



UNIVERSIDADE FEDERAL DE RORAIMA
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA

JOÃO DOS SANTOS PANERO

DISSERTAÇÃO DE MESTRADO

DETERMINAÇÃO DE PROTEÍNA, ÓLEO E UMIDADE POR ESPECTROSCOPIA NIR
EM GRÃOS DE SOJA DO ESTADO DE RORAIMA.

Boa Vista
Novembro de 2007

JOÃO DOS SANTOS PANERO

DETERMINAÇÃO DE PROTEÍNA, ÓLEO E UMIDADE POR ESPECTROSCOPIA NIR
EM GRÃOS DE SOJA DO ESTADO DE RORAIMA.

Dissertação apresentada ao Programa de Pós-graduação em Química da Universidade Federal de Roraima, como pré-requisito para obtenção do título de Mestre em Química.

Orientador: Prof^o. Dr. Henrique Eduardo B. da Silva

Co-orientador: Prof^o. Dr. Jarbas José R. Rohwedder

Área de concentração: Química Ambiental

Boa Vista
Novembro de 2007

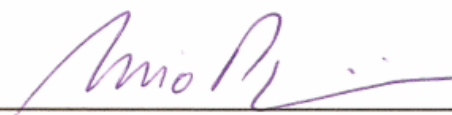
JOÃO DOS SANTOS PANERO

DETERMINAÇÃO DE PROTEÍNA, ÓLEO E UMIDADE POR ESPECTROSCOPIA NIR EM GRÃOS DE SOJA DO ESTADO DE RORAIMA

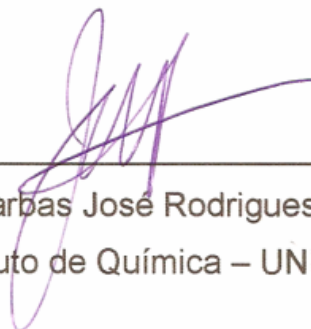
Dissertação apresentada como Pré-requisito para a conclusão do Curso de mestrado em Química da universidade Federal de Roraima, defendida em 21 de novembro de 2008 e avaliada pela seguinte banca examinadora:



Prof. Dr. Henrique Eduardo Bezerra da Silva
Orientador/Departamento de Química -UFRR



Prof. Dr. Celio Pasquini
Instituto de Química – UNICAMP



Prof. Dr. Jarbas José Rodrigues Rohwedder
Instituto de Química – UNICAMP

Aos meus pais e irmãos por estarem sempre ao meu lado; aos meus amigos pelo incentivo, amizade e convívio.

AGRADECIMENTOS

Primeiramente à Deus, por minha vida;

Aos meus pais e irmãos pelo quanto se sacrificaram no decorrer da minha vida.

À salvadora dos meus dias...;

Ao Prof. Dr. Henrique pela orientação;

Ao Prof. Dr. Jarbas pela co-orientação;

Aos companheiros de curso pelo convívio e descontração durante esses anos;

Aos professores do PPGQ pelo compartilhamento de conhecimentos;

À EMBRAPA Roraima, à UFRR e à Unicamp por disponibilizarem toda estrutura física, tecnológica e humana para realização deste trabalho;

À CAPES pela bolsa concedida;

Ao CNPq, à FINEP, à FAPESP e ao PROCAD/CAPES pelo auxílio durante a pesquisa;

E à todas as outras pessoas envolvidas direta ou indiretamente neste trabalho.

“Só se vê bem com o coração. O essencial é invisível aos olhos”.
(Antoine de Saint-Exupéry)
Livro O Pequeno Príncipe

RESUMO

Neste trabalho são combinadas as potencialidades das técnicas quimiométricas e espectroscópica (NIR - infravermelho próximo) visando à determinação dos teores de proteína, óleo e umidade em grãos de sojas cultivadas no campo experimental da Embrapa Roraima, para os quais utilizou-se como referência o método Kjeldahl, o método de extração por Soxhlet e o método da estufa, respectivamente. Na construção dos modelos de calibração NIR estudou-se vários tipos de pré-processamento dos dados espectrais, faixas espectrais e conjuntos de amostras. O método de regressão usado na construção dos modelos foi o PLS (mínimos quadrados parciais) utilizando dados centrados na média para um conjunto de 90 espectros (calibração e validação interna). A avaliação dos modelos foi feita observando os valores de coeficiente de correlação (R) e os erros quadrados médios. Sendo os valores de previsão estimados para os demais 45 espectros não empregados na calibração. Para a determinação do teor de proteína, o melhor resultado foi obtido empregando correção multiplicativa de sinal (MSC) na faixa 1510-1980 nm, tendo RMSEP igual a 0,66. Para a determinação do teor de óleo o melhor resultado foi obtido empregando MSC com correção de linha base (BL) na faixa 1010-2500 nm, com RMSEP de 0,40. E para a determinação do teor de umidade, o melhor modelo foi obtido empregando a Transformação Padrão Normal de Variação (SNV) com correção de linha base (BL) na região de combinação das faixas 1870-2020 nm e 1400-1500 nm, que forneceu RMSEP de 0,12. Os erros relativos médios entre os valores previstos e medidos foram inferiores a 2 %. Com estes resultados concluiu-se que podem ser obtidos bons modelos visando à previsão dos teores de proteína, óleo e umidade em grãos de sojas cultivadas em Roraima, e que a técnica espectroscópica NIR aliada à regressão por PLS é adequada por permitir além de determinações rápidas e precisas, a não destruição das amostras nem a utilização de reagentes e por não gerar resíduos que sejam nocivos ao ambiente.

Palavras-chave: Soja, Espectroscopia no Infravermelho Próximo, PLS, Umidade, Óleo, Proteína.

ABSTRACT

In this study the potentialities of (NIR – Near Infrared) spectroscopy and chemometrics techniques are combined aiming to determination of protein, oil and moisture in whole-grain soybean cultivated in the experimental field of the Roraima Embrapa, for which it was used as reference the Kjeldahl method, the extraction method by Soxhlet and the oven method, respectively. In the development of the models were studied several pre-processing types of the spectra data, spectral range and sample set. The regression method used in the development of the models was PLS (Partial Least Square) using mean-centered data, for a set of 90 spectra (calibration and cross validation). The evaluation of the models was made observing the values of correlation coefficient (R) and the root mean square error. The prevision values were estimated for the other 45 spectra nor used in the calibration set. For the determination of protein level, best results were obtained using Multiplicative Signal Correction (MSC) in the 1510-1980 nm range, obtaining RMSEP with values 0.66. For the determination of oil level, best results were obtained using MSC with transform Base Line (BL) in the 1110-2500 nm range, obtaining RMSEP with values 0.40. And for the determination of moisture level, best results were obtained using Standard Normal Variate Transformation (SNV) with BL, in the 1870-2020 nm and 1400-1500 nm combined range, obtaining RMSEP with values 0.12. The medium relative errors among predicted values and the measured values were inferior to 2%. With these results conclude that good models aiming the prediction of protein, oil and moisture levels in the whole-grain soybean from Roraima can be obtained. And that the technique NIR allied to the PLS regression is adequate to allow besides fast and precise determinations, the not destruction of the samples nor the use of reagents use and for not create wastes which are harmful to the environment.

Key-words: Soybean, Near Infrared Spectroscopy, PLS, Moisture, Oil, Protein.

LISTA DE TABELAS

TABELA 01 - Comparação da média de produtividade de algumas lavouras no Estado de Roraima com a média nacional.....	16
TABELA 02 - Regiões espectrais do Infravermelho.....	18
TABELA 03 - Atribuição das regiões espectrais às possíveis transições vibracionais.....	21
TABELA 04 - Espectros usados no conjunto de calibração e previsão – Conjunto 1.....	57
TABELA 05 - Espectros usados no conjunto de calibração e previsão – Conjunto 2.....	57
TABELA 06 - Espectros usados no conjunto de calibração e previsão – Conjunto 3.....	58
TABELA 07 - Resultados médios e RSD dos teores de proteína, óleo e umidade (n=3).....	69
TABELA 08 - Conjuntos de espectros escolhidos para o desenvolvimento dos modelos.	79
TABELA 09 - Porcentagem de variância capturada por cada VL do modelo PLS.....	84
TABELA 10 - Comparação dos resultados de umidade (%) para as amostras do conjunto de previsão obtidos pelos métodos da estufa e de espectroscopia NIR, aplicado o modelo de regressão de melhor ajuste obtido.....	84
TABELA 11 - Porcentagem de variância capturada por cada VL do modelo PLS.....	87
TABELA 12 - Comparação dos resultados de óleo (%) para as amostras do conjunto de previsão obtidos pelos métodos de extração por Soxhlet e de espectroscopia NIR, aplicado o modelo de regressão de melhor ajuste obtido.....	88
TABELA 13 - Porcentagem de variância capturada por cada VL do modelo PLS.....	90
TABELA 14 - Comparação dos resultados de proteína total (%) para as amostras do conjunto de previsão obtidos pelos métodos de Kjeldahl e de espectroscopia NIR, aplicado o modelo de regressão de melhor ajuste obtido.....	91

LISTA DE FIGURAS

FIGURA 01 - Interação da radiação NIR com partículas sólidas na amostra.....	22
FIGURA 02 - Diagrama de bloco de um espectrômetro.....	24
FIGURA 03 - Configuração da instrumentação básica do NIR: a) Transmitância; b) Reflectância.....	26
FIGURA 04 - Representação da reflexão especular e difusa de uma onda eletromagnética em uma amostra particulada.....	27
FIGURA 05 - Esquema da organização dos dados para as etapas de calibração e previsão...32	32
FIGURA 06 - Representação: (a) Matriz X - Espectros NIR e (b) Matriz Y - Teores do analito.....	36
FIGURA 07 - Gráfico dos <i>scores</i> (PC1 <i>versus</i> PC2) a) para 2 variáveis e b) para 3 variáveis.....	39
FIGURA 08 - Representação gráfica do cálculo para centrar os dados na média.....	44
FIGURA 09 - Processo de filtragem utilizando média móvel.....	47
FIGURA 10 - Sistema Soxhlet em refluxo usado para a determinação de óleo.....	65
FIGURA 11 - Espectros originais de todas as amostras de grãos de soja.....	70
FIGURA 12 - Faixa de comprimento de onda de 1400 a 1610 nm.....	72
FIGURA 13 - Faixas de comprimentos de ondas de 1400 a 1610 nm, 2040 a 2080 nm, 2160 a 2200 nm, e 2450 a 2490 nm.....	73
FIGURA 14 - Faixa de comprimento de onda de 1870 a 2020 nm.....	74
FIGURA 15 - Faixas de comprimentos de ondas de 1400 a 1500 nm e de 1870 a 2020 nm..	74
FIGURA 16 - Faixas de comprimentos de ondas de 1400 a 1500 nm, 1740 a 1840 nm e 1870 a 2020 nm.....	75
FIGURA 17 - Faixa de comprimento de onda de 2231 a 2427 nm.....	76
FIGURA 18 - Faixas de comprimentos de ondas de 1402 a 1418 nm, 2054 a 2087 nm, 2290 a 2331 nm, e 2358 a 2402 nm.....	77
FIGURA 19 - Faixa de comprimento de onda de 1510 a 1980 nm.....	77
FIGURA 20 - Faixa de comprimento de ondas de 1010 a 2500 nm.....	78
FIGURA 21 - Faixas de comprimento de ondas: a) de 1413 a 1010 nm b) de 2500 a 1413 nm c) de 2250 a 1950 nm com 1545 a 1350 nm d) de 2100 a 1690 nm com 1495 a 1160 nm.....	79
FIGURA 22 - Exemplo de um dos gráficos do número de variáveis latentes em função da variância.....	80

FIGURA 23 - Exemplo de um dos gráficos do número de variáveis latentes em função do <i>PRESS</i>	81
FIGURA 24 - Representação do comportamento do modelo em relação a presença de uma amostra anômala.....	82
FIGURA 25 - Gráfico de resíduos Student pela <i>leverage</i>	82
FIGURA 26 - Conjunto de espectros do modelo de calibração obtido após MSC empregado na determinação da umidade.....	83
FIGURA 27 - Conjunto de espectros do modelo de calibração obtido após MSC e BL empregado na determinação do óleo.....	87
FIGURA 28 - Conjunto de espectros do modelo de calibração obtido após SNV e BL empregado na determinação da proteína.....	90

LISTA DE SIGLAS

- BL – Correção de Linha Base (*Transform Base Line*).
- CV – Validação Cruzada (*Cross Validation*).
- EMSC – Correção do Espalhamento Multiplicativo Estendido (*Extended multiplicative signal correction*).
- FIR – Infravermelho Distante (*Far Infrared*).
- HCA – Análise de Agrupamentos Hierárquicos (*Hierarchical Cluster Analysis*).
- MIR – Infravermelho Médio (*Middle Infrared*).
- MSC – Correção do Espalhamento Multiplicativo (*Multiplicative Signal Correction*).
- NIR – Infravermelho Próximo (*Near Infrared*).
- NIRRS – Espectroscópica de Infravermelho Próximo com Reflexão Difusa (*Near Infrared Reflectance Spectroscopy*).
- PC – Componente Principal (*Principal Component*).
- PCA – Análise por Componentes Principais (*Principal Component Analysis*).
- PCR – Regressão nas Componentes Principais (*Principal Component Regression*).
- PLS – Mínimos Quadrados Parciais (*Partial Least-Square*).
- PRESS – Somatório Quadrático dos Erros de Predição (*Prediction Residual Error Sum of Squares*).
- R – Coeficiente de correlação.
- RMSEC – Raiz Quadrada dos Erros Médios de Calibração (*Root Mean Square Error Calibration*).
- RMSECV – Raiz Quadrada dos Erros Médios de Validação Cruzada (*Root Mean Square Error Cross Validation*).
- RMSEP – Raiz Quadrada dos Erros Médios de Previsão (*Root Mean Square Error Prediction*).
- SNV – Transformação Padrão Normal de Variação (*Standard Normal Variate Transformation*).
- VL – Variável Latente.

SUMÁRIO

1	INTRODUÇÃO	14
1.1	SOJA.....	14
1.1.1	Soja em Roraima	15
1.2	ESPECTROSCOPIA NO INFRAVERMELHO.....	17
1.2.1	NIR e Aplicações	20
1.2.2	Reflexão Difusa	25
1.3	QUIMIOMETRIA.....	29
1.3.1	Modelos de Regressão	30
1.3.1.1	Calibração.....	31
1.3.1.2	Previsão.....	33
1.3.2	Métodos Descritivos e Exploratórios e métodos de Calibração Multivariada 34	
1.3.2.1	<i>PCA</i>	34
1.3.2.2	<i>PLS</i>	39
1.4	PROCESSAMENTO E TRATAMENTO DOS DADOS ESPECTRAIS.....	42
1.4.1	Dados centrados na média	43
1.4.2	Correções de Linha base	44
1.4.3	Correção do Espalhamento Multiplicativo (MSC)	44
1.4.4	Filtro de Média Móvel	45
1.4.5	Derivadas	47
1.4.6	Transformação Padrão Normal de Variação (SNV)	48
1.4.7	Correção do Espalhamento Multiplicativo Estendido (EMSC)	49
1.5	QUÍMICA VERDE.....	49
2	OBJETIVO	51
3	JUSTIFICATIVAS	52
4	PARTE EXPERIMENTAL	53
4.1	AMOSTRAGEM.....	53
4.2	MÉTODOS DE REFERÊNCIA.....	53
4.2.1	Determinação da Umidade	53
4.2.1.1	Reagentes.....	53
4.2.1.2	Materiais e equipamentos.....	53
4.2.1.3	Procedimentos experimentais.....	53
4.2.2	Determinação do Óleo	54
4.2.2.1	Reagentes.....	54
4.2.2.2	Materiais e equipamentos.....	54
4.2.2.3	Procedimentos experimentais.....	54
4.2.3	Determinação da Proteína	54
4.2.3.1	Reagentes.....	54
4.2.3.2	Materiais e equipamentos.....	55
4.2.3.3	Procedimentos experimentais.....	55
4.3	OBTENÇÃO DOS ESPECTROS.....	56
4.4	DESENVOLVIMENTOS DOS MODELOS DE CALIBRAÇÃO.....	56
4.4.1	Seleção de faixa espectral	56
4.4.2	Seleção de espectros	56
4.4.3	Número de variáveis latentes	58
4.4.4	Amostras anômalas	59
4.5	OTIMIZAÇÃO E VALIDAÇÃO DOS MODELOS DE CALIBRAÇÃO.....	60
5	RESULTADOS E DISCUSSÃO	64
5.1	TEORES DE UMIDADE, ÓLEO E PROTEÍNA PELOS MÉTODOS DE REFERÊNCIA.....	64

5.2	SELEÇÃO DA FAIXA ESPECTRAL E AMOSTRAS.....	70
5.3	NÚMERO DE VARIÁVEIS LATENTES.....	79
5.4	AMOSTRAS ANÔMALAS.....	81
5.5	DETERMINAÇÃO DA UMIDADE POR ESPECTROSCOPIA NIR.....	83
5.6	DETERMINAÇÃO DO ÓLEO POR ESPECTROSCOPIA NIR.....	86
5.7	DETERMINAÇÃO DA PROTEÍNA POR ESPECTROSCOPIA NIR.....	89
6	CONSIDERAÇÕES FINAIS	94
7	PERSPECTIVAS FUTURAS	96
8	REFERÊNCIAS	97

1 INTRODUÇÃO

1.1 SOJA

A soja, originária da China, é uma planta que pertence à família das leguminosas denominada cientificamente de *Glycine max* (L). Segundo as escrituras chinesas, o consumo humano desta leguminosa é datado de 30 a.C. Introduzida na Europa no século XVIII e, no início do século XIX nos Estados Unidos da América, a soja tem seu registro datado no Brasil em 1882, na Bahia (YUAN et al., 2002; PENHA, 2003; SIMAS, R., 2005; ZARKADAS et al., 2007).

A soja é o produto de maior importância agrícola no Brasil, o qual ocupa o segundo lugar entre os grandes produtores mundiais, atrás apenas dos EUA (GIANLUPPI et al., 2000a; PENHA, 2003; SMIDERLE, 2003).

Sendo um dos alimentos mais completos conhecidos pelo homem, a soja é considerada um alimento funcional, fornece nutrientes ao organismo e traz benefícios para a saúde. É uma fonte rica em proteínas, em energia, e possui isoflavonas e ácidos graxos poli-insaturados que têm ação na redução de riscos de doenças crônico-degenerativas. Também é excelente fonte de minerais (como: ferro, potássio, fósforo e cálcio), fibras, vitaminas, e outros nutrientes. A ingestão de soja é aconselhada não somente para pessoas com reações alérgicas ao leite animal, mas também recomendada para prevenir doenças de coração, obesidade, hipercolesterolemia, câncer, diabetes, doenças de rins, e osteoporose (GARCÍA et al., 1998; PENHA, 2003; ZARKADAS et al., 2007).

A soja é uma importante mercadoria da agricultura devido sua composição química única, está incluída entre as cinco principais fontes protéicas, sendo também fonte calórica. Dentre os cereais e outras espécies leguminosas, as sementes de soja contêm a maior quantidade de proteínas (cerca de 40 %) e um nível relativamente alto de óleo (cerca de 20 %) (BOYDAK et al., 2002; MAEHLER et al., 2003). Como fonte barata de proteínas, a soja pode amenizar a situação da população que sofre com a pobreza, subnutrição e deficiência protéica (YAMADA et al., 2003; GREENPEACE, 2007).

A composição da soja depende de muitos fatores, incluindo genótipo, estação de crescimento, localização geográfica, práticas agrônômicas, efeitos causados pelo espaçamento entre as fileiras na plantação e a irrigação, todos esses fatores afetam a composição da proteína e óleo na soja (BOYDAK et al., 2002; PENHA, 2003; ZARKADAS et al., 2007).

Além do teor de proteína, o teor de óleo nas sementes de soja, também determina seu valor comercial. Dentre diversos estudos realizados mundialmente, as sementes de soja

contêm cerca de 33-45 % de proteínas e 18-25 % de óleo com base na matéria seca. E aproximadamente 8,5 % de umidade nos grãos (MORGANO; QUEIROZ; FERREIRA, 1999; BOYDAK et al., 2002; MAEHLER et al., 2003; PENHA, 2003; YAMADA et al., 2003; MORAES et al., 2006; ZARKADAS et al., 2007).

1.1.1 Soja em Roraima

O Estado de Roraima, localizado no extremo norte do Brasil, vai conquistando seu lugar no mapa do *agrobusiness* nacional e mundial sob o impulso da soja. Em Roraima, o cultivo de soja começou na década de 90. Na busca do desenvolvimento econômico, o governo do Estado estabeleceu como prioritária a produção de soja no cerrado, mas a soja só se instalou definitivamente no lavrado roraimense a partir de 1999, com a chegada dos primeiros produtores do Sul e Sudeste, atraídos pelo potencial da região e pelo baixo preço da terra (SOUSA, R.; SMIDERLE; SILVA, 2005; MAXETRON, 2006; O ESTADÃO, 2006a, 2006b).

A soja é uma das alternativas mais promissoras para o aproveitamento agrícola dessas áreas de cerrado, em face de seu valor econômico, e também pela disponibilidade de tecnologias para seu cultivo nas condições locais (GIANLUPPI et al., 2003).

O Estado acumula uma série de vantagens para a soja, como: o regime de chuvas rigorosamente regular, maior luminosidade devido à proximidade da linha do Equador, temperatura estável em níveis elevados durante todo o ano, grande oferta de água para irrigação, campos de vegetação rasteira e grandes extensões de terra plana, a pavimentação da rodovia BR-174, energia elétrica abundante e uma localização geográfica privilegiada em relação aos mercados, que caracterizam o estado de Roraima como uma nova fronteira agrícola. Complementam os atrativos da região: o baixo preço das terras, a facilidade de mecanização, disponibilidade de uma base tecnológica para a produção e o alto potencial de produtividade das culturas já identificados pela Embrapa (GIANLUPPI et al., 2002; GIANLUPPI et al., 2003; O ESTADÃO, 2006b; PORTAL BRASIL, 2006).

