

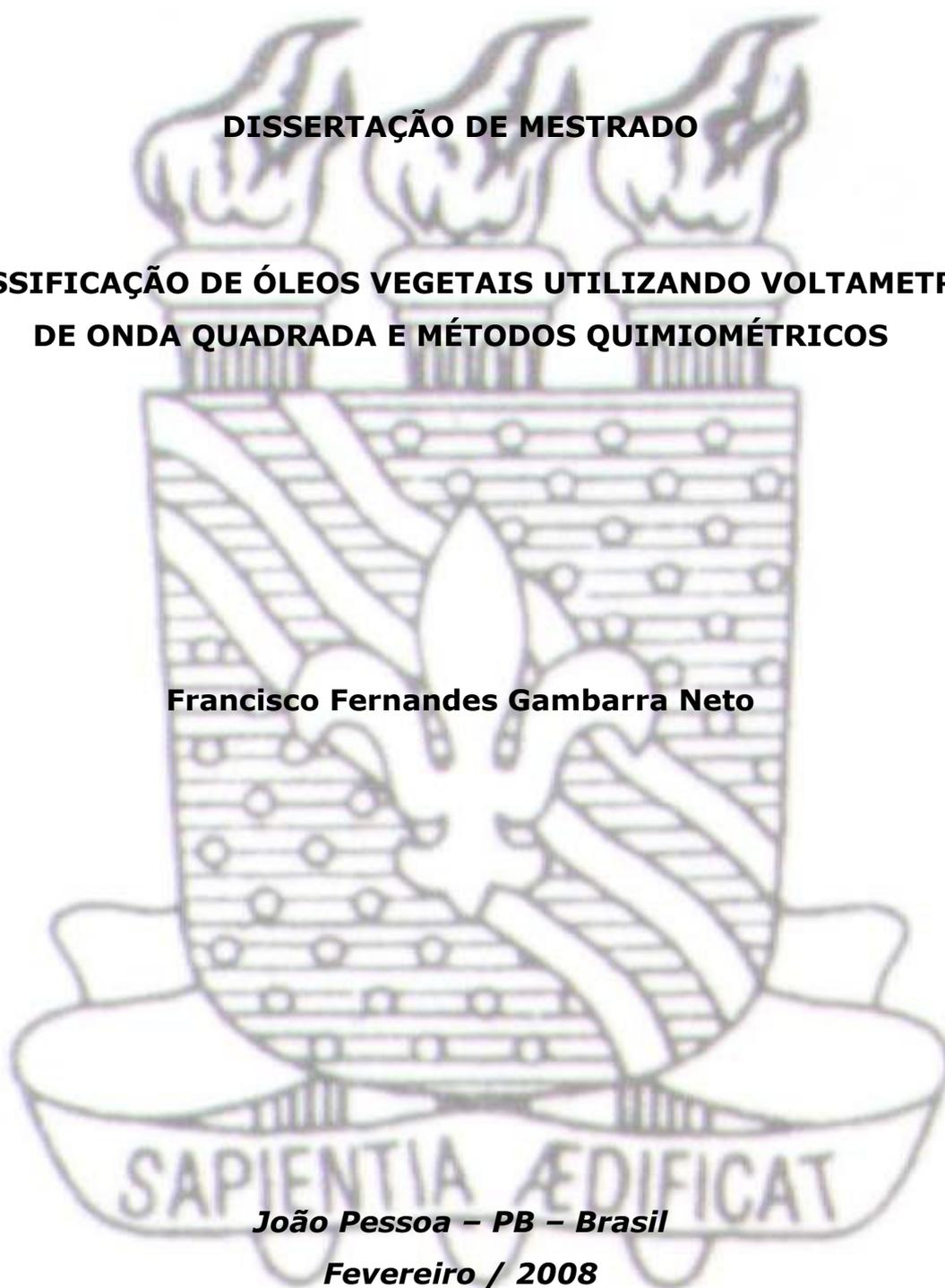


**UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
DEPARTAMENTO DE QUÍMICA
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA**

DISSERTAÇÃO DE MESTRADO

**CLASSIFICAÇÃO DE ÓLEOS VEGETAIS UTILIZANDO VOLTAMETRIA
DE ONDA QUADRADA E MÉTODOS QUIMIOMÉTRICOS**

Francisco Fernandes Gambarra Neto



João Pessoa – PB – Brasil

Fevereiro / 2008



**UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
DEPARTAMENTO DE QUÍMICA
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA**

DISSERTAÇÃO DE MESTRADO

**CLASSIFICAÇÃO DE ÓLEOS VEGETAIS UTILIZANDO VOLTAMETRIA
DE ONDA QUADRADA E MÉTODOS QUIMIOMÉTRICOS**

Francisco Fernandes Gambarra Neto

**Dissertação apresentada como
requisito para a obtenção do
título de Mestre em Química pela
Universidade Federal da Paraíba.**

Orientador: Prof. Dr. Mário César Ugulino de Araújo

2º Orientador: Dr. Glimaldo Marino

Bolsista (CAPES)

João Pessoa – PB – Brasil

Fevereiro / 2008

Classificação de Óleos Vegetais Utilizando Voltametria de Onda Quadrada e Métodos Quimiométricos.

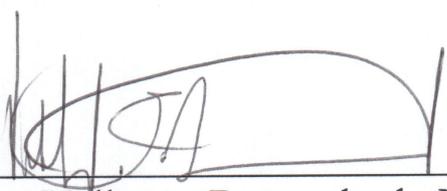
Aprovada pela banca examinadora:



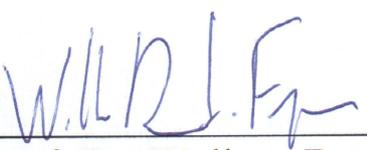
Prof. Dr. Mário César Ugulino de Araújo
Orientador/Presidente



Prof. Dr. Glimaldo Marino
2º. Orientador



Prof. Dr. Valberes Bernardo do Nascimento
Examinador



Prof. Dr. Wallace Duarte Fragoso
Examinador

Aos meus pais, Fernando Antônio Casimiro Gambarra e Ana Maria Fontes Gambarra, à minha avó, Maria de Lourdes Casimiro Fernandes e às minhas irmãs, Fernanda Fontes Gambarra e Fabiana Fontes Gambarra, por todo amor, amizade, apoio, ajuda e confiança.

À minha namorada, Wanessa Porto Tito, por todo amor, companheirismo e compreensão.

Com muito carinho e gratidão,
Eu dedico.

AGRADECIMENTOS

- Em primeiro lugar agradeço a Deus por me dar disposição, força de vontade e saúde para realizar este trabalho.
- Aos meus grandes amigos Márcio José Coelho de Pontes e Gledson Emídio José, que estão sempre do meu lado acreditando, ajudando, debatendo, colaborando e sempre prontos para escutarem e compartilharem os bons e maus momentos.
- Ao Professor Dr. Mário César Ugulino de Araújo, pela orientação, confiança e por ter apostado em mim.
- Aos pesquisadores Dr. Glimaldo Marino e Dr. Everaldo Paulo de Medeiros, pelo aprendizado, ajuda, incentivo e, principalmente, pela amizade.
- À Renato Sousa Lima, pela amizade e ajuda em uma parte do experimental.
- A todos do LAQA, que de uma forma ou de outra, ajudaram-me e fizeram-me sentir bem-vindo ao laboratório.
- Aos demais professores do LAQA e do Departamento de Química, que me forneceram aprofundamentos disciplinares durante o Mestrado.
- Aos Professores Germano Veras e Ana Cláudia, pelos conselhos, amizade e confiança desde a graduação pela Universidade Estadual da Paraíba.
- A CAPES pelo fornecimento da bolsa.

SUMÁRIO

ÍNDICE DE FIGURAS	v
ÍNDICE DE TABELAS	x
LISTA DE ABREVIATURAS E SIGLAS	xii
RESUMO	xiv
CAPÍTULO I INTRODUÇÃO	1
1.1 Óleos Vegetais	2
1.2 Composição dos óleos vegetais	3
1.3 Óleos vegetais e a saúde humana.....	5
1.4 Tipos de óleos vegetais	6
1.4.1 Óleo de canola.....	6
1.4.2 Óleo de girassol	7
1.4.3 Óleo de milho	8
1.4.4 Óleo de soja	8
1.5 Controle de qualidade de óleos vegetais.....	10
1.6 Voltametria de onda quadrada	11
1.7 Quimiometria.....	13
1.7.1 Pré-processamento dos dados	14
1.7.2 Planejamentos experimentais	16
1.7.2.1 Planejamento composto central	16
1.7.2.2 Parâmetro de Taguchi	17
1.7.3 Análise de Componentes Principais (PCA).....	18
1.7.4 SIMCA	20
1.7.5 Análise discriminante linear (LDA)	22
1.7.6 Seleção de variáveis.....	23
1.7.7 Algoritmo das projeções sucessivas (SPA)	23
1.7.8 SPA no contexto da classificação	25
1.7.9 Quimiometria aplicada à voltametria.....	26
CAPÍTULO II OBJETIVOS	32
2.1 Objetivo Geral.....	33
2.1.1 Objetivos Específicos:	33

CAPÍTULO III EXPERIMENTAL	34
3.1 Amostras.....	35
3.2 Divisão do conjunto de amostras	35
3.3 Materiais e reagentes.....	36
3.4 Equipamentos	36
3.5 Procedimento analítico	37
3.5.1 Manipulação das amostras	38
3.5.2 Eletrólito suporte, eletrodo de trabalho e sentido da varredura	39
3.5.3 Limpeza do eletrodo de trabalho e parâmetros da SWV.....	39
3.6 Softwares	42
 CAPÍTULO IV RESULTADOS E DISCUSSÃO	 43
4.1 Eficiência da extração alcoólica	44
4.2 Eletrólito suporte, eletrodo de trabalho e sentido da varredura.....	45
4.3 Limpeza eletroquímica do eletrodo de trabalho.....	46
4.4 Otimização dos parâmetros da SWV.....	49
4.5 Seleção da região de trabalho no voltamograma	51
4.6 Perfis dos voltamogramas do etanol e de etanol + óleo	52
4.7 Pré-processamento dos voltamogramas	53
4.7.1 Pré-processamento aplicado às variáveis.....	54
4.8 Análise exploratória dos dados	55
4.8.1 Aplicação da PCA.....	55
4.8.1.1 Aplicação da PCA às amostras não vencidas	56
4.8.1.2 Comparação dos óleos em relação à composição química...57	
4.8.1.3 Aplicação de PCA às amostras da classe PVE	59
4.9 Modelagem e Classificação SWV-SIMCA.....	60
4.9.1 Construção dos modelos SWV-SIMCA	61
4.9.2 Validação dos modelos SWV-SIMCA.....	64
4.9.2.1 Validação do modelo SWV-SIMCA da classe Canola	64
4.9.2.2 Validação do modelo SWV-SIMCA da classe Girassol	67
4.9.2.3 Validação do modelo SWV-SIMCA da classe Milho	67
4.9.2.4 Validação do modelo SWV-SIMCA da classe Soja	68
4.9.2.5 Validação do modelo SWV-SIMCA da classe PVE	69
4.9.3 Uso dos modelos SWV-SIMCA para o Conjunto de Previsão	72
4.10 Construção dos modelos SWV-LDA-SPA.....	74

4.10.1	Uso do modelo SWV-LDA-SPA no Conjunto de Previsão.....	76
4.10.2	Uso do SWV-LDA-SPA para a previsão de todas as amostras .	76
4.11	Comparação entre os modelos SWV-SIMCA e SWV-LDA-SPA.....	78
CAPÍTULO V CONCLUSÕES.....		80
5.1	Conclusões	81
5.2	Propostas futuras	82
CAPÍTULO VI REFERÊNCIAS		83

ÍNDICE DE FIGURAS

Figura 1.1 Estrutura molecular dos principais ácidos graxos dos óleos vegetais	4
Figura 1.2 Variação de potencial em função do tempo para a SWV. Estão representados o potencial inicial, E_i , a amplitude de pulso, ΔE_p , o incremento de potencial, ΔE_s , a largura de pulso, t_p e os pontos A e B onde as correntes direta e reversa são medidas, respectivamente.	12
Figura 1.3 Voltamogramas esquemáticos de onda quadrada (SOUZA et al., 2003) para um processo redox de um sistema reversível (1) e de um sistema irreversível (2).	12
Figura 1.4 Representação de uma matriz de dados de natureza multivariada.	14
Figura 1.5 Primeira e segunda PCs no espaço tridimensional das variáveis.	19
Figura 1.6 Previsão de uma amostra desconhecida A usando o modelo SIMCA da classe B construído com uma PC. A distância "y" entre a amostra A e o eixo da componente corresponde ao resíduo da PC1, "x" é a distância entre a projeção de A na direção da PC1 e a fronteira da classe e "z" é a distância de A até o limite da classe.	20
Figura 1.7 Gráfico de $S_i \times H_i$, detalhando os limites de uma determinada classe 1 e a disposição das demais amostras ao seu redor. Neste caso, uma das amostras da própria classe 1 foi mal classificada como visualizado.	21
Figura 1.8 Representação esquemática da análise discriminante linear. Os centros de gravidade (os asteriscos) dos grupos formados por círculos e quadrados são inicialmente determinados. Em seguida traça-se um plano de decisão eqüidistante e perpendicular à linha que une os centros de gravidade dos grupos, formando as regiões R1 e R2.	

Se um ponto qualquer estiver presente na região R1, o mesmo é classificado como pertencente ao grupo 1. Caso contrário, obrigatoriamente pertencerá à categoria 2.....22

Figura 1.9 Exemplo do funcionamento do SPA para selecionar potenciais a partir de cinco variáveis voltamétricas: a) Projeções no plano ortogonal a X_3 ; b) Visão espacial dos vetores no plano horizontal.24

Figura 3.1 Potenciostato/Galvanostato μ AUTOLAB[®] Type II à esquerda e o módulo 663 VA Stand[®] da cela eletroquímica à direita.37

Figura 3.2 Esquema da extração de ácidos graxos das amostras de óleo e tomada de alíquota para realização das medidas.....38

Figura 3.3 Quadrado do voltamograma médio (a) e perfil da variância (b).....41

Figura 4.1 Perfil dos voltamogramas após a subtração do perfil de etanol P.A. no sentido de varredura anódico para 4 diferentes proporções de uma amostra de óleo de soja em etanol.44

Figura 4.2 Perfil dos voltamogramas após a subtração do perfil de etanol P. A. nos sentidos de varredura anódica (linha preta) e catódico (linha vermelha), empregando RDE-Au e o eletrólito suporte NaOH (a) e o CH_3COOH (b); RDE-Pt e o eletrólito suporte NaOH (c) e CH_3COOH (d). A concentração dos eletrólitos de suporte foi $0,1 \text{ mol L}^{-1}$45

Figura 4.3 Voltamograma de uma amostra de óleo de canola (linha vermelha), na região de 0,300 a -0,900 V, após subtrair a resposta do branco ($\text{NaOH } 0,1 \text{ mol L}^{-1}$ — linha azul) do voltamograma originalmente obtido pela SWV (linha verde).....51

Figura 4.4 Voltamograma de uma amostra de óleo de canola na região de trabalho selecionada (-0,046 a -0,900 V).52

- Figura 4.5** Perfis, após subtração do branco, dos voltamogramas do etanol (linha azul), de um óleo de canola em etanol na proporção 1:1 (linha vermelha) e a subtração destes dois perfis (linha verde). O detalhe no canto superior esquerdo dessa figura é uma ampliação da área marcada nos voltamogramas.52
- Figura 4.6** Voltamogramas das 114 amostras de óleo vegetal na região de trabalho, após a subtração do branco: Voltamogramas das 66 amostras de óleo não vencidas (a) e das 48 amostras vencidas (b)...54
- Figura 4.7** Gráfico dos escores de PC1 *versus* PC2 das 114 amostras de óleo vegetal.55
- Figura 4.8** Gráfico dos escores de PC1 *versus* PC2 para as 66 amostras não vencidas de óleo vegetal distribuídas em quatro classes. ..56
- Figura 4.9** Voltamogramas médios das classes Canola, Girassol, Milho e Soja. A seta indica uma ampliação da região de menor intensidade de corrente e o círculo destaca a região de -580 mV a -650 mV onde o perfil da classe Canola se diferencia das demais.57
- Figura 4.10** Gráfico dos escores de PC1 *versus* PC2 das 48 amostras de óleo vegetal da classe PVE.59
- Figura 4.11** Perfil voltamétrico das amostras de óleo da classe PVE.....60
- Figura 4.12** Gráficos da variância residual e explicada *versus* nº de PCs das classes: a) Canola; b) Girassol; c) Milho; d) Soja; e) PVE.61
- Figura 4.13** Gráfico dos escores de PC1 *versus* PC2 *versus* PC3 para as classes: a) Canola; b) Girassol; c) Milho; d) Soja; e) PVE. As amostras dos conjuntos de treinamento e de teste estão pintadas de azul e vermelho, respectivamente.63
- Figura 4.14** Gráfico de Si x Hi (95% de confiança) do modelo da classe Canola construído com 6 PCs. A seta indica uma ampliação das

fronteiras deste modelo evidenciando a classificação incorreta da amostra 11.....	65
Figura 4.15 Teste F aplicado aos resíduos das amostras de treinamento e teste da classe Canola para modelos com 4 (a), 5(b) e 6 (c) PCs.....	65
Figura 4.16 Gráfico de $S_i \times H_i$ (95% de confiança) do modelo da classe Canola construído com 4 PCs após a etapa de validação. A seta indica uma ampliação das fronteiras deste modelo evidenciando a classificação correta de todas as amostras.....	66
Figura 4.17 Gráfico de $S_i \times H_i$ (95% de confiança) do modelo da classe Girassol construído com 4 PCs. A seta indica uma ampliação das fronteiras deste modelo evidenciando a classificação incorreta de sete amostras de óleo de soja.....	67
Figura 4.18 Gráfico de $S_i \times H_i$ (95% de confiança) do modelo da classe Milho construído com 3 PCs. A seta indica uma ampliação das fronteiras deste modelo evidenciando a classificação incorreta de uma amostra de óleo de soja.	68
Figura 4.19 Gráfico de $S_i \times H_i$ (95% de confiança) do modelo da classe Soja construído com 4 PCs. A seta indica uma ampliação das fronteiras deste modelo evidenciando a classificação incorreta de seis amostras de óleo de girassol e de uma amostra de óleo de milho.	69
Figura 4.20 Gráfico de $S_i \times H_i$ (95% de confiança) do modelo da classe PVE construído com 3 PCs. A seta indica uma ampliação das fronteiras do modelo evidenciando a classificação incorreta das amostras 10 e 19.	70
Figura 4.21 Teste F aplicado aos resíduos das amostras de treinamento e teste da classe PVE para modelos com 1 (a), 2(b), 3 (c) e 4 (d) PCs.	70

Figura 4.22 Perfil dos resíduos das amostras da classe PVE <i>versus</i> variáveis para modelos com 2 (a) e 4 (b) PCs.	71
Figura 4.23 Número de variáveis selecionadas pelo SPA <i>versus</i> os valores da função de custo.	75
Figura 4.24 Variáveis selecionadas pelo LDA-SPA.....	76
Figura 4.25 Gráfico dos escores de FD1 <i>versus</i> FD2 das 114 amostras de óleo vegetal.	77
Figura 4.26 Gráfico dos escores de FD2 <i>versus</i> FD3 para as 66 amostras não vencidas de óleo vegetal distribuídas em quatro classes. ..	78

ÍNDICE DE TABELAS

Tabela 1.1 Principais oleaginosas e seus respectivos conteúdos em óleo.	6
Tabela 1.2 Características físico-químicas e composição em ácidos graxos dos óleos refinados de canola, girassol, milho e soja (Fonte: ANVISA, 2007).....	9
Tabela 3.1 Classes e quantidade de amostras de óleos vegetais analisados.	35
Tabela 3.2 Número de amostras de treinamento, teste e previsão selecionadas pelo KS para as cinco classes de óleos vegetais.	36
Tabela 4.1 Detalhamento dos fatores e níveis utilizados para a construção do planejamento composto central, em que $\alpha = 1,5$	47
Tabela 4.2 Planejamento composto central 2^4 para a otimização da etapa de limpeza eletroquímica do eletrodo de trabalho.	48
Tabela 4.3 Parâmetros ótimos escolhidos para a limpeza eletroquímica do eletrodo.	49
Tabela 4.4 Detalhamento dos fatores e níveis utilizados para a construção do planejamento composto central 2^3 , em que $\alpha = 1,67$	49
Tabela 4.5 Planejamento composto central 2^3	50
Tabela 4.6 Parâmetros escolhidos para a SWV.	50
Tabela 4.7 Matriz da composição química dos óleos vegetais de acordo com os dados da RDC 492/99 (ANVISA, 2007).	58
Tabela 4.8 Correlação entre os diferentes tipos de óleo vegetal de acordo com a composição de química.....	58

- Tabela 4.9** Número de PCs e a variância explicada acumulada dos modelos SWV-SIMCA.62
- Tabela 4.10** Número de PCs e variância explicada acumulada para cada modelo SWV-SIMCA de óleo vegetal após a etapa de validação.72
- Tabela 4.11** Classificação SIMCA das amostras de previsão das cinco classes de óleo vegetal. O símbolo * indica que a amostra foi classificada no modelo.73
- Tabela 4.12** Erro de classificação dos modelos SWV-SIMCA (com quatro níveis de significância) e SWV-LDA-SPA para as amostras de óleo do conjunto de previsão.79

LISTA DE ABREVIATURAS E SIGLAS

AcG	Ácido Graxo
AcGs	Ácidos Graxos
Ag/AgCl	Eletrodo de referência de prata/cloreto de prata
ALS	Mínimos quadrados alternados
ANN	Redes neurais artificiais
ANVISA	Agência Nacional de Vigilância Sanitária
CLS	Mínimos quadrados clássicos
CNNPA	Conselho Nacional de Normas e Padrões para Alimentos
CO ₂	Dióxido de Carbono
CV	Voltametria cíclica
CVA	Análise de variáveis canônicas
DA	Análise de Discriminante
DI	Dimensionalidade inerente
DPP	Polarografia de pulso diferencial
DPV	Voltametria de pulso diferencial
f	Freqüência da onda quadrada
FA	Fase alcoólica
F_{cal}	Valor calculado para o teste F
F_{crit}	Valor crítico adotado para o teste F
FT	Transformada de Fourier
G	Risco médio de uma classificação incorreta
HCA	Análise de agrupamentos hierárquicos
Hi	Influência
HMDE	Eletrodo de gota pendente de mercúrio
k	Número de fatores no planejamento experimental
KF	Kalman filter
KNN	k -vizinhos mais próximos (<i>k-nearest neighbours</i>)
LDA	Análise discriminante linear
LDL	Lipoproteína de baixa densidade (mau colesterol)

LSV	Voltametria de varredura linear
MCR	Resolução de curva multivariada
MIR	Espectroscopia no infravermelho médio
MLR	Regressão linear múltipla
n_F	Número de experimentos de um planejamento fatorial
NaOH	Hidróxido de sódio
NIR	Espectroscopia no infravermelho próximo
NPV	Voltametria de pulso normal
PC	Componente principal
PCA	Análise de componentes principais
PCR	Regressão por componente principal
PLS	Mínimos quadrados parciais
PVE	Prazo de Validade Expirado
RDC	Resolução da Diretoria Colegiada
RDE-Pt	Eletrodo rotatório de disco de platina
RMSEP	Erro quadrático médio de previsão
Si	Raiz quadrada da variância residual
SIMCA	Modelagem Independente Flexível por Analogia de Classe
SPA	Algoritmo das projeções sucessivas
SWV	Voltametria de onda quadrada
t_p	Largura ou tempo do pulso na onda quadrada
v	Velocidade de varredura da onda quadrada
WT	Transformada Wavelet
Xcal	Matriz das amostras de calibração
x	Distância entre a projeção de A na direção da PC e a fronteira
y	Distância entre a amostra desconhecida A e o eixo da PC
z	Proximidade de A até a caixa
Z	Parâmetro de Taguchi
α	Valor da distância axial dos pontos no planejamento estrela
ΔE_p	Amplitude de pulso
ΔE_s	Incremento de potencial

RESUMO

A autenticidade de óleos vegetais comestíveis tem se tornado um assunto muito importante, tanto do ponto de vista comercial, devido à evasão fiscal, como do prejuízo à saúde do consumidor. Assim, é necessário o desenvolvimento de novas metodologias analíticas que possibilitem análises simples, rápida e de baixo custo. A combinação de métodos voltamétricos e quimiométricos é uma boa alternativa para atingir este objetivo. A quimiometria tem sido muito usada em dados espectroscópicos para uma variedade de aplicações no campo de alimentos, medicamentos, ambiental, etc. Todavia, poucos trabalhos têm explorado as características atrativas dos métodos voltamétricos, principalmente com relação à sua natureza multivariada associada à quimiometria. Neste trabalho é proposto um método rápido, simples e barato para fiscalização de óleos vegetais comestíveis que combina a voltametria de onda quadrada com análise discriminante linear associada ao algoritmo das projeções sucessivas para a seleção de variáveis. Sua eficiência foi avaliada comparando-o ao método conhecido tradicionalmente como SIMCA. Cinco classes de óleo: Canola, Girassol, Milho, Soja e os com prazo de validade expirado (PVE) de diferentes fabricantes foram analisadas, utilizando um eletrodo de disco de platina. Para as 32 amostras do conjunto de previsão, o SIMCA apresentou 6 erros de classificação (81% de classificação correta), enquanto que o método proposto classificou corretamente todas as amostras sendo, portanto, uma alternativa promissora para a classificação de óleos vegetais e a fiscalização quanto ao prazo de validade.

Palavras chave: óleos vegetais, voltametria de onda quadrada, análise discriminante linear, algoritmo das projeções sucessivas.

ABSTRACT

The authenticity of edible vegetable oils has become a very important issue, both in trade due to tax evasion, and the damage caused to the health of the consumer. Thus, it is necessary to develop new analytical methodologies that allow simple, fast and low cost analysis. The combination of voltammetric and chemometric methods is a good alternative to achieve this goal. Chemometrics has been widely used in spectroscopic data for a variety of applications in the food, medicine, and environmental areas, etc. However, few studies have investigated the attractive characteristics of the voltammetric methods, particularly with respect to its multivariate nature, with chemometrics. This work proposes a fast, simple and cheap monitoring method for edible vegetable oils that combines the Square Wave Voltammetry (SWV) with Linear Discriminant Analysis (LDA) associated with Successive Projection Algorithm (SPA) employed in the variable selection. Its performance was evaluated by comparing it to the traditional method known as SIMCA (Soft Independent Modeling of Class Analogy). Five classes of oil: canola, sunflower, corn, soybean and the ones with expired shelf-life, from different manufacturers were analyzed with a platinum disk working electrode. The SWV-SIMCA method presented 6 errors (81% of correct classification), whereas the SWV-LDA-SPA method proposed classified correctly all the prediction samples, showing to be a promising alternative for classification and inspection of vegetable oils.

Key words: vegetable oil, SWV, LDA, SPA, SIMCA.

CAPÍTULO I

INTRODUÇÃO

1.1 Óleos Vegetais

Óleos e gorduras são substâncias lipofílicas bastante utilizadas pelos organismos vivos, podendo ser de origem animal, vegetal ou mesmo bacteriana (MORETTO e FEET, 1998) e formam uma das três maiores classes de produtos alimentícios, sendo as outras constituídas por proteínas e carboidratos, e são conhecidos desde os ancestrais do homem pela facilidade de extração e pelas diversas fontes e propriedades (WEISS, 1970). A diferença entre óleos e gorduras é definida pela resolução nº 20/77 do CNNPA (Conselho Nacional de Normas e Padrões para Alimentos) que impõe a temperatura de 20 °C como limite inferior para o ponto de fusão das gorduras, ou seja, quando o ponto de fusão se situa abaixo deste limite, a substância é classificada como óleo (líquido a temperatura ambiente). Contudo, o termo gordura é o mais empregado quando o estado físico não é significativo.

O óleo vegetal é uma gordura obtida de plantas. É extraído na maioria das vezes das sementes ou grãos destes vegetais conhecidos como oleaginosos. Os óleos destinados ao consumo humano são submetidos a um processo de refino cuja finalidade é melhorar sua aparência, sabor, odor e estabilidade por remoção de alguns componentes como: ácidos graxos livres, proteínas, corantes naturais, umidade e compostos voláteis e inorgânicos (MORETTO e FEET, 1998).

A cadeia agroindustrial de oleaginosas abrange um conjunto de mercados, dos quais os mais importantes são: produção e comércio de grãos; produção de óleos vegetais (óleo bruto e refinado) e produção de rações (COUTINHO et al., 2008). No Brasil, desde 1972 é feito um levantamento anual sobre a safra de grãos. No ano de 2003 foi observada uma produção de 123,6 milhões de toneladas, recorde nacional até então. No ano de 2007 um novo recorde foi atingido com uma produção de 133 milhões de toneladas, com soja e milho correspondendo a 82% do total. Isto ocorreu devido ao elevado preço dos grãos no mercado internacional, fazendo do Brasil um dos maiores produtores mundiais de soja

(BECKSTEIN, 2008). Para 2008, está previsto um aumento de 2,1% na safra rendendo 135,7 milhões de toneladas. A crescente demanda das oleaginosas é responsável direto pelo ascendente cultivo de grãos (IBGE, 2008).

Além do setor alimentício, o Governo Federal brasileiro vem incentivando o aproveitamento dos óleos vegetais como fonte energética destinada a diversos segmentos, abrindo oportunidades nos setores social, econômico e, sobretudo, ambiental. A produção está sendo destinada para a estruturação do programa de produção e uso do biodiesel em muitos países, principalmente no Brasil, já que diversas espécies oleaginosas, tais como a palma, nabo forrageiro, mamona, soja, girassol e babaçu, entre outras são cultivadas e posteriormente utilizadas como matéria-prima na produção de biocombustível (COUTINHO et al., 2008, BIODIESELBR, 2008).

