e Sobering [76], Windham *et alii* [77] e Delwiche *et alii* [78] são exemplos desta aplicação.

Além disso, foram ainda publicados trabalhos que estimam o parâmetro cor de farinhas de trigo empregando a técnica NIRRS, entre os principais

trabalhos encontramos os realizados por Osborne *et alii* [14], por Sgrulletta e De Stefanis [41-42] e por Dowell [79].

Amido danificado, que em determinada quantidade pode ser benéfico na panificação, tem sido um dos parâmetros mais estudados utilizando espectroscopia no infravermelho. Dentre os principais trabalhos encontramos os publicados por Osborne *et alii* [14], Hong *et alii* [20], Szalánczy [49], Osborne e Douglas [80], Finney *et alii* [81], Orman e Schumann Jr. [82], Delwiche e Weaver [83], Morgan e Williams [84], Guy *et alii* [85], Zeng *et alii* [86] e Millar *et alii* [87].

Quanto aos ensaios empíricos (ou reológicos), como por exemplo o número de queda (*falling number*), a farinografia, a extensiografia e a alveografia, nenhum trabalho que envolva correlação com técnicas espectroscópicas foi localizado na revisão da literatura.

Técnicas de Reflexão no Infravermelho

4.1 Introdução

Com certeza, o avanço da espectroscopia no infravermelho médio, como técnica para análise quantitativa, deve-se a combinação da transformada de Fourier e da nova geometria dos espectrofotômetros com a utilização do interferômetro de Michelson [88-91], tornando-os mais rápidos e robustos. Um esquema do "coração" do espectrofotômetro FT-IR pode ser visualizado na figura 4.1

Embora, existam outras concepções óticas para a instrumentação do infravermelho próximo [91], o presente trabalho empregou os princípios do interferômetro de Michelson, tanto para as aquisições de espectros realizadas na região do infravermelho médio, quanto nas análises que empregaram a região do infravermelho próximo.

Um detalhamento sobre as configurações utilizadas nas técnicas espectroscópicas que são avaliadas neste trabalho, será oportunamente apresentado no capítulo que descreve os procedimentos experimentais adotados.

Os anos 90 foram marcados pela utilização da espectroscopia no infravermelho médio para a análise qualitativa e quantitativa de alimentos. Wilson [92], em 1990, introduz as potencialidades da região do infravermelho médio e as técnicas de reflexão: reflexão total atenuada (ATR), detecção fotoacústica e reflexão difusa (DRIFTS), para análise de alimentos e sua utilização no controle industrial.

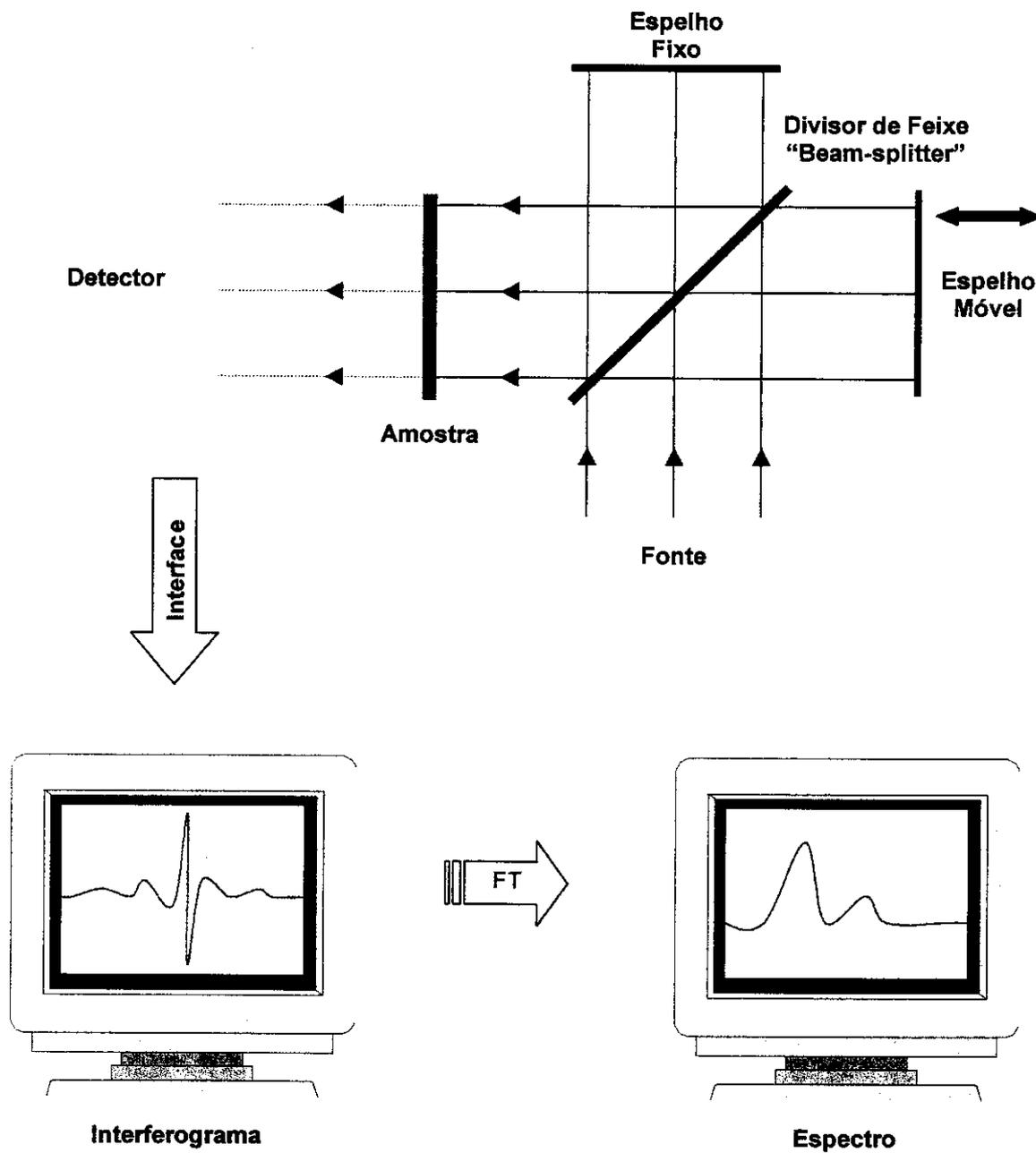


Figura 4.1 - Esquema ilustrativo para o interferômetro de Michelson e do espectro resultante da aplicação da transformada de Fourier.

Em 1991, van de Voort e Ismail [93] revelam as vantagens do emprego da espectroscopia por reflexão total atenuada (ATR) na análise quantitativa de leite, exaltando a necessidade de investigar o emprego desta técnica em outros alimentos. Van de Voort [94], no ano seguinte, descreve as principais aplicações da espectroscopia ATR na análise de leite, carne, óleos, gorduras, manteiga, margarina, leite condensado e frutas, avaliando tanto aspectos quantitativos quanto qualitativos, como a certificação da qualidade.

Nos últimos anos, Reeves III e Zapf [95] empregam a espectroscopia por reflexão difusa no infravermelho com transformada de Fourier (DRIFTS) para discriminar vários ingredientes alimentares.

Em 1999, Wilson e Tapp [96] trazem uma revisão para o emprego da espectroscopia no infravermelho médio na análise de alimentos. Destacam-se as aplicações da espectroscopia por transmissão empregada para estimar o índice de peróxido de óleos vegetais, por reflexão difusa sendo principalmente utilizada na certificação dos alimentos e por reflexão total atenuada onde os trabalhos foram concentrados na determinação de açúcares.

Todavia, o emprego da espectroscopia no infravermelho próximo é consagrada desde o início da década de sessenta, quando Hart *et alii* [97] apresentam um método para determinar umidade em sementes baseado na análise espectrofotométrica no infravermelho próximo. Diversas outras aplicações são descritas por Watson [98], onde o infravermelho próximo é empregado na análise de produtos agrícolas.

A aplicação da espectroscopia no infravermelho próximo na análise de alimentos foi inicialmente restrita ao emprego da reflexão difusa (NIRR), que foi amplamente empregada na análise de grãos, farinhas e produtos industrializados [99-100]. Entre estas aplicações foi consagrada o emprego da espectroscopia NIRR na determinação de proteína em substituição ao método

Kjeldahl, que apesar de ser bastante preciso é de procedimento lento e gerador de resíduos [101].

No início dos anos 80, Osborne [13, 102, 103] traz umas das primeiras retrospectivas da espectroscopia por reflexão no infravermelho próximo discutindo sobre suas principais aplicações e, particularmente, enfocando o trigo, a farinha e os produtos industrializados.

Em 1984, Davies [104] apresenta os principais avanços da técnica NIRR, discutindo os novos métodos de calibração e regressão multivariada, de seleção de comprimentos de onda para a calibração e as novas formas de utilização da espectroscopia NIRR na indústria de alimentos. McShane em 1989 [105] reforça a perspectiva da utilização da espectroscopia NIRR no controle de processos na indústria alimentícia.

Nos anos 90, os princípios das aplicações da técnica NIRR foram reportados por Williams e Stevenson [106], que também apresentam algumas das aplicações típicas desta técnica na análise de alimentos. McQueen *et alii* [107], em 1995, acrescentam a técnica fotoacústica para a espectrofotometria no infravermelho próximo. E ainda neste mesmo período, Dupuy *et alii* [108] empregaram as fibras ópticas para a classificação de vegetais utilizando espectroscopia no infravermelho.

Recentemente, Archibald *et alii* [109, 110] revisitam a espectroscopia NIRR, comparando-a com a espectroscopia Raman, e descrevem métodos de análise e de regressão multivariada para a determinação de fibra em cereais.

Segundo a maioria dos autores citados, três métodos de reflexão no infravermelho têm apresentado aplicações práticas na aquisição de espectros que contenham informações químicas de determinada matriz [88-91, 111, 112]. São eles: o método por reflexão especular (ou externa), o método por reflexão difusa e o método por reflexão interna.

4.2 Reflexão Especular

A reflexão externa ou especular não é uma técnica muito popular em espectroscopia no infravermelho considerando o seu potencial de utilização *in situ*. Uma das principais razões reside na forma dos espectros que, geralmente, são muito diferentes daquela quando os espectros são visualizados no formato transmitância ou absorbância [113].

A reflexão especular ocorre na maioria das superfícies que constituem a fronteira entre fases. A radiação refletida depende da natureza do feixe incidente, bem como das propriedades ópticas dos meios em questão [114, 115].

A reflexão especular ou externa de uma superfície é definida como a razão da intensidade de radiação refletida pela intensidade da radiação incidente, na qual o ângulo de incidência sobre a amostra é exatamente igual ao ângulo de reflexão.

A reflexão especular ocorre em certa extensão para todas as superfícies que constituem interfaces de fases condensadas. No entanto para que ela seja observada predominantemente é necessário que a superfície seja lisa e plana em toda a extensão da área coberta pelo feixe de luz incidente [116, 117].

A espectroscopia de reflexão especular ou externa é claramente descrita pelas equações de Maxwell e de Fresnel. Contudo as fórmulas resultantes para reflexão requerem que os parâmetros tais como ângulo de incidência, a polarização da radiação incidente, a espessura de materiais (filmes sob substratos) e as constantes ópticas sejam conhecidos [118].

As dependências do espectro de reflexão especular com as funções ópticas (ângulo de incidência, polarização da luz, etc.) podem ser entendidas a

partir das condições de fronteira que governam a passagem da radiação eletromagnética de um meio para outro.

Quando a radiação eletromagnética incide em uma interface separando dois meios, conforme ilustrado na figura 4.2, uma parte da radiação é refletida para o meio 1, enquanto que parte continua no meio 2, mas a direção da propagação é alterada, esse fenômeno é chamado de refração.

Na figura 4.2 os índices i , t , e r referem-se ao feixe incidente, transmitido e refletido, respectivamente, n_1 e n_2 são os índices de refração da fase 1 e 2, e θ_i e θ_r são os ângulos de incidência e de reflexão, enquanto que θ_t é o ângulo de refração, medidos da superfície normal.

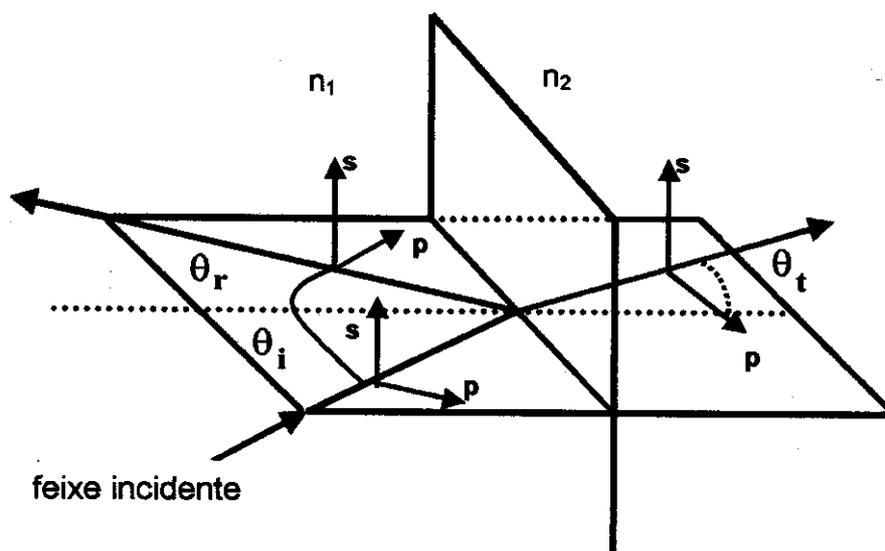


Figura 4.2 - Representação da reflexão e refração de uma onda eletromagnética no plano para uma interface.

Neste esquema foi considerado que $n_2 > n_1$, e que os ângulos θ_i e θ_t são relacionados pela lei de Snell, conforme expressão 4.1, e conseqüentemente o ângulo refletido é igual ao ângulo incidente.

$$n_1 \cdot \text{sen} \theta_i = n_2 \cdot \text{sen} \theta_t \quad (4.1)$$

Assume-se que a normal a superfície está ao longo do eixo z, e a radiação incidente no plano de incidência (x,z). A direção da propagação da radiação incidente é descrita por um vetor unitário.

As intensidades refletidas e refratadas dependem do estado de polarização em relação a radiação incidente. Considerando a lei da propagação nas interfaces pode-se encontrar a amplitude relativa de reflexão.

A descrição da reflexão da luz para uma simples interface entre dois meios é dada pelas equações de Fresnel. As amplitudes relativas para reflexão numa interface entre dois materiais (meios 1 e 2) não absorventes são expressas pelas equações 4.2 e 4.3.

$$r_s = \frac{n_1 \cos \theta_1 - n_2 \cos \theta_2}{n_1 \cos \theta_1 + n_2 \cos \theta_2} \quad (4.2)$$

$$r_p = \frac{n_2 \cos \theta_1 - n_1 \cos \theta_2}{n_2 \cos \theta_1 + n_1 \cos \theta_2} \quad (4.3)$$

onde $r = E_r/E_i$ e E_r e E_i são as amplitudes dos vetores campo elétrico refletido e incidente, respectivamente, θ_1 é o ângulo de incidência no primeiro meio com índice de refração n_1 e θ_2 é o ângulo de refração no segundo meio com índice de refração n_2 .

Quando a radiação incidente é polarizada no plano, aplica-se a equação r_p , enquanto r_s refere-se a polarização perpendicular.

Assumindo que o primeiro meio é o ar ($n_1 = 1$) e se o segundo meio consiste em uma amostra absorvente, então o índice de refração é complexo e é expresso conforme equação 4.5.

$$n_2 = n + i.k \quad (4.5)$$

onde $i = \sqrt{-1}$, n é o índice de refração e k é o índice de absorção, o qual está relacionado com a absorvidade (α), conforme apresentado na equação 4.6.

$$k = \alpha \cdot \ln(10/4\pi\nu) \quad (4.6)$$

onde ν é o número de onda.

A reflexão R é dada pelo quadrado do valor absoluto da reflectividade (r). Os valores reais para R são expressos por 4.7, onde os valores de r seguem a equação 4.8.

$$R = |r|^2 \quad (4.7)$$

$$r_{(\lambda)} = \frac{E_{r(\lambda)}}{E_{i(\lambda)}} \quad (4.8)$$

A reflexão R para incidência normal e o meio n_1 correspondendo ao ar, isto é $\theta_i = \theta_r \cong 0$, pode ser reduzida, conforme expressão 4.9, sendo independente da polarização da luz.

$$R = (n_2 - 1)^2 / (n_2 + 1)^2 \quad (4.9)$$

A aplicação do processo matemático conhecido como a transformada de Kramers-Krönig pode converter o espectro de reflexão especular em uma estimativa das constantes ópticas (n e k). A alta razão sinal ruído melhora a qualidade da transformação do espectro [119, 120].

A maior parte dos acessórios para reflexão especular são projetados de forma a dirigir a radiação infravermelha de baixo para cima, refletindo da amostra que é colocada com a face refletora em um suporte horizontal.

A intensidade de luz refletida é usualmente baixa, fazendo com que essa técnica tenha potencialidade de aplicação no infravermelho com transformada de Fourier (FT-IR).

Quando da obtenção do espectro de reflexão especular, outras formas reflexões podem estar presentes contribuindo para o sinal. Isto pode ser resultante do espalhamento devido a rugosidade da amostra, ou por defeitos no interior da amostra. De forma semelhante, a reflexão especular geralmente ocorre simultaneamente a outras formas de reflexão.

4.3 Reflexão Difusa

A técnica de reflexão difusa está bem fundamentada na literatura [92, 94, 97, 114, 120-123], sendo amplamente aplicada associada aos equipamentos que operam no infravermelho próximo, com os quais geralmente é designada de espectroscopia de reflexão no infravermelho próximo (NIRS ou NIR); ou associada àqueles que operam na região do infravermelho médio, sendo conhecida por espectroscopia por reflexão difusa no infravermelho médio com transformada de Fourier (DRIFTS).

A reflexão difusa ocorre em superfícies não totalmente planas, podendo o substrato ser contínuo ou fragmentado (na forma de pó). Neste processo de reflexão, o feixe incidente penetra a superfície da amostra interagindo com a matriz, retornando a superfície da mesma, após absorção parcial e múltiplos espalhamentos, conforme ilustrado pela figura 4.3.

Na reflexão difusa, a radiação incidente entra em contato diversas vezes com as partículas da amostra sendo conseqüentemente atenuada. Com isto a radiação que sofre a reflexão difusa fornece informações qualitativas e quantitativas, sobre a natureza química da amostra.

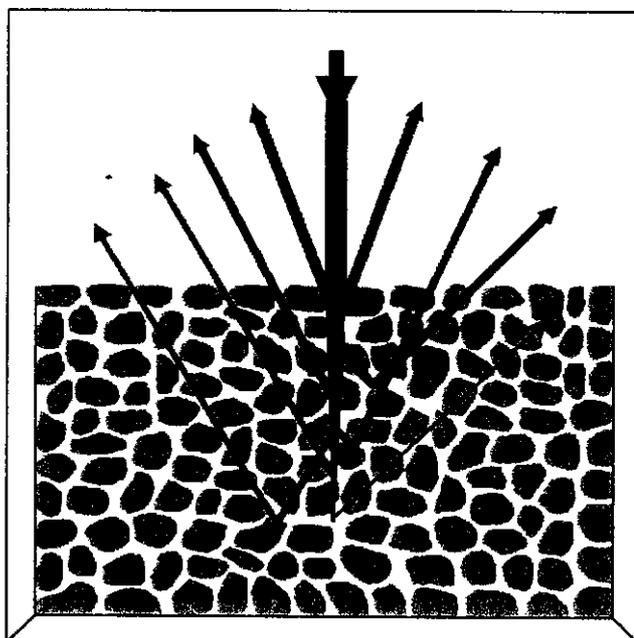


Figura 4.3 - Representação da reflexão especular e difusa de uma onda eletromagnética em uma amostra particulada [123].

É importante salientar que a radiação que retorna de forma difusa de um substrato é geralmente muito inferior em magnitude que a radiação incidente, em função disso, a maior parte dos acessórios de reflexão difusa apresentam esquemas óticos que visam concentrar a radiação, para posteriormente a mesma ser focada sobre o sistema de detecção dos instrumentos.

Outro aspecto importante é a ocorrência simultânea da reflexão especular (reflexão que se dá na superfície do substrato e que é igual ao ângulo de incidência), que pode ser mais ou menos pronunciada em função das características da matriz com a qual o feixe incidente interage. Em geral, este processo pode ser visualmente detectado ao se examinar o espectro de reflexão difusa no infravermelho, e constatar a presença de sinais em forma de derivada, conforme ilustrado na figura 4.4.

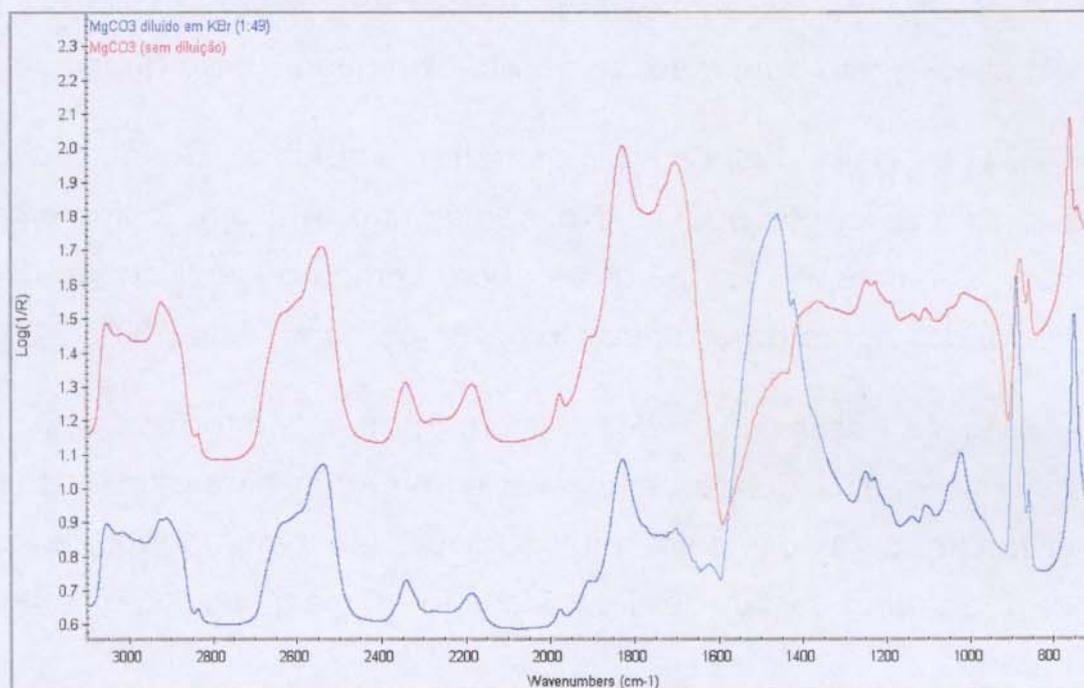


Figura 4.4 – Espectros obtidos por DRIFTS de amostras de carbonato de magnésio com (em azul) e sem (em vermelho) diluição em KBr.

A ocorrência simultânea da reflexão especular é mais freqüente em regiões do espectro onde a absorção da radiação é mais intensa. Outra consequência da ocorrência da reflexão especular é o deslocamento das bandas.

Do ponto de vista da análise quantitativa, além das anomalias que podem ser atribuídas à presença do fenômeno da reflexão especular, merece ainda destaque o efeito que os diferentes tamanhos de partículas podem ter sobre o espectro. Estas anomalias estão amplamente descritas nos trabalhos publicados por Olinger e Griffiths [124-126].