Estes fatores têm atraído os produtores de várias regiões do país em busca de informações, sendo que muitos deles estão se fixando aqui, para o cultivo de grãos, em especial a soja, por representar perspectivas de competitividade quanto aos mercados importadores da Venezuela, Estados Unidos da América, Europa e Ásia (GIANLUPPI et al., 2000c, 2002, 2004a, 2006).

Em função dos baixos teores de matéria orgânica e dos argilominerais, que apresentam baixa capacidade de armazenar água e nutrientes, para o uso desses solos na produção de soja

com qualidade, há necessidade, de fazer a correção completa com calcário, macro e micronutrientes (EMBRAPA, 2005).

Em Roraima, os dias são mais longos, com uma a duas horas adicionais de luz solar sobre as lavouras em relação ao resto do País. Com isso, o ciclo da cultura é mais curto e os teores de óleo (+2%) e proteína (+3%) mais elevados também em comparação a outras regiões do Brasil e do mundo (O ESTADÃO, 2006b).

Segundo levantamento do Centro de Pesquisas Agroflorestais da Embrapa em Boa Vista, a produtividade no lavrado roraimense é das melhores, se destacando acima das médias nacionais, como mostrada na Tabela 01. O potencial produtivo inflaciona o valor da terra, aumentando de pouco menos de R\$ 50 no começo da implantação da lavoura de soja para R\$ 500 por hectare atualmente (O ESTADÃO, 2006a).

Tabela 01 - Comparação da média de produtividade de algumas lavouras no Estado de Roraima com a média nacional.

	Arroz	Milho	Soja
Produção média em Roraima (kg/hectare)	7 mil	6 mil	3 mil
Produção média nacional (kg/hectare)	5 mil	4,2 mil	2,4 mil

Já se somam 20 cultivares de soja adaptadas pela Embrapa para os cerrados de Roraima, das quais 15 são as que prevalecem. Entre as cultivares de soja para o cerrado de Roraima, destaca-se como a mais cultivada a BRS Tracajá (produtividade média de 3,6 toneladas por hectare) (GIANLUPPI et al., 2000c, 2005; MAXETRON, 2006). Além da cultivar Tracajá, as cultivares Raimunda, Luziânia, Mirador, Sambaíba e Candeia têm em média uma produtividade superior a 3 toneladas por hectare. Atualmente, 90 % da área plantada com soja em Roraima é com sementes de tecnologia desenvolvida pela Embrapa (EMBRAPA, 2006).

O atual cenário da soja em Roraima tem despertado interesse internacional. Técnicos do Departamento de Agricultura dos EUA (USDA) já estiveram em Boa Vista em Agosto de 2003 avaliando o potencial da região. O Grupo Mitsubishi pretendia cultivar soja convencional (não-transgênica) em Roraima exclusivamente para exportação, entretanto, o investimento do grupo japonês só aconteceria se o governo do Estado garantisse ao grupo a posse definitiva da terra, o que não foi possível (O ESTADÃO, 2006c).

O Estado de Roraima aparece na relação dos produtores de grãos como uma área de fronteira agrícola onde a soja começou ser plantada recentemente. As duas primeiras safras

foram pequenas em Roraima, e a produção de 2000 foi toda armazenada em sacas e vendida junto com a soja do ano seguinte. O ponto de virada ocorreu em 2001, quando o Grupo Maggi inaugurou o porto de Itacoatiara e passou a comprar toda a produção roraimense. A soja do lavrado percorre 1 mil quilômetros de caminhão até Itacoatiara, em rodovias de asfalto. A localização geográfica do lavrado é estratégica, próxima de importantes centros de consumo representados por Manaus e pela Venezuela, além do porto de Itacoatiara. A próxima rota de escoamento deverá ser o porto de Georgetown, capital da Guiana, distante apenas 500 quilômetros do lavrado, abrindo o caminho marítimo mais curto para os mercados do Caribe e da Europa. Segundo a Embrapa Roraima, em 2004 foram produzidas aproximadamente 33 mil toneladas de soja numa área de 12 mil hectares, e estima-se um crescimento tanto de área plantada como em produtividade (EMBRAPA, 2005; O ESTADÃO, 2006b).

A safra de Roraima é vendida a preço de entressafra, pois a soja é cultivada durante a entressafra brasileira, porque Roraima está no Hemisfério Norte, o que garante suprimento doméstico para exportação num período de escassez. Em 2001 e 2002 o Grupo Maggi pagou em média US\$ 12,00 por saca de soja entregue em Itacoatiara. É um valor bem próximo ao verificado no porto de Paranaguá, o principal embarcadouro de soja da América do Sul (O ESTADÃO, 2006b).

Em razão do chamado "frete de retorno" (ocorre quando caminhões que trouxeram mercadorias a Boa Vista e que precisam retornar a Manaus de qualquer modo aproveitam a oportunidade para fazer transportes) o produtor de Roraima ganha vantagem no custo do frete, com despesa inferior a US\$ 1,00 para levar uma saca até o porto do Grupo Maggi (O ESTADÃO, 2006b).

O governo estadual pretende viabilizar a exportação de soja para a Venezuela a custos inferiores aos do produto comprado nos Estados Unidos, na Bolívia e na Argentina (PORTAL BRASIL, 2006).

Assim, o cultivo de soja em Roraima vem ganhando cada vez mais destaque e se tornando uma importante fonte econômica.

1.2 ESPECTROSCOPIA NO INFRAVERMELHO

Situada depois da região do visível, a região espectral do infravermelho abrange a radiação com números de onda no intervalo de aproximadamente 12.800 a 10 cm^{-1} ou comprimento de onda de 780 a 100.000 nm. Esta região é dividida basicamente em três faixas: a radiação no infravermelho próximo – NIR (do inglês, *Near Infrared*) que compreende de 12.800 a 4.000 cm^{-1} ; a radiação no infravermelho médio - MIR (do inglês,

Middle Infrared) de 4.000 a 200 cm^{-1} ; e a radiação no infravermelho distante - FIR (do inglês, *Far Infrared*) que compreende de 200 a 10 cm^{-1} . Os limites aproximados para cada região são apresentados na Tabela 02 (SKOOG; HOLLER; NIEMAN, 2002; BORIN, 2003; NETO, 2005; VALDERRAMA, 2005).

Tabela 02 - Regiões espectrais do Infravermelho.

Região	Intervalo de número de onda ($\bar{\nu}$), cm^{-1}	Região em comprimento de onda (λ), nm	Região de frequência (ν), Hz
Próximo (NIR)	12.800 a 4.000	780 a 2.500	$3,8 \times 10^{14}$ a $1,2 \times 10^{14}$
Médio (MIR)	4.000 a 200	2.500 a 5.000	$1,2 \times 10^{14}$ a $6,0 \times 10^{12}$
Distante (FIR)	200 a 10	5.000 a 100.000	$6,0 \times 10^{12}$ a $3,0 \times 10^{11}$

(Fonte: VALDERRAMA, 2005. p. 21)

Não existe um limite específico definindo onde termina uma região e começa outra, é comum encontrar na literatura diferentes valores para cada faixa, como a do NIR: 800 a 2.500 nm (SOUSA, E., 2005), 780 a 2.500 nm (BLANCO; VILLARROYA, 2002), 770 a 3000 nm (FIDÊNCIO, 2001), 750 a 2500 nm (PASQUINI, 2003), 715 a 2500nm (RODIONOVA et al., 2005), entretanto, a IUPAC (*International Union of Pure and Applied Chemistry*) define a região NIR sendo de 780 a 2500 nm (SILVA, H., 1999).

A espectroscopia no infravermelho teve, por ocasião da segunda guerra mundial, sua primeira aplicação no monitoramento de qualidade e controle da produção na indústria petroquímica. Combustíveis, lubrificantes e polímeros foram os principais compostos monitorados pelos equipamentos utilizando a região do MIR (COSTA FILHO; POPPI, 2002; VALDERRAMA, 2005).

Com o aparecimento de espectrofotômetros dispersivos, no fim dos anos 50, apenas algumas espécies químicas, devido às limitações instrumentais da época, eram identificadas e quantificadas através da espectroscopia em química (COSTA FILHO, 2003; NETO, 2005; VALDERRAMA, 2005).

A região do MIR era a mais utilizada para aplicações qualitativas, devido a possibilidade de se observar bandas de absorção de grupos orgânicos específicos como N-O, N-H, C-H, O-H, C-C, C=C e S-H (COSTA FILHO, 2003; NETO, 2005; ROGGO et al., 2007).

A região do FIR é principalmente útil para estudos de compostos inorgânicos, onde as vibrações de estiramento e deformação angular de átomos metálicos e ligantes, tanto inorgânicos como orgânicos, têm vibrações que podem ser observadas abaixo de 650 cm^{-1} (VALDERRAMA, 2005).

As aplicações da região FIR se dão também no estudo de gases que apresentam momentos de dipolo permanentes como, por exemplo, H₂O, O₃, HCl e AsH₃. Moléculas compostas apenas por átomos leves também absorvem na região FIR, desde que estas moléculas possuam modos de deformação angular da estrutura que envolvam mais de dois átomos que não sejam o hidrogênio (VALDERRAMA, 2005).

A partir da década de 70, com o desenvolvimento dos espectrofotômetros com Transformada de Fourier (FT-IR), da informática, do interfaceamento de instrumentos eletrônicos com computadores e a introdução de recursos matemáticos mais sofisticados, as análises quantitativas expandiram-se para as regiões espectrais no NIR e MIR (NETO, 2005; VALDERRAMA, 2005).

Essa expansão desencadeou o interesse de vários grupos de pesquisas governamentais e industriais, visando ao desenvolvimento de metodologias de análise, tendo como consequência um impacto direto no aumento da produtividade industrial, redução de gastos e minimização de resíduos industriais (NETO, 2005; VALDERRAMA, 2005).

Apesar de a região NIR ter sido detectada antes da região MIR, esta última foi a que ganhou maior aceitação e seu uso consagrou-se para identificação de grupos funcionais orgânicos. A região NIR foi negligenciada até por volta dos anos 80, quando essa ganhou maior atenção (PASQUINI, 2003).

O trabalho pioneiro envolvendo exploração analítica na região do espectro NIR, datada de 1938, verificava a determinação de água em gelatina. Em seguida, Barchewitz foi o primeiro a aplicar a espectroscopia NIR para a determinação de combustível. Inicialmente, visavam encontrar um comprimento de onda para aplicar a Lei de Lambert e Beer na determinação quantitativa. Esta idéia foi abandonada depois que Karl Norris reportou a adoção de modelos com mais de um comprimento de onda para correlacionar os dados dos espectros com a concentração do analito (PASQUINI, 2003).

Na década de 70, o grupo de pesquisa liderado pelo professor Karl Norris, atuando pelo Departamento de Agricultura dos estados Unidos, tornou-se o pioneiro na utilização do espectro NIR como uma ferramenta para determinações analíticas em amostras complexas e industriais, provando ser possível a aplicação do NIR para determinação de moléculas orgânicas, porém acompanhado de limitações tecnológicas da época (BURNS; CIURCZAK, 2001; BLANCO; VILLARROYA, 2002; SOUSA, E., 2005). Norris e sua equipe investigaram tópicos como: sangue em ovos, maturação de melões, proteína e umidade em trigo e “dureza” de trigo (BURNS; CIURCZAK, 2001).

Os anos 90 foram marcados pela utilização da espectroscopia no MIR para a análise qualitativa e quantitativa de alimentos, onde a reflexão total atenuada (ATR), detecção fotoacústica e reflexão difusa (DRIFTS) foram introduzidas em análise de alimentos e utilização no controle industrial (FERRÃO, 2000).

A espectroscopia no infravermelho vem ganhando muito espaço nos laboratórios analíticos e em análises de controle de qualidade nos processos industriais, como indústrias farmacêuticas, de alimentos, têxteis, etc. (FIDÊNCIO, 2001; BORIN, 2003).

1.2.1 NIR e Aplicações

A região NIR pode ser dividida em NIR de ondas curtas (*short-wavelength NIR*) compreendendo de 750 a 1100 nm; e NIR de ondas longas (*long-wavelength NIR*) compreendendo a faixa de 1100 a 2500 nm. Estas subdivisões são baseadas somente nos tipos de detectores usados, sendo que os detectores de silício são usados para as ondas curtas e detectores de PbS, InGaAs, ou detectores de germânio para as ondas longas (CHRISTIAN, 2004).

O espectro NIR está associado principalmente à anarmocidade das ligações envolvendo átomos de hidrogênio que vibram com maior amplitude devido à sua pequena massa e o grande momento de dipolo das suas ligações com átomos de carbono, nitrogênio e oxigênio (COSTA FILHO, 2003; CHRISTIAN, 2004).

Os sinais observados para a região do NIR são devidos principalmente a sobretons (do inglês, *Overtones*) e bandas fundamentais de ligações O-H, N-H, C-H e S-H, e bandas de ligações como C=O, C-C e C-Cl que são muito mais fracas ou até mesmo ausentes (SILVA, H., 1999; SCAFI, 2000; BURNS; CIURCZAK, 2001; BLANCO; VILLARROYA, 2002; SIMAS, R., 2005; COZZOLINO et al., 2006).

Os sobretons são transições entre estados vibracionais não consecutivos, apresentando intensidades muito mais fracas do que a transição entre o estado fundamental e o primeiro estado excitado. Bandas de combinação surgem quando uma molécula poliatômica possui dois ou mais modos vibracionais que podem interagir entre si. Daí, a banda de combinação resultante é dada pela soma de cada uma das frequências dos módulos vibracionais envolvidos na interação (COSTA FILHO, 2003).

Um exemplo deste fato é a interação entre os módulos vibracionais de um estiramento assimétrico a 3500 cm^{-1} e a deformação angular a 1645 cm^{-1} da ligação O-H da molécula da água, o que provoca a formação da banda de combinação correspondente à água, ocorrendo

em 1940 nm ($3500 + 1645 = 5145 \text{ cm}^{-1}$ que é o equivalente á 1943,6 nm) (COSTA FILHO, 2003).

O estiramento é promovido durante a mudança da distância interatômica dos átomos de uma ligação química podendo acontecer de forma simétrica ou assimétrica, enquanto a deformação angular envolve a mudança no ângulo da ligação entre os átomos, podendo acontecer no plano ou fora do plano da molécula (BORIN, 2003; COSTA FILHO, 2003; VALDERRAMA, 2005).

A região do NIR é caracterizada por conter bandas fracas e sobrepostas, tipicamente mais fracas (aproximadamente um fator de 10 a 100 vezes) que as bandas fundamentais do MID (CHRISTIAN, 2004; SOUSA, E., 2005). Essas intensidades inferiores, se por um lado dificultam em parte a obtenção de medidas, por outro permite, por exemplo, o seu uso em amostras que apresentam alto teor de umidade e garante uma alta penetração da radiação em medidas de reflectância de amostras sólidas (SCAFI, 2000).

As bandas de absorção referentes aos sobretons ocorrem em aproximadamente metade e um terço do comprimento de onda fundamental. Logo a ligação C-H, que tem no estado fundamental uma vibração de estiramento em 3200 nm, terá seu primeiro sobreton perto de 1600 nm e, seu segundo sobreton ao redor de 1070 nm (SOUSA, E., 2005).

A Tabela 03 ilustra as principais regiões espectrais do NIR e a natureza da transição vibracional associada a elas (COSTA FILHO, 2003).

Tabela 03 - Atribuição das regiões espectrais às possíveis transições vibracionais.

Região espectral (nm)	Natureza da transição vibracional
2200 – 2450	Combinação de estiramento C-H
2000 – 2200	Combinação de estiramento N-H, O-H
1650 - 1800	Primeiro sobreton do estiramento C-H
1400 - 1500	Primeiro sobretons de combinações N-H, O-H
1100 - 1225	Primeiro sobretons do estiramento C-H
950 - 1100	Segundo sobretons de combinações N-H, O-H
850 - 950	Terceiro sobretons do estiramento C-H
775 - 850	Terceiro sobretons de combinações N-H, O-H

(Fonte: COSTA FILHO, 2003. p. 94)

O espectro do infravermelho próximo é dependente de parâmetros físicos como: tamanho das partículas, polimorfismo, solventes residuais, umidade, temperatura, pressão, densidade, textura superficial, forma, compactação, etc. (COSTA FILHO, 2003; LUYPAERT et al., 2004).

A interação da radiação eletromagnética na faixa do NIR com materiais sólidos granulares pode ser interpretada de forma que cada partícula no sistema granular atua como um ponto de espalhamento, no qual as partículas vizinhas afetam a interação luz/partícula. A incidência do feixe de luz sobre a superfície da amostra pode provocar os fenômenos de reflexão, transmissão, refração, difração e absorção da luz incidente (BURNS; CIURCZAK, 2001; COSTA FILHO, 2003), Como mostra a Figura 01.

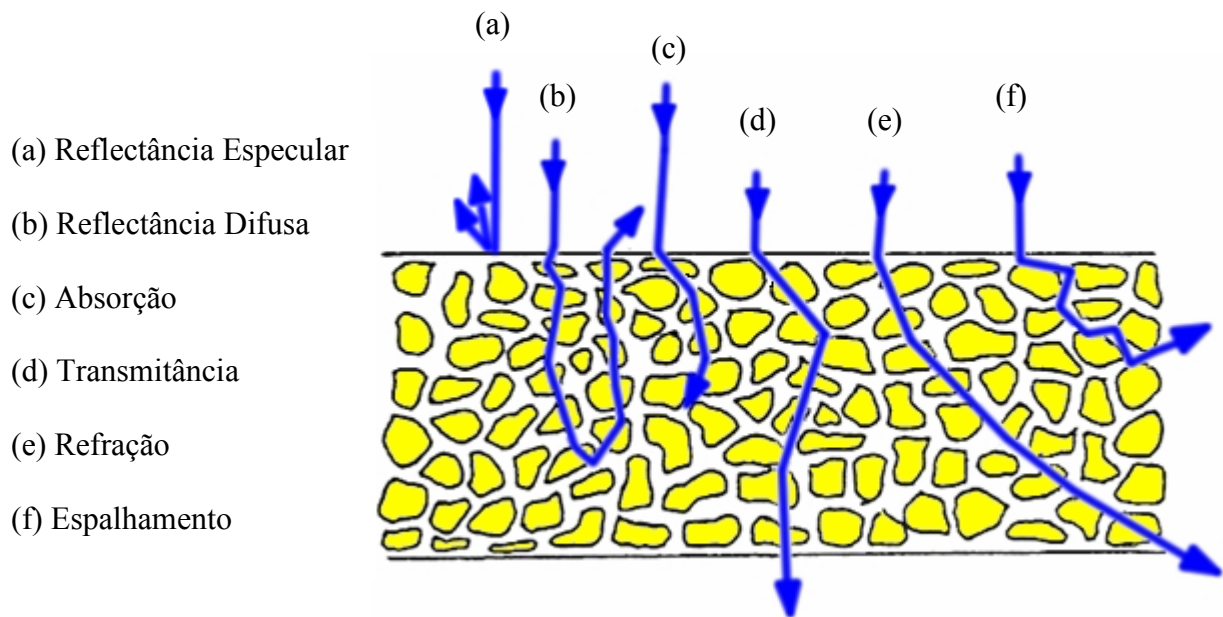


Figura 01 - Interação da radiação NIR com partículas sólidas de uma amostra.

As perdas de energia na amostra podem ocorrer devido a reflexão especular, refração interna, absorção completa, acarretadas devido reflexão de raio em sólidos de grande abertura angular. Se a amostra não absorve nenhuma fração da energia incidente, ocorre a reflexão total. Produtos agrícolas absorvem seletivamente radiação NIR rendendo informação sobre as ligações moleculares dentro do material estudado (BURNS; CIURCZAK, 2001).

O emprego da espectroscopia no NIR, consagrado desde a década de 60 com aplicações na determinação de umidade em sementes e nas análises de produtos agrícolas, foi amplamente empregado nas análises de grãos, farinhas e produtos industrializados, consagrando-se também na determinação de proteína (FERRÃO, 2000). Atualmente a técnica que utiliza transformada de Fourier tem sido constantemente aplicada para a determinação de proteína, umidade, amido, óleos, lipídeos e celulose em produtos provenientes da agricultura, tais como, grãos e óleos vegetais (FIDÊNCIO, 2001).

A informação contida no espectro NIR pode ser empregada para ter uma estimativa da concentração de uma dada substância na amostra, ou ainda, a estimativa de uma propriedade física quando esta pode ser, de alguma forma, correlacionada com alterações significantes na intensidade e/ou comprimento de onda das características espectrais produzidas pela amostra. Tais características podem ser indiretamente associadas com o analito (PASQUINI, 2003).

Dentre as aplicações na região do NIR encontram-se análises quantitativas nos mais diversos setores: clínico e biomédico (KONDEPATI et al., 2007a, 2007b); processos *on-line*, de monitoramento industrial e controle de qualidade (PAGÈS, 2002; HARBECK; FAURIE; SCHEPER, 2004; BLANCO; IÑÓN; GARRIGUES; GUARDIA, 2006; ORTIZ et al., 2006); monitoramento de reações orgânicas (HEIKKA et al., 1997); ambiental; têxtil; petroquímico (LITANI-BARZILAI et al., 1997; MACHO; LARRECHI, 2002; SOUSA, E., 2005); agrícola e agro-alimentício (FERRÃO, 2000; FIDÊNCIO, 2001; AZZOUZ et al., 2003; HERRERA; GUESALAGA; AGOSIN, 2003), Controle de qualidade de formulações agroquímicas (MOROS et al., 2005, 2006); indústria farmacêutica - visando monitoramento de produção ou adulteração de remédios - (BLANCO et al., 1999; KRÄMER; EBEL, 2000; SCAFI, 2000; RODIONOVA et al., 2005; ROGGO et al., 2007); cosmético; artes e etc. Sendo a espectroscopia NIR também utilizada para monitorar de forma rápida todos os estágios do processo de fabricação e, com isso, minimizar o tempo de fabricação e garantir em cada etapa do processo uma melhor qualidade do produto (FIDÊNCIO, 2001; PASQUINI, 2003; NETO, 2005; SOUSA, E., 2005; VALDERRAMA, 2005).

