

**Universidade Estadual de Campinas
Instituto de Química
Laboratório de Quimiometria em Química Analítica**

***“Aplicação de Métodos Quimiométricos para
Análises de Controle de Qualidade de Óleos Vegetais
utilizando Espectroscopias no Infravermelho e
Raman ”***

Dissertação de Mestrado

Rosângela Cristina Barthus.

Orientador : Prof. Dr. Ronei Jesus Poppi.

**Campinas
Setembro/ 1999**

1

UNIDADE	I Q
N.º CHAMADA:	UNICAMP
	B283a
V. Ex.	
TOMBO BC/	40932
PROC.	278/00
C <input type="checkbox"/>	D <input checked="" type="checkbox"/>
PREÇO	\$11,00
DATA	13/04/00
N.º CPD	

CM-00138989-9

**FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DO INSTITUTO DE QUÍMICA
UNICAMP**

B283a

Barthus, Rosângela Cristina

Aplicação de métodos quimiométricos para análises de controle de qualidade de óleos vegetais utilizando espectroscopias no infravermelho e Raman / Rosângela Cristina Barthus. -- Campinas, SP: [s.n.], 1999.

Orientador: Ronei Jesus Poppi.

Dissertação (mestrado) – Universidade Estadual de Campinas. Instituto de Química.

1. PLS(mínimos quadrados parciais). 2. Cartas multivariadas. 3. PCA(análise de componentes principais).
I. Poppi, Ronei Jesus. II. Universidade Estadual de Campinas. Instituto de Química. III. Título.

" Quando se quer uma coisa, todo o Universo conspira para que a pessoa consiga realizar o seu sonho. E sempre antes de realizar um sonho, a Alma do Mundo resolve testar tudo aquilo que foi aprendido durante a caminhada. Ela não faz isso porque seja má, mas para que possamos, junto com o sonho, conquistar também as lições que aprendemos seguindo em direção a ele."

de "O Alquimista"

*Aos meus pais Valter e Dirma
e à minha avó Elza
pelo imenso apoio em todos os momentos da minha vida.....*

*Ao Roberto, Jane, Lucas,
Robson e Elaine.*

Agradecimentos

- *Ao Prof .Dr. Ronei Jesus Poppi pela paciência, amizade e orientação.*
- *Aos membros do grupo LAQQA : Marcelo, Jacqueline, Paulo Augusto, Karla e, em especial, à Cleidiane, Sívio, Eduardo e Paulo Fidêncio que em muito contribuíram para a realização deste trabalho.*
- *Ao Técnico Renato Grimaldi da Faculdade de Engenharia de Alimentos da UNICAMP (Laboratório de Óleos e Gorduras) pelas amostras de óleos vegetais e informações concedidas.*
- *À Técnica Márcia do Instituto de Química da UNICAMP pela obtenção dos espectros Raman dos óleos vegetais.*
- *Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq pela bolsa concedida.*
- *A todos aqueles que direta ou indiretamente contribuíram para a realização deste trabalho.*
- *Ao Fábio, pelo apoio, incentivo, amizade, pelos e-mails que tornaram mais alegres minhas tardes de 1998 Saudades e Até mais*

CURRICULUM VITAE

ROSÂNGELA CRISTINA BARTHUS

DADOS PESSOAIS

Endereço : Rua Paulínia n^o 48
Bairro Jardim dos Oliveiras CEP: 13043-420.
Campinas –SP fone: (0XX19) 230-0137

RG-19.314.423

Nasc-21/01/70

Est.Civil : Solteira.

Nacionalidade : Brasileira

Naturalidade : Campinas-SP

FORMAÇÃO ACADÊMICA

Mestrado em Química , área de concentração em Química Analítica,
Universidade Estadual de Campinas - UNICAMP - Campinas/SP .

Título : Aplicação de Métodos Quimiométricos para Análises de Controle de Qualidade de Óleos Vegetais utilizando Espectroscopias no Infravermelho e Raman.

Orientador : Prof. Dr. Ronei Jesus Poppi.

Início : Agosto de 1997

Término : 20 de Setembro de 1999 (data da defesa).

Bacharelado em Química Tecnológica, Universidade Estadual de Campinas-
UNICAMP – Campinas /SP

Início : 1^o Semestre de 1992.

Término: dezembro de 1996.

**Licenciatura em Química, Universidade Estadual de Campinas- UNICAMP-
Campinas –SP**

Término : Dezembro de 1999.

**Técnica em Química , Escola Técnica Estadual Conselheiro Antônio Prado-
ETECAP – Campinas –SP.**

Início: 1985

Término: 1988.

ARTIGOS CIENTÍFICOS

**Título : Avaliação do Desempenho Mecânico de Latas de Óleo Comestível
quanto ao Impacto Estacionário e Queda Livre.**

Revista :Metalurgia e Materiais

Fevereiro de 1997, vol 53 nº 462.

Co-autores- Dantas, S .T , Jaime ,S.B.M, Anjos ,V.D.A

TRABALHOS APRESENTADOS EM REUNIÕES CIENTÍFICAS.

- **22ª Reunião da Sociedade Brasileira de Química**

**Título: Determinação do Nível total de Insaturação de Óleos Vegetais por
Espectroscopia Raman e Calibração Multivariada.**

Co- Autor : Ronei J. Poppi.

Data: 25 a 28 de Maio de 1999.

- **10º Encontro Nacional de Química Analítica.**

**Título: O Emprego de Cartas de Qualidade Multivariadas na
Caracterização de Processos Oxidativos de Óleos Vegetais.**

Co-Autor : Ronei J. Poppi.

Data : 31 de Agosto a 3 de Setembro de 1999.

EXPERIÊNCIAS PROFISSIONAIS

Centro de Tecnologia de Embalagens de Alimentos CETEA - ITAL.

Campinas –SP

Função: Estagiária de nível Superior.

Período: 09 /96 –12/96

BRASWEY -S/A - Indústria de Alimentos.

Campinas –SP

Função: Analista Química.

Período: 1989/1990

RHODIA –S/A - Indústria Química.

Paulínia-SP

Função: Estagiária de Nível Médio.

Período : 04/88-12/88.

CURSOS

- **História da Química .**

Encontro Nacional dos Estudantes de Química.

Data : 4 a 11 de fevereiro de 1993.

LÍNGUA ESTRANGEIRA

- **Inglês**

Nível intermediário.

INFORMÁTICA

- Windows 98 , Origin 4.1, Uncrabler 6.0, Matlab 4.0 .

Rosângela Cristina Barthus.

Índice

Resumo.....i

Abstract.....ii

Capítulo 1- Uma Breve Introdução às Análises de Controle de Qualidade de Óleos Vegetais.

1-Introdução 1

Capítulo 2- Fundamentos e Aspectos Quantitativos das Espectroscopias no Infravermelho e Raman.

2 - Os princípios básicos das espectroscopias no infravermelho e Raman..... 7

2.1- As espectroscopias com transformada de Fourier9

2.2- Fatores experimentais que afetam as análises espectroscópicas..... 10

2.2.1 - Espectroscopia no infravermelho 10

2.2.2 - Espectroscopia Raman 11

Capítulo 3 – Descrição dos Métodos Quimiométricos.

3- Organização dos dados para análise multivariada..... 14

3.1- Análise de componentes principais..... 15

3.2- A carta de controle de qualidade univariada, segundo Shewart.....21

3.3- A construção das cartas univariadas.....21

3.4- As indicações da falta de controle.....22

3.5- As cartas de controle de qualidade multivariadas 24

3.6- O problema do T^2 24

3.7- Cartas de controle de qualidade multivariadas usando componentes principais.....27

3.8- A carta de controle de qualidade multivariada T^228

3.9- A carta de controle de qualidade multivariada Q.....29

3.10- O que podem ser determinados através das cartas de controle : Um exemplo.....	31
3.11- Introdução às análises de regressão.....	33
3.12- Pré-tratamento dos dados	36
3.13- O método dos mínimos quadrados parciais.....	37
3.14- A previsão.....	40
3.15- O número de variáveis latentes.....	41
3.16- A identificação das amostras anômalas.....	42
3.17- "Leverages"	43
3.18- Análise residual.....	44
3.19- O gráfico de influência.....	45

Capítulo 4 – Determinação do Nível Total de Insaturação dos Óleos Vegetais por Espectroscopia Raman e Calibração Multivariada.

4 - Parte experimental.....	46
4.1- Determinação do índice de iodo dos óleos vegetais.....	46
4.2- Método padrão para a determinação do índice de iodo.....	48
4.3- Os espectros Raman.....	49
4.4- Resultados e discussões da determinação do índice de iodo dos óleos vegetais	51
4.5- O método padrão.....	52
4.6- A organização de dados para a análise multivariada.....	55
4.7- Resultados e discussão da análise multivariada	55
4.8- Método dos mínimos quadrados parciais.....	56
4.9- A normalização.....	56
4.10- Centrar na média	57
4.11- Outros tipos de pré-tratamento.....	57
4.12- A construção do modelo de calibração (PLS).....	57
4.13- O número ideal de variáveis latentes utilizadas para a construção do modelo.....	58
4.14- A identificação de amostras que não pertencem ao modelo	60
4.15- Construção de modelos baseados na seleção das variáveis.....	63

Capítulo 5 - O Emprego de Cartas de Qualidade Multivariadas na Caracterização do Processo Oxidativo dos Óleos Vegetais.

5 - Parte experimental.....	66
5.1 - Caracterização do processo oxidativo dos óleos vegetais.....	66
5.2 - Resultados e discussões	69
5.2.1- Cartas de controle de qualidade multivariadas	69
5.2.2- O estudo com quatro componentes principais.....	77
5.3- As cartas de controle univariadas	81
Conclusões.....	85
Referências Bibliográficas.....	86

Resumo

Autora : Rosângela Cristina Barthus

Orientador : Ronei Jesus Poppi

O controle de qualidade dos óleos vegetais é extremamente importante para diversos setores industriais, como o de produtos alimentícios, cosméticos, tintas, vernizes e lubrificantes, entre outros. Para assegurar a qualidade destes óleos existem inúmeros métodos de controle.

Neste trabalho, duas novas metodologias para o controle de qualidade de óleos vegetais foram propostas. O desenvolvimento destes métodos visa tornar as análises de controle de qualidade mais práticas e rápidas, além de não consumirem reagentes com alto grau de toxicidade.

Na primeira metodologia foi proposta a determinação do nível total de insaturação de diversos tipos de óleos vegetais empregando calibração multivariada, mais especificamente o método dos mínimos quadrados parciais (PLS) a dados obtidos através de espectroscopia Raman. Foram obtidos bons resultados, comparáveis com o do método oficial da *American Oil Chemist' Society* (AOCS).

Em uma segunda metodologia procurou-se caracterizar o processo oxidativo do óleo de soja de diferentes procedências utilizando cartas de controle de qualidade multivariadas (T^2 e Q). Os dados utilizados para a construção destas cartas foram obtidos por espectroscopia no infravermelho médio através de medidas por reflectância total atenuada (ATR). Através destas cartas foi possível identificar óleos deteriorados e um óleo que não era de soja.

Abstract

Authoress : Rosângela Cristina Barthus

Adviser : Ronei Jesus Poppi

The quality control of vegetable oils is very important in several industrial sectors such as food products, cosmetics, dyes, varnishes and lubricants. Intending to assure the quality of these oils there are a lot of control methods.

In this work two new methodologies for vegetable oil quality control were proposed. The development of these methods is to let the analysis faster and more practical, moreover not use toxic chemicals.

In the first methodology it was proposed the determination of the total insaturation level of several vegetable oils types using multivariate calibration, based on the partial least squares method (PLS) to data obtained by Raman spectroscopy. Good results were obtained, comparable to the American Oil Chemist' Society (AOCS) official method.

In a second methodology it was characterized the soybean oil oxidative process of different sources using multivariate control quality charts (T^2 and Q). The data to build these charts were obtained by medium infrared spectroscopy with measurements by attenuated total reflectance (ATR). From these charts it was possible to identify deteriorated oils and one of them that was not soybean oil.

CAPÍTULO I

*Uma Breve Introdução às Análises de Controle de
Qualidade de Óleos Vegetais.*

1-Introdução

Os óleos vegetais utilizados como produtos alimentícios, cosméticos, matéria-prima para tintas, vernizes e lubrificantes, entre outras aplicações, são conhecidos desde a antiguidade. O uso destes óleos depende, muitas vezes, de que sua qualidade esteja assegurada, existindo para este fim, inúmeros métodos de controle [1].

O presente trabalho busca desenvolver novas metodologias para a análise de dois parâmetros de qualidade dos óleos vegetais : a caracterização do processo oxidativo ou deterioração e o índice de iodo [2]. Em ambos os casos, existe uma dependência direta em relação às duplas ligações do carbono, presentes no triglicerídeo, que são os centros reativos nos quais ocorre a oxidação e para o índice de iodo correspondem à medida do nível total de insaturação dos óleos .

No caso do índice de iodo, o Método de Wijs - usado como referência, é uma técnica reconhecida oficialmente pela AOCS (American Oil Chemists' Society) [3] e baseia-se no método indireto de titulação do iodo. Além dessa, a cromatografia também vem se destacando como uma importante técnica instrumental para este tipo de determinação.

Para a determinação do índice de iodo, neste estudo, utiliza-se a espectroscopia Raman como uma técnica alternativa que tem como vantagens o fato de ser rápida, não destrutiva e não requerer nenhum pré-tratamento no material a ser analisado, podendo ser aplicada, inclusive, para o monitoramento de processos industriais "on-line" [4].

O uso da espectroscopia Raman em produtos alimentícios ou agrícolas só foi possível devido às inovações desta técnica nos últimos anos. Este tipo de técnica apresentava problemas, como o fato de ocasionar emissão de fluorescência e possibilidade da amostra sofrer fotodecomposição ou fotoisomerização em virtude da forte iluminação do laser na região do visível. Um outro problema era que os

antigos equipamentos apresentavam dificuldades na obtenção do espectro devido à maneira como os mesmos eram construídos. O surgimento da espectroscopia Raman com transformada de Fourier, empregando laser na região do infravermelho próximo, resolveu muitos dos problemas apresentados, tornando possível a aplicação desta técnica em novas situações [5].

No entanto, apesar de todos estes avanços, este tipo de espectroscopia não era vista como um instrumento analítico, em contraste com a espectroscopia no infravermelho médio e próximo onde a informação quantitativa podia ser obtida diretamente do espectro de absorção. Isto porque a espectroscopia Raman baseia-se no fenômeno de espalhamento inelástico da luz do laser por moléculas, o que torna qualquer tipo de quantificação mais difícil. Os espectros Raman apresentam normalmente desvios da linha base e não são reprodutíveis [6]. Em virtude disso, os dados necessitam de pré-tratamentos para estas correções. Após corrigidos e com a aplicação de métodos de calibração multivariada, bons resultados podem ser obtidos, tornando esta técnica um poderosa ferramenta analítica.

O segundo parâmetro estudado foi o processo oxidativo dos óleos vegetais. A oxidação lipídica é um fenômeno espontâneo e inevitável que, embora relativamente bem estudado em termos gerais, ainda se apresenta como um processo complexo. A oxidação depende de inúmeros fatores como por exemplo do tipo de óleo, da presença de íons metálicos e de antioxidantes e das condições em que ocorre.

A degradação oxidativa dos ácidos graxos insaturados pode ocorrer por várias vias. Aqui discutir-se-á o processo de autoxidação que é um processo dinâmico que evolui com o passar do tempo [7]. Trata-se de um fenômeno puramente químico, onde a sequência reacional pode ser classicamente dividida em : iniciação, propagação e terminação.

A iniciação do processo oxidativo é caracterizada pela formação de radicais livres de ácidos graxos . Isso ocorre quando um hidrogênio é removido de um carbono α -

metileno por ação da luz, calor, metais ou de outros radicais livres de ácidos graxos. O radical graxo formado é altamente suscetível ao ataque do oxigênio atmosférico formando o radical livre de peróxido. Este, por sua vez, é um forte iniciador de novos radicais livres de ácidos graxos e formação dos hidroperóxidos (produtos primários da oxidação).

A formação de radicais livres de peróxidos e de novos radicais livres de ácidos graxos pode ser repetida milhares de vezes, o que caracteriza a etapa de propagação. A propagação pode ser seguida da fase de terminação se os radicais livres reagirem entre si, produzindo produtos inativos na reação em cadeia.

Os hidroperóxidos formados não têm importância direta na alteração do odor ou do sabor. Contudo, eles são muito instáveis e se decompõem, com o rompimento da cadeia de hidrocarboneto, gerando uma variedade de aldeídos, alcóois e cetonas (produtos secundários da oxidação), estes sim, os principais agentes responsáveis pelo gosto e odor desagradáveis. Os aldeídos são também suscetíveis à oxidação e transformam-se em ácidos graxos (produtos terciários da oxidação). Além destes, a umidade, hidrocarbonetos, lactonas, furanas, e outros produtos menores podem também ser produzidos.

Para a identificação do processo oxidativo existem diferentes métodos. Estes podem ser físicos, químicos ou físico-químicos. Um destes métodos consiste no emprego da espectroscopia no infravermelho médio.

A região do infravermelho médio do espectro foi durante um tempo desprezada em análises de alimentos devido à problemas instrumentais e de amostragem. Muitas amostras de alimentos são opacas e espalham luz, além de conter água. A água representa um grande problema porque apresenta fortes bandas de absorção nesta região e prejudicam a conservação das celas, acessórios, etc [8] .

Avanços na instrumentação de infravermelho, particularmente com a incorporação da transformada de Fourier e novas técnicas como a Reflectância Total Atenuada (ATR)

têm levado ao surgimento do interesse na aplicação do infravermelho médio para amostras de alimentos. Esta região tem uma vantagem significativa: as bandas de absorção são bem resolvidas e podem ser determinadas para um grupo específico. Cabe ressaltar, ainda, que estes métodos são rápidos, podem ser automatizados e reduzem o uso de produtos químicos tóxicos. Os óleos puros podem ser aplicados diretamente sobre o cristal de ATR [9].

Os dados obtidos a partir da espectroscopia no infravermelho médio foram utilizados, neste trabalho, para a construção das cartas de controle de qualidade [10,11].

Conforme o caso, requisitos bastante diversos podem caracterizar a qualidade de um produto: uma propriedade física ou química, uma temperatura, uma pressão, etc. Em fases intermediárias e na fase final deve-se verificar a conformidade do processo com a especificação. A especificação indica aquilo que é desejado, como passo preliminar para se estabelecer uma norma de qualidade. Nesta norma são consideradas as possibilidades e o custo de fabricação, as condições existentes no mercado, a finalidade e o modo de utilização de um artigo, as necessidades do consumidor e a garantia que lhe possa ser dada. Por isso nenhuma norma de qualidade poderá atender todas as condições que possam existir, mas toda norma deverá ser eficiente, tecnicamente exequível e comercialmente econômica. O controle de fabricação é exercido pelo produtor durante o processo produtivo, sendo o objetivo manter a qualidade do produto satisfatoriamente uniforme, prevenindo a produção de amostras fora da especificação. A verificação de que o processo está ou não sob controle é feita pelo exame da característica de qualidade das amostras extraídas periodicamente. Se o processo estiver sob controle, estas características amostrais apresentarão aquela variabilidade correspondente a de uma população normal, isto é a variabilidade devida ao acaso na amostragem. Estas causas de variação não provocam alterações apreciáveis na qualidade do produto, sua eliminação é impossível ou antieconômica, e por isso, as causas aleatórias são consideradas como parte natural do processo de fabricação. Com isso, tem-se que a distribuição normal constitui noção estatística indispensável no controle de fabricação. Sabe-se que os itens de uma distribuição normal (de média μ e desvio-padrão σ) se distribuem em torno da média nas seguintes proporções

aproximadas : 68% dos valores no intervalo $\mu \pm \sigma$, 95% no intervalo $\mu \pm 2\sigma$ e 99,7% no intervalo $\mu \pm 3\sigma$. Quando existe a presença de amostras anômalas, elas concentram-se fora destes intervalos. Estas amostras indicam que o processo de fabricação (isto é, a população) se modificou e ficou fora de controle. As causas das modificações podem ser descobertas e por isso são chamadas causas identificáveis.

A presença de causas identificáveis é indicada pela ocorrência de diferenças significativas entre o valor observado e a média do processo, isto é de valores amostrais fora da faixa de controle; essa presença é revelada por meio de um dispositivo prático : a carta de controle de qualidade.

As cartas de controle de qualidade, tanto univariadas quanto multivariadas constituem um instrumento do diagnóstico da existência ou não do controle. Estas cartas baseiam-se na distribuição normal e distribuição normal multivariada, respectivamente.

Uma grande desvantagem das cartas de controle univariadas (Cartas Shewart, aqui utilizadas, as quais se baseiam na média e no desvio padrão) aparece quando o número de variáveis aumenta muito, pois cada variável tem de ser monitorada individualmente em cartas de controle próprias. Na prática, estas cartas demandam muito trabalho para o analista checar o processo, e a probabilidade de cometer erros é bem maior. Uma outra desvantagem é que situações fora de controle podem ser esquecidas devido às correlações entre os dados. Isto significa que as medidas podem estar em controle em todas as cartas univariadas e de fato estar numa situação fora de controle [12].

Devido ao grande número de dados fornecidos pelos computadores instalados nos processos de produção modernos, monitorar apenas algumas variáveis de qualidade é totalmente inadequado para muitos dos processos modernos. Somente métodos multivariados que tratam de todos os dados simultaneamente podem extrair informações na direção da variação do processo, que é como as variáveis estão se comportando umas em relação às outras [13].

As cartas multivariadas correspondem a uma extensão natural das cartas univariadas. Elas podem ainda ser baseadas na análise de componentes principais .

A análise de componentes principais (PCA) faz a combinação linear das variáveis originais, levando a um número relativamente pequeno de fatores que não são correlacionados. Uma vantagem deste método de controle multivariado é que somente duas cartas de controle são utilizadas : Carta de Hotelling's T^2 e a Carta Q. A carta T^2 é baseada nas componentes principais significativas enquanto que a carta Q representa o erro do modelo baseado nas componentes principais remanescentes [13].