Com base na expansão do mercado de óleos vegetais, seu controle de qualidade tem se tornado um assunto importante tanto do ponto de vista comercial como da saúde dos consumidores. Portanto, é de fundamental importância o constante aperfeiçoamento de técnicas analíticas simples, rápidas, robustas e de baixo custo.

1.2 Composição dos óleos vegetais

Os óleos vegetais são constituídos majoritariamente por triglicerídeos, que são compostos de três moléculas de ácido graxo (AcG) com uma molécula de glicerol. As unidades acilas são os principais constituintes dos óleos vegetais, representando cerca de 95% da massa molecular dos triglicerídeos. Em pequena quantidade aparecem mono e diglicerídeos, ácidos graxos livres, glicolipídeos, proteínas, fosfolipídeos, esteróis entre outras substâncias (MORETTO e FEET, 1998, WEISS, 1970, KRAUSE e MAHAN, 1991).

Os AcGs são ácidos monocarboxílicos de cadeia longa e com número par de átomos de carbono (4 a 30 átomos) dispostos em uma cadeia

linear, em virtude da sua produção biológica, que se inicia a partir de unidades acetato da acetil coenzima A. Podem ser saturados ou insaturados, com estes predominando nos óleos vegetais. As ligações duplas raramente se apresentam conjugadas (-CH=C=CH-), e sim, separadas por um grupo metilênico (-CH=CH-CH₂-CH=CH-). As duplas ligações impedem a rotação livre dos átomos gerando os isômeros nas configurações *cis*, que é a mais comum entre os AcGs dos óleos vegetais, e *trans*. AcGs são geralmente representados pelo símbolo C_x:y(z), em que C_x indica o número de átomos de carbono que forma a cadeia linear; y e z indicam a quantidade de ligação dupla existente na molécula e sua(s) posição(ões), respectivamente. (MORETTO e FEET, 1998, WEISS, 1970, KRAUSE e MAHAN, 1991).

Nos óleos vegetais, os AcGs mais comuns são o palmítico (C16:0) e o esteárico (C18:0), entre os saturados, e oléico (C18:1(9)), linoléico (C18:2(9,12)) e linolênico (C18:3(9,12,15)), entre os insaturados (Figura 1.1).

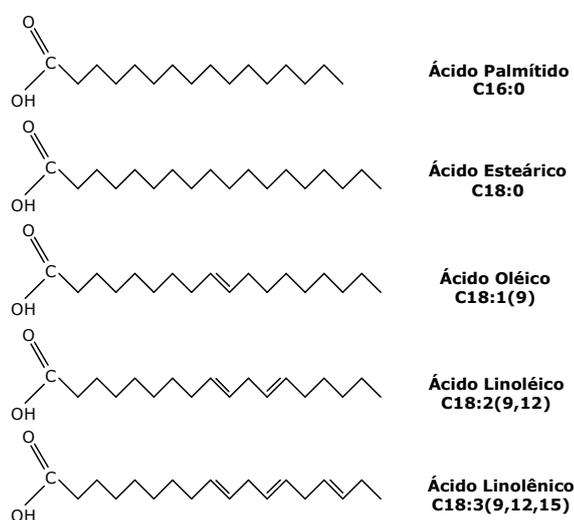


Figura 1.1 Estrutura molecular dos principais ácidos graxos dos óleos vegetais

Nas áreas de bioquímica e nutrição, os ácidos graxos são agrupados em famílias conhecidas como ômega (ω). Nestas, o carbono do grupamento metila terminal é reconhecido como posição número um, logo, a posição da primeira ligação dupla caracteriza a família. Por exemplo, o ácido oléico e o linoléico pertencem às famílias ω 9 e ω 6,

respectivamente (Figura 1.1). Os AcGs das famílias $\omega 6$ e $\omega 3$ não são produzidos pelo organismo humano e por isso são considerados como ácidos graxos essenciais ao homem, sendo o ácido linoléico, $\omega 6$, e o ácido linolênico, $\omega 3$, os mais importantes (MORETTO e FEET, 1998, KRAUSE e MAHAN, 1991). Todavia, o ácido linolênico é encontrado em baixas concentrações na maioria dos óleos vegetais, sendo o óleo de linhaça uma exceção com cerca de 60% (m/m) de $\omega 3$ (CEOTTO, 2008).

1.3 Óleos vegetais e a saúde humana

Os óleos, juntamente com as gorduras, são denominados geralmente de lipídios e são encontrados no sangue e em diferentes órgãos do homem. Eles desempenham funções cruciais como co-fatores enzimáticos, transportadores de elétrons, pigmentos que absorvem radiações luminosas, âncoras hidrofóbicas, agentes emulsificantes, hormônios, mensageiros intracelulares, componentes da estrutura celular e como reservas de energia (LEHNINGER et al., 1995, KRAUSE e MAHAN, 1991). Esta importante fonte de combustível para os seres humanos contribui, em alguns países, com 30-40% do total de energia consumida na alimentação. Entretanto, quando ingeridos em excesso e/ou sem levar em consideração a qualidade e o tipo da gordura, podem causar graves problemas como a obesidade, o aumento de colesterol no sangue e cardiopatias. Estas complicações são atribuídas à maior proporção de ácidos graxos saturados que aparecem principalmente em gorduras animais, manteigas e em produtos derivados da hidrogenação de óleos como as margarinas. Por outro lado, as gorduras insaturadas são capazes de reduzir o índice de colesterol, diminuindo os problemas cardiovasculares. Como os óleos vegetais são ricos em ácidos graxos insaturados, eles são os mais indicados para uma dieta saudável (MORETTO e FEET, 1998, KRAUSE e MAHAN, 1991).

Os ácidos graxos linoléico ($\omega 6$) e linolênico ($\omega 3$), ambos poliinsaturados, fazem parte de um grupo especial de ácidos graxos, cuja

ingestão é muito importante, pois ajudam na redução dos níveis de colesterol no sangue e desempenham funções fisiológicas imprescindíveis nos organismos vivos, sendo, portanto, essenciais à vida (LEHNINGER et al., 1995, KRAUSE e MAHAN, 1991).

1.4 Tipos de óleos vegetais

Existe uma grande diversidade de oleaginosas em todo o mundo. Algumas das mais importantes estão listadas na [Tabela 1.1](#).

Tabela 1.1 Principais oleaginosas e seus respectivos conteúdos em óleo.

Material oleaginoso	Óleo (% em massa)
Copra	66-68
Babaçu	60-65
Gergelim	50-55
Polpa de palma (dendê)	45-50
Caroço de palma	45-50
Amendoim	45-50
Colza (canola)	40-45
Girassol	35-45
Açafrão	30-35
Oliva	25-30
Algodão	18-20
Soja	18-20

De acordo com a Agência Nacional de Vigilância Sanitária, em sua resolução RDC 482/99 ([ANVISA, 2007](#)), são comercializados no Brasil os seguintes óleos vegetais:

- algodão
- amendoim
- arroz
- canola
- girassol
- milho
- soja
- uva
- babaçu
- côco
- palma
- palmiste
- oliva
- gergelim
- azeite saborizado

A seguir, são apresentadas características de quatro destes óleos, a saber, canola, girassol, milho e soja.

1.4.1 Óleo de canola

Este óleo é conhecido desde os primeiros séculos da Era Cristã como produto de extração de uma planta chamada colza. No entanto, ela

começou a participar da alimentação humana somente a partir da década de 1970, quando cientistas conseguiram, através de melhoramento genético, desenvolver espécies dessa planta (*Brassica napus* e *Brassica campestris*) com teores aceitáveis de ácido erúico e glucosilatos. O nome canola (**C**anadian **O**il **L**ow **A**cid) foi oficialmente aceito pela Canadian Grain Commission, em 1987 (MORETTO e FEET, 1998).

Atualmente a canola é cultivada em diversos países, sendo China, Canadá, Índia, Alemanha, França, Austrália e Polônia os maiores produtores mundiais (GAZZONI, 2007). No Brasil, os trabalhos de pesquisa com a cultura de colza nos estados do Rio Grande do Sul e de Santa Catarina foram iniciados em 1974, em meados dos anos 80 no Paraná (PR) e em 2003 no estado de Goiás (GO). Os grãos de canola produzidos no país possuem cerca de 24 a 27% de proteína e de 34 a 40% de óleo (EMBRAPA, 2007a).

O óleo de canola é um dos mais saudáveis, pois possui elevada quantidade de $\omega 3$, gorduras monoinsaturadas e vitamina E, que ajudam a diminuir o LDL (*low density lipoprotein*), mais conhecido como "mau colesterol". Além disso, possui menor quantidade de gordura saturada (6%) quando comparados com óleos de soja (15%) e de milho (11%) (MORETTO e FEET, 1998, EMBRAPA, 2007a, EMBRAPA, 2007b).

1.4.2 Óleo de girassol

O girassol (*Helianthus annus* L.) é uma planta de origem americana, que foi utilizada como alimento pelos índios. É um dos poucos vegetais que pode ser cultivado em clima frio (ANVISA, 2007, WEISS, 1970).

Atualmente, os maiores produtores mundiais são a Rússia, Estados Unidos, Argentina e China. O girassol ocupa o quarto lugar como fonte de óleo comestível, perdendo para soja, palma e canola. A produção brasileira ainda é pouco expressiva, mas o país tem importado este grão, principalmente da Argentina, para atender a demanda comercial deste produto, uma vez que o óleo de girassol vem sendo indicado em dietas por sua baixa quantidade de ácidos graxos saturados, alto teor de ácido

linoléico e vitaminas, além de apresentar características organolépticas agradáveis (ABOISSA, 2007, CAMPESTRE, 2006).

1.4.3 Óleo de milho

O milho (*Zea mays* L.) é um dos cereais mais cultivados por todo o mundo. Originou-se nas Américas e logo foi difundido pela Europa, África e Ásia. Atualmente, os maiores produtores mundiais são os Estados Unidos, entretanto, outros países como Canadá, China, Países Bálticos, Brasil, México, Argentina e África do Sul têm uma produção significativa deste grão. No Brasil, o milho é o cereal mais cultivado depois da soja (MORETTO e FEET, 1998, IBGE, 2008).

Os principais ácidos graxos que compõem o óleo de milho são o linoléico (até 62%) e oléico (até 42%). O ácido linolênico (tri-insaturado) é altamente sujeito à oxidação e está presente em pequena percentagem neste óleo, quando comparado ao de soja. Geralmente, os triglicerídeos do óleo de milho apresentam a cadeia poliinsaturada na posição central do glicerol (carbono 2), cercada por duas cadeias saturadas, ou por uma cadeia saturada e uma insaturada nas extremidades (carbono 1 e 3). Esta configuração fornece à cadeia poliinsaturada uma proteção contra os processos oxidativos, aumentando a estabilidade, já que as posições 1 e 3 são a mais reativas (MORETTO e FEET, 1998, RODRIGUES et al., 2003, REGINA e SOLFERINI, 2007).

1.4.4 Óleo de soja

A soja (*Glycine max* L.) é a oleaginosa mais cultivada no Brasil, que ocupa a 2ª posição do *ranking* mundial da produção, perdendo apenas para os Estados Unidos (IBGE, 2008, GAZZONI, 2007). Esta oleaginosa domina o mercado mundial tanto de proteína vegetal como de óleo comestível (MORETTO e FEET, 1998). O grão de soja possui cerca de 40% de proteínas, 20% de lipídios (óleo), 5% de minerais e 34% de carboidratos, além de minerais como potássio, cálcio, fósforo, ferro, cobre, magnésio e sódio. Os principais AcGs que compõem o óleo de soja

são o linoléico, oléico, palmítico e linolênico. Estas características fazem da soja e seus derivados um importante fator para a balança comercial brasileira. A vantagem do óleo de soja em relação a outros se deve ao seu baixo preço aliado à sua boa qualidade. No Brasil, além de ser utilizado na alimentação e em diversos outros setores, o óleo de soja vem se destacando na produção de biodiesel, abrindo também as possibilidades para a utilização dos demais óleos vegetais, uma vez que eles se tornaram o principal insumo para a produção de biodiesel (EMBRAPA, 2007c, CAMPESTRE, 2007).

A Tabela 1.2 abaixo detalha as características físico-químicas e os ácidos graxos presentes nos óleos refinados de canola, girassol, milho e soja de acordo com a RDC 482/99 da ANVISA (2007).

Tabela 1.2 Características físico-químicas e composição em ácidos graxos dos óleos refinados de canola, girassol, milho e soja (Fonte: ANVISA, 2007).

	Canola	Girassol	Milho	Soja	
Características físico-químicas	Índice de Refração (40°C)	1,465-1,467	1,467-1,469	1,465-1,468	1,466-1,470
	Índice de Iodo (Wijs)	110-126	110-143	103-128	120-143
	Índice de Saponificação	182-193	188-194	187-195	189-195
	Matéria Insaponificável (g/100g)	< 2,0	< 1,5	< 2,8	< 1,5
	Acidez (g de ácido oléico/100g)	< 0,3	< 0,3	< 0,3	< 0,3
	Índice de Peróxido (meq/Kg)	< 10,0	< 10,0	< 10,0	< 10,0
	Brassicasterol (g/100g)	> 5,0	—	—	—
Ácidos graxos (g/100g)	C < 14	—	< 0,4	< 0,3	< 0,1
	Mirístico - C14:0	< 0,2	< 0,5	< 0,1	< 0,5
	Palmítico - C16:0	2,5-6,5	3,0-10,0	9,0-14,0	7,0-14,0
	Palmitoléico - C16:1	< 0,6	< 1,0	< 0,5	< 0,5
	Esteárico - C18:0	0,8-3,0	1,0-10,0	0,5-4,0	1,4-5,5
	Oléico (ω 9) - C18:1	53,0-70,0	14,0-35,0	24,0-42,0	19,0-30,0
	Linoléico (ω 6) - C18:2	15,0-30,0	55,0-75,0	34,0-62,0	44,0-62,0
	Linolênico (ω 3) - C18:3	5,0-13,0	< 0,3	< 2,0	4,0-11,0
	Araquídico - C20:0	0,1 - 1,2	< 1,5	< 1,0	< 1,0
	Eicosenóico - C20:1	0,1 - 4,3	< 0,5	< 0,5	< 1,0
	Behênico - C22:0	< 0,6	< 1,0	< 0,5	< 0,5
	Lignocérico - C24:0	< 0,2	< 0,5	< 0,5	—
	Erúcido- C22:1	< 2,0	< 0,5	—	—
	Tetracosenóico - C24:1	< 0,2	< 0,5	—	—

1.5 Controle de qualidade de óleos vegetais

A comunidade científica mundial tem proposto diversos métodos para o controle de qualidade de óleos vegetais. Critérios de qualidade químicos e físico-químicos têm sido determinados por espectrometria no infravermelho próximo (*Near Infrared* — NIR) e no médio (*Mid Infrared* — MIR) (HENDL et al., 2001, FRANCO et al., 2006, BARTHUS e POPPI, 2001, CHRISTY e EGEBERG, 2006, WANG et al., 2006, GUILLÉN e CABO, 1997). Também são empregadas outras técnicas como cromatografia (MALAVIA, 2007), fluorescência (SIKORSKA et al., 2005, POULLI et al., 2005, POULLI et al., 2006) e quimiluminescência (STEPANYAN et al., 2005, PAPADOPOULOS, 2002).

Métodos quimiométricos têm sido comumente aplicados a dados cromatográficos e espectroscópicos com a finalidade de comparação, discriminação e classificação na avaliação da autenticidade de óleos vegetais. Neste contexto, as técnicas estatísticas multivariadas mais utilizadas são: *k*-vizinhos mais próximos (*k-Nearest Neighbours* — KNN) (SIKORSKA, 2005), análises de discriminante (*Discriminant Analysis* — DA) (SIKORSKA et al., 2005, LANKMAYR et al., 2004, MARTÍN et al., 1999, LAI et al., 1994, BAETEN et al., 1998, VONCINA et al., 2005, NAGY et al., 2005, LEE et al., 2001, KEMSLEY, 1996), análise de agrupamentos hierárquicos (*Hierarchical Cluster Analysis* — HCA) (POULLI et al., 2005), análise de componente principal (*Principal Component Analysis* — PCA) (POULLI et al., 2005, DIAZ et al., 2005, LEE et al., 1998, LANKMAYR et al., 2004, LAI et al., 1994, LEE et al., 2001, KEMSLEY, 1996), modelagem flexível independente por analogia de classe (*Soft Independent Modelling of Class Analogy* — SIMCA) (DIAZ et al., 2005), análise de variáveis canônicas (*Canonical Variables Analysis* — CVA) (ALVES et al., 2005), mínimos quadrados parciais (*Partial Least Squares* — PLS) (CHRISTY et al., 2004, MARQUEZ et al., 2005, KEMSLEY, 1996), redes neurais artificiais (*Artificial Neural Networks* — ANN) (VONCINA et al., 2005, HUSAIN et al., 1996) e algoritmo das projeções sucessivas (*Successive Projection Algorithm* — SPA) (PONTES et al., 2005).

A maioria das técnicas citadas acima possui como desvantagem o alto custo dos equipamentos e manutenção, quando comparadas com a voltametria de onda quadrada (*Square Wave Voltammetry* – SWV) que utiliza instrumentação simples e de custo relativamente baixo.

1.6 Voltametria de onda quadrada

A SWV foi criada na década de 1950 quando Barker aplicou pulsos de potencial a um eletrodo gotejante de mercúrio e chegou à polarografia de onda quadrada (BARD e FAULKNER, 2001, SOUZA et al., 2003).

Ramaley e Krause basearam-se nos trabalhos de Barker para desenvolver a teoria da voltametria de onda quadrada utilizando eletrodos estacionários com o intuito de eliminar os ruídos provenientes do capilar de mercúrio. A partir daí, a SWV pôde ser utilizada para a análise de traços com a vantagem da possibilidade de substituição do eletrodo de mercúrio por um eletrodo sólido (RAMALEY e KRAUSE, 1969, KRAUSE e RAMALEY, 1969). Nestes trabalhos, os autores se limitaram a saltos de pequeno incremento de potencial, resultando em baixas velocidades de varreduras.

As limitações da metodologia desenvolvida por RAMALEY e KRAUSE (1969) foram analisadas por CHRISTIE et al. (1977) e TURNER et al. (1977), para então, chegar ao modelo atual da SWV, em que as leituras de correntes em função dos pulsos de potenciais aplicados são feitas a velocidades maiores.

A SWV consiste em aplicar, ao eletrodo de trabalho, dois pulsos de potenciais simétricos sobre cada degrau de uma escada de potencial. Um dos pulsos é aplicado no sentido da escada de potencial (direto) e o outro, no sentido reverso. A corrente é medida no final de cada pulso, fechando um ciclo de onda quadrada (SKOOG et al., 2002, WANG, 2006, BARD e FAULKNER, 2001).

A Figura 1.2 ajuda a definir os principais parâmetros da onda quadrada, que se caracteriza por uma amplitude de pulso, ΔE_p , medida

com relação a cada degrau da escada, e uma largura de pulso, t_p , dada em tempo e que também pode ser expressa em termos da frequência da onda quadrada, $f = 1/2t_p$. A varredura começa num potencial inicial, E_i , então, no final de cada ciclo a escada aumenta instantaneamente por um incremento de potencial, ΔE_S . Conseqüentemente, a velocidade de varredura, $v = \Delta E_S/2t_p$, pode ser dada por $v = \Delta E_S \times f$. O sinal é obtido como uma intensidade de corrente de forma diferencial entre as correntes nos sentidos direto e reverso, apresentando excelente sensibilidade e alta rejeição a correntes capacitivas (BARD e FAULKNER, 2001, SOUZA et al., 2003, SOUZA et al., 2004).

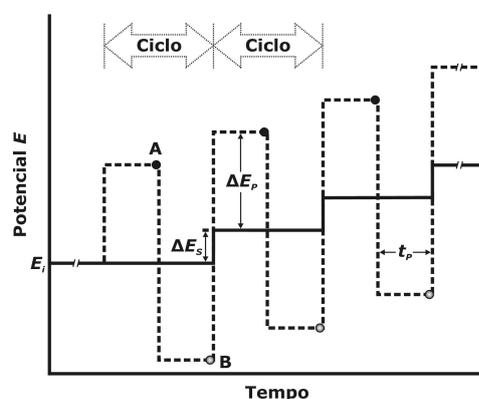


Figura 1.2 Variação de potencial em função do tempo para a SWV. Estão representados o potencial inicial, E_i , a amplitude de pulso, ΔE_p , o incremento de potencial, ΔE_S , a largura de pulso, t_p e os pontos A e B onde as correntes direta e reversa são medidas, respectivamente.

O pico voltamétrico resultante apresenta posição, largura e altura características do tipo de sistema redox (WANG, 2006, SOUZA et al., 2004). A Figura 1.3 apresenta os voltamogramas teóricos associados a um sistema reversível e irreversível com a separação observada das correntes direta, reversa e resultante.

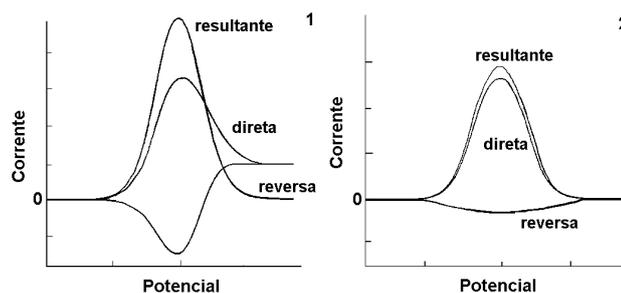


Figura 1.3 Voltamogramas esquemáticos de onda quadrada (SOUZA et al., 2003) para um processo redox de um sistema reversível (1) e de um sistema irreversível (2).

A principal vantagem da SWV é a possibilidade de se obter correntes de pico bem definidas em experimentos executados em alta velocidade de varredura, resultando em análises mais rápidas e com alta sensibilidade. Além disso, por tratar-se de uma técnica de pulso, a corrente faradaica pode ser registrada no final do pulso, instante no qual a contribuição da corrente capacitiva é minimizada (BARD e FAULKNER, 2001).

A voltametria ganhou grande importância na química analítica à medida que os equipamentos e as técnicas foram se aperfeiçoando (SKOOG et al., 2002). Entretanto, estas técnicas ficaram limitadas, principalmente, a aplicações em calibração univariada. Um perfil voltamétrico apresentando várias sobreposições de picos analíticos que, em geral, parece ser um grave problema, é na verdade um excesso de informações valiosas muito próximas umas das outras, que podem ser extremamente úteis quando métodos quimiométricos são empregados para extrair o máximo de informações possíveis (HERRERO e ORTIZ, 1999).

1.7 Quimiometria

O ascendente avanço tecnológico faz com que a química, sobretudo a química analítica moderna, ocupe um lugar de destaque na Ciência. Atualmente, existem diversas técnicas analíticas que podem gerar grande quantidade de dados em curto intervalo de tempo. Estes dados apresentam uma gama de informação de natureza multivariada e, muitas vezes, complexa, o que torna importante o uso de ferramentas que permitam extrair, de um modo mais criterioso, o maior número possível de informações relevantes. Neste contexto, torna-se extremamente útil o uso da Quimiometria (MASSART et al., 1997, KOWALSKI, 1984, BEEBE et al., 1998), que pode ser definida como um conjunto de métodos matemáticos e/ou estatísticos utilizados para planejar ou selecionar experimentos de forma otimizada e extrair o máximo de informação química dos dados obtidos (FERREIRA et al., 1999).

A quimiometria agrupa várias ferramentas extremamente úteis para planejamento e otimização de experimentos (PIMENTEL et al., 1998, HONORATO et al., 2002), processamento de sinais (GALVÃO et al., 2004, GALVÃO et al. 2001a), seleção de variáveis e amostras (PONTES et al., 2005, DANTAS-FILHO et al., 2004, ARAUJO et al., 2001, GALVÃO et al., 2001b), reconhecimento de padrões e classificação de amostras (PONTES et al., 2005, COSTA et al., 2004), calibração multivariada (DANTAS-FILHO et al., 2004, ARAUJO et al., 2001, GALVÃO et al., 2001b), transferência de calibração (HONORATO et al., 2005), entre outros.

Uma Tabela, representada por uma matriz $\mathbf{A}_{m \times n}$ (Figura 1.4) é utilizada para apresentar os dados de natureza multivariada. Nela, geralmente as amostras (objetos) estão localizadas nas linhas e as variáveis, que são medidas de alguma propriedade das amostras, são distribuídas nas colunas. No caso de dados voltamétricos, as variáveis são os valores de corrente registrados em cada potencial que compõe a janela de varredura.

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

Figura 1.4 Representação de uma matriz de dados de natureza multivariada.

1.7.1 Pré-processamento dos dados

Os dados obtidos de uma análise, também conhecidos como dados brutos, podem apresentar diversas fontes de variação causando irregularidades sistemáticas na linha de base, ruídos instrumentais ou variáveis de diferentes unidades e magnitudes. Então, antes da realização da modelagem multivariada propriamente dita, faz-se necessário um pré-processamento para remover as fontes indesejáveis de variações que possam atingir as variáveis e/ou amostras do conjunto de dados como um todo (MASSART et al., 1997, FERREIRA et al., 1999).

Existem basicamente três tipos de pré-processamento aplicáveis às variáveis: a centralização dos dados na média, que consiste na subtração dos elementos de cada linha pela média da sua respectiva coluna, o escalonamento, onde cada elemento de uma linha é dividido pelo desvio padrão da sua respectiva variável, fazendo com que todos os eixos da coordenada sejam conduzidos à mesma escala e, conseqüentemente, cada variável fica com o mesmo peso, e o auto-escalonamento, que consiste em centralizar os dados na média e efetuar o escalonamento. Desta forma, as variáveis terão médias zero e desvios padrões igual a um. Tanto o escalonamento quanto o auto-escalonamento são utilizados quando se pretende atribuir a mesma importância às variáveis do sistema de investigação (MASSART et al., 1997, BEEBE et al., 1998).

Além do pré-processamento dos dados aplicado às variáveis, devem-se considerar também as variações existentes entre as amostras. Para contornar tais problemas, métodos como a normalização, a derivação, a filtragem de ruído aleatório, entre outros são bastante utilizados. A normalização é aplicada quando se deseja remover variações sistemáticas presentes na amostra. Ela é realizada dividindo cada variável por uma constante que é escolhida a partir de uma análise preliminar dos dados. Os métodos de derivação são comumente aplicados quando, em um conjunto de dados, deseja-se corrigir problemas relacionados com a variação da linha de base. Além disso, o método da derivação é muito útil para permitir uma melhor visualização de picos existentes nos sinais originais. Contudo, uma avaliação prévia dos dados deve ser realizada, pois as respostas instrumentais são compostas de sinais verdadeiros e ruídos aleatórios que podem aparecer com baixa ou alta intensidade. Portanto, o método derivativo pode, em muitos casos, super-dimensionar a presença de eventuais ruídos presentes nas amostras (MASSART et al., 1997, BEEBE et al., 1998).

Vale salientar que, o uso de um ou outro método de pré-processamento, seja ele aplicado às amostras e/ou variáveis, deve ser previamente avaliado procurando sempre tomar as decisões mais

coerentes de acordo com a natureza dos dados. Para isto, é muito importante ter conhecimento aprofundado sobre o sistema em estudo.

1.7.2 Planejamentos experimentais

Diversas variáveis influenciam a qualidade dos dados e resultados obtidos por voltametria. Ainda assim, técnicas quimiométricas de planejamento experimental têm sido pouco exploradas na otimização dessas variáveis.

Planejamentos experimentais típicos podem ser aplicados no contexto da química eletroanalítica, às vezes, sendo necessário um ou mais planejamentos para atingir o objetivo final. Sempre que possível, os parâmetros destas técnicas como incremento de potencial e amplitude de pulso, entre outros no caso da voltametria, devem ser avaliados de modo a garantir que o método seja sensível, preciso e exato.

Por convenção, adotam-se as seguintes notações para o uso de planejamento experimental:

- Fatores: são as variáveis controladas
- Valor das variáveis (Níveis): são valores dos fatores compreendendo a região de investigação.
- Resposta do experimento: é o valor, alvo da otimização, obtido para cada experimento.

1.7.2.1 Planejamento composto central

O planejamento composto central é uma opção vantajosa ao planejamento fatorial completo de três ou mais níveis, por fornecer igual ou menor número de experimentos. Em geral, um planejamento composto central para k fatores é formado por três partes:

- Um planejamento fatorial completo ou fracionário de dois (2) níveis e k fatores, contendo $n_F = 2^k$ experimentos;
- Um planejamento axial ou estrela, formado por $2k$ experimentos e com todos os fatores nulos, exceto um, que é igual a um certo valor α ou $-\alpha$;
- Um ponto central, que deve ser repetido determinado número de vezes.

Desta forma, o total de experimentos distintos será: $n_F + 2k + 1$.

O valor da distância axial dos pontos no planejamento estrela, α , geralmente é estabelecido usando-se o conceito de rotabilidade em que $\alpha = \sqrt[4]{2^k}$. Um planejamento é “rodável” se a variância de suas estimativas só depender da distância em relação ao ponto central, ou seja, se a precisão da resposta prevista for a mesma em todos os pontos situados numa hipersfera com centro no próprio centro do planejamento. Esta propriedade esférica é um critério de projeto usado quando o planejamento fatorial é completo. No entanto, não é importante ter uma rotabilidade exata para se ter um bom planejamento. (BARROS-NETO et al., 2003, CALADO e MONTGOMERY, 2003).