A possibilidade de realizar medidas de amostras em todos os estados e formas como: gases, líquidos, sólidos, sistemas binários e terciários como as amostras semi-sólidas, pastas, comprimidos, géis e outras, faz com que as regiões do infravermelho sejam de fundamental importância (SCAFI, 2000; NETO, 2005; VALDERRAMA, 2005).

A instrumentação utilizada na obtenção dos espectros NIR é composta basicamente por cinco componentes: uma fonte de radiação contínua, um seletor de comprimento de onda (monocromador), um porta amostra, um detector de radiação que converte a energia luminosa em energia elétrica e um dispositivo para ler a resposta do detector, como mostra a Figura 02 (SILVA, H., 1999; CHRISTIAN, 2004; SOUSA, E., 2005; VALDERRAMA, 2005).

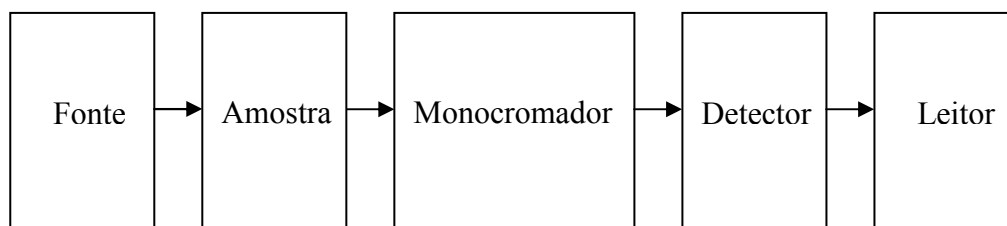


Figura 02 - Diagrama de blocos de um espectrômetro

Como fonte de radiação, são utilizados, geralmente, lâmpadas de Tungstênio – Halogênios, que emitem radiação contínua em toda a faixa espectral, caracterizadas por fornecer grande quantidade de energia na região entre 700 e 2500 nm (SOUSA, E., 2005; VALDERRAMA, 2005). As células ou porta-amostra para medidas de absorção são normalmente de quartzo ou sílica fundida, transparentes até 3000 nm. O caminho óptico varia de 0,1 a 10 cm (VALDERRAMA, 2005).

A evolução tecnológica proporcionou a substituição dos filtros de interferência, que eram empregados nos primeiros espectrofotômetros NIR e que acarretavam dificuldades para se obter um espectro com resolução apropriada, por monocromadores baseados em redes de difração, sistemas interferométricos (FT) e Filtros Ópticos Acústicos Sintonizáveis (AOTF) (GUCHARDI, 2000; SCAFI, 2000; SOUSA, E., 2005). Este último é um dispositivo óptico-acústico baseado na difração da luz pela onda acústica em um cristal anisotrópico, geralmente TeO_2 e se caracteriza-se por fazer as ondas acústicas e ópticas se propagarem em diferentes ângulos através do cristal quando disposto sob uma configuração não-colinear. A radiofrequência aplicada no transdutor associado ao cristal TeO_2 é convertida em onda acústica e propagada pelo cristal fazendo com que o índice de refração varie. Desta forma é feita a seleção do comprimento de onda proveniente da fonte de radiação. (GUCHARDI, 2000; PASQUINI, 2003; SOUSA, E., 2005).

Outra parte fundamental para a obtenção dos espectros é o detector, baseado geralmente em materiais fotocondutores como PbS (SOUSA, E., 2005; VALDERRAMA, 2005) e InGaAs, tais detectores, quando empregados na região do espectro NIR, são caracterizados pela quantidade mínima de energia radiante que eles podem detectar, faixa espectral de resposta e velocidade de resposta (SOUSA, E., 2005).

As fontes de alta intensidade e os eficientes detectores compensam a desvantagem da baixa intensidade das absorções mencionadas anteriormente, contribuindo assim para o aumento da relação sinal/ruído de medidas efetuadas nessa região espectral (SCAFI, 2000).

Os métodos por reflexão especular (ou externa), por reflexão difusa e por reflexão interna no infravermelho apresentam aplicações práticas na detenção informações químicas de determinada matriz e possuem a vantagem de não serem destrutivas para a amostra (FERRÃO, 2000; SIMAS, R., 2005).

1.2.2 Reflexão Difusa

Medidas por reflectância são obtidas a partir da radiação espalhada na superfície da amostra, sendo uma importante ferramenta para determinação de constituintes de amostras na forma de partículas pequenas. Um corpo branco tal como KBr, BaSO₄, MgO, são usados como material de referência, uma vez que estes refletem praticamente toda radiação incidente (FIDÊNCIO, 2001). A reflectância é definida como sendo a razão das potências ou fluxos luminosos da radiação refletida (I_R) e da radiação de referência (I_0) numa superfície. Assim, a reflectância é medida comparando-se a quantidade medida com a referência de acordo com a seguinte expressão (BURNS; CIURCZAK, 2001; FIDÊNCIO, 2001):

$$R = \frac{I_R}{I_0} \quad (10)$$

A reflectância é amplamente associada aos equipamentos que operam no infravermelho próximo, e bem fundamentada na literatura, é geralmente designada de espectroscopia de reflexão no infravermelho próximo (NIRS ou NIR); já associada aos equipamentos que operam no infravermelho médio é conhecida por espectroscopia por reflexão difusa no infravermelho médio com transformada de Fourier (DRIFTS) (FERRÃO, 2000).

O espectro de reflectância não é idêntico ao de absorção, porém apresenta certa semelhança e carrega as mesmas informações. A reflectância difusa não requer grande preparo da amostra, apenas moagem visando transformá-la em pó, devendo ter um tamanho uniforme de partícula e a moagem deve ser refrigerada para não alterar as características químicas da amostra, e o grau de compactação não deve sofrer variações (SIMAS, R., 2005).

Análises de reflexão NIR em produtos de cereais são normalmente realizadas usando partículas com um diâmetro médio de 100 μm . É obviamente impossível descrever como cada fóton interage com um pó refletindo difusamente, devido ao número de reflexões, refrações e difrações em uma amostra contendo partículas de formas irregulares. Não obstante, é possível conceituar certos eventos e efeitos destes eventos na reflectância medida (BURNS; CIURCZAK, 2001).

A Figura 03 mostra as duas instrumentações básicas mais comuns em análises por NIR (BURNS; CIURCZAK, 2001).

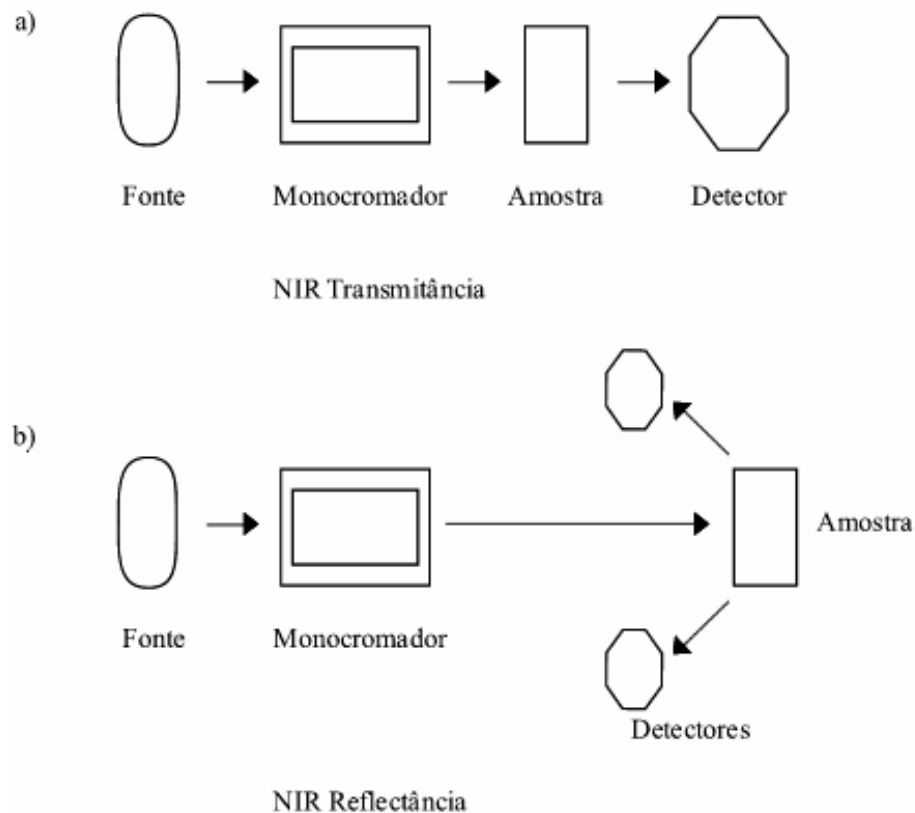


Figura 03 - Configuração da instrumentação básica do NIR: a) Transmitância; b) Reflectância.

Esse processo de reflexão é baseado na interação do feixe incidente com a matriz, qual penetra na superfície, sofre absorção parcial e múltiplos espalhamentos e retorna à superfície da amostra. A radiação incidente sofre atenuação após entrar em contato diversas vezes com as partículas das amostras, essa reflexão difusa, que são os raios que emergem da superfície após terem penetrado no material, fornece informações qualitativas e quantitativas da amostra. A reflexão difusa ocorre em superfícies não totalmente planas, podendo a amostra estar na forma de material particulado ou não (FERRÃO, 2000; FIDÊNCIO, 2001; COSTA FILHO, 2003; SIMAS, R., 2005). A Figura 04 ilustra esse processo.

A maior parte dos acessórios para a determinação da reflexão difusa apresenta esquemas ópticos que concentram a radiação para, posteriormente, ser focada sobre o sistema de detecção dos instrumentos. Isso se deve ao fato da magnitude da radiação difusa que retorna ser muito inferior a da radiação incidente (FERRÃO, 2000).

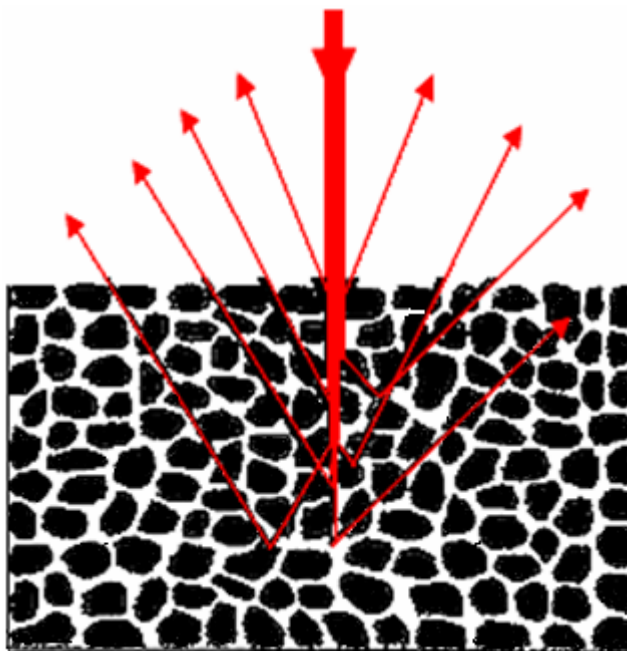


Figura 04 - Representação da reflexão especular e difusa de uma onda eletromagnética em uma amostra particulada. (Fonte: MESSERSCHMIDT, 1999. p. 17)

Simultaneamente com a reflexão difusa ocorre a reflexão especular, essa ocorrência é mais freqüente nas regiões dos espectros de intensa absorção da radiação, a qual é governada pelas equações de Fresnel, sendo definida como a razão entre a intensidade de radiação refletida e a intensidade da radiação incidente, cujo o ângulo de incidência da radiação sobre a amostra é exatamente igual ao ângulo de reflexão (MESSERSCHMIDT, 1999; FERRÃO, 2000; COSTA FILHO, 2003).

Diferentes tamanhos de partículas podem causar efeitos sobre o espectro, que podem ser reduzidos por transformações no espectro, através da primeira e segunda derivadas ou correção do espalhamento multiplicativo (MSC), ou transformação padrão normal de variação (SNV) (FERRÃO, 2000).

Parâmetros como a geometria ótica, compactação da amostra, índice de refração e outros podem afetar o formato, posição e intensidade da banda dos espectros de reflexão difusa associada ao NIR e MIR. A concentração do material, através do qual a radiação é transmitida, é afetada pelo vazio entre as partículas, bem como, pelo tamanho, forma e compactação da mesma. A alteração do tamanho das partículas faz com que a razão entre as bandas fracas e as fortes não sejam constantes (MESSERSCHMIDT, 1999; FERRÃO, 2000; COSTA FILHO, 2003).

A atenuação de parte da reflexão difusa pela absorção no meio é dada pela lei de Lambert-Bourguer:

$$I = I_0 \cdot e^{-\varepsilon \cdot \delta} , \quad (11)$$

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

A facilidade em se obter espectros de pós, sólidos e espécies em sólidos difundiu o uso da NIRRS. E, a utilidade desta, foi ainda mais aumentada com o uso de novos métodos de amostragem, proporcionando a obtenção de informações sobre as amostras em sua forma mais próxima da original (MESSERSCHMIDT, 1999; FERRÃO, 2000).

A obtenção de informações quantitativas pode ser alcançada preparando-se adequadamente as amostras e conhecendo-se a limitação desta técnica na obtenção de uma boa qualidade dos espectros (MESSERSCHMIDT, 1999; FERRÃO, 2000).

Dentre as diversas aplicações, encontram-se determinações de aminoácidos, fibras e investigação da textura em grãos de trigos e determinação de umidade em amostras de milho; caracterização de polímeros sintéticos, carvão e catálise, bem como a identificação de frações separadas por cromatografia (MESSERSCHMIDT, 1999; FERRÃO, 2000).

A temperatura, a umidade, os solventes residuais, a espessura das amostras e a estrutura polimórfica são os principais fatores que influenciam nos trabalhos quantitativos na região NIR (COSTA FILHO, 2003; LUYPAERT et al., 2004).

Em trabalho de alta precisão o parâmetro temperatura deve ser considerado, uma vez que, a variação da mesma em uma solução, mesmo que de poucos graus, pode resultar em mudanças substanciais no perfil do espectro (COZZOLINO, 2007). Entretanto este parâmetro se torna menos crítico para espectros de reflectância NIR de sólidos. Em caso de trabalhos envolvendo determinação de água, deve-se manter constante a umidade da sala de aquisição de espectros. Também é necessário cuidado com possíveis contaminações por solventes residuais na amostra, pois o mesmo pode acarretar mudanças relevantes no perfil do espectro (COSTA FILHO, 2003; COZZOLINO, 2007).

A variação na espessura da amostra sólida implica na variação do caminho óptico. Quando a amostra possui uma espessura na ordem de milímetros este fator é mais sensível. O motivo é que a luz pode penetrar vários milímetros no interior da amostra, já que a radiação NIR é absorvida muito menos do que a radiação MIR. Outro fator é a estrutura polimórfica, uma vez que variações na estrutura cristalina podem ser observadas, pois a reflectância NIR mede diretamente substâncias sólidas cristalinas (COSTA FILHO, 2003).

É praticamente impossível se fazer uso quantitativo ou qualitativo da técnica NIR sem a utilização de faixas espectrais, ou do espectro total devido a não especificidade das

ocorrências espectrais na região NIR. É necessário o estabelecimento de uma relação matemática entre o espectro e valores de concentrações medidos por um método de referência. Relação essa, que quando bem estabelecida, propicia a baixo custo, com alta precisão e repetitividade, resultados de uma dada concentração através de um modelo de calibração. Para o tratamento dos dados espectrais obtidos utilizam-se as técnicas quimiométricas (SCAFI, 2000; SIMAS, R., 2005).

1.3 QUIMIOMETRIA

A composição química de uma amostra não é uma grandeza observável, sendo assim obtida de forma indireta por processos analíticos (tais como absorção da luz, condutividade, pesos, etc.). Em qualquer análise se faz necessário fazer uma calibração, que se baseia em encontrar uma função que relacione as medidas efetivamente realizadas com a concentração ou propriedade desejada e, assim, chegar ao valor da concentração (PIMENTEL; NETO, 1996; NETO; PIMENTEL; ARAÚJO, 2002).

As técnicas instrumentais incitadas pela invasão de microprocessadores e microcomputadores no laboratório químico, tornam-se cada vez mais sofisticadas, e requerem tratamentos de dados mais complexos do ponto de vista estatístico e matemático, sempre visando relacionar os sinais obtidos com os resultados desejados (BRUNS, 1985; SENA, 2000).

A análise de grandes quantidades de dados, onde são encontrados diversos valores para muitas variáveis obtidos por medidas instrumentais de amostras torna-se mais fácil quando feita através de técnicas matemáticas da Estatística Multivariada. Essas técnicas são utilizadas na fase de tratamento de dados químicos, onde são conhecidos como métodos quimiométricos, no qual seu uso é indispensável (BRUNS, 1985).

Em geral, o estudo da quimiometria tem sido voltado para três grandes focos: calibração multivariada; análises discriminatórias de classificação e de reconhecimento de padrão; e monitoramento e modelagem de processos de reação e reagentes (HASWELL; WALMSLEY, 1999).

Uma das mais importantes áreas da quimiometria é a conversão de dados multivariados em uma informação útil, incluindo processos como a Análise por Componentes Principais (PCA) e métodos de regressão, como a Regressão Linear Múltipla (MLR), a Regressão nas Componentes Principais (PCR) e os Mínimos Quadrados Parciais (PLS). Tais métodos, são

usados na construção de modelos de regressão, que permitem estimar com eficiência propriedades das amostras que se deseja investigar.

O uso da espectroscopia NIR, combinada com o tratamento de dados por métodos quimiométricos, tem se tornado cada vez mais rotineiros, encontrando aplicações nas mais diversas áreas (OLIVEIRA, 2004).

A quimiometria não é uma disciplina da matemática, mas sim da química, isto é, os problemas que ela se propõe a resolver são de interesse e originados na química, ainda que as ferramentas de trabalho provenham principalmente da matemática, estatística e computação (FERREIRA et al., 1999). Assim a Quimiometria é uma área da Química que utiliza métodos estatísticos e matemáticos para planejar, selecionar ou definir as melhores condições para a obtenção de medidas experimentais, buscando uma otimização, e permitir a obtenção mais completa possível de informações, a partir de dados físico-químicos (ARAUJO, 2005).

A quimiometria é usada para a classificação de amostras, de acordo com suas propriedades, bem como a quantificação de variáveis e até mesmo para previsão de resultados, através de análises de resultados relevantes.

É importante não esquecer dos princípios químicos implícitos na quimiometria. Isto é, se quimicamente a resposta não faz sentido, o método quimiométrico foi mal aplicado ou o seu planejamento foi mal elaborado (MENDHAM et al., 2002).

1.3.1 Modelos de Regressão

Devido a enorme quantidade de objetos (amostras químicas) e variáveis (parâmetros ou propriedades químicas medidas) é necessário, para facilitar a interpretação dos dados, ordená-los na forma de matriz, onde os objetos ou amostras são representados por linhas da matriz e as variáveis por colunas da matriz. Assim, seja a matriz X de n linhas e m colunas (MESSERSCHMIDT, 1999):

$$X = (a_{ij})_{m \times n} , \quad (12)$$

em que

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & & & \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix} \quad (13)$$

de modo que cada linha corresponde uma amostra e cada coluna uma variável da amostra.

O princípio básico da calibração multivariada consiste em encontrar uma relação entre as variáveis independentes e as variáveis dependentes. Entenda-se como variáveis independentes x_1, x_2, \dots, x_n (ex.: valores de absorvância a vários comprimentos de onda) na matriz X . E as variáveis dependentes y_1, y_2, \dots, y_n (ex.: concentração de um analito, teor de proteína bruta) na matriz Y . Esse modelo nada mais é do que uma função matemática que pode ser linear ou não-linear, dependendo da complexidade do sistema em estudo (FIDÊNCIO, 2001).

Em uma determinação por NIR, os espectros de cada amostra são organizados numa matriz, X ($n \times m$), de variáveis independentes, onde cada linha representa uma amostra e contém as respostas medidas para a mesma. Nos espectros obtidos, a unidade do eixo vertical, $\log(1/\text{reflectância})$ – abreviado como $\log(1/R)$ – é relacionada a absorção; um maior valor de $\log(1/R)$ significa que mais radiação tem sido absorvida (menos refletida) pela amostra naquele comprimento de onda.

Para a análise dos dados obtidos no exemplo mencionado anteriormente, são empregados os métodos multivariados que permitem um estudo com as várias espécies presentes ao mesmo tempo, não importando a existência ou ausência de diferenças espectrais marcantes entre elas nem a existência de alta correlação nos dados. É possível também, a identificação de problemas eventuais com linha base ou interferentes nas amostras usadas na calibração e nas novas amostras, de previsão (FERREIRA et al., 1999).

A construção dos modelos de regressão pode ser dividida basicamente em duas etapas: calibração, onde as concentrações conhecidas são usadas para construir um modelo que relacione a grandeza medida com a concentração da espécie de interesse; e previsão, onde o modelo construído na etapa anterior é usado para prever concentrações de novas amostras, a partir dos sinais analíticos medidos para elas (PIMENTEL; NETO, 1996).

1.3.1.1 Calibração

De posse dos conjuntos dos dados, inicia-se o procedimento por calibração, que consiste em encontrar um algoritmo matemático que reproduza a(s) propriedade(s) de interesse a partir dos resultados registrados pelo instrumento. Isto é, é onde se encontra uma relação entre as medidas físicas (variáveis independentes da matriz X) e as propriedades de interesse (variáveis dependentes da matriz Y) (PIMENTEL; NETO, 1996; FERREIRA et al., 1999; NETO; PIMENTEL; ARAÚJO, 2002).