Para reduzir o efeito do tamanho de partícula, a exemplo de Devaux et alii [127], muitos autores têm proposto transformações no espectro, tais como a primeira e a segunda derivadas ou correção do espalhamento multiplicativo (MSC).

Os efeitos da reprodutibilidade da técnica de reflexão difusa também são discutidos, tanto para medidas qualitativas, quanto para quantitativas.

Fuller e Griffiths [128, 129] demonstraram o potencial da técnica em várias áreas para a aplicação quantitativa, como por exemplo na caracterização de polímeros sintéticos, carvão e catálise, bem como na identificação de frações separadas por cromatografia descrita por Boroumand *et alii* [130].

Olinger e Griffiths [124, 125] ressaltam ainda, que a radiação não interage da mesma forma, podendo percorrer diferentes caminhos através da amostra, como resultado de diferentes profundidades de penetração, de diferentes números de choques e de um diferente número de espalhamento nas interfaces.

Uma variedade de parâmetros afetam o formato da banda, a posição, a intensidade em ambos os espectros de reflexão difusa (infravermelho próximo e infravermelho médio). Entre eles destacam-se: geometria ótica, absorção pela matriz, índice de refração, absorvidade, empacotamento da amostra, morfologia (tamanho da partícula), concentração do analito.

A principal causa de desvio da linearidade da intensidade da banda da reflexão difusa versus concentração é a radiação refletida da superfície das partículas da amostra, chamada reflexão especular (ou reflexão de Fresnell) [125].

Na ausência de absorção e com espessura infinita e múltiplos espalhamentos, um máximo de radiação retorna superfície. Quando ocorre absorção, a radiação é transmitida e espalhada nas interfaces no interior da amostra, antes de retornar a superfície.

O tamanho a forma das partículas, o vazio entre elas e a compactação afetam a quantidade (concentração) do material através do qual a radiação é transmitida.

Alguns desses fatores também afetam o caminho entre as interfaces individuais. E a morfologia da superfície da amostra é um consideração importante na magnitude do componente especular [123].

A atenuação da parte da reflexão difusa pela absorção no meio é dada pela lei Lambert-Bourguer, conforme expressão 4.10.

$$I = I_0 \cdot e^{-\varepsilon \cdot \delta} \quad (4.10)$$

onde I representa a intensidade da radiação, I_0 corresponde a intensidade da radiação incidente, ε é o coeficiente de extinção molar da amostra e δ é a espessura média da camada.

A profundidade de penetração da radiação, δ , em um meio absorvente depende de vários fatores. Para uma dada matriz e concentração do analito, o tamanho da partícula do analito e da matriz diluente governam a profundidade de penetração efetiva [131, 132].

Uma das conseqüências do tamanho das partículas é relatado por TeVrucht e Griffiths [131], ressaltando que a razão entre as bandas fracas e fortes não permanece constante com a alteração do tamanho das partículas.

Uma aproximação para descrever o comportamento da luz num meio difuso é dado pela solução de equações diferenciais de primeira ordem, as quais são descritas pela diminuição da intensidade da radiação no interior da amostra devido ao espalhamento e a absorção.

Kubelka e Munk, desenvolveram uma teoria fenomenológica em 1931, considerando que uma camada é composta de partículas absorventes e

espalhativas uniformemente e aleatoriamente distribuídas, cuja dimensão das partículas é muito menor que a espessura da camada [120].

Eles consideram somente a direção da radiação incidente e refletida perpendicular a superfície da camada, assumindo que a radiação é monocromática e que a espessura é infinita (o aumento da espessura não resulta em diferença na reflexão).

A resolução das equações diferenciais obtidas pela concepção desta teoria [120], resulta na expressão 4.11 conhecida como equação de Kubelka-Munk.

$$F(R_{\infty}) = \frac{(1 - R_{\infty})^2}{2R_{\infty}} = \frac{k}{s} \quad (4.11)$$

onde R_{∞} é reflexão da amostra para profundidade infinita, k é o coeficiente de absorção da amostra (que é proporcional a concentração) e s é o coeficiente de espalhamento da amostra.

O coeficiente de espalhamento determina a extensão da interação da luz incidente com a amostra antes da radiação retornar a superfície. Para um certa extensão, s controla a profundidade na qual a luz penetra na amostra. Quanto menor valor de s , maior é o valor $F(R_{\infty})$.

Entretanto, existe um limite para este comportamento, sendo somente válido para as partículas que efetivamente espalham a radiação incidente [133-134].

Como consequência, o efeito do tamanho da partícula desloca a linha de base, comportamento que se torna muito pronunciado em comprimentos de onda de grande absorção pela amostra. Por exemplo, em duas amostras com mesma composição, porém com diferente granulometria, há maior re-

flexão das partículas menores (mais finas), e os efeitos da reflexão especular são minimizados em uma superfície composta por partículas de menor tamanho.

A função de Kubelka-Munk pode relacionar os espectros de reflexão difusa com a concentração da amostra, transformando o espectro de reflexão em formato semelhante a um espectro de absorbância.

Por esta razão, a função de Kubelka-Munk é conhecida como a Lei de Beer da espectroscopia de reflexão.

A correta quantificação de analitos em amostras em pó é dependente do efetivo poder de penetração, sendo grande o suficiente para fornecer um espectro representativo da amostra por inteiro. A equação de Kubelka-Munk é mais simples e é geralmente a forma mais usada para definir uma afinidade linear entre intensidade da banda e a concentração da amostra na espectroscopia por reflexão difusa [126].

O rápido aumento na utilização da espectroscopia no infravermelho por reflexão difusa surgiu devido a facilidade em se obter espectros de pós, sólidos e espécies adsorvidas em sólidos.

Também novos métodos de amostragem têm introduzido abrasivos para coletar espectros de amostras que não são pós, aumentando a utilidade da técnica sem a necessidade de extensiva preparação da amostra, podendo-se obter informações de materiais na forma mais próximo da original [121].

Vários acessórios comerciais tem sido empregados para obter os espectros DRIFTS, onde a radiação espalhada é coletada em montagem óptica própria e é dirigida para o detetor, conforme ilustra a figura 4.5 [131].

Estudos comparando diferentes geometrias óticas para os acessórios de reflexão difusa também foram estudados por Yang *et alii* [135]. Um dos

principais critérios é a capacidade do esquema óptico proposto minimizar a importância da componente especular no espectro final.

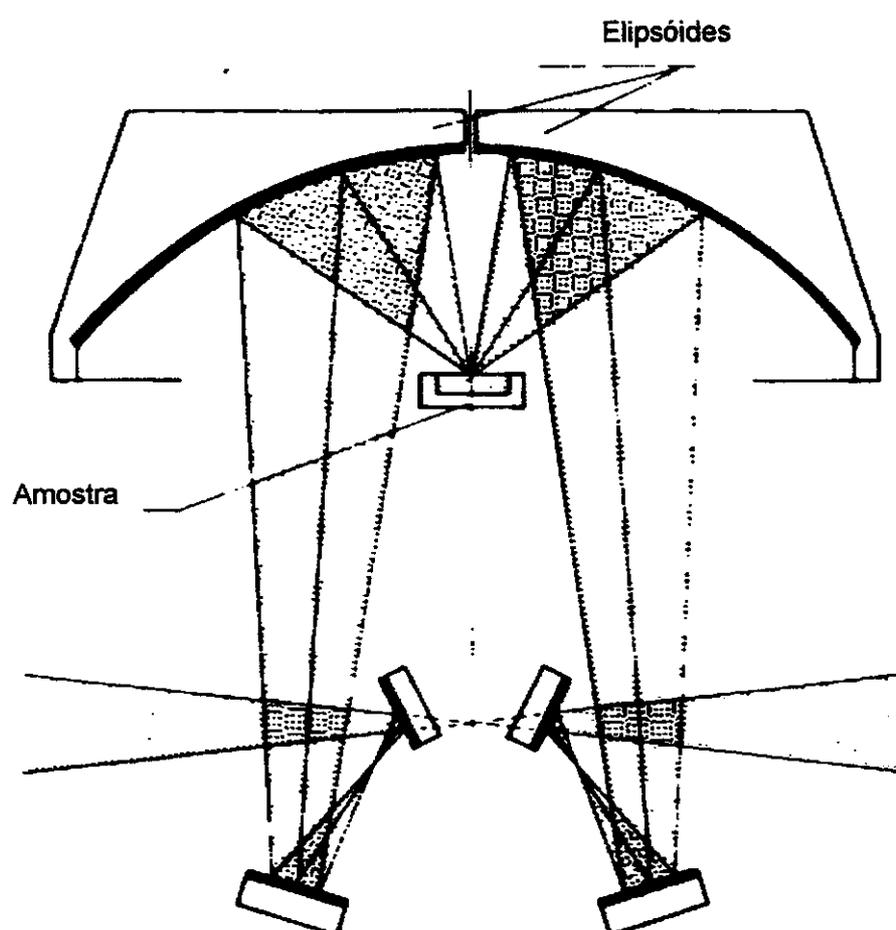


Figura 4.5 – Esquema comum para a obtenção de espectros DRIFTS .

O potencial da técnica DRIFTS foi reconhecido como um novo método de amostragem para aplicações quantitativas. Sua aplicação requer um entendimento teórico das limitações da técnica para obtenção de espectros de boa qualidade. Resultados quantitativos podem ser obtidos, se for feita uma preparação adequada da amostra [121].

Várias aplicações têm sido descritas na literatura para a análise de alimentos, como àquelas que já foram apresentadas, em parágrafos anteriores [9, 11, 14, 17, 18, 21, 28, 32-34, 36, 38, 48, 78-83, 86, 102, 109, 110].

Outras aplicações envolvem a determinação de umidade em amostras de milho realizada por Stermer *et alii* [136], a investigação da textura em grãos de trigo realizada por Osborne *et alii* [137], a determinação de amino-ácidos em amostras de trigo introduzida por Williams *et alii* [138], a caracterização qualitativa da pureza de amostras de farinha de trigo empregando espectroscopia no infravermelho médio estudada por Renard *et alii* [139].

Também foram estudadas, por espectroscopia de reflexão difusa, a determinação de fibra em farinha de trigo cujos resultados foram apresentados por Horváth *et alii* [140].

Recentemente, as técnicas de reflexão difusa foram empregadas na classificação de farinhas de trigo, publicados por Delwiche e Norris [141] e por Delwiche e Massie [142]. Hareland [143] emprega a reflexão difusa para estudar a distribuição do tamanho de partícula em amostras de farinha.

Outra aplicação apresentada em 1999, investiga as propriedades radioativas em materiais granulados [144]. Neste mesmo ano, Reeves III e Zapf [145] empregam pesquisa com bibliotecas de espectros no infravermelho próximo e médio para classificar amostras em pó de ingredientes alimentícios.

4.4 Reflexão Interna Atenuada

A reflexão total atenuada (ATR) é uma técnica de amostragem rápida que requer uma mínima preparação da amostra e, principalmente, que apresenta fácil obtenção de espectros de materiais espessos e fortemente absorventes, difíceis de serem analisados por espectroscopia por transmissão [124].

Existem múltiplos elementos de ATR e uma variedade de acessórios desenvolvidos para obter espectros de materiais líquidos, sólidos e viscosos para larga faixa de tipos de amostras, incluindo filmes, resíduos, papel, revestimentos sobre papel, pós, tintas, tecidos, espumas, minerais, vidros etc. Alguns dos mais recentes acessórios permitem o monitoramento de fluxo [146].

A técnica ATR se caracteriza pelas múltiplas reflexões da radiação infravermelha que ocorrem no interior de cristais, de materiais com alto índice de refração (como por exemplo ZnSe), interagindo apenas com a amostra que estiver superficialmente no cristal.

A figura 4.6 ilustra a reflexão total atenuada através do emprego de um elemento de ATR. Este cristal deve ser composto de um material com alto índice de refração, para que somente uma pequena parte do feixe de radiação incidente seja refletido ao atingir o cristal.

Após o feixe incidente sofrer a difração penetrando no meio n_2 , a geometria do cristal deve promover a reflexão e para tanto $\theta < \text{sen}^{-1}(n_3/n_2)$. Esta reflexão interna cria uma onda evanescente a qual se estende na amostra em contato com o cristal. Na região do infravermelho onde a amostra absorve energia, a onda evanescente será atenuada.

Num acessório de ATR tem-se um elemento de material adequado, por exemplo de ZnSe, que deverá produzir múltiplas reflexões.

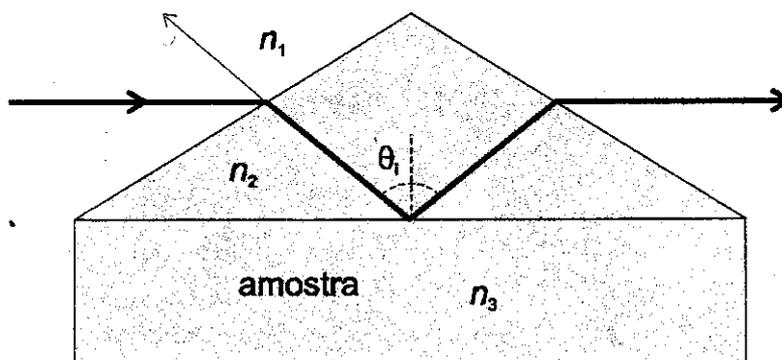


Figura 4.6 - Esquema da reflexão interna em um elemento de ATR.

Os princípios teóricos básicos da espectroscopia de reflexão interna (IRS), foram publicados no início dos anos 60 por Harrick [147] e Fahrenfort [148], que descrevem detalhes da teoria e apresentam dados experimentais.

O fenômeno da reflexão interna é observado sob certas condições. Quando a radiação entra em um prisma feito com um índice de refração alto em relação ao meio externo (cristal de ATR) a radiação será refletida totalmente internamente. Para tanto, considere a figura 4.7, na qual um feixe de radiação encontra-se no interior do elemento de ATR.

Para um índice de refração tal que $n_2 > n_3$, aplicando-se a lei de Snell o ângulo de refração é imaginário para valores do ângulo de incidência tal que satisfaça a expressão 4.12.

$$\theta_i > \text{sen}^{-1}(n_3/n_2) \quad (4.12)$$

O ângulo acima do qual esta onda refratada deixa de ser real é denominado de ângulo crítico (θ_c) e é expresso conforme equação 4.13.

$$\theta_c = \text{sen}^{-1}\left(\frac{n_3}{n_2}\right) \quad (4.13)$$

Nestas condições o feixe é totalmente refletido na interface n_3/n_2 .

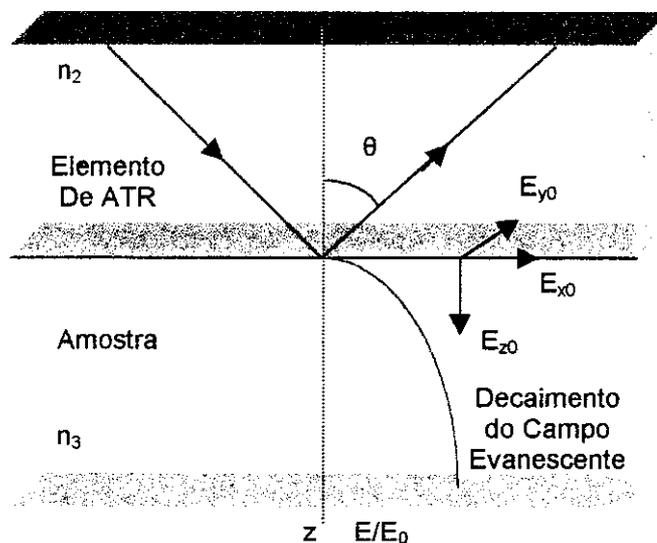


Figura 4.7 - Representação da propagação da radiação infravermelha através do elemento de ATR.

As propriedades do campo evanescente podem ser visualizadas na figura 4.7, onde a propagação da radiação infravermelha do meio 2, com índice de refração n_2 , sofre reflexão interna na interface com o meio 3, com índice de refração n_3 , quando o ângulo de incidência excede o ângulo crítico.

Esse fenômeno de reflexão interna total é mais facilmente descrito para uma onda plana infinita de uma interface entre um meio não absorvente semi-infinito [146].

A propagação do campo evanescente no meio 3 ocorre sob certas condições, sendo elas [146] :

- a intensidade do campo no meio 3 não é zero, mas há um componente normal instantâneo de fluxo de energia no meio 3 cujo termo médio é zero. Assim, não há diminuição da energia e a propagação da radiação no meio 2 é totalmente refletida no interior do elemento de ATR.

- o campo evanescente no meio 3 é uma onda não transversa e tem componentes em todas orientações espaciais.
- o campo evanescente é confinado a vizinhança da superfície do meio 3, e diminui em intensidade com a distância no meio 3 normal a superfície (ao longo do eixo z na figura 4.7).
- há um fluxo de energia diferente de zero paralelo a superfície, ao longo do eixo x na figura 4.7, resultando num deslocamento da onda incidente e refletida.

A profundidade de penetração da radiação no segundo meio, pode ser quantitativamente definida em termos da distância requerida para que a intensidade do campo elétrico decaia de um fator de $1/e$ do valor inicial [149], podendo ser calculada pela equação 4.14.

$$d_p = \frac{\lambda}{2\pi n_1 \left(\sin^2 \theta_i - \left(\frac{n_3}{n_2} \right)^2 \right)^{1/2}} \quad (4.14)$$

Onde θ_i é o ângulo de incidência, n_2 é o índice de refração do elemento de ATR e n_3 o índice de refração da amostra.

Esta equação mostra que a profundidade de penetração depende do comprimento de onda da radiação no meio com índice de refração mais alto (elemento de ATR).

A radiação atinge a amostra penetra poucos microns de profundidade. Desta forma, qualquer material que esteja em contato com o cristal de ATR pode absorver a radiação incidente atenuando sua intensidade, dando origem ao espectro infravermelho. Em função disto, a composição da superfície pode ser estudada e o espectro torna-se independente da espessura da amostra.

Contudo, a profundidade de penetração é proporcional ao comprimento de onda e, portanto, é observado um aumento constante na profundidade de penetração na amostra, com a varredura do espectro de comprimento de onda maiores para menores. Como consequência, as intensidades relativas podem ser diferentes das observadas pelo emprego da espectroscopia por transmissão.

A profundidade de penetração também varia com o índice de refração da amostra. Esse efeito é menos pronunciado quando o índice de refração do cristal de ATR é maior.

A escolha do cristal pode resultar em distorções da banda no espectro. Quando o ângulo de incidência for muito maior que o ângulo crítico, estas distorções resultantes no espectro são minimizadas.

Não existe nenhum material universal para ATR. Um dos materiais mais comumente usados é o ZnSe. O seu índice de refração é 2,4 e sua faixa de utilização é de 20.000 até 650 cm^{-1} , sendo recomendado para medidas em soluções aquosa por ser insolúvel em água .

Outra opção é o silício (Si), indicado quando as amostras são altamente espalhativas. Como os cristais de silício tem alto índice de refração (3,5), a profundidade de penetração é menor. A faixa de utilização do Si é de 9,000 até 400 cm^{-1} , sendo relativamente inerte e também útil em temperaturas elevadas (300°C) [121].

Dois parâmetros controláveis são o ângulo de incidência e a escolha do elemento de ATR. A espessura efetiva e a profundidade de penetração aumentam, quando o ângulo aproxima-se do ângulo crítico, podendo-se variar a profundidade de penetração ajustando o ângulo de incidência. Contudo o número de reflexões também varia com o ângulo de incidência.

Um problema geralmente encontrado, é a dificuldade de obter-se boa reprodutibilidade no contato da amostra com o elemento de ATR. Esse efeito é observado na variação da intensidade das bandas com a pressão aplicada. Aumentando a pressão, a eficiência de contato é aumentada e, conseqüentemente, as intensidades das bandas aumentam.

A área de contato entre o cristal e a amostra é um fator que também influencia na intensidade das bandas, para uma boa reprodutibilidade. Para medidas quantitativas, deve-se colocar toda a área do cristal em contato com a amostra. Irregularidades na superfície da amostra vão dificultar um contato mais efetivo [121,146].

Por outro lado, este efeito pode ser útil, por dar informações sobre as propriedades ou condições da superfície de um material, incluindo identificação, modificações e adsorção superficiais [121].

A espectroscopia por reflexão total atenuada foi amplamente empregada na década de 90. Inicialmente os trabalhos se concentraram na determinação de açúcares [150-157].

Heise *et alii* [150], Ward *et alii* [151] e Budínová *et alii* [157] centraram o emprego da espectroscopia ATR na determinação de glicose em sangue e plasma sangüíneo.

Dupuy *et alii* [152] propõem um controle de qualidade laboratorial para a análise de açúcares em pó. Neste mesmo período empregam a técnica de ATR para a análise quantitativa de açúcares em xaropes a base de glicose [153].

Os estudos com açúcares culminaram com o trabalho de Kemsey *et alii* [154], que empregam fibras óticas em determinações quantitativas, e com os resultados de Bellon-Maurel *et alii* [155], que estudam a decomposição do amido avaliando simultaneamente os açúcares presentes.

Em 1997, Cadet *et alii* [156], empregam a técnica de ATR na região do infravermelho médio para determinar açúcares em amostras biológicas, como sucos de frutas.

Paralelamente, um outro nicho para o emprego da espectroscopia por ATR, foi caracterizado pelo emprego desta técnica na determinação quantitativa e qualitativa de óleos e gorduras de origem vegetal e animal.

Van de Voort *et alii* empregam espectroscopia por reflexão total atenuada na determinação do número de iodo e do grau de saponificação de óleos e gorduras [158], no controle de qualidade de manteiga [159], na quantificação dos ácidos livres de óleos e gorduras [160] e na quantificação de gordura e umidade em produtos com alta quantidade de gorduras [161].

A técnica de ATR, também foi empregada no monitoramento do processo de oxidação de óleos vegetais [162], na caracterização de óleos, manteigas e margarinas [163] e na autenticação de óleos vegetais [164], em especial do azeite de oliva [165].

O grupo de Van de Voort [166] também desenvolveu métodos para o controle de qualidade de leite condensado visando principalmente a determinação de gorduras e sólidos.

Outras aplicações da espectroscopia por ATR envolvem a análise de látex por Dupuy *et alii* [167], o estudo de amostras de polímeros em geral [168-169], bem como no monitoramento de processos de polimerização [170].

O potencial da técnica de ATR no infravermelho médio foi ainda demonstrado por Rossi *et alii* [171] como uma técnica analítica para a determinação estrutural das moléculas de água em soluções aquosas, assim como por Schönherr *et alii* [172], para a análise quantitativa de misturas solventes.

Um único estudo relacionado à utilização da espectroscopia por ATR com farinha de trigo é descrito na literatura por Popineau *et alii* [173], onde a técnica de ATR é empregada para identificar glúten em soluções aquosas.

Técnicas de Análise Multivariada

5.1 Introdução

A Quimiometria pode ser entendida como uma área multidisciplinar do conhecimento, onde são aplicadas ferramentas matemáticas e estatísticas à análise de dados relativos aos processos químicos, que apresentam natureza multivariada, como por exemplo as informações analíticas disponibilizadas na áreas da espectroscopia e da cromatografia.

As técnicas quimiométricas visam o delineamento e análise de experimentos, tendo como objetivo utilizar de forma eficiente, correta e econômica os recursos disponíveis.

Um grande problema, muitas vezes enfrentado pelos químicos, é a necessidade de conhecer as propriedades e a composição de sistemas, que apresentam difícil solução pelos métodos convencionais de laboratório. Todavia, estas propriedades podem ser obtidas de forma indireta, a partir de outras variáveis descritas como variáveis latentes. Desta forma, um dos objetivos dos métodos quimiométricos é o de encontrar as variáveis latentes, bem como as relações existentes entre os dados físico-químicos e o sistema em questão [174].