1.7.2.2 Parâmetro de Taguchi

Taguchi (MASSART et al., 1997) interessou-se pela problemática de “como usar planejamento experimental para diminuir a dispersão em torno de um valor desejado e conseqüentemente aumentar a robustez de um produto”. Sua idéia consiste basicamente de dois elementos: 1) resposta e 2) planejamento experimental que permite a maximização da resposta.

A resposta é caracterizada pelo valor e sua dispersão quando o experimento é replicado. Neste caso, o valor da resposta e a dispersão são dados pelas Equações (1.1) e (1.2), respectivamente:

$$\bar{y} = \frac{\sum y_i}{n} \quad (1.1)$$

$$s^2 = \frac{\sum (y_i - \bar{y})^2}{(n - 1)} \quad (1.2)$$

onde \bar{y} é a média das respostas, chamada de sinal e s é a variância, chamada de ruído. A proposta de Taguchi (Equação 1.3) foi maximizar a relação sinal/ruído.

$$Z = 10 \times \log \left(\frac{\bar{y}^2}{s^2} \right) \quad (1.3)$$

Isto é equivalente a minimizar o desvio padrão relativo, ou seja, melhorar a precisão do método utilizado.

Uma vez que um método analítico é otimizado através de planejamentos estatisticamente bem estabelecidos, os sinais, por ele obtidos, tornam-se mais adequados ao emprego de ferramentas quimiométricas para a construção de modelos de classificação e/ou calibração multivariada.

1.7.3 Análise de Componentes Principais (PCA)

A PCA é uma técnica quimiométrica de reconhecimento de padrões não-supervisionado cujo princípio é fatorar um conjunto de dados de natureza multivariada, objetivando a representação das variações existentes na forma de componentes principais (PCs) (Figura 1.5). As PCs são as representações dos novos eixos formados a partir da combinação linear das variáveis originais e nelas está contida toda a informação dos dados na forma de variância. Elas são sempre ortogonais (perpendiculares) umas as outras, e sucessivas PCs descrevem quantidades decrescentes da variância explicada dos dados, de modo que a informação relevante está contida nas primeiras componentes. Por esta razão, poucas PCs são necessárias para que toda ou a maior parte da variância dos dados seja explicada, permitindo que os dados multivariados sejam interpretados em dimensionalidades plausíveis à percepção humana (MASSART et al., 1997, BEEBE et al., 1998, FERREIRA et. al., 1999).

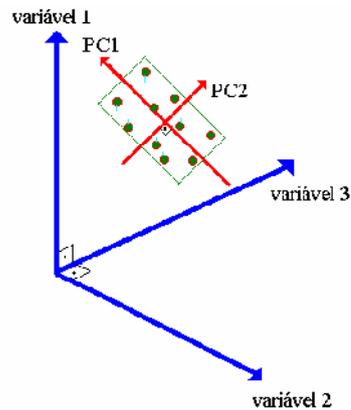


Figura 1.5 Primeira e segunda PCs no espaço tridimensional das variáveis.

A contribuição que cada variável original exerce sobre uma determinada PC é denominada peso, que pode ser definido, matematicamente, como sendo o cosseno do ângulo entre o eixo da variável e o eixo da PC. Da mesma forma que as amostras têm coordenadas definidas pelas variáveis originais, elas também as possuem em relação aos novos eixos (PCs), onde são denominadas escores das amostras. Um determinado número (menor do que o total) de PCs não consegue descrever toda a variância dos dados, então essa parte não descrita será representada em uma nova matriz chamada de resíduos (MASSART et al., 1997, BEEBE et al., 1998).

O número de PCs suficiente para descrever a informação dos dados é definido como sendo a dimensionalidade inerente (DI) ou posto da matriz. A DI deve-se às características físicas e/ou químicas encontradas no conjunto de dados, que estão mascaradas pela presença de correlação entre as variáveis e ruídos. Para que a DI seja alcançada, é imprescindível definir o número adequado de PCs que fornecem as informações de interesse para o analista, excluindo o máximo de ruído sem comprometer as informações pertinentes à melhor interpretação dos dados. Por isso, recorre-se ao uso de ferramentas de diagnóstico, que investigam e exploram três aspectos do conjunto de dados: o modelo, as amostras e as variáveis. Dentre estas ferramentas, pode-se citar o gráfico dos resíduos, o gráfico da variância explicada, gráfico dos escores e o gráfico dos pesos.

Outro ponto tão importante quanto à determinação do número de PCs é a detecção de amostras anômalas (*outliers*), que apresentam características diferentes e afetam os resultados obtidos. Estas amostras devem, portanto, ser removidas do conjunto de dados (MASSART et al., 1997, BEEBE et al., 1998). É importante ressaltar que o descarte de uma ou mais amostras do modelo deve ser efetivamente avaliado utilizando diversos tipos de ferramentas de diagnóstico. Desconsiderar a variabilidade existente do conjunto de dados como um todo e adotar, com pouca informação *a priori*, uma amostra como sendo *outlier* pode prejudicar o desempenho dos modelos.

PCA é uma técnica bastante utilizada como base para outras técnicas quimiométricas multivariadas como, por exemplo, o SIMCA, que é uma técnica de reconhecimento de padrão supervisionado.

1.7.4 SIMCA

O SIMCA é um método de reconhecimento de padrão supervisionado no qual cada classe é individualmente modelada por PCs (KOWALSKI, 1984). Para cada classe ou grupo de amostras previamente estabelecido, um modelo PCA é construído delimitando estatisticamente uma região espacial. Amostras futuras podem ser avaliadas como pertencentes à sua classe verdadeira se possuírem características que possam ser inseridas no espaço multidimensional de um dos agrupamentos (BEEBE et al., 1998, FLATEN et al., 2004), ou seja, a amostra é classificada como pertencente a uma dada classe se apresentar variância dentro de um valor crítico determinado pelo modelo. Obtém-se este valor em função do valor de z^2 (Figura 1.6) calculado pela Equação (1.4).

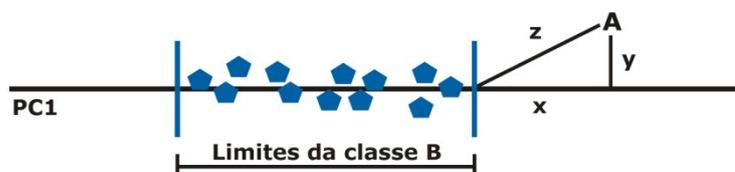


Figura 1.6 Previsão de uma amostra desconhecida A usando o modelo SIMCA da classe B construído com uma PC. A distância “y” entre a amostra A e o eixo da componente corresponde ao resíduo da PC1, “x” é a distância entre a projeção de A na direção da PC1 e a fronteira da classe e “z” é a distância de A até o limite da classe.

$$z^2 = x^2 + y^2 \quad (1.4)$$

O valor z^2 é convertido a uma variância e dividido pela variância das amostras da classe B para formar um valor F calculado, F_{cal} . Utiliza-se, então, o teste F , em que um valor crítico, F_{crit} , é escolhido empiricamente ou a partir de uma tabela do teste F . Se o F_{cal} for menor do que o valor crítico adotado, a amostra desconhecida pode ser classificada como pertencente à classe B (MASSART et al., 1997, BEEBE et al., 1998).

Uma das ferramentas utilizadas pelos pacotes quimiométricos para mostrar os resultados da classificação SIMCA é uma tabela de asteriscos, em que a presença dos mesmos representa que a amostra foi classificada como pertencente ao modelo elaborado. Outra forma é o gráfico bidimensional $S_i \times H_i$ (Figura 1.7). Este gráfico detalha os dois limites usados para a classificação: a distância da amostra em relação ao modelo da classe selecionado (raiz quadrada da variância residual, S_i) e a influência (Leverage, H_i) que é a distância das amostras projetada ao centro do modelo. Então, amostras que se encontrarem dentro de ambos os limites definidos pelo modelo específico, pertencerão à classe em questão (CAMO, 1998).

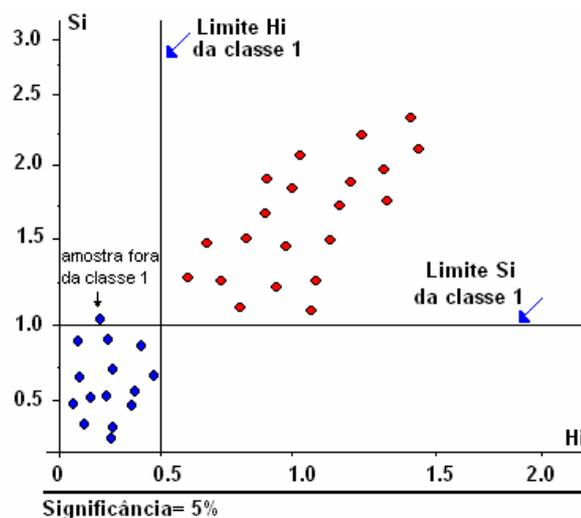


Figura 1.7 Gráfico de $S_i \times H_i$, detalhando os limites de uma determinada classe 1 e a disposição das demais amostras ao seu redor. Neste caso, uma das amostras da própria classe 1 foi mal classificada como visualizado.

Dois tipos de erros podem ocorrer em um estudo de classificação:

- Erro do Tipo I (Falso-Positivo): amostra avaliada não é classificada em sua própria classe.
- Erro do Tipo II (Falso-Negativo): amostra avaliada é classificada em uma classe errada.

Em SIMCA, a mesma amostra pode apresentar erros do Tipo I e do Tipo II simultaneamente, quando deixar de classificar em sua classe verdadeira e classificar em uma ou mais classes erradas.

1.7.5 Análise discriminante linear (LDA)

A análise discriminante linear (LDA) é uma técnica que consiste em estimar uma combinação linear de duas ou mais variáveis independentes obtendo uma função linear, que pode distinguir os pontos relativos aos dados de dois ou mais grupos. Se essa função de fato existe, então é possível dizer que os pontos pertencentes a esses grupos são linearmente separáveis. A discriminação é feita determinando o conjunto ótimo de pesos para as variáveis independentes de tal maneira que se maximize a variância entre os grupos em relação à variância dentro dos grupos. Uma das maneiras mais simples de se obter uma função linear discriminante (MASSART et al., 1997, KOWALSKI et al., 1984, BRUNS e FAIGLE, 1985) é ilustrada na [Figura 1.8](#).

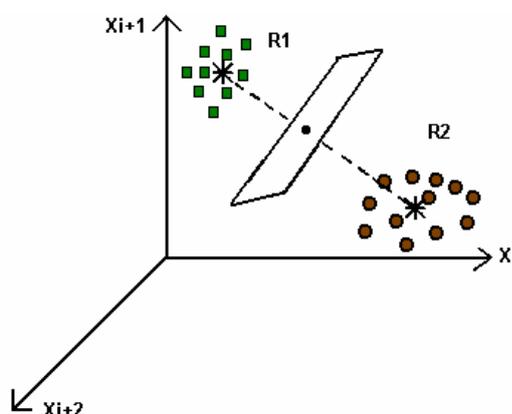


Figura 1.8 Representação esquemática da análise discriminante linear. Os centros de gravidade (os asteriscos) dos grupos formados por círculos e quadrados são inicialmente determinados. Em seguida traça-se um plano de decisão equidistante e perpendicular à linha que une os centros de gravidade dos grupos, formando as regiões R1 e R2. Se um ponto qualquer estiver presente na região R1, o mesmo é classificado como pertencente ao grupo 1. Caso contrário, obrigatoriamente pertencerá à categoria 2.

Vale salientar que em LDA, se a amostra apresentar erro do Tipo I, conseqüentemente ocorrerá erro do Tipo II (BRUNS e FAIGLE, 1985).

Diferentemente do SIMCA, método de classificação consolidado para trabalhar em altas dimensões, a LDA limita-se ao uso de dados com baixa dimensionalidade. Neste contexto, torna-se necessário o emprego de métodos de seleção de variáveis que são úteis para aumentar a capacidade preditiva dos modelos, eliminar variáveis não informativas e reter aquelas que, provavelmente, resultam em uma boa exatidão.

1.7.6 Seleção de variáveis

As técnicas quimiométricas tradicionalmente utilizadas no contexto da classificação possibilitam trabalhar com todo o conjunto de dados, reduzindo sua dimensionalidade, por meio de fatoração, para facilitar a visualização e interpretação destes dados. Por exemplo, PCA transforma as variáveis e distribui os objetos em novos sistemas de eixos ortogonais em função da maior variância dos dados (MASSART et al., 1997, BEEBE et al., 1998), permitindo a utilização destes com bastante eficiência em técnicas como SIMCA e LDA (DIAZ et al., 2005, PONTES et al., 2006, KEMSLEY, 1996). A principal desvantagem deste tipo de compressão de dados é a dificuldade de interpretações físicas ou químicas do sistema, tendo em vista que trabalha no domínio das variáveis transformadas. Vários algoritmos têm sido propostos para uma seleção mais conveniente de variáveis (GESTAL et al., 2004, WU e MASSART, 1997, STEEL et al., 2000), entre eles, pode-se destacar o SPA desenvolvido por ARAÚJO et al. (2001) para seleção de variáveis, no contexto da calibração multivariada, aplicada aos modelos de regressão linear múltipla (*Multiple Linear Regression* – MLR).

1.7.7 Algoritmo das projeções sucessivas (SPA)

A essência do SPA consiste em buscar um conjunto pequeno e representativo de variáveis com ênfase na minimização de colinearidade entre elas. Desta maneira, torna-se possível usar modelos MLR que,

embora relativamente simples e de fácil interpretação, são severamente afetados por problemas de colinearidade (ARAÚJO et al., 2001).

Os fundamentos teóricos, bem como os detalhes da rotina operacional do SPA têm sido extensivamente tratados em outros trabalhos (ARAÚJO et al., 2001, GALVÃO et al., 2001b, PONTES et al., 2005, DANTAS-FILHO et al., 2004, HONORATO et al., 2005). Apesar disto, é importante a apresentação de um exemplo ilustrativo no contexto da voltametria.

Considere um conjunto $J = 5$ variáveis voltamétricas: $X_1 = 400$ mV; $X_2 = 450$ mV; $X_3 = 500$ mV; $X_4 = 550$ mV e $X_5 = 600$ mV para a determinação de dois analitos ($A=2$) usando três amostras de calibração ($X_{cal} = 3$). Então, a rotina de operações do algoritmo para a seleção das variáveis é realizada na seguinte ordem:

- 1º) Para o cálculo das projeções, usa-se como vetor ($K(0)$) de partida aquele que possuir a maior norma. Contudo, os demais vetores são posteriormente testados e aqueles mais ortogonais serão incorporados para a formação de cadeias de variáveis candidatas (Figura 1.9).

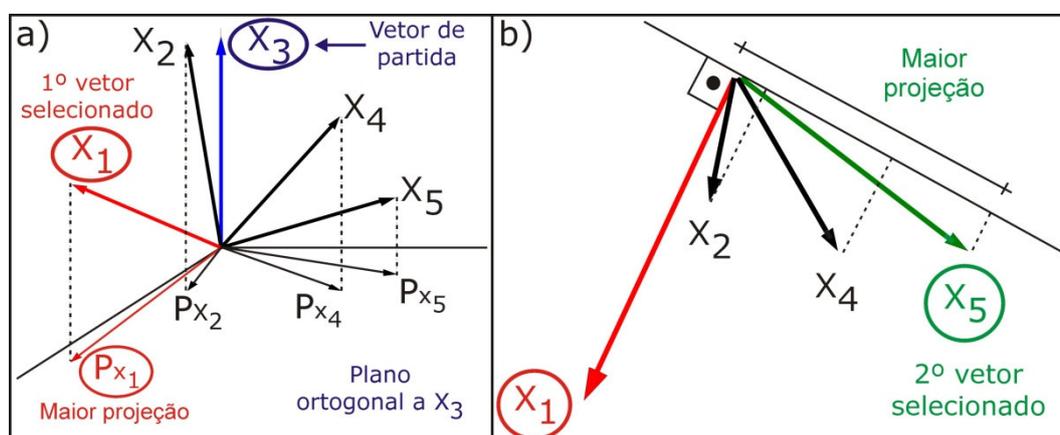


Figura 1.9 Exemplo do funcionamento do SPA para selecionar potenciais a partir de cinco variáveis voltamétricas: a) Projeções no plano ortogonal a X_3 ; b) Visão espacial dos vetores no plano horizontal.

- 2º) O programa gera $J \times [(X_{cal} - A) + 1]$ cadeias ($K(0)$, N), onde $K(0)$ pode variar de 1 até J , e N , número de variáveis em cada cadeia, pode variar de A até X_{cal} . Para este exemplo: serão geradas $5 \times [(3 - 2) + 1] = 10$ cadeias, ou seja, todas as possíveis combinações entre $K(0)$ e N :

$X_{1,2}$	$X_{2,2}$	$X_{3,2}$	$X_{4,2}$	$X_{5,2}$
$X_{1,3}$	$X_{2,3}$	$X_{3,3}$	$X_{4,3}$	$X_{5,3}$

- 3º) A primeira cadeia de variáveis a ser formada é composta pelo vetor de partida, neste caso X3 (500 mV), além das duas variáveis mais ortogonais X1 (400 mV) e X5 (600 mV).
- 4º) Para se determinar qual a melhor cadeia de variáveis, faz-se uma regressão por MLR para cada uma delas, empregando-se como matriz X as intensidades de corrente da cadeia de variáveis e como matriz Y as concentrações dos dois analitos avaliados. Realiza-se, através da Equação (1.5), o cálculo do erro quadrático médio de previsão (RMSEP) de amostras de um conjunto de teste para cada um dos modelos MLR construídos.

$$RMSEP = \sqrt{\frac{1}{M} \sum_{m=1}^M (y_m - \hat{y}_m)^2} \quad (1.5)$$

Em que y^m e \hat{y}^m são os valores de referência e previsto pelo modelo, de um parâmetro de interesse. O algoritmo escolhe a cadeia de variáveis cujo modelo MLR resulta no menor RMSEP.

Vale salientar que o número de variáveis selecionadas em cada cadeia, N, é limitado pela quantidade de amostras do conjunto de calibração para evitar problemas com a inversão de matrizes nos modelos MLR.

1.7.8 SPA no contexto da classificação

Recentemente, [PONTES et al. \(2005\)](#) adaptaram o SPA para o contexto dos métodos de reconhecimento de padrão. Neste caso, a informação da modelagem consiste nos dados das respostas instrumentais (matriz X) e um vetor contendo o índice de classe para cada amostra. Como se trata de classificação, os valores de referência dos y parâmetros não são disponíveis para processar a estatística do RMSEP que é a função de custo adotada pelo SPA-MLR para a seleção das variáveis ([DANTAS-FILHO et al., 2004](#)). Portanto, quando um subconjunto de variáveis sob avaliação é empregado, a função de custo adotada é o risco médio G de uma classificação incorreta, sendo este um critério da Análise Discriminante Linear.

Esta função é calculada no conjunto de teste como:

$$G = \frac{1}{Kv} \sum_{k=1}^{Kv} g_k \quad (1.6)$$

Onde Kv é o número de amostras do conjunto de teste e g_k é risco de uma classificação incorreta da K^{th} validação da amostra X_k . Logo, g_k é definido como:

$$g_k = \frac{r^2(X_k, \mu_{Ik})}{\min_{Ij \neq Ik} r^2(X_k, \mu_{Ij})} \quad (1.7)$$

Na Equação (1.7), o numerador $r^2(X_k, \mu_{Ik})$ é o quadrado da distância de Mahalanobis entre a amostra X_k (de índice de classe Ik) e a amostra média, μ_{Ik} , de sua classe verdadeira. Esta distância é calculada como:

$$r^2(X_k, \mu_{Ik}) = (X_k - \mu_{Ik}) \sum^{-1} (X_k - \mu_{Ik})^T \quad (1.8)$$

Em que μ_{Ik} e Σ correspondem à amostra média e a covariância, respectivamente, e são calculadas com os dados das amostras do conjunto de calibração. O denominador corresponde ao quadrado da distância de Mahalanobis entre a amostra X_k e o centro da classe incorreta mais próxima. Idealmente, g_k é o maior valor possível, de modo a validar melhor os limites entre as classes. Portanto, será escolhida a cadeia de variáveis que resultar no menor risco médio G de uma classificação incorreta.

1.7.9 Quimiometria aplicada à voltametria

HERRERO e ORTIZ (1999) avaliaram os aspectos qualitativos e quantitativos da aplicação do algoritmo genético (*Genetic Algorithm* – GA) no contexto da seleção de variáveis em alguns exemplos envolvendo dados polarográficos e voltamétricos. Em todos os casos o desempenho dos modelos PLS com e sem seleção de variáveis pelo GA foi avaliada em função da média do erro absoluto relativo para a previsão. No exemplo da

determinação de benzaldeído por DPP, a média do erro absoluto relativo para o modelo PLS foi de 2,10% enquanto para o modelo GA-PLS foi de 1,85%.

[BESSANT e SAINI \(1999\)](#) desenvolveram uma metodologia para determinação simultânea de glicose, frutose e etanol empregando como eletrodo de trabalho um eletrodo de platina não modificado, a voltametria de escada de pulso duplo (*Dual Pulse Staircase Voltammetry – DPSV*) e ANN para a calibração multivariada. Os voltamogramas utilizados são resultados da subtração da resposta do branco, que é o voltamograma do próprio eletrólito suporte ($\text{NaOH} - 0,1 \text{ mol L}^{-1}$). Posteriormente, os mesmos autores ([BESSANT e SAINI, 2000](#)) testaram estes dados empregando outras técnicas de calibração multivariada, tais como: MLR, regressão por componentes principais (*Principal Component Regression – PCR*) e PLS. Ainda assim, os resultados não superaram os obtidos com ANN em virtude do caráter não linear do etanol e de sua interação com frutose.

[PALACIOS-SANTANDER et al. \(2003\)](#) aliaram a técnica ANN a métodos de redução de dimensionalidade matricial, tais como: transformadas de Fourier (*Fourier Transform – FT*) e Wavelet (*Wavelet Transform – WT*) para resolver problemas de sobreposição de sinais eletroquímicos obtidos pela voltametria de pulso diferencial (*Differential Pulse Voltammetry – DPV*). Neste estudo, utilizou-se um eletrodo de gota pendente de mercúrio (*Hanging Mercury Drop Electrode – HMDE*) para a determinação simultânea dos íons Ti^+ e Pb^{2+} . Este método foi comparado com as técnicas PLS e PCR e, baseando-se na raiz do erro médio quadrático (*Root Mean Square Error – RMSE*), estabeleceu-se que o modelo proposto é mais eficiente. Paralelamente, [COCCHI et al. \(2003\)](#) utilizaram esta mesma metodologia e os mesmos analitos para comparar os resultados da calibração multivariada empregando PLS, MLR e ANN aos coeficientes wavelets gerados após a compressão dos voltamogramas. Os melhores resultados foram obtidos por MLR.

A DPV também foi utilizada para elaboração de um método capaz de determinar pesticidas organofosforados, utilizando HMDE e quimiometria. Devido à alta sobreposição dos voltamogramas para estes compostos, a quantificação foi realizada empregando comparativamente as técnicas PCR, PLS, mínimos quadráticos clássicos (*Classical Least Square* – CLS), Kalman filter (KF) e ANN. Recuperação satisfatória para um conjunto de amostras sintéticas foi obtida por ANN (NI et al., 2004).

NI et al. (2001a) desenvolveram um método utilizando HMDE e DPV para determinação simultânea de nitrobenzeno, 2-, 3-, 4-nitrofenol e 2,4-dinitrofenol com PLS, PCR e CLS. Foram alcançados erros relativos de previsão de $\pm 10\%$ para PLS e PCR e $\pm 38\%$ para CLS. Posteriormente, esta metodologia foi utilizada na construção de um modelo de calibração multivariada para quatro pesticidas (carbamatos) incluindo, desta vez, o uso de ANN. Neste caso, o erro relativo de previsão foi de 5,6% (NI et al., 2005). Este mesmo grupo utilizou eletrodo de carbono vítreo e as técnicas mais comuns de voltametria e calibração multivariada para determinação simultânea dos antioxidantes hidroxianisol butilado, hidroxitolueno butilado, propil galato e tert-butilhidroquinona (NI et al., 2000) e, dos antipsicóticos cloridrato de clorpromazina e cloridrato de prometazina (NI et al., 2001b).

Ácido nalidíxico (antibacteriano) e seu principal metabólico, ácido 7-hidroximetilnalidíxico, foram determinados em urina aplicando ANN e PLS aos voltamogramas obtidos por meio da polarografia de pulso diferencial (*Differential Pulse Polarography* – DPP), SWV e métodos de corrente contínua (*Sampled DC*). Modelo PCR foi construído e os escores obtidos foram utilizados para o treinamento das redes neurais. Os melhores resultados foram obtidos com ANN e DPP (GUIBERTEAU et al., 2004).

CUKROWSKA et al. (2001) desenvolveram um método para determinação simultânea de adenina e citosina utilizando padrões destes analitos para elaborar modelos de calibração multivariada empregando ANN. Voltametria de varredura linear (*Linear Sweep Voltammetry* – LSV) e DPP foram testadas neste trabalho. A técnica LSV apresentou melhores

resultados, com erros absolutos de 3,7% e 5,2% para adenina e citosina, respectivamente.

[BARTHUS et al. \(2005\)](#) aplicaram o método PLS aos dados obtidos por SWV com eletrodo de carbono vítreo para misturas sintéticas de ácido ascórbico e b-nicotinamida adenina dinucleotídeo (NADH). Foram obtidos erros relativos máximos de 1,7% e 2,1% para NADH e ácido ascórbico, respectivamente.

PLS foi utilizado para quantificação de benzaldeído empregando a DPP com intuito de determinar a capacidade de discriminação. Neste caso, a calibração univariada não pôde ser aplicada devido ao aumento característico do pico com a concentração e as leves oscilações do pH ([SANZ et al., 2001](#)).

[SAURINA et al. \(2000\)](#) utilizaram um eletrodo de grafite-metacrilato, voltametria cíclica (*Cyclic Voltammetry* – CV) e as técnicas quimiométricas PCR e PLS para a determinação simultânea de aminoácidos oxidáveis (cisteína, tirosina e triptofano). Desvios-padrão relativos de 10% foram obtidos para as previsões de cisteína e triptofano e 18% para tirosina. Posteriormente, este mesmo grupo ([SAURINA et al., 2004](#)) desenvolveu uma nova metodologia utilizando DPV e um eletrodo de platina para a quantificação dos mesmos aminoácidos. Esta estratégia apresentou um bom compromisso entre sensibilidade, seletividade e reprodutibilidade para estes analitos.

Vinhos, cervejas, cafés, leites e sucos de fruta foram analisados por SWV com eletrodo de platina, que passou por um procedimento de limpeza por ultra-som antes de cada amostra. PCA e HCA foram utilizadas como técnicas de reconhecimento de padrão. Para o emprego de calibração multivariada por PCR e PLS, uma amostra de suco de laranja foi diluída sucessivas vezes para construir os conjuntos de amostras necessários. Os resultados em termos de RMSEP foram de 4,0% e 5,4% para PCR e PLS, respectivamente ([SCHREYER e MIKKELSEN, 2000](#)).

Um arranjo de microeletrodos de carbono vítreo reticulado foi desenvolvido para a quantificação e classificação de bactérias utilizando

voltametria de pulso normal (*Normal Pulse Voltammetry* – NPV), SIMCA e PLS. Resultados aceitáveis foram obtidos em calibração multivariada transformando a variável dependente em escala logarítmica ($\log(1+y_i)$) e auto-escalando todas as variáveis independentes. (Berrettoni et al., 2004a, Berrettoni et al., 2004b).

Um algoritmo para resolução de curva multivariada (*Multivariate Curve Resolution* – MCR) por mínimos quadrados alternados (*Alternating Least Square* – ALS) tem sido utilizado com bastante eficiência para resolução de sobreposição de picos em sinais eletroanalíticos, inclusive, servindo como ferramenta para investigar o processo eletroquímico envolvido (ESTEBAN et al., 2000, CRUZ et al., 2000, DÍAZ-CRUZ et al., 2000, DÍAZ-CRUZ et al., 20001, FERNÁNDEZ et al., 2001, DÍAZ-CRUZ et al., 2002, ANTUNES et al., 2002, CRUZ et al., 2002).

DIAZ et al. (2004) avaliaram as técnicas DPV, SWV e DC com intuito de desenvolver uma metodologia para a determinação de tocoferol (vitamina E) em amostras de óleos vegetais (canola, milho, soja e oliva) usando PLS. Utilizou-se um eletrodo de carbono vítreo e o voltamograma do branco (eletrólito de suporte) foi registrado antes de cada medida. Os melhores resultados foram alcançados pelo PLS-DPV, quando comparados com os obtidos por cromatografia.

ESTEBAN et al. (2006a e b) publicaram duas revisões reunindo o uso de dados voltamétricos com diferentes aplicações quimiométricas: otimização e planejamento experimental, preparação e transformação de dados para tratamentos posteriores, exploração de dados, classificação de amostras e calibração multivariada.

Recentemente, JAKUBOWSKA (2007) desenvolveu uma nova função wavelet mãe dedicada ao pré-processamento de voltamogramas levando em consideração seu perfil, bem como as necessidades básicas para a análise desse tipo de sinal como: redução de ruído experimental, correção da linha de base e sobreposição de picos.

A revisão literária até aqui realizada permitiu observar que poucos trabalhos têm sido desenvolvidos explorando a união de técnicas

voltamétricas com a análise multivariada, sobretudo no contexto dos métodos de classificação de amostras. Diante disto, a SWV, técnica voltamétrica caracterizada por apresentar uma boa sensibilidade em alta velocidade de varredura, combinada com métodos de reconhecimento de padrão e seleção de variáveis pode ser uma alternativa promissora para classificação de diferentes tipos de óleos vegetais refinados.