Para a construção do modelo de calibração é usado um conjunto de amostras, denominado conjunto de calibração (*Calibration Set*) ou de treinamento (ARAÚJO, 2005).

Os parâmetros do modelo (função) são chamados de coeficientes de regressão (b). Tais parâmetros são matematicamente determinados a partir de dados experimentais. A Figura 05 esquematiza os procedimentos para calibração e previsão, em que é mostrada a obtenção da matriz X através das medidas instrumentais e sua relação com a matriz Y por meio da calibração. Posteriormente a esta etapa é feita a previsão, isto é, a partir do modelo de calibração construído (COSTA FILHO, 2003).

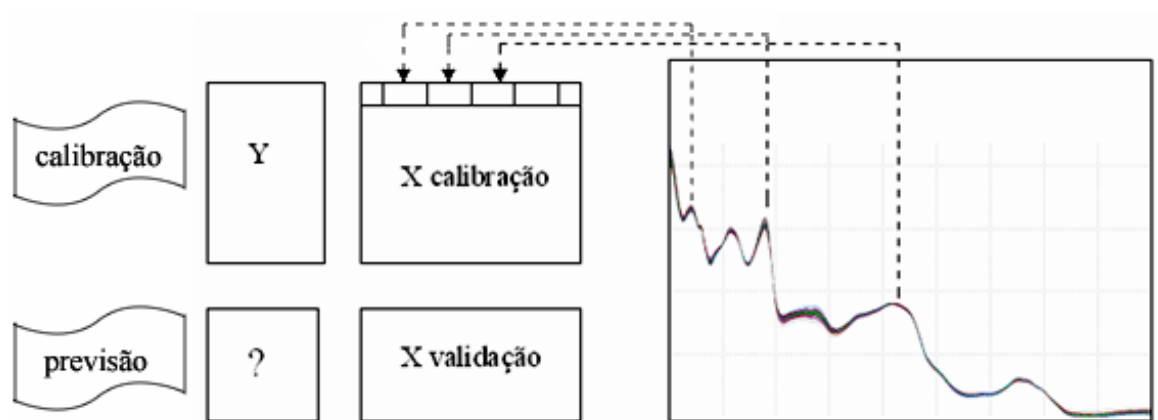


Figura 05 – Esquema da organização dos dados para as etapas de calibração e previsão. (Fonte: COSTA FILHO, 2003. p. 9).

Dependendo da dimensão da matriz, os métodos de calibração podem ser univariados (quando de ordem zero); multivariados (quando de primeira ordem); e multidimensionais (quando de segunda ordem, ou terceira ordem). No caso da segunda ordem, os dados para uma única amostra química incluem uma matriz que pode ser visualizada como uma superfície bi-dimensional ou “paisagem” (SIMAS, R., 2005; SENA; TREVISAN; POPPI, 2005; ESCANDAR et al., 2007).

O processo geral de regressão consiste, basicamente, de duas etapas: modelagem, que estabelece uma relação matemática entre X (variáveis independentes, ex.: valores de absorvância) e Y (propriedade de interesse, ex.: teores de proteína) no conjunto de calibração; e a validação, que otimiza a relação no sentido de uma melhor descrição do(s) analito(s) ou propriedade(s) em interesse (FERREIRA et al., 1999; SIMAS, R., 2005).

Após a etapa de calibração é usado outro grupo de amostras para validar a calibração, chamado de conjunto de validação (*Validation Set*). Essa validação, feita através de variáveis

independentes (X) juntamente com os coeficientes de regressão, e serve para calcular os valores estimados para as variáveis dependentes (Y) (MENDHAM et al., 2002).

Uma vez concluída a regressão, o sistema está apto a ser utilizado para previsão em outras amostras. Portanto, um modelo de regressão que relaciona a intensidade de radiação e a concentração do analito estudado, correlacionando os dados medidos com a propriedade de interesse da amostra através da modelagem e validação, pode assim, ser capaz de prever uma propriedade de interesse a partir das respostas de um equipamento (FERREIRA et al., 1999; SIMAS, R., 2005).

Durante a etapa de calibração dois fatores devem ser considerados: 1 - O número de fatores A a ser utilizado no modelo (número de componentes principais - PCs ou número de variáveis latentes - VLs). 2 - Detecção de "outliers" (amostras anômalas) que são amostras que sofreram algum tipo de interferência durante o processo de determinação de suas propriedades (FERREIRA et al., 1999).

O processo de validação é usado para avaliar a eficiência e desempenho do modelo de calibração, Por isso, o conjunto de validação, deve ter variáveis (Y_p) conhecidos, com a finalidade de comparar com os valores previstos, através de cálculos de desvios (LIMA, 2000).

1.3.1.2 Previsão

Na literatura NIR, o termo "previsão" é comumente usado no lugar de "medidas" ou "determinação". Instrumentos NIR determinam proteína e outros componentes pela medida de valores $\log(1/R)$ que podem então ser relacionados à quantidade do componente quando determinados por outros métodos, chamados de métodos de referência ou método padrão (WILLIAMS; NORRIS, 2001).

A previsão consiste no processo de estimar parâmetros de uma amostra aplicando um modelo de regressão a um vetor de medidas da amostra. Com ferramentas de diagnósticos de previsão é possível determinar se o modelo não é aplicável devido a falhas instrumentais ou a presença de amostras anômalas (ARAUJO, 2005).

Na primeira etapa tem-se a função

$$Y=f(X): \tag{14}$$

$$Y_c=X_c * b \tag{15}$$

Em que o conjunto de calibração é: Y_c e X_c , conhecidos inicialmente, b é a matriz dos coeficientes de regressão, que é calculado matematicamente.

Na segunda etapa, com o valor de b calculado, determina-se Y_p através da expressão:

$$Y_p = X_p * b \quad (16)$$

Em que X_p é o conjunto de previsão.

Deste modo, duas matrizes são empregadas, o padrão que contém as informações sobre o analito de interesse (em concentração conhecida), e o analito de interesse em concentração desconhecida (a ser determinada pelo método de calibração) (PIMENTEL; NETO, 1996; REIS, 2002). Entenda-se como analito, a espécie química de interesse contida em uma amostra, cuja concentração é determinada em uma análise química (ARAUJO, 2005).

Como os métodos de calibração multivariada usam vários canais de dados, obtêm-se precisão nos valores estimados, sendo possível determinar espécies, mesmo na presença de interferentes. Para o desenvolvimento dos modelos de calibração têm-se diversos métodos, cada um possui suas vantagens e desvantagens, são exemplos de métodos de calibração: o MLR (regressão linear múltipla), o PCR (regressão por componentes principais) e o PLS (Regressão por mínimos quadrados parciais) (SIMAS, R., 2005).

1.3.2 Métodos Descritivos e Exploratórios e métodos de Calibração Multivariada

Os métodos quimiométricos podem ser aplicados em dados multivariados com propósitos qualitativos onde destacamos a Análise por componentes Principais (PCA) e a Análise por Agrupamentos Hierárquicos (HCA), e quantitativos (calibração) como o PLS, o RLM, e o PCR (PARREIRA, 2003).

Neste trabalho serão abordados apenas os métodos PCA e o PLS, os quais foram utilizados neste estudo.

1.3.2.1 PCA

O PCA é um método matemático usado para maximizar as informações que podem ser extraídas de um conjunto de dados e assim, reduzir o número de variáveis da matriz original dos dados envolvidos no modelamento. Atuando de modo a simplificar a série de dados e reduzir a dimensão dos dados com a finalidade de gerar novas variáveis que não são correlacionadas. Por causa dessa habilidade, diz-se grosso modo que acontece uma

“compressão de dados” (REIS, 1997; BARTHUS, 1999; MORGANO; QUEIROZ; FERREIRA, 1999; ROGGO et al., 2007).

O PCA é um dos métodos de Reconhecimento de Padrões (PR - *Pattern Recognition*) Não Supervisionados, os quais são usados para examinar similaridades ou diferenças entre amostras (ROGGO et al., 2007). Outro PR não supervisionado é o HCA (do inglês, *Hierarchical Cluster Analysis*), que não será abordado nesse trabalho.

A análise por PCA é utilizada quando se tem uma grande quantidade de dados que serviriam para representar o sistema e construir o modelo de calibração, entretanto esse excesso de dados acaba dificultando o tratamento (processamento e armazenamento) dos mesmos, isto é, muitos desses dados podem ser irrelevantes para a análise quimiométrica. Assim, a análise por PCA pode reduzir essa quantidade de dados, descartando as informações redundantes e, fornecendo um conjunto de dados com informações relevantes, ou seja, só dados realmente úteis para a análise e construção do modelo do sistema (NÆS; MARTENS, 1989; MELLO, 1998; BARTHUS, 1999; PARREIRA, 2003).

Por esse motivo, o PCA é a base fundamental da maioria dos métodos modernos para tratamento de dados multivariados (FERREIRA et al., 1999; MORGANO; QUEIROZ; FERREIRA, 1999).

Uma das principais vantagens da análise de componentes principais (PCA) é a eliminação de grande parte dos ruídos experimentais. Isto é possível, pois os ruídos constituem uma população de erros aleatórios, ou seja, não estão correlacionados com as informações contidas na matriz de dados. Devido a ausência de correlação, uma quantidade maior de componentes principais é necessária para representar a matriz de dados (REIS, 1997).

Para efetuar tal redução, sem perda de informações, a matriz original passa a ser representada por novas variáveis, ortogonais entre si e direcionadas em função da distribuição espacial dos dados. Essas novas variáveis são chamadas de Componentes Principais – PC (do inglês, *Principal Component*) (REIS, 1997).

O método PCA tem por princípio básico obter uma coleção menor de dados a partir da coleção de dados inicial, mas que possa representar as principais características, isto é, maior parte das informações estatísticas úteis da matriz inicial. Para isso, são encontradas combinações lineares nas variáveis independentes e/ou nas variáveis dependentes (NÆS; MARTENS, 1989; MELLO, 1998; NAGATA; BUENO; PERALTA-ZAMORA, 2001; FIDÊNCIO, 2001).

A matriz de dados onde será aplicada o método PCA vai depender dos métodos a serem seguidos, posteriormente, como por exemplo o PCR ou PLS podendo ser aplicado apenas em uma das matrizes ou em duas.

Assim, o método PCA possibilita que a matriz dessas variáveis independentes seja decomposta em uma soma de matrizes menores, que são oriundas de produtos vetoriais entre t_i (*scores*) e p_i (*loadings* ou “pesos”) e uma matriz residual (E) (MESSERSCHMIDT, 1999; NAGATA; BUENO; OTTO, 1999; SENA, 2000; PERALTA-ZAMORA, 2001), de acordo com a expressão:

$$X = t_1 p_1^t + t_2 p_2^t + \dots + t_n p_n^t + E \quad (17)$$

Em que t_i são os escores (*scores*), p_i os pesos (*loadings*), e E representa a matriz de erros da matriz das variáveis independentes. Os *scores* correspondem as coordenadas de cada amostras nas PCs e os *loadings* a contribuição de cada variável nas PCs (FIDÊNCIO, 2001).

A expressão (17) pode ser representada por:

$$X = TP^t + E \quad (18)$$

Em que P^t é uma matriz com p^t linhas e T é uma matriz com t colunas.

Por exemplo, para determinar o teor de proteína bruta em grãos de soja, através da espectroscopia NIR, necessita-se de diversas amostras e, para cada amostra, pelo menos um espectro, obtendo as seguintes matrizes, X e Y, como mostra esquematicamente as Figura 06 (a) e (b) .

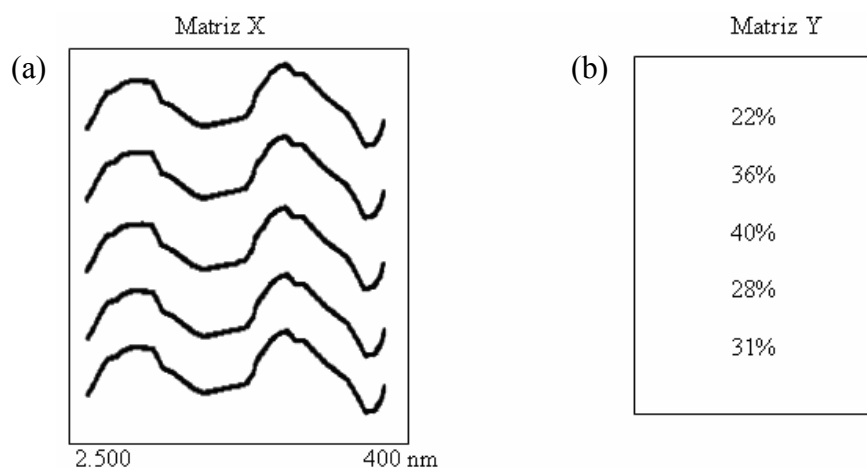


Figura 06. Representação: (a) Matriz X - Espectros NIR e (b) Matriz Y - Teores do analito.

Onde na matriz X , estão os espectros NIR correspondente a cada amostra e as absorvâncias lidas em diferentes comprimentos de onda, e na matriz Y , estão os respectivos valores de proteínas determinados por um método de referência..

Os valores das colunas da matriz X são centrados na média para obter \bar{X} , é necessário subtrair o valor médio de cada elemento da coluna e, em seguida, calcular a covariância de \bar{X} , utilizando a seguinte equação (MELLO, 1998):

$$Cov(\bar{X}) = \frac{\bar{X}^T \bar{X}}{m-1} \quad (19)$$

É necessário também obter a maior variação da matriz, $Cov(\bar{X})$, pela seguinte equação (MELLO, 1998):

$$Cov(\bar{X})p_i = \lambda_i p_i \quad (20)$$

Em que λ_i , é o i -ésimo autovalor associado ao i -ésimo autovetor p_i .

Os autovalores mostram o quanto da variabilidade que cada fator remove de X , sendo denominados por $\lambda_1, \lambda_2, \dots, \lambda_a$. E os autovetores representam os vetores que explicam a máxima variância na calibração (MESSERSCHMIDT, 1999).

O que de fato acontece é que ao colocar a matriz de covariância, substitui-se aqueles valores de absorvância lida num comprimento de onda e que se relaciona com outra absorvância lida em outro comprimento de onda, por um valor que corresponde a combinação linear destes. Assim, se o grupo de comprimentos de ondas possuem alta covariância, podemos substituir o grupo todo pela combinação linear de suas absorvâncias. Para obter esses grupos de alta covariância é necessário empregar a equação (20) e obter assim seus autovalores, e em seguida, seus autovetores (MELLO, 1998).

A solução da equação (20) apresenta vários autovalores, o primeiro autovetor corresponde ao maior autovalor, fornece a direção no espaço das variáveis (colunas de X) que descreve a maior porcentagem da variância das amostras. Se um único autovetor não puder descrever toda as variáveis do sistema, pode se obter um segundo autovetor e sucessivamente, até que todas as variações das amostras sejam obtidas (MELLO, 1998; OTTO, 1999).

Para minimizar os erros residuais em cada passo, os autovetores são calculados consecutivamente. Assim, os autovetores consistem num sistema mutuamente otimizado de coordenadas ortogonais. Cada autovetor explica a máxima variância possível nos dados.

Cada fator é responsável pela fração sucessiva de variância dos dados (MESSERSCHMIDT, 1999; OTTO, 1999).

Após obter a direção de maior variância das amostras, faz-se uma mudança de base da matriz representada pelas variáveis originais pelas componentes principais e tal mudança é feita através de uma transformação linear. Em seguida, projeta-se a matriz X nesta nova base (MELLO, 1998), de acordo com a expressão:

$$Xp_i = t_i, \quad (21)$$

Através do método dos mínimos quadrados, tem-se (MELLO, 1998):

$$X = t_i p_i^T (p_i p_i^T)^{-1} \quad (22)$$

Na equação 21 os vetores t_i são conhecidos como *scores* e contém informações de como as amostras estão relacionadas umas com as outras. Os vetores p_i são conhecidos como *loadings* ou “pesos” e contém informação de como as variáveis estão relacionadas umas com as outras (MELLO, 1998).

O máximo de *scores* necessários para a descrição da matriz X é escolhido em função da variância capturada pelo autovalor e, geralmente, reduz severamente a dimensão da matriz de dados. Por isso, os *scores* podem ser usados como “absorbâncias” em procedimentos de calibração (MELLO, 1998).

No PCR, é estabelecida uma relação linear entre os *scores* das variáveis independentes e as concentrações da matriz Y . No PLS, são obtidos *scores* da matriz X e da matriz Y . E assim, é feita a calibração usando os *scores* das duas matrizes (MELLO, 1998).

Os *scores* podem ser usados, com vantagens, em substituição de outras variáveis.

Na modelagem em PCs, a 1ª reta obtida (chamada de PC1 : Primeiro Componente Principal) contém o máximo da variância dos dados possível na direção de maior espalhamento dos dados (BORIN, 2003), ou seja, contém o máximo da informação estatística, sendo descrita como:

$$X = t_1 p_1^T + e_1 \quad (23)$$

Se PC1 não descrever toda a variância dos dados, a parte não descrita fica na matriz dos resíduos e_1 , que é modelada dando PC2 (Segunda Componente Principal) ortogonal a PC1 mais e_2 e dois novos vetores p_2^T e t_2 , descrita como:

$$X = t_1 p_1^T + t_2 p_2^T + e_2 \quad (24)$$

A Figura 07a apresenta o gráfico dos *scores*, ou seja, dos objetos ou amostras projetados em um plano formado por 2 PCs (PC1 versus PC2) e a Figura 07b apresenta os *scores* para 3 variáveis.

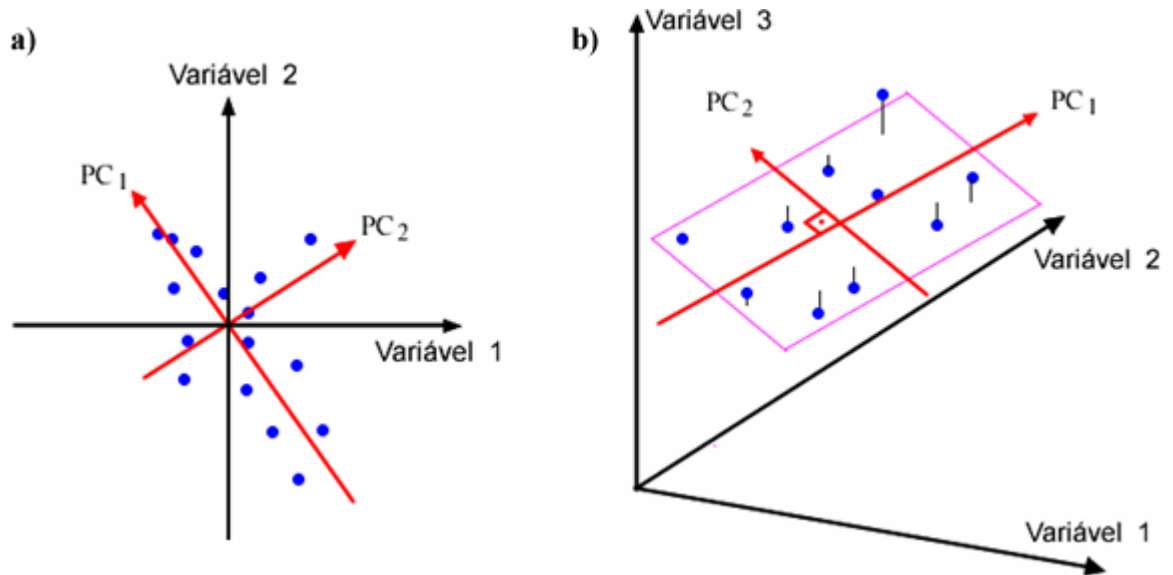


Figura 07 - Gráfico dos *scores* (PC1 versus PC2) a) para 2 variáveis e b) para 3 variáveis (Fonte: ARAUJO, 2005)

Se PC2 ainda não descrever toda a variância dos dados, a matriz dos resíduos e_2 é modelada dando PC3 ortogonal a PC2, e_3 , p_3^T e t_3 e, assim, sucessivamente, até que os resíduos seja zero ou desprezíveis (ARAUJO, 2005).

A PCA é comumente empregada, não apenas na identificação de grupos distintos, mas também na seleção de amostras anômalas na construção de modelos para calibração multivariada, contribuindo na seleção de variáveis. Os resíduos é a porção dos dados originais que não é descrita por um dado número de PCs. Sua magnitude e forma podem ser usadas para determinar o número de PCs do modelo (ARAUJO, 2005).

1.3.2.2 PLS

O método PLS (do inglês, *Partial Least Squares*), foi desenvolvido por Herman Wold na década de 70 e baseia-se nas propriedades do algoritmo NIPALS (*Nonlinear Iterative Partial Least Squares*) (NÆS; MARTENS, 1989; COSTA FILHO, 1998; MESSERSCHMIDT, 1999; PARREIRA, 2003).

O método de calibração multivariada mais comumente usada é indubitavelmente o PLS, e vem sendo cada vez mais utilizado em modelagem de dados, podendo ser usado para modelar sistemas que possuem comportamento linear ou não-linear (CHEN; CAI; SHAO, 2007).

O método PLS pode ser dividido em três etapas básicas: 1- montagem da matriz de dados X a partir dos dados espectrais e Y a partir das propriedades de interesse; 2- Eliminação de informações que não são relevantes ao sistema, através de combinações lineares das colunas da matriz X, e das combinações lineares da matriz Y, feita por PCA; 3- Obtenção de uma relação entre as matrizes reduzidas X e Y (COSTA FILHO, 1998; FIDÊNCIO, 2001).

O PLS decompõe as matrizes de dados X (variáveis independentes – onde estão os dados espectrais) e Y (variáveis dependentes – onde estão contidas as propriedades a serem analisadas) em uma soma de produtos de dois vetores (os “*scores*” e os “*loadings*”). Isso é possível, quando feita a redução de variáveis da matriz X e Y, assim, representando a matriz X e Y pelos seus *scores* sem perder informação estatística útil (FRANK; KALIVAS; KOWALSKI, 1983; NÆS; MARTENS, 1989; OTTO, 1999, SIMAS, R., 2005).