No caso deste estudo, não sendo possível acompanhar um processo de fabricação e refinamento dos óleos vegetais em nível industrial, bem como controlar o produto acabado durante um longo período, optou-se por adquirir produtos que estavam dentro do padrão de qualidade adequado e promover suas deteriorações. Após isso, o que se faz é verificar se as cartas de controle são capazes de identificar estas modificações.

CAPÍTULO II

*Fundamentos e Aspectos Quantitativos
das Análises Espectroscópicas*

2-Os princípios básicos das espectroscopias no infravermelho e Raman

Em uma molécula, o número de vibrações, a descrição dos modos vibracionais e sua atividade em cada tipo de espectroscopia vibracional (Raman e no infravermelho) podem ser previstas a partir da simetria da molécula e da aplicação da teoria de grupo [14]. Embora ambas as espectroscopias estejam relacionadas às vibrações moleculares, os mecanismos básicos de sondagem destas vibrações são essencialmente distintos em cada uma. Em decorrência disso, os espectros obtidos apresentam diferenças significativas: quando da ocorrência de um mesmo pico nos espectros Raman e no infravermelho observa-se que o seu tamanho relativo nos espectros é muito diferente. Existe, também, o caso onde um certo pico aparece em um espectro e é totalmente ausente em outro [15]. Devido a estas diferenças, a espectroscopia no infravermelho é superior em alguns casos e em outros a espectroscopia Raman oferece espectros mais úteis. De modo geral, pode-se dizer que as espectroscopias Raman e no infravermelho são técnicas complementares.

Uma breve discussão sobre as teorias envolvidas em cada tipo de espectroscopia será agora apresentada.

A condição para que ocorra absorção da radiação infravermelha é que haja variação do momento de dipolo elétrico da molécula como consequência de seu movimento vibracional ou rotacional (o momento de dipolo é determinado pela magnitude da diferenças de carga e a distância entre dois centros de carga). Somente nestas circunstâncias, o campo elétrico alternante da radiação incidente interage com a molécula, originando os espectros. De outra forma, pode-se dizer que o espectro de absorção no infravermelho tem origem quando a radiação eletromagnética incidente tem uma componente com frequência correspondente a uma transição entre dois níveis vibracionais [14,15].

A espectroscopia de Reflexão Interna ou Reflectância Total Atenuada (ATR) é uma técnica utilizada para se obter espectros no infravermelho de amostras como :

pastas, adesivos e pó que não podem ser analisados pelos métodos normais, como pastilhas ou filmes [15].

O princípio deste tipo de espectroscopia baseia-se no fato de que quando um feixe de radiação passa de um meio mais denso (cristal de ATR) para um meio menos denso (amostra), ocorre reflexão. A fração do feixe de luz incidente que é refletida aumenta conforme aumenta o ângulo de incidência, e quando excede um determinado ângulo crítico (α) a reflexão é completa. No ponto de reflexão (de acordo com observações experimentais) o feixe atua como se penetrasse a uma pequena distância dentro da amostra . A profundidade de penetração (d_p) é dada pela equação 1: [16].

$$d_p = \lambda / \{2\pi n_1 [\text{sen}^2(\Theta) - (n_2 / n_1)]^{1/2}\} \quad (1)$$

onde

n_1 - índice de refração (cristal ATR).

n_2 - índice de refração (amostra).

θ - o ângulo de incidência.

λ - o comprimento de onda da radiação.

A radiação de penetração é chamada onda evanescente. Se o meio menos denso absorve a radiação evanescente, ocorre atenuação do feixe a comprimentos de onda correspondentes às bandas de absorção no infravermelho [15].

Nos espectros Raman tem-se um processo fisicamente diferente, o de espalhamento de luz. Sua atividade depende da variação do momento de dipolo induzido (pelo campo eletromagnético incidente) com a vibração [14].

No espalhamento inelástico Raman Stokes um fóton de energia $h\nu_0$ interage com uma molécula no estado vibracional fundamental E_0 , levando-a a um estado intermediário E_i (estado vibracional virtual). Ocorre o espalhamento com energia $h(\nu_0 - \nu_i)$ retornando a

molécula em um nível de maior energia E_2 , onde ν_v é a frequência vibracional correspondente à transição $E_2 - E_1 = h\nu_v$. Caso o fóton incidente interaja com a molécula no nível vibracional excitado E_2 o fóton espalhado pode, por decaimento do nível intermediário E_i ao estado fundamental E_1 , ter energia $h(\nu_0 + \nu_v)$, maior do que a do fóton incidente, dando origem ao espalhamento inelástico Raman Anti-Stokes. Se o fóton incidente sofre uma colisão puramente elástica será espalhado com a mesma frequência, originando a radiação Rayleigh (espalhamento elástico).

Dessa forma optou-se pela utilização da espectroscopia Raman para a determinação da insaturação C=C (variação do momento de dipolo induzido) e a espectroscopia no infravermelho para estudar C=O (variação do momento de dipolo elétrico) para o caso dos processos oxidativos. A espectroscopia Raman poderia ser utilizada para estudar o processo oxidativo, onde as variações nos espectros seriam devidos principalmente a diminuição da intensidade da banda correspondente à dupla ligação (reação com o oxigênio). Do mesmo modo a espectroscopia no infravermelho também poderia ser utilizada para determinar a insaturação, pois com os métodos quimiométricos, as informações, mesmo que mínimas, correspondentes às duplas ligações poderiam ser obtidas dos espectros totais.

2.1-As espectroscopias com transformada de Fourier

Estas espectroscopias são técnicas que empregam a operação matemática conhecida como transformada de Fourier para calcular espectros obtidos a partir de um interferograma. O princípio do interferograma de Michelson utilizado é o seguinte: a radiação vinda de uma fonte passa através de um "beam splitter" (divisor de feixe) que transmite metade do feixe para um espelho móvel e reflete a outra metade para um espelho fixo. Estes dois feixes são refletidos dos respectivos espelhos e recombinados ambos construtivamente ou destrutivamente novamente no "beam splitter", dependendo da posição do espelho móvel. Os detetores utilizados devem ser sensíveis aos fenômenos e ainda existe acoplado ao equipamento um microcomputador que registra os espectros [17].

A utilização de espectrofotômetros com transformada de Fourier permitiu que as antigas deficiências dos espectrofotômetros convencionais como elevados ruídos instrumentais e baixa sensibilidade fossem contornadas. É muito vantajoso o fato de que todos os dados das frequências espectrais são obtidos simultaneamente. Isto diminui o tempo de registro dos espectros e, assim, são obtidos espectros de maior razão sinal / ruído. Os espectros apresentam, também, frequências mais exatas e reprodutivas [16].

2.2-Fatores experimentais que afetam as análises espectroscópicas.

2.2.1-Espectroscopia no infravermelho.

Na técnica de ATR, alguns cuidados devem ser tomados para que os espectros obtidos sejam reprodutivos. O alinhamento do acessório de ATR e a temperatura são fatores que devem ser mantidos constantes para que as análises não sejam afetadas. Estes fatores sozinhos ou em combinação podem afetar seriamente a sensibilidade para análises quantitativas, pois eles influenciam o caminho ótico efetivo, que é o produto do grau de penetração e do número de reflexões internas. Pode-se programar o aparelho através de rotinas que consigam reconhecer estas condições (temperatura, alinhamento ótico e grau de penetração) [16] .

Outros fatores como remover e recolocar amostras no espectrofotômetro frequentemente resultam em grande variação espectral. Algumas vezes, isto pode ser minimizado tentando-se reproduzir a introdução da amostra na mesma posição de medida [18]. Pequenas variações na composição da amostra, sua superfície e diferenças de reflexão também podem contribuir para as variações. Para o caso de líquidos é importante que não haja a remoção do acessório de ATR do espectrofotômetro quando as celas são mudadas. Contudo, nos experimentos realizados estes efeitos não foram significativos uma vez que não houve variação na temperatura e nem no alinhamento do acessório de ATR.

2.2.2 - Espectroscopia Raman

A espectroscopia Raman tem atraído pouca atenção para análises quantitativas, principalmente pelo fato de que a reprodutibilidade dos valores de intensidades para os espectros de uma mesma amostra é algo de difícil obtenção. Uma razão para isso é que as intensidades dependem não somente dos parâmetros amostrais (α), como o espalhamento que ocorre quando da interação da radiação eletromagnética com a amostra (fatores intrínsecos), mas também dependem de fatores instrumentais (β).

$$I = \alpha \cdot \beta \quad (2)$$

com

$$\beta = B \cdot V \cdot F_{instr} \quad (3)$$

Sendo "B", a intensidade de radiação do laser, "V" o volume do material que espalha a radiação incidente e F_{instr} outros parâmetros que levam em conta fatores instrumentais como a sensibilidade do detetor, eficiência do monocromador, etc.

O espalhamento ocorre em todas as direções mas somente uma pequena porção da radiação que é espalhada é "coletada" pelo espectrofotômetro, sendo que em alguns casos os espectrofotômetros são ineficientes ao interconverter a luz espalhada em sinal. Além disso, pode ocorrer perda de radiação em todas as interfaces da cela. Estes e outros problemas, contudo, podem ser controlados [19].

A cela deve ter alta performance óptica. A fonte do laser deve ter uma potência constante e a mudança na posição de qualquer componente espectrofotométrico deve ser evitada para que a reprodutibilidade seja a melhor possível. Além disso, para análises quantitativas a correção do valor das intensidades é essencial. Isto é feito através de pré-tratamento de dados como a normalização dos espectros ou, então, empregando-se um padrão interno [6].

Tomados os devidos cuidados, esta técnica espectroscópica aliada a métodos quimiométricos constitui um poderoso instrumento de análises, tanto de classificação quanto quantitativas. Maiores informações sobre a espectroscopia Raman pode ser encontrada nas referências 20-22.

A seguir são mostrados nas Figuras 1 e 2 os espectros Raman e no infravermelho médio de um óleo vegetal, onde as principais diferenças podem ser notadas.

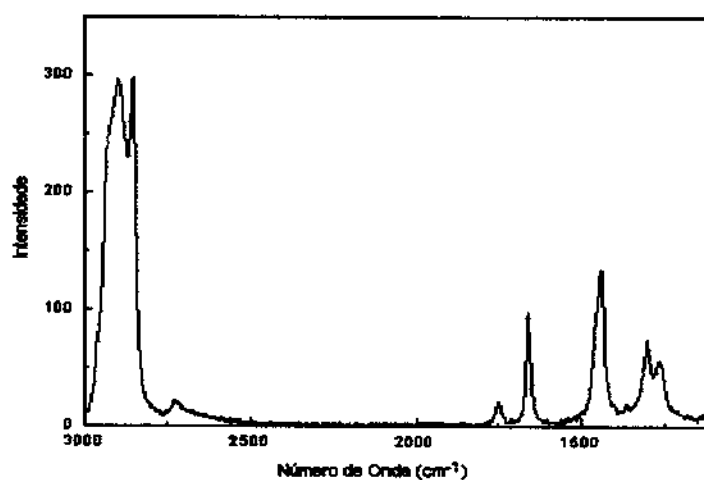


Figura1 - Espectro Raman de um óleo vegetal

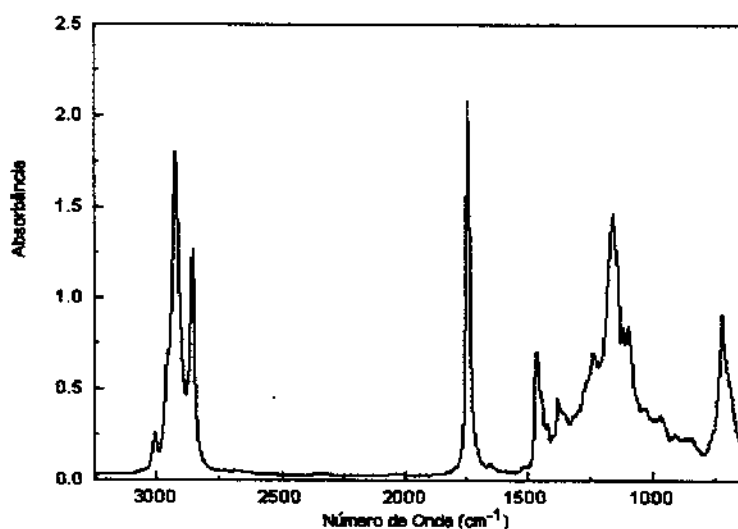


Figura2- E espectro no infravermelho médio de um óleo vegetal

CAPÍTULO III

Descrição dos Métodos Quimiométricos.

3 - A organização dos dados para análise multivariada

Uma tabela de dados ou um gráfico são os exemplos mais comuns de formas em que os dados multivariados podem ser arranjados .

Quando na forma de tabela, tem-se os objetos dispostos em linhas e as variáveis em colunas. No caso de dados químicos, os objetos vão corresponder aos compostos químicos (amostras) e as variáveis às medidas de alguma propriedade destes mesmos compostos.

Em geral, uma tabela de dados com " m " linhas e " n " colunas pode ser representada na forma de uma matriz de dimensão $m \times n$. A representação geral de uma matriz [23] de dados é mostrada na Figura 3 :

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & \dots & a_{2n} \\ \vdots & \vdots & & & \vdots \\ \vdots & \vdots & & & \vdots \\ a_{m1} & a_{m2} & \dots & \dots & a_{mn} \end{bmatrix}$$

Figura 3 - Representação de uma Matriz de Dados.

Um vetor é uma matriz que tem somente uma linha ou uma coluna sendo chamado de vetor linha ou coluna respectivamente. Os vetores são geralmente designados, por letras minúsculas em negrito (\mathbf{v}) . As matrizes são designadas por letras maiúsculas em negrito (\mathbf{A}) . Na matriz \mathbf{A} (figura 3), os termos representados por (a) correspondem aos elementos da matriz e seus índices (mn) representam respectivamente a linha e a coluna onde estes elementos aparecem. Os elementos das matrizes, dos vetores, ou ainda constantes como o coeficiente de regressão (b) são representados por letras comuns.

Todas estas notações são utilizadas em álgebra linear e adotadas em quimiometria.

Depois de organizados, pode-se analisar estes dados aplicando diversos métodos. A análise de componentes principais, um destes métodos, foi utilizada na construção das cartas multivariadas e, também, constitui a base do método de calibração multivariada baseado nos mínimos quadrados parciais (PLS). Ou melhor dizendo o PLS pode ser entendido como uma extensão do Algoritmo NIPALS, utilizado na análise de componentes principais. A partir daqui serão apresentados estes métodos.

3.1-Análise de componentes principais.

Os instrumentos de análises utilizados em laboratórios ou nos modernos processos industriais produzem uma grande quantidade de dados. Devido a esta vasta quantidade, estes dados podem muitas vezes ser redundantes ou estarem correlacionados. Um procedimento que permite solucionar estes problemas consiste na aplicação da análise de componentes principais [24,25, 26].

O principal objetivo da aplicação da análise de componentes principais (PCA) é reduzir a dimensão dos dados e formar novas variáveis que não são correlacionadas. Isto é feito através da combinação linear das variáveis originais $X = \{ x_k, 1, \dots, k \}$ que descrevem a maior tendência dos dados. Os novos fatores obtidos $T = \{ t_a, 1, \dots, a \}$ contém a maior parte da informação estatística útil presente nos dados originais .

A análise de componentes principais pode ser feita com base em autovalores e autovetores de uma matriz de variância – covariância, que é a matriz definida como: $C(X) = X'X$. De outra forma a PCA pode ser obtida através de diversos algoritmos, aplicados diretamente na matriz de dados. Estas duas situações serão demonstradas a seguir.

Primeiramente, como exemplo, serão calculados os autovalores e os autovetores de uma matriz de variância-covariância que suponha-se representam as variâncias – covariâncias da concentração de dois metais. Os autovetores e autovalores são utilizados a seguir para obter a PCA.

A matriz de variância-covariância (C_{x_1, x_2}) é a seguinte :

$$C_{(x_1, x_2)} = \begin{bmatrix} sx_1^2 & Cov \\ Cov & sx_2^2 \end{bmatrix}$$

Onde sx_1^2 e sx_2^2 representam as variâncias de x_1 e x_2 respectivamente e Cov é a covariância entre x_1 e x_2 .

Em termos numéricos, tem-se:

$$C_{(x_1, x_2)} = \begin{bmatrix} 3,07 & 2,47 \\ 2,47 & 2,43 \end{bmatrix}$$

Esta matriz foi encontrada a partir da concentração original dos dois metais aplicando a definição dada anteriormente para a matriz de covariância.

A matriz de variância-covariância (C) multiplicada por um autovetor x é igual a uma constante, o autovalor l , multiplicado por x (equação 4). Esta é a base de cálculos de autovalores e autovetores .

$$C \cdot x = l \cdot x \quad (4)$$

Para determinar estes autovalores , a equação 4 pode ser reescrita , como:

$$C \cdot x - l \cdot x = 0 \quad (5)$$

Ou $(C - l I) \cdot x = 0 \quad (6)$

onde I é a matriz identidade .

Se x não é zero então o determinante da matriz deve ser zero, isto é :

$$|C - l I| = 0 \quad (7)$$

Então:

$$|C - l I| = \begin{vmatrix} sx_1^2 - l & Cov \\ Cov & sx_2^2 - l \end{vmatrix} = 0 \quad (8)$$

A equação 8 origina a equação quadrática (equação 9)

$$(sx_1^2 - l)(sx_2^2 - l) - Cov^2 = 0 \quad (9)$$

Substituindo os valores das variâncias-covariâncias para os dois metais (matriz C_{x_1, x_2}), a equação fica:

$$(3,07 - l)(2,43 - l) - 2,47^2 = 0 \quad (10)$$

Ou
$$l^2 - 5,5l + 1,36 = 0 \quad (11)$$

Resolvendo-se a equação 11 encontra-se os dois autovalores (ou raízes características). O valor de l_1 é igual a 5,24 e valor de l_2 é igual a 0,26. Associado a cada autovalor está o autovetor (ou vetor característico) $v = [v_1, v_2]$ que pode ser encontrado através das equações 12 e 13.

$$\begin{bmatrix} sx_1^2 - l_1 & Cov \\ Cov & sx_2^2 - l_1 \end{bmatrix} \begin{bmatrix} v_{1_1} \\ v_{1_2} \end{bmatrix} = 0 \quad (12)$$

$$\begin{bmatrix} sx_1^2 - l_2 & Cov \\ Cov & sx_2^2 - l_2 \end{bmatrix} \begin{bmatrix} v_{2_1} \\ v_{2_2} \end{bmatrix} = 0 \quad (13)$$

onde v_1 e v_2 são os autovetores associados com o primeiro autovalor e v_2 e v_2 são os autovetores associados com o segundo autovalor.

Resolvendo-se as equações e normalizando da seguinte forma: $(v/\sqrt{v'v})$ obtém-se os resultados:

$$v_1 = [0,75 \quad 0,66] \quad (14)$$

$$v_2 = [-0,66 \quad 0,75] \quad (15)$$

As componentes principais são calculadas a partir desta base de autovalores e autovetores :

Para a primeira componente principal, tem-se :

$$PC1 = 0,75 X_{1i} + 0,66 X_{2i} \quad (16)$$

E para a segunda componente principal:

$$PC2 = -0,66 X_{1i} + 0,75 X_{2i} \quad (17)$$

Então, os elementos dos autovetores (0,75 0,66) vem a ser os coeficientes requeridos para a combinação linear das variáveis originais e são conhecidos como os "loadings". Os elementos individuais das novas variáveis (PC1 PC2) são derivados de X_1 e X_2 e são conhecidos como os "scores".

A variância total dos dados originais (metais A e B) é dada pela soma de suas variâncias individuais : $3,07 + 2,43 = 5,5$. Dessa forma o metal A contribui com 56% da variância e o metal B com os 44% restantes. No caso da PCA, tem-se que os autovalores (5,24 e 0,26) representam a variância associada a cada componente principal PC1 e PC2. A primeira componente principal contém, por esta razão, $5,24/5,50$ ou mais de 95% do total de variância e o segundo componente principal menos do que 5% $0,26/5,50$.

Como dito anteriormente, a análise de componentes principais, pode ser feita também através de diferentes algoritmos decompondo-se uma matriz de dados X de "rank" r , em uma soma de matrizes de posto igual a 1, como na equação 18.

$$X = M_1 + M_2 + M_3 + \dots + M_n \quad (18)$$

onde "rank" representa o número de vetores linearmente independentes de uma matriz. Essas novas matrizes com "rank" igual a 1, correspondem aos produtos dos vetores "scores" t_h e "loadings" p_h . Os "scores" e os "loadings" são calculados aos pares por um processo iterativo.

$$X = t_1 p'_1 + t_2 p'_2 + \dots + t_h p'_h \quad (19)$$

Ou $X = TP'$ (20)

A representação gráfica é mostrada na figura 4.

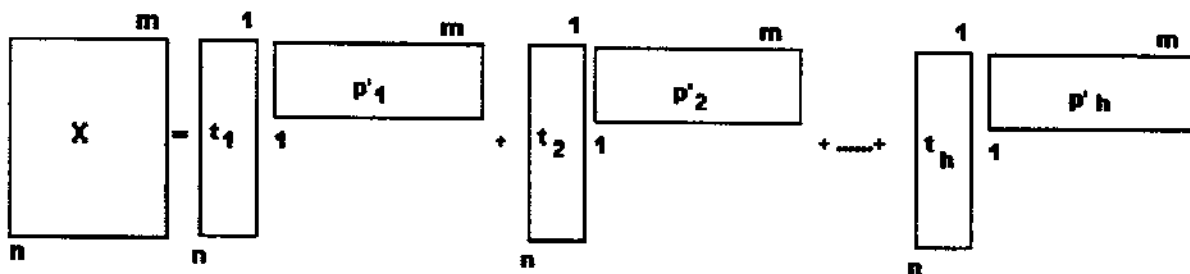


Figura 4- Representação da somatória dos produtos de matrizes dos vetores "scores" e "loadings".