CAPÍTULO II

OBJETIVOS

2.1 Objetivo Geral

Neste trabalho é proposto um método rápido, simples e barato para a classificação e fiscalização de óleos vegetais comestíveis quanto ao prazo de validade, que combina a voltametria de onda quadrada (SWV) com análise discriminante linear (LDA) associada ao algoritmo das projeções sucessivas (SPA) para a seleção de variáveis.

2.1.1 Objetivos Específicos:

- i) Utilizar técnicas de planejamento experimental para otimizar as condições de análise em SWV.
- ii) Utilizar a combinação da SWV, cuja instrumentação e manutenção não são dispendiosas quando comparadas às demais técnicas freqüentemente adotadas (cromatografia e espectroscopia), com métodos de reconhecimento de padrão supervisionado (SIMCA e LDA) e seleção de variáveis.
- iii) Aplicar PCA a partir dos voltamogramas visando à discriminação de amostras de óleos vegetais refinados não vencidos e vencidos.
- iv) Elaborar e validar modelos SWV-SIMCA para classificação de óleos vegetais não vencidos e vencidos.
- v) Utilizar a LDA com seleção prévia de variáveis (potenciais voltamétricos) pelo SPA para a classificação de óleos vegetais refinados.
- vi) Comparar o desempenho dos modelos SWV-SIMCA e SWV-LDA-SPA em termos do número de erros de classificação.

CAPÍTULO III

EXPERIMENTAL

3.1 Amostras

Cento e quatorze amostras de óleos vegetais refinados, de diferentes lotes e fabricantes, foram adquiridas em estabelecimentos comerciais da cidade de João Pessoa, Paraíba. Dessas amostras, 66 são aqui denominadas não vencidas, pois foram adquiridas no período de agosto a setembro de 2006 e estavam dentro do prazo de validade, e 48 são amostras com o prazo de validade expirado (PVE), que foram adquiridas em 2004. As classes e a quantidade de amostras analisadas são detalhadas na [Tabela 3.1](#).

Tabela 3.1 Classes e quantidade de amostras de óleos vegetais analisados.

Classes de Óleo	Número de Amostras
Canola	15
Girassol	16
Milho	17
Soja	18
PVE	48

3.2 Divisão do conjunto de amostras

Antes da construção dos modelos quimiométricos, as amostras de cada uma das classes de óleos delineadas na [Tabela 3.1](#) foram divididas em três conjuntos (treinamento, teste e previsão) utilizando o algoritmo KS ([KENNARD e STONE, 1969](#)), que é usado para extrair um conjunto representativo de amostras de um determinado conjunto de dados. Inicialmente, o KS foi aplicado de modo a gerar o conjunto de treinamento de cada classe. Em seguida, o KS foi aplicado às amostras remanescente de cada classe de modo a gerar o conjunto de teste. As amostras restantes foram então usadas como amostras do conjunto de previsão. O número de amostras de cada classe e de cada conjunto é mostrado na [Tabela 3.2](#).

Tabela 3.2 Número de amostras de treinamento, teste e previsão selecionadas pelo KS para as cinco classes de óleos vegetais.

Classes	Conjuntos		
	Treinamento	Teste	Previsão
Canola	9	3	3
Girassol	10	3	3
Milho	10	3	4
Soja	10	4	4
PVE	20	10	18

3.3 Materiais e reagentes

Todos os reagentes utilizados foram de grau analítico P.A. e são listados abaixo juntamente com os materiais:

- Álcool etílico absoluto (ISOFAR);
- Hidróxido de sódio (MERCK KGaA);
- Ácido acético (MERCK KGaA);
- Água deionizada Milli-Q (Millipore Inc);
- Pipetas automáticas de 2-20 e 20-200 μL (Eppendorf) e de 1,00-10,0 mL (VOLAC – John Poulten Ltd);
- Ponteiros (Eppendorf) descartáveis de polietileno;
- Vidraria;
- Tubos de polipropileno de 1,5 mL (Eppendorf).

3.4 Equipamentos

A parte experimental foi realizada empregando a SWV em um estudo eletroanalítico que possibilitasse a elaboração e avaliação de modelos quimiométricos de classificação de óleos vegetais refinados. As medidas foram realizadas em um potenciostato/galvanostato $\mu\text{AutoLab}^{\text{®}}$ Type II da Eco Chemie acoplado ao módulo polarográfico 663 VA Stand[®] da Metrohm (Figura 3.1), equipado com um eletrodo de referência de Ag/AgCl – KCl (3.0 mol L^{-1}) e um fio de platina como contra eletrodo. O eletrodo de trabalho foi escolhido entre os eletrodos de disco de platina e

de ouro e, todos os parâmetros foram otimizados com base em planejamentos experimentais.

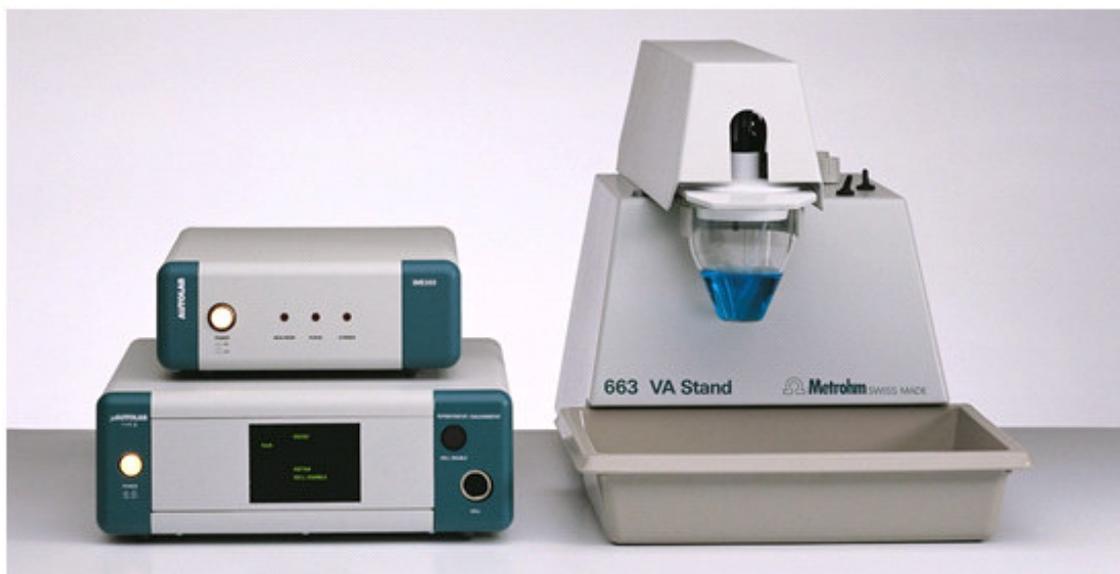


Figura 3.1 Potenciostato/Galvanostato μ AUTOLAB[®] Type II à esquerda e o módulo 663 VA Stand[®] da cela eletroquímica à direita.

3.5 Procedimento analítico

Os parâmetros da voltametria de onda quadrada selecionados para os estudos preliminares foram os seguintes:

- Potencial - tempo de condicionamento: 0,7 V – 3 s (sob agitação);
- Potencial - tempo de deposição: -0,9 V – 3 s (sob agitação);
- Tempo de equilíbrio: 10s;
- Frequência: 20 Hz;
- Janela de varredura de potencial: -0,9 a 0,6 V (eletrodo rotatório de disco de platina VS. Ag/AgCl);
- Incremento de potencial: 5 mV;
- Amplitude de pulso: 25 mV.
- Voltamogramas registrados em triplicata.

Os dois primeiros parâmetros (tópicos) foram escolhidos com base no trabalho realizado por [BESSANT e SAINI \(1999\)](#) que desenvolveram um método para determinação eletroquímica multivariada de compostos alifáticos com eletrodo de disco de platina, NaOH como eletrólito de suporte e adaptaram uma etapa de limpeza eletroquímica do eletrodo de trabalho.

3.5.1 Manipulação das amostras

Cada amostra passou por um processo de extração de ácidos graxos (AcGs) adaptado do método oficial CA 5a-40, da AOCS (*The American Oil Chemists' Society*), que rege a quantificação de AcGs livres em óleos vegetais refinados por titulação. Para permitir tal procedimento, o óleo é previamente misturado com etanol na proporção 1:1 m/v.

Devido à alta sensibilidade da SWV, pequenas quantidades de amostras são necessárias. Assim, a extração foi realizada misturando-se, sob agitação mecânica, alíquotas de 200 μL de óleo com 200 μL de etanol em tubos de propileno de 1,5 mL. Em seguida, a mistura foi mantida em repouso por aproximadamente trinta minutos até a formação de um sobrenadante translúcido (fase alcoólica - FA) contendo, predominantemente AcGs livres. Então, uma alíquota de 10 μL da FA foi adicionada à cela eletroquímica juntamente com 10 mL de NaOH 0,1 mol L⁻¹, agitando-se por 60 s com o eletrodo rotatório para homogeneizar. Nestas condições, o voltamograma de cada amostra foi registrado em triplicata. Um esquema ilustrando este procedimento é mostrado na Figura 3.2.

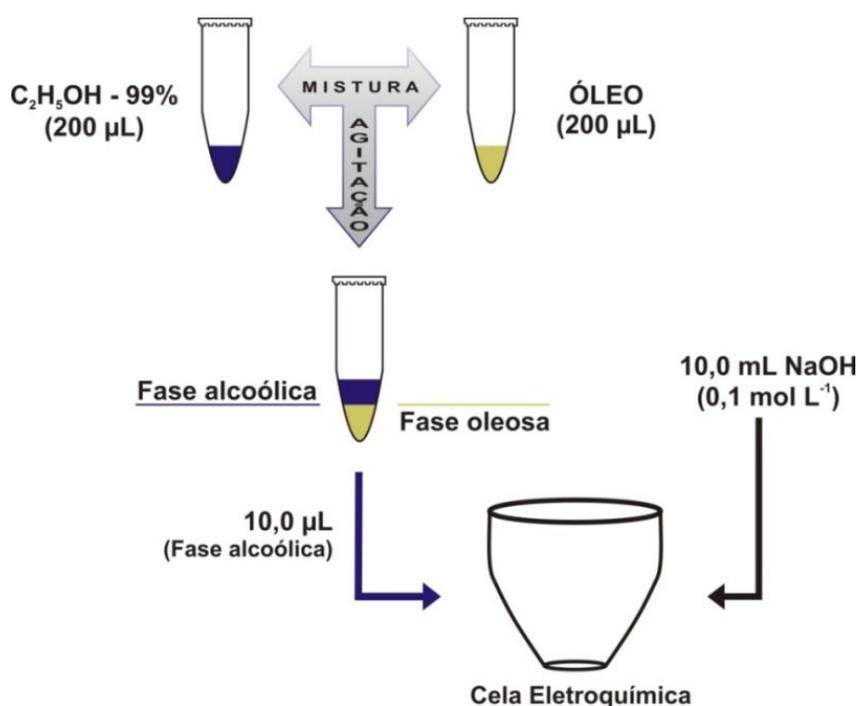


Figura 3.2 Esquema da extração de ácidos graxos das amostras de óleo e tomada de alíquota para realização das medidas.

O sucesso na manipulação das amostras depende da eficiência da etapa de extração alcoólica. Portanto, quatro diferentes proporções de óleo em álcool (1:4, 1:2, 1:1 e 1:½ v/v) foram avaliadas e foi escolhida, para dar continuidade ao trabalho, a que apresentou o perfil voltamétrico mais informativo.

3.5.2 Eletrólito suporte, eletrodo de trabalho e sentido da varredura

Dois tipos de eletrólito de suporte, o NaOH e CH₃COOH 0,1 mol L⁻¹; de eletrodo de trabalho, o eletrodo rotatório de disco (*Rotating Disc Electrode*) de platina, RDE-Pt, e de ouro, RDE-Au, e do sentido de varredura, catódico e anódico, foram estudados. Neste estudo e durante a realização de todas as medidas foram utilizados Ag/AgCl, como eletrodo de referência e um fio de platina como eletrodo auxiliar, trabalhando-se sempre à temperatura ambiente (25 °C) e sem desaerar o conteúdo da cela eletroanalítica.

3.5.3 Limpeza do eletrodo de trabalho e parâmetros da SWV

Uma das grandes vantagens de trabalhar com planejamentos experimentais é a possibilidade de uma melhor interpretação do sistema em estudo. Entretanto, eles foram utilizados neste trabalho somente para fornecerem a melhor combinação dos níveis e a seqüência de experimentos que deveriam ser realizados para otimizar a etapa de limpeza eletroquímica do eletrodo de trabalho e os parâmetros da SWV.

A limpeza eletroquímica do eletrodo (sempre imerso na solução do eletrólito suporte com amostra) consistiu de duas etapas. Na primeira, o eletrodo foi mantido num potencial anódico, PA, por um determinado tempo, t_A , para remover agentes contaminantes adsorvidos em sua superfície. Posteriormente, na segunda etapa, o eletrodo foi mantido num potencial catódico, PC, durante um tempo, t_C , para remover a camada de óxido de platina formada na primeira etapa. Em ambas as etapas, o eletrodo foi mantido sob agitação. Assim, regenerou-se a superfície do

eletrodo, deixando-o em condições para o registro dos voltamogramas sem agitação (BESSANT e SAINI, 1999, FUNG e MO, 1996). Os fatores PA, t_A , PC e t_C foram avaliados utilizando um planejamento composto central.

Freqüência (FR), incremento de potencial (IP) e amplitude de pulso (AP) são parâmetros que caracterizam a técnica da SWV. Teoricamente, existe um efeito de correlação intenso entre eles, de modo que é interessante estudá-los simultaneamente. Portanto, os fatores FR, IP e AP também foram avaliados utilizando um planejamento composto central.

Para a realização dos planejamentos experimentais, os valores dos pontos axiais (alfa), do ponto central e os níveis do planejamento cúbico foram codificados de acordo com a Equação (3.1):

$$x_i = \frac{v_i - \bar{v}}{\frac{\Delta v}{2}} \quad (3.1)$$

Onde x_i é o valor codificado; v_i o valor experimental do nível; \bar{v} o valor médio entre os níveis mais (1) e menos (-1), que é exatamente o valor do nível zero (0) e Δv é a diferença entre os níveis 1 e -1.

As medidas foram seqüencialmente efetuadas seguindo o procedimento operacional padrão mostrado abaixo:

- a. Limpar a cela;
- b. Passar água deionizada nos eletrodos e enxugá-los;
- c. Adicionar 10,0 mL do NaOH - 0,1 mol L⁻¹;
- d. Adicionar 10,0 µL da fase àlcoólica da extração 1:1 (óleo:etanol);
Obs.: Utilizar a mesma amostra para todos os experimentos.
- e. Agitar por 60 segundos com o eletrodo rotatório;
- f. Gerar uma varredura sem limpeza eletroquímica para manter uma poluição constante da superfície do eletrodo de trabalho;
Obs.: Este item é válido somente para a otimização da etapa de limpeza eletroquímica do eletrodo.
- g. Registrar os voltamogramas em triplicata, randomicamente, com os parâmetros de limpeza eletroquímica indicados pelo planejamento.

A resposta de cada experimento do planejamento foi o valor obtido a partir do parâmetro de Taguchi, Z (MASSART et al., 1997), mostrado na

Equação (1.3) e obtido através de modificações nas Equações (1.1) e (1.2).

Primeiramente, para cada experimento do planejamento, foram calculados o quadrado do voltamograma médio (Figura 3.3a) através da Equação (3.2) e o perfil da variância (Figura 3.3b) através da Equação (3.3).

$$\bar{I}_k = \frac{1}{n_{rep}} \sum_{i=1}^{n_{rep}} I_{i,k} \quad (3.2)$$

$$S_k^2 = \frac{1}{n_{rep} - 1} \sum_{i=1}^{n_{rep}} (I_{i,k} - \bar{I}_k)^2 \quad (3.3)$$

Onde $n_{rep} = 3$ é o número de replicatas do voltamograma para cada experimento; $I_{i,k}$ é o k-ésimo valor de corrente medido para o i-ésimo voltamograma replicado; \bar{I} e s^2 são as respectivas média da intensidade de corrente registrada e a variância para cada potencial varrido ao longo de cada ensaio.

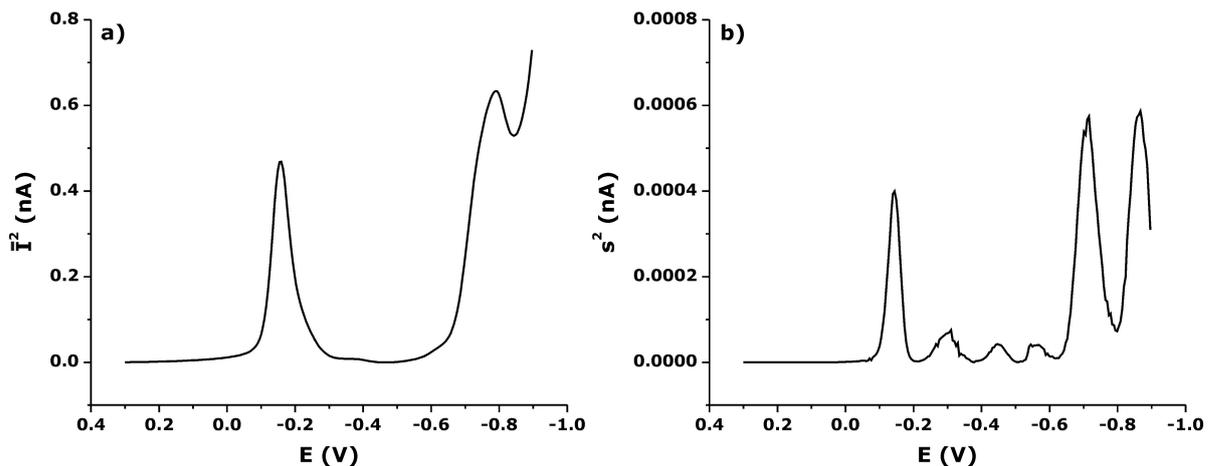


Figura 3.3 Quadrado do voltamograma médio (a) e perfil da variância (b).

Na seqüência, o voltamogram médio (Figura 3.3a) foi dividido pela variância (Figura 3.3b) através da Equação (3.4), calculando o parâmetro de Taguchi (Z) para cada variável.

$$Z_j = 10 \times \log \left(\frac{\bar{I}_j^2}{S_j^2} \right) \quad (3.4)$$

Finalmente, os valores de Z foram somados para calcular o Z_R (Equação 3.5), um número representativo da relação sinal/ruído de cada experimento e que será utilizado como resposta para o planejamento.

$$Z_R = \sum_j^{np} (Z_j) \quad (3.5)$$

Onde np é o número de potenciais varridos ou o número de variáveis independentes que compõem um voltamograma e Z_R é a resposta de cada experimento.

O fato do eletrodo de trabalho entrar em contato com o analito e sua matriz pode causar contaminações na superfície do eletrodo a cada medida, resultando em voltamogramas instáveis. Este problema é bastante conhecido, em meio aos Químicos Analíticos, como "efeito de memória", daí a importância de utilizar o parâmetro de Taguchi como resposta para minimizar a dispersão das medidas de corrente em cada potencial da janela de varredura.

3.6 Softwares

O programa Statistica[®] 6.0 foi empregado para a elaboração dos planejamentos experimentais. Pré-processamento dos dados, PCA e SIMCA foram realizados utilizando o programa Unscrambler[®] 9.7 (CAMO S.A.). Os algoritmos KS e LDA-SPA foram aplicados em Matlab[®] 6.5.

CAPÍTULO IV

RESULTADOS E

DISCUSSÃO

4.1 Eficiência da extração alcoólica

Quatro diferentes proporções de uma mesma amostra de óleo de soja em álcool etílico (1:4, 1:2, 1:1 e 1:½ v/v) foram testadas para avaliar a eficiência do procedimento de extração. Antes de registrar os voltamogramas de cada extração em estudo (10 µL da FA em 10 mL NaOH 0,1 mol L⁻¹), foi registrado o sinal do etanol (voltamograma para 10 µL de etanol P. A. em 10 mL de NaOH 0,1 mol L⁻¹). Os voltamogramas obtidos para a FA das extrações foram subtraídos do sinal do álcool para evidenciar as alterações causadas no sinal do etanol, pelos componentes do óleo presentes na FA (Figura 4.1). Assim, foi possível escolher a melhor proporção em função do perfil mais informativo.

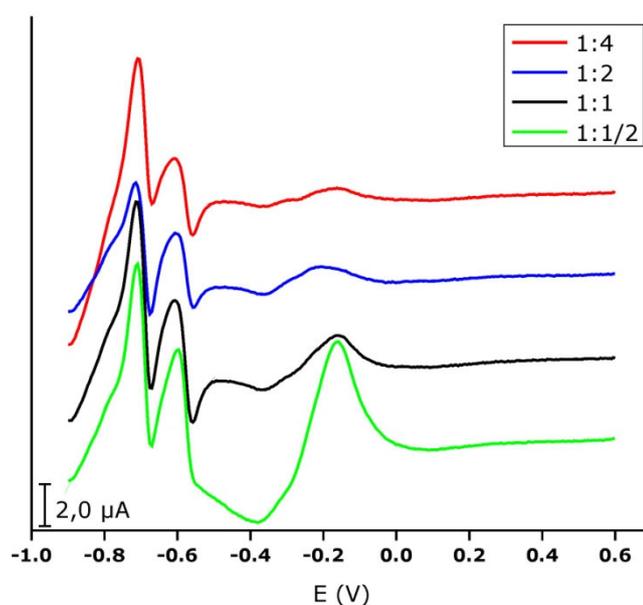


Figura 4.1 Perfil dos voltamogramas após a subtração do perfil de etanol P.A. no sentido de varredura anódico para 4 diferentes proporções de uma amostra de óleo de soja em etanol.

Apesar da proporção 1:½ (linha verde da Fig. 4.1) apresentar maiores valores de corrente ao longo de todo perfil, a região entre -0,570 e -0,400 V não tem informação relevante, diferentemente do que acontece no perfil das outras proporções. A proporção 1:1 de óleo em álcool para a extração dos AcGs livres foi escolhida porque o perfil obtido ao longo de todo o voltamograma (linha preta da Fig. 4.1) apresenta

sinais com mais informações (picos mais bem definidos nos voltamogramas).

Vale salientar que após o procedimento de extração, a FA contém, além de ácidos graxos livres, uma série de outros compostos como, por exemplo, triglicerídeos. SMITH e CIRCLE (1972) mostraram que a solubilidade do óleo de soja em álcool etílico absoluto à temperatura ambiente é de aproximadamente 5%. Como os demais óleos em estudo têm a composição relativamente semelhante ao de soja, pode-se concluir que eles apresentam solubilidade, mesmo que baixa, em etanol à temperatura ambiente.

4.2 Eletrólito suporte, eletrodo de trabalho e sentido da varredura

O principal objetivo da análise destes parâmetros foi investigar a condição experimental que permitisse, após a subtração do sinal de etanol P. A., a melhor relação entre sensibilidade e a presença de informação. Os resultados experimentais são mostrados na Figura 4.2.

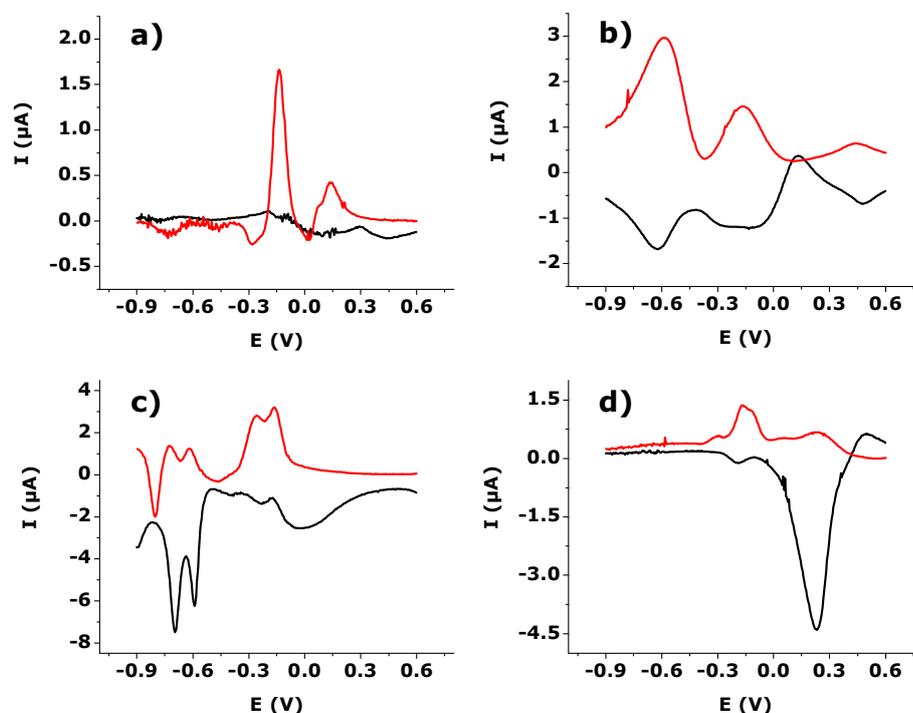


Figura 4.2 Perfil dos voltamogramas após a subtração do perfil de etanol P. A. nos sentidos de varredura anódica (linha preta) e catódica (linha vermelha), empregando RDE-Au e o eletrólito suporte NaOH (a) e o CH₃COOH (b); RDE-Pt e o eletrólito suporte NaOH (c) e CH₃COOH (d). A concentração dos eletrólitos de suporte foi 0,1 mol L⁻¹.

Os voltamogramas registrados em NaOH 0,1 mol L⁻¹ com o RDE-Pt no sentido catódico (linha vermelha da [Figura 4.2c](#)) apresentaram valores de corrente maiores e mais informativos na região entre os potenciais 0,0 e -0,500 V. Provavelmente, as respostas obtidas com o eletrólito NaOH foram as melhores porque ele facilita a limpeza eletroquímica do RDE-Pt em técnicas pulsadas, como também foi observado por [BESSANT e SAINI, \(2000\)](#).

Três diferentes concentrações de NaOH (0,1 mol L⁻¹, 0,01 mol L⁻¹ e 0,001 mol L⁻¹) foram testadas. Concentrações maiores não foram testadas de modo a evitar o ataque químico do NaOH ao eletrodo de referência cuja cápsula é de vidro. Observou-se que em concentrações de NaOH menores que 0,1 mol L⁻¹ ocorre perda de sinal e informação. Assim, todos os voltamogramas usados no tratamento quimiométrico foram registrados usando NaOH 0,1 mol L⁻¹ como eletrólito suporte.

Soluções de ácidos fortes não foram testadas para evitar a quebra das ligações insaturadas dos ácidos graxos.

Como pode ser observado nos voltamogramas apresentados nas [Figuras 4.1](#) (linha preta) e [4.2c](#) (linha vermelha), a região compreendida entre 0,300 e 0,600 V, independentemente do sentido da varredura, não apresenta perfil informativo. Portanto, a janela de varredura adotada neste trabalho foi sempre de 0,300 a -0,900 V (sentido catódico).

Os parâmetros otimizados até este momento no trabalho estão listados abaixo:

- Eletrólito suporte: NaOH 0,1 mol L⁻¹;
- Eletrodo de Trabalho: RDE-Pt
- Janela de varredura no sentido catódico: 0,300 a -0,900 V

4.3 Limpeza eletroquímica do eletrodo de trabalho

A essência de todo método voltamétrico é a parte sensora que entra sempre em contato com o analito, possibilitando interações químicas e

físicas com o mesmo. Portanto, para obter a maior precisão das medidas, é de fundamental importância que a superfície de resposta do eletrodo de trabalho seja reproduzida. Com esta finalidade, um planejamento experimental composto central 2^4 foi elaborado para otimizar a etapa de limpeza eletroquímica.

Os fatores PA, t_A , PC e t_C , já apresentados na [Seção 3.5.3](#), e os níveis utilizados para a elaboração do planejamento são mostrados na [Tabela 4.1](#).

Tabela 4.1 Detalhamento dos fatores e níveis utilizados para a construção do planejamento composto central, em que $\alpha = 1,5$.

Fatores	Níveis (nível codificado)									
	0,400 (- α)		0,500 (-1)		0,700 (0)		0,900 (1)		1,000 (α)	
PA (V)	0,400	(- α)	0,500	(-1)	0,700	(0)	0,900	(1)	1,000	(α)
t_A (s)	0	(- α)	1	(-1)	3	(0)	5	(1)	6	(α)
PC (V)	-1,2	(- α)	-1,1	(-1)	-0,900	(0)	-0,700	(1)	-0,600	(α)
t_C (s)	0	(- α)	1	(-1)	3	(0)	5	(1)	6	(α)

O valor do alfa ($\alpha = 1,5$) para determinar os pontos axiais do planejamento foi escolhido levando-se em consideração os limites da região de investigação, tendo em vista que, de acordo com estudos preliminares, o RDE-Pt não precisava permanecer mais do que seis segundos sob os potenciais aplicados e, estes não poderiam ser mais positivos ou mais negativos do que os adotados, pois poderiam provocar a formação de um excesso de óxido de platina e/ou desprendimento de hidrogênio, o que prejudica a precisão das análises.

O planejamento composto central 2^4 é mostrado na [Tabela 4.2](#). O maior valor de Z_R (8068,34) foi obtido pelo experimento 21, que indicou, então, os parâmetros ótimos ([Tabela 4.3](#)) para a etapa de limpeza eletroquímica. Vale salientar que, o maior valor de Z_R implica na minimização da dispersão dos valores de corrente em torno do voltamograma médio, o que leva a medidas mais precisas, sendo, portanto, um indício de que a superfície do eletrodo está sendo reproduzida.

Tabela 4.2 Planejamento composto central 2^4 para a otimização da etapa de limpeza eletroquímica do eletrodo de trabalho.