A dificuldade encontrada no método PCR é contornada pelo PLS, pois esse, além de utilizar a informação das variáveis independentes, também utiliza a informação das variáveis dependentes (*y*) (PARREIRA, 2003; ROGGO et al., 2007).

A performance do método depende das variáveis selecionadas para a construção do modelo de calibração, sendo apresentado no Subitem 4.4.1 deste trabalho alguns métodos para seleção dessas variáveis, tais como, comprimentos de ondas característicos das espécies químicas de interesse e eliminação de variáveis desprovidas de informações úteis (PASTI et al., 1998).

No PLS é utilizado, comumente, o termo “Variável Latente” (VL) para designar as componentes principais. Isto se deve ao fato da construção das mesmas serem feitas a partir de informações contidas no vetor das variáveis dependentes. O número de variáveis latentes que será utilizado é normalmente determinado durante o processo de validação cruzada (PARREIRA, 2003).

Quando se retiram ou se acrescentam novas amostras ao conjunto de calibração, os parâmetros do modelo criado inicialmente não se alteram significativamente. Essa é uma grande vantagem do método PLS (FERREIRA et al., 1999).

Para o cálculo dos *scores* e *loadings* podem ser usados vários algoritmos. O modelo resultante pode ser expresso como:

$$X = TP^T + E = \sum t_n p_n^T + E \quad (25)$$

$$Y = UQ^T + F = \sum u_n q_n^T + F \quad (26)$$

Em que T e U são os “scores” de X e Y respectivamente, P e Q são os “loadings” de X e Y respectivamente e, E e F são os erros residuais, ou seja, a variância que o modelo não conseguiu explicar (OTTO, 1999; LIMA, 2000; NAGATA; BUENO; PERALTA-ZAMORA, 2001; SIMAS, R., 2005).

As variáveis independentes podem ser relacionadas com as variáveis dependentes, correlacionando os “scores” do bloco X (T) com os “scores” do bloco Y (U) para cada componente principal (COSTA FILHO, 1998; BARTHUS, 1999; SENA, 2000; SIMAS, R., 2005).

Para tal é utilizado o seguinte modelo linear:

$$U = bT + \varepsilon \quad (27)$$

Assim,

$$U = b_n t_n \quad (28)$$

Em que $n = 1, 2, \dots, n$ para n componentes principais e,

$$b_n = u_n t_n / t_n^T t_n \quad (29)$$

Na equação (28), os valores de b_n são agrupados na matriz diagonal B que contém os coeficientes de regressão entre a matriz de scores U de Y e a matriz de scores T de X .

Na equação (27), b é o coeficiente de ajuste, ou seja, o coeficiente de regressão para cada VL, também conhecido como coeficiente angular. Finalmente, podemos substituir o valor de U da equação 27 na equação 26 e obter (COSTA FILHO, 1998; MESSERSCHMIDT, 1999):

$$Y = bTQ^T + G \quad (30)$$

E, portanto podemos obter os scores da matriz Y a partir dos scores da matriz X e vice-versa (MELLO, 1998).

O melhor modelo possível é aquele que consegue obter as menores matrizes de resíduo E e F , ao mesmo tempo que obtém a melhor relação linear entre t e u (LIMA, 2000).

Como essas componentes principais seriam calculadas para os dois blocos separadamente, este modelo não é o melhor possível, pois resultaria numa relação não muito satisfatória entre os *scores* desses dois blocos. Para manipular as informações desses dois blocos simultaneamente e obter a melhor correlação linear possível, no PLS são criadas as Variáveis Latentes (VLs) (COSTA FILHO, 1998; MESSERSCHMIDT, 1999), isso é obtido fazendo-se uma leve rotação dos eixos das PCs, ocasionando mudança nos valores dos *scores*. Essas VLs já não descrevem a direção de maior variância (mas sim uma direção próxima àquela), descrevendo as amostras nos espaços individuais e aumentando a correlação entre t e u (BARTHUS, 1999; FIDÊNCIO, 2001; SIMAS, R., 2005). A relação mista entre os dois blocos é feita na equação 30, onde o modelo final consiste das matrizes dos *scores* T e U que são linearmente relacionadas por um coeficiente b (b é diagonal).

Feita a calibração, o modelo é utilizado para a previsão do bloco dependente a partir do bloco independente conhecido na equação (30):

$$Y = bTQ^T + G$$

Sendo que, a “concentração” (as propriedades que se quer analisar) de novas amostras podem ser previstas a partir dos novos *scores*, T^* (SENA, 2000; NAGATA; BUENO; PERALTA-ZAMORA, 2001), substituindo na equação 30, tem-se:

$$Y = bT^*Q^T \tag{31}$$

O modelo PLS pode ser considerado como uma relação externa entre as matrizes X e Y individualmente e, de uma relação interna ligando os dois blocos, das variáveis independentes (X) e dependentes (Y) (MESSERSCHMIDT, 1999; ROGGO et al., 2007).

O método PLS foi escolhido para este trabalho devido ser o mais utilizado, tendo sido observado também o seu uso nos trabalhos que envolveram determinações de proteína em grãos inteiros de trigo (COZZOLINO et al., 2006), em café cru (MORGANO et al., 2005), em farelo de soja (SIMAS, R., 2005), em farinha de trigo (FERRÃO, 2000).

1.4 PROCESSAMENTO E TRATAMENTO DOS DADOS ESPECTRAIS

O tratamento dos dados é aplicado antes da fase de calibração para melhorar a eficiência do método, produzindo um modelo de regressão com a melhor capacidade de previsão de novas amostras. Os sinais analíticos são os dados obtidos de procedimentos e técnicas analíticas. Tais sinais provenientes de instrumentos apresentam-se misturados com ruído e

com informação não relevante, o que dificulta a interpretação e modelagem das propriedades associadas a esses. Por isso, faz-se o uso de processos matemáticos a fim de solucionar ou minimizar efeitos devido variações aleatórias ou sistemáticas que não estão relacionadas ao parâmetro de interesse. Diversos pré-tratamentos da resposta (sinal medido) têm sido relatados na literatura, como por exemplo: normalização, correção de linha base, cálculo da derivada, alisamento, correção multiplicativa de sinal. Determinações quantitativas a partir de espectros obtidos no infravermelho podem apresentar problemas de deformação da linha de base, variações no caminho óptico, ruídos, entre outros (MELLO, 1998; PEDRO, 2004; MORGANO, 2005; SOUSA, E., 2005).

Os espectros NIR devem ser convenientemente pré-processados e pré-tratados, de modo que reduzam informações indesejadas. Uma vez que informações espectrais importantes podem ser significativamente removidas a etapa de pré-processamento deve ser realizada com cautela (SOUSA, E., 2005). Alguns métodos usados nessa etapa são: centrar os dados espectrais na média; escalonamento (ou escalamento) pela variância e o autoescalamento dos dados (combinação dos dois métodos anteriores). Alguns métodos usados para o pré-tratamento dos espectros são: suavização e remoção de ruído instrumental, primeira e segunda derivadas (FREDERICO, 2003).

Os dados originais do conjunto de calibração e validação apresentam, muitas vezes, problemas de espalhamento e variação de linha base que precisam ser corrigidos se eles não contiverem informação relevante. Os espectros de validação devem receber o mesmo tratamento do que aqueles de calibração por serem idênticos (SIMAS, R., 2005).

1.4.1 Dados centrados na média

Este tipo de pré-processamento é um dos mais empregados na construção de modelos de calibração multivariada e consiste em fazer com que para cada variável seus valores tenham média zero, calculando para cada coluna da matriz X (comprimento de onda) o valor médio e subtrair cada intensidade do valor médio calculado. Dessa forma, ocorre a mudança do sistema de coordenadas para o centro dos dados (BARTHUS, 1999; FREDERICO, 2003, MORGANO, 2005; DU et al., 2005). A equação (32) é utilizada para centrar os dados na média.

Para uma coluna j :

$$\bar{x}_j = \sum_{i=1}^n \frac{x_{ij}}{n} \quad (32)$$

$$x_{ij(c)} = x_{ij} - \bar{x}_j \quad , \quad (33)$$

Em que $x_{ij(c)}$ é o valor centrado na média para a variável j na amostra i ; x_{ij} é o valor da variável j na amostra i ; e \bar{x}_j é a média dos valores das amostras na coluna j .

Como ocorre o deslocamento para o centro da informação descrita pelos dados originais, não há qualquer distorção da informação disponível, não alterando a informação para o modelo (PEDRO, 2004; MORGANO, 2005).

Uma representação gráfica do efeito de centralização na média sobre um conjunto de amostras com 3 variáveis (x_1, x_2, x_3) é ilustrada na Figura 08.

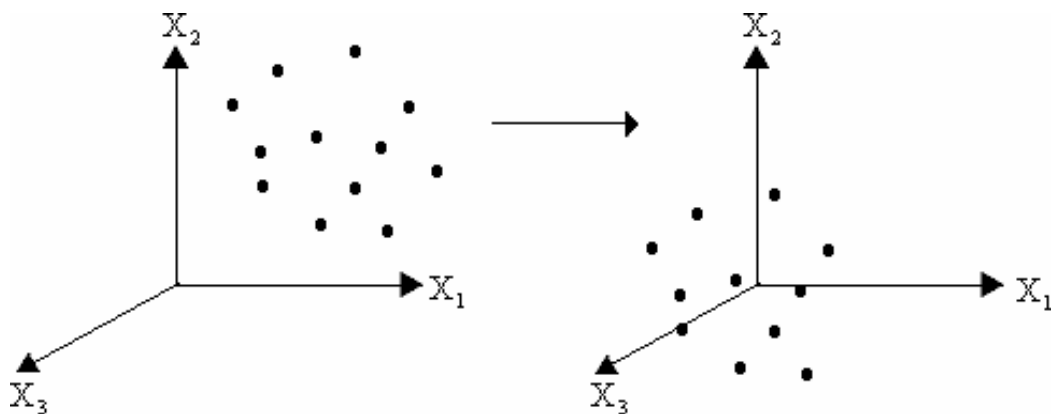


Figura 08 - Representação gráfica do cálculo para centrar os dados na média. (Fonte: BORIN, 2003. p. 32)

1.4.2 Correções de Linha base

Com intuito de corrigir a linha de base das amostras, utiliza-se o método de correção de linha base (*Transform Base Line*), que é amplamente utilizada como pré-tratamento em dados espectroscópicos e fundamenta-se em subtrair o valor de menor intensidade dos dados dos demais valores do espectro, daí o valor mínimo passa a ser zero e os demais valores tornam-se próximos a este valor (SOUSA, E., 2005).

1.4.3 Correção do Espalhamento Multiplicativo (MSC)

A não homogeneidade da amostra, principalmente pelas diferenças de granulometria, geometria, empacotamento e orientação das partículas, provoca durante a aquisição dos espectros por técnicas de reflexão fenômenos não desejados, como por exemplo, o espalhamento da luz. A presença de ruídos das mais diversas origens, que acompanham os sinais analíticos obtidos por técnicas instrumentais, pode dificultar a interpretação e assim, a

etapa de modelagem das propriedades associadas aos sinais espectrais (MELLO, 1998; FERRÃO, 2000; COSTA FILHO, 2003; ROGGO et al., 2007).

Para corrigir o efeito do espalhamento de luz presente nos espectros obtidos por técnicas de reflectância, utiliza-se o método de Correção do Espalhamento Multiplicativo – MSC (do inglês, *Multiplicative Signal Correction*), originalmente chamada de *multiplicative scatter (ou scattering) correction* (NÆS; MARTENS, 1989). Tal método é um processo matemático que minimiza a interferência relativa ao espalhamento de luz corrigindo a linha base dos espectros de reflectância difusa de modo que todos os espectros possuam o espalhamento “ideal” (entenda-se ideal como o espalhamento médio de todos os espectros formadores na matriz de dados \mathbf{X} , assim sendo, um espectro de referência) (LUYPAERT et al., 2002; AZZOUZ et al., 2003; PEDRO, 2004; PIZARRO et al., 2004).

O procedimento realizado pela técnica inicia-se calculando o espectro médio \bar{x} , daí, para cada espectro individual x_i , os parâmetros a_i e b_i são estimados por regressão em \bar{x} , resultando na equação da reta do tipo apresentado pela expressão abaixo (CANDOLFI et al., 1999; LUYPAERT et al., 2002; BORIN, 2003; MORGANO, 2005).

$$x_i = a_i + \bar{x}b_i \quad (34)$$

Por fim, os coeficientes linear (a_i) e angular (b_i) estimados são utilizados para corrigir cada um dos espectros originais, conforme a relação a seguir (CANDOLFI et al., 1999; FERRÃO, 2000; BORIN, 2003; MORGANO, 2005):

$$x_{i\text{ corrigido}} = \frac{x_i - a_i}{b_i} \quad (35)$$

Visto que neste estudo utilizou-se a técnica de infravermelho próximo de reflectância difusa em grãos de soja, obviamente apresentando partículas de granulometria elevada e diferente, bem como, considerando que a amostra não é homogênea, fez-se o emprego da correção do espelhamento de luz (que é uma das principais fontes de erro encontrada em determinações quantitativas nesta técnica) (MELLO, 1998; SIMAS, R., 2005), dentre outros métodos testados, a fim de minimizar o efeito do espalhamento de luz.

1.4.4 Filtros de Média

O ruído instrumental pode ser minimizado ou removido do sinal analítico gerado pelo instrumento, a suavização é realizada empregando-se certos tipos de filtros incorporados em

diversos *softwares* comerciais. Essas ferramentas de suavização são freqüentemente usadas para melhorar a relação sinal/ruído (LAGO; JULIANO; KASCHERES, 1995; MELLO, 1998; CERQUEIRA; POPPI; KUBOTA, 2000; PEDRO, 2004).

Dentre as técnicas de suavização que utilizam janelas, três apresentam grande destaque: suavização pela média, suavização pela média móvel e suavização Savitzky-Golay. O filtro de média móvel (*Moving Window Averaging Method*) é o tipo mais simples de alisamento utilizado para eliminar ruído de alta freqüência, e foi o estudado no tratamento dos espectros deste trabalho. Esse método consiste na aplicação de um algoritmo utilizado para tratamento dos dados, calculando a média dos valores contidos dentro de uma janela móvel que é deslocada ponto a ponto até atingir o final do espectro, assim, os dados contidos na janela são substituídos pela média calculada (NÆS; MARTENS, 1989; LAGO; JULIANO; KASCHERES, 1995; MELLO, 1998; CERQUEIRA; POPPI; KUBOTA, 2000).

O processo é executado na seguinte seqüência:

- a- Define-se a origem do sinal ($x_{i=1}$);
- b- Define-se o número de pontos do intervalo (n);
- c- Obtém-se a média dos pontos do intervalo:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (36)$$

- d- \bar{x} é o primeiro ponto do novo sinal \bar{x}_i
- e- Repete-se o processo (a-d) para $x_{i=2}$
- f- Atualiza-se o valor de i e repete-se a etapa até todo sinal ter sido suavizado

Um exemplo do processo de filtragem dos dados utilizando uma janela móvel de tamanho igual a 5 aplicado a um vetor \mathbf{X}_i , que poderia representar dados espectrais ou dados de qualquer outras espécies, é ilustrado na Figura 09. As medidas suavizadas vêm formar, então, o vetor \mathbf{X}_i (Filtrado).

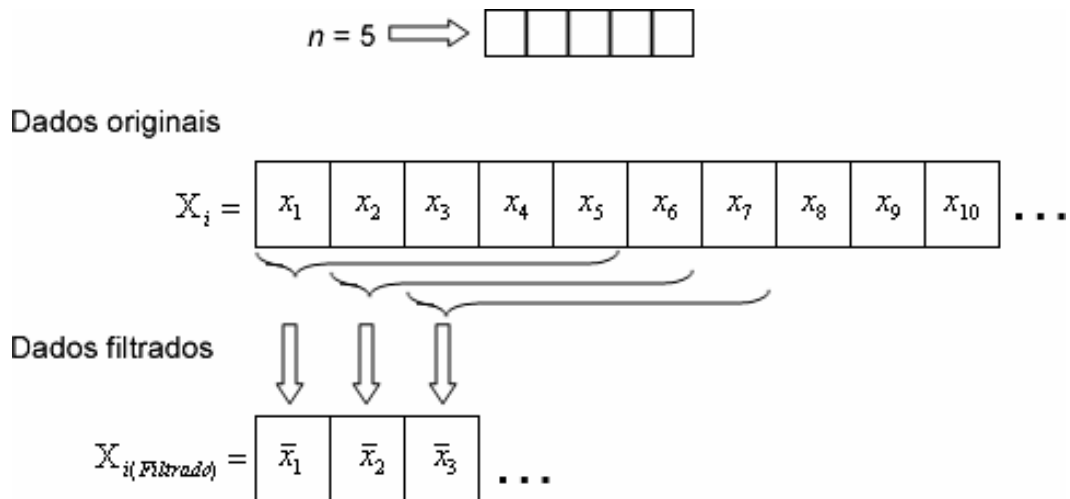


Figura 09 - Processo de filtragem utilizando média móvel.

O problema deste tipo de filtro é a distorção que ele provoca no sinal analítico. Este tipo de distorção é crítica em picos estreitos, porém não causa grandes transtornos em sinais de banda larga (LAGO; JULIANO; KASCHERES, 1995; CERQUEIRA; POPPI; KUBOTA, 2000; FREDERICO, 2003). Uma janela muito grande remove ruído, mas por outro lado, a forma da banda pode ser alterada causando perdas de informações relevantes (SOUSA, E., 2005). Assim sendo, a aplicação do filtro de média móvel é limitada a sinais de banda larga, em que a largura do intervalo necessário a redução do ruído, não é crítica (MELLO, 1998).

1.4.5 Derivadas

Devido a morfologia das partículas pode-se ter efeito de inclinações provocadas na linha de base dos espectros. Para minimizar esses efeitos e podendo também destacar ombros espectrais faz-se a aplicação da primeira ou segunda derivada sobre os dados espectrais. Com isso, as informações contidas ao longo dos diferentes comprimentos de onda são geralmente acentuadas. Contudo, a principal desvantagem desse processo é que o ruído uma vez presente nos espectros originais, pode se tornar acentuado, diminuindo a relação sinal/ruído por isso do cuidado com a qualidade dos espectros quando se pretende aplicar o cálculo das derivadas (FERRÃO, 2000; PIZARRO et al., 2004; MORGANO, 2005; CHEN, 2006).

Além de correção da linha base, a utilização de espectros de primeira e segunda derivadas evidencia a estrutura fina e permitem isolar bandas que são indistintas nos espectros originais, o que ajuda a resolver problemas de sobreposição espectral (CANDOLFI et al., 1999; FREDERICO, 2003; LUYPAERT et al., 2004).

Os gráficos resultantes das derivações não possuem mais a mesma forma dos espectros originais, guardando as características inerentes das derivadas. Os métodos de regularização de linha base pela aplicação de derivadas apresentam a desvantagem de diminuir a razão sinal/ruído e assim, quando os espectros apresentam regiões ruidosas, deve-se realizar um alisamento prévio eficiente (este pode ser facultado uma vez que a derivada de Savtzky-Golay é acompanhada de um filtro) (PEDRO, 2004).

Vários algoritmos têm sido empregados para o cálculo das derivadas, entre eles destacam-se a derivada de Norris e a derivada de Savtzky-Golay (SOUSA, E., 2005), essa última, foi a empregada para a aplicação de primeira derivada nos espectros aqui em estudo visando encontrar o melhor modelo.

1.4.6 Transformação Padrão Normal de Variação (SNV)

É um método de transformação matemática aplicada para corrigir as interferências de espalhamento e os efeitos do espalhamento multiplicativo entre diferentes amostras (GUO; WU; MASSART, 1999; AZZOUZ et al., 2003; OLIVEIRA, 2004; LUYPAERT et al., 2004), a transformação padrão normal de variação (SNV – do inglês, *Standard Normal Variate Transformation*) tem função análoga à MSC. Contudo, apesar do SNV e MSC terem a mesma finalidade, ou seja, corrigir a linha base, estas duas técnicas são distintas (COSTA FILHO, 2003). A correção feita pelo SNV remove o espalhamento através da normalização de cada espectro pelo desvio padrão; isto é, ela é realizada subtraindo-se o espectro inteiro ($X_{i,j}$) pela média desse mesmo espectro (\bar{x}_i) e então o valor é dividido pelo seu próprio desvio padrão (CANDOLFI et al., 1999; LUYPAERT et al., 2002; HERRERA; GUESALAGA; AGOSIN, 2003; PIZARRO et al., 2004), sendo assim, o SNV não necessita de um espectro “ideal” para fazer a correção dos espectros (COSTA FILHO, 2003).

As equações a seguir, ilustram o processo realizado pelo SNV:

$$\text{Média do espectro: } \bar{x}_i = \sum_{j=1}^p \frac{X_{i,j}}{p} \quad (37)$$

$$\text{Espectro corrigido: } x_{i(SNV)} = \frac{(X_{i,j} - \bar{x}_i)}{\sqrt{\frac{\sum_{j=1}^p (X_{i,j} - \bar{x}_i)^2}{p-1}}} \quad (38)$$

Em que \mathbf{X} representa uma matriz com n espectros e p comprimentos de onda, \bar{x}_i é a média do vetor contendo o espectro i da matriz \mathbf{X} .

Como este método corrige cada espectro independentemente, ele pode ser aplicado em casos onde ocorra maior variabilidade entre um conjunto de espectros (COSTA FILHO, 2003).