Para exemplificar o significado de t_h e p'_h tem-se a figura 5 que ilustra a análise de componentes principais para o caso de duas variáveis x_1 e x_2 . A figura (5A) mostra uma componente principal que é a reta que aponta na direção de maior variabilidade das amostras (pontos 1 a 6) da figura 5B. Os "scores" t_h são as projeções das amostras na direção da componente principal e os "loadings" p'_h são os cossenos dos ângulos formados entre a componente principal e cada variável.

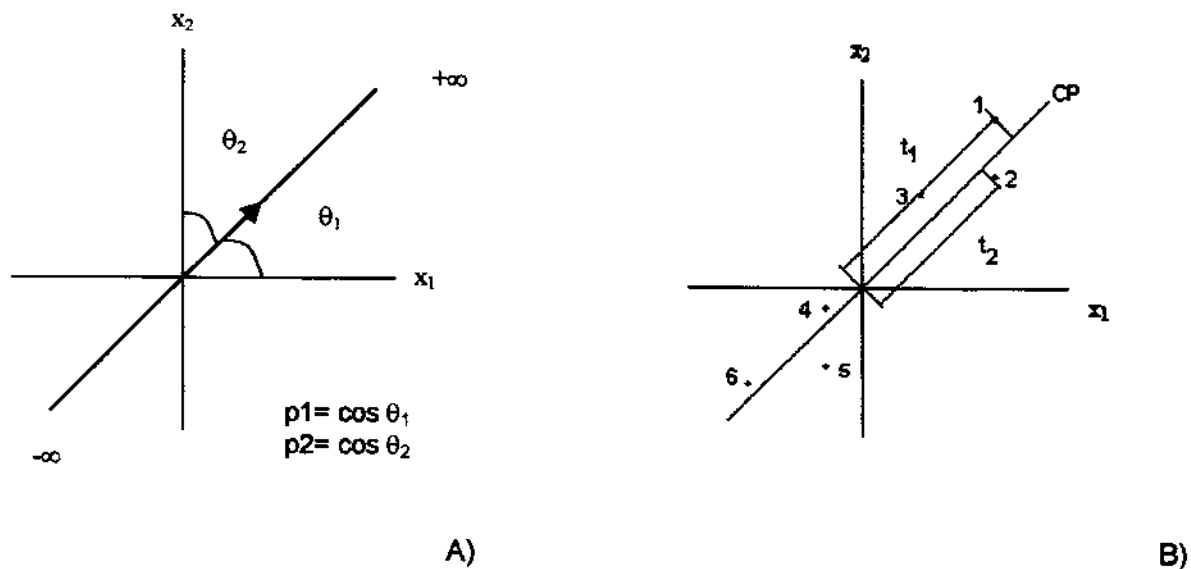


Figura 5 .Uma componente principal no caso de duas variáveis : (A) Os " loadings " são cossenos dos ângulos entre a componente principal e cada variável ; (B) " scores " são as projeções das amostras (1-6) na direção da componente principal.

3.2-A carta de controle de qualidade univariada , segundo Shewart

Baseando-se na distribuição normal as cartas de controle univariadas, segundo Shewart, têm duas finalidades específicas :

1. Verificar se o processo está sob controle, ou seja, a sua estabilidade. Para isso, somente valores observados nas variáveis características de qualidade das amostras são disponíveis. Não existem padrões de comparação, ou seja, as informações são fornecidas pelas próprias amostras. A norma baseia-se na linha média e nos limites de controle (estabelecidos com o uso da média e dos desvios padrões dos valores das variáveis de qualidade).
2. Verificar se o processo permanece sob controle. Corresponde a verificar a conformidade do processo de fabricação de acordo com a norma pré-existente. Para isso, além dos valores das variáveis características de qualidade observados nas amostras dispõe-se de um norma de controle (definida em 1). Os valores amostrais, além daqueles admitidos na norma, indicarão que o processo saiu de controle. A situação é a que existe quando se deseja manter a fabricação ou o controle de qualidade de um produto em um determinado nível de qualidade.

3.3 - A construção das cartas univariadas.

A construção destas cartas obedecem o seguinte esquema geral: [13].

- Na abscissa coloca-se os números das amostras ou então o tempo.
- Na ordenada existem três valores importantes (correspondentes à média e aos desvios-padrões) pelos quais são traçadas linhas horizontais :

1. O valor médio que determina a linha média (ou linha central) do gráfico representada por μ .
2. O limite superior de controle que determina a linha superior de controle representando o desvio de $+2\sigma$ (limite de confiança de 95%).
3. O limite inferior de controle que determina a linha inferior de controle representando o desvio de -2σ (limite de 95% de confiança).

O limite de controle especificado neste trabalho foi de 95%, porém os limites de 99,7% de confiança (correspondente a $\pm 3\sigma$) ou 68% (correspondente a $\pm \sigma$) também poderia ter sido utilizado.

O gráfico fica dividido em duas regiões :

- a região 1 (interior aos limites de controle de 2σ) correspondente à faixa de normalidade ou de controle .
- a região 2 (fora dos limites de controle de 2σ) correspondente a faixa de anormalidade e/ou de descontrole .

3.4-As indicações da falta de controle são as seguintes :

1. Pontos situados fora dos limites de controle, na região 2.
2. Pontos situados na região 1, muito próximos a uma linha de controle (inferior ou superior).
3. Uma disposição peculiar dos pontos na região 1, muito embora nenhum deles esteja próximo de nenhuma linha de controle.

Destas indicações a mais importante é a primeira citada.

Estas situações são melhor visualizadas através dos gráficos, mostrados nas figuras de 6 a 8, onde estão exemplificadas as três situações apresentadas acima.

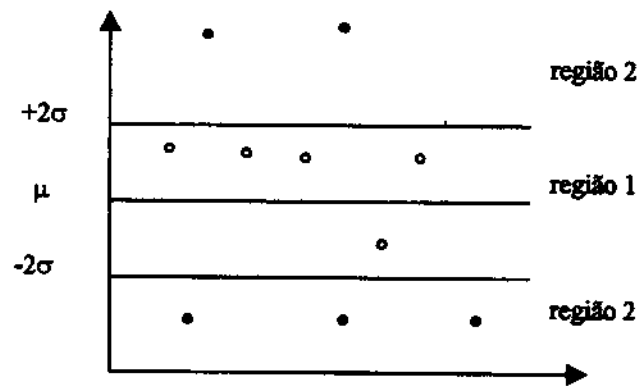


Figura 6 –Carta Univariada mostrando a situação fora de controle (pontos escuros) referente a 1.

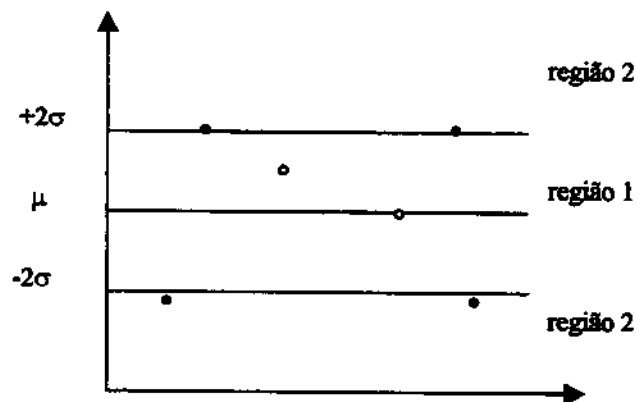


Figura 7-Carta Univariada mostrando situação fora de controle (pontos escuros) referente a 2

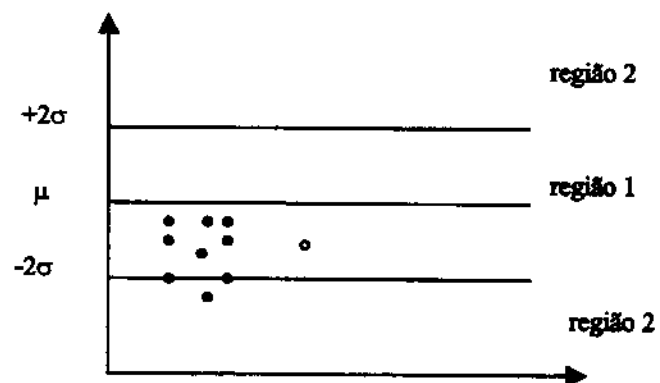


Figura 8-Carta Univariada mostrando situação fora de controle (pontos escuros) referente a 3

3.5- As cartas de controle de qualidade multivariadas

Para a apresentação das cartas multivariadas uma breve discussão sobre as cartas T^2 será feita .

3.6- O problema do T^2

Para introduzir o conceito estatístico de Hotelling's T^2 e a construção das cartas multivariadas utiliza-se primeiramente a teoria da distribuição t .

Usando primeiramente esta teoria pode-se fazer a seguinte análise : verificar se um valor específico μ_0 é um valor plausível para a média populacional μ . Com isso duas hipóteses são aceitas :

$$H_0: \mu_0 = \mu \quad \text{e} \quad H_1: \mu_0 \neq \mu$$

Aqui H_0 é a hipótese que diz que o valor é plausível e H_1 é a hipótese alternativa. Se x_1, x_2, \dots, x_n correspondem a amostras de uma população normal, então usando o teste t :

$$t = \frac{(\bar{X} - \mu_0)}{s/\sqrt{n}} \quad (21)$$

onde

$$\bar{X} = \frac{1}{n} \sum_{j=1}^n X_j \quad (22)$$

e

$$s^2 = \frac{1}{n-1} \sum_{j=1}^n (X_j - \bar{X})^2 \quad (23)$$

Rejeita-se H_0 se o valor de t em módulo encontrado exceder ao valor correspondente àquele de uma porcentagem específica (nível de confiança de 95%, 99,7% , etc) da distribuição t com (n-1) graus de liberdade.

De outra forma, esta análise pode ser feita utilizando t^2 , então elevando ao quadrado a equação 21, tem-se:

$$t^2 = \frac{(\bar{X} - \mu_0)^2}{s^2/n} = n(\bar{X} - \mu_0) (s^2)^{-1} (\bar{X} - \mu_0) \quad (24)$$

A variável t^2 na equação acima indica uma medida de distância, ou seja, ela corresponde à distância entre o valor da amostra média \bar{X} e o valor μ_0 . As unidades de distância são expressas em termos de s/\sqrt{n} (desvio padrão estimado de \bar{X}). Uma vez que \bar{X} e s^2 são observados, o teste vem Rejeitar H_0 (μ_0 não é um valor plausível para a média populacional) em favor de H_1 , ao nível de significância α , se:

$$n(\bar{X} - \mu_0) (s^2)^{-1} (\bar{X} - \mu_0) > t_{n-1}^2(\alpha/2) \quad (25)$$

onde $t_{n-1}^2(\alpha/2)$ corresponde à região do limite de confiança da distribuição t com (n-1) graus de liberdade.

Se H_0 não é rejeitado conclui-se que μ_0 é um valor adequado para a média da população normal. Existe, na verdade, um intervalo de valores para os quais μ_0 pode ser considerado um valor adequado para a média populacional. Este intervalo é dado pela expressão 26.

$$\bar{x} - t_{n-1}(\alpha/2) s/\sqrt{n} \leq \mu_0 \leq \bar{x} + t_{n-1}(\alpha/2) s/\sqrt{n} \quad (26)$$

μ_0 está na região do limite de confiança.

O procedimento pode ser expandido para o caso de vetores (para determinar se existe dentro do vetor valores plausíveis para a média multivariada). Por analogia com o que foi feito até agora pode-se utilizar as equações 27 a 29 para esta análise:

$$T^2 = (\bar{\mathbf{X}} - \boldsymbol{\mu}_o) \left(\frac{1}{n} \mathbf{S} \right)^{-1} (\bar{\mathbf{X}} - \boldsymbol{\mu}_o) = n (\bar{\mathbf{X}} - \boldsymbol{\mu}_o) \mathbf{S}^{-1} (\bar{\mathbf{X}} - \boldsymbol{\mu}_o) \quad (27)$$

onde

$$\bar{\mathbf{X}} = \frac{1}{n} \sum_{j=1}^n \mathbf{X}_j \quad (28)$$

$$\mathbf{S} = \frac{1}{n-1} \sum_{j=1}^n (\mathbf{X}_j - \bar{\mathbf{X}}) (\mathbf{X}_j - \bar{\mathbf{X}})' \quad (29)$$

$$\boldsymbol{\mu}_o = \begin{bmatrix} \mu_{10} \\ \mu_{20} \\ \dots \\ \mu_{p0} \end{bmatrix}$$

onde \mathbf{S} é a matriz de covariância estimada de $\bar{\mathbf{X}}$.

Se a distância entre $\bar{\mathbf{x}}$ e $\boldsymbol{\mu}_o$ origina um valor de T^2 (equação 27) que ultrapassa o valor limite dado pela equação 30, a hipótese $H_o: \boldsymbol{\mu}_o = \boldsymbol{\mu}$ é rejeitada.

Neste caso não se utilizam as tabelas de valores para t , uma vez que T^2 é distribuído como uma distribuição F :

$$\frac{(n-1)p}{n-p} F_{p, n-p}(\alpha) \quad (30)$$

onde $F_{p, n-p}$ é uma variável com uma distribuição F com p e $n-p$ graus de liberdade e o (α) representa a porcentagem da região do limite de confiança da distribuição F .

Uma discussão sobre a distribuição estatística da equação 30 pode ser encontrada na referência 23.

Então, tem-se que ocorre a rejeição de H_0 em favor de H_1 , em um nível de significância α se :

$$T^2 = n(\bar{X} - \mu_0)'(S)^{-1}(\bar{X} - \mu_0) > \frac{(n-1)p}{n-p} F_{p, n-p}(\alpha) \quad (31)$$

As cartas multivariadas T^2 verificam se um determinado valor pertence ou não a uma dada população. Nestas cartas só há um limite de controle (dado pelo termo que vem após $>$ na equação 31), ao contrário das cartas univariadas, onde dois limites são necessários.

3.7- Cartas de controle de qualidade multivariadas usando componentes principais

As cartas multivariadas T^2 (Hotelling's) e Q foram utilizadas neste trabalho para monitorar a qualidade dos óleos vegetais. As cartas T^2 podem ser construídas com ou sem a utilização da análise de componentes principais. Já a carta Q só tem aplicação quando baseada nessa análise pois ela mede os desvios em relação ao modelo PCA .

Quando um grande número de variáveis são disponíveis existe grande probabilidade destas estarem correlacionadas. Com isso, a matriz de covariância S na equação 31 toma-se ou é bem próxima de uma matriz singular e sua inversa consequentemente não poderia ser obtida. Um procedimento comum para evitar problemas de correlação é, como dito anteriormente, aplicar a análise de componentes principais.

Com o uso de algoritmos a decomposição da matriz X em "scores" (t) e "loadings" (p) e ainda uma matriz de resíduos (E) é novamente representada na equação 32.

$$X = TP' + E = \sum_{i=1}^k t_i p_i' + E \quad (32)$$

Na maioria das vezes, devem ser realizados pré-tratamentos para os dados antes da aplicação da PCA. Estes podem ser: centrar na média, escalonamento para variância unitária, etc.

Para a construção das cartas multivariadas um problema crítico consiste em estabelecer o número ideal de componentes principais que deve ser utilizado em cada caso. Existem inúmeras técnicas para isso. O método adotado neste trabalho é baseado na análise do gráfico que apresenta em sua ordenada os autovalores (ou seja, o quanto de variância explica cada componente principal) e na abscissa o número de componentes principais. Isto é importante, pois as componentes principais significativas (mínimo do gráfico) serão utilizadas para a construção das cartas T^2 enquanto as componentes principais remanescentes serão utilizados para a construção das cartas Q.

3.8-A carta de qualidade multivariada T^2

O valor de T^2 é estimado pelos "scores" (t) das medidas multivariadas no espaço definidos pelas A ($\leq k$) componentes principais. O T^2 é uma medida da variação de cada amostra dentro do modelo PCA e é definido como [27]:

$$T^2 = \sum_{i=1}^A \frac{t_i^2}{s_i^2} \quad (33)$$

onde t_i são os "scores" e s_i^2 é a medida da variância destes "scores". A variância é dada pelos autovalores (λ).

Quando se "divide" os "scores" pela variância, ou seja pelos autovalores estabelece-se que cada um dos componentes principais contribuem igualmente para o cálculo de T^2 . Caso todos os componentes principais fossem utilizados para a construção das cartas T^2 , quando efetuada a divisão dos scores (t) por uma pequena variância, ocorreria o seguinte: valores de variância mínimos teriam grande importância no cálculo

e desse modo o valor limite de T^2 obtido seria muito alto. Provavelmente nenhuma amostra analisada posteriormente teria um valor maior que este limite e assim nenhuma anomalia poderia ser identificada. Por esta razão a carta T^2 baseada nas primeiras componentes principais consiste no cálculo de T em relação às variáveis qualidade do produto que são de maior importância.

O cálculo do limite de confiança para T^2 é calculado com base na distribuição F , como segue:

$$T^2 = \frac{(m-1)A}{m-A} \cdot F_{A, m-A, \alpha} \quad (34)$$

onde m é o número de amostras usadas para desenvolver o modelo PCA, A é o número de componentes retidas no modelo e α a porcentagem da região do limite de confiança.

Contudo o monitoramento via T^2 somente baseados nas primeiras componentes principais não é suficiente. Este irá detectar se existe ou não alguma amostra anômala em relação ao modelo ou carta de qualidade construída baseada nas primeiras componentes principais. Porém se um evento totalmente novo que ocorre não está previsto pelos dados de referência usados para desenvolver o modelo PCA em controle, então novas PC(s) precisariam ser utilizadas para obter estas informações. É necessário, então analisar as demais componentes principais. Isto é feito através das cartas Q.

3.9- A carta de qualidade multivariada Q

A falta de ajuste estatístico Q, onde Q é simplesmente a soma dos quadrados dos erros (E) de cada amostra [27] é dada pela equação 35.

$$Q_i = \sum_{i=A+1}^k \frac{t_i^2}{S_{\theta}^2} \quad (35)$$

Onde $\sum_{i=A+1}^k \frac{t_i^2}{S_i^2}$ representa o erro (e) ou resíduo. No caso da carta Q são utilizadas as componentes principais restantes (de A até k), ou seja são utilizadas as informações das componentes principais não significativas. O Q pode ser entendido como uma medida da diferença ou resíduo entre o valor das observações e o valor de sua projeção nas componentes principais retidos no modelo. Quando o processo está sob controle os valores de Q devem ser pequenos.

Os limites de confiança para Q são calculados a partir das equações que utilizam os autovalores :

$$Q_{\alpha} = \Theta \left[\frac{c\alpha \sqrt{2\Theta_2 h_o^2}}{\Theta_1} + 1 + \frac{\Theta_2 h_o (h_o - 1)}{\Theta_1^2} \right]^{\frac{1}{h_o}} \quad (36)$$

onde

$$\Theta_i = \sum_{j=A+1}^k \lambda_j^i \quad \text{para } i=1,2,3 \quad (37)$$

$$\text{e} \quad h_o = 1 - \frac{2\Theta_1 \Theta_3}{3\Theta_2^2} \quad (38)$$

Na equação 36 $c\alpha$ é o desvio padrão normal, A na equação 37 corresponde ao número de componentes principais retidos no modelo e k ao número total de componentes principais.

Para o cálculo do limite de controle de Q usa-se os autovalores. Outro modo de se calcular o limite a partir de Q seria aproximar esta distribuição de uma distribuição qui-quadrada e calcular o limite dado por esta última.

Uma discussão sobre esta distribuição e cálculos do limite de controle pode ser encontrada na referência 28.

O limite Q define a distância do modelo (a partir da qual tem-se dados não usuais) baseada nos dados utilizados para se fazer o modelo PCA. T^2 por outro lado é a medida da distância das variáveis em relação a média multivariada. Os limites de controle indicam a faixa de variação considerada normal para estas cartas.

É importante notar que T^2 e Q são estabelecidos considerando que os dados seguem uma distribuição multivariada normal. Isto é admitido com base no teorema do limite central [23]. Os "scores" tendem a produzir medidas que são normalmente distribuídas em relação às variáveis originais. Já os erros tendem mais comumente a uma distribuição normal.

Para a construção das cartas de controle de qualidade multivariadas seleciona-se amostras representativas do processo em estudo que devem ser livres de situações anormais (ou seja, espera-se que todas as amostras apresentem valores próximos para a variável característica de qualidade que esta sendo analisada). Acaso seja detectada alguma amostra anormal esta deve ser excluída e os limites devem ser recalculados.

3.10-O que podem ser determinados através das cartas: Um Exemplo

Diferentes tipos de desvios podem ser definidos com estas cartas : Quando um determinado processo é " perturbado " em uma ou mais variáveis modeladas , sem mudança , contudo, na relação deste modelo ocorre o seguinte : o processo vai estar sem controle nas cartas T^2 mas estará sob controle nas cartas Q .Por exemplo, em um dado processo tem-se amostras com duas espécies. A partir destas amostras, respostas instrumentais são obtidas e destas respostas constróem-se as cartas de qualidade. Supondo que em um determinado dia, houve uma mudança na resposta instrumental para a espécie A , a mesma mudança também foi notada para a espécie B e ainda existe uma estrutura de correlação entre A e B (prevista pelo modelo). Neste caso a estrutura do modelo foi mantida e somente haverá uma situação fora de controle na carta T^2 .

Quando novos eventos ocorrem e que não são contemplados pelo modelo, um desvio significativo será visto na carta Q. Neste caso a carta T^2 permanecerá em controle. Um novo evento relacionado com o nosso exemplo é o seguinte: houve novamente mudanças para a resposta instrumental da espécie A enquanto observa-se que as mudanças ocorridas para a resposta instrumental do componente B vão em direção contrária. Neste caso, este tipo de evento não foi considerado pelo modelo e assim uma situação fora de fora de controle será constatada apenas na carta Q. Qualquer combinação de efeitos do primeiro e do segundo tipo irão causar uma situação fora de controle em ambas as cartas T^2 e Q .