	Ensaio	PA	t_A	PC	t_C	Respostas (Z_R)
Planejamento Cúbico	1	-1	-1	-1	-1	6752,70
	2	1	-1	-1	-1	7574,18
	3	-1	1	-1	-1	7505,33
	4	1	1	-1	-1	7897,29
	5	-1	-1	1	-1	5709,19
	6	1	-1	1	-1	6062,23
	7	-1	1	1	-1	5752,70
	8	1	1	1	-1	6485,12
	9	-1	-1	-1	1	7003,26
	10	1	-1	-1	1	6973,48
	11	-1	1	-1	1	7084,23
	12	1	1	-1	1	7163,81
	13	-1	-1	1	1	6081,17
	14	1	-1	1	1	6374,20
	15	-1	1	1	1	7842,54
	16	1	1	1	1	7312,00
Pontos Axiais	17	-1,5	0	0	0	6791,33
	18	1,5	0	0	0	7053,75
	19	0	-1,5	0	0	6621,40
	20	0	1,5	0	0	6837,41
	21	0	0	-1,5	0	8068,34
	22	0	0	1,5	0	6407,56
	23	0	0	0	-1,5	6181,43
	24	0	0	0	1,5	6839,99
Pontos Centrais	25	0	0	0	0	6786,28
	26	0	0	0	0	6847,69
	27	0	0	0	0	6636,69

Tabela 4.3 Parâmetros ótimos escolhidos para a limpeza eletroquímica do eletrodo.

Parâmetros	Valores reais
Potencial Anódico	0,7 V
Tempo de permanência t_A	3 s
Potencial Catódico	-1,2 V
Tempo de permanência t_C	3 s

4.4 Otimização dos parâmetros da SWV

Diferentemente do que geralmente é feito em estudos eletroanalíticos, onde todos as variáveis experimentais são avaliadas individualmente, neste trabalho também foi utilizado um planejamento composto central para otimizar os parâmetros da SWV.

Os fatores FR, IP e AP, já apresentados na [Seção 3.5.3](#), e os níveis utilizados para a elaboração do planejamento experimental são mostrados na [Tabela 4.4](#).

Tabela 4.4 Detalhamento dos fatores e níveis utilizados para a construção do planejamento composto central 2^3 , em que $\alpha = 1,67$.

Fatores		Níveis (nível codificado)									
FR	(Hz)	9,95	($-\alpha$)	20,00	(-1)	35,00	(0)	50,00	(1)	60,05	(α)
IP	(mV)	1,00	($-\alpha$)	2,00	(-1)	3,50	(0)	5,00	(1)	6,01	(α)
AP	(mV)	16,63	($-\alpha$)	25,00	(-1)	37,50	(0)	50,00	(1)	58,38	(α)

O valor do alfa ($\alpha = 1,67$) para determinar os pontos axiais foi estabelecido porque o uso de velocidades de varredura (produto dos fatores IP e FR) acima de 360 mV s^{-1} prejudica o perfil dos voltamogramas. O planejamento experimental é mostrado na [Tabela 4.5](#).

Tabela 4.5 Planejamento composto central 2^3 .

	Ensaio	FR	IP	AP	Respostas (Z_R)
Planejamento Cúbico	1	-1	-1	-1	17290.8
	2	1	-1	-1	16527.4
	3	-1	1	-1	6953.8
	4	1	1	-1	7641.4
	5	-1	-1	1	18074.4
	6	1	-1	1	16160.1
	7	-1	1	1	6330.1
	8	1	1	1	7359.1
Pontos Axiais	9	-1,67	0	0	8900.2
	10	1,67	0	0	10493.1
	11	0	-1,67	0	32715.6
	12	0	1,67	0	6050.4
	13	0	0	-1,67	10122.3
	14	0	0	1,67	10552.3
Pontos Centrais	15	0	0	0	10720.4
	16	0	0	0	10676.4
	17	0	0	0	10176.6

Analogamente ao tópico anterior, os parâmetros ótimos (Tabela 4.6) foram aqueles provenientes do experimento 11 que forneceu o maior valor de Z_R (32715.6).

Tabela 4.6 Parâmetros escolhidos para a SWV.

Parâmetros	Valores reais
Frequência	35 Hz
Incremento de Potencial	1,0 mV
Amplitude de Pulso	37,5 mV

4.5 Seleção da região de trabalho no voltamograma

Os voltamogramas das 114 amostras de óleo vegetal refinado foram obtidos na faixa de 0,300 a -0,900 V. A [Figura 4.3](#) mostra o voltamograma de uma amostra de óleo (linha vermelha) que resulta da subtração do voltamograma originalmente obtido pela SWV (linha verde) menos o voltamograma do branco (NaOH 0,1 mol L⁻¹ – linha azul).

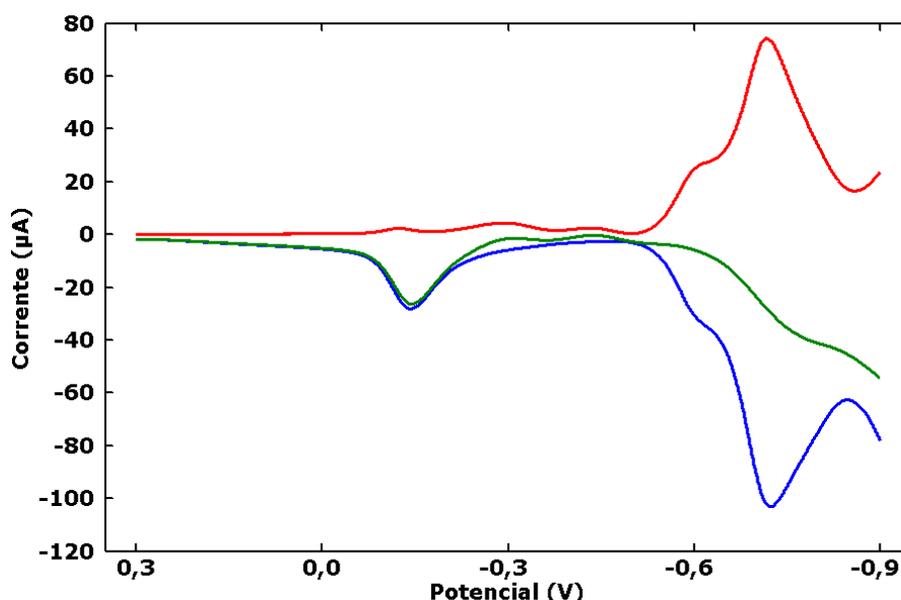


Figura 4.3 Voltamograma de uma amostra de óleo de canola (linha vermelha), na região de 0,300 a -0,900 V, após subtrair a resposta do branco (NaOH 0,1 mol L⁻¹ – linha azul) do voltamograma originalmente obtido pela SWV (linha verde).

Todos os voltamogramas obtidos apresentam 1124 variáveis (potenciais voltamétricos), para as quais os valores de corrente foram registrados. Para todas as amostras, as intensidades de corrente no intervalo de 0,300 a -0,045 V se confundiram com o sinal do branco), de modo que após sua subtração, os valores de corrente ficaram muito próximos do zero e, portanto, sem informação ([Figura 4.3](#)). Por esta razão, esse intervalo foi eliminado e a faixa compreendida entre -0,046 a -0,900 V foi selecionada como região de trabalho ([Figura 4.4](#)) para todos os voltamogramas usados no tratamento quimiométrico, que foram exatamente os voltamogramas subtraídos da resposta do branco. Vale salientar que, o voltamograma de uma amostra ([Figura 4.3](#) – linha

vermelha) contém informações do etanol e dos compostos do óleo nele presentes.

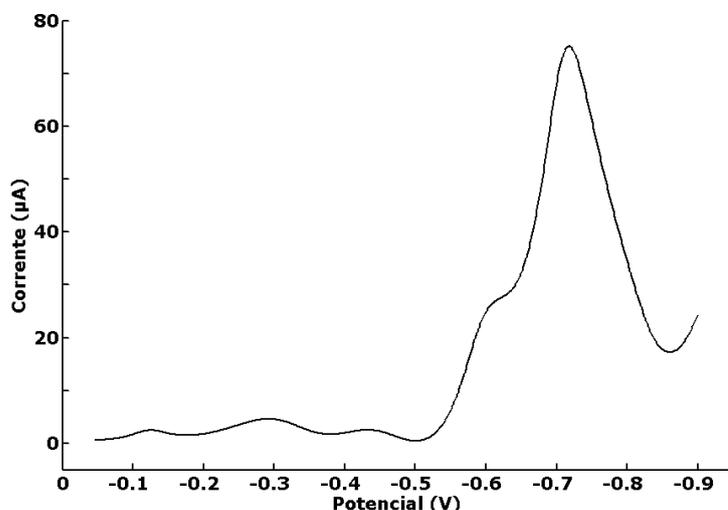


Figura 4.4 Voltamograma de uma amostra de óleo de canola na região de trabalho selecionada (-0,046 a -0,900 V).

4.6 Perfis dos voltamogramas do etanol e de etanol + óleo

A [Figura 4.5](#) mostra os perfis, após subtração do branco, dos voltamogramas do etanol (linha azul), de um óleo de canola em etanol na proporção 1:1 (linha vermelha) e a subtração destes dois perfis (linha verde).

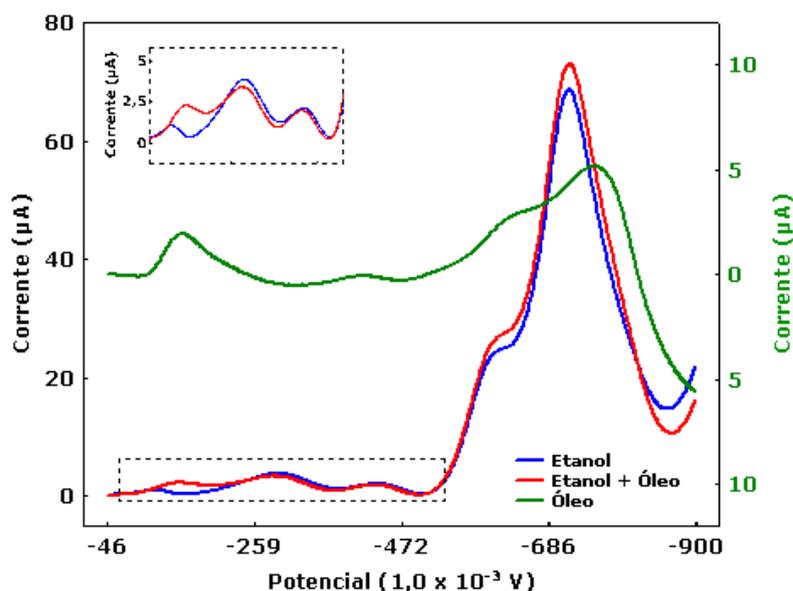


Figura 4.5 Perfis, após subtração do branco, dos voltamogramas do etanol (linha azul), de um óleo de canola em etanol na proporção 1:1 (linha vermelha) e a subtração destes dois perfis (linha verde). O detalhe no canto superior esquerdo dessa figura é uma ampliação da área marcada nos voltamogramas.

Devido à grande diversidade de compostos (etanol, ácidos graxos livres, ésteres etílicos, carboxilatos graxos, triglicerídeos, sabão, matéria insaponificável, glicerol etc.) eventualmente extraídos pelo etanol, torna-se difícil atribuir os picos e potenciais a analitos específicos. Para que isso fosse possível, seria necessário utilizar padrões de todos os componentes do óleo vegetal e obter a resposta, tanto para cada um deles, como para misturas de tais padrões. No entanto, na literatura existem trabalhos que utilizam técnicas eletroanalíticas para determinação de ácidos graxos. Neste contexto, alguns potenciais tais como -0,33 V (TAKAMURA et al., 1995), -0,415 V (KOTANI et al. 2000) e -0,6 V (MENDONÇA et al., 2003) são atribuídos a presença de ácidos graxos. Assim, pode ser deduzido que existe no perfil da amostra de óleo de canola em etanol informações que podem ser usadas para a caracterização de óleos vegetais na janela de potencial de trabalho entre -0,046 a -0,900 V.

4.7 Pré-processamento dos voltamogramas

Os voltamogramas das amostras de óleo em etanol 1:1 não sofreram variações significativas na linha de base e apresentaram pouco ruído, portanto, não foi necessário aplicar técnicas de pré-processamento. Os voltamogramas das 114 amostras de óleos vegetais analisados neste trabalho são apresentados na [Figura 4.6](#). Nesta figura, os menores valores de corrente no pico estão associados às amostras vencidas (PVE), isto pode estar acontecendo porque, à medida que os óleos envelhecem, ocorre um aumento na concentração de AcGs livres e, conseqüentemente, uma maior formação de sabão dentro da cela eletroquímica causando uma depleção do sinal analítico.

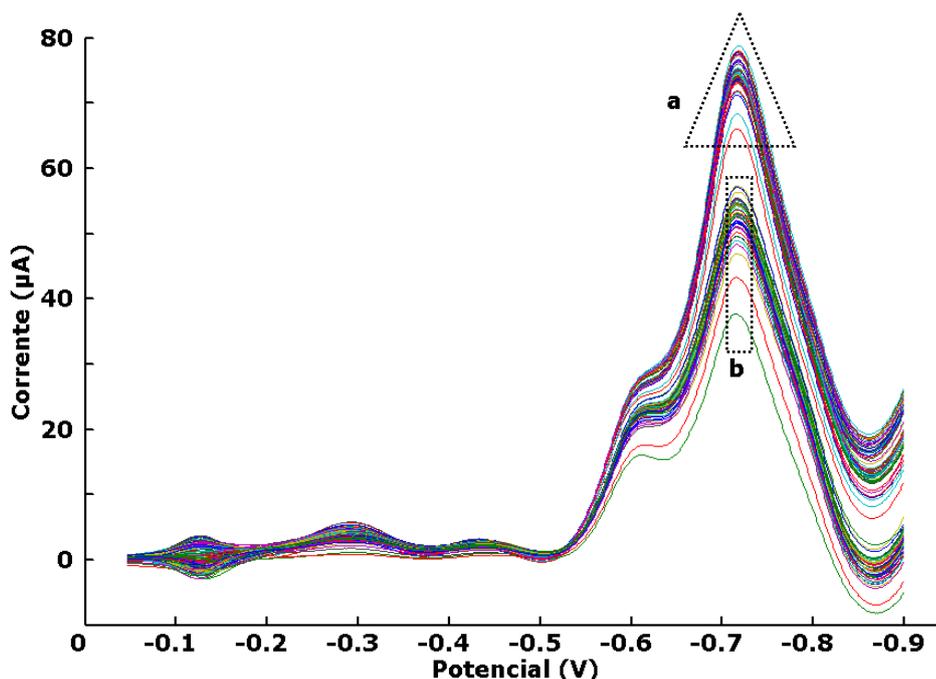


Figura 4.6 Voltamogramas das 114 amostras de óleo vegetal na região de trabalho, após a subtração do branco: Voltamogramas das 66 amostras de óleo não vencidas (a) e das 48 amostras vencidas (b).

4.7.1 Pré-processamento aplicado às variáveis

A eliminação de algumas variáveis não informativas (0,300 a -0,04501 V) e a escolha da região de trabalho (-0,04608 a -0,900 V), conforme descrito na [Seção 4.5](#), foi o primeiro pré-processamento realizado sobre as variáveis. Além deste, fez-se necessário a centralização dos dados na média.

Todas as amostras analisadas possuem perfis voltamétricos semelhantes ([Figura 4.6](#)). Isto está relacionado à grande semelhança na composição dos óleos analisados, bem como à natureza complexa do sinal obtido por SWV. Observa-se nos voltamogramas apenas a tendência de formação de dois grandes grupos que correspondem às amostras não vencidas e vencidas. Entretanto, torna-se impossível a distinção entre os diferentes tipos de óleos vegetais por meio da inspeção visual dos voltamogramas. Para isto, tornou-se indispensável o uso de ferramentas quimiométricas, conforme descrito na [Seção 1.7](#).

4.8 Análise exploratória dos dados

Com o propósito de investigar a existência de similaridades e/ou diferenças entre as amostras e variáveis (potenciais eletroquímicos) analisadas por SWV, realizou-se uma análise exploratória dos dados aplicando PCA como técnica de reconhecimento de padrão não-supervisionado.

4.8.1 Aplicação da PCA

Inicialmente foi realizada uma PCA em todo o conjunto das 114 amostras de óleo vegetal para avaliar possíveis sobreposições entre as classes. Essa análise mostrou que existe uma separação em dois grandes grupos ao longo de PC1, que detém 98 % da variância explicada dos dados (Figura 4.7). Os escores mais negativos em PC1 pertencem às amostras da classe PVE, enquanto que os mais positivos correspondem às amostras não vencidas.

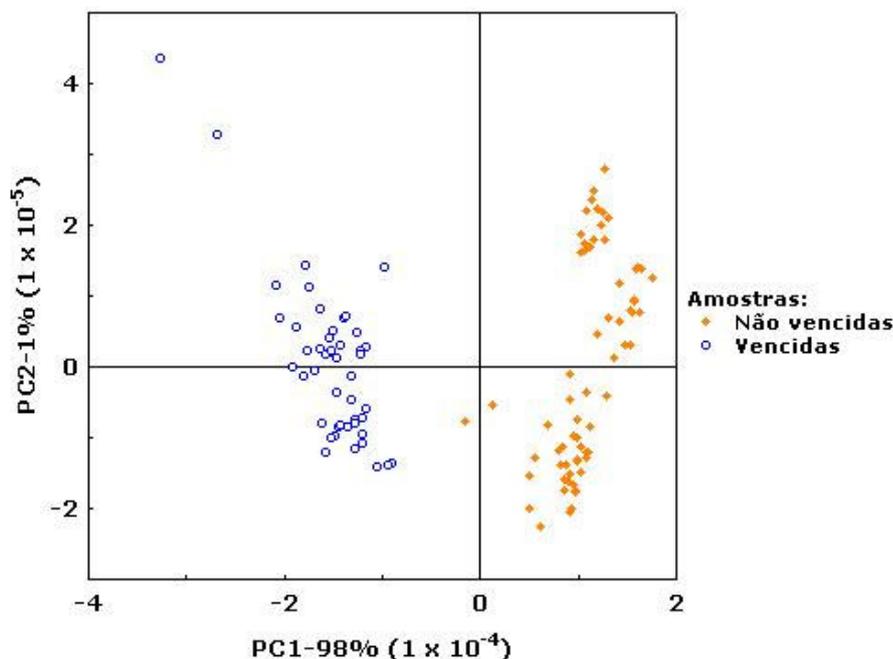


Figura 4.7 Gráfico dos escores de PC1 versus PC2 das 114 amostras de óleo vegetal.

Em virtude do perfil da PCA para todas as amostras, aplicou-se novamente esta técnica para cada um dos dois grandes grupos acima evidenciados.

4.8.1.1 Aplicação da PCA às amostras não vencidas

A [Figura 4.8](#) mostra o gráfico dos escores para a PC1 *versus* PC2 para todas as classes de amostras não vencidas. A PC1 e PC2 explicam 87% e 9% da variância dos dados, respectivamente. Valores positivos de escores ao longo de PC1 distinguem dois grupos: um constituído pelas classes Canola e Milho e o outro formado por Soja e Girassol. Além disso, valores positivos de escores para PC1 e PC2 caracterizam claramente a classe Canola, assim como os valores de escores positivos para PC1 e negativos para PC2 os fazem para a classe Milho. É importante observar que as classes Girassol e Soja estão se sobrepondo ao longo dos eixos formados pela PC1 e PC2, ainda assim, a classe Girassol é mais amplamente distribuída e apresenta uma tendência de deslocar-se mais para a esquerda. Possivelmente, o motivo desta sobreposição se deve ao fato de que estes óleos possuem características semelhantes com relação à sua composição em ácidos graxos.

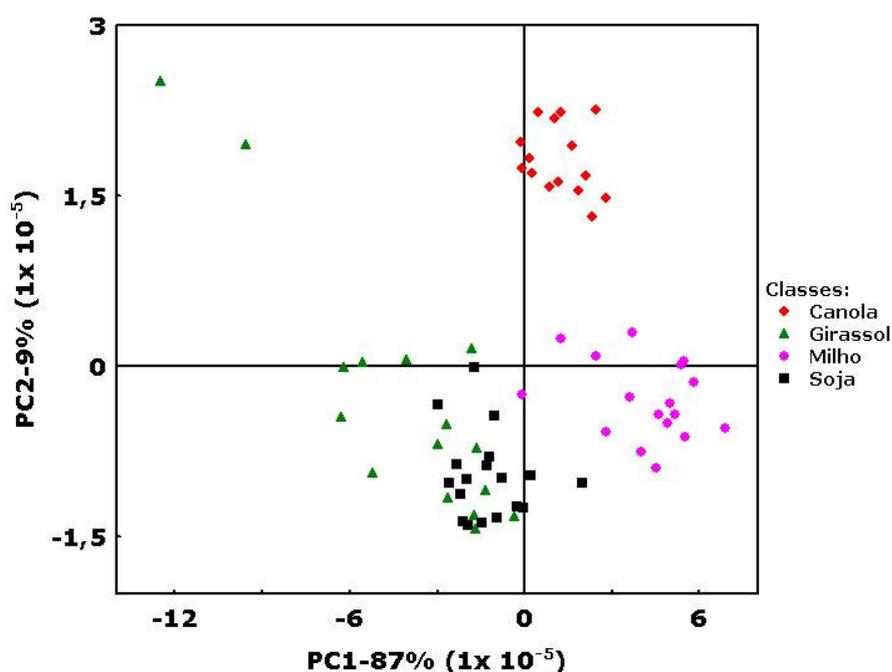


Figura 4.8 Gráfico dos escores de PC1 *versus* PC2 para as 66 amostras não vencidas de óleo vegetal distribuídas em quatro classes.

Os voltamogramas médios das classes apresentados na [Figura 4.9](#) corroboram com os resultados obtidos pela PCA. Nota-se que existem

semelhanças entre os perfis voltamétricos das classes Girassol e Soja, assim como das classes Canola e Milho. Na região de -580 mV a -650 mV, é evidenciada uma pequena diferença entre o perfil do voltamograma da classe Canola em relação às demais.

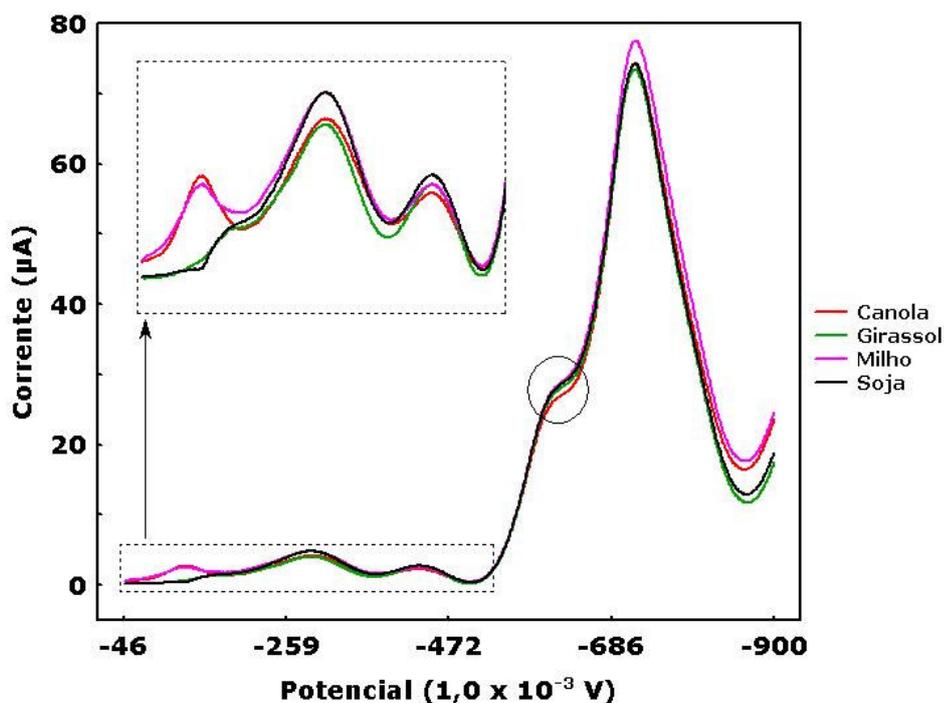


Figura 4.9 Voltamogramas médios das classes Canola, Girassol, Milho e Soja. A seta indica uma ampliação da região de menor intensidade de corrente e o círculo destaca a região de -580 mV a -650 mV onde o perfil da classe Canola se diferencia das demais.

4.8.1.2 Comparação dos óleos em relação à composição química

Um estudo de correlação foi realizado entre os quatro tipos de óleos avaliados. Para isso, os valores médios de sua composição química apresentados na [Tabela 1.2](#) foram organizados em uma matriz de dados ([Tabela 4.7](#)). Os tipos de óleos foram considerados como variáveis e os valores de correlação entre eles foram calculadas ([Tabela 4.8](#)). As duas maiores correlações foram obtidas entre os óleos girassol e soja (98,5%) e soja e milho (97,7%). A maior correlação para o óleo de canola, 76,2%, ocorreu com o óleo de milho. Estes valores corroboram com o comportamento da dispersão das classes em estudo apresentado na [Figura 4.8](#).

Tabela 4.7 Matriz da composição química dos óleos vegetais de acordo com os dados da RDC 492/99 (ANVISA, 2007).

Componentes Químicos (g/100g)	Soja	Milho	Girassol	Canola
Matéria Insaponificável	1,5	2,8	1,5	2,0
Brassicasterol	0	0	0	5,0
C4aC12	0,1	0,3	0,4	0
Mirístico	0,5	0,1	0,5	0,2
Palmítico	10,5	11,5	6,5	4,5
Palmitoléico	0,5	0,5	1,0	0,6
Esteárico	3,4	2,2	5,5	1,9
Oléico (ω -9)	24,5	33,0	24,5	61,5
Linoléico (ω -6)	53,0	48,0	65,0	22,5
Linolênico (ω -3)	7,5	2,0	0,3	9,0
Araquídico	1,0	1,0	1,5	0,65
Eicosenóico	1,0	0,5	0,5	2,2
Behênico	0,5	0,5	1,0	0,6
Lignocérico	0	0,5	0,5	0,2
Erúcico	0	0	0,5	2,0
Tetracosenóico	0	0	0,5	0,2

Tabela 4.8 Correlação entre os diferentes tipos de óleo vegetal de acordo com a composição de química.

VARIÁVEIS	Correlação (%)			
	SOJA	MILHO	GIRASSOL	CANOLA
SOJA	100,0			
MILHO	97,7	100,0		
GIRASSOL	98,5	95,9	100,0	
CANOLA	64,3	76,2	57,8	100,0

Estes resultados estão de acordo com alguns trabalhos de classificação de óleos vegetais que utilizam outras técnicas analíticas, tais como: espectrometria de massas com ionização por "electrospray"

(CATHARINO et al., 2005) e FTIR, FT-NIR e FT-Raman (YANG et al., 2005).

4.8.1.3 Aplicação de PCA às amostras da classe PVE

Como a classe PVE (Tabela 3.1) apresenta amostras dos mesmos tipos de óleos vegetais não vencidos (canola, girassol, milho e soja), uma PCA (Figura 4.10) foi realizada com intuito de investigar a formação de grupos entre as amostras com o prazo de validade expirado.

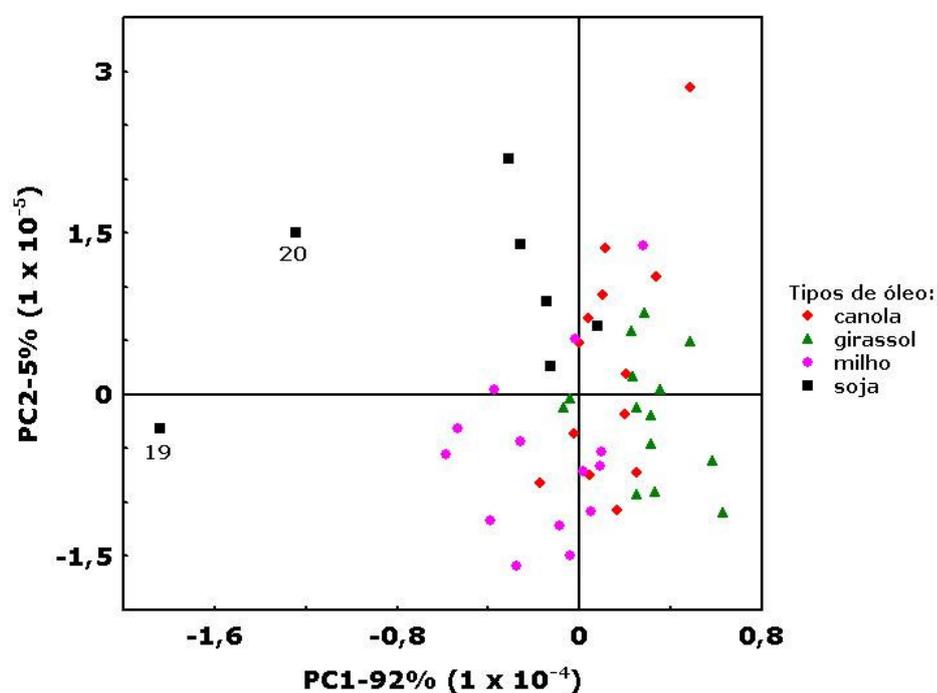


Figura 4.10 Gráfico dos escores de PC1 versus PC2 das 48 amostras de óleo vegetal da classe PVE.

Pode-se observar na Figura 4.10 que todas as amostras pertencentes à classe PVE encontram-se com pouca distinção. Nota-se também que duas amostras de soja (19 e 20) estão deslocadas, com escores negativos, ao longo de PC1. Este comportamento é esperado em virtude de um perfil voltamétrico diferenciado para essas duas amostras (Figura 4.11). Possivelmente, isto ocorreu devido a uma maior degradação dessas amostras ao longo do tempo. Então, estas amostras não foram consideradas *outliers* e permaneceram no conjunto dos óleos da classe PVE para futuras investigações.