1.4.7 Correção do Espalhamento Multiplicativo Estendido (EMSC)

Outro método que permite corrigir espalhamentos de sinais é a Correção do Espalhamento Multiplicativo Estendido (EMSC - do inglês, *Extended multiplicative signal correction*), Desenvolvido por Martens *et al.* esse método de pré-processamento permite a separação dos efeitos físicos de espalhamento de luz dos efeitos químicos (vibracional) de absorção da luz no espectro (WILLIAMS; NORRIS, 2001; MARTENS; NIELSEN; ENGELSEN, 2003, SÁIZ-ABAJO, 2005; CHEN; MORRIS; MARTIN, 2006).

Esta técnica é capaz de estimar e separar efeitos físicos multiplicativos (tamanho da amostra, espalhamento de luz) dos efeitos químicos (absorbância de analitos e interferentes) e físicos (mudança de temperatura, variações de linha base, etc.) adicionais (SÁIZ-ABAJO, 2005).

A EMSC também pode ser usada para remover efeitos de interferentes físicos e químicos identificados, porém indesejados, enquanto mantendo identificados, mas efeitos desejados como também efeitos indesejados nos dados. Por esses propósitos, a EMSC permite o uso de conhecimento prévio sobre o sistema e seus componentes (constituintes espectrais) na correção, que pode às vezes ser bem útil e bons rendimentos nos resultados de calibração (SÁIZ-ABAJO, 2005; CHEN; MORRIS; MARTIN, 2006).

A EMSC parece ser aplicável para diferentes tipos de dados espectroscópicos (UV, VIS, NIR, IR, Raman), cromatografia, eletroforeses e dados sensoriais (SÁIZ-ABAJO, 2005).

1.5 QUÍMICA VERDE

A utilização de substâncias muitas vezes tóxicas e/ou inflamáveis, que após um processo químico normalmente geram um “lixo tóxico” (resíduo), tornam algumas atividades na área de química potenciais causadoras de risco e poluição. Dentre as alternativas prováveis/possíveis para o destino dos resíduos, esses podem ser encaminhados para as

estações de tratamento, podem ser reciclados ou reutilizados, ou ainda, serem incinerados depois de tratados (SILVA, F.; LACERDA; JUNIOR, 2005).

A Química Verde (do inglês, “*Green Chemistry*”), também conhecida como Química Limpa, é um tipo de prevenção da poluição causada por atividades na área de química, sendo uma estratégia importante no tocante ao problema do meio ambiente (SILVA, F.; LACERDA; JUNIOR, 2005). A Química Verde busca o desenvolvimento e utilização de produtos químicos e processos para reduzir ou eliminar o uso de geração de substâncias nocivas à saúde e ao ambiente (LENARDÃO et al., 2003). Os principais pontos (ou os princípios elementares) da química verde são basicamente definidos por autores (LANCASTER, 2002; LENARDÃO et al., 2003; AHLUWALIA; KIDWAI, 2004; SILVA, F.; LACERDA; JUNIOR, 2005) em 12 tópicos, os quais são citados a seguir:

1. Prevenção de resíduos;
2. Economia de átomos (máxima incorporação de toda a massa dos reagentes no produto);
3. Síntese de produtos menos perigosos;
4. Projeto de produtos seguros;
5. Solventes e reagentes auxiliares mais seguros;
6. Eficiência energética;
7. Uso de fontes renováveis de matéria-prima;
8. Evitar a formação de derivados;
9. Uso de catalisadores ou subprodutos sempre que possível;
10. Produtos biodegradáveis;
11. Análise em tempo real para a prevenção da poluição;
12. Química inerentemente segura para a prevenção de acidentes.

O princípio número 1 baseia-se no fato de que evitar a produção de resíduo é melhor do que tratá-lo ou limpá-lo após sua geração (LANCASTER, 2002; JUNIOR, 2005; LENARDÃO et al., 2003; AHLUWALIA; KIDWAI, 2004; SILVA, F.; LACERDA). Os resíduos (ou subprodutos) se descartados podem causar poluição, além de requererem, posteriormente, tratamento. Assim, a redução na fonte, é um modo mais eficiente de minimizar o impacto ambiental, além disso, o investimento em tecnologias mais limpas, diminui os investimentos no tratamento de resíduos, os quais nem sempre resolvem satisfatoriamente o problema (LENARDÃO et al., 2003; AHLUWALIA; KIDWAI, 2004).

2 OBJETIVO

Este trabalho teve o seguinte objetivo:

- Determinar os teores de proteína, de óleo e de umidade em grãos inteiros de 09 (nove) cultivares de soja produzidas em Roraima utilizando a técnica de espectroscopia na região do infravermelho próximo (NIR) com reflectância difusa aliada à regressão por mínimos quadrados parciais (PLS).

3 JUSTIFICATIVAS

O método Kjeldahl é um método convencional, oficialmente recomendado para quantificação de proteína, baseado na determinação de nitrogênio total, sendo o mais comum e largamente utilizado para análises de rotina de um largo número de amostras (grãos, carne, e outros materiais biológicos). Apesar de ser considerado um dos mais exatos, o método Kjeldahl requer muitas etapas, demandando tempo, além de ser perigoso e dispendioso, utilizando mais de um equipamento, consumindo muitos reagentes, e catalisadores tóxicos, gerando resíduos nocivos ao ambiente (FERRÃO, 2000; WILLIAMS; NORRIS, 2001; SKOOG et al., 2004).

Para a quantificação do teor de óleo é comumente utilizado o método de extração por Soxhlet, que tem como grande incômodo o tempo gasto nas análises e grande quantidade de solvente utilizado, além do gasto energético. Já na quantificação do teor de umidade o método mais usado é o da estufa, a qual deve permanecer ligada em média por 24 horas (FERRÃO, 2000; WILLIAMS; NORRIS, 2001).

A técnica espectroscópica de infravermelho próximo com reflexão difusa (NIRRS) aliada ao PLS pode ser utilizada como um método alternativo para determinação do teor de proteínas, de óleo e de umidade, apresentando diversas vantagens sobre os métodos de referência como a não destruição da amostra, o menor tempo de análise, a não utilização de reagentes e a não geração de resíduos, assim, não agredindo o ambiente, tornando-se uma técnica limpa e de acordo com a visão de Química Limpa, a qual inclui, dentre outros, o princípio de evitar a produção do resíduo em vez de tratá-lo ou “limpá-lo” após sua geração (LENARDÃO et al., 2003; SILVA, F.; LACERDA; JUNIOR, 2005)

A aplicação da espectroscopia NIR em soja representa um avanço uma vez que esta é um dos principais itens de exportação agrícola do Brasil e vem ganhando espaço econômico em Roraima, sendo sua produção praticamente toda voltada para a exportação, além de ser considerada um alimento funcional (SCAFI, 2000). Além do fato de não ter sido encontrado na literatura nacional estudos referentes à utilização da técnica NIRRS visando a determinação de proteína, óleo, e umidade em grãos inteiros se soja.

4 PARTE EXPERIMENTAL

4.1 AMOSTRAGEM

Estudou-se 9 cultivares de soja, sendo elas: BRS Raimunda, BRS MA Tracajá, BRS 252 Serena, Embrapa 63 Mirador, BRS Luziânia, BRS MG Nova Fronteira, Celeste, BRS Candeia, MA/BR 65 Sambaiba (GIANLUPPI et al., 2000a, 2000b, 2000c, 2002, 2003, 2004a, 2004b, 2005, 2006). Essas cultivares foram numeradas de 1 a 9, respectivamente.

As amostras foram obtidas da Embrapa Roraima (Empresa Brasileira de Pesquisa Agropecuária), cultivadas no Campo Experimental Água Boa (com exceção da nº 9, que foi cultivada na Fazenda de Bonfim).

As cultivares foram colhidas entre setembro e outubro de 2004 e acondicionadas a $25 \pm 3^\circ\text{C}$ e aproximadamente 75% de Umidade Relativa (com exceção da nº 8, colhida entre setembro e outubro de 2005 e acondicionada em câmara fria a $15 \pm 2^\circ\text{C}$ e Umidade Relativa de 65%).

4.2 MÉTODOS DE REFERÊNCIA

4.2.1 Determinação da Umidade

4.2.1.1 Reagentes

O método utilizado não necessitou de nenhum reagente específico.

4.2.1.2 Materiais e equipamentos

Dessecador de vidro; Cápsulas de metal com tampas; Balança Analítica (Tecnal); Estufa (Biomatic); Termômetro; Luvas descartáveis; Pinça metálica; Grãos de soja.

4.2.1.3 Procedimentos experimentais

Inicialmente colocou-se as cápsulas de metal com as tampas em estufa a 130°C por uma hora. Posteriormente, retirou-se as cápsulas da estufa e transferiu-se para o dessecador por uma hora, em seguida, retirou-se uma a uma, pesado-as. Pesou-se 10 gramas de grãos de soja em cada cápsula. Colocou-se as cápsulas novamente na estufa a uma temperatura de 105°C por 30 horas. Depois de retirar as amostras da estufa, tampou-se os recipientes e os colocou por 30 minutos no dessecador, por fim, pesou-se os recipientes com as amostras.

4.2.2 Determinação do Óleo

4.2.2.1 Reagentes

Éter sulfúrico.

4.2.2.2 Materiais e equipamentos

Aparelho do tipo Soxhlet; Balança Analítica (Tecnal); Estufa (Biomatic); Grampeador; Dessecador apropriado; Capela para exaustão de gases; Papel-filtro qualitativo; Pinça de metal; Rotaevaporador; Amostra;

4.2.2.3 Procedimentos experimentais

Inicialmente, pesou-se 2,0 gramas (M_1) em farinha de uma amostra diretamente no papel-filtro, enrolando-o, de forma que a amostra ficasse em seu interior, dobrando e grampeando as extremidades para que evitasse perdas da amostra. Em seguida, o cartucho (papel-filtro + amostra) foi colocado em estufa a 105 °C durante 2 horas.

O balão do extrator de Soxhlet também foi colocado em estufa a 105 °C durante 1 hora; depois, deixou-se esfriar em dessecador por 10 minutos, e pesou-se o balão (M_2).

Após retirar o cartucho da estufa, o mesmo foi colocado dentro da parte intermediária do extrator de Soxhlet. Então adicionou-se aproximadamente 170 mL de Éter ao balão do extrator conectando-o a parte intermediária do extrator e ajustando o conjunto ao condensador.

Abriu-se a torneira para refrigeração do condensador e iniciou-se o aquecimento de tal forma a obter de 2 a 4 gotas por segundo.

O sistema em refluxo foi mantido por 6 horas. O éter foi recuperado em frasco apropriado, e sempre que possível reutilizado nas demais análises.

O balão com o extrato foi colocado em estufa a 80° C por 2,5 horas; depois colocado em dessecador por 10 minutos e pesado (M_3).

4.2.3 Determinação da Proteína

4.2.3.1 Reagentes

Ácido sulfúrico (H_2SO_4 $d = 1,84$), p.a., concentrado; Hidróxido de sódio, a 40% (m/v); Ácido sulfúrico 0,05 mol L^{-1} ; Mistura catalisadora: Sulfato de sódio (Na_2SO_4), Sulfato de cobre (Cu_2SO_4) e Selênio em pó (Se); Solução receptora indicadora: Solução de ácido bórico (H_3BO_3), a 2% (m/v), Solução alcoólica de verde de bromocresol, a 0,1% (m/v) e Solução

alcoólica de vermelho de metila, a 0,1% (m/v). Todas as soluções preparadas utilizando-se água desionizada.

4.2.3.2 Materiais e Equipamentos

Bloco digestor; Capela para exaustão de gases; Microdestilador de Kjeldahl; Bureta de 5 mL; Erlenmeyers de 125 mL; Espátula de aço; Balança Analítica (Quimis Q-510-1500); Tubos de ensaio com borda reforçada (25 x 250 mm); Pipeta automática de 5 mL; Pipeta graduada de 5 e 10 mL; Frasco dosador de reagentes; Pêra de borracha e Amostra.

4.2.3.3 Procedimentos experimentais

Inicialmente foi realizada a digestão da amostra, em que foram pesados aproximadamente 0,2000 g em farinha da amostra e, em seguida, colocou-a no tubo de ensaio (com exceção de 3 tubos, os quais eram os “brancos” e não contiveram as amostras). Depois adicionou-se a cada tubo 2 g de mistura catalisadora e, na capela de exaustão, foram adicionados em cada tubo 5 mL de ácido sulfúrico concentrado, em seguida os tubos foram transferidos para o bloco digestor, onde aumentou-se gradativamente a temperatura até atingir 350 °C, e foram mantidos em aquecimento até a mistura ficar completamente clara (com coloração verde brilhante), o tempo total sob aquecimento foi de 2,5 horas. Durante a digestão agitou-se os tubos por duas vezes de modo que qualquer material aderido nas paredes voltasse à mistura e então fosse digerido completamente.

Posteriormente, retirou-se os tubos do bloco de digestão deixando-os esfriar à temperatura ambiente, colocou-se 20 mL de água desionizada em cada tubo para evitar a cristalização da mistura. Então os tubos foram levados imediatamente para a etapa de destilação.

Então colocou-se 10 mL da solução receptora-indicadora em erlenmeyer de 125 mL, Mergulhou-se o bico de saída do condensador do sistema de destilação na solução receptora indicadora, contida no erlenmeyer. Depois pegou-se um tubo com a mistura, adicionou 15 mL de hidróxido de sódio ao tubo, encaixou-se o tubo diretamente no microdestilador e iniciou-se a destilação, somente finalizando-se quando a solução do erlenmeyer totalizou 50 mL. Procedimento feito para cada um dos 30 tubos.

Por fim realizou-se a titulação, onde o conteúdo do erlenmeyer foi titulado usando uma bureta carregada com ácido sulfúrico diluído e padronizado; observou-se a coloração rósea clara no ponto de viragem e anotou-se o volume gasto na titulação.

4.3 OBTENÇÃO DOS ESPECTROS

Os espectros de reflectância na região NIR foram obtidos no Instituto de Química da UNICAMP, num espectrofotômetro BOMEM MD-160, com resolução de 8 cm^{-1} , sendo realizada 5 réplicas (ou seja, 9 amostras de cultivares de soja, de cada amostra obteve-se 3 replicatas, e de cada replicata obteve-se 5 espectros, totalizando 135 espectros) devido as amostras terem propriedades físicas - empacotamentos, granulometria, etc. - diferentes por serem grãos inteiros e, assim, obter espectros mais representativos das amostras para cada replicata foi obtido o espectro médio de 100 varreduras na região espectral de 14000 a 4000 cm^{-1} (714 a 2500 nm). O instrumento foi equipado com um acessório de reflectância difusa, e os sinais expressos em $\log(1/R)$. O programa utilizado foi o Win-Bomem Easy 3.04b. Para a aquisição dos espectros, as amostras foram colocadas em frascos de vidros de 30 mL com paredes transparentes e uniformes.

Cada espectro foi importado para o programa computacional OriginPro 7.5 resultando em uma matriz 135x2595, depois a matriz foi transportada para o programa computacional The Unscrambler® v9.2. Este programa foi empregado para o pré-tratamento dos espectros e construção dos modelos de calibração e de validação.

4.4 DESENVOLVIMENTOS DOS MODELOS DE CALIBRAÇÃO

4.4.1 Seleção de faixa espectral

Neste trabalho, foram explorados a seleção de faixas espectrais pelos comprimentos de ondas onde ocorrem as absorções características dos analitos em estudo, espectros completos (sem os comprimentos de ondas de 714 a 1010 nm para os quais os sinais são muito afetados pelo ruído instrumental), e outras faixas do espectro (a fim de se investigar as melhores variáveis para a construção do modelo de calibração).

4.4.2 Seleção de espectros

Os espectros foram nomeados da seguinte maneira, tomando como base *XYZ*, em que *X* adquiriu valor de **1** a **9** (representando cada uma das nove cultivares); *Y* adquiriu valor de **1** a **3** (representando cada uma das três amostras referentes a cada cultivar); e *Z* a qual foi atribuído letras **a**, **b**, **c**, **d** e **e** (representando cada um dos cinco espectros tirados para cada amostra). Com isso, o *espectro 42e* se refere ao quinto espectro tirado da segunda amostra

proveniente da cultivar quatro. Assim, dividiu-se os 135 espectros em 3 conjuntos diferentes a fim de se buscar o melhor conjunto de espectros para os modelo de calibração e validação.

Conjunto 1: Neste conjunto fez-se uso dos espectros *X1Z* e *X2Z* (sombreados e em negrito na Tabela 04), no conjunto de calibração; e os demais fizeram parte do conjunto de previsão.

Tabela 04 - Espectros usados no conjunto de calibração e previsão – Conjunto 1.

11a	21a	31a	41a	51a	61a	71a	81a	91a
11b	21b	31b	41b	51b	61b	71b	81b	91b
11c	21c	31c	41c	51c	61c	71c	81c	91c
11d	21d	31d	41d	51d	61d	71d	81d	91d
11e	21e	31e	41e	51e	61e	71e	81e	91e
12a	22a	32 ^a	42a	52a	62a	72a	82a	92a
12b	22b	32b	42b	52b	62b	72b	82b	92b
12c	22c	32c	42c	52c	62c	72c	82c	92c
12d	22d	32d	42d	52d	62d	72d	82d	92d
12e	22e	32e	42e	52e	62e	72e	82e	92e
13a	23a	33 ^a	43a	53a	63a	73a	83a	93a
13b	23b	33b	43b	53b	63b	73b	83b	93b
13c	23c	33c	43c	53c	63c	73c	83c	93c
13d	23d	33d	43d	53d	63d	73 d	83d	93d
13e	23e	33e	43e	53e	63e	73e	83e	93e

Conjunto 2: Neste conjunto fez-se uso dos espectros *X1Z* e *X3Z* (sombreados e em negrito na Tabela 05), no conjunto de calibração; e os demais fizeram parte do conjunto de previsão.

Tabela 05 - Espectros usados no conjunto de calibração e previsão – Conjunto 2.

11a	21a	31 ^a	41a	51a	61a	71a	81a	91a
11b	21b	31b	41b	51b	61b	71b	81b	91b
11c	21c	31c	41c	51c	61c	71c	81c	91c
11d	21d	31d	41d	51d	61d	71d	81d	91d
11e	21e	31e	41e	51e	61e	71e	81e	91e
12a	22a	32a	42a	52a	62a	72a	82a	92a
12b	22b	32b	42b	52b	62b	72b	82b	92b
12c	22c	32c	42c	52c	62c	72c	82c	92c
12d	22d	32d	42d	52d	62d	72d	82d	92d
12e	22e	32e	42e	52e	62e	72e	82e	92e
13a	23a	33a	43a	53a	63a	73a	83a	93a
13b	23b	33b	43b	53b	63b	73b	83b	93b
13c	23c	33c	43c	53c	63c	73c	83c	93c
13d	23d	33d	43d	53d	63d	73d	83d	93d
13e	23e	33e	43e	53e	63e	73e	83e	93e

Conjunto 3: Neste conjunto fez-se uso dos espectros *X2Z* e *X3Z* (sombreadas e em negrito na Tabela 06), no conjunto de calibração; e os demais fizeram parte do conjunto de previsão.

Tabela 06 – Espectros usados no conjunto de calibração e previsão – Conjunto 3.

11a	21a	31a	41a	51a	61a	71a	81a	91a
11b	21b	31b	41b	51b	61b	71b	81b	91b
11c	21c	31c	41c	51c	61c	71c	81c	91c
11d	21d	31d	41d	51d	61d	71d	81d	91d
11e	21e	31e	41e	51e	61e	71e	81e	91e
12a	22a	32a	42a	52a	62a	72a	82a	92a
12b	22b	32b	42b	52b	62b	72b	82b	92b
12c	22c	32c	42c	52c	62c	72c	82c	92c
12d	22d	32d	42d	52d	62d	72d	82d	92d
12e	22e	32e	42e	52e	62e	72e	82e	92e
13a	23a	33a	43a	53a	63a	73a	83a	93a
13b	23b	33b	43b	53b	63b	73b	83b	93b
13c	23c	33c	43c	53c	63c	73c	83c	93c
13d	23d	33d	43d	53d	63d	73d	83d	93d
13e	23e	33e	43e	53e	63e	73e	83e	93e

4.4.3 Número de variáveis latentes

O método mais usado para a determinação do número mais adequado de VLS é a validação cruzada (CV – do inglês, *Cross Validation*). Neste procedimento, deixa-se uma amostra de fora do processo (método conhecido como “*leave one out*”, validação cruzada completa), de uma série de **n** amostras, **n-1** amostras são usadas na calibração e a seguir essa amostra é prevista pelo modelo. O processo é repetido até que todas as amostras tenham sido excluídas uma vez e previstas. Compara-se a concentração de cada amostra prevista pelo modelo com o valor de referência e determina-se o erro, como o da raiz quadrada da soma do quadrado dos erros da validação cruzada - RMSECV (do inglês, *Root Mean Square Error of Cross Validation*), representada pela equação (39) (HERRERA; GUESALAGA; AGOSIN, 2003; PARREIRA, 2003; PEDRO, 2004, BRAGA, 2004).

$$RMSECV = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \quad (39)$$

Os resultados de RMSECV são plotados em um gráfico em função do número de VLs (de 1 até A), o comportamento típico para esses gráficos é a observação de um mínimo ou um patamar, que indica a melhor dimensionalidade do modelo de regressão, ou seja, o número de VLs que produziu o menor erro de previsão sem perda significativa de variância dos dados (BRAGA, 2004; VALDERRAMA, 2005).

Um outro método usado para escolha do número mais adequado de variáveis latentes é observar os valores de *PRESS* (do inglês, *Prediction Residual Error Sum of Squares*), que é o Somatório Quadrático dos Erros de Predição, calculado utilizando a equação (40). Assim, como o cálculo é baseado nos resíduos, quanto menor o valor do *PRESS*, mais bem ajustado está o modelo com a quantidade de VLs utilizada (CHAN et al., 2007).

$$PRESS = \sum (y_i - \hat{y}_i)^2 \quad (40)$$

Em que y_i é o valor medido para cada uma das amostras, e \hat{y}_i é o valor previsto para cada amostra.