As análises qualitativas e quantitativas, utilizando espectroscopia na região do infravermelho, expandiram-se a partir do momento em que os dados gerados por um espectrofotômetro FT-IR puderam ser digitalizados, habilitando os métodos estatísticos na resolução de problemas de análise química.

Inicialmente muitas amostras não comportavam o isolamento de uma banda para análise, o que tornava necessário o uso de métodos de separação química tal como a cromatografia. Porém, a possibilidade de utilizar várias frequências do espectro, tornou crescente o tipo de amostras que podem ser analisadas por espectroscopia no infravermelho.

A análise dos dados obtidos por técnicas espectroscópicas, podem ser realizadas por métodos multivariados, que permitem o estudo de várias espécies presentes, sem importar a existência de variações marcantes nos espectros, ou mesmo a presença de alta correlação entre os mesmos [175].

Métodos de análise multivariada para o controle de qualidade de produtos alimentícios têm sido apresentados na literatura [176-189]. Em particular o emprego da espectroscopia no infravermelho próximo, tem apresentado resultados satisfatórios no controle de adulterações via análise discriminante [177, 179, 180], ou na substituição de métodos tradicionais para a quantificação de determinado parâmetro de qualidade via métodos de regressão multivariada [184-186].

5.2 Análise por Componentes Principais

A análise vetorial, que permite descrever a variação (ou dispersão) de um determinado conjunto de dados, é conhecida como análise por componentes principais ou simplesmente como análise de fatores. Entre as primeiras aplicações destes métodos destacam-se os controles estatísticos de processos, que usualmente necessitam avaliar diferentes variáveis de processo, simultaneamente [190-192].

A análise por componentes principais (PCA) consiste num método matemático que reduz o número de variáveis originais, gerando uma nova representação destas variáveis (espectros) através de variáveis latentes (ou componentes principais) que são ortogonais entre si [193-195].

A análise de fatores é aplicada numa matriz de dados analíticos, que relaciona um conjunto de variáveis (os espectros) a diversos experimentos (as amostras). Em geral, a matriz de dados X é centrada na média, podendo também os dados sofrerem outros tipos de tratamento como por exemplo os escalamentos.

Antes de submeter os dados a modelagem, alguns autores sugerem que o conjunto de dados seja ajustado de forma a facilitar as operações posteriores [196-197]. Uma forma de melhorar a manipulação destes dados é conhecida como centrar na média, que consiste em subtrair os valores médios calculados para cada variável (colunas da matrix X) de cada uma das variáveis correspondentes.

No PCA, a matriz X , após ter seus dados centrados na média, é decomposta no produto de três matrizes, através do algoritmo de decomposição de valor singular (SVD)[198], resultando na expressão 5.1.

$$X = U \cdot S \cdot V^t \quad (5.1)$$

A matriz V é denominada de matriz dos pesos (*loadings*), as colunas da matriz V correspondem aos autovetores, e S é uma matriz diagonal que indica o quanto de variância é descrita por cada um dos vetores bases correspondentes e que contém a raiz quadrada dos autovalores. As matrizes U e V são ortogonais entre si, e o produto entre as matrizes U e S é representado por T , sendo denominado de escores (*scores*). O algoritmo SVD ordena cada uma destas informações, de forma a colocar em ordem decrescente de variância explicada, cada uma das dimensões (componentes principais) originadas.

Como resultado da análise de componentes principais, o conjunto de dados originais é agrupado em função da correlação existente entre as variáveis, gerando um novo conjunto de eixos (componentes principais) ortogonais entre si e que, geralmente, apresentam uma forma mais simplificada para a visualização dos dados. Na figura 5.1 podem ser observadas algumas das potencialidades dessa compressão dos dados espectroscópicos.

O método de análise por componentes principais, foi um dos primeiros recursos quimiométricos a ser amplamente empregado pelo químicos. Apesar de, nos últimos anos, terem sido apresentados na literatura uma gama bastante significativa de novas ferramentas quimiométricas, o PCA continua sendo cada vez mais empregado.

A tabela 5.1 resgata o número de referências bibliográficas que apresentam em seu resumo, título, ou palavras-chave, o termo PCA, nos 5 últimos anos, consultadas todas as bases de dados do *Institute for Scientific Information*.

Tabela 5.1 – *Freqüência do termo PCA nas literaturas indexadas para os últimos 5 anos.*

Período	Freqüência
1996	605
1997	676
1998	700
1999	783
2000	289
1996-2000	3053

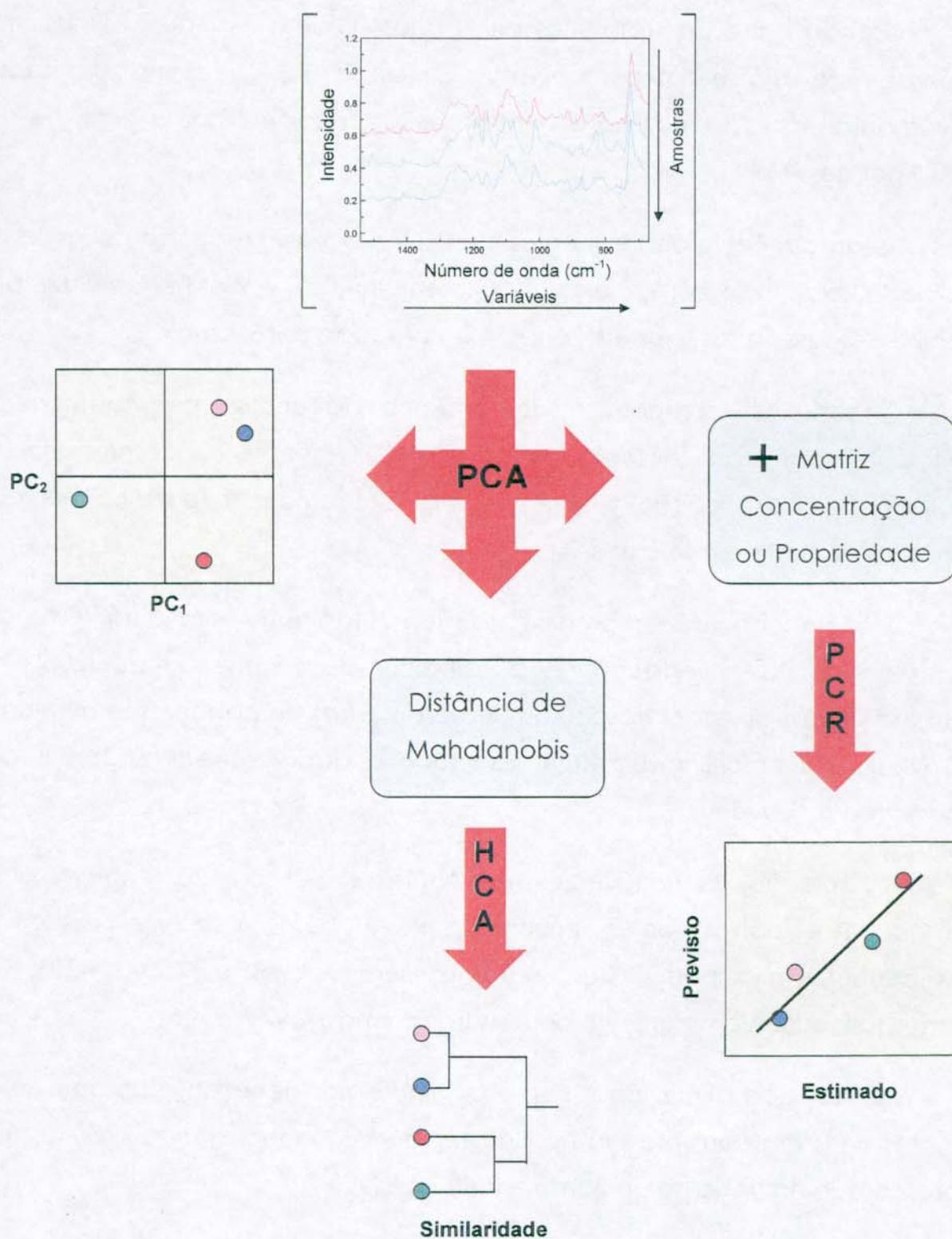


Figura 5.1 – Esquema das potencialidades da aplicação da análise por componentes principais.

Grande parte dos trabalhos publicados durante os cinco últimos anos, empregando análise por componentes principais (PCA), utilizam informações instrumentais das técnicas de espectroscopia vibracional como base de suas investigações [199].

As aplicações mais freqüentes do PCA envolvem redução da dimensionalidade, explorado por Wu *et alii* [200], Kemsley [201] e Wold *et alii* [202], e na análise de agregados (clusters), onde é empregado por Szczubialka *et alii* [203].

Cadet [204] compara o desempenho da análise por componentes principais, com a aplicação de polinômios de Legendre no desempenho de métodos de quantificação, empregando espectros no infravermelho médio de amostras de matrizes biológicas.

A detecção de anomalias (*outliers*), também tem sido estudada empregando PCA. Destacam-se os trabalhos de Centner *et alii* [205], que estudam dados de espectroscopia NIR de amostras de polímeros, e de Stork *et alii* [206], que empregam o PCA associado a dados espectroscópicos com transformada *wavelet*.

Na seleção de variáveis, Verdú-Andrés e Massart [207] comparam o método por componentes principais (PCA) com o de regressão por componentes principais (PCR), enquanto que Walczak e Massart [208, 209] empregam o PCA em conjunto com as transformadas *wavelet*.

Métodos de otimização, como os algoritmos genéticos, também foram empregados em conjunto com a análise de componentes principais, como é apresentado no trabalho de Barros e Rutledge [210].

Verdú-Andrés *et alii* [211], também propõem a correção da falta de linearidade, presente em conjuntos de dados espectroscópicos, empregando componentes principais para selecionar as melhores variáveis latentes.

A análise por componentes principais também teve sucesso junto aos dados espectroscópicos por emissão Raman, conforme apresentados nos trabalhos de Stellman *et alii* [212], de Hayden e Morris [213] e de Andrew e Hancewicz [214].

Armanino e Festa [215] empregam o método PCA para classificar o trigo a partir dos dados de parâmetros analíticos como o teor de cinza, umidade, gordura e proteína.

Associada a espectroscopia no infravermelho médio (MID), a técnica PCA tem sido empregada por Dupuy *et alii* [216] e por Dahlberg *et alii* [217] para classificação de óleos e gorduras.

Em 1997, Stallard [218] emprega a análise por componentes principais, paralelamente às técnicas de análise discriminante linear (LDA) e de regressão por mínimos quadrados parciais (PLS), para comparar o desempenho das técnicas de espectroscopia no infravermelho próximo e médio na discriminação de distintas classes de compostos orgânicos.

A espectroscopia por reflexão no infravermelho próximo (NIRR), também foi utilizada em conjunto com o método PCA por Bacci *et alii* [219] para discriminar amostras de calcários provenientes de distintas transformações, por Fischer e Eichhorn [220] no monitoramento *on-line* e *off-line* do processo de extrusão de polímeros, e por Blanco *et alii* [221] visando o controle de qualidade de produtos farmacêuticos.

Recentemente, Ferreira *et alii* tem apresentado excelentes resultados no emprego da análise por componentes principais, em conjunto com a análise hierárquica de agrupamentos (HCA), para caracterizar o perfil oceanográfico do litoral norte paulista [222] e para relacionar o conteúdo de ácidos graxos e minerais de diferentes produtos alimentícios embutidos [223].

5.3 Métodos de Regressão Multivariada

A calibração ou regressão é uma das mais importantes ferramentas empregadas na análise quantitativa, em particular na análise instrumental. Os procedimentos matemáticos utilizados tem como objetivo descrever as relações quantitativas existentes entre duas ou mais variáveis.

O processo de calibração consiste de duas etapas [224]: a primeira, dita descritiva, emprega um conjunto de medidas realizadas para uma série de amostras (os padrões), cuja concentração é conhecida, para construir um modelo que relacione a grandeza medida com a concentração da espécie de interesse; e uma segunda, preditiva, que utiliza-se deste modelo para prever as concentrações de novas amostras, com base nas medidas analíticas realizadas para cada uma destas amostras.

Os métodos de regressão multivariada são assim chamados, pois permitem manipular dados de um conjunto de medidas analíticas, geralmente sinais instrumentais, associados à uma ou mais variáveis de interesse ao mesmo tempo.

Os métodos de calibração multivariada tem sido usados, principalmente, quando as espécies químicas presentes em uma mistura precisam ser determinadas. Porém a informação analítica disponível não apresenta seletividade, ou seja, em uma mistura não é possível identificar cada espécie separadamente a partir da resposta instrumental [225].

Esses métodos têm, recentemente, tornado possível estimar propriedades químicas e físicas de diversas matrizes, a partir de seus espectros no infravermelho.

O fundamento destes métodos é então, estabelecer uma relação entre duas matrizes (ou blocos) de dados químicos e/ou físicos, quando existir uma dependência (ou correlação) entre as propriedades que descrevem cada uma destas matrizes de dados.

Em espectroscopia, os métodos de calibração multivariada podem ser usados, mesmo quando há sobreposição de informações espectrais de vários componentes, sobre todas as regiões espectrais medidas.

Ao contrário dos métodos de calibração univariada, técnicas multivariadas obtêm maior precisão a partir de informações redundantes no espectro, podendo minimizar variações da linha de base e/ou detectar pontos divergentes (*outliers*).

Nestes métodos a sensibilidade da detecção de *outliers* é geralmente aumentada pelo acréscimo do número de frequências usadas na análise. O número de impurezas, que pode ser avaliado pela análise, é também maior pelo aumento do número de frequências empregadas [225].

A concepção da aplicação da regressão multivariada pode ser observada na figura 5.2, onde são apresentados três dos principais métodos de regressão multivariada, mais freqüentemente utilizados: o método de regressão linear múltipla (MLR), o método de regressão por componentes principais (PCR) e o método por mínimos quadrados parciais (PLS).

Os métodos de regressão por componentes principais (PCR) e por mínimos quadrados parciais (PLS) são dois métodos baseados em análise fatorial [226-228]. O processo de análise fatorial é usado para decompor a matriz dos dados espectrais, em um produto de duas matrizes menores, com o objetivo de simplificar (compactar) a representação do conjunto de dados.

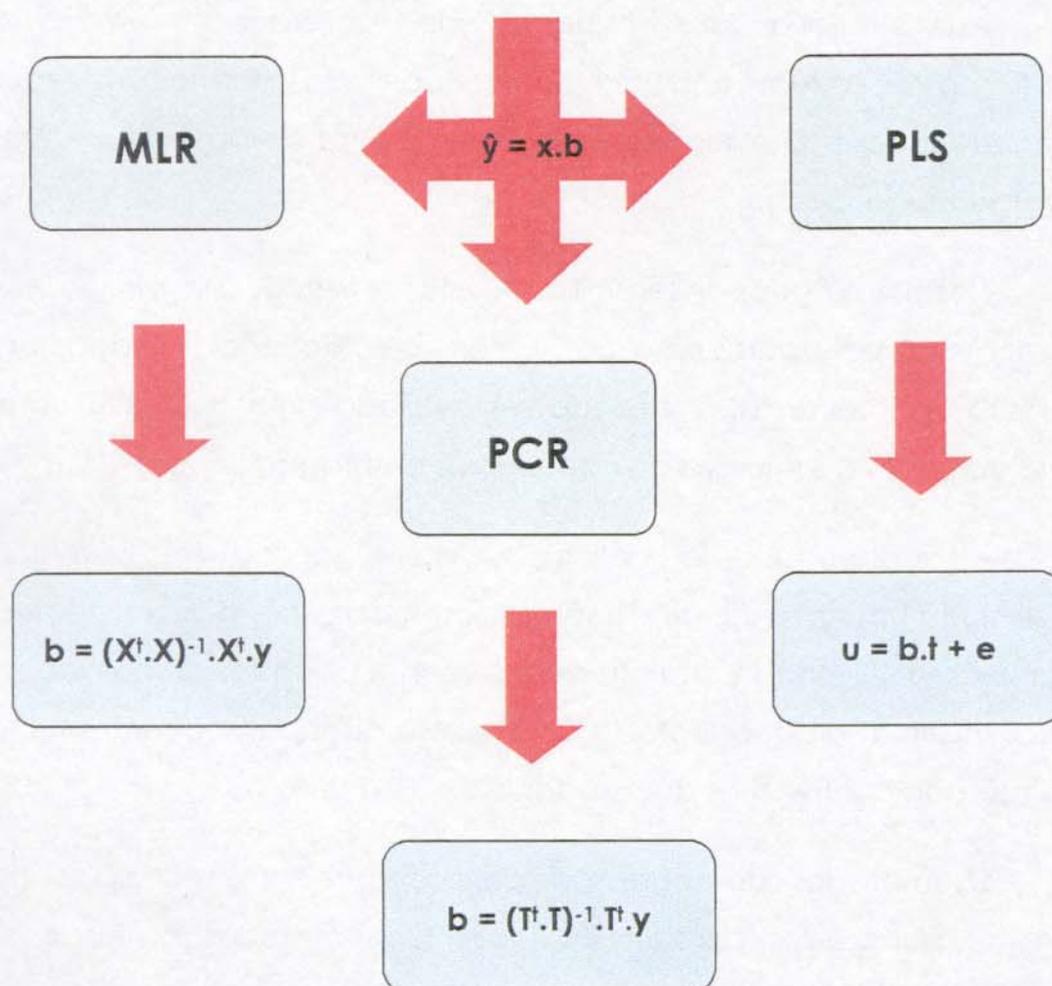
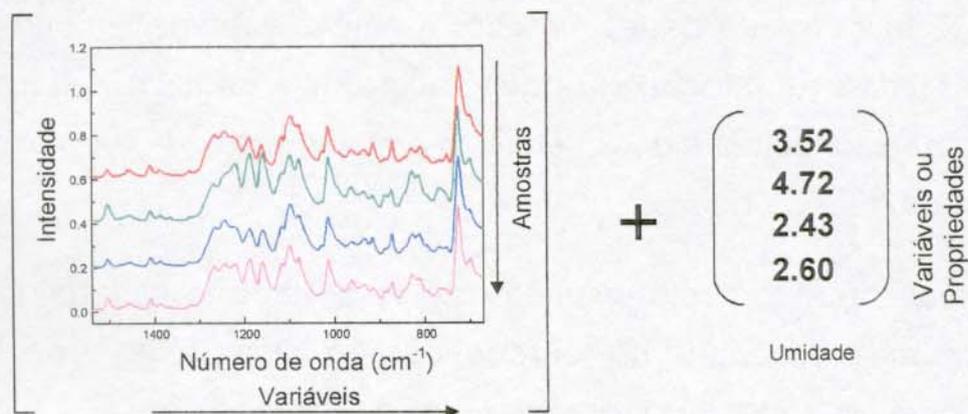


Figura 5.2 – Esquema da calibração multivariada empregando os algoritmos de regressão linear múltipla (MLR), regressão por componentes principais (PCR) e regressão por mínimos quadrados parciais (PLS).

5.3.1 Regressão Linear Múltipla (MLR)

O método da regressão linear múltipla busca a combinação linear entre as variáveis independentes (espectros ou medidas instrumentais) e as variáveis dependentes (propriedades), de forma que seja minimizado o erro na estimativa destas propriedades [195].

No algoritmo da regressão linear múltipla a solução de sistemas lineares é obtida segundo as seguintes condições, considerando n como o número de amostras e v o número de variáveis:

- Se $n < v$, o sistema não tem solução, uma vez que ele apresenta mais variáveis que amostras;
- Se $n = v$, existe uma única solução para o sistema;
- Se $n > v$, o sistema não tem uma única solução, pois ele tem mais amostras que variáveis.

Quando formulamos um problema para ser resolvido por regressão linear múltipla, onde trabalhamos com algum dos métodos espectroscópicos para gerar as variáveis independentes, temos que planejar tal método de forma que exista uma solução para o sistema, e assim possamos encontrar uma relação satisfatória para a predição de novas amostras.

De fato, a relação resultante do planejamento adequado para este tipo de modelagem, resulta na equação 5.2, que geralmente é apresentada em sua forma mais geral (expressão 5.3).

$$\hat{y} = x_1.b_1 + x_2.b_2 + x_3.b_3 + \dots + x_v.b_v + e \quad (5.2)$$

$$\hat{y} = \mathbf{x}.\mathbf{b} + \mathbf{e} \quad (5.3)$$

onde \hat{y} é a estimativa da propriedade de interesse (variável independente), x representa o vetor das medidas instrumentais (variáveis independentes), para medidas no infravermelho este vetor corresponde aos valores de absorvância medidos a cada frequência (ou comprimento de onda), b refere-se ao vetor dos coeficientes de regressão, e e é vetor dos resíduos da modelagem.

O vetor dos coeficientes de regressão b , para o método MLR, é calculado utilizando-se o critério dos mínimos quadrados, resultando na expressão apresentada na figura 5.2.

Como limitação para a aplicação da regressão linear múltipla encontramos a determinação da matriz inversa, que pode ser de difícil determinação, quando as variáveis independentes estiverem muito correlacionadas.

O método de regressão linear múltipla, já foi soberano quando do início da aplicação da regressão multivariada aos problemas da química analítica. Mais precisamente, entre os anos 60 a 80, e particularmente pela simplicidade na obtenção da solução do sistemas lineares, possíveis de serem planejados com este método.

Esta realidade vem mudando década a década, em virtude da expansão da microinformática, aliada à evolução nos sistemas de análise instrumental, permitindo que outros modelos de regressão possam ser facilmente implementados.

Poder-se-ia pensar então, que o MLR está com seus dias contatos. Mas não é isso que, aparentemente, os quimiométricos vem demonstrando pensar. De fato, poucos são os trabalhos publicados que empregam a técnica de regressão linear múltipla na sua forma mais crua, a exemplo do trabalho publicado por Hall *et alii* [229], onde a técnica MLR é comparada ao PLS no monitoramento de processos biofermentativos.

A grande maioria das publicações destes últimos 5 anos, tem utilizado as facilidades matemáticas do MLR associadas a métodos de otimização e seleção de variáveis, como nos trabalhos reportados por Massart *et alii* [230-232], que estabelecem um novo conceito de simbiose entre o MLR e os algoritmos genéticos.

Mesmos as revisões mais recentes sobre o panorama da regressão multivariada, por vezes descrevem ou apenas citam a técnica de regressão linear múltipla, como uma ferramenta potencial, como pode ser constatado nas publicações de Chamidade *et alii* [233], de Weidemann *et alii* [234] e de Scarminio *et alii* [235].

Neste panorama, ainda foi recentemente publicado o trabalho realizado por Ge *et alii* [236], onde o MLR é empregado no reconhecimento de padrões, ao lado da análise discriminante linear e das redes neurais.

No estudo do trigo, Delwiche e Massie [142], empregaram as técnicas de regressão linear múltipla (MLR) e por mínimos quadrados parciais (PLS), na identificação das classes de trigo, utilizando dados de espectroscopia no visível e por reflexão no infravermelho próximo.

Em 1998, Bruno-Soares *et alii* [54] empregam o regressão linear múltipla e a regressão por mínimos quadrados parciais para a previsão da composição química e nutricional de diversos cereais.

Neste mesmo período, Delwiche [237] realiza um dos mais amplos estudos sobre a quantificação de proteína em trigo, empregando a espectroscopia no infravermelho próximo, encontrando resultados comparáveis entre a regressão linear múltipla (MLR) e a regressão por mínimos quadrados parciais (PLS). Apesar de ter constatado que o PLS produzir valores de erro de previsão (SEP) menores que MLR, o pesquisador sugere que a técnica de MLR continue a ser preferencialmente utilizada, pois permite o emprego de instrumentos com

recursos mais simplificados, sem significativa perda da qualidade dos valores de previsão.