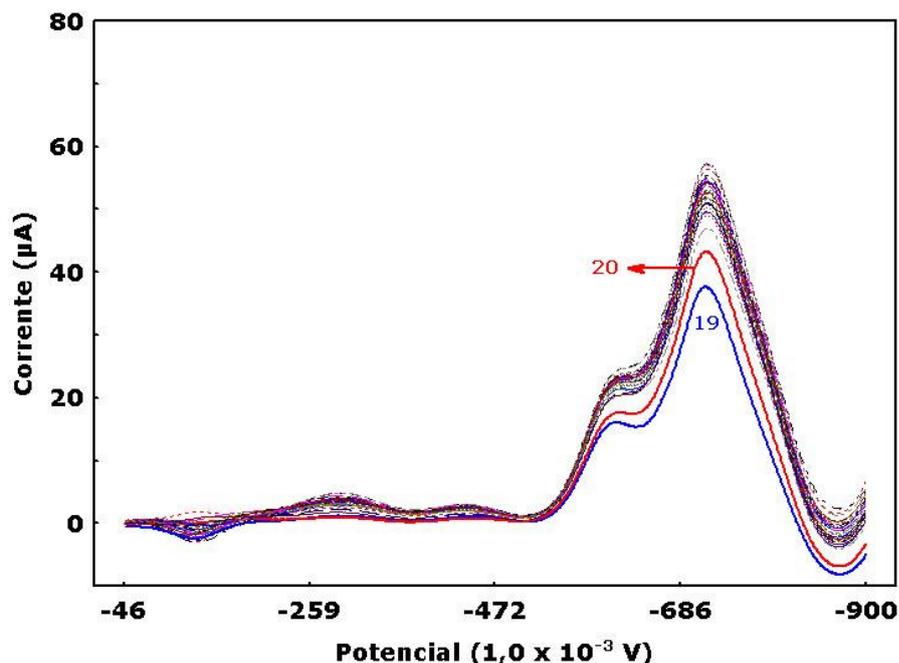


Figura 4.11 Perfil voltamétrico das amostras de óleo da classe PVE.

A análise exploratória realizada em todo o conjunto de dados mostrou claramente a distinção entre óleos vencidos e não vencidos (Figura 4.7). Contudo, para a inspeção sobre as amostras não vencidas, três grupos de amostras foram caracterizados: Canola, Milho e um último formado pela sobreposição parcial das classes Girassol e Soja (Figura 4.8). Para as amostras vencidas, não foi possível obter uma boa discriminação entre os quatro tipos de amostras.

Torna-se, portanto, necessário o desenvolvimento de modelos baseados em métodos de reconhecimento de padrão supervisionado. Nestes métodos, é possível construir modelos individuais para cada classe e utilizá-los para a previsão de amostras futuras. Como visto na Seção 1.7, SIMCA e LDA são as técnicas mais utilizadas para este propósito.

4.9 Modelagem e Classificação SWV-SIMCA

Modelos de classificação SWV-SIMCA para cada tipo de óleo vegetal foram construídos com as amostras do conjunto de treinamento e validados com as amostras do conjunto teste.

4.9.1 Construção dos modelos SWV-SIMCA

Três ferramentas de diagnóstico foram utilizadas para a construção destes modelos: 1) gráfico da variância explicada *versus* número de PCs; 2) gráfico da variância residual *versus* número de PCs e 3) gráfico dos escores. Como a determinação do número de variáveis latentes é um fator preponderante na avaliação da capacidade preditiva dos modelos, conhecer o número ótimo de PCs para cada classe modelada é essencial. Então, os gráficos (Figura 4.12a-e) para as cinco classes foram construídos para esta finalidade.

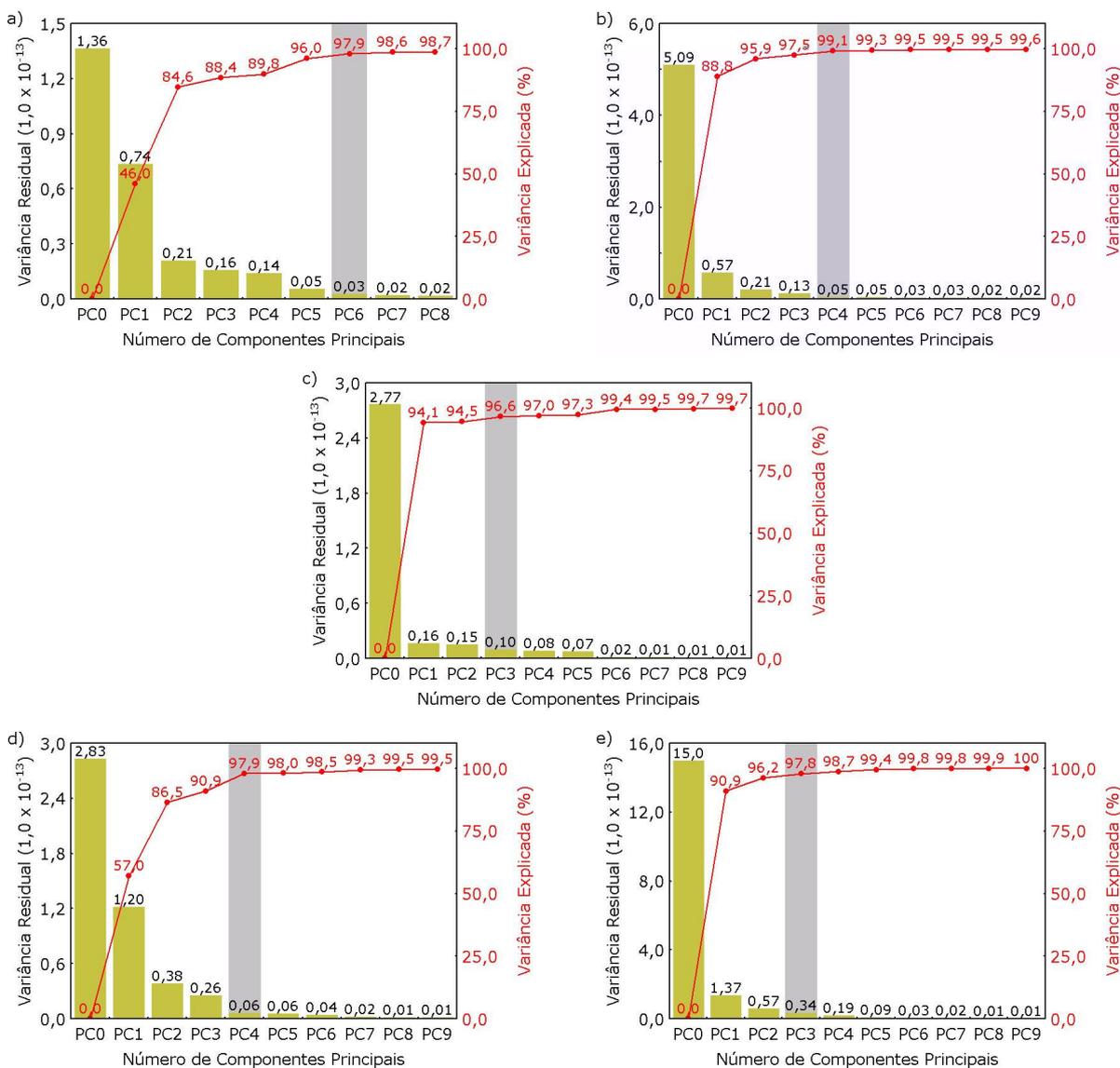


Figura 4.12 Gráficos da variância residual e explicada *versus* nº de PCs das classes: a) Canola; b) Girassol; c) Milho; d) Soja; e) PVE.

A [Tabela 4.9](#) reúne o número de PCs e o total da variância explicada para cada modelo depois da análise dos gráficos apresentados na figura [Figura 4.12a-e](#). Para todos os casos, não houve um aumento significativo da variância explicada, assim como uma diminuição da variância residual depois do número de PCs estabelecido para a construção dos modelos. Este resultado coincidiu com o adotado pelo *default* do Unscrambler®.

Tabela 4.9 Número de PCs e a variância explicada acumulada dos modelos SWV-SIMCA.

Modelo	nº. de PCs	Variância Explicada (%)
Canola	6	97,9
Girassol	4	99,1
Milho	3	96,1
Soja	4	97,9
PVE	3	97,8

Por fornecer uma visão privilegiada das amostras ao longo das PCs, a análise dos gráficos dos escores ([Figura 4.13](#)) é também uma ferramenta de diagnóstico importante na validação dos modelos construídos. Com ela, torna-se possível examinar as diferenças, semelhanças, localização e anomalias das diferentes amostras usadas no processo de modelagem.

Os cinco gráficos de escores apresentados na [Figura 4.13a-e](#) mostram que todas as amostras do conjunto de treinamento, escolhidas separadamente para cada classe pelo algoritmo KS, encontram-se bem distribuídas ao longo das 3 PCs, com amostras de teste dentro dos espaço varrido pelas amostras do conjunto de treinamento.

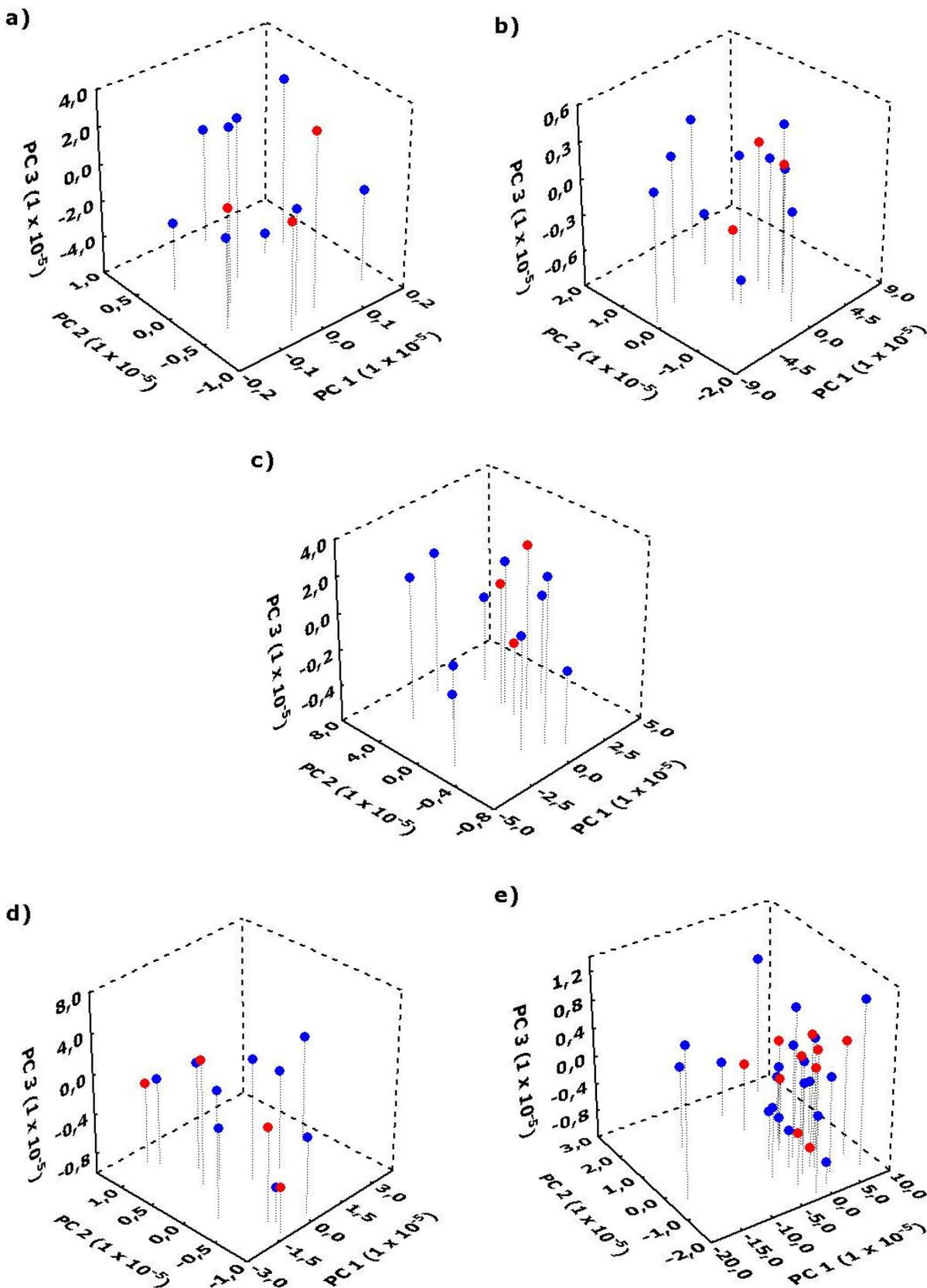


Figura 4.13 Gráfico dos escores de PC1 versus PC2 versus PC3 para as classes: a) Canola; b) Girassol; c) Milho; d) Soja; e) PVE. As amostras dos conjuntos de treinamento e de teste estão pintadas de azul e vermelho, respectivamente.

4.9.2 Validação dos modelos SWV-SIMCA

O critério utilizado pelas ferramentas de diagnósticos na escolha do número ótimo de PCs para cada modelo é passivo de erros. Isto porque, de acordo com [LEE et al. \(1998\)](#), esta diretriz é uma tentativa visual de fazer a melhor escolha entre os maiores e menores autovalores. Portanto, a validação é uma etapa muito importante para analisar estatisticamente se houve alguma incongruência na construção dos modelos. Neste contexto, os modelos SWV-SIMCA construídos para as cinco classes de óleos vegetais foram aplicados às amostras dos conjuntos de treinamento e de teste (nesta ordem e de acordo com a [Tabela 3.2](#)) com intuito de avaliar sua capacidade preditiva. Para isso foram construídos gráficos $S_i \times H_i$ ([Seção 1.7.4](#)), empregando um nível de confiança de 95%. As avaliações do modelo SWV-SIMCA para cada classe são apresentadas nas seções a seguir.

4.9.2.1 Validação do modelo SWV-SIMCA da classe Canola

No gráfico $S_i \times H_i$ da classe Canola ([Figura 4.14](#)), verifica-se que a amostra 11, pertencente ao conjunto de teste, está sendo classificada incorretamente. Possivelmente, o motivo desta má classificação pode ser atribuído à escolha inadequada do número de PCs adotado, 6 neste caso ([Tabela 4.9](#)), para a construção do modelo ([BEEBE et al., 1998](#)). Devido a isso, um teste F ([Figura 4.15](#)) foi aplicado aos resíduos das amostras da classe Canola para modelos com 4, 5 e 6 PCs, uma vez que a variância residual dos modelos com 2, 3 e 4 PCs aparenta-se ainda significativa ([Figura 4.12a](#)).

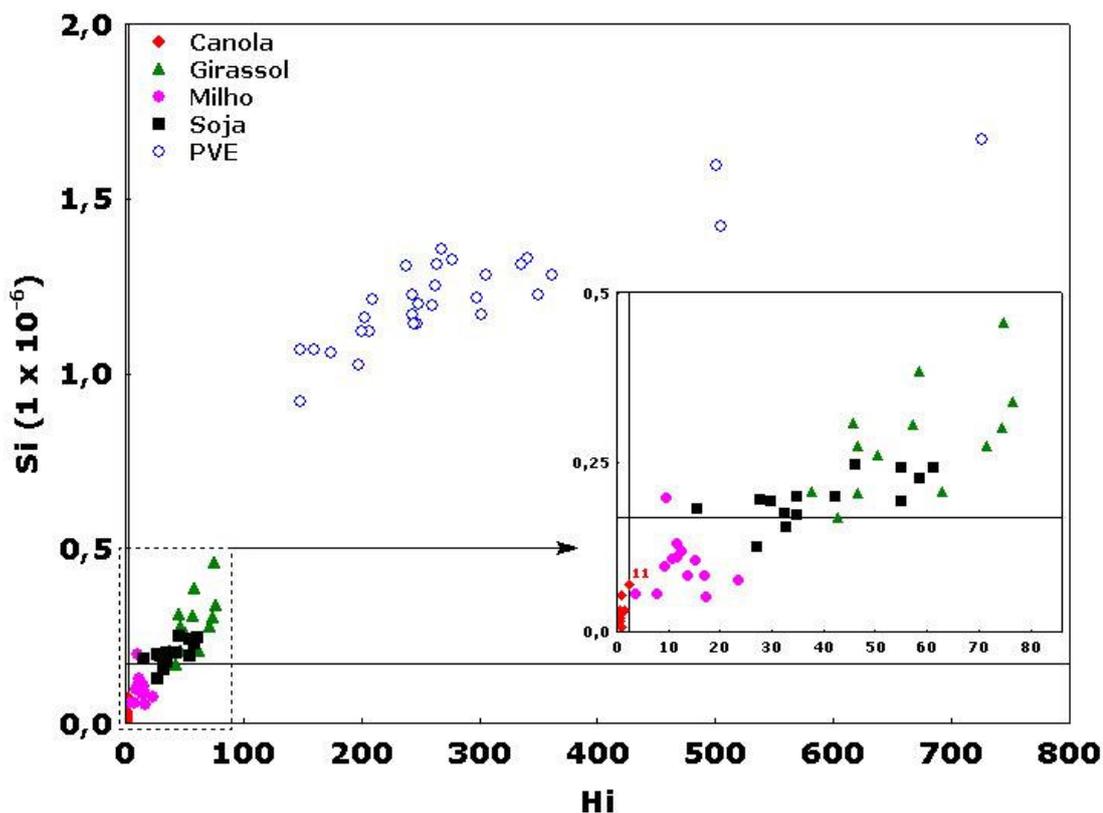


Figura 4.14 Gráfico de Si x Hi (95% de confiança) do modelo da classe Canola construído com 6 PCs. A seta indica uma ampliação das fronteiras deste modelo evidenciando a classificação incorreta da amostra 11.

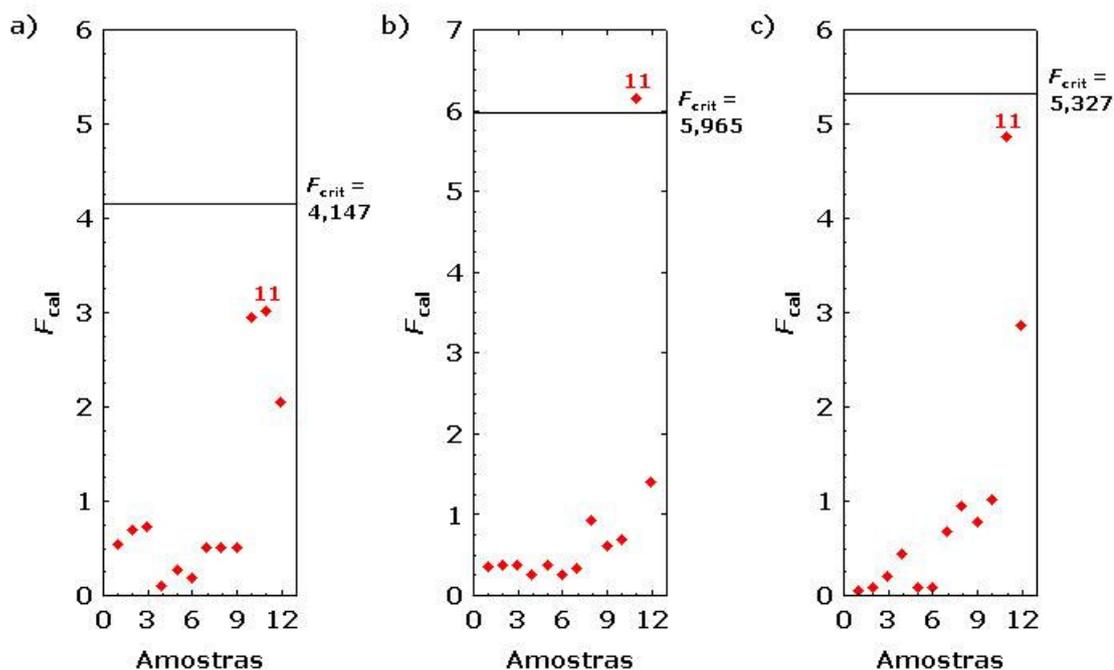


Figura 4.15 Teste F aplicado aos resíduos das amostras de treinamento e teste da classe Canola para modelos com 4 (a), 5(b) e 6 (c) PCs.

Analisando os resultados do teste F e lembrando-se do número restrito de amostras disponíveis (9 no caso), a utilização de 6 PCs para estabelecer o modelo da classe Canola é provavelmente um caso de modelo sobre-ajustado (*overfitting*). Para os resíduos do modelo com 5 PCs (Figura 5.15b), a amostra 11 apresentou valor de F_{cal} maior que o de F_{crit} , indicando diferenças significativas entre ela e as demais. Permanecer com modelos construídos com 5 ou 6 PCs poderia comprometer a habilidade preditiva do modelo. Portanto, optou-se por utilizar apenas 4 PCs para elaborar o modelo SWV-SIMCA definitivo da classe Canola. Para efeito de comparação, um novo gráfico $S_i \times H_i$ para este modelo foi construído (Figura 4.16) e uma maior definição da classe Canola é claramente observada.

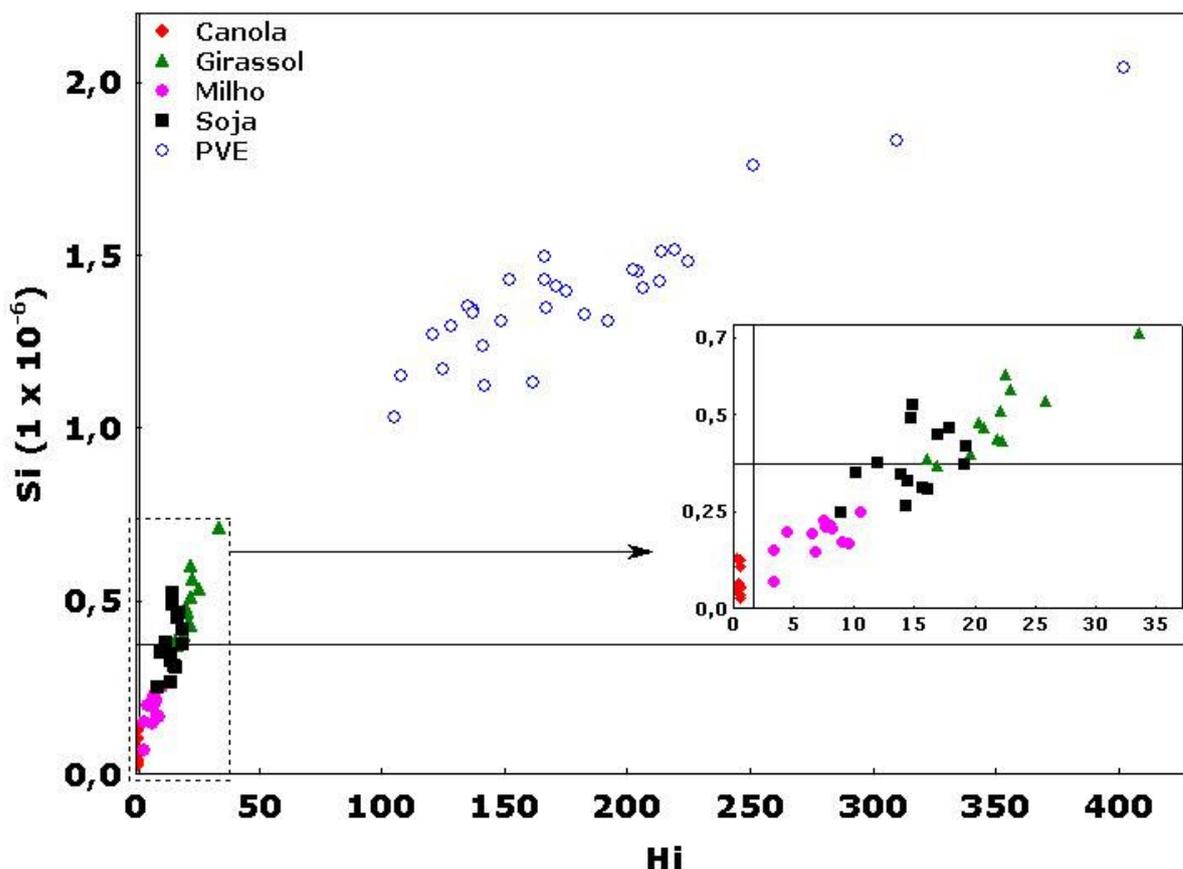


Figura 4.16 Gráfico de $S_i \times H_i$ (95% de confiança) do modelo da classe Canola construído com 4 PCs após a etapa de validação. A seta indica uma ampliação das fronteiras deste modelo evidenciando a classificação correta de todas as amostras.

4.9.2.2 Validação do modelo SWV-SIMCA da classe Girassol

O gráfico Si x Hi para a classe Girassol (Figura 4.17) mostra que nenhuma amostra de óleo vegetal pertencente a esta classe foi classificada incorretamente, usando um modelo com 4 PCs. Entretanto, sete amostras da classe Soja foram classificadas como sendo pertencentes à classe Girassol. Estes resultados corroboram com aqueles obtidos pela análise exploratória (Figura 4.8), onde foram encontradas fortes sobreposições entre as classes Girassol e Soja.

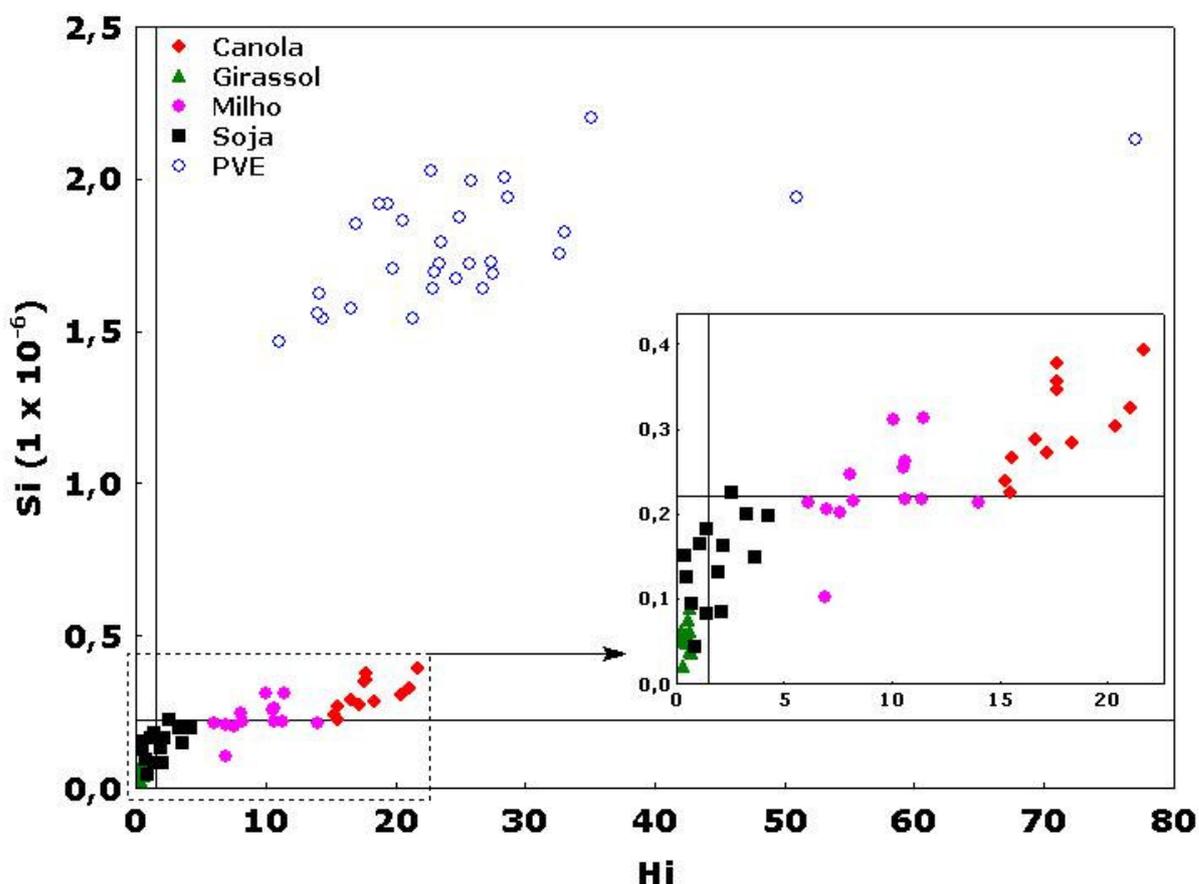


Figura 4.17 Gráfico de Si x Hi (95% de confiança) do modelo da classe Girassol construído com 4 PCs. A seta indica uma ampliação das fronteiras deste modelo evidenciando a classificação incorreta de sete amostras de óleo de soja.

4.9.2.3 Validação do modelo SWV-SIMCA da classe Milho

Para a classe Milho, o gráfico Si x Hi (Figura 4.18) do modelo com 3 PCs indica que nenhuma amostra é classificada incorretamente. Porém,

uma amostra da classe Soja foi classificada como pertencente à classe Milho.