3.11-Introdução às análises de regressão

Criar um modelo de calibração significa desenvolver expressões matemáticas que, de algum modo, possam descrever o comportamento de uma variável de interesse. Em todos os casos estas variáveis serão chamadas de variáveis dependentes e designadas pela letra Y. Como exemplo, em química, pode-se citar as concentrações de um analito de interesse.

Para a construção do modelo deve-se ter, primeiramente, alguma informação sobre a variável dependente. As informações sobre o comportamento destas variáveis são fornecidas por outras, as chamadas variáveis independentes. Exemplos dessas são os sinais analíticos que podem ser relacionados com as concentrações das substâncias. Estas novas variáveis serão designadas pela letra X. Os modelos envolvem ainda constantes que serão chamadas de coeficientes que controlam o comportamento do modelo e são estimadas a partir dos dados (variáveis). Tem-se o modelo matemático : $Y=BX$.

O modelo matemático consiste em estabelecer relações entre as variáveis dependentes e independentes, via alguma função de transferência. A complexidade matemática do modelo irá depender do processo que está sendo estudado. De qualquer forma, o modelo construído deve ser o mais realístico possível, ou seja, ele deve descrever adequadamente a relação entre X e Y.

Atualmente, com o avanço da instrumentação e o uso de computadores tem sido estimulado o surgimento de modelos de calibração multivariada [29,30].

Exemplos de métodos de calibração multivariada são : PLS (Mínimos Quadrados Parciais), PCR (Regressão a Componentes Principais), *MLR (Regressão Linear Múltipla), PLS N – Way, Parafac, Redes Neurais, Algoritmos Genéticos em conjunto com regressão linear múltipla, etc.

A calibração multivariada envolve duas fases : a construção do modelo - **calibração** e a verificação de sua validade, ou seja, se o modelo construído é ou não realístico - **previsão** (ou validação).

Na fase de calibração forma-se uma matriz com as respostas instrumentais (variáveis independentes - X) e uma outra matriz com as concentrações das substâncias (variáveis dependentes -Y). Os dados utilizados nesta etapa constituem o conjunto de treinamento.

A partir disso, desenvolve-se um modelo matemático adequado que melhor possa reproduzir Y a partir dos dados de X.

O modelo matemático desenvolvido na fase de calibração é, então, utilizado na fase de previsão para estimar as concentrações dos constituintes de novas amostras, a partir de suas respostas instrumentais (variáveis independentes). Os dados utilizados nesta fase formam o conjunto teste.

Na figura 9 está representado o procedimento adotado para a construção das matrizes X e Y .

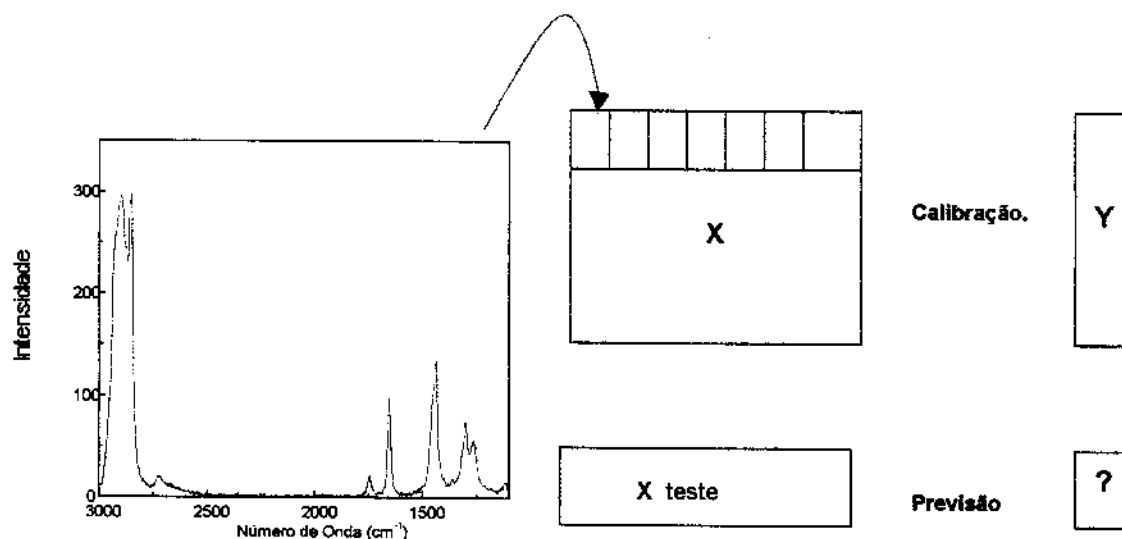


Figura 9 - Dados de espectros organizados para a calibração multivariada.

Nesta ilustração, X é a matriz de variáveis independentes. As linhas desta matriz correspondem às amostras e as colunas são formadas pelos valores das intensidades do espectro em cada número de onda (a flecha, mostrada na figura 9, indica esse procedimento – formação da primeira linha da matriz de dados com o espectro de uma amostra, as demais linhas são formadas pelos dados de espectros de outras amostras). Y é a matriz de variáveis dependentes. Suas linhas correspondem às mesmas amostras daquelas de X e as colunas são formadas pelos valores das concentrações destas amostras.

Na fase de previsão, X teste é a matriz formada com o uso de novas amostras da mesma forma que a matriz X . Esta, juntamente com o modelo desenvolvido é utilizada para prever os valores das concentrações destas novas amostras e conseqüentemente determina a matriz ? na ilustração 9.

3.12-Pré-tratamento de dados

Um procedimento comum e muito importante, realizado antes de qualquer método multivariado, como os de classificação ou calibração, são os de pré-tratamento de dados [31]. Estes procedimentos permitem correções dos dados obtidos (de outra forma interpretar e obter informações destes dados seria muito difícil em alguns casos), tornando os cálculos computacionais de análises multivariadas mais precisos e exatos.

A seguir serão descritos os procedimentos realizados para o tratamento dos dados obtidos através das análises espectroscópicas e utilizados neste trabalho.

Os dados obtidos através dos espectros Raman foram submetidos à normalização. A normalização média [32], aqui empregada, consiste em dividir cada variável pelo seu valor médio. Dessa forma para cada linha é obtido o respectivo valor médio (o valor médio para cada variável é encontrado pela soma de todas as variáveis e consequente divisão pelo número total de variáveis) e a seguir divide-se cada variável, dessa mesma linha, pelo valor médio encontrado. Para as demais linhas é feito o mesmo procedimento. A equação 39 é utilizada para o cálculo de normalização média.

$$X(i,k) = X(i,k) / \bar{X} \quad (39)$$

onde \bar{X} é o valor médio correspondente a linha.

Um último pré-tratamento utilizado neste trabalho e aplicado tanto para os dados dos espectros Raman quanto para os dos espectros no infravermelho médio foi o de centrar os dados na média [31]. Este, corresponde a fazer com que para cada variável seus valores tenham média zero. Isto é feito, obtendo-se para cada coluna o valor médio. Após isso subtrai-se este valor de cada variável dessa mesma coluna. Para as demais colunas o procedimento é o mesmo. Com isso, ocorre a mudança do sistema de coordenadas para o centro de dados. A equação 40 é a utilizada para centrar os dados na média.

$$X(i,k) = X(i,k) - \bar{X}_1 \quad (40)$$

onde \bar{X}_1 é o valor médio correspondente a coluna.

Existem muitas outras formas de pré-tratamento de dados como a utilização de filtros para eliminação da influência dos ruídos, outros tipos de normalizações, escalonamento de variáveis, emprego de derivadas, correção do espalhamento multiplicativo (MSC), etc. O significado de cada uma destas formas de pré-tratamentos bem como o modo de utilização pode ser encontrado nas referências 25 e 32. A decisão da aplicação de qualquer forma de pré-tratamentos deve ser avaliada e estudada caso a caso, tomando certos cuidados para que a escolha seja adequada ao sistema com que se está trabalhando. Nesta fase pode ser incluída ainda a redução ou eliminação de variáveis.

3.13- O método dos mínimos quadrados parciais (PLS)

O modelo de calibração baseado no método dos mínimos quadrados parciais (PLS) é um método relativamente moderno, desenvolvido e popularizado na ciência analítica por Wold [25,31].

O método PLS consiste em decompor as matrizes de dados X (variáveis independentes) e Y (variáveis dependentes) em uma soma de produtos de dois vetores (os " scores " e os " loadings "). Sabe-se que é possível representar uma matriz de dados, sem a perda de informação estatística útil, pela sua matriz de "scores" , com a vantagem de não haver correlação entre as variáveis. Isto é exatamente o que se faz no PLS, ou seja, tanto a matriz de variáveis independentes (X) como a das variáveis dependentes (Y) são representadas por seus " scores " , utilizando a redução de variáveis pela análise de componentes principais. Diversos algoritmos podem ser utilizados, como no caso da PCA.

O modelo resultante é :

$$\begin{aligned} X &= TP' + E \\ Y &= UQ' + F \end{aligned} \quad (41)$$

onde os elementos das matrizes de T e U são os "scores" de X e Y respectivamente e os elementos P e Q são os "loadings". As matrizes E e F correspondem aos erros, ou seja, o quanto o modelo não "consegue" explicar.

Uma relação entre os dois blocos (variáveis independentes X e variáveis dependentes Y) pode ser realizada correlacionando-se os "scores" do bloco X (T) com os "scores" do bloco Y (U) para cada componente principal .Um modelo linear é utilizado para esta relação:

$$u = b_n t_n \quad n=1,2,\dots,n \text{ para } n \text{ componentes principais.} \quad (42)$$

Este modelo, entretanto não é o melhor possível. Isto porque as componentes principais seriam calculadas para os dois blocos separadamente, podendo resultar numa relação não muito satisfatória entre os " scores " dos dois blocos. Para que se encontre uma melhor correlação possível torna-se importante manipular a informação desses dois blocos simultaneamente. No PLS isto é feito por uma leve rotação das componentes principais, com conseqüente mudança nos valores dos " scores ". Devido ao fato de ocorrer esta rotação, o termo componente principal já não é mais utilizado, no lugar deste o termo designado é o de variáveis latentes. As variáveis latentes já não descrevem a direção de maior variância (mas sim uma próxima). O que ocorre aqui é um compromisso entre a habilidade das variáveis latentes em descrever as amostras nos espaços individuais e o aumento na correlação entre t e u . A figura 10 exemplifica o modelo de calibração baseado no PLS.

Assim pode ser produzida uma relação mista entre os dois blocos :

$$Y = TBQ' + F \quad (43)$$

O modelo final consiste das matrizes dos "scores" T e U que são linearmente relacionadas por um coeficiente B (B é diagonal).

A figura 10 representa o modelo de calibração baseado nos mínimos quadrados parciais.

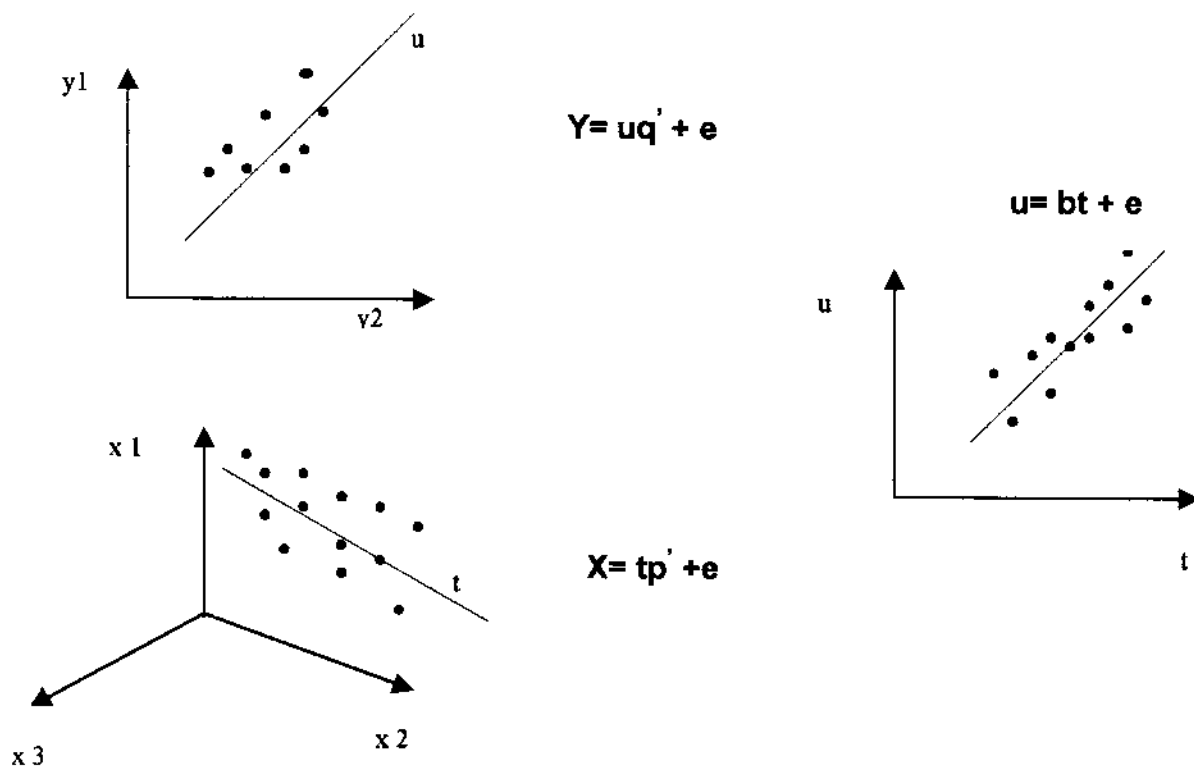


Figura 10- Representação geométrica do PLS com uma variável latente modelando os dados de X (variáveis independentes) e Y (variáveis dependentes). Os dados de X são formados por três variáveis e os dados de Y por duas.

3.14- Previsão.

A parte fundamental da regressão é a verificação de sua validade, ou seja, se o modelo construído é capaz de prever corretamente ou com pequena margem de erro os valores das concentrações das novas amostras.

Tomando como exemplo dados obtidos de espectros Raman, isto é feito da seguinte forma:

Para a fase de previsão é formada uma nova matriz. Então, os dados obtidos dos espectros Raman, ou melhor, os valores das intensidades em cada número de onda formam as colunas desta matriz (matriz X_{teste}) a qual tem seis linhas e cada linha corresponde a uma amostra. Após isto, os dados passam por pré-tratamentos matemáticos, assim como os dados de calibração.

O passo seguinte consiste em obter os vetores de "scores" (t) e os resíduos (e) da matriz X_{teste} utilizando os loadings (p) calculados na fase de treinamento.

$$X_{teste} = \sum_{a=1}^A t_a p_a' + e, \text{ para "A" variáveis latentes na modelagem} \quad (44)$$

Com os valores dos "scores" (t) da matriz X_{teste} , além dos coeficientes de regressão b e os loadings q do bloco Y calculados na fase de calibração, pode-se fazer a estimativa das concentrações da seguinte forma :

$$Y = \sum_{a=1}^A t_a b_a q_a', \text{ para "A" variáveis latentes.} \quad (45)$$

3.15-O número de variáveis latentes

O número de variáveis latentes utilizado em um modelo é de fundamental importância nos resultados a serem obtidos. A utilização de um número menor de variáveis latentes poderá fornecer resultados não satisfatórios, uma vez que toda a informação disponível dos dados originais não estará sendo utilizada. No caso porém de se usar um número de variáveis latentes superior ao necessário ocorrerá a modelagem de ruídos podendo ocasionar problemas no modelo [31].

Existem várias maneiras de se determinar o número ideal de variáveis latentes. O método adotado neste trabalho é o método de validação cruzada.

Neste método de validação, as mesmas amostras são usadas para modelar e validar o modelo construído. Isto é feito da seguinte forma: do conjunto de treinamento, usado para a construção do modelo de calibração, são retiradas certas amostras. Um novo modelo é construído com as amostras que sobraram. A seguir, as amostras que foram retiradas são utilizadas para a previsão, sendo calculados os resíduos referentes à esta previsão. O processo é repetido com outros grupos de amostras, até que todas as amostras tenham sido colocadas fora do modelo de calibração. Após isso todos os resíduos de previsão são combinados para dar a variância residual da validação para cada variável latente.

A primeira versão deste método e adotada neste trabalho consiste em retirar do conjunto de treinamento uma amostra de cada vez. Este método é chamado validação cruzada completa [32].

A verificação do número ideal de variáveis latentes é feita através de gráficos de variância residual contra o número de variáveis latentes. O mínimo deste gráfico é tomado como o número ideal de variáveis latentes pois a partir deste não ocorrem valores de variância significativos que possam melhorar o modelo a ser construído.

pode-se decidir qual procedimento adotar: eliminá-las ou mantê-las no processo. Para esta decisão é importante entender, quando possível, o que significa cada anomalia.

Amostras com altos ou baixos níveis de analitos são, geralmente, amostras muito diferentes do restante do conjunto. Em casos porém, onde a resposta instrumental é linear, ou melhor, o modelo de calibração é linear, amostras desse tipo são mais informativas do que propriamente anomalias e devem ser mantidas no modelo de calibração. Outros tipos de anomalias que não podem ser explicadas e que afetam o modelo de calibração devem ser eliminadas. A identificação do problema pode estar na análise experimental ou instrumental, sendo necessário repetir os procedimentos. Dentro de certos critérios, como não existir nenhum problema com os equipamentos e o número de amostras diferentes ser reduzido o meio mais fácil consiste na eliminação destas amostras.

Os métodos utilizados para a detecção de anomalias neste trabalho, foram baseados no estudo de "leverages" e análise residual [25].

3.17- Leverages.

Em um espaço p-dimensional quanto mais próximos dois pontos estiverem mais semelhantes serão as amostras ou variáveis relativas a esses pontos. A leverage retrata exatamente isso: a similaridade entre as amostras. A "leverage" pode ser interpretada como a distância entre um ponto ou variável e o centro do modelo, ou seja, o quão distante um ponto ou uma variável esta quando comparada com o conjunto total.

A "leverage" de um objeto descreve sua influência ou a contribuição deste para o modelo de calibração. Uma "leverage" próximo a zero indica que a amostra ou variável correspondente têm muito pouca influência no modelo de calibração. A observação contrária, ou seja altas "leverages", indica que determinado ponto têm alta influência no modelo. Com isso, tem-se que as "leverages" indicam se uma amostra é ou não diferente das demais mas é necessário outros estudos para se afirmar que esta amostra corresponde realmente a uma anomalia.

Nos modelos de regressão, como o PLS, a "leverage" (h_i) pode ser encontrada através da seguinte expressão:

$$h_i = 1/I + \sum_{a=1}^A \left(t_{ia}^2 / \hat{t}_a' \hat{t}_a \right) \quad (46)$$

que corresponde a distância mahalanobis de um objeto para o centro do modelo. Nesta expressão, t_{ia} é o valor do score para o objeto i , \hat{t}_a são os vetores scores, A o número de variáveis latentes e o termo $1/I$ representa a contribuição a partir da intercessão, para o caso de modelos de regressão.

A identificação de anomalias através das "leverages" é feita graficamente. Um gráfico que tem em um eixo os valores das "leverages" e em outro números das amostras.

3.18- Análise residual em Y.

Os resíduos constituem uma outra forma de se detectar anomalias. Estes correspondem a desvios entre os dados de referência e os estimados pelo modelo. Resíduos grandes significam que o modelo não está conseguindo modelar os dados de forma adequada. Os resíduos podem ser calculados tanto para X quanto Y . Contudo aqui, será discutida a análise residual somente para os dados de Y pois esta foi a análise utilizada neste estudo.

Os valores referentes aos dados de Y são conhecidos nos dados de calibração e em alguns casos também nos dados de validação. Desse modo, os resíduos podem ser calculados através da diferença entre o valor de referência y_j e o valor previsto \hat{y}_{ij} pelo modelo, segundo a equação abaixo :

$$f_{\hat{y}} = y_{\hat{y}} - \hat{y}_{\hat{y}} \quad (47)$$

Através dos resíduos pode ser calculada a variância residual . A variância residual é definida como a média quadrática residual corrigida pelo número de graus de liberdade .

$$RESVAR = \frac{1}{d} \sum_{i=1}^i f_{\hat{y}}^2 \quad (48)$$

O número de graus de liberdade (d na equação acima) é definido depois de feita a regressão. A variância dos resíduos de Y expressa como muito da variação permanece nas respostas observadas depois de construído o modelo.

3.19- O gráfico de Influência

O gráfico de influência mostra em uma coordenada a variância residual de X ou Y contra a "leverage" (Se a variância de X for usada, a influência é no espaço X e se a variância de Y for usada, a influência é na relação de regressão entre X e Y).

Estes gráficos foram construídos para mostrar dois tipos de diagnósticos de anomalias simultaneamente, mas podem ser usados também para encontrar quais amostras tem mais influência no modelo. Amostras com altas leverages e alta variância residual são mais influentes e em geral este é o pior caso representando uma anomalia (neste caso tais amostras devem ser excluídas do modelo de calibração).

CAPÍTULO IV

*Determinação do Nível Total de Insaturação de
Óleos Vegetais por Espectroscopia Raman e
Calibração Multivariada.*

PARTE EXPERIMENTAL

4-Parte experimental

4.1-Determinação do índice de iodo dos óleos vegetais.