Em 1999, Reeves III [40] utiliza os métodos de regressão MLR, PCR e PLS para comparar o desempenho das regiões no infravermelho médio e próximo na análise de fibra e proteína de diversos produtos agrícolas.

5.3.2 Regressão por Componente Principais (PCR)

O algoritmo utilizado no método de regressão por componentes principais (PCR), emprega etapas semelhantes à análise por componentes principais (PCA), para decompor a matriz de dados X (de variáveis independentes) e então relacionar os resultados do PCA com os dados das propriedades de interesse (variáveis dependentes ou matriz Y).

Para os modelos que envolvem dados espectroscópicos, a matriz X é composta por valores de absorvância para medidas por transmissão, ou por $\text{Log}(1/R)$ para medidas de reflexão. A matriz X terá tantas linhas quanto for o número de amostras modeladas e tantas colunas quanto for o número de frequências (ou números de ondas) selecionados para o modelo.

A matriz Y , de variáveis dependentes terá o mesmo número de linhas que a matriz X , e seu número de colunas será respectivo aos número de parâmetros de interesse, que podem ser relativos às concentrações de diversas espécies químicas, assim como propriedades físico-químicas e físicas relacionadas com a informação química obtida nos espectros.

Este método de regressão tem por vantagem o fato de não ser sensível aos problemas induzidos pela correlação entre variáveis, que geralmente estão presentes em conjuntos de dados espectroscópicos.

Para um bom desempenho do método de regressão que emprega componentes principais (PCR), devem ser observadas as seguintes particularidades, no que se refere a escolha das variáveis independentes [238]:

- seleciona-se, sempre que possível, o menor número de variáveis. A introdução de variáveis desnecessárias no modelo provocará a redução da precisão dos valores durante a etapa de previsão;
- escolhe-se as variáveis independentes que não estejam altamente correlacionadas uma com as outras. A colinearidade pode causar instabilidade na determinação dos coeficientes de regressão.

O método de regressão por componentes principais (PCR) é um processo realizado em duas etapas. Primeiramente, a informação espectral (variáveis independentes) é utilizada para calcular o espectro dos componentes principais (variáveis latentes).

Então, os espectros dos componentes principais e as informações das propriedades (variáveis dependentes) são empregadas para gerar o modelo de calibração. O resultado deste processo é a equação linear 5.4, proveniente da regressão empregando os termos da expressão 5.1.

$$\hat{y} = t.b + e \quad (5.4)$$

onde **b** é o vetor dos coeficientes de regressão a serem determinados e **e** o vetor de erro.

Os valores referentes aos coeficientes de regressão são calculados pela modelagem das componentes principais (ou variáveis latentes) originadas via processo de compressão da matriz **X**, realizada através do algoritmo SVD. Este vetor de coeficientes é geralmente expresso pela equação ilustrada na figura 5.2, anteriormente apresentada.

Entre os trabalhos publicados nos últimos 5 anos, encontramos a aplicação do método PCR por Cadet [204], na quantificação de sacarose em amostras de açúcar proveniente de cana de açúcar, empregando dados de espectroscopia no infravermelho próximo.

Centner *et alii* [231], também se valem da regressão por componentes principais para propor um novo método de seleção de variáveis, baseado na eliminação das variáveis não informativas.

Ainda no ano de 1996, o método PCR foi empregado para na seleção do número mínimo de amostras para calibração por Ferré e Rius [239], com dados de espectroscopia nas regiões do UV-visível e infravermelho próximo, na seleção de comprimentos de onda por Xu e Schechter [240] para a análise multicomponente com medidas de fluorescência, no efeito das características da matriz para a determinação de açúcares por Reeves III [241] na região do infravermelho próximo, e na genérica avaliação da potencialidade desta técnica na análise multivariada de propriedades, empregando a região do infravermelho próximo estudada por Sun [242].

Na análise de corantes em alimentos, o método PCR foi utilizado com sucesso por Ni e Gong [188], ao lado dos métodos por mínimos quadrados parciais (PLS) e por mínimos quadrados clássicos (CLS), não sendo observada diferença significativa entre o PCR e o PLS.

Neste mesmo período, Lorber *et alii* [243], utilizam-se do PCR para estudar sinais de analitos puros via calibração multivariada, Díaz *et alii* [244] comparam o desempenho dos modelos de regressão CLS, PCR e PLS estudando misturas de pesticidas, enquanto que Xu e Schechter [245] buscam estabelecer um método de calibração livre da necessidade de seleção do número ótimo de fatores.

Xie e Kalivas [246, 247] estudam métodos para selecionar as componentes principais a serem utilizadas nos modelos de regressão PCR,

enquanto que Kalivas [248] faz uso dos métodos de regressão por componentes principais (PCR) e por mínimos quadrados parciais (PLS), para introduzir suas idéias com respeito a uma nova proposta para a regressão multivariada: o método de regressão do subespaço cíclico (CSR).

Ainda em 1998, Centner e Massart [249] comparam a eficiência do método de regressão por peso localizado (LWR) com os então clássicos métodos de regressão por mínimos quadrados parciais (PLS) e por componentes principais (PCR).

Recentemente, Faber [250] utiliza-se dos métodos de calibração, entre os quais o PCR, para interpretar a sensibilidade dos métodos de regressão multivariada que empregam dados de espectroscopia NIR, com relação aos pré-tratamentos dos espectros.

Novos modelos híbridos, onde os conceitos do PCR são empregados, foram apresentados por Goicoechea e Olivieri [251] onde a seleção de comprimentos de onda é realizada com a análise linear híbrida (HLA), e Burnham *et alii* [252] que introduzem a modelagem por variáveis latentes multivariada (LVMR).

Em 1999, o método de regressão por componentes principais (PCR) foi ainda empregado por Scarminio *et alii* [253] para estimar frequências vibracionais fundamentais no infravermelho e comparar com valores de parâmetros espectrais observados experimentalmente.

5.3.3 Regressão por Mínimos Quadrados Parciais (PLS)

Esta técnica foi desenvolvida na década de 70 por Herman Wold [254, 255]. Foi usada primeiramente na região do infravermelho próximo, onde é difícil designar bandas para componentes particulares. No modelo PLS, há algumas suposições feitas em torno dos dados que serão analisados. As informações espectrais e as informações das concentrações são usadas ao mesmo tempo na fase de calibração.

O alicerce do método dos mínimos quadrados parciais (PLS), reside na decomposição de uma matriz de dados X , em termos da soma das várias matrizes M , que apresentem dimensionalidade um, e que são acrescidas de uma matriz de resíduos (que corresponde a parte não modelada de X), conforme equação 5.5.

$$X = M_1 + M_2 + \dots + M_a + E \quad (5.5)$$

onde a corresponde ao número de fatores (componentes principais ou variáveis latentes) selecionados para truncar a igualdade, e E corresponde a matriz de resíduos, relacionada ao número de fatores escolhido.

Na figura 5.3, encontra-se apresentada uma seqüência de diagramas que visam ilustrar a concepção, bem como este processo de modelagem. As matrizes M constituem os chamados componentes principais (ou variáveis latentes), e são formadas pelo produto de dois vetores, t (os escores) e p (os pesos) conforme as expressões 5.6 e 5.7.

$$X = t_1.p_1^t + t_2.p_2^t + \dots + t_v.p_a^t + E \quad (5.6)$$

$$X = TP^t + E \quad (5.7)$$

$$\begin{array}{|c|} \hline \mathbf{X} \\ \hline \end{array} = \begin{array}{|c|} \hline \mathbf{M}_1 \\ \hline \end{array} + \begin{array}{|c|} \hline \mathbf{M}_2 \\ \hline \end{array} + \dots + \begin{array}{|c|} \hline \mathbf{M}_a \\ \hline \end{array} + \begin{array}{|c|} \hline \mathbf{E} \\ \hline \end{array}$$

$$\begin{array}{|c|} \hline \mathbf{X} \\ \hline \end{array} \begin{array}{l} m \\ n \end{array} = \begin{array}{|c|c|} \hline 1 & \mathbf{p}_1^t \quad m \\ \hline \mathbf{t}_1 & \\ \hline \end{array} \begin{array}{l} m \\ n \end{array} + \begin{array}{|c|c|} \hline 1 & \mathbf{p}_2^t \quad m \\ \hline \mathbf{t}_2 & \\ \hline \end{array} \begin{array}{l} m \\ n \end{array} + \dots + \begin{array}{|c|c|} \hline 1 & \mathbf{p}_a^t \quad m \\ \hline \mathbf{t}_v & \\ \hline \end{array} \begin{array}{l} m \\ n \end{array} + \begin{array}{|c|} \hline \mathbf{E} \\ \hline \end{array} \begin{array}{l} m \\ n \end{array}$$

$$\begin{array}{|c|} \hline \mathbf{X} \\ \hline \end{array} \begin{array}{l} m \\ n \end{array} = \begin{array}{|c|c|} \hline 1 & \mathbf{P}^t \quad m \\ \hline \mathbf{T} & \\ \hline \end{array} \begin{array}{l} m \\ n \end{array} + \begin{array}{|c|} \hline \mathbf{E} \\ \hline \end{array} \begin{array}{l} m \\ n \end{array}$$

Figura 5.3 – Esquema via diagramas para o algoritmo de regressão multivariada por mínimos quadrados parciais (PLS).

A dimensionalidade do espaço original é igual ao número de colunas em \mathbf{X} , ou seja, o número de variáveis originais, expresso por m . No novo modelo, a dimensionalidade é descrita pelo número de matrizes \mathbf{M}_i necessárias para descrever \mathbf{X} . Assim, se for possível descrever uma matriz \mathbf{X} que tenha muitas variáveis, por um número pequeno dessas matrizes \mathbf{M}_i haverá um decréscimo na dimensionalidade, sem perda significativa de informação.

Na modelagem por mínimos quadrados parciais (PLS), tanto a matriz das variáveis independentes \mathbf{X} , como a das variáveis dependentes \mathbf{Y} são representadas pelos escores e pelos pesos, conforme as expressões 5.7 e 5.8.

$$\mathbf{Y} = \mathbf{UQ}^t + \mathbf{F} \quad (5.8)$$

A relação entre as duas matrizes de dados \mathbf{X} e \mathbf{Y} pode ser obtida, correlacionando-se os escores de cada bloco, afim de obter uma relação linear descrita pela expressão 5.9, ou de forma equivalente por 5.10.

$$\mathbf{U}_a = \mathbf{b}_a \mathbf{t}_a \quad (5.9)$$

$$\mathbf{U} = \mathbf{bT} + \mathbf{e} \quad (5.10)$$

Na equação linear representada em 5.10, \mathbf{T} é uma matriz de resposta (assim como um conjunto de espectros) para uma série de amostras de calibração; \mathbf{U} é uma matriz contendo as propriedades (variáveis dependentes) de todas estas amostras; \mathbf{b} é um vetor contendo os parâmetros do modelo, e \mathbf{e} é um vetor que representa o ruído do espectro e os erros do modelo.

Entretanto, pode ocorrer que o modelo assim concebido não seja totalmente satisfatório. Isto porque cada matriz (\mathbf{X} e \mathbf{Y}) é decomposta separadamente, podendo resultar numa relação não muito adequada (não linear) entre os escores dos dois blocos.

Em virtude disto, deve-se buscar um modelo onde as matrizes de resíduos **E** e **F** sejam as menores possível e, ao mesmo tempo, conseguir uma relação linear entre **t** e **u**.

No método de regressão PLS, este ajuste é concretizado através de uma leve mudança nos valores dos escores, de forma a produzir a melhor correlação possível.

Existe um compromisso entre a capacidade em descrever as amostras nos espaços individuais (modelagem dos blocos **X** e **Y**), e o aumento na correlação **t** e **u** [238].

Assim que o modelo de regressão tenha sido gerado, estamos aptos a utilizar este modelo na previsão das propriedades de interesse.

Na figura 5.4, encontra-se ilustrado o conjunto de ações, que geralmente são realizadas, para obtermos a previsão de uma amostra semelhante àquelas modeladas pelo processo de regressão.

O método de regressão por mínimos quadrados parciais foi amplamente utilizado na última década, tendo se tornado um dos mais populares em função do seu desempenho.

Na tabela 5.2, podemos verificar o número de referências bibliográficas que apresentam em seu resumo, título, ou palavras-chave, o termo PLS, nos últimos 10 anos.

Além de algumas aplicações recentes do método PLS para a análise qualitativa e quantitativa de alimentos, pode ser citado o trabalho que determina anti-oxidantes sintéticos, realizado por Espinosa-Mansilla *et alii* [256], empregando espectrofotometria UV-visível.

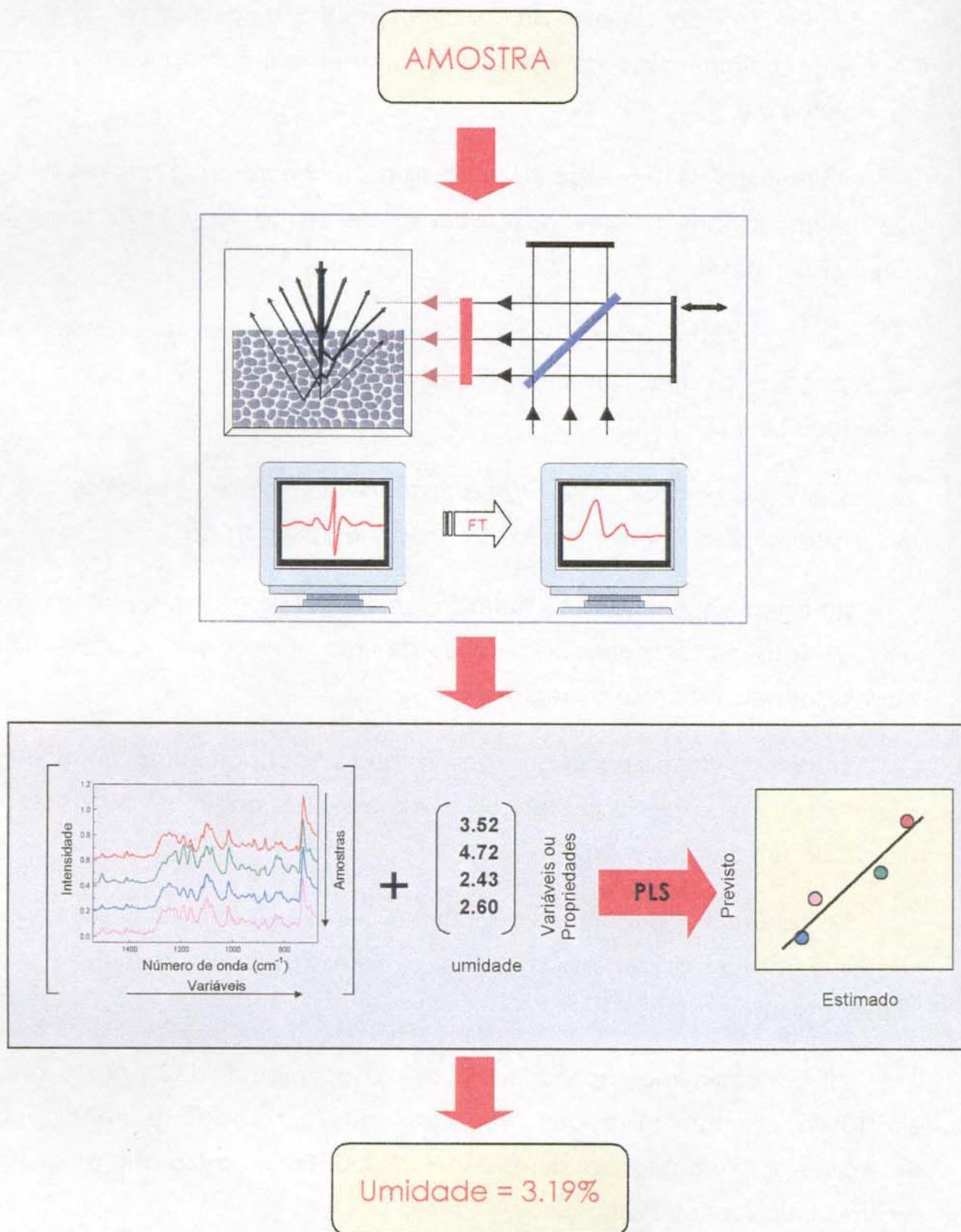


Figura 5.4 – Esquema da previsão de amostras externas.

O PLS também foi empregado na tipificação de vinhos, em conjunto com o modelo SIMCA por Ortiz *et alii* [257] e na quantificação de misturas binárias de ésteres de ácidos graxos por Haines *et alii* [258].

Kellner *et alii* [259] compararam métodos univariados e multivariados, em particular o PLS, para a determinação de sacarose em sucos de frutas empregando análise por injeção em fluxo e detecção no infravermelho com transformada de Fourier.

Tabela 5.2 – Freqüência dos termos MLR, PCR e PLS nas literaturas indexadas para os últimos 10 anos.

Período	MLR	PLS
1991-92	236	193
1993-94	228	289
1995-96	262	389
1997-98	268	569
1999-2000	196	481
1991-2000	1190	1921

Institute for Scientific Information – junho 2000

Em 1997, Reeves III [39], também utiliza a regressão por mínimos quadrados parciais para estudar o desempenho da reflexão difusa na quantificação de fibra, teor de proteína, lignina e digestibilidade, quando do emprego simultâneo das regiões do infravermelho médio e próximo em rações.

Neste mesmo ano, Dahlberg *et alii* [217] empregam o método PLS para desenvolver modelos de classificação de óleos vegetais, utilizando dados de espectroscopia FT-IR.

Em 1998, entre as diversas aplicações, que fazem uso da regressão por mínimos quadrados parciais, foram publicados os estudos de Archibald *et alii* [109] que aplicam o PLS na determinação de fibras em cereais a partir de dados NIRR, e a proposta de classificação de trigo associando as informações das espectrofotometrias nas regiões do visível e no infravermelho próximo.

Ainda neste período, medidas de espectroscopia por reflexão no infravermelho próximo foram empregadas para a determinação de amino ácidos presentes em rações, publicado por van Kempen e Bodin [260], para a determinação de proteína e suas propriedades bioquímicas em trigo, relatado por Delwiche *et alii* [261], e na determinação de nitrogênio em folhas de trigo, recentemente realizada por Mello *et alii* [262].

5.4 Critérios para Avaliação de Modelos de Regressão

Para que um modelo seja eficiente, é necessário que ele descreva da melhor forma possível a situação real, levando-se em conta o maior número possível de variações. Quanto maior o número de fatores no modelo menor o desvio da reta de calibração. Porém, o aumento de fatores também aumenta o ruído e os erros de modelagem.

O número ótimo de fatores a ser usado num modelo será o número de fatores que corresponda ao ponto no qual a diminuição do erro (produzida pelo aumento da complexidade do modelo) é compensada pelo aumento do

erro de superavaliação. Na figura 5.5 este ponto ótimo é representado pelo mínimo na curva, e para o qual um erro mínimo na previsão é produzido.

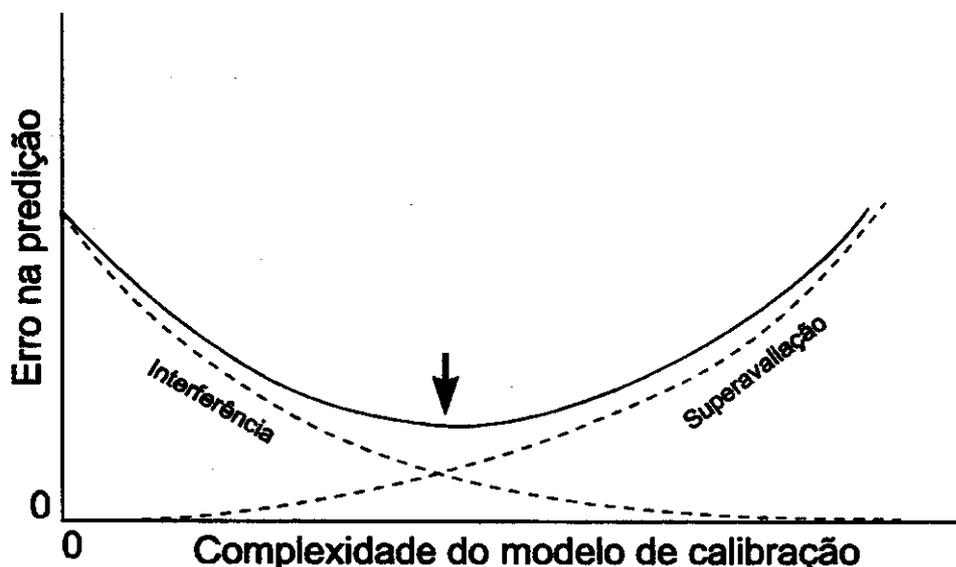


Figura 5.5 - Seleção do ponto ótimo para o modelo de calibração multivariada.

O número apropriado de fatores, variáveis latentes (VL), na calibração PLS pode ser determinada pelo PRESS (*predictive residual errors sum of squares*), que é a soma dos desvios quadrados para os valores estimados em relação aos valores atuais conforme a equação 5.11.

$$\text{PRESS} = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (5.11)$$

Nesta equação, y_i é o valor de referência para a i -ésima amostra, e \hat{y}_i é a previsão para o valor desta amostra. Um decréscimo no PRESS, então, indica um modelo com aperfeiçoada habilidade de previsão (robusto). O modelo otimizado é formado usando o número mínimo de fatores que tem um PRESS com um erro padrão mínimo.

Outra medida que é empregada para avaliar a habilidade de previsão de um modelo é o SEC (standard error of calibration), que apresenta informações semelhantes aos do PRESS, sendo expressa conforme equação 5.12, onde n representa o número de amostras utilizadas.

$$SEC = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (5.12)$$

Na avaliação do conjunto de amostras de validação podemos ajustar a expressão empregada para determinar o SEC, sendo denominada de SEV. Este critério é bastante útil quando deseja-se avaliar se o modelo tende a superdimensionar o número de variáveis latentes a serem empregados. Os valores de SEV, são calculados com os y_i de referência e os valores estimados com o conjunto de validação, conforme equação 5.13.

$$SEV = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (5.13)$$

Uma outra notação bastante empregada é o SEP, que apresenta informação semelhante ao SEV, porém se refere ao conjunto de amostras empregados para a previsão.

5.5 Seleção de Variáveis e Otimização de Modelos de Regressão

As técnicas espectroscópicas na região do infravermelho, geralmente contém um grande número de variáveis (sinal ou absorbância para cada comprimento de onda, no caso do infravermelho). Por exemplo, os espectrofotômetros que operam no infravermelho médio (região entre 4000-400 cm^{-1}) quando utilizam resolução de 4 cm^{-1} , geram informações sobre a intensidade de absorção em mais de 1800 frequências (variáveis).

Todavia, o grande número de variáveis, provindo das técnicas espectroscópicas, pode tornar o desempenho dos processadores muito lento ou conter informações sobre o sistema estudado que venham a interferir na modelagem, como por exemplo bandas que não contenham nenhuma informação das espécies ou propriedades a serem analisadas.

Por outro lado, alguns métodos, empregados na calibração multivariada de dados espectroscópicos, apresentam dificuldade para a solução quando o número de variáveis é superior ao número de amostras modelas, como é o caso da Regressão Linear Múltipla (MLR).

Em virtude disto, tornaram-se indispensáveis a utilização de critérios e métodos para selecionar os comprimentos de onda (ou números de onda) correspondentes aos sinais do detector (absorbâncias), que modeladas apresentassem melhor sensibilidade e linearidade para os analitos e/ou propriedades de interesse.

Na busca do subconjunto de variáveis, que melhor ajustasse determinado modelo multivariado, foram apresentados diversos critérios para proceder a seleção, sendo a grande maioria deles advindo do próprio processo de análise multivariada, onde muitos desses buscam avaliar e comparar o erro nos conjuntos de modelagem e de validação.