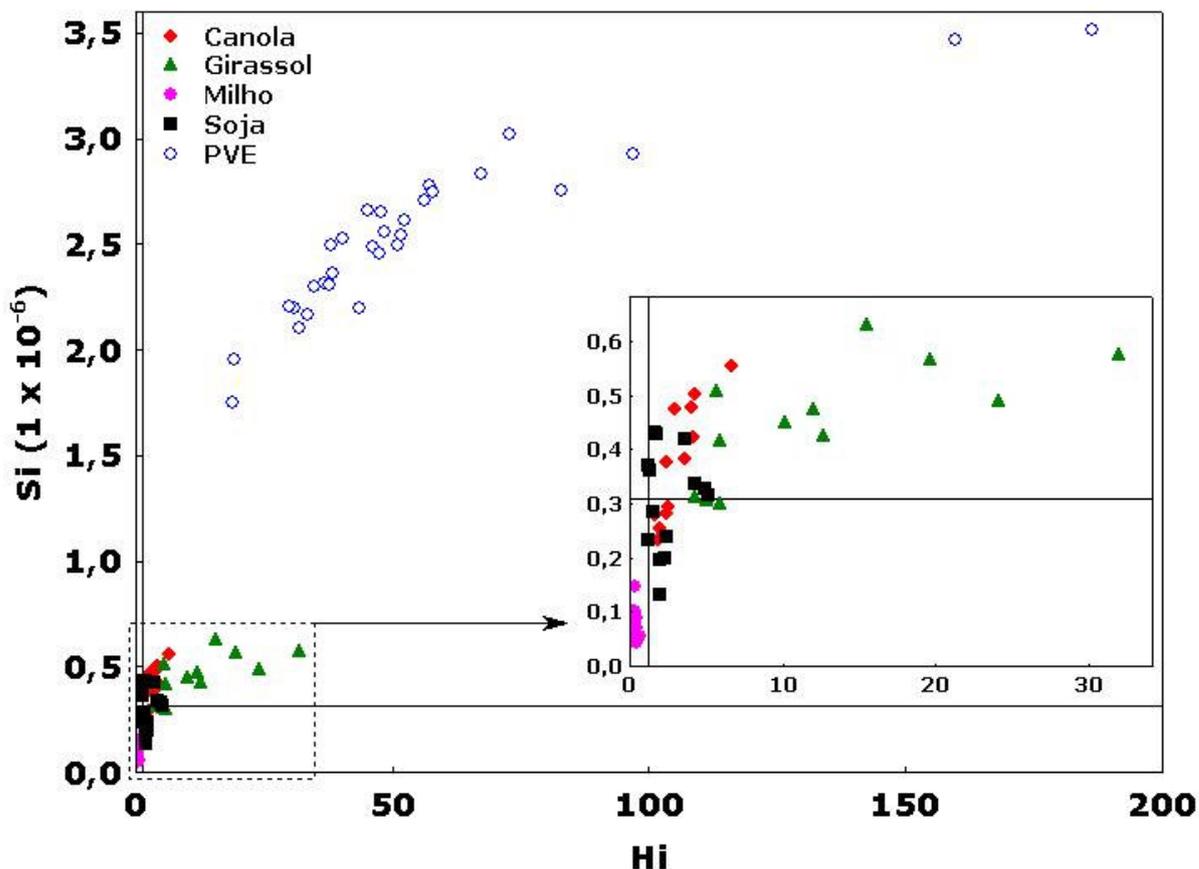


Figura 4.18 Gráfico de Si x Hi (95% de confiança) do modelo da classe Milho construído com 3 PCs. A seta indica uma ampliação das fronteiras deste modelo evidenciando a classificação incorreta de uma amostra de óleo de soja.

4.9.2.4 Validação do modelo SWV-SIMCA da classe Soja

De acordo com o exposto no gráfico Si x Hi para a classe Soja (Figura 4.19), observa-se que todas as amostras foram classificadas corretamente. Todavia, seis amostras da classe Girassol e uma amostra da classe Milho foram classificadas como pertencentes à classe Soja.

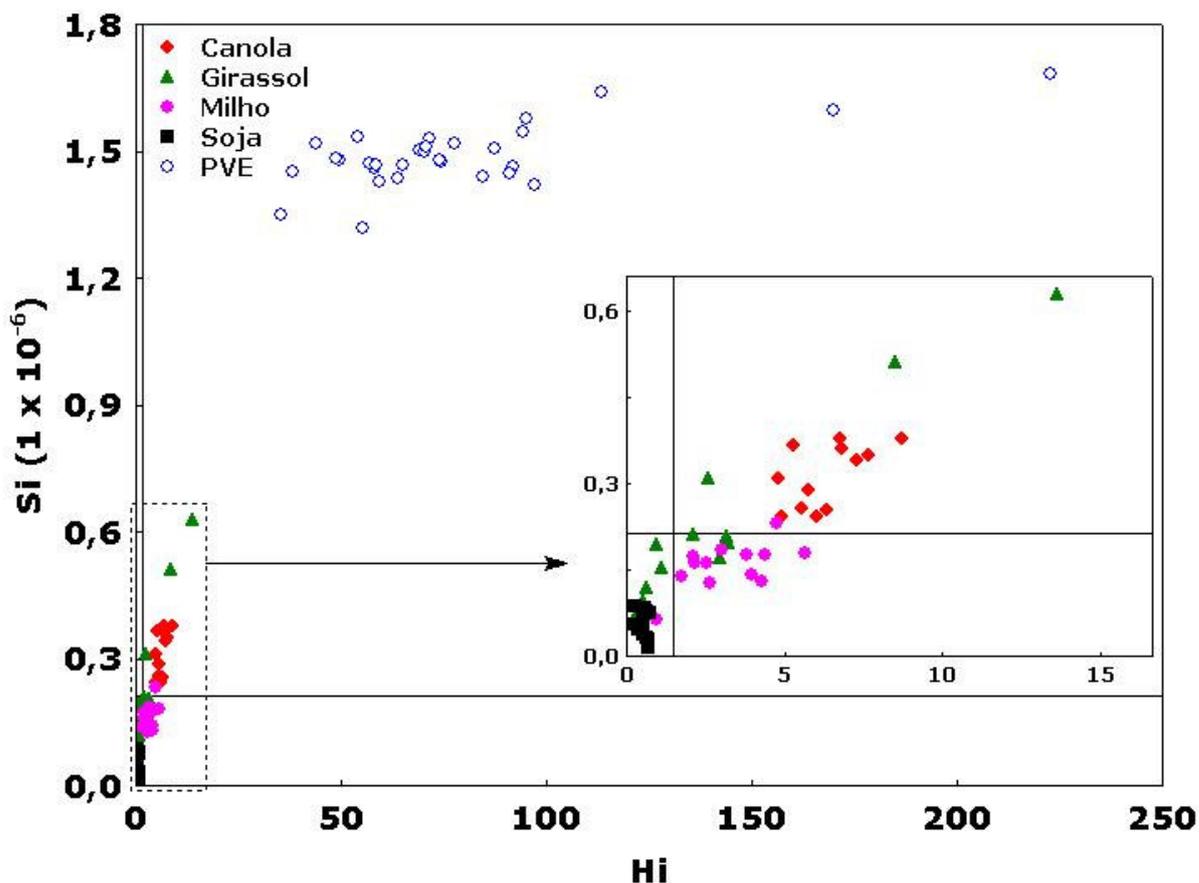


Figura 4.19 Gráfico de Si x Hi (95% de confiança) do modelo da classe Soja construído com 4 PCs. A seta indica uma ampliação das fronteiras deste modelo evidenciando a classificação incorreta de seis amostras de óleo de girassol e de uma amostra de óleo de milho.

4.9.2.5 Validação do modelo SWV-SIMCA da classe PVE

No gráfico Si x Hi da classe PVE (Figura 4.20), percebe-se que as amostras de treinamento 10 e 19 não foram classificadas corretamente. Analogamente ao que aconteceu para o modelo inicial da classe Canola, o motivo desta má classificação pode ser atribuído à escolha inadequada do número de PCs adotado, 3 neste caso (Tabela 4.9), para a construção do modelo (BEEBE et al., 1998). Portanto, para investigar o comportamento das amostras empregadas na etapa de validação, foi aplicado um teste *F* com 95% de confiança sobre os resíduos dos modelos construídos com 1, 2, 3 e 4 PCs (Figura 4.21).

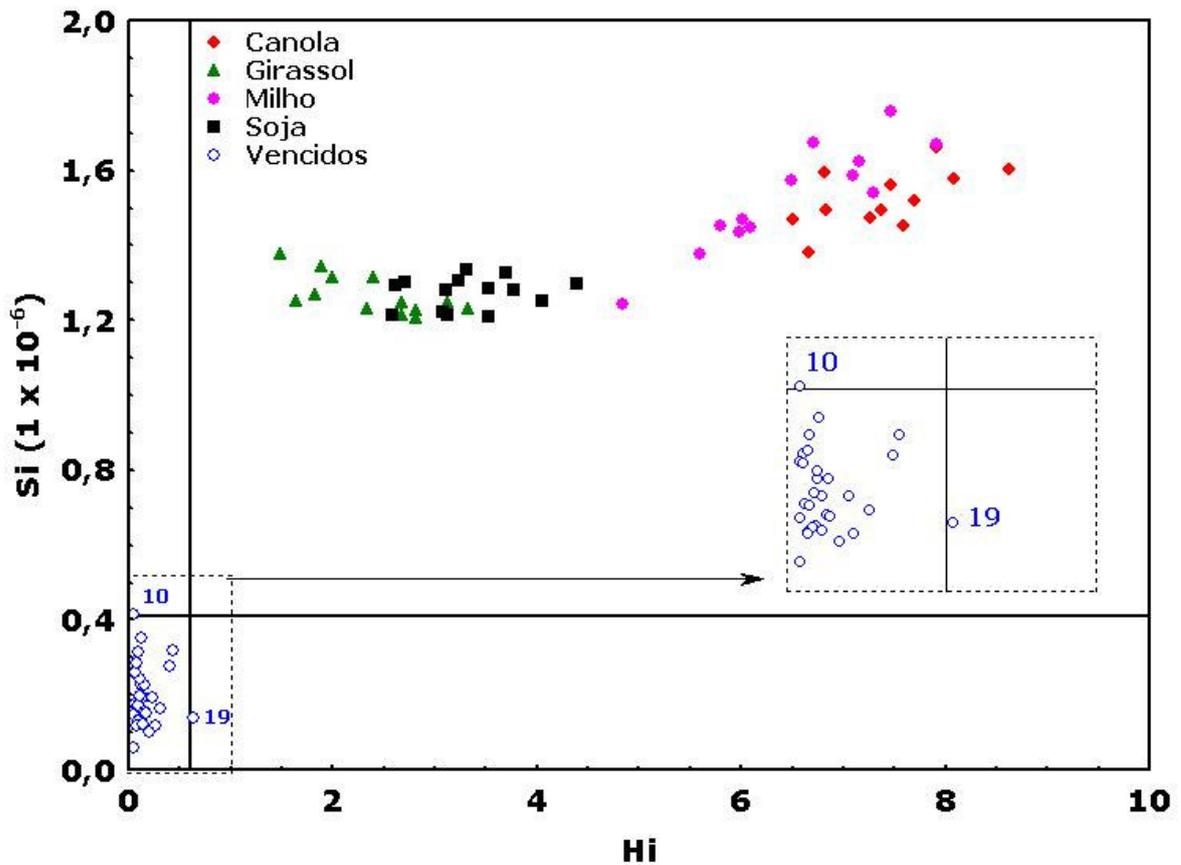


Figura 4.20 Gráfico de $S_i \times H_i$ (95% de confiança) do modelo da classe PVE construído com 3 PCs. A seta indica uma ampliação das fronteiras do modelo evidenciando a classificação incorreta das amostras 10 e 19.

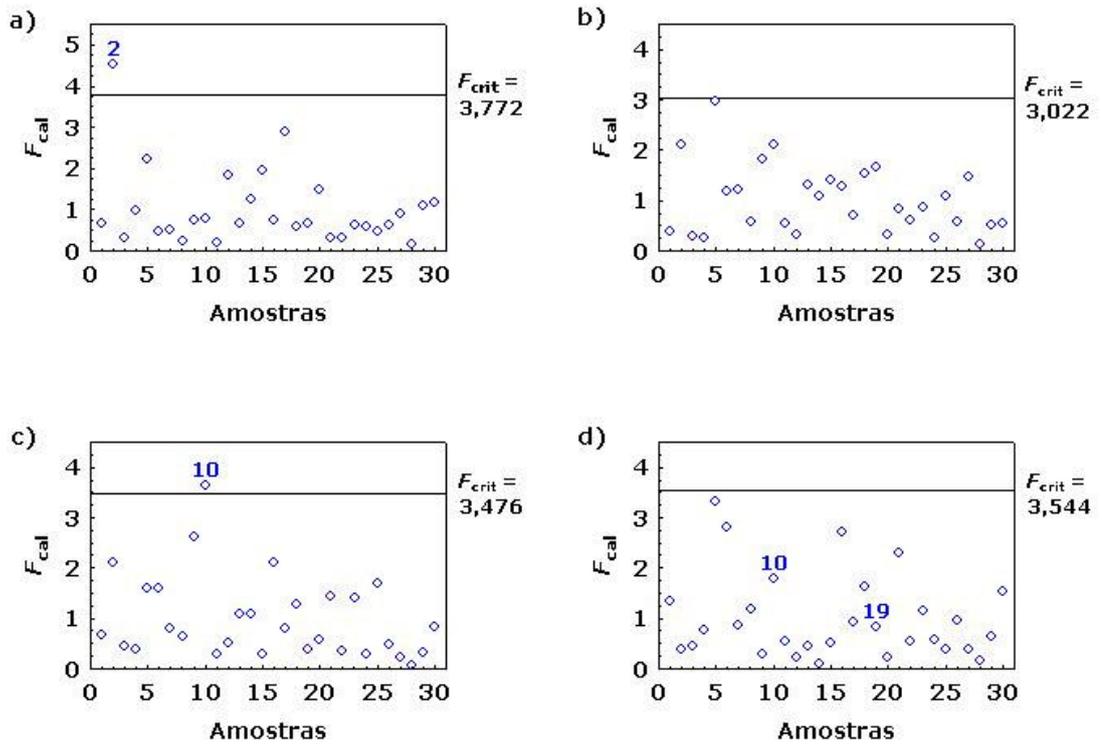


Figura 4.21 Teste F aplicado aos resíduos das amostras de treinamento e teste da classe PVE para modelos com 1 (a), 2 (b), 3 (c) e 4 (d) PCs.

Analisando os resultados do teste F para os resíduos do modelo com 3 PCs (Figura 4.21c), apesar da amostra 19 ser classificada incorretamente, seu valor de F_{cal} (0,40) é menor do que o F_{crit} (3,48). Isto indica que não existem diferenças significativas entre ela e as demais. Com relação à amostra 10, o F_{cal} (3,64) é maior do que o F_{crit} , indicando, portanto, diferenças significativas entre esta amostra e as demais. Assim, o modelo da classe PVE construído com 3 PCs parece não ser adequado, sendo necessário avaliar os modelos com 2 ou 4 PCs.

Os testes F para os resíduos dos modelos construídos com 2 e 4 PCs (Figura 4.21b e d) mostram que todas as amostras apresentam, em ambos os casos, valor de F_{cal} menor que F_{crit} . No entanto, os resíduos das amostras para o modelo com 2 PCs *versus* variáveis (Figura 4.22a) apresentam um perfil com estrutura, portando ainda informações. Para o modelo com 4 PCs (Figura 4.22b), percebe-se um perfil residual das amostras mais aleatório e, portanto, menos informativo. Face ao exposto, 4 PCs parece ser um número adequado para construir o modelo SWV-SIMCA da classe PVE.

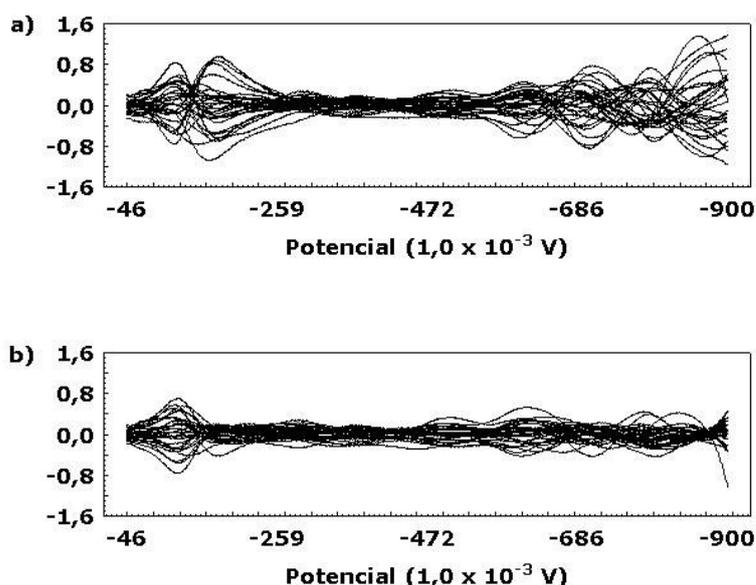


Figura 4.22 Perfil dos resíduos das amostras da classe PVE *versus* variáveis para modelos com 2 (a) e 4 (b) PCs.

Após a validação de todos os modelos, o número de PCs e a variância explicada acumulada são mostrados na Tabela 4.10 para cada um deles.

Tabela 4.10 Número de PCs e variância explicada acumulada para cada modelo SWV-SIMCA de óleo vegetal após a etapa de validação.

Modelo	nº. de PCs	Variância Explicada (%)
Canola	4	89,8
Girassol	4	99,1
Milho	3	96,1
Soja	4	97,9
PVE	4	98,7

Os modelos construídos para as cinco classes na etapa de validação foram avaliados com todas as amostras pertencentes aos conjuntos de treinamento e teste. A classe PVE, que engloba óleos dos quatro tipos com o prazo de validade expirado, encontrou-se sempre bem separada das demais classes em estudo (Canola, Girassol, Milho e Soja), que são constituídas por óleos não vencidos. Desta forma, torna-se evidente a possibilidade de fiscalização desses quatro tipos de óleos vegetais analisados com relação ao seu prazo de validade. No entanto, alguns problemas de sobreposição ocorreram entre as classes Girassol e Soja (Figuras 4.17 e 4.19) e entre Milho e Soja (Figuras 4.18 e 4.19).

Isto pode ter ocorrido devido a:

1. Existência de características em comum entre os óleos de Girassol e Soja e entre Milho e Soja com relação à composição em ácidos graxos como mostra o estudo da correlação entre os tipos de óleo de acordo com as especificações da ANVISA (Seção 4.8.1.2);
2. A combinação da técnica SWV com a modelagem SIMCA não foi suficientemente discriminatória para a classificação;
3. Necessidade de um número maior de amostras representativas nos conjuntos de treinamento e de teste;

Os modelos SWV-SIMCA construídos foram usados na classificação de amostras de um conjunto externo de previsão.

4.9.3 Uso dos modelos SWV-SIMCA para o Conjunto de Previsão

Os modelos SWV-SIMCA das classes Canola, Girassol, Milho, Soja e PVE foram usados na classificação das 32 amostras do conjunto de previsão. Os resultados são apresentados na Tabela 4.11. É importante

ressaltar que os erros apresentados numa classificação podem ser do tipo I, quando a amostra avaliada não é classificada na sua classe verdadeira (falso negativo), e do tipo II, quando a amostra é classificada como pertencente a uma classe errada (falso positivo). Estes dois tipos de erro podem acontecer simultaneamente quando uma amostra não é classificada em sua classe verdadeira (erro tipo I), mas é classificada em uma ou mais classes erradas (erro tipo II).

Tabela 4.11 Classificação SIMCA das amostras de previsão das cinco classes de óleo vegetal. O símbolo * indica que a amostra foi classificada no modelo.

Amostras	Modelo (nº de PCs)				
	Canola (4)	Girassol (4)	Milho (3)	Soja (4)	PVE (4)
Canola-1	*				
Canola-2	*				
Canola-3	*				
Girassol-1		*		*	
Girassol-2		*			
Girassol-3		*		*	
Milho-1			*		
Milho-2			*		
Milho-3			*		
Milho-4			*		
Soja-1		*		*	
Soja-2			*	*	
Soja-3		*		*	
Soja-4		*		*	
PVE-1					*
PVE-2					*
PVE-3					*
PVE-4					*
PVE-5					*
PVE-6					*
PVE-7					*
PVE-8					*
PVE-9					*
PVE-10					*
PVE-11					*
PVE-12					*
PVE-13					*
PVE-14					*
PVE-15					*
PVE-16					*
PVE-17					*
PVE-18					*

Verifica-se na [Tabela 4.11](#) que todas as amostras de previsão foram classificadas corretamente, não apresentando, portanto, erro do tipo I. Não obstante, duas amostras da classe Girassol (Girassol-1 e Girassol-3) e todas as quatro amostras da classe Soja apresentaram erro do tipo II (3 classificadas como Girassol e 1 como Milho), a um nível de 95% de confiança.

4.10 Construção dos modelos SWV-LDA-SPA

O mesmo banco de dados utilizado para a análise exploratória e modelagem SWV-SIMCA foi empregado na elaboração dos modelos SWV-LDA-SPA. Vale salientar que uma modelagem baseada em LDA gera sempre N° de Classes menos 1 funções lineares discriminantes. Assim, serão geradas na construção do modelo SWV-LDA-SPA, 4 funções lineares discriminantes (FD1 a FD4) para as cinco classes consideradas (Canola, Girassol, Milho, Soja e PVE).

O gráfico apresentado na [Figura 4.23](#) mostra o número de variáveis selecionadas pelo SPA *versus* os valores da função de custo. O número ideal de variáveis é aquele que apresenta o menor risco médio G de uma classificação incorreta pela LDA, conforme discutido em maiores detalhes na [Seção 1.7.8](#) do Capítulo 1. Portanto, o modelo SWV-LDA-SPA que apresenta o menor risco médio G (0,184) de uma classificação incorreta é aquele que usa apenas dez variáveis das 800 variáveis (potenciais voltamétricos) do conjunto de dados. Essas variáveis correspondem às medidas de corrente nos potenciais -0,1208, -0,1732, -0,2917, -0,4274, -0,5043, -0,6293, -0,7361, -0,7884, -0,8557 e -0,8995 V.

O modelo construído com 10 variáveis não apresentou erros utilizando a série de teste como técnica de validação, ou seja, nenhuma amostra deste conjunto foi classificada incorretamente.

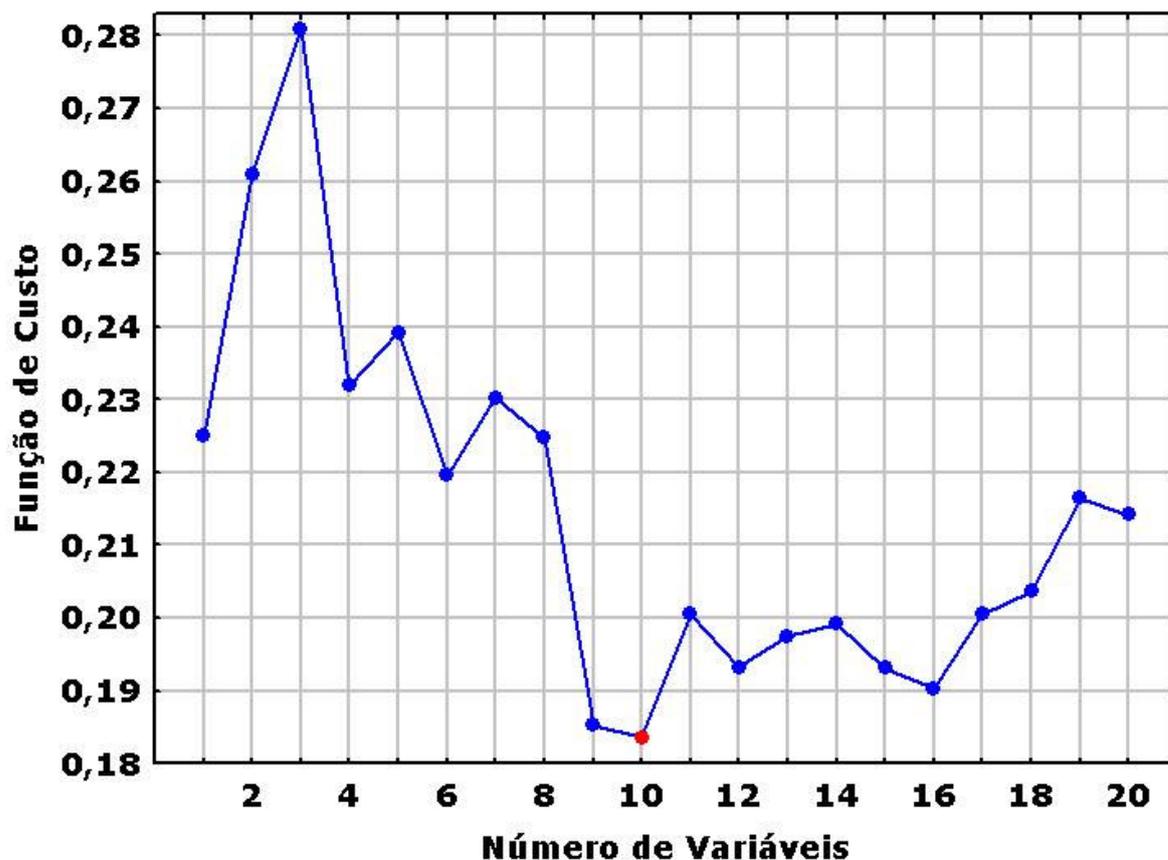


Figura 4.23 Número de variáveis selecionadas pelo SPA *versus* os valores da função de custo.

A [Figura 4.24](#) contém o voltamograma de uma amostra de óleo com as variáveis selecionadas pelo SPA em destaque. Elas estão distribuídas sobre as regiões mais informativas (próximos a picos, vales e regiões de potencial de meia onda) distanciadas em intervalos que variam de 43,8 a 135,7 mV. De acordo com [COCCHI et al. \(2003\)](#), quando a diferença entre dois potenciais de pico é menor que 100 mV, sérios problemas de sobreposição está ocorrendo no voltamograma da técnica SWV. Isto mostra a eficiência do SPA em buscar as variáveis mais adequadas, contornando o problema de sobreposição de picos.

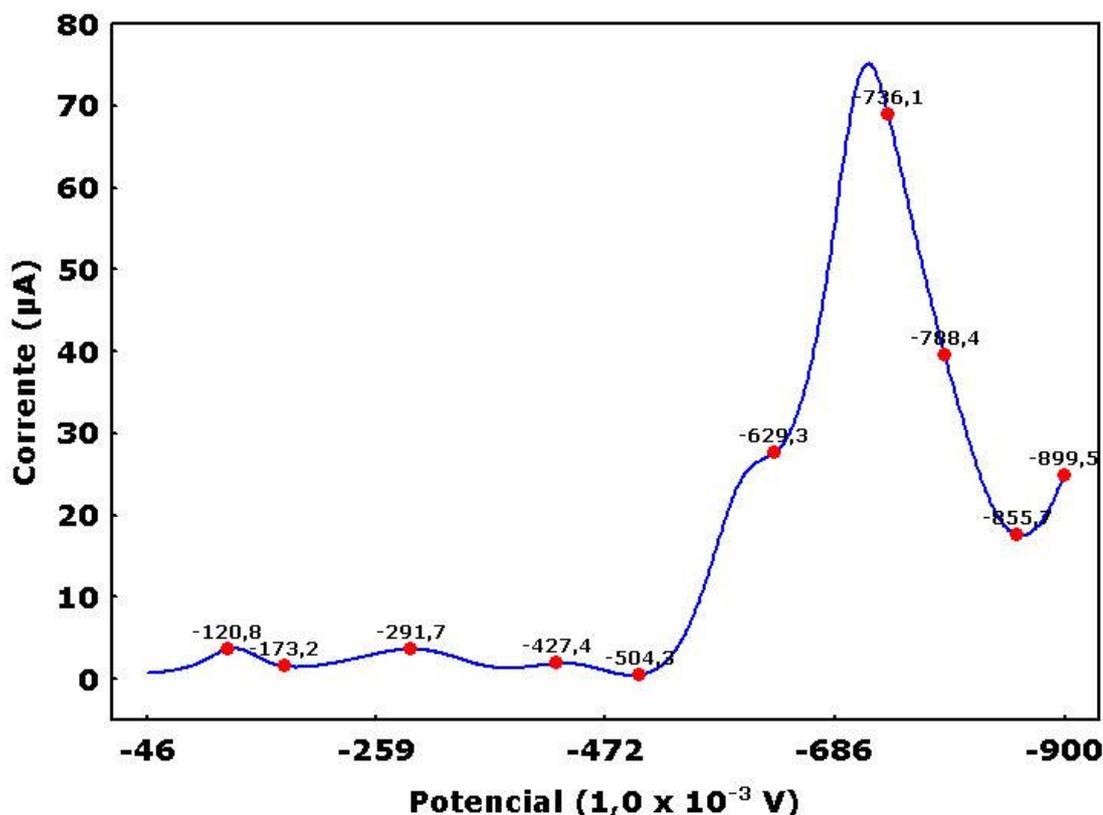


Figura 4.24 Variáveis selecionadas pelo LDA-SPA.

4.10.1 Uso do modelo SWV-LDA-SPA no Conjunto de Previsão

O modelo SWV-LDA-SPA construído com as dez variáveis expostas na Figura 4.24, foi aplicado na classificação das amostras do conjunto de previsão. Nenhum erro ocorreu, ou seja, todas as amostras deste conjunto foram corretamente classificadas.

4.10.2 Uso do SWV-LDA-SPA para a previsão de todas as amostras

O modelo SWV-LDA-SPA construído para as cinco classes foi aplicado para todas as 114 amostras analisadas e novamente nenhum erro foi encontrado. Para ilustrar esse resultado, os escores da função discriminante 1 (FD1) versus os escores da função discriminante 2 (FD2) são mostrados na Figura 4.25 e, como esperado, a classe PVE está bem discriminada das demais ao longo dos escores positivos da FD1. A FD2 discrimina dois grupos, um formado pelas classes Canola e Milho com escores positivos, e o outro, formado pelas classes Soja e Girassol com

escores negativos. Todavia, como nenhum erro de classificação foi encontrado com o modelo SWV-LDA-SPA, provavelmente as informações necessárias para discriminar estes dois grupos encontram-se nos escores das funções FD3 e FD4.