Nesta parte do trabalho, foram utilizadas amostras de óleos vegetais de diferentes variedades como : milho, oliva e soja (de diferentes procedências), bem como, arroz, canola, gergelim, girassol, dendê e misturas de óleos de soja e oliva e óleos de girassol e oliva adquiridas em um supermercado local. Além destes, foram utilizados os óleos de palma e palmiste fornecidos pela Faculdade de Engenharia de Alimentos da UNICAMP (laboratório de Óleos e Gorduras) e outras misturas preparadas em laboratório, a partir do óleo de palma com os óleos de soja e de palmiste. A quantidade final destas misturas foi expressa em termos de porcentagem de massa (esses resultados constam da tabela 3). No total, vinte e quatro amostras de óleos vegetais, abrangendo a faixa de índices de iodo entre 17 e 130 foram analisadas .

Reagentes e soluções

Solução de Tiosulfato de Sódio 0,1M

Para a preparação da solução de tiosulfato de sódio 0,1M pesou-se 25 gramas de tiosulfato de sódio pentahidratado e dissolveu-se em 1 litro de água recentemente fervida e resfriada. Em seguida foi adicionado a solução 0,1 grama de carbonato de sódio. A solução foi deixada em repouso por um dia antes de ser padronizada.

Solução de Iodeto de Potássio 10%

Para a preparação desta solução dissolveu-se 10 gramas do sal em 100 mL de água destilada.

Solução de Amido 1%.

A solução de amido 1% foi preparada para uso diário. Dessa forma dissolveu-se aproximadamente 1 grama de amido em água quente, sob agitação. Aqueceu-se esta solução por mais alguns minutos até se obter uma cor clara e filtrou-se. Após o resfriamento, esta solução foi armazenada em recipiente fechado.

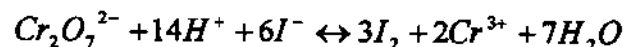
Reagente de Wijs

O reagente de Wijs corresponde a uma solução de monocloreto de iodo [ICl] em ácido acético. É um reagente adquirido comercialmente.

Padronização do Tiosulfato de Sódio.

Na padronização do tiosulfato de sódio 0,1 M pesou-se cerca de 0,2000 gramas de dicromato de potássio puro e seco em estufa a 120°C por 2 ½ horas (padrão primário) e dissolveu-se esta substância em 50 mL de água destilada. A seguir foram adicionados ao meio 2 gramas de iodeto de potássio e 8 mL de ácido clorídrico concentrado. Homogeneizou-se e titulou-se com a solução de tiosulfato, sob agitação constante, até que a cor da solução castanha mude para verde amarelado. Neste ponto adicionou-se 3 mL da solução de amido e prosseguiu-se a titulação até mudança brusca da cor azul para verde.

Em meio ácido os íons dicromato reagem com iodeto de acordo com a equação :



Com os dados obtidos pode-se calcular a concentração da solução padrão de tiosulfato.

4.2 - Método padrão para a determinação do índice de iodo (AOCS)

Para a determinação do índice de iodo nas amostras de óleos vegetais foi tomado como referência o método oficial da American Oil Chemists' Society (AOCS) [3].

Os materiais utilizados foram limpos e secos. Não deve haver água em momento algum no início deste procedimento.

De acordo com o método padrão, pesou-se uma quantidade de óleo adequada ao índice de iodo esperado. A tabela 2 mostra o valores dos índices de iodo seguido das respectivas massas .

Tabela 2- Massa de óleo a ser pesada de acordo com o valor do índice de iodo.

Índice de iodo	Massa (g)
<3	10,0000
3	10,5760
5	6,3460
10	3,1730
20	1,5865
40	0,7935
60	0,5288
80	0,3966
100	0,3173
120	0,2644
140	0,2266
160	0,1983
180	0,1762
200	0,1586

Após a pesagem, as amostras de óleos vegetais foram transferidas para erlenmeyers de 500 mL com tampa esmerilhada. A seguir foram adicionadas às amostras 20 mL de ciclohexano P.A e 25,00 mL de solução de Wijs 0,1 mol/L.

Os erlenmeyers foram fechados, levemente agitados e finalmente estocados em local escuro por cerca de uma hora à temperatura ambiente (25° C).

Após este período, os erlenmeyers foram retirados deste local e foram adicionados à estes 20 mL de solução de iodeto de potássio 10% e 100 mL de água destilada. Para cada grupo de amostras analisadas fez-se um branco.

As amostras foram tituladas, em seguida, com solução de tiosulfato de sódio 0,1 mol/L , previamente padronizada. Inicialmente a cor da própria solução serviu como indicador. Esta titulação foi feita até que a cor amarela da solução tivesse quase desaparecido. Adicionou-se então amido 1% e prosseguiu-se a titulação até o desaparecimento da cor azul.

Os valores de índices de iodo obtidos por este método vão corresponder às variáveis dependentes utilizadas posteriormente para a construção e validação do modelo de calibração .

4.3 - Os espectros Raman.

Os espectros Raman das amostras de óleos vegetais foram obtidos em um espectrofotômetro Bomen DA8 FT spectrometer (Hartmann & Braun). O laser utilizado foi o Nd:YAG que emite luz a um comprimento de onda de 1064 nm . O instrumento é equipado com beamsplitter de quartzo e detetor InGaAs refrigerado com nitrogênio líquido. Os espectros foram obtidos com resolução de 4cm^{-1} em um total de 256 varreduras.

Para as análises, todas as amostras foram colocadas em uma cela de bulbo de vidro que tem metade de sua parede interna recoberta com alumínio .Isto permite que a

luz do laser espalhada seja melhor coletada pelo detetor do espectrofotômetro. As celas foram preenchidas e esvaziadas com o auxílio de uma pipeta pasteur, sendo limpas com diclorometano e secas ao ar.

Os óleos vegetais foram analisados na forma líquida. Quando necessário procedeu-se a um aquecimento das amostras que se mostravam sólidas à temperatura de aproximadamente 50°C em estufa.

Os espectros Raman foram adquiridos em duplicata, sendo utilizado para fins de cálculos, os valores dos espectros médios. A faixa espectral do deslocamento Raman utilizada neste trabalho foi de 3000 a 1000 cm^{-1} .

Os valores das intensidades, em cada número de onda, obtidos dos espectros Raman corresponderão às variáveis independentes utilizadas para a construção e validação do modelo.

A figura 11 mostra os espectros Raman dos diferentes tipos de óleos vegetais utilizados neste trabalho.

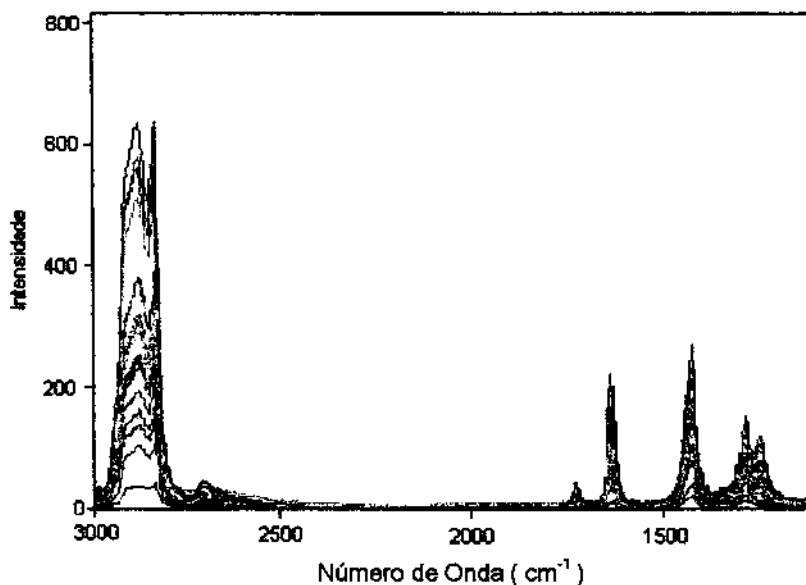


Figura 11- Espectros Raman dos Óleos Vegetais

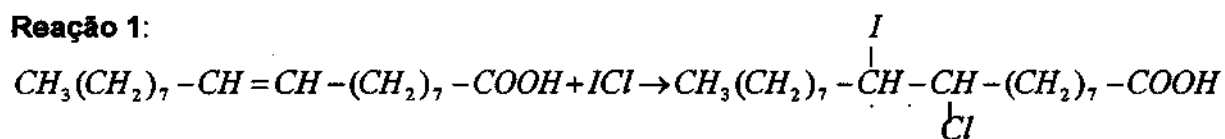
RESULTADOS E DISCUSSÕES

4.4- Resultados e discussões da determinação do índice de iodo dos óleos vegetais.

A adição quantitativa de monocloreto ou monobrometo de iodo constitui a reação básica para a determinação do índice de iodo de óleos e gorduras. O índice de iodo é uma medida do nível total de insaturação e é expresso em termos da quantidade de iodo (em gramas) "absorvido" por cem gramas do material analisado. Este índice representa a verdadeira insaturação dos óleos vegetais quando as duplas ligações entre os carbonos não são conjugadas e também não se situam em posição adjacente à carboxila, pois nestes casos a adição de halogênios é incompleta ou seja não quantitativa.

A reação de adição dos halogênios como cloro ou bromo às duplas ligações do carbono nos ácidos graxos insaturados é bastante favorável, mesmo em casos onde estes ácidos estejam combinados, isto é, mesmo em triglicerídeos, como nos óleos vegetais. O iodo é o menos reativo e só é adicionado na forma de monobrometo ou monocloreto de iodo. A reação de adição deste último é mostrada na reação 1:

Reação 1:



Esta reação ocorre simplesmente misturando-se os reagentes, em solvente inerte. A reação de adição procede rapidamente à temperatura ambiente ou inferiores, não necessitando de exposição à luz ultravioleta; na verdade até se procura evitar temperaturas mais altas e exposição à luz, assim como a presença de excesso de halogênio porque, do contrário, a reação de substituição torna-se uma reação secundária importante [33].

O mecanismo da reação de adição, envolve dois passos: No primeiro passo, o iodo transfere-se da molécula do monocloreto de iodo (ICl) para a dupla ligação do carbono, formando um íon cíclico. Este passo representa a adição eletrófila pois o iodo transferido tem carga positiva. No segundo passo o íon cloreto formado pela quebra da molécula de ICl reage com o íon cíclico formado, dando origem ao produto da reação.

O íon cíclico formado é capaz de reagir não só com o íon cloreto, mas com outros íons e mesmo água se esta estiver presente, daí a necessidade do material volumétrico estar seco no início do processo.

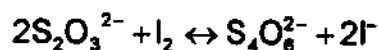
4.5- O método padrão :

O método padrão (AOCS) para a determinação do nível total de insaturação dos óleos vegetais baseia-se no método indireto de titulação do iodo.

Primeiramente faz-se a dissolução dos óleos vegetais com ciclohexano (solvente inerte) e posterior reação com a solução de Wijs (exemplificada na reação 1). Esta reação processa-se em recipiente fechado, no período de aproximadamente uma hora à temperatura ambiente e na ausência de luz. O reagente que não foi consumido na reação (1) em meio a iodeto de potássio libera iodo que é titulado então com solução padrão de tiosulfato de sódio 0,1 M.

A reação entre iodo e tiosulfato é a reação de oxi-redução básica da iodometria indireta. Em solução neutra ou levemente ácida (como é o caso), o íon tiosulfato é quantitativamente oxidado por iodo à íon tetratioato. A reação entre o tiosulfato e o iodo é mostrada na reação 2.

Reação 2



Como é típico das determinações indiretas, no início da reação, nenhum tipo de indicador é adicionado. A própria cor da solução (o iodo em solução aquosa de iodeto de potássio é de cor amarelo a marrom relativamente intensos) serve para indicar as mudanças (reações) que estão ocorrendo. Com a proximidade do ponto final, adiciona-se amido como indicador. O amido produz uma coloração azul com o iodo (ou melhor com o íon triiodeto), devido à adsorção do íon triiodeto no amido.

Em soluções ácidas e em presença de iodo, este indicador sofre hidrólise formando produtos que dão coloração vermelha ao amido " mascarando " o verdadeiro ponto final, deste modo sua adição é adiada ao máximo quando a concentração de iodo já é baixa [34]. O ponto final da titulação é visualizado pela descoloração da solução .

As determinações do índice de iodo pelo método titulométrico foram realizadas em triplicata, sendo usados, para fins de cálculos, os valores médios destas determinações.

A partir dos dados obtidos, aplicou-se o teste t para determinar em que intervalo se encontra a média com um nível de confiança de 95%. Neste caso, conhece-se a média (\bar{x}) e o desvio padrão (s), que são as estimativas de μ e σ . A equação utilizada para os cálculos do teste t foi a de número 21 que consta do capítulo 3.

Na tabela 3 são mostrados os resultados médios dos índices de iodo obtidos com nível de confiança de 95%, fazendo uso da análise titulométrica [35] .

Tabela 3- Resultados dos Índices de Iodo obtidos pela Análise Titulométrica.

TIPO DE AMOSTRA	RESULTADOS ÍNDICE DE IODO	LIMITE CONF. 95%
1 Óleo de arroz	101	3
2 Azete de Dendê	56	3
3 Óleo de Canola	110	4
4 Óleo de Gergelim	111	3
5 Óleo de girassol	118	1
6 Óleo de milho	114	3
7 Mist11 (óleo de palma 25% e óleo de soja 75%)	79	3
8 Mist21 (óleo de palma 70% e óleo de soja 30%)	73	4
9 Mist31 (óleo de palma 57,5% e óleo de soja 42,5%)	85	3
10 Mist4 (óleo de palma 50% e óleo de soja 50%)	88	4
11 Mist5 (óleo de palma 43,8% e óleo de soja 56,3%)	95	0
12 Mist6 (óleo de palma 47,5% e óleo de soja 52,5%)	91	1
13 Mist7 (óleo de palma 97% e óleo de soja 3%)	57	2
14 Mist8 (óleo de palma 90% e óleo de palmiste 10%)	48	2
15 Mminas (óleo de milho)	116	0
16 Olgr (óleo de girassol 70% e óleo de oliva 30%)	120	1
17 Óleo de oliva	88	4
18 Olive(óleo de oliva)	84	1
19 Olpur (óleo de oliva)	87	2
20 Óleo de Palma	53	1
21 Óleo de Palmiste	17	0
22 Sjoj (óleo de soja)	121	4
23 Soliv (óleo de soja 85% e óleo de oliva 15%)	121	1
24 Sperd (óleo de soja)	130	1

4.6- Organização de dados para a Análise Multivariada

Das vinte e quatro amostras de óleos vegetais, dezoito foram selecionadas para a construção do modelo de calibração e seis para a verificação de sua validade.

Para a fase de calibração, os dados instrumentais, obtidos a partir dos espectros Raman, foram dispostos na forma de uma matriz 18X1000 que corresponde à matriz de variáveis independentes (matriz X). As colunas desta matriz são formadas pelos valores das intensidades em cada número de onda do espectro (faixa de 3000 a 1000 cm^{-1} , resolução de 4 cm^{-1}) e o número de linhas é determinado pelo total de amostras selecionadas para esta fase. Uma outra matriz 18X1 é formada com os valores dos índices de iodo, correspondentes às amostras da matriz X, obtidos na análise titulométrica. Esta matriz 18X1 corresponde à matriz de variáveis dependentes (matriz Y).

Para a fase de previsão, os dados dos outros seis espectros Raman, não utilizados na fase de calibração, foram dispostos na forma de uma matriz 6x1000 correspondente à matriz de variáveis independentes (matriz X_{teste}). Novamente aqui tem-se que as colunas da matriz são formadas pelos valores das intensidades em cada número de onda do espectro e o número de linhas correspondem ao número de amostras utilizadas para esta fase. A partir destes dados foram obtidos os valores dos índices de iodo para estas amostras.

As amostras escolhidas para a fase de previsão são aquelas que apresentam valores de índices de iodo abrangendo toda a faixa de índices estudados e que são os mais distintos possíveis.

4.7-Resultados e Discussões da Análise Multivariada

4.8 - Método dos Mínimos Quadrados Parciais

Depois de organizados, aplicou-se o método dos mínimos quadrados parciais (PLS) a estes dados para a construção do modelo de calibração. Em virtude da não aplicação de nenhum pré-tratamento, os resultados obtidos não foram considerados bons. Isto mais uma vez, evidencia a necessidade de pré-tratamentos a dados obtidos através dos espectros Raman, pois de outra forma seria impossível obter qualquer informação quantitativa.

As variáveis independentes passaram, então, por pré-tratamentos matemáticos. Dentre todos os tipos considerados, a normalização média (para a correção no valor das intensidades) e também centrar na média (ou seja todas as variáveis ficaram com média zero) foram os pré-tratamentos a partir dos quais melhores resultados foram obtidos.

4.9- A Normalização

A normalização média foi obtida dividindo-se cada linha da matriz de variáveis independentes pela sua respectiva média (equação 39 - capítulo3), neutralizando assim a influência de fatores irrelevantes ao processo. Com isso, somente os valores relativos das variáveis são usados para descrever as amostras o que permite que as informações específicas sejam encontradas.

Esta normalização é aplicada principalmente em casos onde todas as variáveis tenham as mesmas unidades e sejam positivas, como os dados deste trabalho .Quando da aplicação em casos onde todos os valores sejam negativos deve-se dividir cada variável pelo valor absoluto da média ao invés da média pura. Então o sinal original é mantido .

Uma propriedade importante dos dados das amostras normalizadas pela média é que a área sob a curva, ou seja a área sob os espectros Raman, vem a ser a mesma para todas as amostras.

4.10- Centrar na Média

Centrar na média consiste em transladar o eixo de coordenadas para um centro. Os dados dessa maneira ficam melhor condicionados. Sem este procedimento a direção de maior variância descrita pela primeira variável latente seria a direção média dos dados. A equação 40 (capítulo 3) mostra como os dados foram centrados na média.

4.11-Outros tipos de pré-Tratamentos considerados.

Pela figura 11 (Parte Experimental) verifica-se que os espectros Raman não apresentam desvios de linha base que possam ser visualizados. A aplicação de uma correção para desvios de inclinação usando MSC (correção de espalhamento multiplicativo) [32] foi, contudo, realizada. Nenhuma mudança nos resultados foi alcançada. Dessa forma, esse último pré-tratamento não foi considerado.

Para a correção do valor das intensidades nos espectros Raman, além da normalização média, estudou-se a possibilidade de aplicação de outros tipos de normalizações, como normalização para área unitária, normalização pelo pico de maior intensidade, bem como da aplicação de derivadas. Estas correções não apresentaram bons resultados. O mesmo acontecendo com o uso de filtros para eliminação de ruídos.

Isto tudo foi observado nos resultados obtidos na fase de previsão, construindo modelos com estes diferentes pré-tratamentos e analisando os resultados obtidos através do cálculo do erro relativo e do RMSEP, utilizados como critérios para avaliação do modelo construído.

4.12- A construção e validação do modelo de calibração (PLS)

Depois de organizar os dados na forma de matrizes: Matriz X (variáveis independentes) e matriz Y (variáveis dependentes) e submeter a matriz X aos pré -

tratamentos : normalização média e centrar na média, pode-se construir o modelo de calibração multivariada baseada nos mínimos quadrados parciais, utilizando as dezoito amostras selecionadas para esta fase (calibração). Este modelo foi construído utilizando-se o programa computacional UNSCRAMBLER 6.0. .

Para a fase de previsão, os dados dos espectros Raman das outras seis amostras organizados também na forma de matriz (matriz X_{teste}) e da mesma forma que os anteriores, estes dados passaram pelos mesmos pré-tratamentos. A seguir, utilizando o modelo PLS desenvolvido na fase de calibração pode-se encontrar os valores dos índices de iodo destas novas amostras. Estes índices são previamente conhecidos (análise titulométrica) e com isso, empregando-se novamente os critérios de avaliação, ou seja os cálculos de erros relativos e RMSEP, pode-se selecionar o melhor modelo.

O Erro Relativo é dado por:

$$Er = ((y_{real} - y_{prev}) / y_{real}) \cdot 100 \quad (49)$$

Sendo o RMSEP:

$$RMSEP = \sqrt{\frac{\sum (y_{real} - y_{prev})^2}{n}} \quad (50)$$

onde " y_{prev} " é o valor do índice de iodo previsto pelo modelo, " y_{real} " é o valor do índice obtido pela análise titulométrica e " n " corresponde ao número total de amostras utilizadas na previsão.

4.13- O número ideal de variáveis latentes utilizados para a construção do modelo.

Para a construção do modelo PLS é essencial encontrar o número ideal de variáveis latentes que devem ser incluídos no modelo. Este número foi encontrado utilizando o procedimento da validação cruzada completa , cujo gráfico é mostrado na figura 12. É possível verificar através deste gráfico que três variáveis latentes (mínimo do gráfico) são necessárias para a construção do modelo de calibração, pois a partir desse valor não se tem mais uma variância residual significativa, ou seja, a maior parte da informação estatística disponível nos dados já foi " capturada ". Se acaso mais variáveis latentes fossem incorporadas ao modelo ocorreria a inclusão de ruídos. Porém no caso de se

usar um número de variáveis latentes menor do que três informações importantes deixariam de ser utilizadas.

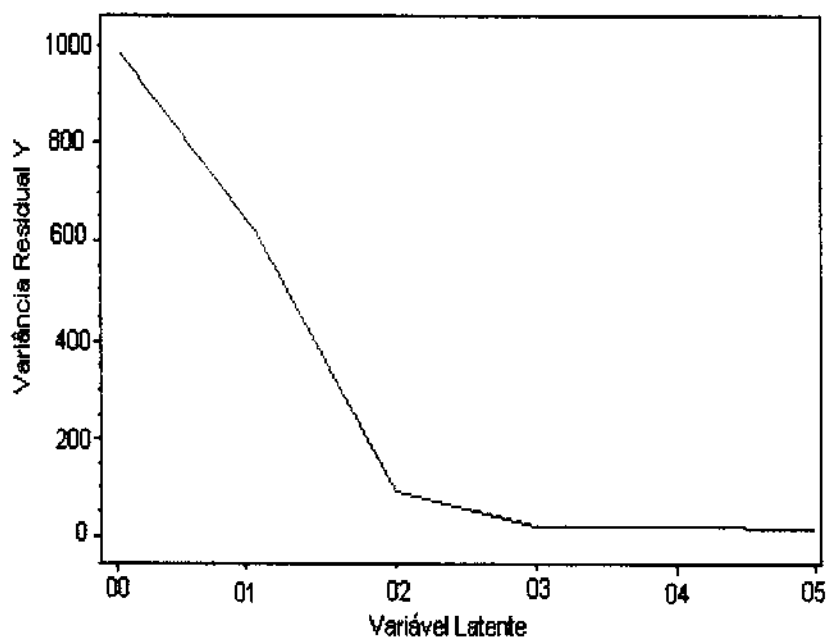


Fig.12-Gráfico de Variáveis Latentes

O gráfico apresentado na figura 12 mostra a variável latente 0. O significado deste número pode ser entendido como o resultado da variância residual média inicial de Y, ou seja, quando não se realizou ainda o procedimento da validação cruzada completa.