Intuitivamente pode-se imaginar que o melhor subconjunto de variáveis possa ser gerado pela análise combinatória de todas as variáveis disponíveis e a partir daí aplicar algum critério para sua seleção. Entretanto, diante do número elevado de variáveis que as técnicas espectroscópicas em questão apresentam, este procedimento torna-se impraticável seja pelo tempo e/ou custo computacionais.

Da mesma forma, procedimentos manuais baseados na observação do conjunto de dados, ou de parâmetros modelados a partir destes, como o erro, continuam sendo empregados na seleção preliminar de sub-regiões do espectro, às quais devem conter a melhor qualidade de informação. Porém para muitos casos é necessário conhecimento prévio do conjunto de dados e dos procedimentos experimentais empregados em sua obtenção.

Como consequência disto, nos últimos anos, diversos pesquisadores têm empregado métodos de otimização combinatória, buscando uma forma mais genérica e precisa para a obtenção do conjunto de variáveis, cuja modelagem resulte de forma rápida e eficaz num modelo mais robusto.

No início dos anos 80, Lindberg *et alii* [263] e Sjöström *et alii* [255] empregam o método de regressão dos mínimos quadrados parciais (PLS) em comparação ao método de regressão múltipla (MR) e se deparam com o problema de selecionar as variáveis para a modelagem via MR. Neste caso empregam o algoritmo de decomposição de valor singular (SVD) cujo resultado é um novo conjunto de variáveis independentes, ortogonais e passíveis de modelagem via MR. Entretanto utilizam o método de validação cruzada (CV), já intensivamente estudado por Stone [264], para selecionar em ambos os modelos o número de fatores (variáveis latentes) a serem empregados comparando o erro de validação.

Neste mesmo período, Honigs *et alii* [265, 266], trabalhando com dados de reflectância no infravermelho próximo avaliam o desempenho do erro padrão de previsão (SEP) e mais tarde utilizam o critério do erro quadrático mínimo (MSE). Enquanto isto, Rossi *et alii* [267, 268], apresentam um modelo para a seleção de distintos intervalos de comprimento de onda na quantificação de misturas, de forma que o intervalo selecionado enfatize as informações espectrais do componente de interesse, sendo utilizadas também a primeira e segunda derivadas para melhor localizar estas sub-regiões.

Thijssen *et alii* [269, 270] discutem a otimização em problemas de mínimos quadrados propondo um critério baseado nas teorias de informação recursivas, resultando em um eficiente método para a seleção de variáveis para determinações espectrofotométricas multicomponente.

Na segunda metade dos anos 80, Otto e Wegscheider [271] discutem a seletividade característica de sistemas multicomponente buscando justificar um critério para assegurar a exatidão e a precisão nas determinações multivariadas. Eles elegem o número de condição de calibração ($\text{Cond}(X)$), também empregado por Juhl e Kalivas [272, 273], como o critério mais genérico para expressar a seletividade, sendo este adequado na seleção de comprimentos de onda em determinações multicomponente.

Krzanowski [274, 275] empregou a análise por componentes principais (PCA) em conjunto com sub-rotinas via Monte Carlo na seleção de variáveis de dados multivariados, sendo mais tarde este mesmo método utilizado por Scarponi *et alii* [179] num estudo quimiométrico de vinhos e por Robert *et alii* [276] em dados de NIR de amostras de leite, para selecionar os melhores comprimentos de onda para a determinação de lactose, gordura e proteína.

Lorber e Kowalski [277] são os primeiros a empregar os algoritmos de *stepwise elimination* e *backward elimination* utilizando como critério a minimização do erro de previsão na seleção de subconjuntos de variáveis.

Em 1990, Kowalski [278] compara o desempenho dos diversos métodos de regressão: *ridge regression* (RR), *principal component regression* (PCR), *partial least squares* (PLS) e *multiple linear regression* (MLR), empregando como critério o PRESS. Para fazer a seleção das variáveis para o método MLR empregou os algoritmos *backward elimination* e *forward selection*.

Diante dos bons resultados apresentados pelos algoritmos de otimização combinatória na seleção de comprimentos de onda, como por exemplo o *branch-and-bound*, o início da década de 90 se caracterizou pelo visível duelo entre o algoritmo genético e o *simulated annealing*, ambos métodos de otimização combinatória não determinísticos. Kalivas *et alii* [279], utilizando dos conceitos do *simulated annealing* generalizado, propõem a seleção de comprimentos de onda utilizando inicialmente um modelo no qual eram simuladas gaussianas. Entretanto o emprego do algoritmo de *generalized simulated annealing* só foi realizado em análises multicomponente a partir do trabalho de Kalivas [280], no qual o autor fazia uma seleção de amostras para posterior modelagem em métodos de calibração multivariada.

Sutter *et alii* [281] também investigaram, comparativamente os métodos *backward elimination* e *forward selection*, a seleção de subconjuntos de variáveis visando a otimização de modelos de regressão utilizando conjuntos de espectros na região do infravermelho próximo. Mas o trabalho decisivo na seleção de comprimentos de onda, para a espectroscopia no infravermelho próximo, foi apresentado por Brenchley *et alii* [282], no qual o algoritmo *generalized simulated annealing* com tamanho de passo modulado foi empregado em um conjunto de dados de amostras trigo e outro de amostras de gasolina.

Recentemente os algoritmos genéticos tem sido empregados como ferramentas de otimização em Quimiometria. Um exemplo da aplicação destes algoritmos é apresentada por Lucasius e Kateman [283], na seleção de comprimentos de onda de dados espectrais na região do ultravioleta em análises de um sistema multicomponente .

Seguiram-se trabalhos que investigaram o desempenho dos algoritmos genéticos frente a outras técnicas na seleção de comprimentos de onda. Lucasius, Beckers e Kateman [284] apresentam um estudo comparativo do algoritmo genético com os métodos *simulated annealing* e *stepwise elimination*, tendo apresentado o algoritmo genético melhor desempenho quando foram empregados como critérios de seleção a seletividade, precisão e o erro mínimo quadrático médio. Da mesma forma, Hörchner e Kalivas [285] aprofundaram o estudo comparativo entre os algoritmos genéticos e variantes do *simulated annealing* na seleção de comprimentos de onda. Porém, seus resultados apresentam excelentes desempenhos para as variantes do *simulated annealing* na seleção de variáveis de dados espectroscópicos nas regiões do ultravioleta-visível.

A implementação dos algoritmos genéticos na seleção de variáveis (comprimentos de onda) na espectroscopia na região do infravermelho próximo é recente. Destacam-se os trabalhos desenvolvidos por Leardi [286], Jouan-Rimbaud *et alii* [228,287], Shaffer *et alii* [288], Bangalore *et alii* [289] e da Costa Filho [290].

Nos últimos anos, um emergente método de seleção de variáveis para técnicas espectroscópicas tem se mostrado promissor. Este método, denominado de eliminação de variáveis não informativas (UVE), teve sua base no recente trabalho publicado por Frenich *et alii* [291] utilizando-se o modelo de calibração PLS e avaliando seu desempenho segundo a média quadrática dos erros de previsão de validação cruzada (RMSECV).

Seguiu o trabalho de Centner *et alii* [205] descrevendo um método para detecção de não homogeneidade em conjunto de espectros no infravermelho próximo e eliminação de amostras anômalas (*outliers*). O desenvolvimento do UVE por Centner *et alii* [231], o qual consiste em eliminar variáveis cuja informação seja correspondente ou inferior às variáveis randômicas convenientemente adicionas, resultou num outro trabalho publicado nesta área por Jouan-Rimbaud *et alii* [292] denominado de extração de componentes relevantes para PLS (RCE-PLS) no qual a recente ferramenta UVE é associada a transformada discreta wavelet.

Parte Experimental

6.1 Amostragem

Amostras de farinha de trigo coletadas entre agosto de 1998 e abril de 1999, oriundas de diferentes moinhos, foram cedidas pela empresa Filler S.A., sediada no município de Santa Cruz do Sul – RS. Um total de 104 amostras foram selecionadas para elaborar uma rotina que possibilitasse estimar os parâmetros de qualidade em amostras de farinha de trigo.

6.2 Análises através dos Métodos de Referência

As análises empregando métodos de referência foram realizadas segundo os métodos apresentados na tabela 6.1, bem como o número de réplicas realizadas para cada parâmetro. Para os métodos onde foram realizadas replicatas, os valores de referência foram considerados como a média.

A descrição detalhada dos métodos e seus respectivos procedimentos são apresentados no apêndice A deste trabalho.

As análises para determinar os teores de cinza e de proteína foram realizadas nos laboratórios da UNISC (Universidade de Santa Cruz do Sul), enquanto que as determinações do número de queda foram realizadas conforme a metodologia descrita na tabela 6.1, no laboratório da Empresa Filler em Santa Cruz do Sul.

Tabela 6.1 – Métodos de referência empregados e o número de replicatas.

Parâmetro	Método	Réplicas
umidade	44-15A da AACC ^a (1981)	1
proteína	2055 da AOAC ^b (1984)	3
cinzas	08-03 da AACC ^a (1984)	3
número de queda	56-81B da AACC ^a (1982)	4

a – AACC – American Association of Cereal Chemists

b – AOAC – Association of Official Analytical Chemists

Os valores de umidade, não foram empregados na modelagem, mas sim, para implementar correções na determinação do número de queda, aplicando-se a tabela de correção do peso da amostra para umidade de 14%, conforme apresentado no apêndice A.

6.3 Aquisição dos Espectros no Infravermelho

Os espectros no infravermelho foram coletados conforme descritos nas tabelas 6.2, 6.3 e 6.4, sendo realizadas 2 réplicas para cada amostra em cada técnica, 64 varreduras e com resolução de 4 cm^{-1} .

Tabela 6.2 – Instrumentação e regiões empregadas na aquisição dos espectros HATR.

Instrumental	Espectrofotômetro Nicolet 520
Região Espectral	650-4000 cm^{-1}
Freqüência do Laser	15798 cm^{-1}
Apodização	Happ-Genzel
Número de pontos	1738
Formato	Log (1/R)
Detector	DTGS KBr
Divisor de Feixe	KBr
Acessório	HATR Spectra-Tech
Aplicativo	OMNIC 4.1

Tabela 6.3 – Instrumentação e regiões empregadas na aquisição dos espectros DRIFTS.

Instrumental	Espectrofotômetro Nicolet Magna 550
Região Espectral	450-4000 cm^{-1}
Frequência do Laser	15798 cm^{-1}
Apodização	Happ-Genzel
Número de pontos	1842
Formato	Log (1/R)
Detector	DTGS KBr
Divisor de Feixe	KBr
Acessório	Reflexão Difusa Easydif® Pike
Aplicativo	OMNIC 4.1

Tabela 6.4 – Instrumentação e regiões empregadas na aquisição dos espectros NIRR.

Instrumental	Espectrofotômetro Bomem DA-08
Região Espectral	5555-10000 cm^{-1}
Freqüência do Laser	15798 cm^{-1}
Apodização	Bartlet
Número de pontos	2307
Formato	Log(1/R)
Detector	InSb
Divisor de Feixe	Quartzo
Acessório	Reflexão Difusa Colector Spectra-Tech
Aplicativo	DA Series acquisition program v1.2h

6.3.1 Aquisição dos Espectros HATR

O espectro de referência foi obtido com o acessório de HATR tipo bote vazio, equivalente ao apresentado na figura 6.1.

Para a obtenção dos espectros, $0,60 \pm 0,01$ g de cada amostra de farinha de trigo foi pesada em balança analítica e cuidadosamente transferida para o bote do acessório de HATR (figura 6.1).

O valor de massa de farinha a ser colocada no acessório foi determinada através da obtenção de uma série de espectros com crescente quantidade de farinha, até que o ruído fosse minimizado e os contornos das bandas espectrais ficassem melhor definidos.



Figura 6.1 – Acessório de reflexão total atenuada horizontal (HATR).

A farinha foi espalhada por toda a superfície do cristal de ATR e, em seguida, coberta e pressionada para garantir um melhor empacotamento da amostra e maior reprodutibilidade.

Os espectros assim obtidos são apresentados na figura 6.2.

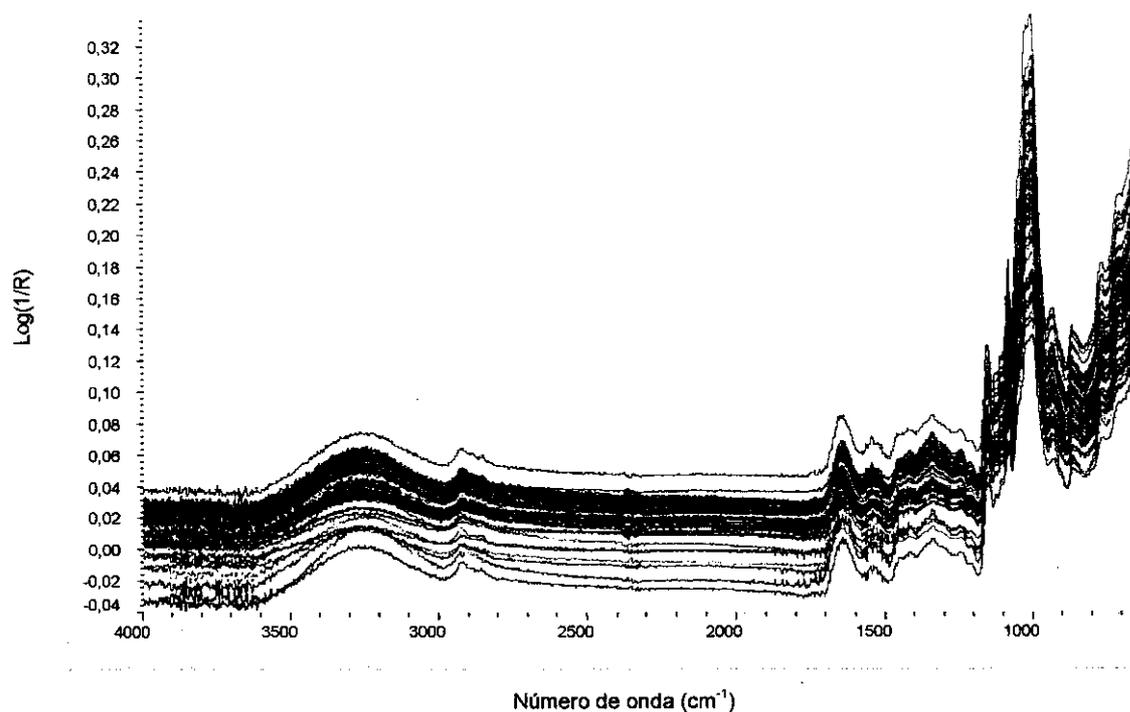


Figura 6.2 – Espectros obtidos no infravermelho médio empregando HATR.

6.3.2 Aquisição dos Espectros DRIFTS

O espectro de referência foi obtido com o acessório DRIFTS colocando-se no lugar da amostra uma peça maciça construída em liga de alumínio comercial. A superfície da peça foi cuidadosamente lixada de forma a tornar-se fosca, conforme ilustrado na figura 6.3.

Foram pesadas em balança analítica, uma mesma massa em gramas de cada amostra de farinha de trigo e cuidadosamente transferidas para o amostrador do acessório de DRIFTS (figura 6.4).

Diante das dificuldades encontradas na reprodutibilidade de espectros por reflexão difusa, a literatura tem apresentado algumas alternativas para um melhor empacotamento das amostras e conseqüente melhora dos espectros [292,293].

Depois de uma série de tentativas não bem sucedidas para reproduzir os espectros de reflexão difusa, um sistema para empacotar as amostras de farinha, possibilitando uma amostragem da superfície mais homogênea, foi idealizado e é apresentado pelo esquema da figura 6.5.

Este empacotador foi inicialmente conformado na oficina mecânica do Instituto de Química da Unicamp. O mesmo foi construído tanto em poliamida (nylon), como em polifluoretileno (teflon), tendo a peça em teflon apresentado melhor desempenho.

A farinha foi empacotada no porta amostra, colocando a massa de amostra pela parte superior, conforme vista superior ilustrada na figura 6.5, e na seqüência foi acoplado o embolo cuidadosamente, seguido por um giro, com o objetivo de manter a superfície uniforme e a aquisição do espectro mais reprodutível.

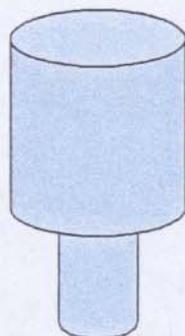


Figura 6.3 – Dispositivo em alumínio com superfície fosca.

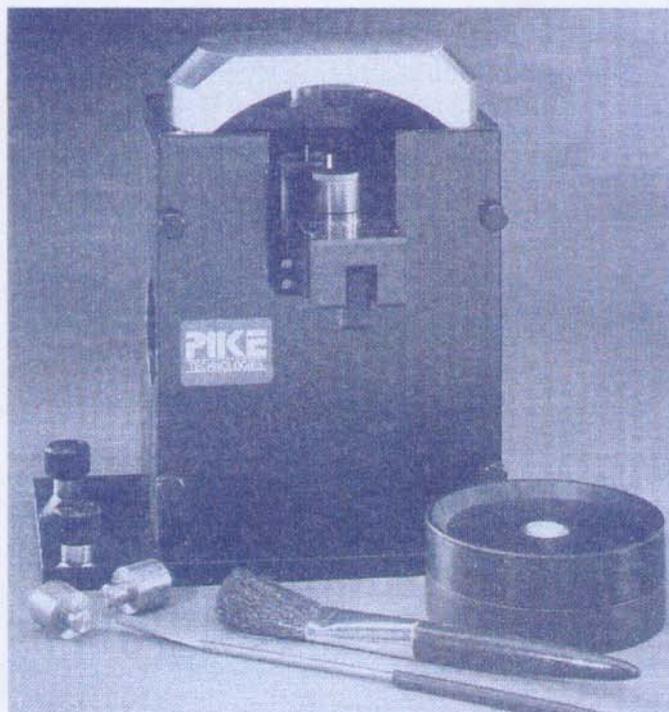


Figura 6.4 – Acessório para reflexão difusa (DRIFTS).

O valor de massa de farinha a ser colocada no acessório, também foi determinado através da obtenção de uma série de espectros com crescente quantidade de farinha, até que o empacotamento da amostra fosse reprodutível e os contornos das bandas espectrais ficassem melhor definidos.

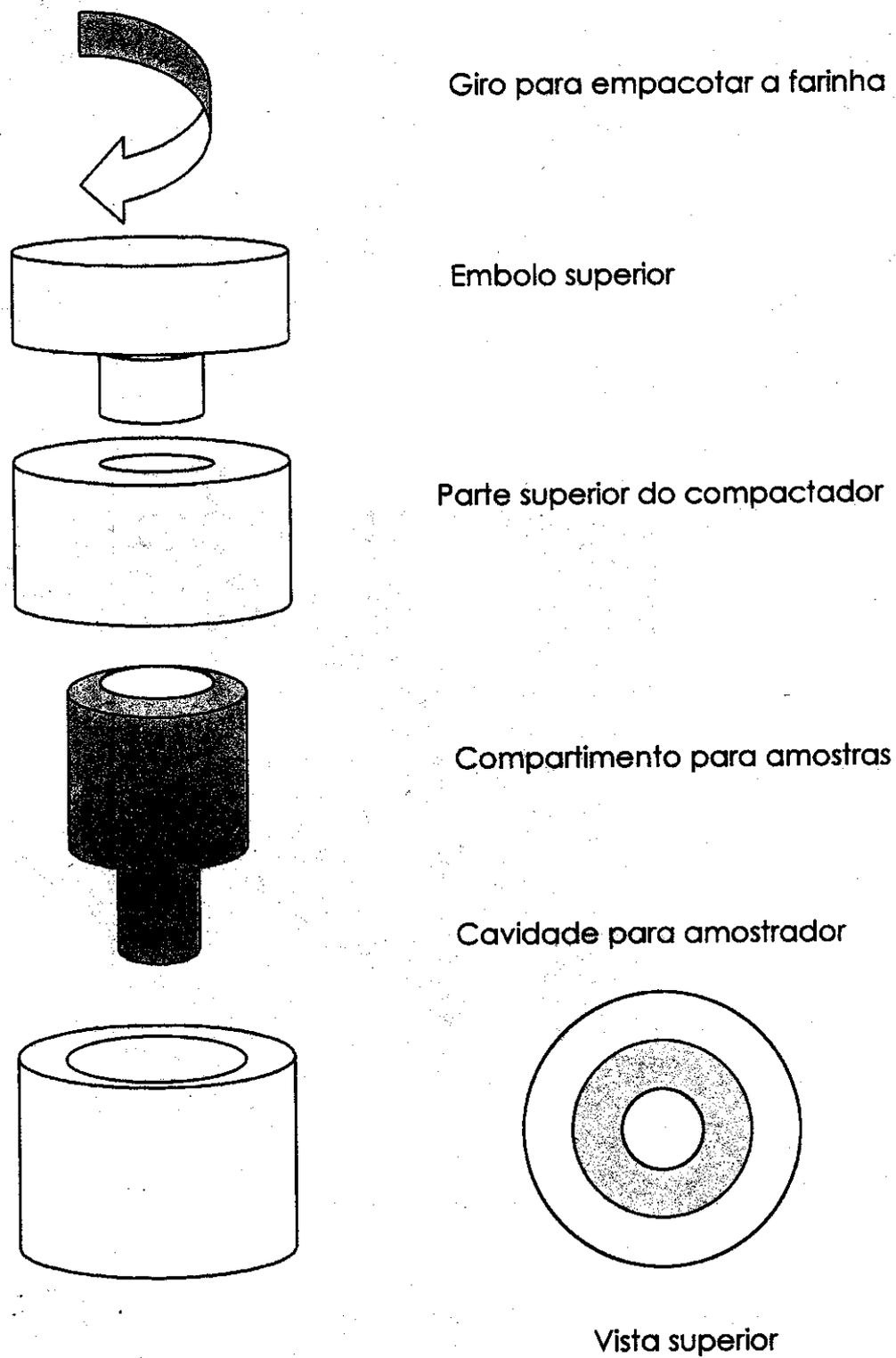


Figura 6.5 – Representação esquemática do empacotador de amostras.

O conjunto de espectros obtidos por reflexão difusa no infravermelho médio com transformada de Fourier é apresentado na figura 6.6.