4.4.4 Amostras anômalas

A identificação de amostras anômalas (*outliers*) pode ser feita através de um gráfico de resíduos Student pela *leverage* (PEDRO, 2004). O *leverage* representa o grau que uma amostra está distante da média do conjunto de dados, ou seja, o “peso relativo” de uma amostra em relação às demais presentes em um mesmo conjunto. Tomando como exemplo dados espectrais, qualitativamente, o *leverage* mede o quanto o espectro de uma amostra difere dos espectros das demais amostras presentes no conjunto de dados. O *leverage* pode ser representado pela equação a seguir (BRAGA, 2004; VALDERRAMA, 2005):

$$h_i = \hat{t}_i^T (\hat{T}^T \hat{T})^{-1} \hat{t}_i \quad (41)$$

Onde T são os valores de todas as amostras de calibração; t_i é o vetor de *scores* de uma amostra em particular.

Um parâmetro de *leverage* crítico é calculado (equação 42) para facilitar a identificação destas amostras (PEDRO, 2004, BRAGA, 2004).

$$h_{crit} = \frac{\alpha \times A}{N} \quad (42)$$

Em que a definição de α varia de autor pra autor, sendo mais usual $\alpha=2$ ou $\alpha=3$ (PEDRO, 2004) (neste trabalho empregou-se $\alpha=3$); N representa o número de amostras da

calibração; e A é o número de componentes principais, também sendo relatado o uso de $(A + 1)$ (VALDERRAMA, 2005).

Amostras com altos valores de h podem ser consideradas como amostras anômalas (VALDERRAMA, 2005). Amostras extremas podem apresentar *leverage* acima do valor crítico, desde que possuam resíduos de Student moderados. Neste caso, podem ser essenciais à calibração, não constituindo, todavia, amostras anômalas (MESSERSCHMIDT, 1999; BORIN, 2003; PEDRO, 2004). Este caso é comum quando um conjunto de amostras apresenta poucos valores elevados ou diminutos para a propriedade de interesse (PEDRO, 2004). O gráfico da concentração estimada na validação cruzada pela concentração dada pelo método de referência também é utilizado na localização de amostras anômalas (MESSERSCHMIDT, 1999).

4.5 OTIMIZAÇÃO E VALIDAÇÃO DOS MODELOS DE CALIBRAÇÃO

Os modelos de calibração multivariada relacionado a matriz X (valores de reflectância em vários comprimentos de onda) e a matriz Y (teores dos analitos determinados em experimentos) foram construídos utilizando o método de calibração multivariada PLS, com o programa The Unscrambler v9.2.

Em todos os modelos construídos empregaram-se os dados centrados na média, e testaram-se diversos métodos de pré-tratamento e pré-processamento dos dados espectrais, tais como: correção de linha base (BL), filtro média móvel (usando largura do intervalo igual a três, cinco, sete, onze e quinze pontos), MSC, EMSC, SNV, primeira derivada; e algumas combinações destas: média móvel + MSC; SNV + primeira derivada; MSC + primeira derivada; EMSC + primeira derivada; BL + primeira derivada; BL + MSC; MSC + BL; primeira derivada + MSC; e primeira derivada + EMSC; MSC + segunda derivada; EMSC + segunda derivada; segunda derivada + MSC; e segunda derivada + EMSC.

Um modelo de calibração pode estar tão bem ajustado aos dados de calibração que suas previsões são falhas; ou ainda, pode não incluir variância suficiente das amostras de calibração e, assim, suas previsões falharem. Para que um modelo seja eficiente é necessário que ele descreva uma situação real da melhor forma possível. Então, uma vez determinado o modelo de regressão é necessário que se faça a sua otimização e validação (TALHAVINI, 2003; PEDRO, 2004; MORGANO, 2005), com o objetivo de escolher o modelo de calibração mais exato e preciso possível (BURNS; CIURCZAK, 2001). Por isso divide-se o conjunto de amostras em três subconjuntos, sendo um conjunto empregado para a construção dos modelos

de regressão. Um outro conjunto empregado para a escolha do melhor modelo em termos de exatidão e precisão, podendo usar o mesmo conjunto usado na etapa anterior desde que o modelo de calibração “ótimo” seja escolhido usando *cross validation* (que foi o empregado neste trabalho). E um conjunto é empregado para estimar a performance do modelo escolhido em futuras amostras (BURNS; CIURCZAK, 2001).

Para uma avaliação prévia dos modelos de calibração podem ser utilizados como parâmetros os coeficientes de correlação. O coeficiente de correlação entre os espectros NIR e os valores medidos pelo método de referência é calculado como mostrado pela equação (43), onde y é a média dos resultados de referência para todas as amostras no conjunto de teste e de treinamento (BURNS; CIURCZAK, 2001; CHEN, 2006). O valor desejável para o coeficiente de correlação (também chamado de coeficiente de determinação, ou RSQ, ou ainda, R^2 obtido após elevar ao quadrado ambos os membros da equação) é que este se situe o mais perto de 1 (BURNS; CIURCZAK, 2001; SIMAS, R., 2005).

$$R = \sqrt{1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (43)$$

Além do coeficiente de correlação, existem três parâmetros estatísticos geralmente empregados para comparação da performance de modelos de calibração multivariada: o RMSEC (*root mean squared error of calibration*), o RMSECV (*root mean squared error of cross validation*), e o RMSEP (*root mean squared error of prediction*) (BURNS; CIURCZAK, 2001). Todos os esses três parâmetros são baseados na raiz quadrada dos erros médios (RMSE) descrita na equação 44.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (44)$$

Onde \hat{y}_i é o valor estimado da concentração da i -ésima amostra; n é a quantidade de amostras usadas no modelo; $\sum (y_i - \hat{y}_i)^2$ é o Somatório Quadrático dos Erros de Predição (PRESS – *Prediction Residual Error Sum of Squares*).

O RMSEC, RMSECV, e RMSEP diferem na determinação do \hat{y}_i , em virtude do conjunto de amostras utilizado.

O RMSEC é uma medida de como o modelo de calibração ajusta o conjunto de calibração, sendo potencialmente o parâmetro menos informativo das três estatísticas. A melhor estimativa do futuro desempenho do modelo de calibração é o RMSEP (BURNS; CIURCZAK, 2001). É comum utilizar-se como parâmetros, em vez do RMSEC, RMSECV e RMSEP, respectivamente, o erro padrão de calibração SEC (do inglês, *standard error of calibration*), erro padrão de validação cruzada SECV (do inglês, *standard error of cross validation*) e erro padrão de previsão SEP (do inglês, *standard error of prediction*), calculados como sendo a raiz quadrada da soma dos quadrados de erros entre os valores de referência e os valores previstos, dividido pelo número de graus de liberdade (SIMAS, R., 2005), como expressado na equação (45).

$$SE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - A - 1}} \quad (45)$$

Onde A é o número de VLs no modelo.

De qualquer forma, comparando as equações (44) e (45), o valor de SE será matematicamente maior que o valor de $RMSE$ pelo fato do denominador ser menor na equação (45) do que aquele da equação (44), assim sendo, na comparação entre dois modelos, o que apresentar menor RMSEP apresentará também menor SEP. Com isso, resolveu-se aqui avaliar os modelos usando além do R, o RMSEC, RMSECV, e o RMSEP.

O método de validação externa e o de validação interna são os mais usados para validar o modelo de regressão. No primeiro método, como descrito anteriormente, uma amostra é usualmente removida do conjunto de calibração de cada vez e n modelos de calibração construídos, o valor dessa amostra é previsto pelo modelo e comparado com o valor encontrado pelo método de referência, e o RMSECV é calculado. Já no segundo método, um certo número de amostras é removido do conjunto de calibração e utilizado para predição posterior. E a performance do modelo é medida calculando-se o RMSEP. (BURNS; CIURCZAK, 2001; PASQUINI, 2003; PEDRO, 2004; CHEN, 2006). A validação é a fase final no desenvolvimento de um modelo quantitativo. O uso de um outro conjunto de validação externa de amostras e o RMSEP é recomendado porque esse produz resultados mais realísticos que são independentes de um dado empregado na construção do modelo e durante a otimização da dimensão do modelo (PASQUINI, 2003).

O RMSEP é usado para avaliar a capacidade de previsão dos modelos construídos, geralmente, empregado para comparar resultados obtidos por diferentes modelos

(TALHAVINI, 2003). Um bom modelo deve ter um alto valor de R , um baixo RMSEC, um baixo RMSECV e um baixo RMSEP (HERRERA; GUESALAGA; AGOSIN, 2003).

Um outro fator que pode ser levado em conta é a habilidade do modelo em prever resultados que não discordem muito dos valores medidos, para isso, tem-se o Erro Padrão, também chamado de resíduo (ou erro absoluto), que é calculado pela diferença entre o valor previsto e o valor medido (y_{exp} , valor experimental) do conjunto de previsão. O erro padrão (resíduo) é calculado para cada amostra (TALHAVINI, 2003) utilizando a expressão mostrada na equação a seguir:

$$EP = y_{pre} - y_{exp} \quad (46)$$

Ou ainda, o Erro Relativo, dado pela equação (47) (SKOOG et al., 2004):

$$EP = \frac{y_{pre} - y_{exp}}{y_{exp}} \quad (47)$$

6 CONSIDERAÇÕES FINAIS

Foi encontrada uma correlação negativa de -0,38 em relação ao teor de proteína e o teor de óleo nas cultivares estudadas, comportamento reportado em outros trabalhos com outras cultivares (BOYDAK et al., 2002; PENHA, 2003; MORAES et al., 2006; ZARKADAS et al., 2007).

Os modelos visando a determinação dos teores de proteína, óleo e umidade em grãos inteiros de soja mostraram bons resultados. Utilizando apenas 5 VLs, as médias dos erros para o conjunto de amostras previstas foram menores que 2,0 % (1,31 % para a determinação de proteína, 1,65 % visando a determinação de óleo e, 1,24 % para a determinação de umidade); sendo que o maior erro relativo encontrado foi de -4,54 % em relação ao teor de proteína previsto pelo NIR e o medido pelo “Kjeldahl”, de 3,73 % em relação ao teor de óleo previsto pelo NIR e o medido pelo método de extração por Soxhlet e, de 4,09 % em relação ao teor de umidade previsto pelo NIR e o medido pela “estufa”.

A utilização da calibração por PLS, método de regressão largamente utilizado em determinações por espectroscopia NIR, proporcionou bons resultados, ratificando sua potencialidade na construção de modelos de previsão visando quantificação de teores importantes na avaliação da qualidade de soja.

Os resultados mostram ainda a importância na escolha do tratamento dos dados, bem como no tipo de pré-processamento empregado, na escolha das regiões espectrais e na escolha das amostras do conjunto de calibração.

A técnica espectroscópica NIR com reflexão difusa, estudada neste trabalho para prever teores de proteína, óleo, e umidade em grãos inteiros de sojas cultivadas em Roraima, apresenta como vantagens sobre os métodos de referência a não destruição da amostra, sem a necessidade de pré-tratamento, o menor tempo de análise, a não utilização de reagentes, e a não geração de resíduos, assim, não agredindo o ambiente, sendo uma técnica limpa e de acordo com alguns princípios básicos da Química Verde.

As vantagens citadas e/ou demonstradas servem como estímulo à implantação da técnica espectroscópica NIR, aliada à calibração multivariada, em laboratórios de análises rotineiras que utilizam o método kjeldahl, e/ou os métodos de extração por Soxhlet e da estufa, ou algum outro método que demande tempo ou gasto de reagentes caros, ou ainda passível de geração de resíduos, já que o momento atual é caracterizado por uma postura global voltada para a preservação ambiental, e centros geradores de resíduos não podem adotar uma postura de incoerência quanto à questão dos resíduos gerados, sendo o gerenciamento destes um dever

para com a sociedade. Mesmo que haja gerenciamento de resíduos, tem-se a opção de escolha da melhor tecnologia observando a razão tempo, custo e risco/benefício, haja vista a vasta aplicação da técnica em análises de rotina nos mais diversos tipos de matrizes.

7 PERSPECTIVAS FUTURAS

Este estudo abre a oportunidade de uma futura implantação desta técnica para determinações de alguns parâmetros de monitoramento da qualidade da soja produzida no Estado, tendo em vista que praticamente toda a produção é voltada para exportação.

8 REFERÊNCIAS

- AHLUWALIA, W.K.; KIDWAI, M. **New Trends in Green Chemistry**. Kluwer Academic. 1ª Ed. 2004. (ISBN 140201872X).
- ARAÚJO, M. C. U. **Curso “Quimiometria II”**, Universidade Federal da Paraíba. Laboratório de Instrumentação e Automação em Química Analítica/Quimiometria-LAQA. João Pessoa, 2005. 1 CD-ROM
- AZZOUZ, T.; PUIGDOMÉNECH, A.; ARAGAY, M.; TAULER, R. Comparison between different data pre-treatment methods in the analysis of forage samples using near-infrared diffuse reflectance spectroscopy and partial least-squares multivariate calibration method. **Analytica Chimica Acta**, 484, 121–134, 2003.
- BARTHUS, R.C. **Aplicação de Métodos Quimiométricos para Análises de Controle de Qualidade de Óleos Vegetais utilizando Espectroscopias no Infravermelho e Raman**. 1999. Dissertação (Mestrado), Universidade Estadual de Campinas - UNICAMP. Campinas, São Paulo.
- BLANCO, M.; COELLO, J.; EUSTAQUIO, A.; ITURRIAGA, H.; MASPOCH, S. Analytical control of pharmaceutical production steps by near infrared reflectance spectroscopy. **Analytica Chimica Acta**, 392, 237-246, 1999.
- BLANCO, M.; PAGÈS, J. Classification and quantitation of finishing oils by near infrared spectroscopy. **Analytica Chimica Acta**, 463, 295–303, 2002.
- BLANCO, M.; VILLARROYA, I. NIR spectroscopy: a rapid-response analytical tool. **Trends in Analytical Chemistry**, Vol. 21, No. 4, 240-250, 2002.
- BORIN, A. **Aplicação de quimiometria e espectroscopia no infravermelho no controle de qualidade de lubrificantes**. 2003. Dissertação (Mestrado), Universidade Estadual de Campinas - UNICAMP. Campinas, São Paulo.
- BOYDAK, E.; ALPASLAN, M.; HAYTA, M.; GERÇEK, S.; SIMSEK, M. Seed Composition of Soybeans Grown in the Harran Region of Turkey As Affected by Row Spacing and Irrigation. **Journal of Agricultural and Food Chemistry**, 50, 4718-4720, 2002.
- BRAGA, Jez W. B. **Avaliação de Figuras de Mérito em Calibração Multivariada, Aplicada na Determinação de Carbamazepina por Espectroscopia no Infravermelho Próximo e Médio**. 2004. Dissertação (Mestrado), Universidade Estadual de Campinas - UNICAMP. Campinas, São Paulo.
- BRUNS, R. E.; FAIGLE, J. F. G. Quimiometria. **Química Nova**, 84-99, 1985.
- BURNS, D.A.; CIURCZAK, E.W. **Handbook of Near-Infrared Analysis**. Second Edition, Revised and Expanded. Practical spectroscopy series volume 27. Marcel Dekker, Inc. New York. (ISBN 0-8247-0534-3), 2001.

CANDOLFI, A.; MAESSCHALCK, R. De; JOUAN-RIMBAUD, D.; HAILEY, P.A.; MASSART. The influence of data pre-processing in the pattern recognition of excipients near-infrared spectra. **Journal of Pharmaceutical and Biomedical Analysis**, 21, 115–132, 1999.

CECCHI, H.M. **Fundamentos Teóricos e Práticos em Análise de Alimentos**. 2ª Ed. Revista. Editora da Unicamp. Campinas – SP. 2007. ISBN: 85-268-0477-4

CERQUEIRA, E. O.; POPPI, R. J.; KUBOTA, L. T. Utilização de filtro de transformada de Fourier para a minimização de ruídos em sinais analíticos. **Química Nova**, 23(5), 2000.

CHAN, C.-O.; CHU, C.-C.; MOK, D.; CHAU, F.-T. Analysis of berberine and total alkaloid content in *Cortex Phellodendri* by near infrared spectroscopy (NIRS) compared with high-performance liquid chromatography coupled with ultra-visible spectrometric detection. **Analytica Chimica Acta**, 592, 121–131, 2007.

CHEN, D.; CAI, W.; SHAO, X. Removing uncertain variables based on ensemble partial least squares. **Analytica Chimica Acta** xxx (2007) xxx–xxx (article in press).

CHEN, Q.; ZHAO, J.; ZHANG, H.; WANG, X. Feasibility study on qualitative and quantitative analysis in tea by near infrared spectroscopy with multivariate calibration. **Analytica Chimica Acta**, 572, 77–84, 2006.

CHEN, Z.-P.; MORRIS, J.; MARTIN, E. Extracting Chemical Information from Spectral Data with Multiplicative Light Scattering Effects by Optical Path-Length Estimation and Correction. **Analytical Chemistry**, 78, 7674–7681, 2006.

CHRISTIAN, Cary D. **Analytical Chemistry**. Sixth Edition. John Wiley & Sons, Inc. (ISBN 0-471-45162-2), 2004.

COSTA FILHO, P. A. **Algoritmo Genético na Seleção de Variáveis em Calibração Multivariada de Dados Espectroscópicos**. 1998. Dissertação (Mestrado), Universidade Estadual de Campinas - UNICAMP. Campinas, São Paulo.

COSTA FILHO, P. A. **Estudo comparativo entre técnicas de inteligência artificial e modelos lineares em determinações quantitativas no infravermelho próximo**. 2003. Tese (Doutorado), Universidade Estadual de Campinas - UNICAMP. Campinas, São Paulo.

COSTA FILHO, P. A.; POPPI, R. J. Aplicação de algoritmos genéticos na seleção de variáveis em espectroscopia no infravermelho médio. Determinação simultânea de glicose, maltose e frutose. **Química Nova**, Vol. 25, No. 1, 46-52, 2002.

COZZOLINO, D.; DELUCCHI, I.; KHOLI, M.; VÁZQUEZ, D. Use of Near Infrared Reflectance Spectroscopy to Evaluate Quality Characteristics in Whole-Wheat Grain. *Agricultura Técnica (Chile)*, 66(4):370-375 oct.-dic. 2006.

COZZOLINO, D.; LIU, L.; CYNKAR, W.U.; DAMBERGS, R.G.; JANIK, L.; COLBY, C.B.;

GISHEN, M. Effect of temperature variation on the visible and near infrared spectra of wine and the consequences on the partial least square calibrations developed to measure chemical composition. **Analytica Chimica Acta**, 588, 224–230, 2007.

DU, Y. P.; KASEMSUMRAN, S.; MARUO, K.; NAKAGAWA, T.; OZAKI, Y. Improvement of the Partial Least Squares Model Performance for Oral Glucose Intake Experiments by Inside Mean Centering and Inside Multiplicative Signal Correction. **Analytical Sciences**, August 2005, Vol. 21. **The Japan Society for Analytical Chemistry**.

EMBRAPA - **Empresa Brasileira de Pesquisa Agropecuária em Roraima** (Sítio eletrônico: <www.cpafr.embrapa.br>). Notícia disponível em: <www.embrapa.gov.br/noticias/banco_de_noticias/2004/abril/bn.2004-1125.8623360415/mostra_noticia>. Acesso em 19 ago. 2006.

EMBRAPA RORAIMA. **Cultivo de Soja no Cerrado de Roraima**. Sistemas de Produção 01. Boa Vista. Setembro, 2005. (ISSN 1809-2675).

ESCANDAR, G. M.; FABER, N. K. M.; GOICOECHEA, H. C.; PEÑA, A. M.; OLIVIERI, A. C.; POPPI, R. J. Second- and third-order multivariate calibration: data, algorithms and applications. **Trends in Analytical Chemistry**, Vol. 26, No. 7, 2007.

FERRÃO, M.F. **Aplicação de técnicas espectroscópicas de reflexão no infravermelho no controle de qualidade de farinha de trigo**. 2000. Tese, Universidade Estadual de Campinas - UNICAMP. Campinas, São Paulo.

FERREIRA, M. M. C.; ANTUNES, A. M.; MELGO, M. S.; VOLPE, P. L. O. Quimiometria I: calibração multivariada, um tutorial. **Química Nova**, vol. 22, n. 5, Set./Out. 1999.

FIDÊNCIO, P.H. **Análise de solos por espectroscopia no infravermelho próximo e aplicação de métodos quimiométricos**. 2001. Tese (Doutorado), Universidade Estadual de Campinas - UNICAMP. Campinas, São Paulo.

FRANK, I. E.; KALIVAS, J. H.; KOWALSKI, B. R. Partial Least Squares Solutions for Multicomponent Analysis. **Analytical Chemistry**, Vol. 55, No. 11, 1800-1804, 1983.

FREDERICO, F.R. **Desenvolvimento de um Sensor Químico de Fibras Ópticas para a Determinação Simultânea de Metais Pesados Empregando Métodos de Calibração Multivariada**. 2003. Tese (Doutorado), Universidade Estadual de Campinas - UNICAMP. Campinas, São Paulo.

GARCÍA, M.C.; MARINA, M.L.; LABORDA, F.; TORRE, M. Chemical characterization of commercial soybean products. **Food Chemistry**, Vol. 62, No. 3, pp. 325-331, 1998.

GIANLUPPI, V.; GIANLUPPI, D.; SMIDERLE, O. J.; NASCIMENTO Jr., A.; ALMEIDA, L.A.; ARANTES, N.E.; BRS MG (Nova Fronteira): Cultivar de soja para Roraima. **Embrapa Roraima**; Comunicado técnico. Nº. 007, Pág. 1-3, dez. 2000a. (ISSN 0101-8620).

GIANLUPPI, V.; GIANLUPPI, D.; SMIDERLE, O. J.; NASCIMENTO Jr, A.; ALMEIDA, L. A. MA/BR 65 (Sambaíba): Cultivar de soja de ciclo precoce recomendada para Roraima. **Embrapa Roraima**; Comunicado técnico. Nº. 008, Pág.1-3, dez. 2000b. (ISSN 0101-8620).