Os resultados obtidos utilizando validação cruzada podem ser expostos também na forma de uma tabela. A tabela 4 mostra a variância explicada de cada variável latente para as variáveis independentes (X) e também para as variáveis dependentes (Y) e não só, como no gráfico da figura 12, para variáveis dependentes (Y).

Tabela 4 – Variância Explicada em cada Variável Latente .

Variável Latente	% Variância em X	% Variância em Y
1	22,05	79,28
2	81,67	95,31
3	85,67	98,95
4	94,75	99,25

Pela tabela 4 verifica-se que as três primeiras variáveis latentes explicam mais de 85% da variância dos dados para as variáveis independentes (X) e mais de 98% da variância dos dados das variáveis dependentes (Y) . Usando a tabela 4, pode-se concluir, mais uma vez, que três é o número ideal de variáveis latentes necessárias para "capturar" a maior parte das informações estatísticas presentes nos dados.

4.14 – Identificação de amostras que não pertencem ao modelo.

Os gráficos apresentados nas Figuras 13 e 14 são usados para identificar possíveis anomalias no grupo de amostras utilizado para a construção do modelo de calibração.

Pela análise do gráfico da Figura 13 , " Leverage ", pode-se verificar que as amostras 2 (azeite de dendê), 9 (mistura 31), 14 (mistura 8) e 21 (óleo de palmiste) apresentam altas "leverages". Amostras com altas "leverages" são amostras diferentes das demais e com isso poderiam representar anomalias sendo assim prejudiciais ao modelo de calibração. Ao contrário, estas amostras poderiam ser informativas e portanto importantes para a construção do modelo de calibração. Dessa forma, estudos complementares foram feitos para verificar se estas amostras eram ou não anomalias. Isso foi feito através da análise do gráfico 14.

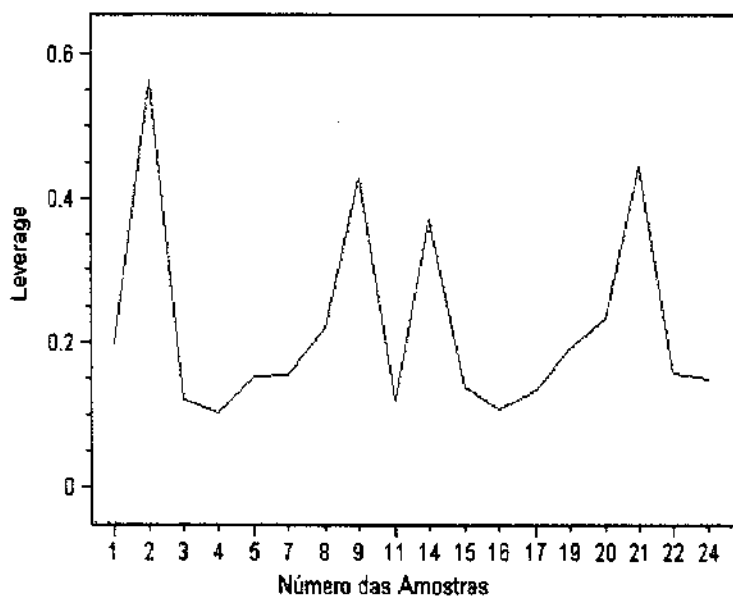


Figura 13-Gráfico das "Leverages"

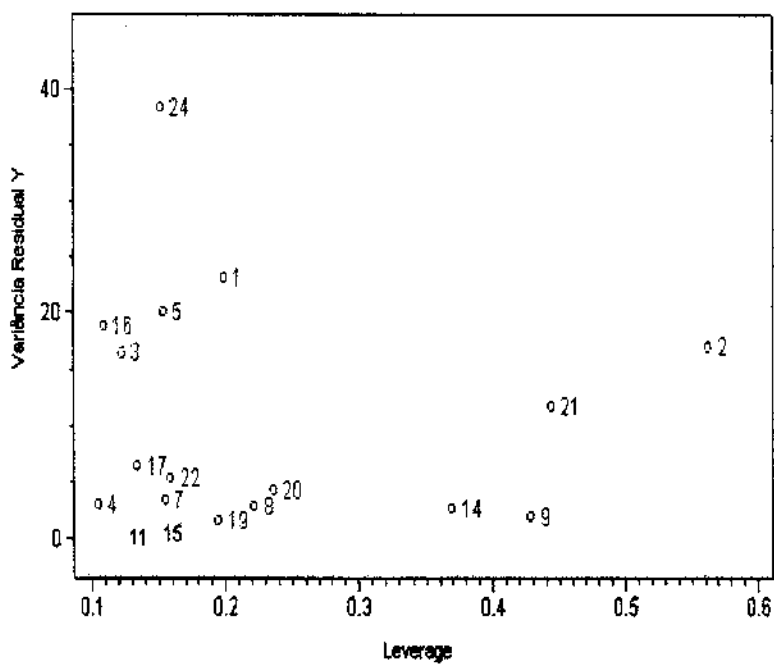


Figura 14-Gráfico de Influência

A figura 14 corresponde ao gráfico de Influência que mostra na ordenada a variância residual Y e na abcissa as "leverages". Pela observação do gráfico, pode-se verificar que a amostra 2 (óleo de dendê) embora apresente alta leverage têm valor de resíduo baixo , o mesmo acontecendo com a amostra 21(óleo de palmiste), a amostra 9 (mistura 31) e a amostra 14 (mistura 8). No capítulo 3 foi dito que as amostras que apresentassem respectivamente altas leverages e altos resíduos correspondem à amostras anômalas. Isto porque elas são diferentes das demais (indicação dada pela análise de leverage) e os valores previstos pelo modelo e seus valores reais estão muito discordantes. Assim sendo, pode-se concluir que não existe nos dados de calibração nenhuma anomalia.

Analisando os dados pode-se classificar as amostras como informativas, ou seja, elas são diferentes das demais amostras: as amostras 2,14 e 21 apresentam baixos índices de iodo ao contrário das demais. Porém, neste caso, em se tratando de um modelo linear (caso exemplificado no capítulo 3), estas amostras conseguem fornecer informações adicionais ao conjunto, melhorando a modelagem. A amostra 9 também não apresenta influência negativa ao modelo.

A amostra 24 (óleo de soja) embora apresente um resíduo um pouco mais alto também não mostra influência negativa em relação ao modelo.

Todas estas observações podem ser comprovadas através do cálculo dos erros relativos e do RMSEP. Construindo o modelo sem a amostra 2 o valor do RMSEP será de 3 (3 também é o resultado do modelo adotado- Tabela 5, que apresenta os resultados finais obtidos) contudo o erro relativo da amostra 10 (utilizada na previsão) irá aumentar. Retirando a amostra 21 ou a amostra 24 o RMSEP será de 4. Retirando as amostras 2,21,24 todas de uma vez, o valor do RMSEP sobe para 5. O mesmo acontece quando se retira do modelo as amostras 9 e/ou 14.

Depois de construído o modelo de calibração baseado nos mínimos quadrados parciais (PLS) utilizando três latentes e verificar se o modelo construído apresentava ou não amostras anômalas, fez-se outra análise: Construiu-se um novo modelo (PLS)

baseado na seleção de variáveis. Estas variáveis são aquelas pertencentes às regiões dos espectros Raman cujos coeficientes de regressão têm maior importância.

4.15 - Construção de modelos baseados na seleção de variáveis

Pela análise da figura 15, o gráfico que apresenta os coeficientes de regressão, pode-se constatar que estes coeficientes refletem basicamente os espectros Raman, onde os mais importantes (aqueles que apresentam maiores valores absolutos) correspondem às bandas espectrais correspondentes ao estiramento das duplas ligações C=C (1661 cm^{-1}), deformação da ligação CH₂ (1439 cm^{-1}), bem como ao estiramento assimétricos -C-H (2926 cm^{-1}) e outros, conforme mostra a tabela 1 (capítulo 2).

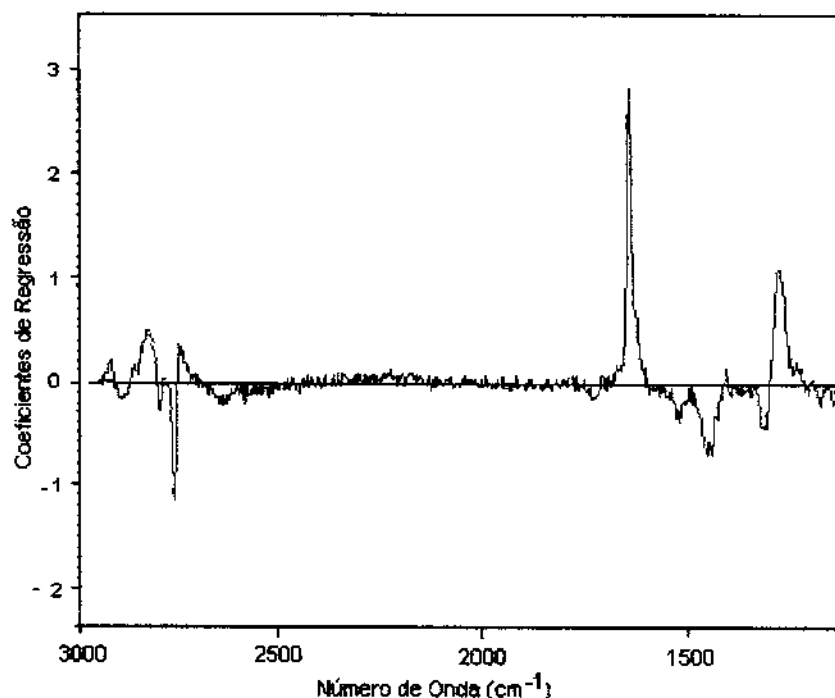


Fig 15 - Gráfico dos coeficientes de Regressão.

No entanto, o modelo construído com a seleção destas regiões não apresentou melhoras significativas em relação ao estudo realizado com todo o espectro. Por este motivo optou-se por fazer o modelo com todas as variáveis dos espectros Raman.

Depois destes estudos conseguiu-se chegar a um bom modelo de calibração multivariada usando o método de mínimos quadrados parciais (PLS).

O gráfico da figura 16 mostra a relação entre o valor previsto do índice de iodo dos óleos vegetais pelo modelo construído e o valor medido (observado), através da análise titulométrica. O gráfico mostra também os desvios destas medidas.

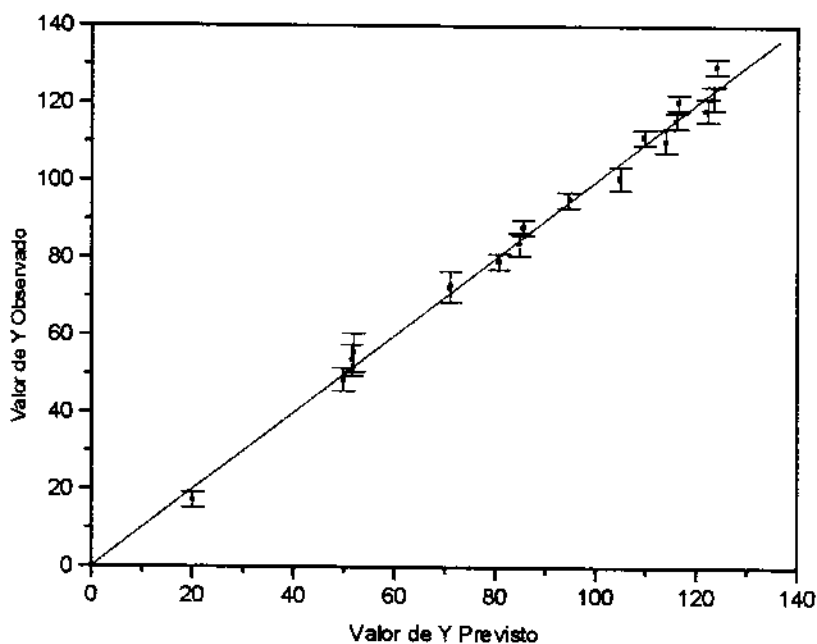


Figura 16 - Índice de Iodo Observado versus Índice de Iodo Previsto

O gráfico apresentou um coeficiente de correlação de 0.996, demonstrando que existe boa concordância entre os dados obtidos pela análise titulométrica e os previstos pelo modelo.

A tabela 5 mostra os resultados finais obtidos com o modelo PLS construído com três variáveis latentes. Este modelo corresponde aos melhores resultados obtidos depois de uma série de modelos construídos. O critério adotado para a escolha deste modelo foi baseado no erro relativo e no RMSEP, como dito anteriormente.

CAPÍTULO V

*O Emprego de Cartas de Qualidade Multivariadas na
Caracterização do Processo Oxidativo de Óleos
Vegetais*

PARTE EXPERIMENTAL

5- Parte Experimental

5.1- Caracterização do Processo Oxidativo dos Óleos Vegetais.

Nesta segunda parte do trabalho procurou-se identificar o processo oxidativo do óleo de soja. Um total de trinta e oito amostras de óleos vegetais, sendo trinta e sete delas correspondentes à óleos de soja e uma correspondendo ao óleo de arroz, foram utilizadas neste estudo. No caso do óleo de soja, quatro das amostras originais foram submetidas a aquecimento em chapa elétrica em um béquer aberto para que o processo de oxidação pelo oxigênio do ar em conjunto com o aumento da temperatura desse origem aos produtos de oxidação.

Os óleos foram oxidados da seguinte maneira:

Óleo 1 : Temperatura 120 °C durante 1 hora (marca A).

Óleo 2 : Temperatura 120 °C durante 1 hora (marca B) .

Óleo 3 : Temperatura 90 °C durante 4 horas (marca C).

Óleo 4 : Temperatura 200 °C durante 40 minutos (marca A).

Uma outra amostra utilizada para fritura foi analisada, neste caso o óleo foi usado por três dias consecutivos em alta temperatura. Uma última amostra de óleo de soja que foi guardada em geladeira por oito meses, depois da embalagem ter sido aberta, também foi utilizada.

Os óleos de soja foram adquiridos em um supermercado local, obedecendo como critério de escolha as diferentes marcas e lotes distintos para que houvesse uma certa heterogeneidade (dentro da variabilidade comum) entre eles.

Os espectros no infravermelho médio dos óleos vegetais foram obtidos em um espectrofotômetro Nicolet 520 FTIR (Nicolet Spectrometer). Para minimizar interferências como vapor de água e o gás carbônico, o equipamento foi purgado com nitrogênio.

Para as análises foi utilizado o acessório de reflectância total atenuada horizontal (HATR) equipado com cristal de seleneto de zinco. Para todas as amostras foram realizadas 254 varreduras com resolução de 4cm^{-1} .

Para a obtenção dos espectros, 1 mL dos óleos puros foram pipetados sobre o cristal de HATR, sendo utilizado o ar como referência.

A faixa espectral selecionada para a construção das cartas de qualidade multivariadas, dentro a faixa total de 4000 a 600 cm^{-1} foi : 3100 a 2400 cm^{-1} e 1900 a 600 cm^{-1} correspondentes às bandas significativas dos espectros para os óleos vegetais. O cristal de HATR foi limpo com diclorometano e acetona.

As figuras 17 e 18 mostram regiões dos espectros no infravermelho médio dos óleos vegetais selecionadas para a construção das cartas de controle de qualidade.

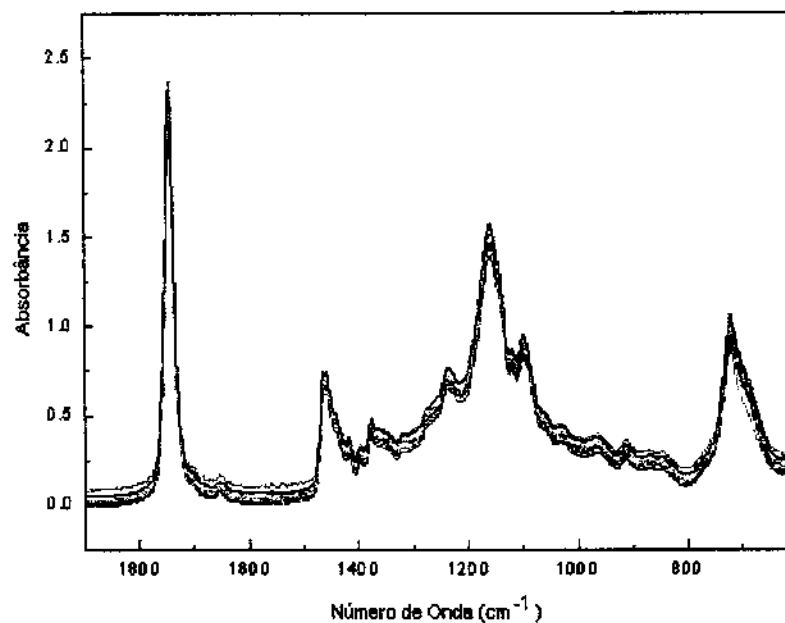


Figura 17 - Região do infravermelho de 1900 a 600 cm^{-1} selecionada para a construção das cartas de controle de qualidade.

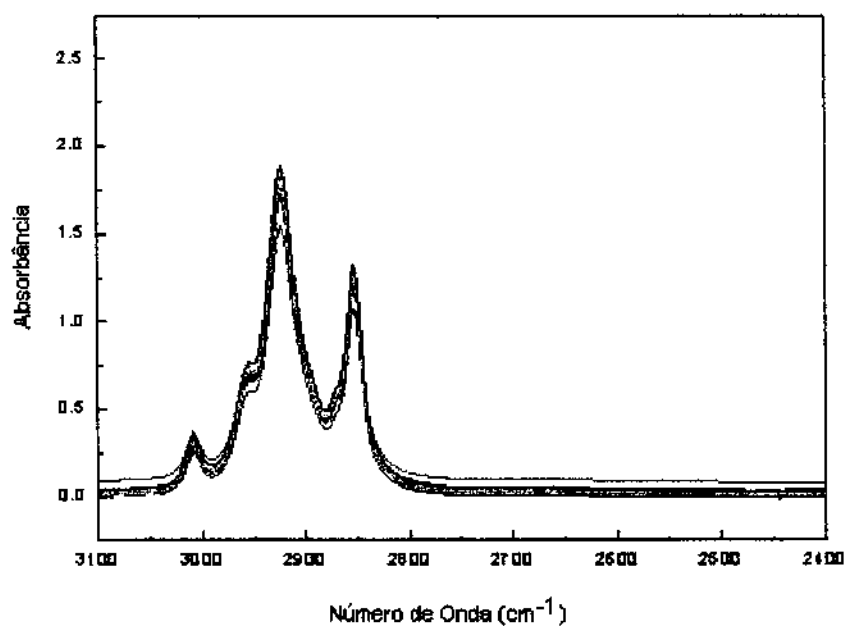


Figura 18 - Região do infravermelho de 3100 a 2400 cm^{-1} selecionada para a construção das cartas de controle de qualidade.

5.2-Resultados e Discussões :

5.2.1- Cartas de Controle de Qualidade Multivariadas.

Das trinta e oito amostras de óleos vegetais, vinte e três foram utilizadas para a construção das cartas multivariadas. Os óleos de soja escolhidos nesta fase foram aqueles que não passaram por nenhum processo oxidativo. As cartas foram construídas selecionando variáveis pertencentes às regiões de bandas representativas de um óleo vegetal, originando uma matriz 23X590 (então, as colunas desta matriz são formadas pelos valores das absorbâncias em cada número de onda selecionado e suas linhas correspondem ao número de amostras utilizadas nesta fase) .

Os dados dos espectros no infravermelho dos óleos vegetais foram primeiramente centrados na média e o modelo construído (ou seja as cartas multivariadas), foi baseado na análise de componentes principais (PCA). Para a construção das cartas multivariadas foi utilizado o programa Matlab 4.0 for Windows.

O primeiro passo para a construção destas cartas consiste em encontrar o número ideal de componentes principais que devem ser utilizados em cada carta. As cartas T^2 baseiam-se nas componentes principais significativas enquanto que as cartas de Q utilizam as componentes principais remanescentes. O número de componentes principais foi determinado através de um gráfico que mostra em um eixo os autovalores (ou o quanto de variância explica cada componente principal) contra as componentes principais. O mínimo deste gráfico indica o número de componentes principais significativos. Pelo gráfico apresentado na figura 19, verifica-se que a partir de três componentes principais não existe mais uma variação significativa no autovalor .Pelo fato de que quatro componentes situa-se também próximo ao mínimo da curva, alguns estudos foram feitos com a utilização de quatro componentes principais.