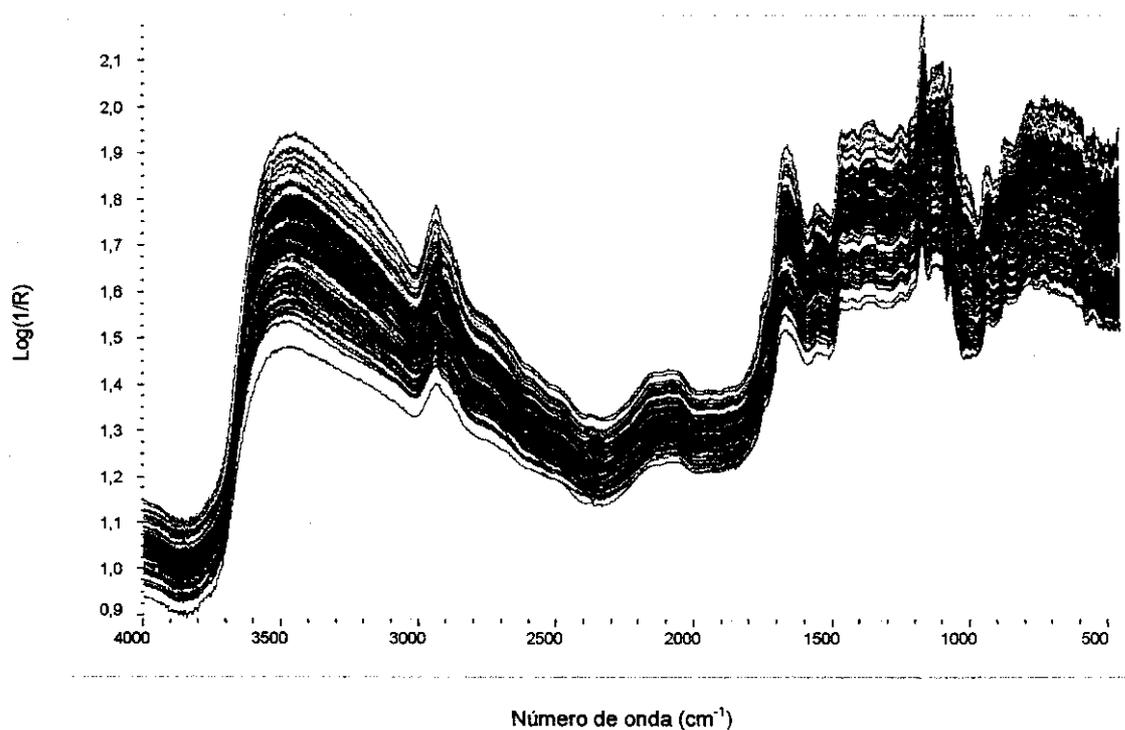


Figura 6.6 – Espectros no infravermelho médio empregando reflexão difusa.

6.3.3 Aquisição dos Espectros NIRRS

Para a técnica de reflexão no infravermelho próximo, o espectro de referência foi obtido com o acessório de reflexão difusa colocando-se no lugar da amostra uma peça maciça construída com polifluoretileno comercial. A superfície da peça foi cuidadosamente lixada de forma a tornar-se fosca.

Os demais detalhes com respeito a aquisição dos espectros NIRRS foram análogos aos realizados para a técnica DRIFTS, diferindo apenas no peso empregado de cada amostra, que neste caso foi de $0,160 \pm 0,001$ g de amostra de farinha de trigo, por trata-se de um outro modelo de acessório de reflexão difusa. Os espectros assim coletados são apresentados na figura 6.7.

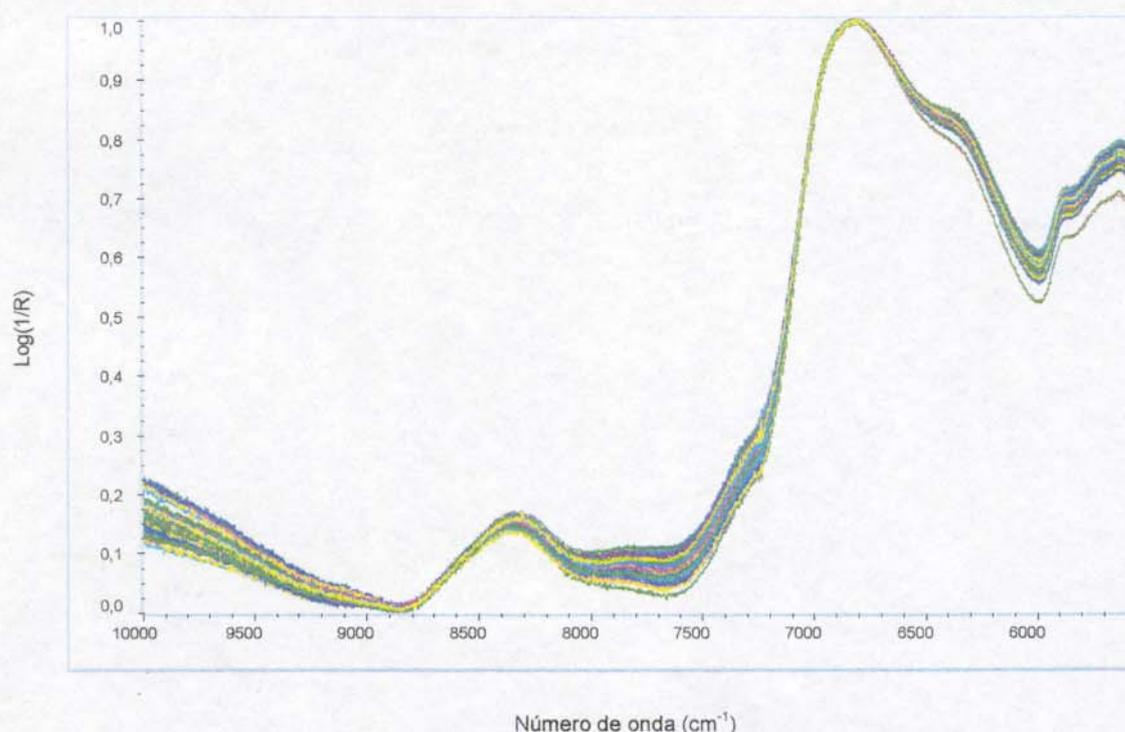


Figura 6.7 – Espectros no infravermelho próximo empregando reflexão difusa.

6.4 Tratamento dos Espectros

Quando da aquisição dos espectros por técnicas de reflexão, alguns fenômenos não desejados podem ocorrer, como por exemplo o espalhamento de luz, provocado pela não homogeneidade da amostra, principalmente pelas diferenças de granulometria, geometria, empacotamento e orientação das partículas. Também costumam acompanhar os sinais analíticos, obtidos por técnicas instrumentais, ruídos das mais diversas origens. A presença destes ruídos podem dificultar a interpretação, bem como a modelagem das propriedades associadas aos sinais espectrais, sendo recomendada sua eliminação ou minimização.

Na tentativa de minimizar os efeitos causados pela dificuldade na obtenção de um espectro ideal, visto que as farinhas de trigo apresentam distribuição granulométrica diferenciada, podem ser empregadas técnicas de tratamento no espectro, antes que o conjunto dos espectros seja modelado pelas técnicas quimiométricas.

Neste trabalho foram experimentados a correção do espalhamento de luz (MSC) e o alisamento empregando um algoritmo de Savitzky-Golay.

Outro tratamento nos dados espectrais comumente empregado, para melhorar o desempenho de modelos de regressão multivariados, trata-se da aplicação da primeira ou segunda derivadas sobre os dados espectrais brutos [295-298].

Estes métodos costumam destacar ombros espectrais bem como minimizar o efeito de inclinações provocadas na linha de base dos espectros, devido a morfologia das partículas.

6.4.1 Correção do Espalhamento de Luz (MSC)

O espalhamento de luz é um fenômeno que altera a relação entre a intensidade das medidas de reflexão e a concentração das espécies absorventes de uma matriz. A reflexão, nestes casos, torna-se função da concentração das espécies absorventes presentes, bem como da falta de homogeneidade ótica da amostra [299-301].

A técnica de correção do espalhamento de luz (MSC) é um processo matemático que visa corrigir o efeito do espalhamento de luz presente no espectros obtidos por técnicas de reflexão. No MSC a interferência relativa ao espalhamento de luz é minimizada com base no espalhamento médio de todos os espectros formadores na matriz de dados X .

Para tanto, é inicialmente calculado o espectro médio \bar{x}_i , fazendo-se a média das absorbâncias para todos as amostras da matriz de dados X , para cada uma das colunas (referentes e cada um dos comprimentos de onda do espectro).

Calcula-se a regressão linear para cada um dos espectros originais contra o espectro médio, em cada um dos comprimentos de onda, o que resulta na equação da reta do tipo apresentado em 6.1.

$$X_k = U_k + v_k \cdot X_i \quad (6.1)$$

Finalmente utilizam-se os coeficientes linear (U_k) e angular (v_k) desta reta para corrigir cada um dos espectros originais, conforme expressão 6.2.

$$X_{\text{corrigido}} = (X_k - U_k) / v_k \quad (6.2)$$

Com o objetivo de minimizar o efeito do espalhamento de luz inerente ao conjunto de espectros, visto a natureza da farinha de trigo não ser homogênea e de apresentar partículas por vezes de granulometria elevada, foi empregado a correção do espalhamento de luz.

Os resultados encontrados com a opção desta correção serão apresentados posteriormente para as técnicas espectroscópicas empregadas, bem como a comparação com outros modelos, onde não são empregadas este tipo de correção.

6.4.2 Alisamento Savitzky-Golay

O método de alisamento empregando algoritmo Savitzky-Golay é um dos mais clássicos empregados para a suavização de sinais analíticos [302]. O ajuste resultante deste algoritmo é originado pelo método de mínimos quadrados, empregando um polinômio de grau variável.

O algoritmo empregado por Savitzky-Golay segue as seguintes etapas:

- seleciona-se o espectro;
- defini-se a largura do intervalo, isto é o número de pontos;
- calcula-se o ponto central do intervalo anteriormente definido;
- remove-se o ponto central deste intervalo;
- ajusta-se um polinômio de grau variável empregando o método dos mínimos quadrados para os demais pontos do intervalo;

- com este polinômio, calcula-se o valor correspondente ao ponto anteriormente excluído, e
- desloca-se o intervalo para o ponto vizinho e repete-se as etapas anteriores.

Neste trabalho foi experimentado o alisamento empregando algoritmo Savitzky-Golay para o conjunto de espectros no infravermelho selecionado-se a opção de alisamento automático disponível no software Omnic E.S.P. versão 4.1 empregado para a modelagem.

Os resultados encontrados com o emprego deste método são oportunamente apresentados e comparados.

6.4.3. Primeira e Segunda Derivadas

Ao aplicarmos as operações de derivação aos espectros, as informações contidas ao longo dos diferentes comprimentos de onda são geralmente acentuadas.

Na figura 6.8 podemos observar que a quantidade de informações, bem como a intensidade relativa entre eles é maior para os espectros com derivada do que no espectro sem tratamento.

Deve-se ter cuidado com a qualidade dos espectros com os quais deseja-se aplicar o cálculo das derivadas, uma vez que não só os sinais espectrais, como também os ruídos, tornam-se acentuados.

Desta forma deve-se avaliar o desempenho não só de modelos onde são empregados os espectros com derivadas, mas também de modelos onde não são empregados nenhum tipo de tratamento.

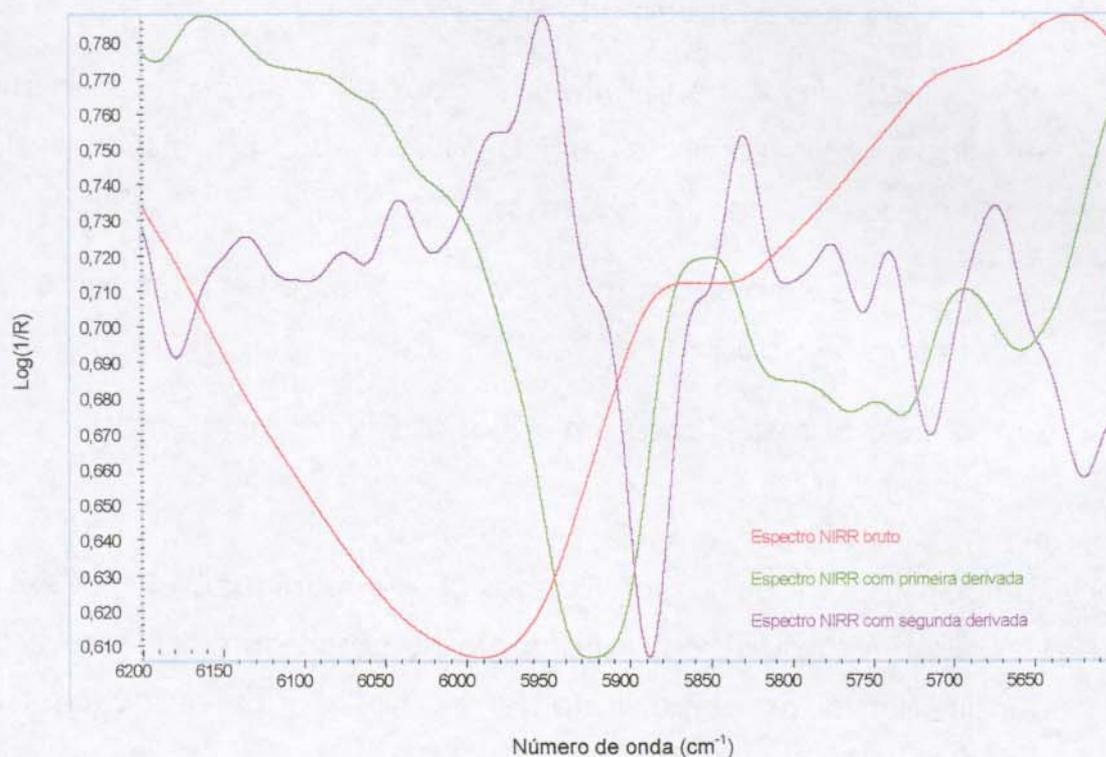


Figura 6.8 – Espectros no infravermelho próximo sem tratamento (em vermelho), com primeira derivada (em verde) e com segunda derivada (em roxo).

6.4.4 Fluxogramas dos Tratamentos Empregados

Para facilitar a identificação dos modelos de regressão multivariada desenvolvidos vamos adotar a seguinte notação: o nome da técnica seguida de um hífen e uma letra que indicará o tipo de pré-processamento empregado nos dados, um número que informará o tipo de tratamento empregado nos espectros antes da modelagem e finalmente entre parênteses o número de variáveis latentes.

Por exemplo para o modelo denominado HATR-B2(5) temos:

- um modelo que emprega dados de HATR;
- que utiliza simultaneamente os dados centrados na média e escalados pela variância (um detalhamento maior será dado no capítulo de resultados e discussões);
- do tipo 2 que emprega também a correção do espalhamento de luz, porém sem derivada;
- um modelo para o qual foram utilizadas 5 variáveis latentes.

a) Modelos tipo 1 – Os modelos identificados com a terminação 1, por exemplo HATR-B1, utilizam os dados sem tratamento ou empregam somente primeira ou segunda derivadas como tratamentos nos dados espectrais, sendo diferenciados conforme ilustrado na figura 6.9.



Figura 6.9 – *Tratamentos empregados nos dados dos modelos tipo 1.*

b) Modelos tipo 2 – Os modelos identificados com a terminação 2, por exemplo DRIFTS-E2, empregam correção para o espalhamento de luz com dados originais, ou dados com primeira ou segunda derivadas, sendo diferenciados conforme ilustrado na figura 6.10.

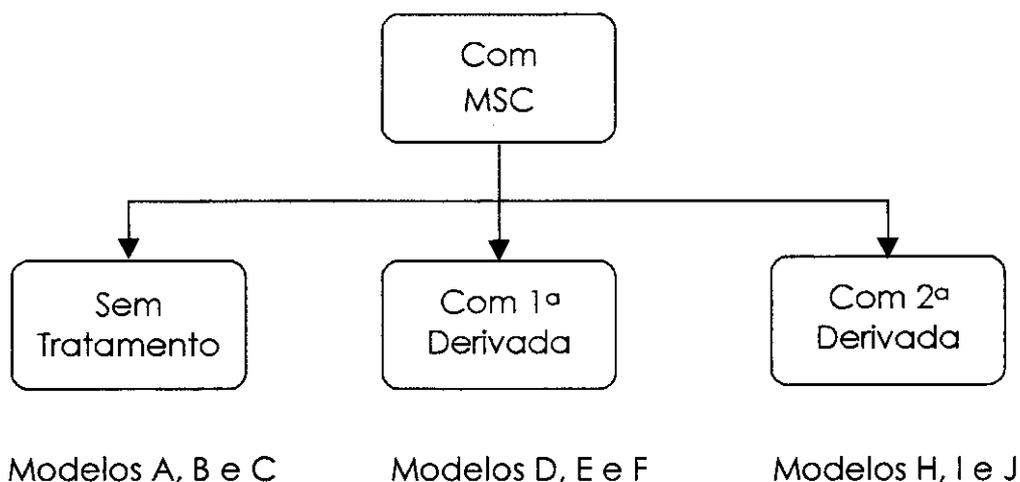


Figura 6.10 – *Tratamentos empregados nos dados dos modelos tipo 2.*

c) Modelos tipo 3 – Os modelos identificados com a terminação 3, por exemplo HATR-B3, empregam alisamento via algoritmo Savitzky-Golay com dados originais, ou dados com primeira ou segunda derivadas, sendo diferenciados conforme ilustrado na figura 6.11.

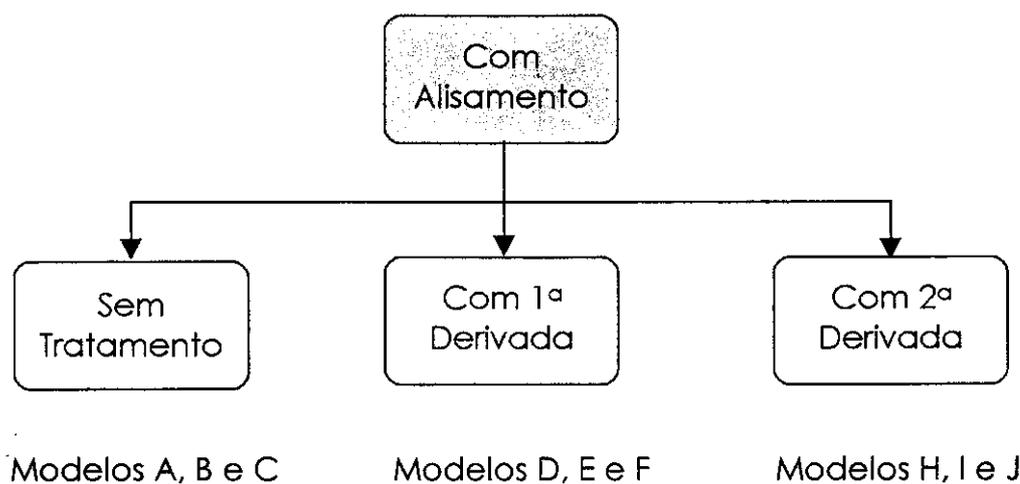


Figura 6.11 – *Tratamentos empregados nos dados dos modelos tipo 3.*

d) Modelos tipo 4 e 5 – Os modelos identificados com o número 4 ou 5, por exemplo NIRR-H4 e NIRR-H5, empregam o conjunto de etapas descritas no fluxograma da figura 6.12 obtendo-se um espectro médio para cada amostra com dados originais, ou dados com primeira ou segunda derivadas, sendo diferenciados conforme ilustrado. Os espectros foram normalizados pelo pico mais intenso de forma a minimizar deslocamentos e compressões de sinais durante a aquisição. As duas réplicas foram então somadas, e sobre o espectro resultante foi empregada a ferramenta de alisamento, baseada no algoritmo de Savitsky-Golay. Na seqüência foi diminuído o número de variáveis pela redução da resolução de 4 cm^{-1} para 8 cm^{-1} , reduzindo-se pela metade o número de pontos de cada espectro. E novamente o espectro foi normalizado ficando compreendido entre 0 e 1.

Nas figuras 6.13 e 6.14 podem ser observados os conjuntos de espectros DRIFTS antes e depois da seqüência de etapas descritas no fluxograma anterior. O conjunto de espectros apresentado na figura 6.14 é bem mais uniforme, além de estar com seus sinais compreendidos entre 0 e 1.

6.5 Modelagem dos Dados Espectrais

Todos os modelos de regressão apresentados neste trabalho foram desenvolvidos utilizando o aplicativo *Turbo Quant Analyst v.1.1a* da Nicolet, sendo baseados no método dos mínimos quadrados parciais (PLS). Algumas tentativas empregando regressão por componentes principais (PCR) também foram experimentadas, porém apresentaram baixa correlação, mesmo com o emprego de mais variáveis latentes ($1 \leq VL \leq 25$). Em função disto o método PCR não foi aprofundado neste estudo.

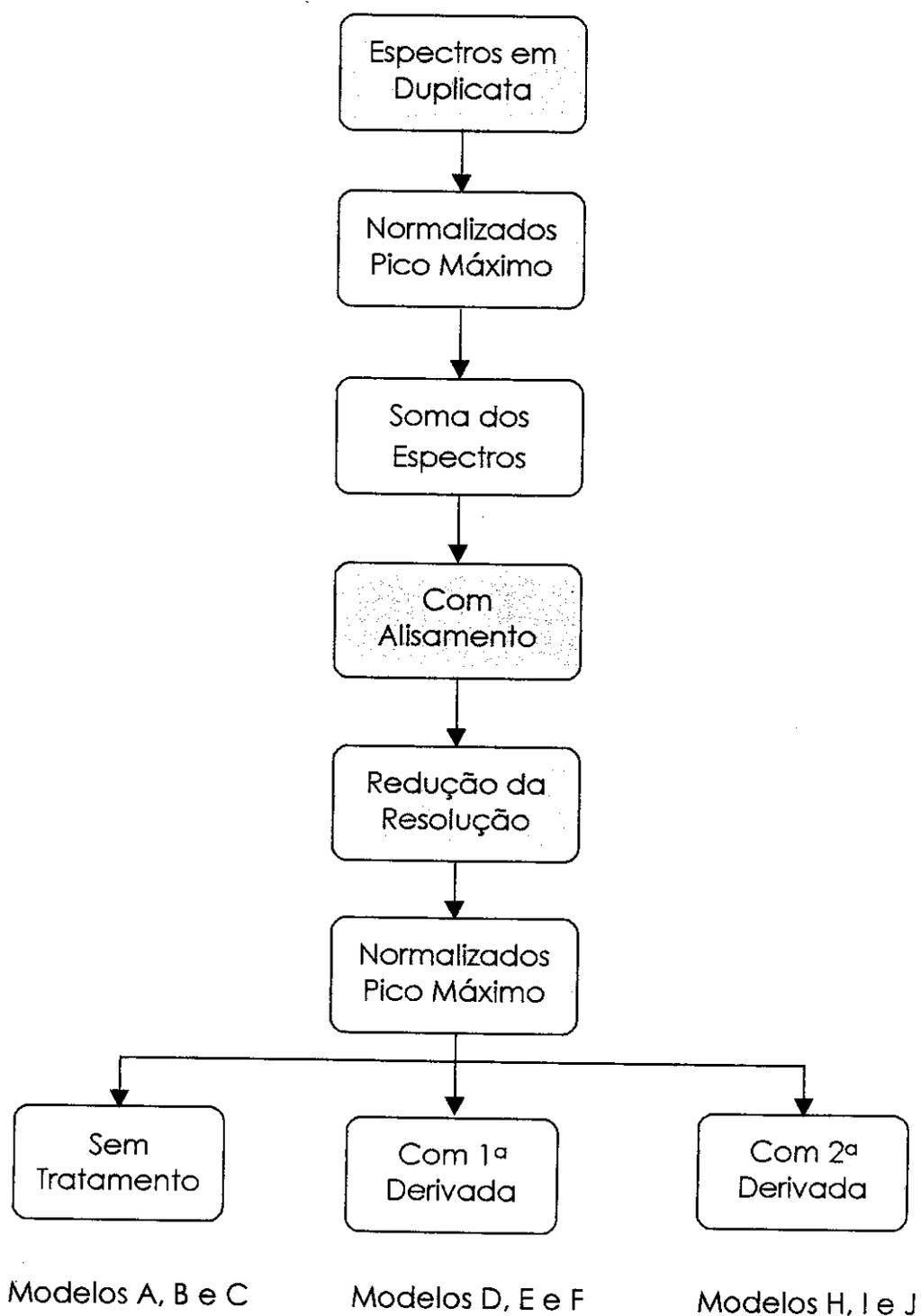


Figura 6.12 – Tratamentos empregados nos dados dos modelos tipos 4 e 5.