GIANLUPPI, V.; GIANLUPPI, D.; SMIDERLE, O. J.; NASCIMENTO Jr., A.; ALMEIDA, L.A. BRS MA (Tracajá): Cultivar de Soja para Roraima. **Embrapa Roraima**; Comunicado técnico. Nº. 009, Pág. 1-3, dez. 2000c. (ISSN 0101-8620).

GIANLUPPI, V.; GIANLUPPI, D.; SMIDERLE, O. J.; ALMEIDA, L.A. BRS Celeste: Nova cultivar de soja para o cerrado roraimense. **Embrapa Roraima**; Comunicado técnico 18. Dezembro, 2002. (ISSN 0102-099).

GIANLUPPI, V.; GIANLUPPI, D.; SMIDERLE, O.J. Orientações técnicas para instalação do cultivo de soja nos cerrados de Roraima. **Embrapa Roraima**; Técnica Circular 02. Dezembro, 2003. (ISSN 0101-9813).

GIANLUPPI, V.; SMIDERLE, O. J.; GIANLUPPI, D.; ALMEIDA, L.A. BRS Luziânia: nova cultivar de soja para o cerrado de Roraima. Boa Vista – RR. **Embrapa Roraima**; Comunicado técnico 04. Agosto, 2004a. (ISSN 0102-099)

GIANLUPPI, V.; SMIDERLE, O. J.; GIANLUPPI, D.; ALMEIDA, L. A.; SOUZA, P. I. M. BRS Serena: Cultivar de soja para os cerrados roraimense. **Embrapa Roraima**; Comunicado técnico 10. Dezembro, 2004b.

GIANLUPPI, V.; SMIDERLE, O.J.; GIANLUPPI, D.; ALMEIDA, L.A.; SOUZA, P.I.M. BRS Raimunda: Cultivar de Soja para os Cerrados de Roraima. **Embrapa Roraima**; Comunicado Técnico 20. Dezembro, 2005 (ISSN 0102-099).

GIANLUPPI, V.; SMIDERLE, O. J.; GIANLUPPI, D.; ALMEIDA, L.A. BRS Candeia: Cultivar de Soja Para o Cerrado de Roraima. **Embrapa Roraima**; Comunicado técnico 09. Boa Vista – RR. Outubro, 2006. (ISSN 1980-4032)

GOLIC, M.; WALSH, K.B. Robustness of calibration models based on near infrared spectroscopy for the in-line grading of stonefruit for total soluble solids content. **Analytica Chimica Acta**, 555, 286–291, 2006.

GREENPEACE. **Soja – um tesouro nacional chinês ameaçado pela engenharia genética**. Campanha de Engenharia Genética. 2004. Disponível em: <http://www.greenpeace.org.br/transgenicos/pdf/7_chinasoja_br.pdf>. Acesso em: 11 set. 2007.

GUCHARDI, R. **Espectroscopia no Infravermelho Próximo com o Uso de Filtro Óptico-Acústico Sintonizável: Instrumentação e Aplicações**. 2000. Tese (Doutorado), Universidade Estadual de Campinas - UNICAMP. Campinas, São Paulo.

GUO, Q.; WU, W.; MASSART, D.L. The robust normal variate transform for pattern recognition with near-infrared data. **Analytica Chimica Acta**, 382, 87-103, 1999.

HARBECK, C.; FAURIE, R.; SCHEPER, T. Application of near-infrared spectroscopy in the sugar industry for the detection of betaine. **Analytica Chimica Acta**, 501, 249–253, 2004.

HASWELL, S. J.; WALMSLEY, A. D. Chemometrics: the issues of measurement and modeling. **Analytica Chimica Acta**, 400, 399–412, 1999.

HEIKKA, R.A.; IMMONEN, K.T.; MINKKINEN, P.O.; PAATERO, E.Y.O.; SALMI, T.O. Determination of acid value, hydroxyl value and water content in reactions between dicarboxylic acids and diols using near-infrared spectroscopy and non-linear partial least squares regression. **Analytica Chimica Acta**, 349, 287-294, 1997.

HERRERA, J.; GUESALAGA, A.; AGOSIN, E. Shortwave–near infrared spectroscopy for non-destructive determination of maturity of wine grapes. **Measurement Science and Technology**, 14, 689–697, 2003.

IÑÓN, F. A.; GARRIGUES, S.; GUARDIA, M. de la. Combination of mid- and near-infrared spectroscopy for the determination of the quality properties of beers. **Analytica Chimica Acta**, 571, 167–174, 2006.

KONDEPATI, V. R.; KEESE, M.; MUELLER, R.; BACKHAUS, J. Near-infrared spectroscopic detection of human colon diverticulitis: A pilot study. **Vibrational Spectroscopy**, 44, 56–61, 2007a.

KONDEPATI, V. R.; KEESE, M.; MUELLER, R.; MANEGOLD, B. C.; BACKHAUS, J. Application of near-infrared spectroscopy for the diagnosis of colorectal cancer in resected human tissue specimens. **Vibrational Spectroscopy**, 44, 236–242, 2007b.

KRÄMER, K.; EBEL, S. Application of NIR reflectance spectroscopy for the identification of pharmaceutical excipients. **Analytica Chimica Acta**, 420, 155–161, 2000.

LAGO, C. L.; JULIANO, V. F.; KASCHERES, C. Applying moving median digital filter to mass spectrometry and potentiometric titration. **Analytica Chimica Acta**, 310 281-288, 1995.

LANCASTER, M. **Green Chemistry: An Introductory Text**. Royal Society of Chemistry. 2002. (ISBN 0854046208).

LENARDÃO, E.J.; FREITAG, R.A.; DABDOUB, M.J.; BATISTA, A.C.F.; SILVEIRA, C.C. “*Green Chemistry*” – Os 12 Princípios da Química Verde e sua Inserção nas Atividades de Ensino e Pesquisa. **Química Nova**, Divulgação, Vol. 26, No. 1, 123-129, 2003.

LIMA, S. L. T. **Seleção de Variáveis em Calibração Multivariada a partir da Hessiana dos erros**. 2000. Dissertação (Mestrado), Universidade Estadual de Campinas - UNICAMP. Campinas, São Paulo.

LITANI-BARZILAI, I.; SELA, I.; BULATOV, V.; ZILBERMAN, I.; SCHECHTER, I. On-line remote prediction of gasoline properties by combined optical methods. **Analytica Chimica Acta**,

339, 193-199, 1997.

LUYPAERT, J.; HEUERDING, S.; HEYDEN, Y.V.; MASSART, D.L. The effect of preprocessing methods in reducing interfering variability from near-infrared measurements of creams. **Journal of Pharmaceutical and Biomedical Analysis**, 36, 495–503, 2004.

LUYPAERT, J.; HEUERDING, S.; JONG, S. de; MASSART, D.L. An evaluation of direct orthogonal signal correction and other preprocessing methods for the classification of clinical study lots of a dermatological cream. **Journal of Pharmaceutical and Biomedical Analysis**, 30, 453-466, 2002.

MACÊDO, J. A. B. **Métodos laboratoriais de análises físico-químicas e microbiológicas**. 3ª edição. CRQ-MG. Belo Horizonte – MG, (ISBN 85-9015687-7), 2005.

MACHO, S.; LARRECHI, M. S. Near-infrared spectroscopy and multivariate calibration for the quantitative determination of certain properties in the petrochemical industry. **Trends in Analytical Chemistry**, Vol. 21, No. 12, 799-806, 2002.

MAEHLER, A.R.; COSTA, J.A.; PIRES, J.L.F.; RAMBO, L. Qualidade de grãos de duas cultivares de soja em função da disponibilidade de água no solo e arranjo de plantas. **Ciência Rural**, Santa Maria, v.33, n.2, p.213-218, mar.-abr. 2003. (ISSN 0103-8478)

MARTENS, H.; NIELSEN, J. P.; ENGELSEN, S. B. Light Scattering and Light Absorbance Separated by Extended Multiplicative Signal Correction. Application to Near-Infrared Transmission Analysis of Powder Mixtures. **Analytical Chemistry**, 75, 394-404, 2003.

MAXETRON – **Serviços, Informações & Representações** (Site eletrônico: <www.maxpressnet.com.br>). Notícia disponível em: <www.maxpressnet.com.br/noticia.asp?TIPO=PA&SQINF=234539>. Acesso em: 19 ago. 2006.

MELLO, C.A. **Redes Neurais e Técnicas de Pré-Processamento de Sinais em Espectroscopia de Reflectância no Infravermelho Próximo**. 1998. Tese (Doutorado), Universidade Estadual de Campinas - UNICAMP. Campinas, São Paulo.

MENDHAM, J. DENNEY, R. C.; BARNES, J. D.; THOMAS, M. J. K. Vogel: Análise Química Quantitativa. In: _____. **Estatística: Introdução à Quimiometria**. 6ª Ed. Rio de Janeiro. LTC. 2002. Cap. 4.

MESSERSCHMIDT, I. **Análise quantitativa por espectroscopia no infravermelho médio empregando técnicas de reflectância e calibração multivariada**. 1999. Tese (Doutorado), Universidade Estadual de Campinas - UNICAMP. Campinas, São Paulo.

MORAES, R.M.A.; JOSÉ, I.C.; RAMOS, F.G.; BARROS, E.G.; MOREIRA, M.A. Caracterização bioquímica de linhagens de soja com alto teor de proteína. **Pesq. agropec. bras.**, Brasília, v.41, n. 5, p.725-729, maio 2006.

MORGANO, M.A. **Aplicação de Métodos Quimiométricos em Análise de Alimentos**. 2005.

Tese (Doutorado em Ciências), Universidade Estadual de Campinas - UNICAMP. Campinas, São Paulo.

MORGANO, M.A.; FARIA, C.G.; FERRÃO, M.F.; BRAGAGNOLO, N.; FERREIRA, M.M.C.; Determinação de proteína em Café Cru por Espectroscopia NIR e Regressão PLS. **Ciência e Tecnologia de Alimentos**. Campinas. 25(1): 25-31, jan.-mar. 2005.

MORGANO, M.A.; QUEIROZ, S.C.N.; FERREIRA, M.M.C. Aplicação da Análise Exploratória na Diferenciação de Vegetais. **Brazilian Journal of Food Technology**. 2(1,2): 73-79, 1999.

MOROS, J.; ARMENTA, S.; GARRIGUES, S.; GUARDIA, M. Near infrared determination of Diuron in pesticide formulations. **Analytica Chimica Acta**, 543, 124–129, 2005.

MOROS, J.; ARMENTA, S.; GARRIGUES, S.; GUARDIA, M. Univariate near infrared methods for determination of pesticides in agrochemicals. **Analytica Chimica Acta**, 579, 17–24, 2006.

NÆS, T.; MARTENS, H. **Multivariate Calibration**. John Wiley & Sons Ltda, 1989.

NAGATA, N.; BUENO, M. I. M. S.; PERALTA-ZAMORA, P. G. Métodos matemáticos para correção de interferências espectrais e efeitos interelementos na análise quantitativa por fluorescência de raios-X. **Química Nova**, Vol. 24, No. 4, 531-539, 2001.

NETO, B. B.; PIMENTEL, M. F.; ARAÚJO, M. C. U. Recomendação para calibração em química analítica – Parte 1. Fundamentos e calibração com um componente (Calibração Univariada). **Química Nova**, Vol. 25, No. 5, 856-865, 2002.

NETO, W.B. **Parâmetros de qualidade de lubrificantes e óleo de oliva através de espectroscopia vibracional, calibração multivariada, e seleção de variáveis**. 2005. Tese, Universidade Estadual de Campinas - UNICAMP. Campinas, São Paulo.

O ESTADÃO (Sítio eletrônico: <www.estadao.com.br>). Notícia disponível em: <www.estadao.com.br/ext/agronegocios/especiais/safra0304/noticias/ultima/01.htm>. Acesso em: 19 ago. 2006a.

O ESTADÃO, Jornal o Estado de S. Paulo (Sítio eletrônico: <www.estadao.com.br>). Notícia disponível em: <www.estadao.com.br/ext/agronegocios/especiais/safra0304/noticias/ultima/03.htm>. Acesso em: 19 ago. 2006b.

O ESTADÃO (Sítio eletrônico: <<http://www.estadao.com.br>>). Notícia disponível em: <www.estadao.com.br/ext/agronegocios/especiais/safra0304/noticias/ultima/04.htm>. Acesso em: 19 ago. 2006c.

OLIVEIRA, F. C. C. A escolha da faixa espectral no uso combinado de métodos espectroscópicos e quimiométricos. **Química Nova**. Vol. 27, No. 2, 218-225, 2004.

ORTIZ, M.C.; SARABIA, L.; GARCÍA-REY, R.; CASTRO, M. D. L. Sensitivity and specificity of PLS-class modelling for five sensory characteristics of dry-cured ham using visible and near infrared spectroscopy. *Analytica Chimica Acta*, 558, 125–131, 2006.

OTTO, M. **Chemometrics. Statistics computer application in Analytical Chemistry**. Wiley-VCH. (ISBN 3-527-29629-X), 1999.

PARREIRA, T.F. **Utilização de Métodos Quimiométricos em Dados de Natureza Multivariada**. 2003. Dissertação (Mestrado), Universidade Estadual de Campinas - UNICAMP. Campinas, São Paulo.

PASQUINI, C. Near infrared spectroscopy: fundamentals, practical aspects e analytical applications. *J. braz. Chem. Soc.*, Vol. 14, No. 2, 198-219, 2003.

PASTI, L.; JOUAN-RIMBAUD, D.; MASSART, D. L.; NOORD, O. E. Application of Fourier transform to multivariate calibration of near-infrared data. *Analytica Chimica Acta*, 364, 253-263, 1998.

PEDRO, A. M. K. **Determinação Simultânea e Não-Destrutiva de Sólidos Totais e Solúveis, Licopeno e Beta-Caroteno em Produtos de Tomate por Espectroscopia no Infravermelho Próximo Utilizando Calibração Multivariada**. 2004. Dissertação (Mestrado), Universidade Estadual de Campinas - UNICAMP. Campinas, São Paulo.

PENHA, L.A.O. **Análise comparativa da composição de soja orgânica e não orgânica**. Dissertação (Curso de Pós-graduação em Agronomia), Universidade Estadual de Londrina, Londrina. 2003.

PIMENTEL, M. F.; NETO, B. B. Calibração: Uma revisão para químicos analíticos. *Química Nova*, 19(3), 1996.

PIZARRO, C.; ESTEBAN-DÍEZ, I.; NISTAL, A.-J.; GONZÁLEZ-SÁIZ, J.-M. Influence of data pre-processing on the quantitative determination of the ash content and lipids in roasted coffee by near infrared spectroscopy. *Analytica Chimica Acta*, 509, 217–227, 2004.

PORTAL BRASIL, a sua biblioteca virtual (Sítio eletrônico: <www.portalbrasil.net>). Notícia disponível em: <www.portalbrasil.net/estados_rr.htm>. Acesso em 19 ago. 2006.

REIS, M.M. **Aplicação de Métodos Quimiométricos em Separação de Espectros e Reconhecimento de Padrões**. 1997. Dissertação (Mestrado), Universidade Estadual de Campinas - UNICAMP. Campinas, São Paulo.

REIS, M.M. **Desenvolvimento e Aplicação de Métodos Quimiométricos de Ordem Superior**. 2002. Tese (Doutorado), Universidade Estadual de Campinas - UNICAMP. Campinas, São Paulo.

RODIONOVA, O. Ye.; HOUMØLLER, L.P.; POMERANTSEV, A.L.; GELADI, P.; BURGER, J.; DOROFEYEV, V.L.; ARZAMASTSEV, A.P. NIR spectrometry for counterfeit drug detection:

A feasibility study. **Analytica Chimica Acta**, 549, 151–158, 2005.

ROGGO, Y.; CHALUS, P.; MAURER, L.; LEMA-MARTINEZ, C.; EDMOND, A.; JENT, N. A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies. **Journal of Pharmaceutical and Biomedical Analysis**, 44, 683–700, 2007.

SÁIZ-ABAJO, M.J.; MEVIK, B.-H.; SEGTNAN, V.H.; NÆS, T. Ensemble methods and data augmentation by noise addition applied to the analysis of spectroscopic data. **Analytica Chimica Acta**, 533, 147–159, 2005.

SCAFI, S.H.F. **Espectroscopia no Infravermelho próximo para identificação de medicamentos falsificados**. 2000. Dissertação, Universidade Estadual de Campinas - UNICAMP. Campinas, São Paulo.

SENA, M. M.; POPPI, R. J.; FRIGHETTO, R. T. S.; VALARINI, P. J. Avaliação do uso de métodos quimiométricos em análise de solos. **Química Nova**, 23(4), 547-556, 2000.

SENA, M.M.; TREVISAN, M.G.; POPPI, R.J. PARAFAC: Uma Ferramenta Quimiométrica para Tratamento de Dados Multidimensionais. Aplicações na Determinação Direta de Fármacos em Plasma Humano por Espectrofluorimetria. **Química Nova**, Vol. 28, No. 5, 910-920, 2005.

SILVA, H.E.B. **Desenvolvimento de Um Espectrofotômetro no Infravermelho Próximo com Transformada de Hadamard**. 1999. Tese (Doutorado), Universidade Estadual de Campinas - UNICAMP. Campinas, São Paulo.

SILVA, F.M.; LACERDA, P.S.B.; JUNIOR, J.J. Desenvolvimento Sustentável e Química Verde. **Química Nova**, Divulgação, Vol. 28, No. 1, 103-110, 2005.

SIMAS, R. C. **Determinação de proteína bruta e aminoácidos em farelo de soja por espectroscopia no infravermelho próximo**. 2005. Dissertação (Mestrado), Universidade Estadual de Campinas - UNICAMP. Campinas, São Paulo.

SKOOG, D. A.; HOLLER, F. J.; NIEMAN, T. A. **Princípios de Análise Instrumental**; Capítulos 16-17. 5ª Edição. Editora Bookman. Porto Alegre, 2002.

SKOOG, D.A.; WEST, D.M.; HOLLER, F.J.; CROUCH, S.R.. **Fundamentals of Analytical Chemistry**. Eighth Edition. Thomson Brooks/cole. (ISBN 0-03-035523-0), 2004.

SMIDERLE, O.J.; GIANLUPPI, V.; MANDARINO, J.M.D. Teores de óleo, Proteínas e Produtividade de Sementes de Soja Produzidas nos Cerrados Roraimense. **Embrapa Roraima**; Comunicado Técnico. Boa Vista. Dezembro, 2006. (ISSN 1980-4032).

SMIDERLE, O.J.; VIANA, M.P.A.; SOUSA, R.C.P.; GIANLUPPI, D. Efeito de doses de zinco e cobre em características físicas e químicas de grãos de soja armazenados em Roraima. **Embrapa Roraima**; Documentos N° 02, dez. 2003. (ISSN 0101 – 9805).

SOUSA, E.R. **Determinação de parâmetros de qualidade de gasolina e estudo de contaminantes utilizando espectros do infravermelho próximo na fase vapor**. 2005. Tese (Doutorado em Ciências), Universidade Estadual de Campinas - UNICAMP. Campinas, São Paulo.

SOUSA, R.C.P.; SMIDERLE, O.J.; SILVA, S.L. Validação do método de reflectância de infravermelho na determinação do teor de água em grãos de soja. In: CONGRESO VIRTUAL IBEROAMERICANO SOBRE GESTIÓN DE CALIDAD EN LABORATORIOS, 3, 2005, Madrid. **Livro de resumos**. Madrid : iberolab, 2005. p. 141-144.

TALHAVINI, I.C. N. **Estudo de Métodos Multivariados para Análise e Calibração de Espectros**. 2003. Tese (Doutorado em Ciências), Universidade Estadual de Campinas - UNICAMP. Campinas, São Paulo.

VALDERRAMA, P. **Avaliação de figuras de mérito em calibração multivariada na determinação de parâmetros de controle de qualidade em indústria alcooleira por espectroscopia no infravermelho próximo**. 2005. Dissertação (Mestrado), Universidade Estadual de Campinas - UNICAMP. Campinas, São Paulo.

VALENTINI, S.R.T.; CASTRO, M.F.P.M.; ALMEIDA, F.H. Determinação do Teor de Umidade de Milho Utilizando Aparelho de Microondas. **Ciência e Tecnologia de Alimentos**. Vol. 18 N°. 2 Campinas. May/July. 1998. (ISSN 0101-2061).

WILLIAMS, P.; NORRIS, K. **Near-Infrared Tecnology; in the Agricultural and Food Industries**. Second edition. AAVV, Inc. St. Paul, Minnesota, USA. (ISBN 1-891127-24-1), 2001.

YAMADA, L.T.P.; BARCELOS, M.F.P.; SOUSA, R.V.; LIMA, A.L. Composição Química e Conteúdo de Ferro Solúvel em Soja [*Glycine max* (L.) Merrill] **Ciênc. agrotec.**, Lavras. V.27, n.2, p.406-413, mar./abr. 2003.

YUAN, Y.J.; VELEV, O.D.; CHEN, K.; CAMPBELL, B.E.; KALER, E.W.; LENHOFF, A.M. Effect of pH and Ca^{2+} - Induced Associations of Soybean Proteins. **Journal of Agricultural and Food Chemistry**, 50, 4953-4958, 2002.

ZARKADAS, C.G.; GAGNON, C.; GLEDDIE, S.; KHANIZADEH, S.; COBER, E.R.; GUILLEMETTE, R.J.D. Assessment of the protein quality of fourteen soybean [*Glycine max* (L.) Merr.] cultivars using amino acid analysis and two-dimensional electrophoresis. **Food Research International**, 40, 129–146, 2007.

ZEAITER, M.; ROGER, J.-M.; BELLON-MAUREL, V. Robustness of models developed by multivariate calibration. Part II: The influence of pre-processing methods. **Trends in Analytical Chemistry**, Vol. 24, No. 5, 2005.