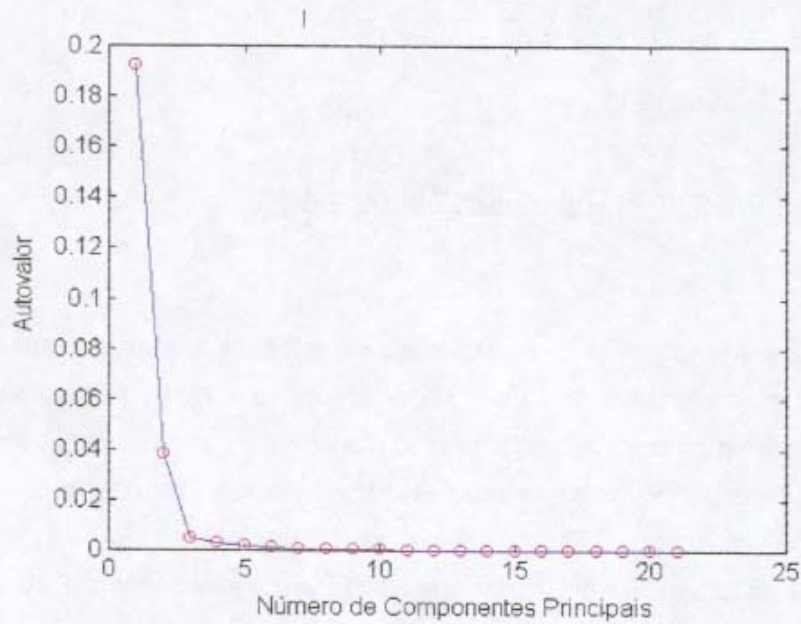


Figura 19. Gráfico de Autovalores versus Número de Componentes Principais.

A Tabela 6 mostra a porcentagem de variância retida em cada uma das componentes principais.

Tabela 6 - Variância Parcial e Total Explicada em cada Componente Principal

Componente Principal	Autovalores Matriz Cov (X)	% Variância Explicada	% Variância Total
1	1,93 e-001	79,11	79,11
2	3,80 e-002	15,60	94,71
3	4,43 e-003	1,82	96,53
4	2,73 e-003	1,12	97,65
5	1,65 e-003	0,68	98,33

Pela Tabela nota-se que não existe muita diferença na variância explicada entre a terceira e a quarta componentes principais. A utilização de três componentes principais permite que o processo estudado fique menos restrito (ou seja não ocorre neste caso um super-ajuste do modelo) e por esse motivo adotou-se três componentes para a construção das cartas T^2 e as restantes para a construção das cartas Q. Do teste comparativo usando quatro componentes principais realizado algumas conclusões puderam ser obtidas. Essa comparação será mostrada mais adiante.

Depois de escolhido o número ideal de componentes principais, construiu-se as cartas T^2 e Q. Para os cálculos de T^2 e Q, bem como para os limites de controle das cartas (com 95% de confiança) foram utilizadas as equações que constam do capítulo 3. Nos cálculos dos limites de controle das cartas T^2 e Q, os dados devem ser livres de situações anormais (ou seja todos os valores de T^2 e Q das amostras devem estar dentro da faixa de controle) e, por esta razão, duas amostras foram retiradas do modelo e os limites tiveram de ser recalculados.

Os gráficos apresentados nas Figuras 20 e 21 mostram que todas as amostras utilizadas na construção do modelo estão dentro da faixa de controle estabelecidas pelos limites. O limite de controle de T^2 (95 % confiança) calculado é de 10,5 e o limite para Q é 0,019.

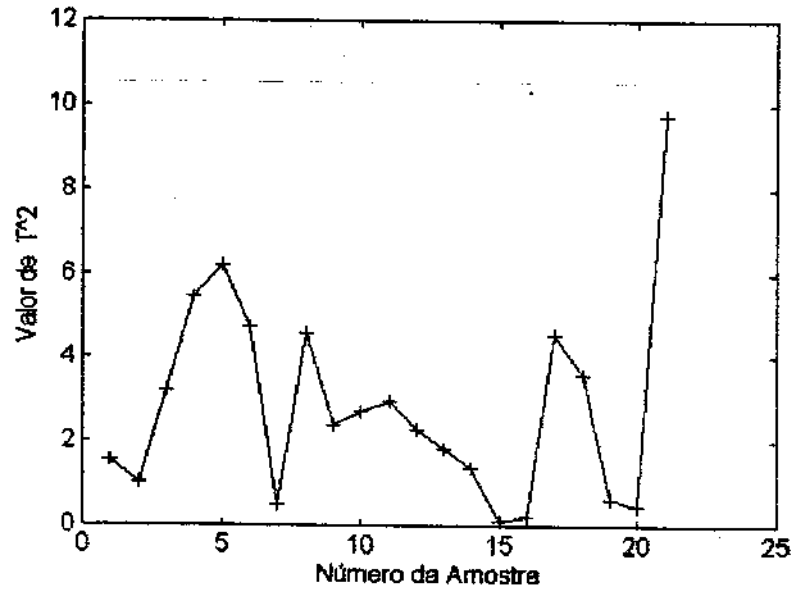


Figura 20- Gráfico do Limite de T^2 com 95% de confiança para as amostras utilizadas na construção do modelo.

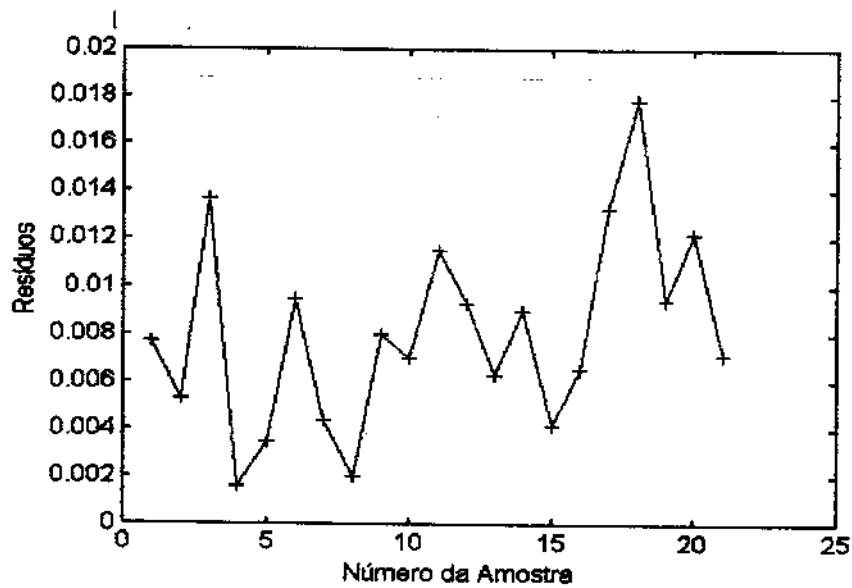


Figura 21 - Gráfico do Limite de Q com 95% de confiança para as amostras utilizadas na construção do modelo.

Os limites calculados serão utilizados a seguir para verificar se os quinze óleos restantes estão ou não dentro da faixa de controle estabelecida. Estes limites são calculados com base na variação normal que existe no conjunto de amostras e conseqüentemente qualquer amostra que estiver fora destes limites (acima da linha de controle dos gráficos) correspondem às amostras fora das normas de qualidade. Os óleos utilizados nesta fase foram àqueles submetidos à oxidação (6 ,7-utilizado para fritura ,8,9,11), ao óleo de arroz(5), ao óleo guardado em geladeira e também a alguns óleos que não sofreram nenhum a alteração (1,2,3,4,12,13,14,15).

Para a comparação dos novos dados com o modelo existente é necessário que estes passem por pré-tratamentos da mesma forma que os dados utilizados para a construção do modelo. Isto é feito, subtraindo o vetor médio (encontrado para os dados do modelo) dos novos dados, ou seja é necessário centrar os dados na média previamente estabelecida.

O próximo passo consiste em projetar estes novos dados nos "loadings" encontrados para o modelo construído e com isso obter os novos " scores" ,Q e T^2 .

Os valores de Q e T^2 podem ser comparados com os limites estabelecidos no modelo construído (cartas multivariadas). Os resultados dessa comparação podem ser visualizados através dos gráficos mostrados nas Figuras 22 e 23.

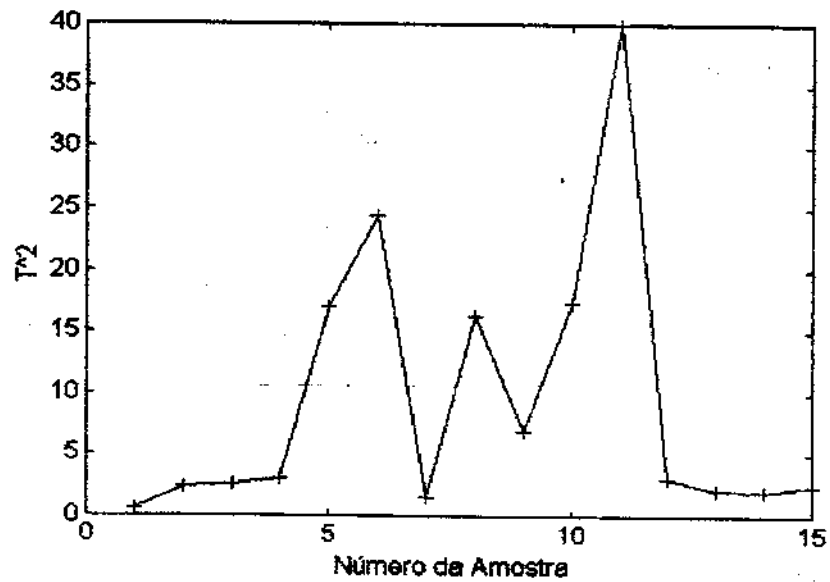


Figura 22 - Carta de Controle de Qualidade Multivariada T^2 contendo as quinze novas amostras.

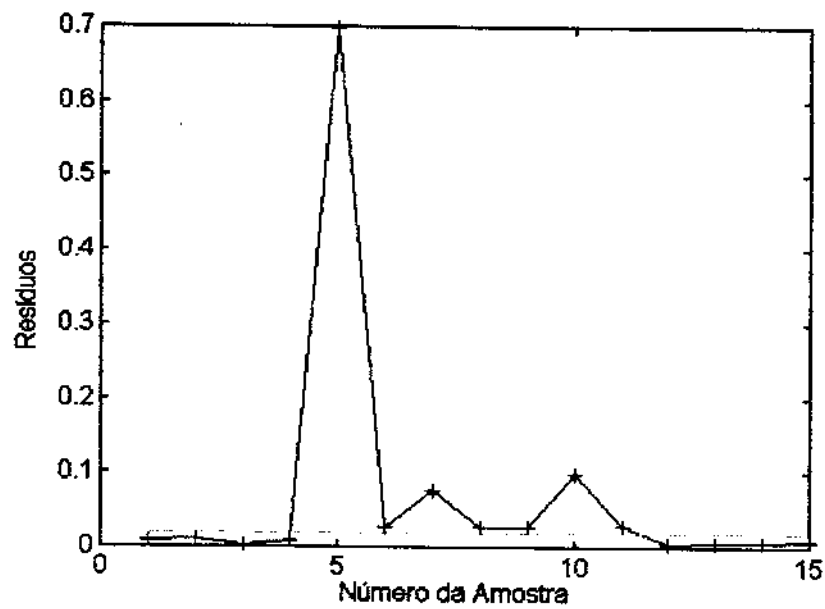


Figura 23 - Carta de Controle de Qualidade Multivariada Q contendo as quinze novas amostras.

Para que uma amostra seja considerada como fora de controle é necessário que ela esteja fora dos limites de controle em uma das duas cartas ou então em ambas.

Pelos gráficos mostrados nas figuras 22 e 23 observa-se que na carta de controle multivariada T^2 , as amostras 5,6,8,10,11 estão fora da faixa de controle (ou seja, elas se encontram acima da linha de controle estabelecida) e na carta Q, tem-se as amostras 5,6,7,8,9,10,11 fora da faixa de controle. As amostras de números cinco a onze correspondem justamente aquelas cinco submetidas a aquecimento (6,7 –utilizada para fritura ,8,9,11), ao óleo de arroz (5) e a amostra guardada em geladeira (10).

Em uma outra análise, o gráfico de "scores" também pode ser utilizado para verificar quais amostras são mais distintas do restante do conjunto.

Pelo gráfico dos " scores " das três componentes principais que explica 96,53 % da variância total dos dados (figura 24) nota-se que as amostras 5,6,8,9,10 e 11 encontram-se distintas das outras amostras do conjunto. Estas observações também podem ser visualizadas, usando dois gráficos em duas dimensões : dos "scores" da primeira componente principal contra a segunda componente principal e dos "scores" da primeira componente principal contra a terceira componente principal, mostrados nas Figuras 25 e 26.

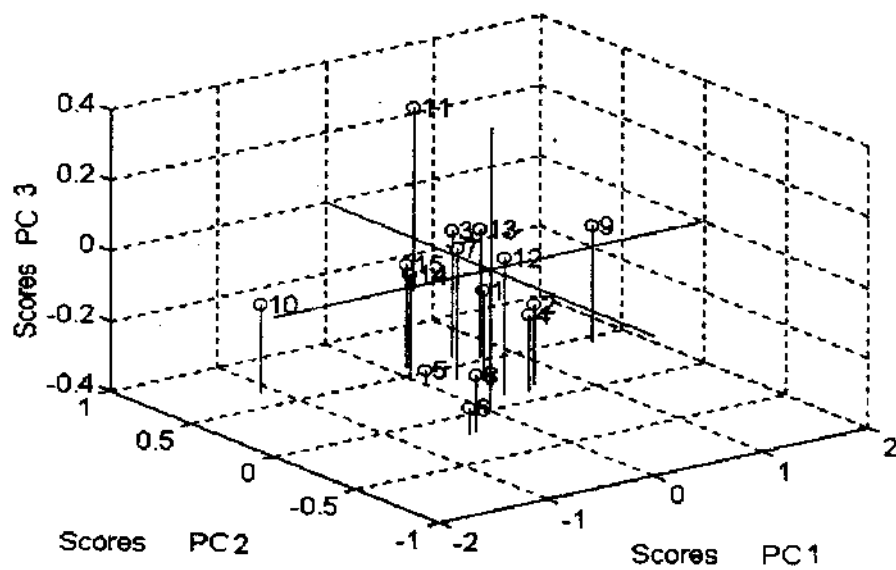


Figura 24- Gráfico dos "scores" da primeira componente principal com relação aos da segunda e terceira componentes principais.

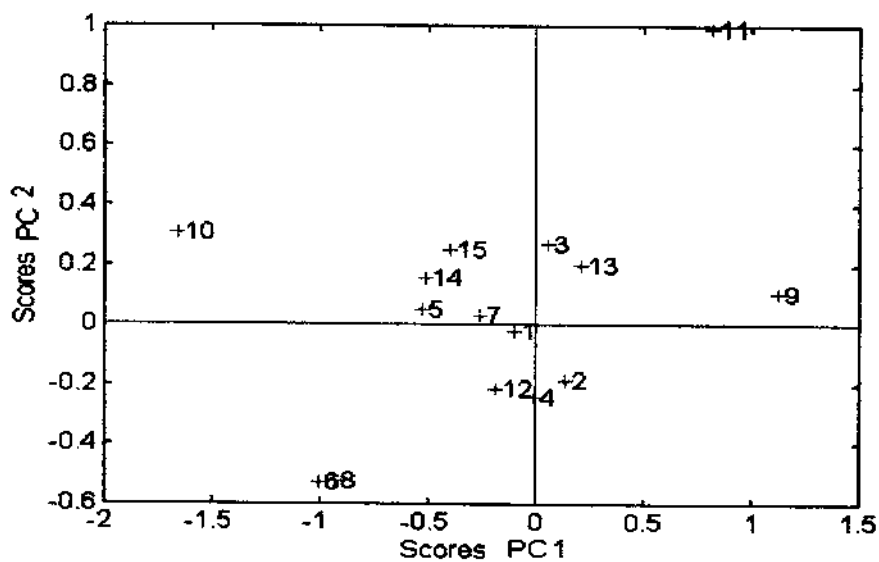


Figura 25- Gráfico dos "scores" da primeira componente principal com relação aos da segunda e terceira componentes principais.

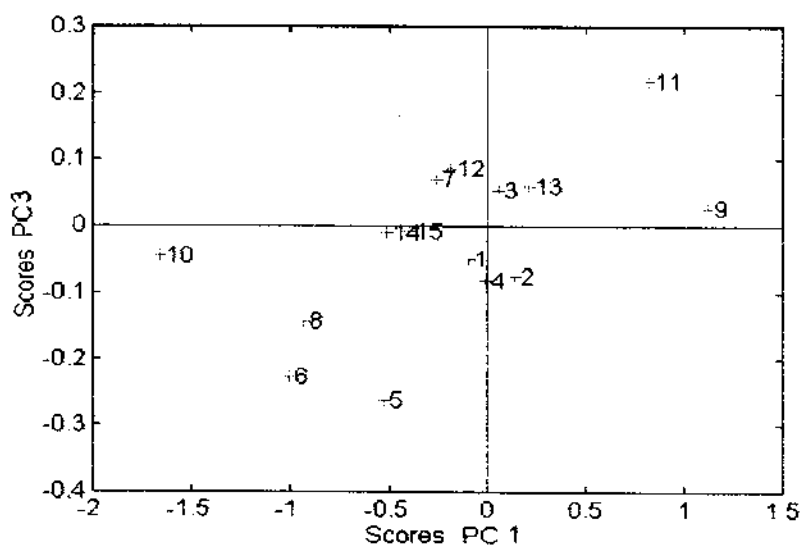


Figura 26 - Gráfico dos " scores " da primeira componente principal com relação aos da terceira componente principal.

Em nenhum dos gráficos de " scores " apresentados foi possível identificar a amostra de número 7 como distinta das amostras que não sofreram nenhuma oxidação. Porém, esta foi corretamente classificada como fora de controle pelas cartas multivariadas.

5.2.2- O Estudo com Quatro Componentes Principais.

As cartas de controle de qualidade multivariadas T^2 e Q utilizando quatro componentes principais estão apresentadas nas Figuras 27 e 28. Neste caso, os limites de controle, são : Para Q , 0.012 e para T^2 13,9.

Verifica-se através destas cartas que as situações fora de controle são as mesmas encontradas quando utiliza-se três componentes principais. Para este caso, no entanto, observa-se que tanto a carta T^2 quanto a carta Q apresentam as mesmas amostras 5,6,7,8,9,10 e 11 como fora de controle.

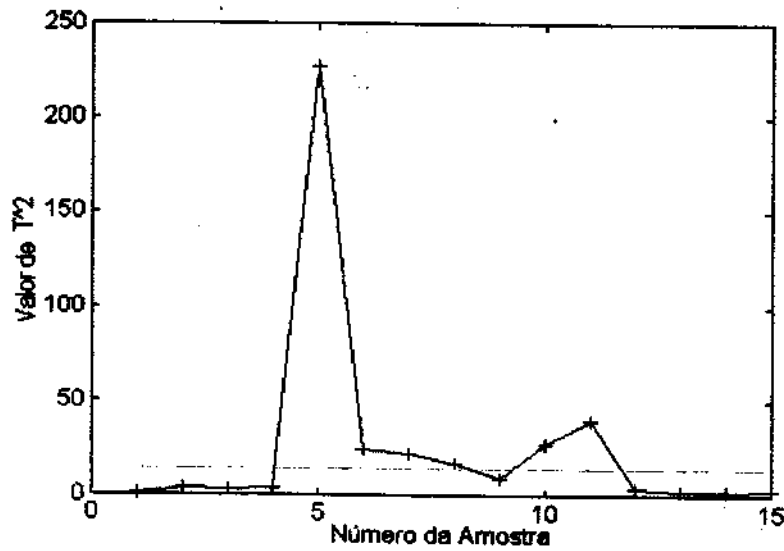


Figura 27 - Carta de Qualidade Multivariada T^2 usando quatro componentes principais.

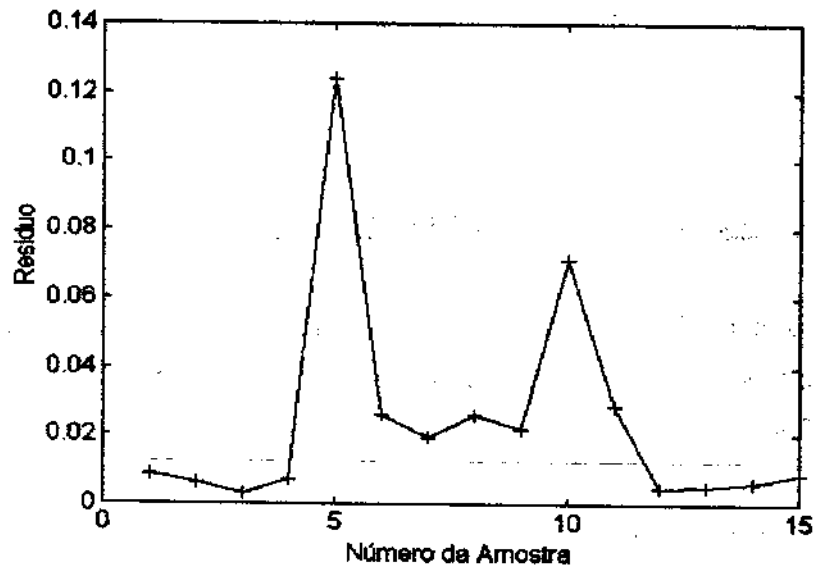


Figura 28 - Carta de Qualidade Multivariada Q usando quatro componentes principais.

A seguir são mostrados nas Figuras 29,30 e 31 os gráficos de "scores" para a primeira componente principal contra a quarta ,a segunda contra a quarta e a terceira contra a quarta.

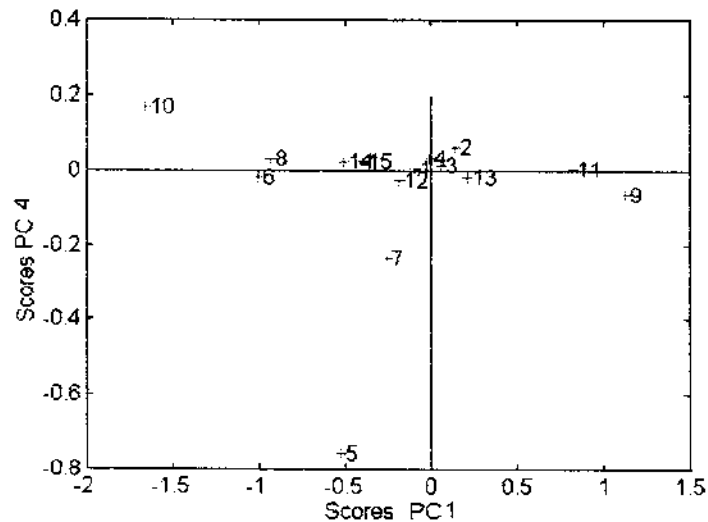


Figura 29 - Gráfico dos " scores " da primeira componente principal com relação aos da quarta componente principal .