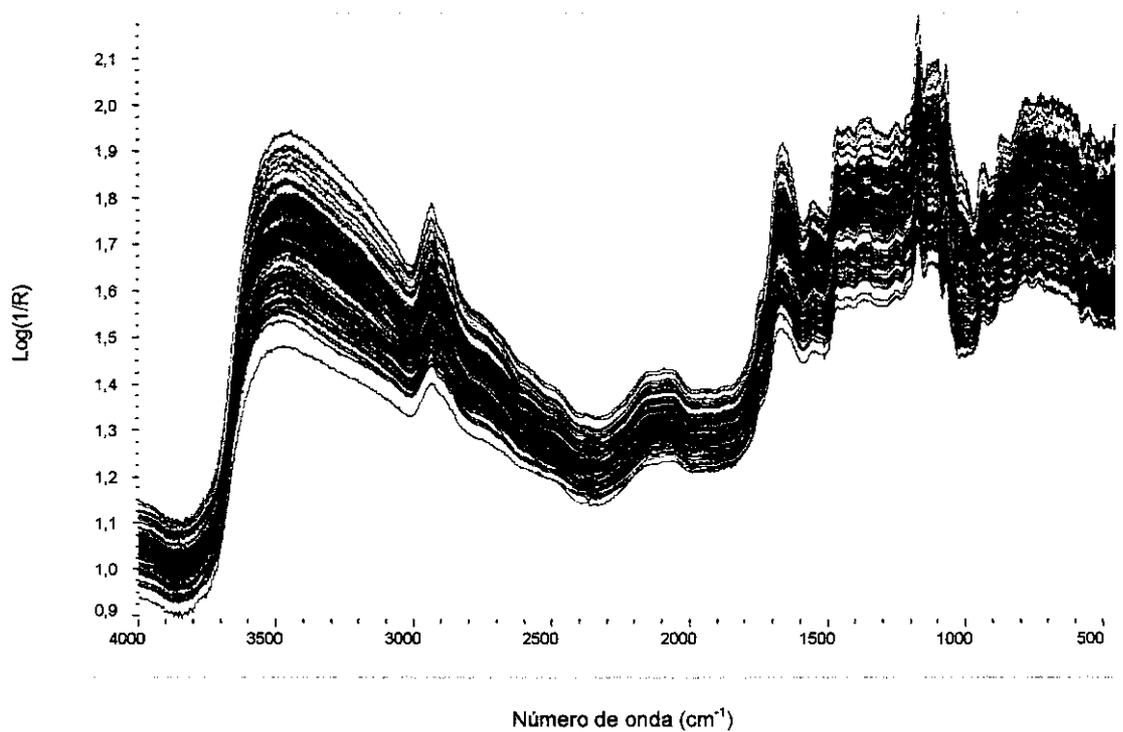


Figura 6.13 – Conjunto de espectros DRIFTS antes do tratamento.

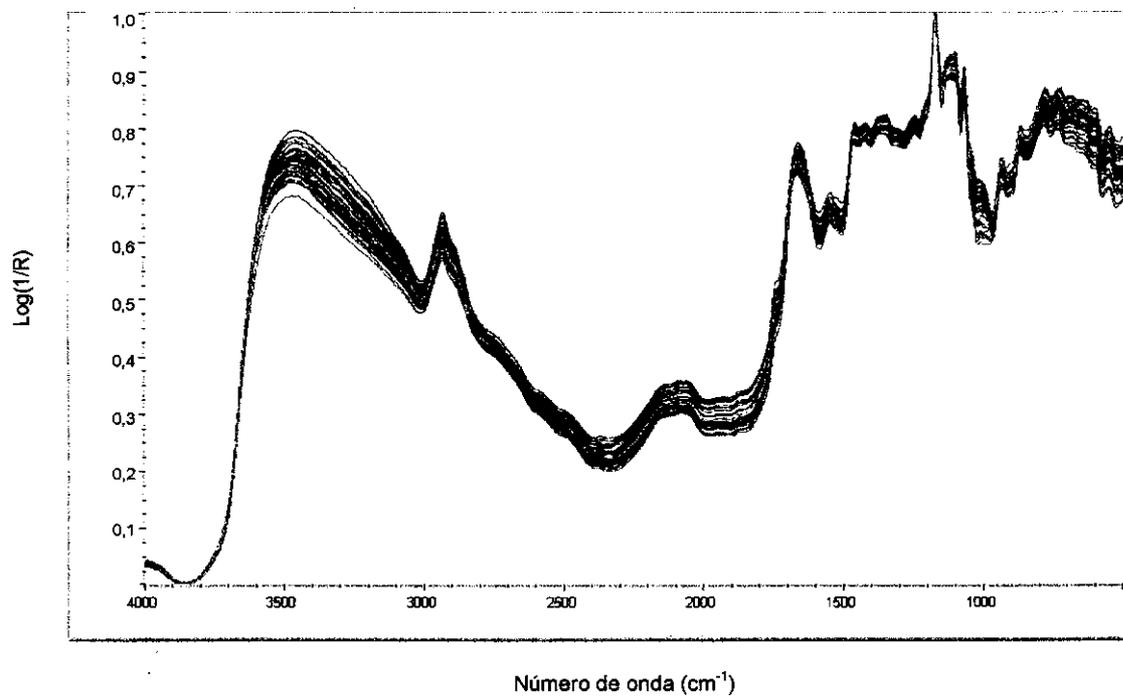


Figura 6.14 – Conjunto de espectros DRIFTS depois do conjunto de tratamentos descrito no fluxograma da figura 6.12.

6.6 Validação dos Espectros e dos Dados de Referência

O conjunto de valores médios dos métodos de referência (matriz de dados Y + valores de umidade) foi estudado no aplicativo Unscrambler 6.11 construindo-se um modelo PCA conforme gráficos bi-variados PC1xPC2 e PC1xPC3 apresentados na figura 6.15. Neste modelo a variância explicada foi 33% (PC1), 19% (PC2), 29% (PC3). A componente principal 1 (PC1), apresenta uma separação para um conjunto de três amostras (92, 93 e 94), que pertencem a um mesmo moinho, e que apresentam elevados valores nos teores de cinzas e proteína.

Modelos de análise por componentes principais também foram empregados para os conjuntos de dados espectrais a serem modelados. As principais informações obtidas encontram-se representadas nos gráficos bi-variados de escores nas figuras 6.16 e 6.17, para os conjuntos de espectros de NIR e HATR, respectivamente.

No caso dos espectros NIR, cujo PCA apresenta as seguintes variâncias explicadas 37% (PC1), 4% (PC2), 4% (PC3), 3% (PC4) e 3% (PC5), as amostras 36 e 59 indicaram diferenças significativas, ratificadas pelo conjunto de espectros, tanto brutos quanto com a primeira derivada.

Já para o conjunto de espectros HATR, cujo PCA apresenta as seguintes variâncias explicadas 23% (PC1), 17% (PC2), 7% (PC3), 7% (PC4) e 5% (PC5), as amostras 68 e 89 destacaram-se do conjunto, sendo também ratificadas pelos conjuntos de espectros sem tratamento e com primeira derivada.

O conjunto de espectros DRIFTS não apresentou, em nenhum de seus gráficos bi-variados, destaque de nenhuma amostra ou conjunto de amostras, sendo omitida sua apresentação.

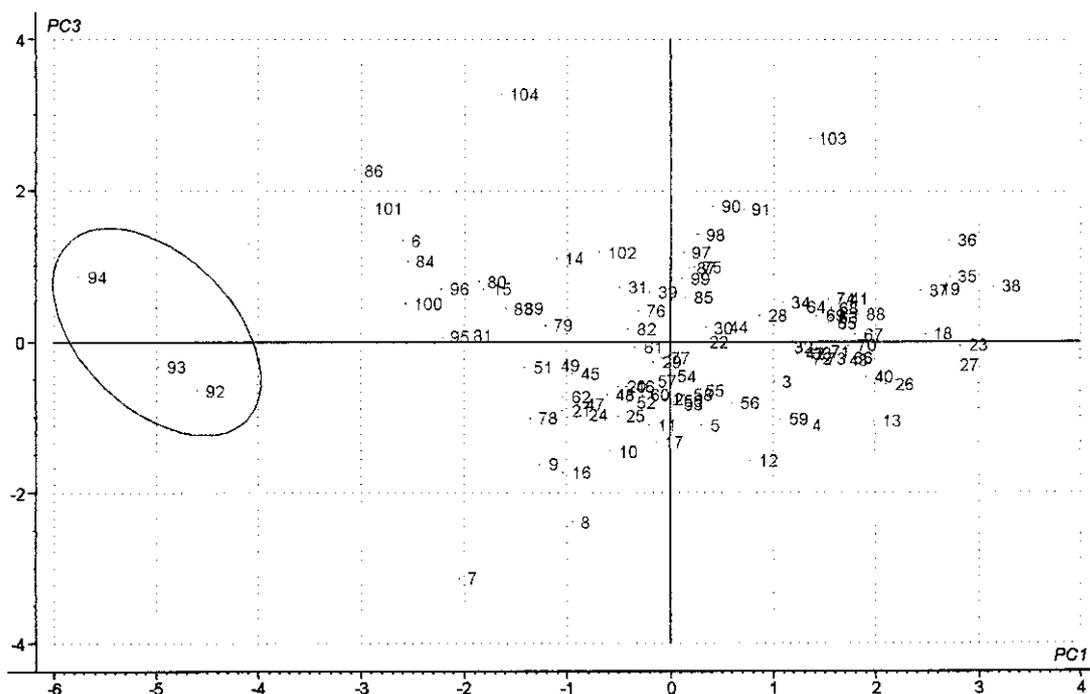
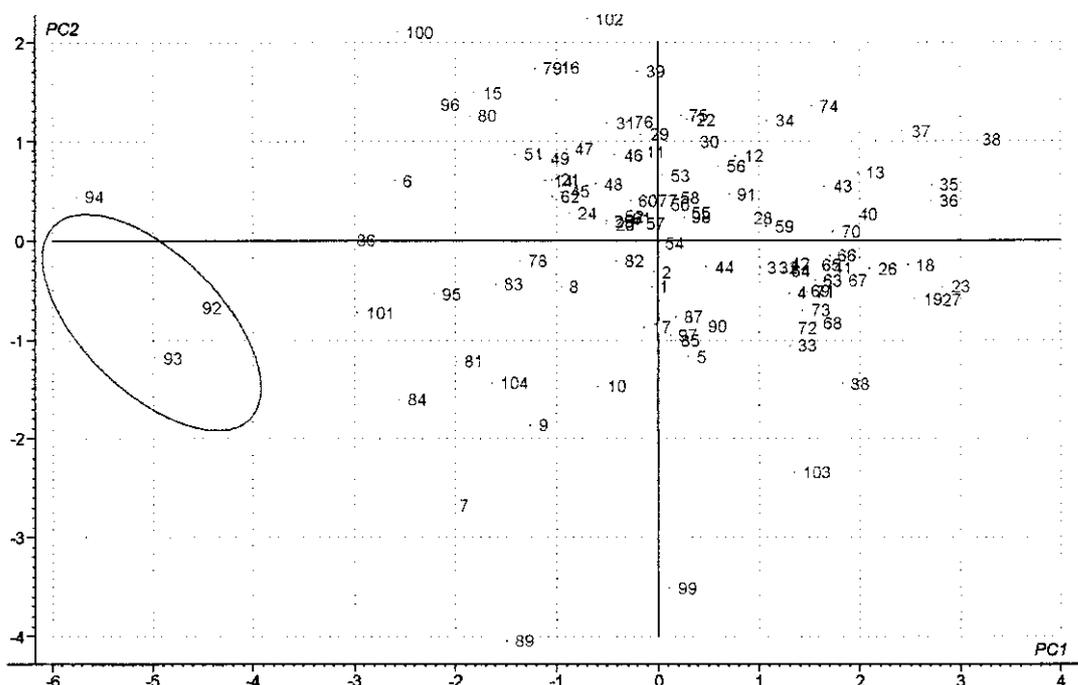


Figura 6.15 – Gráficos bi-variados de escores para o conjunto de valores médios dos métodos de referência.

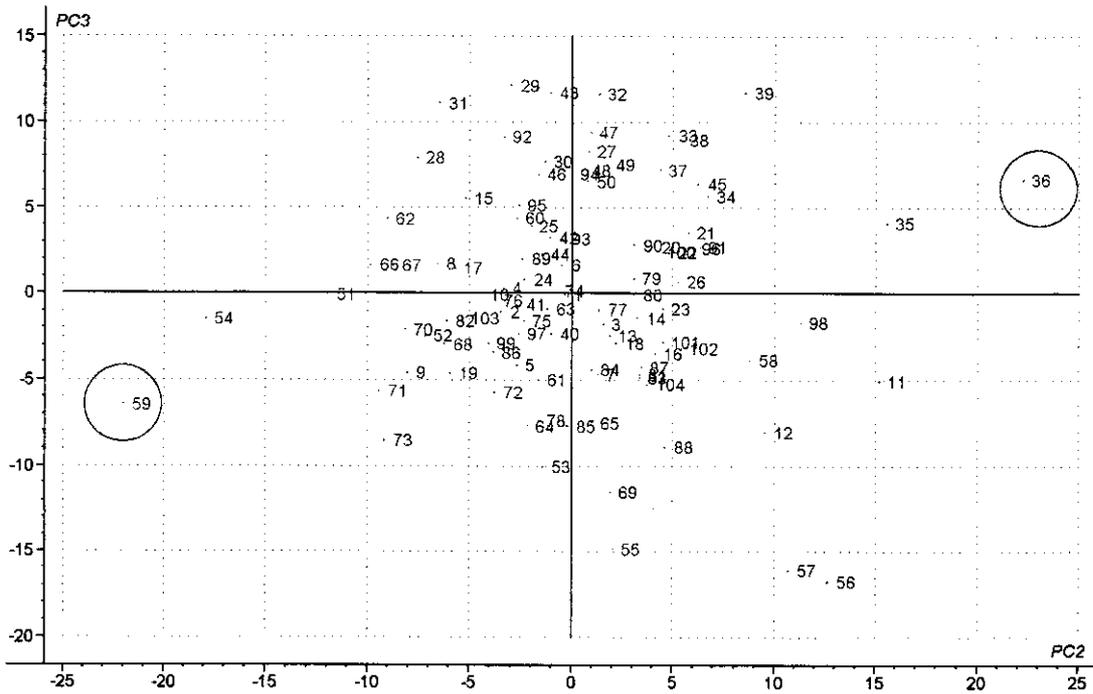
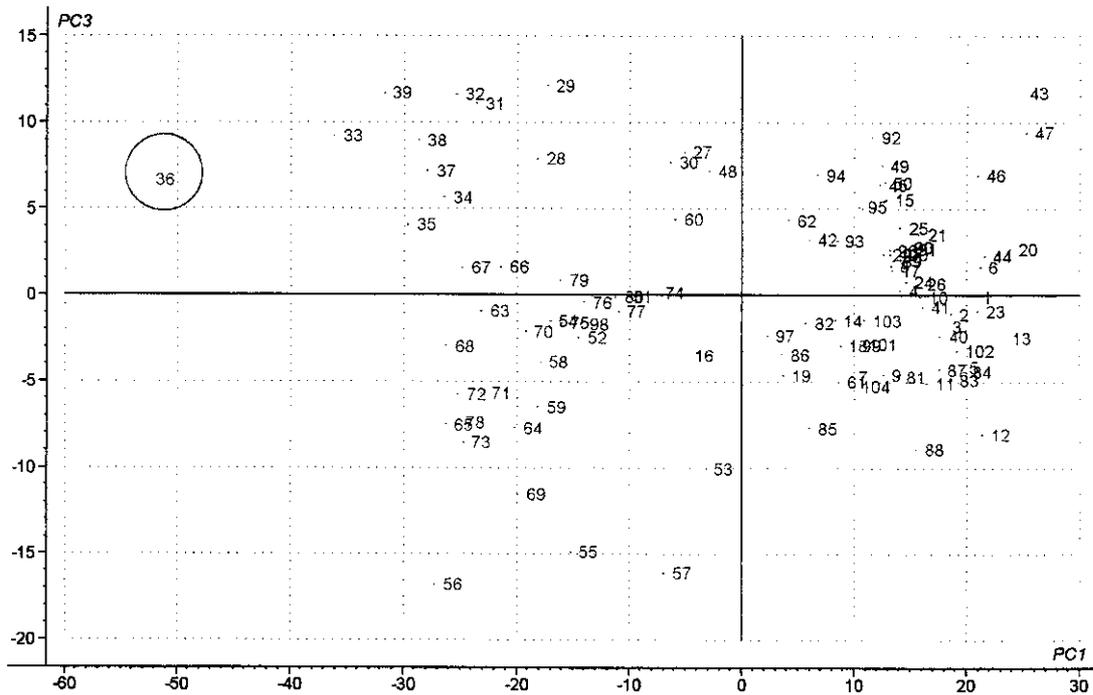


Figura 6.16 – Gráficos bi-variados de escores para o conjunto de espectros médios NIRR.

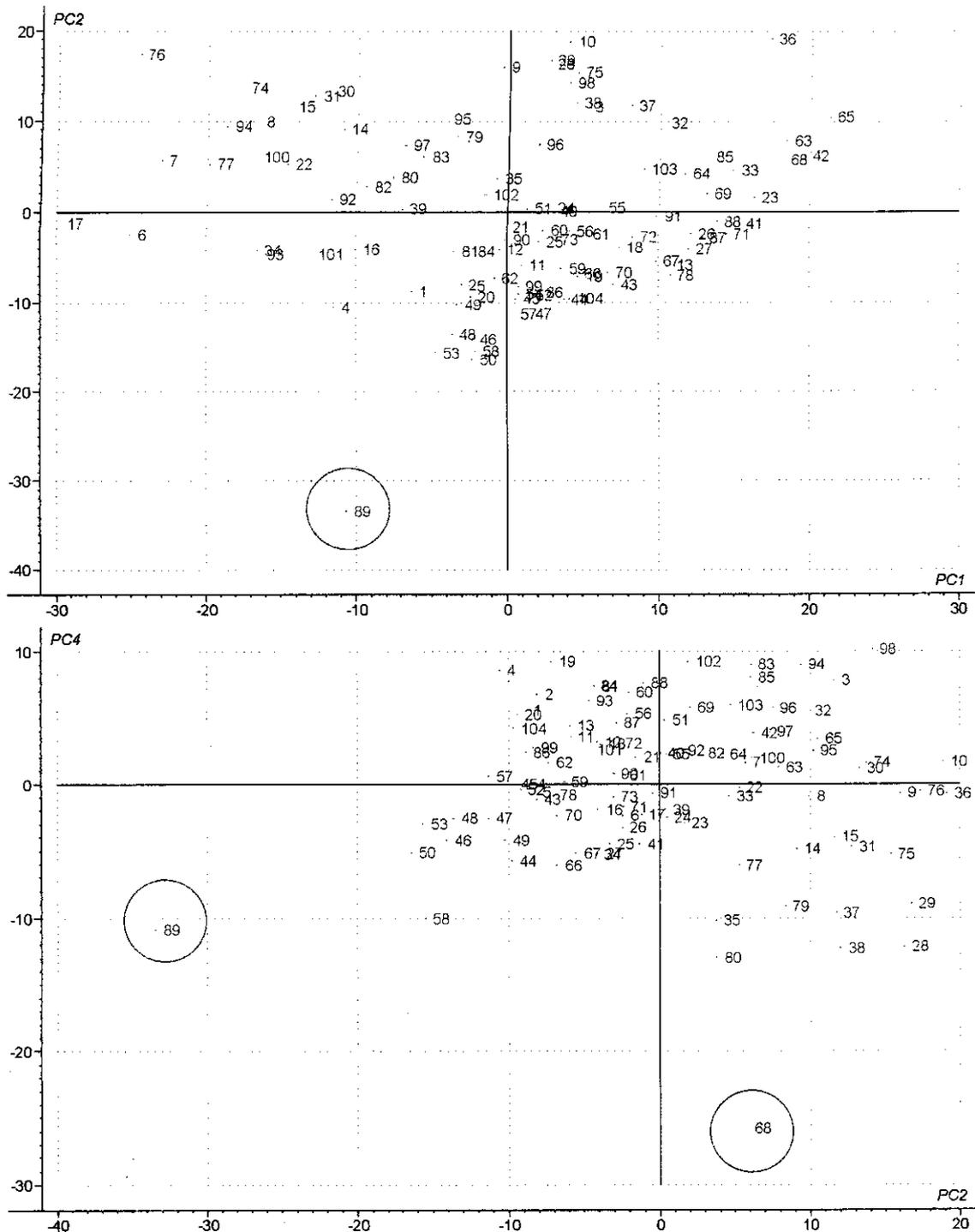


Figura 6.17 – Gráficos bi-variados de escores para o conjunto de espectros médios HATR.

6.7 Seleção das Amostras

Com base nos resultados apresentados para a validação dos dados, foram inicialmente retiradas as amostras 36, 59, 68 e 89, uma vez que seus espectros apresentaram anomalias que podem gerar problemas nas futuras modelagens a que serão submetidas os espectros.

As amostras de número 92, 93 e 94 foram mantidas, uma vez que os valores de teores de cinza e proteína encontravam-se dentro dos limites aceitos para amostras de farinha de trigo, não sendo atribuída nenhuma anomalia a sua determinação via método de referência.

As 100 amostras restantes foram divididas em dois conjuntos: 55 para a construção dos modelos de regressão e 45 para a validação. As amostras foram divididas de forma que a distribuição destas fosse tanto representativa para as diferentes farinhas oriundas dos diferentes moinhos, quanto para sua distribuição ao longo da variação característica de cada parâmetro a ser avaliado, dando-se preferência ao conjunto de calibração.

Para o conjunto de calibração todas as diferentes farinhas estão pelo menos representadas por uma das 100 amostras, bem como pelo menos uma das amostras com valores mais extremos (maior e menor) para cada parâmetro também compõe o conjunto de calibração.

A distribuição das amostras de calibração e de validação para cada parâmetro encontra-se representada nas figuras 6.18 a 6.20, para os teores de proteína, cinzas e número de queda, respectivamente.

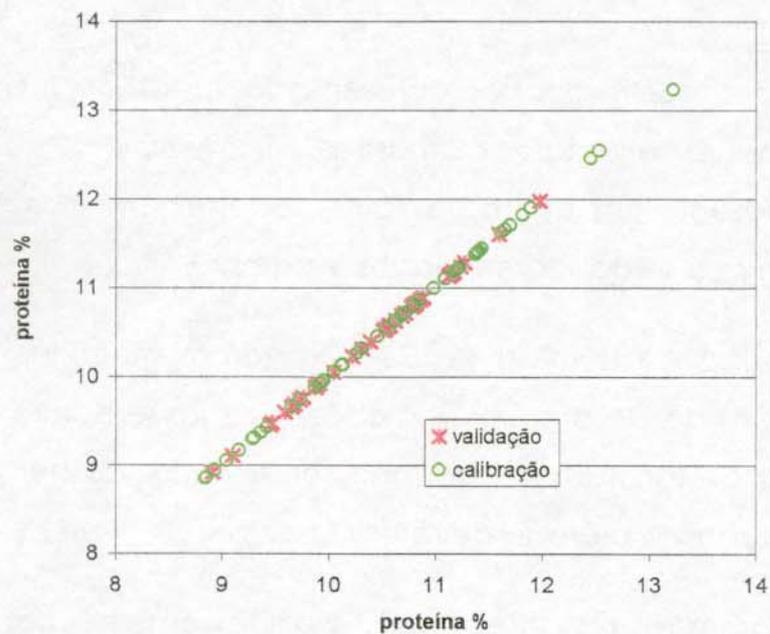


Figura 6.18 – Distribuição das amostras de calibração e de validação para o teor de proteína.