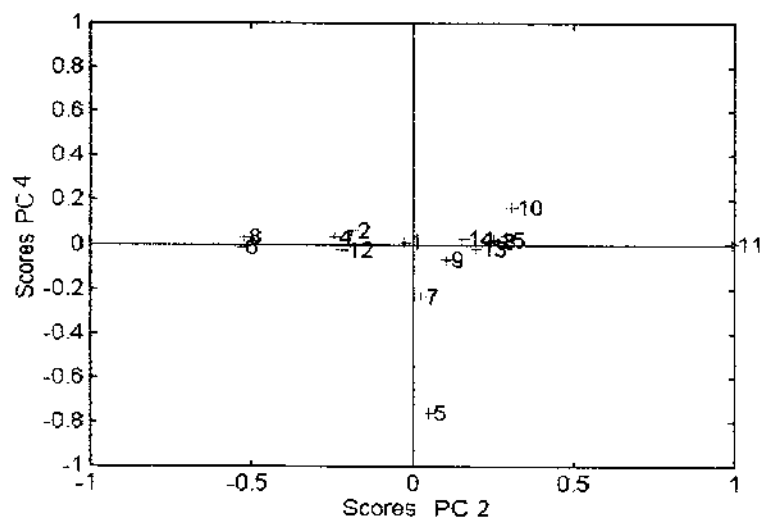


Figura 30 - Gráfico dos " scores " da segunda componente principal com relação aos da quarta componente principal .

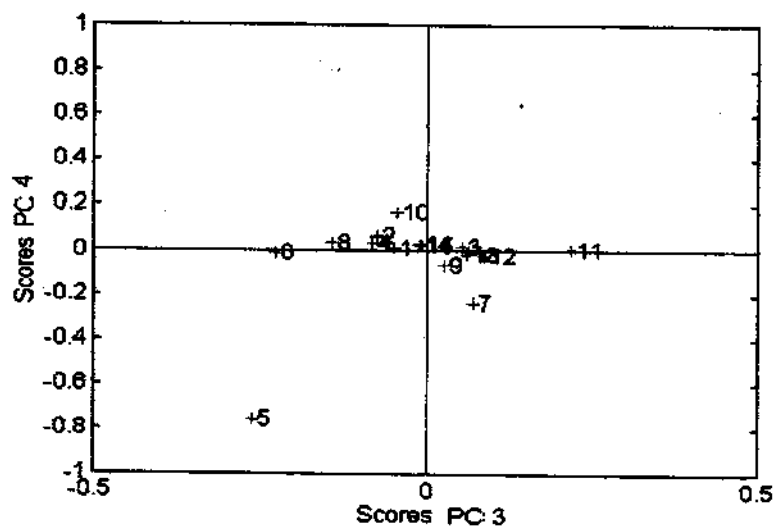


Figura 31 - Gráfico dos "scores" da terceira componente principal com relação aos da quarta componente principal .

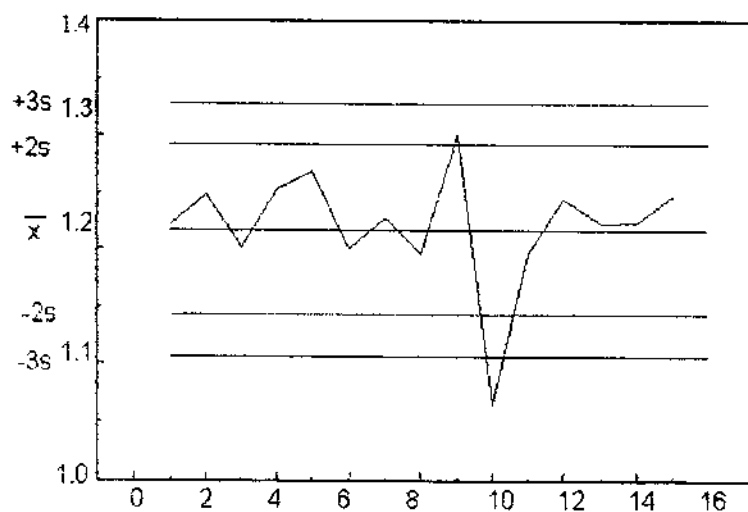
Com o gráfico dos "scores" e com o uso de quatro componentes principais foi possível identificar a amostra 7 como uma amostra distinta das demais. No caso da amostra 7 foi necessário uma componente a mais para conseguir capturar mais informações e assim distingui-la das demais.

5.3-Cartas de Controle Univariadas

Para efeitos comparativos foi feito um estudo construindo cartas univariadas. Para este estudo estatístico, cinco cartas de controle univariadas foram utilizadas. O controle das variáveis, neste caso, baseou-se na escolha das variáveis correspondentes aos picos de absorções características dos óleos vegetais nesta região do infravermelho (infravermelho médio).

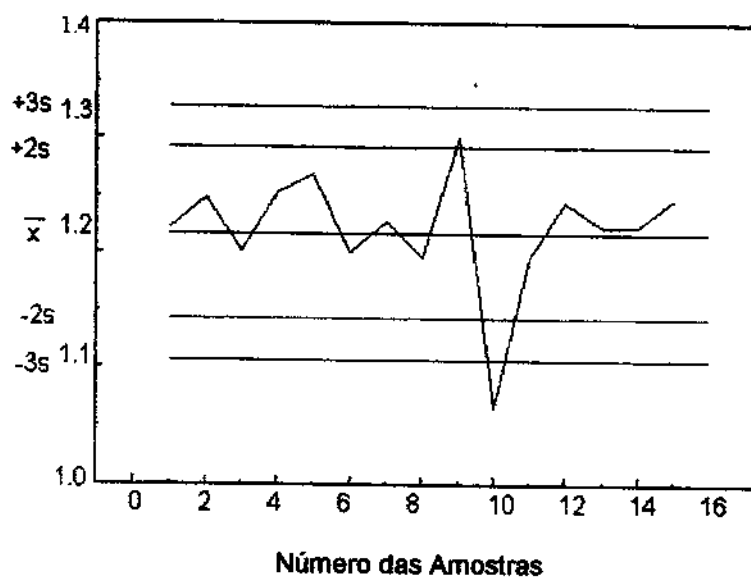
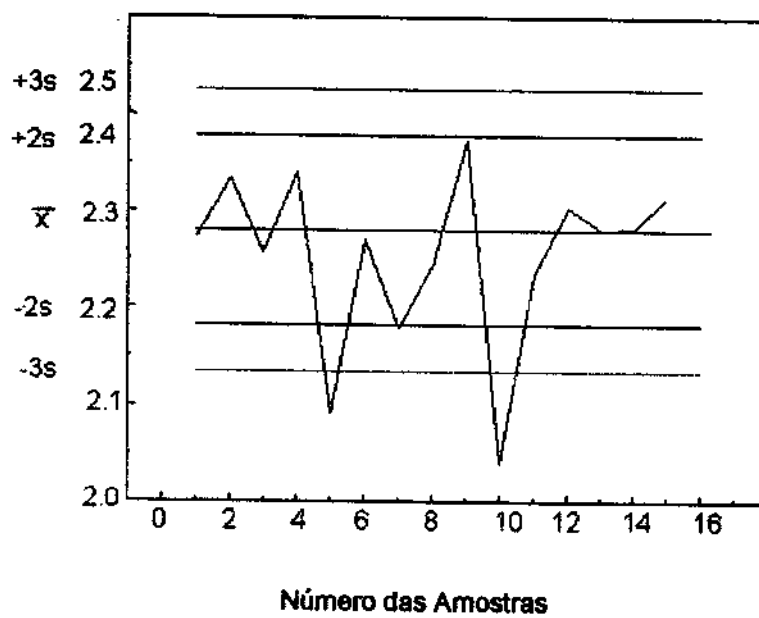
Os picos escolhidos foram: $2923,5 \text{ cm}^{-1}$, $2852,2 \text{ cm}^{-1}$, $1743,3 \text{ cm}^{-1}$, $1151,2 \text{ cm}^{-1}$ e $719,4 \text{ cm}^{-1}$.

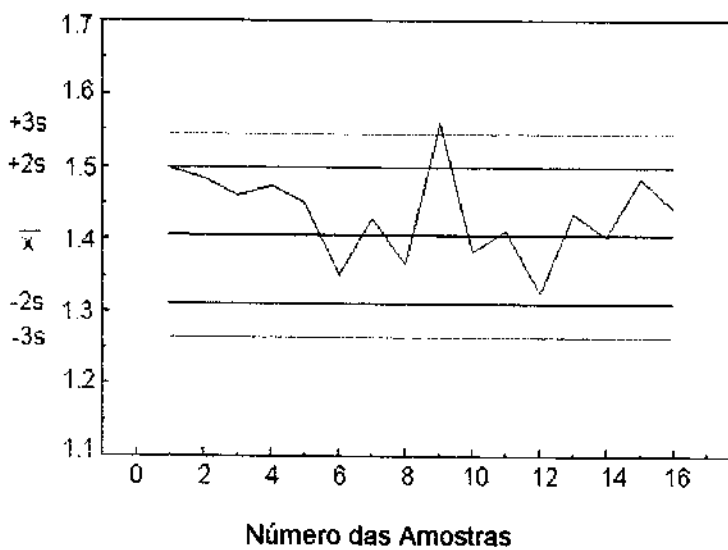
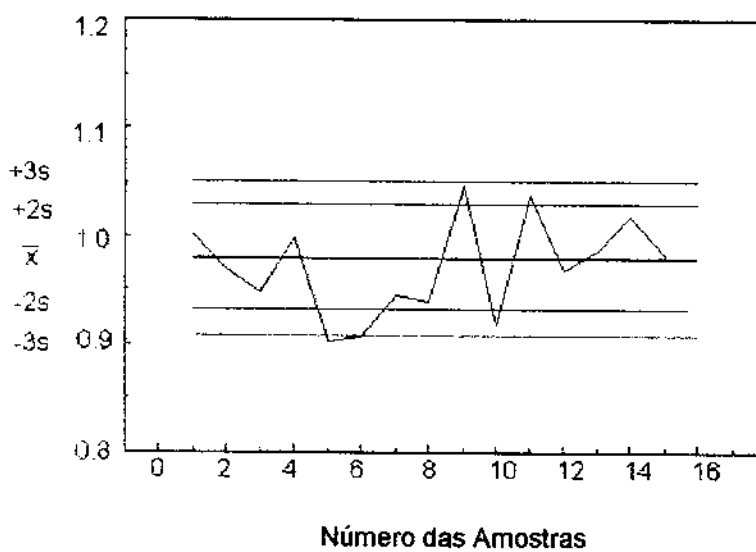
Primeiramente, utilizando as 21 amostras escolhidas inicialmente para o desenvolvimento do modelo os limites de controle foram calculados com base na média \bar{x} (altura média dos picos) e em duas vezes o desvio - padrão (95% de confiança). Os limites baseados na média \bar{x} e em três vezes o desvio padrão (99,7% de confiança) também foi usado. A seguir verificou-se quais das quinze amostras restantes estavam dentro da faixa de controle. As figuras de 32 a 36 mostram as cartas construídas baseadas nos cinco picos escolhidos.



Número da Amostra

Figura 32-Carta Univariada referente a pico em $2923,5 \text{ cm}^{-1}$

Figura 33- Carta Univariada referente a pico em $2852,2 \text{ cm}^{-1}$ Figura 34- Carta Univariada referente a pico em $1743,3 \text{ cm}^{-1}$

Figura 35- Carta Univariada referente a pico em $1151,2 \text{ cm}^{-1}$ Figura 36 - Carta Univariada referente á pico em $719,4 \text{ cm}^{-1}$

Na primeira carta univariada em $2923,5 \text{ cm}^{-1}$ observa-se que a amostra 10 está fora de controle. Na segunda em $2852,2 \text{ cm}^{-1}$ observa-se que as amostras 9 e 10 estão fora de controle. Na terceira carta em $1151,2 \text{ cm}^{-1}$ a amostra 9 está fora de controle. Na quarta carta em $1743,3 \text{ cm}^{-1}$ as amostras 5,7,9 e 10 se encontram fora de controle e na última carta as amostras 5,6,9,10 e 11. Para o caso univariado tem-se que somente com a junção de todas as cartas univariadas foi possível a identificação das situações fora de controle.

Com base nestes dados foi possível identificar as situações fora de controle com o uso das cartas univariadas e multivariadas. Uma das vantagens do modelo multivariado é que se deve analisar apenas duas cartas de controle de qualidade, enquanto no modelo univariado foi preciso de uma carta de controle para cada variável monitorada para se identificar as amostras fora de controle. Em ambos os casos, com exceção da amostra número oito, todas as outras amostras puderam ser identificadas. No caso da amostra oito a não observação do estado fora de controle nas cartas univariadas foi devido ao fato de que as análises das situações univariadas não leva em conta a estrutura de correlação entre os dados.

A espectroscopia no infravermelho tem sido usada na avaliação do ranço, desempenhando um papel importante no reconhecimento dos grupos funcionais não característicos e de compostos tendo duplas ligações com configuração trans.

Neste trabalho, procurou-se utilizar as informações fornecidas pelos espectros para identificar estes danos oxidativos, obtendo bons resultados na análise de óleos deteriorados em temperaturas de 90 a 200°C . A identificação de um óleo deteriorado pelo tempo (o óleo guardado em geladeira) também foi possível, além do óleo de arroz. Apesar disso, estudos utilizando temperaturas mais baixas de oxidação (40 a 80°C) devem ser realizados para se poder avaliar o potencial desta técnica nestes casos.

RESULTADOS E DISCUSSÕES

CONCLUSÕES

Conclusões

Foi possível demonstrar pelo trabalho realizado que a utilização de métodos espectroscópicos, como o infravermelho médio e Raman, em conjunção com métodos quimiométricos pode resultar em aplicações analíticas de grande potencial, tanto para determinações quantitativas como qualitativas.

A espectroscopia Raman aliada à calibração multivariada mostrou-se um bom instrumento para a determinação do índice de iodo dos óleos vegetais, comparável com os resultados obtidos pelo método titulométrico recomendado pela AOCS. A grande vantagem deste método espectroscópico, é que ele não utiliza reagentes tóxicos e é bastante rápido. Porém, ainda a instrumentação para espectroscopia Raman, devido ao seu custo, não é muito utilizada para fins analíticos, mas a tendência é que num futuro próximo comecem a aparecer instrumentos mais acessíveis e baratos.

Na segunda parte do trabalho utilizou-se as cartas de qualidade uni e multivariadas e espectroscopia no infravermelho para identificar o processo oxidativo dos óleos vegetais. As cartas mostraram-se perfeitamente capazes de identificar situações onde havia óleos deteriorados. As cartas multivariadas apresentam como vantagens o fato de que somente duas cartas são necessárias para identificar semelhantes situações. Deve-se destacar que muitas vezes em processos industriais não é tão necessário a quantificação das espécies, mas sim uma indicação qualitativa se o processo está, ou não, dentro dos padrões normalmente encontrados. Nestes casos, as cartas de qualidade multivariadas são importantes ferramentas para o controle destes processos.

REFERÊNCIAS BIBLIOGRÁFICAS

Referências Bibliográficas

1. Rohr R., *Óleos e Gorduras Vegetais e seus Subprodutos Protéicos*, Editora UNICAMP, Campinas, (1973).
2. Hartman L. e Esteves W., *Tecnologia de Óleos e Gorduras Vegetais*, São Paulo, (1982).
3. *Official Methods and Recommended Practices of the American Oil Chemists' Society*, editora Firestone, American Oil Chemists' Society, Champaign, (1989).
4. Baeten V., Meurens M., Morales M. T. e Aparicio R., "Detection of Virgin Olive Oil by Fourier Transform Raman Spectroscopy", *J. Agric. Food Chem.* **44** págs.2225-2230, (1996).
5. Ozaki Y., Chao R., Ikegaya K., Muraishi S. e Kawauchi K., "Potencial of Near-Infrared Fourier Transform Raman spectroscopy in Food Analysis", *Appl. Spectrosc.* **46** págs 1503-1507 ,(1992).
6. Adar F., Geiger R. e Nooman J., "Raman Spectroscopy for Process / Quality Control". *Appl. Spectrosc. Rev.* **32** págs 45-101, (1997).
7. Silva M. A. F., Borges M. F. M. e Ferreira A. M., "Métodos para Avaliação do Grau de Oxidação Lipídica e da Capacidade Antioxidante", *Química Nova* **22** págs 94-103, (1999).
8. Lai W. Y., Kemsley K., Wilson H. R., "Potential of Fourier Transform Infrared Spectroscopy for the Authentication of Vegetable Oils", *J.Agric.Food.Chem.* **42** págs 1154-1159, (1994).
9. Van de Voort F. R., Ismail A. A., Sedaman J. e Emo G., "Monitoring the Oxidation of Edible Oils by Fourier Transform Infrared Spectroscopy", *J. Am. Oil Chem. Soc.* **71**, págs 243-252, (1994).

10. Lourenço F. R. C. B., *Controle Estatístico de Qualidade*, Ao Livro Técnico, Rio de Janeiro, (1964).
11. Ekambaram S. K., *A base Estatística dos Gráficos de Controle de Qualidade*, Editora Polígono, São Paulo (1972).
12. Theodora K. e MacGregor J. F., "Process Analysis , Monitoring and Diagnosis, using Multivariate Projection Methods ", *Chemom. Intell. Lab. Systems* 28 págs 3-21, (1995)
13. Nijhuis A ., Jong S. e Vandeginste, B. G. M., "Multivariate Statistical Process Control in Chromatography", *Chemom. Intell. Lab. Systems* 38 págs 51-62 (1997).
14. Sala O., Santos P. S., Hase Y., *Espectroscopia Raman Princípios e Aplicações*, Oficinas Gráficas da Fundação Valeparaibana de Ensino , São José dos Campos -SP (1984)
15. Skoog A. D., Leary J. J., *Principles of Instrumental Analysis*, Saunders College Publishing 4^a ed. , New York , (1991).
16. Van de Voort F.R., Sedman J. ,Emo G. e Ismail A.A., "Rapid and Direct Iodine Value and Saponification Number Determination of Fats and Oils by Attenued Total Reflectance/FourierTransform Infrared Spectroscopy", *J. Am. Oil Chem. Soc.* 69 págs 118-1123, (1992).
17. Coulthup D. W., *Introduction to Infrared and Raman Spectroscopy*, Academic Press, 2^a ed., New York, (1975).
18. Haaland D. M., "Multivariate Calibration Methods Applied to Quantitative Chemical Analysis" em *Practical Fourier Transform Infrared Spectroscopy - Industrial and Laboratory Chemical Analysis* , editores : Ferraro J. R. e Krishmman K., Academic Press, New York, (1990).
19. Jawhari T., Hendra P. J., Wills H. A. e Judkins M. J., "Quantitative Analysis using Raman Methods", *Spectrochim. Acta* 46A págs 161-170, (1990).

20. Schader B., Hoffman A. e Stefan K., "Near Infrared Fourier Transform Raman Spectroscopy :facing absorption and Background", *Spectrochim. Acta* **47 A** págs 1135-1148 (1991).
21. Li-chan.E. C. Y., "The Aplication of Raman Spectroscopy in Food Science", *Trends Food Sci. Tech.* **7** págs 361-370, (1996).
22. Cutler D. J., " Fourier Transform Raman Instrumentation", *Spectrochim. Acta* **46A** págs 131-151, (1990).
23. Johnson A . R. e Wichem W. D., *Applied Multivariate Statistical Analysis*, Prentice Hall, Upper Saddle River, New Jersey, 4th ed., (1998).
24. Rawlings O . J., *Applied Regression Analysis – A Research Tool*, Wadsworth & Brooks /Cole Advanced Books & Software,Pacific Grove ,Califómia, (1988).
25. Martens H. e Tormod N., *Multivariate Calibration*, John Wiley & Sons ,New York, (1989) .
26. Geladi P. e Kowalski B. R., "Partial Least Square Regression: A Tutorial" , *Anal. Chim. Acta* **185** , págs 1-17 (1986).
27. Wise B. M. e Gallagher N. B., *PLS_Toolbox for use with Matlab*, Eigenvector Research Inc., Manson, WA, (1998).
28. Jackson E. J., Mudholkar S. G., "Control Procedures for Residuals Associated with Principal componentes", *Analysis –Technometrics* **21** págs 341-349, (1979).
29. Kowalski B. R. e Beebe R. K., "An Introduction to Multivariate Calibration and Analysis" *Anal. Chem.* **59** , págs 1007-1017A (1987).
30. Thomas V. E., "A Primer on Multivariate Calibration" *Anal. Chem* **66** , págs 795-804 (1994) .

31. Poppi R. J., "Quantificação de Picos Cromatográficos Superpostos por Métodos de Calibração Multivariada", UNICAMP – dissertação de mestrado (1989).
32. The Unscrambler, CAMO ASA, versão 6.0 (1997).
33. Morrison R. T. e Boyd R. M., *Química Orgânica*, Fundação Calouste Gulbenkian, 12ª edição , Lisboa, (1983).
34. Ohlweiler , O .A –*Química Analítica Quantitativa* ,4ª Edição -Livros Técnicos e Científicos, vol 2.
35. Godinho, O . E. S., Baccan N., de Andrade J. C. e Barone J. S., *Química Analítica Quantitativa Elementar* ,editora UNICAMP, Campinas, 2ª. ed. (1979).