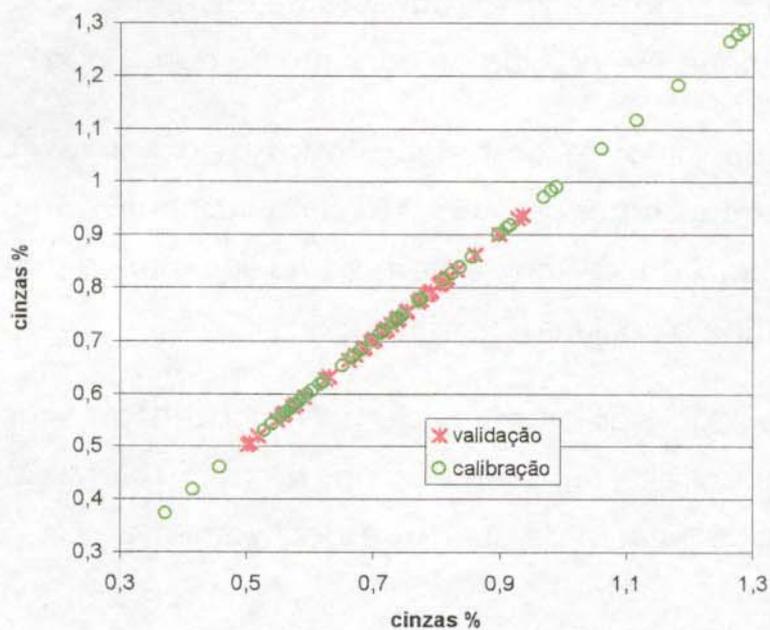


Figura 6.19 – Distribuição das amostras de calibração e de validação para o teor de cinza.

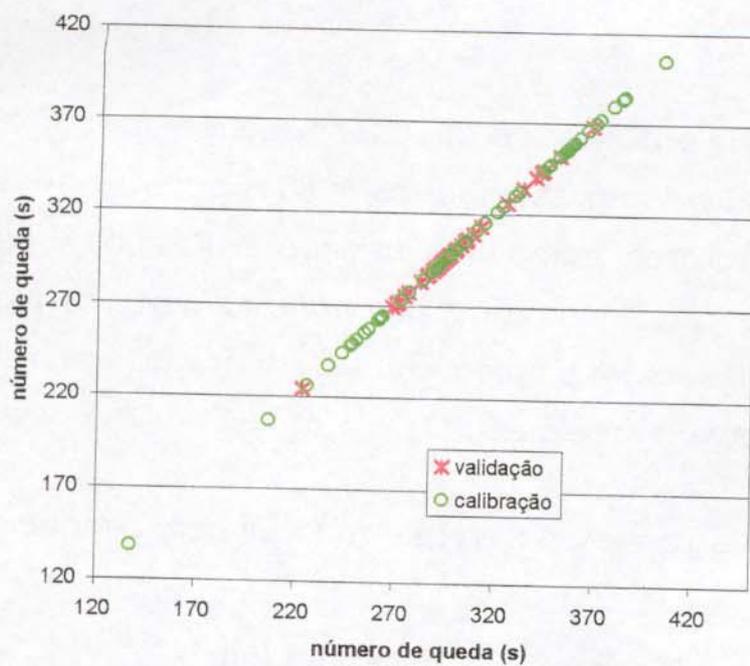


Figura 6.20 – Distribuição das amostras de calibração e de validação para número de queda.

6.8 Seleção de Variáveis

Ao conjunto de espectros de cada técnica espectroscópica foi aplicada a primeira derivada e foram retiradas as regiões onde se evidenciou pouca informação espectral, bem como aquelas onde o ruído no espectro devido ao vapor de água e ao dióxido de carbono estavam presentes. As regiões remanescentes e que foram empregadas na construção dos modelos tipo 1, 2 e 3 são apresentadas na tabela 6.5.

Tabela 6.5 - Regiões espectrais selecionadas para a modelagem.

Técnica	Regiões (cm⁻¹)
NIRR	5600 – 8800
DRIFTS	800-1800 ; 2500-3500
HATR	800-1800 ; 2500-3500

Para os modelos tipo 4 e 5 também foram retiradas as regiões onde se evidenciou pouca informação espectral, bem como aquelas onde o ruído no espectro devido ao vapor de água e ao dióxido de carbono estavam presentes. As regiões remanescentes e que foram otimizadas estão apresentadas na tabela 6.6.

Estudos preliminares empregando modelos de regressão evidenciaram que a utilização das regiões totais descritas na tabela 6.6, não resultam em modelos com boa correlação entre os espectros e os parâmetros de interesse.

É necessário melhorar estes resultados utilizando-se conjuntos de variáveis (comprimentos de onda) que diminuam os antagonismos e que realcem os efeitos sinérgicos resultantes da aplicação da análise multivariada. Desta forma os efeitos que determinados espectros, ou parte de espectros, conferem aos modelos são atenuados.

Diversos métodos de seleção de variáveis têm sido descritos na literatura ao longo dos anos, conforme foi amplamente descrito no capítulo anterior. Entretanto, vamos nos restringir aos métodos mais empíricos, além de serem de fácil implementação, podem ser manualmente monitorados.

Tabela 6.6 - Regiões espectrais selecionadas para a otimização.

Técnica	Regiões (cm⁻¹)	nº pontos
NIRR	5555 – 9000	894
DRIFTS	700-1800 ; 2800-3100	388
HATR	650-1800 ; 2800-3100	410

a) Método das variáveis importantes

Neste método foram utilizadas as informações dos coeficientes de regressão, dos espectros e dos *loadings* de um modelo PLS construído no Unscrambler com as regiões descritas na tabela 6.6. Um exemplo desta seleção é apresentado nas figuras 6.21; 6.22 e 6.23 para a avaliação da proteína e empregando espectros DRIFTS.

A partir da identificação das tendências apresentadas foram selecionadas sub-regiões, como é o caso da sub-região em destaque no espectro, que serão oportunamente avaliadas pelo emprego de um algoritmo de otimização.

Este procedimento foi adotado para cada uma das técnicas em relação a cada um dos parâmetros de interesse, totalizando de 10 a 15 regiões selecionadas em cada caso que encontram-se descritas nas tabelas 6.7 e 6.8.

O método *forward selection* foi utilizado para selecionar qual conjunto de regiões apresentava melhor correlação com cada um dos parâmetros de interesse. No apêndice B, encontramos um exemplo de como este algoritmo foi utilizado para otimizar o conjuntos de regiões realmente modelados.

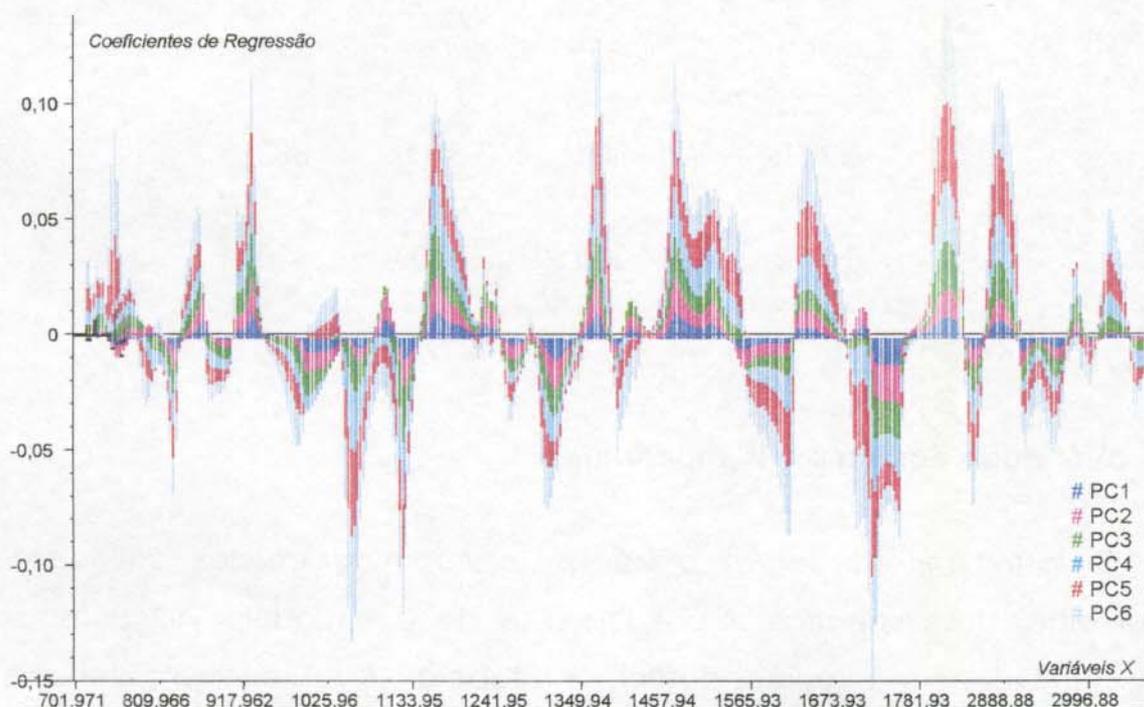


Figura 6.21 – Gráficos dos coeficientes de regressão da matriz X para o conjunto de dados DRIFTS.

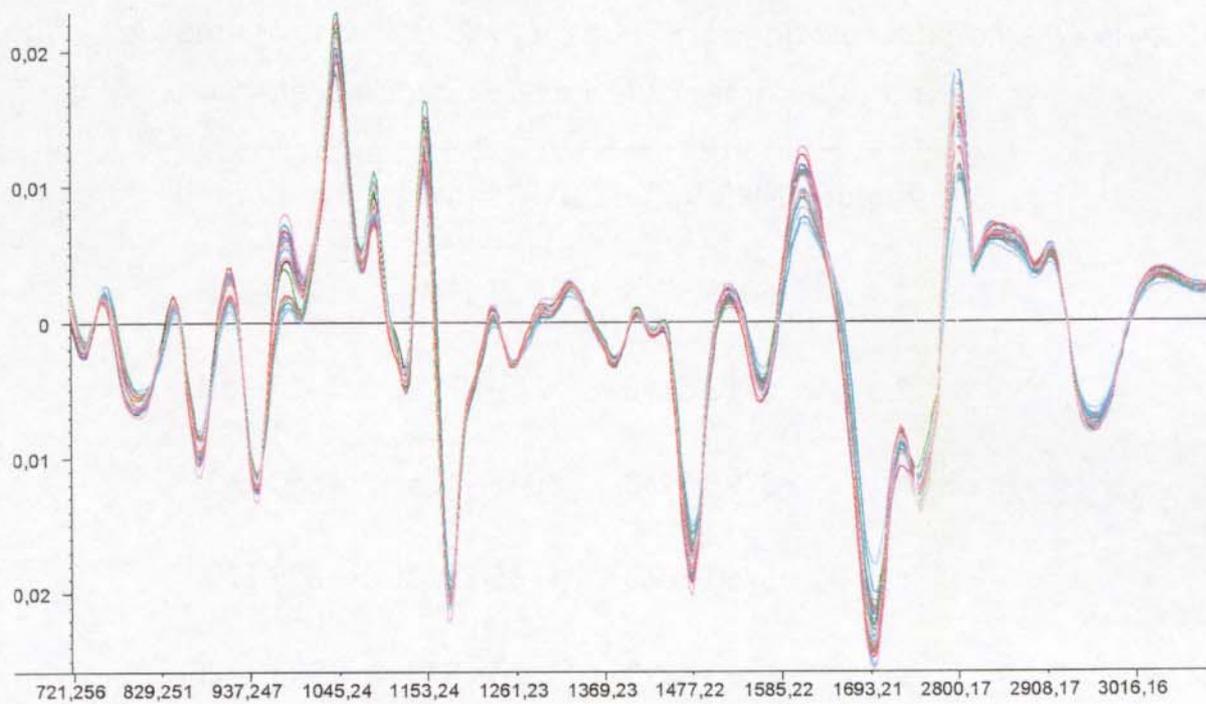


Figura 6.22 – Gráficos dos espectros com primeira derivada da matriz X para o conjunto de dados DRIFTS.

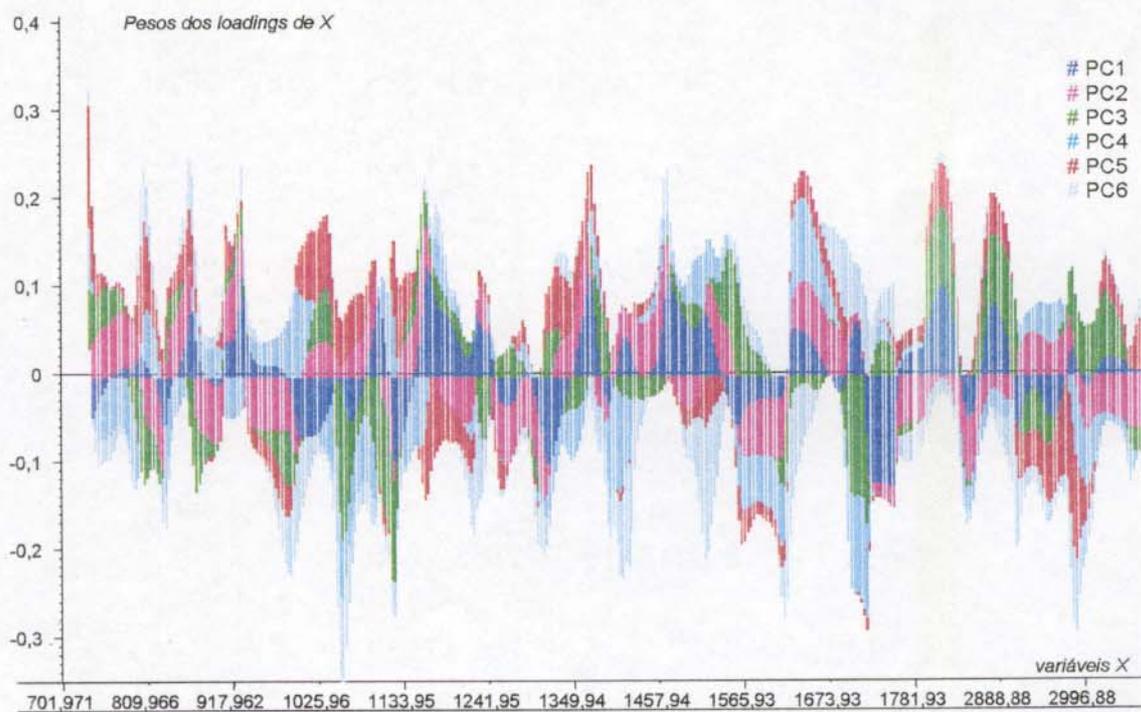


Figura 6.23 – Gráficos dos pesos dos loadings da matriz X para o conjunto de dados DRIFTS.

Tabela 6.7 – Regiões selecionadas pelo método das variáveis mais importantes para os parâmetros teores de cinzas e proteína.

Sub-região	NIRR (cm⁻¹)	DRIFTS (cm⁻¹)	HATR (cm⁻¹)
1	5615-5715	860-970	1250-1420
2	5715-5795	970-1065	1420-1550
3	5795-5950	1065-1165	1550-1635
4	5950-6060	1165-1270	1635-1700
5	6060-6390	1270-1370	1700-1750
6	6390-6670	1370-1420	1750-1800
7	6670-6940	1420-1460	2820-2875
8	6940-7240	1460-1585	2875-2920
9	7240-7400	1585-1650	2920-3010
10	7400-7570	1650-1740	3010-3090
11	7570-7600	1740-1785	
12	7600-8075	2800-2850	
13	8075-8110	2850-2885	
14	8110-8145	2885-2930	
15	8145-8180	2930-2990	

Tabela 6.8 – Regiões selecionadas pelo método das variáveis mais importantes para o parâmetro número de queda.

Sub-região	NIRR (cm⁻¹)	DRIFTS (cm⁻¹)	HATR (cm⁻¹)
1	5630-5785	810-845	735-860
2	5785-6060	845-1020	860-970
3	6460-6660	1020-1070	970-1020
4	6660-6940	1070-1165	1020-1070
5	7190-7235	1165-1270	1115-1290
6	7235-7400	1270-1370	1340-1370
7	7400-7515	1370-1460	1370-1415
8	7580-7825	1460-1585	1470-1525
9	8060-8170	1585-1650	1525-1650
10	8590-8630	1650-1740	1650-1745
11	8810-8850	1740-1780	1745-1820
12		2820-2915	1820-1855
13		2915-2940	2855-2895
14		2940-2985	2940-2975
15		2985-3035	3050-3100

b) Método da subtração dos espectros

Neste método os espectros médios de amostras com valores extremos de determinada propriedade são subtraídos, e a partir do perfil resultante, subdivididos em sub-regiões que serão posteriormente modeladas, buscando a otimização dos modelos para cada parâmetro estudado. Na figura 6.24, podemos observar como os perfis de subtração (sinais em vermelho) de amostras de farinha contendo alto e baixos valores de proteína, estão relacionados com o espectro de uma amostra genérica obtido por HATR (espectro em azul).

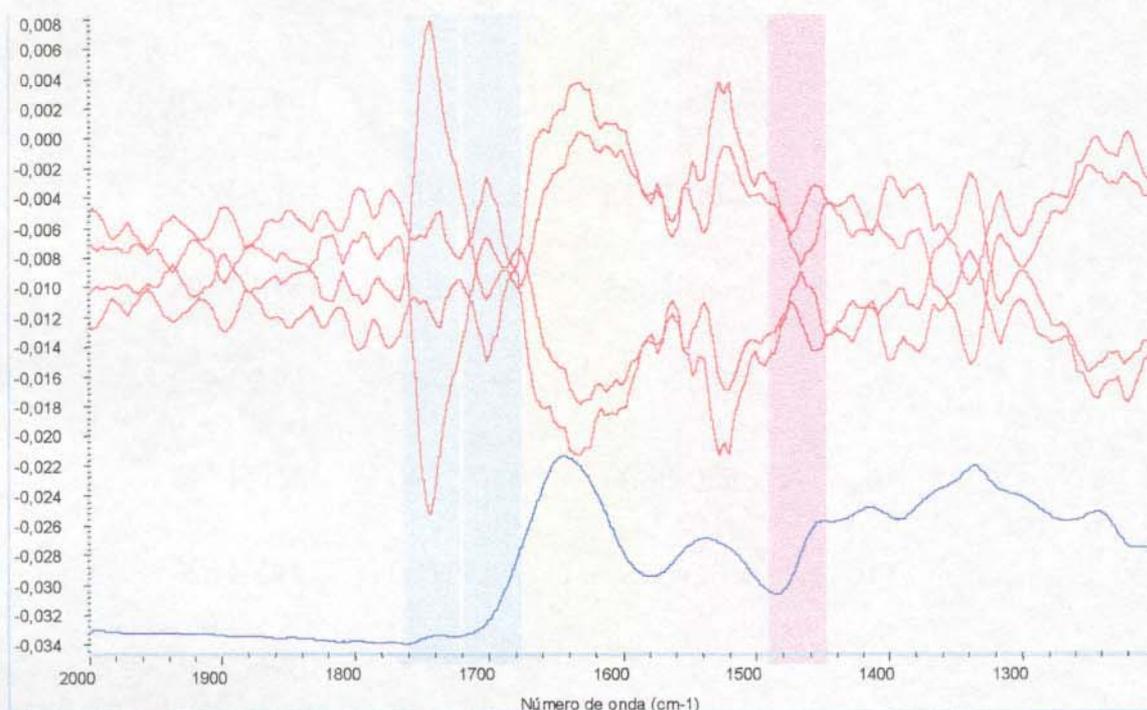


Figura 6.24 – Conjunto de perfis de subtração de amostras com altos e baixos teores de proteína comparado ao espectro HATR de uma amostra de farinha.

Nas tabelas 6.9 e 6.10 são apresentadas as regiões selecionadas para cada parâmetro combinada com as três técnicas espectroscópicas em estudo.

Tabela 6.9 – Regiões selecionadas pelo método da subtração de espectros para os parâmetros teores de cinzas e proteína.

Nº	NIRR (cm⁻¹)	DRIFTS (cm⁻¹)	HATR (cm⁻¹)
1	5600-5700	950-1030	965-1045
2	5700-5800	1080-1140	1045-1080
3	5800-6000	1150-1210	1080-1145
4	6100-6600	1505-1585	1460-1505
5	6600-7100	1650-1700	1505-1550
6	7115-7330	1715-1760	1555-1605
7	7350-7940	2840-2875	1605-1665
8	8100-8415	2875-2905	1725-1770
9	8415-8760	2905-2950	2830-2865
10	8760-8975	2950-2990	2885-2950

Tabela 6.10 – Regiões selecionadas pelo método da subtração de espectros para parâmetro o número de queda.

Nº	NIRR (cm⁻¹)	DRIFTS (cm⁻¹)	HATR (cm⁻¹)
1	5590-5665	800-860	990-1040
2	5665-5880	860-980	1055-1080
3	5960-6060	980-1050	1080-1145
4	6125-7100	1050-1145	1455-1490
5	7130-7275	1145-1190	1505-1570
6	7275-7360	1270-1475	1590-1640
7	7350-7500	1590-1675	1640-1725
8	7655-7790	1675-1715	1725-1755
9	8265-8325	1715-1775	2840-2865
10	8400-8500	2850-2990	2910-2940

6.8.1 Seleção de Variáveis e a Atribuição de Sinais

Uma das vantagens do emprego da modelagem PLS é que não se faz necessário um conhecimento preliminar de quais comprimentos de onda devam ser empregados para determinar a correlação entre os espectros e os parâmetros de interesse.

Esse fato nos permite buscar a correlação entre os dados espectroscópicos (matriz X) e as variáveis dependentes (matriz Y) de parâmetros como o teor de cinza, cujos precursores devem ser de natureza complexa e que provavelmente seus espectros estejam sobrepostos e com intensidades muito inferiores àqueles compostos que são mais representativos na farinha, como por exemplo o amido.

Entretanto, o histórico apresentado pelos profissionais que se dedicaram a utilizar a Quimiometria para construção de modelos de regressão [122], por longo tempo empregaram processos de seleção dos comprimentos de onda, pois a utilização do método MLR fazia necessário uma redução (seleção) dos comprimentos de onda que seriam mais representativos na calibração de determinada propriedade (ou parâmetro).

Neste sentido, a literatura apresenta diversas tabelas para a identificação ou atribuição das bandas que representam as vibrações moleculares de determinadas espécies [122, 303]. Historicamente estas atribuições estão concentradas na região do infravermelho próximo onde a técnica MLR foi intensamente estudada, e em particular o parâmetro proteína exaustivamente explorado [304].

Quanto a região do infravermelho médio, pouco se tem feito no sentido de atribuir as frequências selecionadas por algum processo de otimização, provavelmente em virtude na natureza mais complexa do conjunto de sinais

que representam as vibrações moleculares de matrizes complexas, como é o caso da farinha.

Neste trabalho, foi verificado que independente do método empregado para a seleção de variáveis, as regiões do infravermelho médio compreendidas entre 2800-3000 cm^{-1} (referentes as vibrações de deformação axial das ligações C-H), sempre estiveram entre àquelas de maior contribuição na modelagem do parâmetro proteína, isto é, levavam a melhores valores de R^2 e a menores valores de SEV.

Outra constatação importante foi que nem sempre as sub-regiões compreendidas entre 1600-1700 cm^{-1} , quando selecionadas, levaram a melhora do modelo de calibração para a proteína. Apesar de serem regiões representativas para o estiramento da carbonila, presente em todas as moléculas de proteína.

Outras considerações tornam-se difíceis em função das faixas (sub-regiões) otimizadas serem de largura considerável chegando a compreender até mais de 20 comprimentos de onda.

Uma otimização pontual, isto é, de cada comprimento de onda não foi experimentada, em função do aplicativo *Turbo Quant* não suportar um processo de iterações cíclico (*looping*), bem como não permitir um número elevado de sub-regiões independentes.