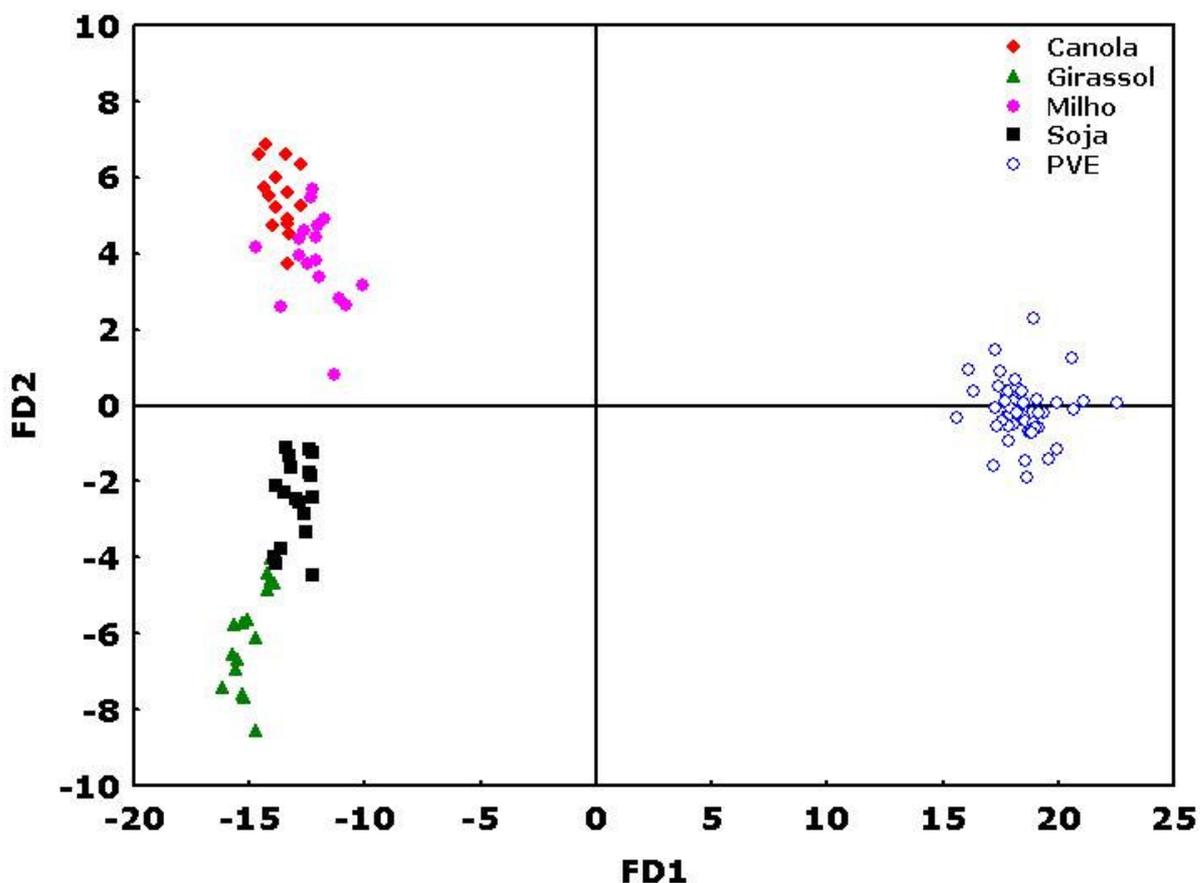


Figura 4.25 Gráfico dos escores de FD1 *versus* FD2 das 114 amostras de óleo vegetal.

Para melhor observar a dispersão entre as quatro classes de óleos não vencidos que formam dois grupos ao longo dos escores de FD2 na [Figura 4.25](#), foi construído o gráfico FD2 *versus* FD3 ([Figura 4.26](#)) onde um comportamento interessante pode ser observado: cada classe, praticamente, está sendo discriminada por um quadrante formando pelos escores positivos e negativos da FD2 e FD3.

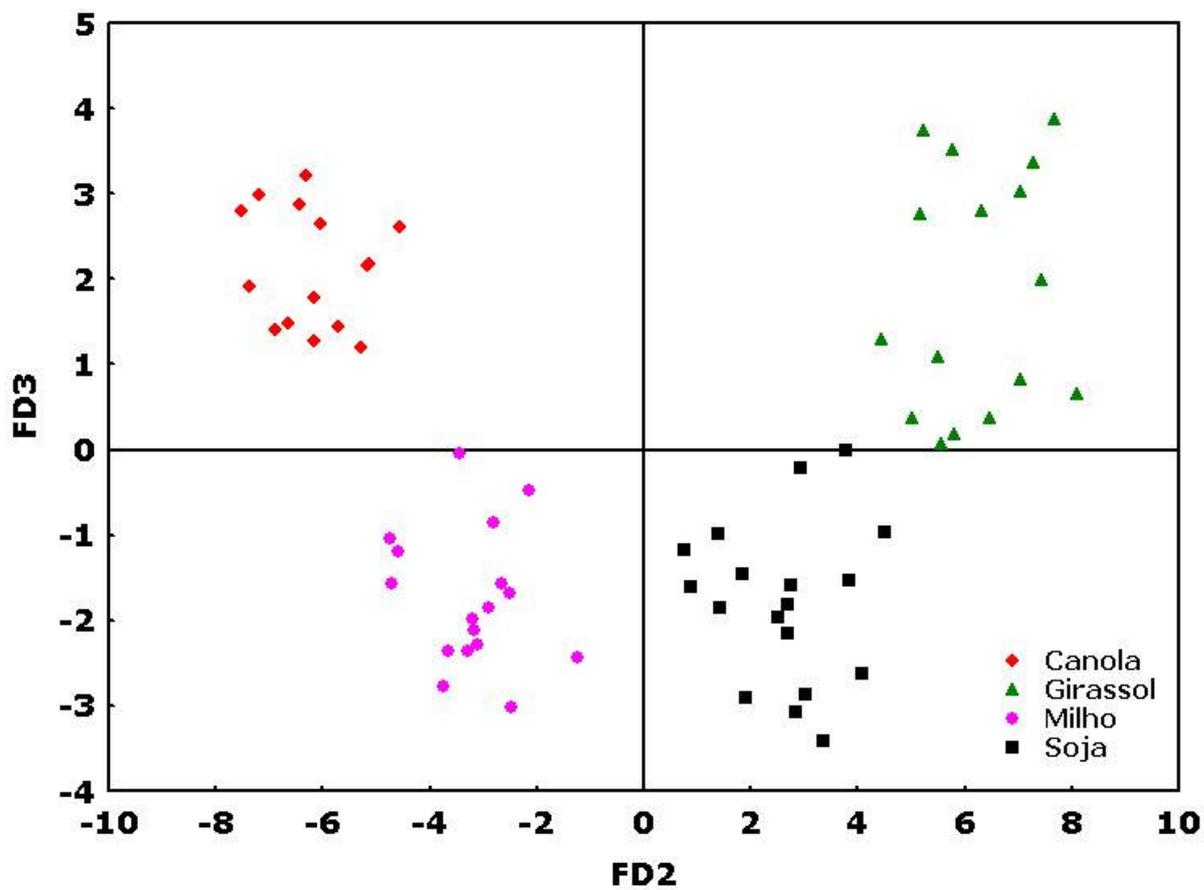


Figura 4.26 Gráfico dos escores de FD2 versus FD3 para as 66 amostras não vencidas de óleo vegetal distribuídas em quatro classes.

Observa-se nas Figuras 4.25 e 4.26 que as diferentes classes encontram-se bem discriminadas pelo modelo SWV-LDA-SPA construído, corroborando com a inexistência de erro na classificação das amostras de todo conjunto de dados.

4.11 Comparação entre os modelos SWV-SIMCA e SWV-LDA-SPA

Para efeito de comparação com o SWV-LDA-SPA, modelos SWV-SIMCA (com quatro níveis de significância do teste F – 1, 5, 10 e 25%) previamente validados foram avaliados quanto ao seu potencial para a classificação de óleos vegetais. Estes foram aplicados ao conjunto de previsão e os resultados são mostrados na Tabela 4.12.

Tabela 4.12 Erro de classificação dos modelos SWV-SIMCA (com quatro níveis de significância) e SWV-LDA-SPA para as amostras de óleo do conjunto de previsão.

	SWV-LDA-SPA	SWV-SIMCA			
		(1%)	(5%)	(10%)	(25%)
Erro Tipo I	0	0	0	0	3
Erro Tipo II	0	6	6	6	3
Total de Erros	0	6	6	6	6

Os erros de classificação SWV-SIMCA para os níveis de significância 1%, 5%, 10% e 25% permaneceram constantes, totalizando um índice de acerto de predição de 81%. Como pode ser observado na [Tabela 4.12](#), os erros do tipo II dos modelos SWV-SIMCA diminuem e do tipo I aumentam com um nível de significância 25%. Isto ocorreu porque, em geral, com o aumento do nível de significância do teste F , as fronteiras definidas por estes modelos diminuem, favorecendo uma má classificação de suas amostras. Por outro lado, o SWV-LDA-SPA não apresentou erros de classificação em qualquer conjunto em que foi testado.

Diante destes resultados, evidencia-se que o método proposto SWV-LDA-SPA é uma alternativa promissora para a classificação e fiscalização de óleos vegetais refinados com relação ao prazo de validade.

CAPÍTULO V

CONCLUSÕES

5.1 Conclusões

- Com intuito de auxiliar no controle de qualidade de óleos vegetais, este trabalho propôs uma nova metodologia que usa a voltametria de onda quadrada (SWV), o algoritmo das projeções sucessivas (SPA) e a análise discriminante linear (LDA) para a classificação de diferentes tipos de amostras de óleos vegetais refinados e a fiscalização destes óleos com relação ao prazo de validade.
- Planejamentos composto central e o parâmetro de Taguchi foram utilizados para otimizar a etapa de limpeza eletroquímica do eletrodo de trabalho e os parâmetros da SWV. Estes estudos garantiram uma boa reprodução da superfície do eletrodo de trabalho, bem como uma boa reprodutibilidade dos voltamogramas das amostras, viabilizando, então, a busca de uma modelagem quimiométrica com poder discriminante suficiente para distinção das classes modeladas.
- A análise exploratória (PCA) realizada para todo o conjunto de dados proporcionou a identificação da viabilidade da construção de modelos quimiométricos para a classificação de óleos vegetais e a fiscalização desses óleos com relação ao prazo de validade dos mesmos.
- Modelos SWV-SIMCA foram construídos para as cinco classes de óleo vegetal analisadas. O índice de acerto de predição foi de 81% para todos os níveis de significância investigados (1%, 5%, 10% e 25%).
- O SWV-LDA-SPA foi avaliado utilizando os mesmos conjuntos de amostras empregados na análise exploratória (PCA) e na modelagem SWV-SIMCA. Apenas 10 variáveis foram selecionadas pelo SPA e todas as amostras foram corretamente classificadas.
- A metodologia proposta (SWV-LDA-SPA) é uma alternativa promissora para classificação e fiscalização de diferentes tipos de óleos vegetais refinados com relação ao prazo de validade dos mesmos. A mesma poderá ser empregada no controle de qualidade de forma rápida, eficiente e com baixo custo por análise.

5.2 Propostas futuras

Pretende-se, como continuidade deste trabalho:

- 1 Automatizar as medidas com o intuito de aumentar a frequência analítica.
- 2 Aumentar o número de classes e amostras de óleos vegetais, além de avaliar a capacidade de previsão dos modelos utilizando amostras adulteradas.
- 3 Construir modelos de calibração multivariada para determinação de diversos parâmetros de interesse no controle de qualidade (índices de acidez, peróxido, saponificação, entre outros).
- 4 Estudar os mecanismos de oxidação e redução eletroquímica dos compostos do óleo presentes na cela eletroanalítica.

CAPÍTULO VI

REFERÊNCIAS

ABOISSA. Composição Química da Semente de Girassol. Disponível em: <<http://www.aboissa.com.br/girassol/compq.htm>>. Acessado em: 07/11/2007.

ALVES, M. R. et al., Classification of PDO olive oils on the basis of their sterol composition by multivariate analysis. *Analytica Chimica Acta*, **549: 166, 2005**.

ANTUNES, M. C. et al., Application of multivariate curve resolution to the voltammetric study of the complexation of fulvic acids with cadmium(II) ion. *Analytica Chimica Acta*, **459: 291, 2002**.

ANVISA. Resolução nº 482, de 23 de setembro de 1999. Disponível em: <http://www.anvisa.gov.br/legis/resol/482_99.htm>. Acessado em: 20/12/2007.

AOCS. *Official methods and recommended practices of the American Oil Chemists' Society*. 5. ed. Champaign, 1998.

ARAUJO, M. C. U. et al., The Successive Projections Algorithm for variable selection in spectroscopy multicomponent analysis. *Chemometrics And Intelligent Laboratory Systems*, **57: 65, 2001**.

BAETEN, V. et al., Oil and fat classification by FT-Raman spectroscopy. *Journal of Agriculture and Food Chemistry*, **46: 2638, 1998**.

BARD, A. J.; FAULKNER, L. R., *Electrochemical Methods Fundamentals and Applications*. 2ª ed. New York: John Wiley & Sons, INC, 2001.

BARROS-NETO, B.; SCARMINIO, I. S.; BRUNS, R. E. *Como Fazer Experimentos*. Campinas: Editora Unicamp, 2003.

BARTHUS, R. C.; MAZO, L. H.; POPPY, R. J., Determinação simultânea de NADH e ácido ascórbico usando voltametria de onda quadrada com eletrodo de carbono vítreo e calibração multivariada. *Eclética Química*, **30: 51, 2005**.

BARTHUS, R. C.; POPPI, R. J., Determination of the total unsaturation in vegetable oils by Fourier transform Raman spectroscopy and multivariate calibration. *Vibrational Spectroscopy*, **26: 99, 2001**.

BECKSTEIN, A. Produção de grãos em 2007 cresce 13,7% e bate recorde de 2003, estima IBGE. Agência Brasil, janeiro de 2008. Disponível em: <<http://www.agenciabrasil.gov.br/noticias/2008/01/08/materia.2008-01-08.9657585151/view>>. Acessado em: 08/01/2008.

BEEBE, K. R.; PELL, R. J.; SEASHOLTZ, M. B. *Chemometrics A Practical Guide*. New York: John Wiley & Sons, 1998.

BERRETTONI, M. et al., Coupling chemometrics and electrochemical-based sensor for detection of bacterial population. *Analytica Chimica Acta*, **509: 95, 2004a**.

BERRETTONI, M. et al., Electrochemical sensor for indirect detection of bacterial population. *Sensors and Actuators B*, **102: 331, 2004b**.

BESSANT, C.; SAINI, S., Simultaneous determination of ethanol, fructose and glucose at an unmodified platinum electrode using artificial neural networks. *Analytical Chemistry*, **71: 2806, 1999**.

BESSANT, C.; SAINI, S., A chemometric analysis of dual pulse staircase voltammograms obtained in mixtures of ethanol, fructose and glucose. *Journal of Electroanalytical Chemistry*, **489: 76, 2000**.

BIODIESELBR, Biodiesel começa a ser comercializado no DF a partir de amanhã. Biodieselbr Online, dezembro de 2005. Disponível em: <<http://www.biodieselbr.com/noticias/biodiesel/biodiesel-comeca-a-ser-comercializado-no-df-a-partir-de-amanha-13-12-05.htm>>. Acessado em: 08/03/2008.

BRUNS, R. E.; FAIGLE, J. F. G., Quimiometria. *Química Nova*, **8: 84, 1985.**

CALADO, V.; MONTGOMERY, D. C., *Planejamento de Experimentos Usando o Statistica*. Rio de Janeiro: E-papers, 2003.

CAMO S.A. *Manual do Usuário*. UNSCRAMBLER, versão 7.5. Noruega, 1998.

CAMPESTRE. Óleo de Girassol. Disponível em: <<http://www.campestre.com.br/oleo-de-girassol.shtml>>. Acessado em: 15/07/2006.

CAMPESTRE. Óleo de Soja. Disponível em: <<http://www.campestre.com.br/oleo-de-soja.shtml>>. Acessado em: 12/06/2007.

CATHARINO, R. R. et al., Characterization of vegetable oils by Electrospray Ionization Mass Spectrometry Fingerprinting: Classification, Quality, Adulteration, and Aging. *Analytical Chemistry*, **77: 7429, 2005.**

CEOTTO, B., A vez da linhaça. Disponível em: <http://nova.abril.com.br/edicoes/35/fechado/beleza_saude/conteudo_92_033.shtml>. Acessado em 08/03/2008

CHRISTIE, J. H.; TURNER, J. A.; OSTERYOUNG, R. A., Square Wave Voltammetry at the Dropping Mercury Electrode: Theory. *Analytical Chemistry*, **49: 1899, 1977.**

CHRISTY, A. A. et al., The detection and quantification of adulteration in olive oil by near-infrared spectroscopy and chemometrics. *Analytical Sciences*, **20: 935, 2004.**

CHRISTY, A. A.; EGEBERG, P. K., Quantitative determination of saturated and unsaturated fatty acids in edible oils by infrared spectroscopy and chemometrics. *Chemometrics and Intelligent Laboratory Systems*, **82: 130, 2006.**

COCCHI, M. et al., Multicomponent analysis of electrochemical signals in the wavelet domain. *Talanta*, **59: 735, 2003.**

COSTA, R. S.; SANTOS et al., A novel strategy to verification of adulteration in alcoholic beverages based on Schlieren effect measurements and chemometric techniques. *Microchemical Journal*, **78: 27, 2004.**

COUTINHO, L. G. et al. Estudo da Competitividade da Indústria Brasileira - Competitividade na Indústria de Óleos Vegetais. Ministério da Ciência e Tecnologia. Disponível em: <http://www.mct.gov.br/upd_blob/0002/2225.pdf>. Acessado em: 09/01/2008.

CRUZ, B. H. et al., Multivariate curve resolution of polarographic data applied to the study of the copper-binding ability of tannic acid. *Analytica Chimica Acta*, **424: 203, 2000.**

CRUZ, B. H. et al., Differential pulse voltammetric study of the complexation of Cd(II) by the phytochelatin (γ -Glu-Cys)₂Gly assisted by multivariate curve resolution. *Journal of Electroanalytical Chemistry*, **520: 111, 2002.**

CUKROWSKA, E. et al., Use of artificial neural networks for the evaluation of electrochemical signals of adenine and cytosine in mixtures interfered with hydrogen evolution. *Journal of Electroanalytical Chemistry*, **503: 117, 2001.**

DANTAS-FILHO, H. A. et al., A strategy for selecting calibration samples for multivariate modelling. *Chemometrics And Intelligent Laboratory Systems*, **72: 83, 2004.**

DIAZ, T. G. et al., Voltammetric behavior and determination of tocopherols with partial least squares calibration: analysis in vegetable oil samples. *Analytica Chimica Acta*, **511: 231, 2004.**

DIAZ, T. G. et al., Characterization of virgin olive oils according to its triglycerides and sterols composition by chemometric methods. *Food Control*, **16: 339, 2005.**

Díaz-Cruz, M. S. et al., Soft- and Hard-Modeling Approaches for the Determination of Stability Constants of Metal±Peptide Systems by Voltammetry. *Analytical Biochemistry*, **279: 189, 2000.**

DÍAZ-CRUZ, M. S.; ESTEBAN, M.; RODRÍGUEZ, A. R., Square wave voltammetry data analysis by multivariate curve resolution: application to the mixed-metal system Cd-Zn-{Lys-Cys-Thr-Cys-Cys-Ala}. *Analytica Chimica Acta*, **428: 285, 2001.**

DÍAZ-CRUZ, M. S. et al., Comparison of the zinc–cadmium exchange properties of the metallothionein related peptide {Lys–Cys–Thr–Cys–Cys–Ala} and a zinc-containing metallothionein: study by voltammetry and multivariate curve resolution. *Journal of Electroanalytical Chemistry*, **523: 114, 2002.**

EMBRAPA. Boletim de Pesquisa Online Nº 6, jun./01. Disponível em: <http://www.cnpt.embrapa.br/biblio/p_bo06_3.htm>. Acessado em: 20/12/2007b.

EMBRAPA. Definição e Histórico de Canola. Disponível em: <<http://www.cnpt.embrapa.br/culturas/canola/definicao.htm>>. Acessado em: 21/12/2007a.

EMBRAPA. Soja na Alimentação. Disponível em: <http://www.cnpso.embrapa.br/soja_alimentacao/index.php?pagina=23>. Acessado em: 15/10/2007c.

ESTEBAN, M. et al., Multivariate curve resolution with alternating least squares optimisation: a soft-modelling approach to metal complexation studies by voltammetric techniques. *Trends in Analytical Chemistry*, **19: 49, 2000.**

ESTEBAN, M.; ARIÑO, C.; DÍAZ-CRUZ, J. M., Chemometrics for the analysis of voltammetric data. *Trends in Analytical Chemistry*, **25: 86, 2006a.**

ESTEBAN, M.; ARIÑO, C.; DÍAZ-CRUZ, J. M., Chemometrics in Electroanalytical Chemistry. *Critical Reviews in Analytical Chemistry*, **36:295, 2006b.**

FERNÁNDEZ, M. et al., Soft modelling approach applied to voltammetric data: study of electrochemically labile metal–glycine complexes. *Journal of Electroanalytical Chemistry*, **505: 44, 2001**.

FERREIRA, M. M. C. et al., Quimiometria I: calibração multivariada, um tutorial. *Química Nova*, **22: 724, 1999**.

FLATEN, G. R.; GRUNG, B.; KVALHEIM, O. M., A method for validation of reference sets in SIMCA modeling, *Chemometrics And Intelligent Laboratory Systems*, **72: 101, 2004**.

FRANCO, D. et al., Applicability of NIR spectroscopy to determine oil and other physicochemical parameters in *Rosa mosqueta* and *Chilean hazelnut*. *European Food Research and Technology*, **222: 443, 2006**.

FUNG, Y.; MO, S., Determination of ethanol in beer by flow injection dual-pulse staircase voltammetric detection. *Analyst*, **121: 369, 1996**.

GALVÃO, R. K. H., et al., Estudo comparativo sobre filtragem de sinais instrumentais usando transformada de Fourier e de Wavelet. *Química Nova*, **24: 874, 2001a**.

GALVÃO, R. K. H. et al., Aspects of the Successive Projections Algorithm for variable selection in multivariate calibration applied to Plasma Emission. *Analytica Chimica Acta*, **443: 107, 2001b**.

GALVÃO, R. K. H. et al., Optimal wavelet filter construction using X and Y data. *Chemometrics And Intelligent Laboratory Systems*, **70: 1, 2004**.

GAZZONI, D. L. As Políticas Públicas de Biocombustíveis e o Mercado de Oleaginosas. EMBRAPA. Disponível em:

<<http://www.iea.usp.br/iea/online/midiateca/politicaspUBLICASgazzoni.pdf>>. Acessado em: 20/12/2007.

GESTAL, M. et al., Classification of apple beverages using artificial neural networks with previous variable selection. *Analytica Chimica Acta* 524: 225, 2004.

GUIBERTEAU, A. et al., Polarography and artificial neural network for the simultaneous determination of nalidixic acid and its main metabolite (7-hydroxymethylnalidixic acid). *Talanta*, **62: 357, 2004.**

GUILLÉN, M. D.; CABO, N., Infrared spectroscopy in the study of edible oils and fats. *Journal of the Science of Food and Agriculture*, **75: 1, 1997.**

HENDL, O. et al., A rapid and simple method for the determination of iodine values using derivative Fourier transform infrared measurements. *Analytica Chimica Acta*, **427: 75, 2001.**

HERREO, A.; ORTIZ, M. C., Qualitative and quantitative aspects of the application of genetic algorithm-based variable selection in polarography and stripping voltammetry. *Analytica Chimica Acta*, **378: 245, 1999.**

HONORATO, F. A. et al., Analytical curve or standard addition method: how to Elect and Design – A strategy applied to copper determination in sugarcane spirits using AAS. *Analyst*, **127: 1520, 2002.**

HONORATO, F. A. et al., Robust modeling for multivariate calibration transfer by the Successive Projections Algorithm. *Chemometrics And Intelligent Laboratory Systems*. **76: 65, 2005.**

HUSAIN, S. et al., Characterization and identification of edible oil blends and prediction of the composition by artificial neural networks – A case study. *Chemometrics and Intelligent Laboratory Systems*, **35: 117, 1996.**

IBGE. Em 2008, safra de grãos deverá crescer 2,1%. Janeiro de 2008. Disponível em: <http://www.ibge.gov.br/home/presidencia/noticias/noticia_visualiza.php?id_noticia=1068&id_pagina=1>. Acessado em: 08/01/2008.

JAKUBOWSKA, M., Dedicated wavelet for voltammetric signals analysis. *Journal of Electroanalytical Chemistry*, **603: 113, 2007**.

KEMSLEY, E. K., Discriminant analysis of high-dimensional data: a comparison of principal components analysis and partial least squares data reduction methods. *Chemometrics and Intelligent Laboratory Systems*, **33: 47, 1996**.

KENNARD, R. W.; STONE, L. A., *Computer Aided Design of Experiments. Technometrics*, **11: 137, 1969**.

KOTANI, A.; FUSE, T.; KUSU, F., Determination of plasma free fatty acids by High-Performance Liquid Chromatography with Electrochemical Detection. *Analytical Biochemistry*, **284: 65, 2000**.

KOWALSKI, B. R. *Chemometrics - Mathematics and Statistics in Chemistry*. Holland: D. Reidel Publishing Company, 1984.

KRAUSE, M. V.; MAHAN, L. K. *Alimentos, Nutrição e Dietoterapia - Um livro-texto do cuidado nutricional*. 7 ed., São Paulo: Roca, 1991.

KRAUSE Jr., M. S.; RAMALEY, L., Analytical Application of Square Wave Voltammetry. *Analytical Chemistry*, **41: 1365, 1969**.

LAI, Y. W.; E. KEMSLEY, E. K.; WILSON, R. H., Potential of Fourier transform infrared spectroscopy for the authentication of vegetable oils. *Journal of Agriculture and Food Chemistry*, **42: 1154, 1994.**

LANKMAYR, E. et al., Chemometrical classification of pumpkin seed oils using UV-Vis, NIR and FTIR spectra. *Journal of Biochemical and Biophysical Methods*, **61: 95, 2004.**

LEE, D. S. et al., Characterization of fatty acids composition in vegetable oils by gas chromatography and chemometrics. *Analytica Chimica Acta*, **358: 163, 1998.**

LEE, D. S. et al., Reversed phase liquid chromatographic determination of triacylglycerol composition in sesame oils and the chemometric detection of adulteration. *Analytica Chimica Acta*, **429: 321, 2001.**

LEHNINGER, A. L.; NELSON, D. L.; COX, M. M. *Princípios de Bioquímica*. 2 ed., São Paulo: Sarvier, 1995.

MALAVIA, J. et al., Analysis of polychlorinated dibenzo-*p*-dioxins, dibenzofurans and dioxin-like polychlorinated biphenyls in vegetable oil samples by gas chromatography-ion trap tandem mass spectrometry. *Journal of Chromatography A*, **1149: 321, 2007.**

MARQUEZ, A. J.; DÍAZ, A. M.; REGUERA, M. I, P., Using optical NIR sensor for on-line virgin olive oils characterization. *Sensors and Actuators B*, **107: 64, 2005.**

MARTÍN, Y. G. et al., Classification of vegetable oils by linear discriminant analysis of Electronic Nose data. *Analytica Chimica Acta*, **384: 83, 1999.**

MASSART, D. L. et al., *Handbook of Chemometrics and Qualimetrics: Part A*. Amsterdam: Elsevier Science B. V., 1997.

MENDONÇA, C. R. B.; BICA, C. I. D.; PIATNICKI, C. M. S., Water in soybean oil microemulsions as medium for electrochemical measurements. *Journal of the Brazilian Chemical Society*, **14: 628, 2003**.

MORENO, L. et al., Analysis of amino acids in complex samples by using voltammetry and multivariate calibration methods. *Analytica Chimica Acta*, **507: 247, 2004**.

MORETTO, E.; FEET, R. *Tecnologia de Óleos e Gorduras vegetais na Indústria de Alimentos*. São Paulo: Varela Editora e Livraria, 1998.

NAGY, K. et al., High performance liquid chromatography-mass spectrometry based chemometric characterization of olive oils. *Journal of Chromatography A*, **1078: 90, 2005**.

NI, Y.; WANG, L.; KOKOT, L., Voltammetric determination of butylated hydroxyanisole, butylated hydroxytoluene, propyl gallate and tert-butylhydroquinone by use of chemometric approaches. *Analytica Chimica Acta*, **412: 185, 2000**.

NI, Y.; WANG, L.; KOKOT, S., Simultaneous determination of nitrobenzene and nitro-substituted phenols by differential pulse voltammetry and chemometrics. *Analytica Chimica Acta*, **431: 101, 2001a**.

NI, Y.; WANG, L.; KOKOT, S., Voltammetric determination of chlorpromazine hydrochloride and promethazine hydrochloride with the use of multivariate calibration. *Analytica Chimica Acta*, **439: 159, 2001b**.

NI, Y.; QIU, P.; KOKOT, S., Simultaneous determination of three organophosphorus pesticides by differential pulse stripping voltammetry and chemometrics. *Analytica Chimica Acta*, **516: 7, 2004**.

NI, Y.; QIU, P.; KOKOT, S., Simultaneous voltammetric determination of four carbamate pesticides with the use of chemometrics. *Analytica Chimica Acta*, **537: 321, 2005**.

PALACIOS-SANTANDER, J. M. et al., Use of artificial neural networks, aided by methods to reduce dimensions, to resolve overlapped electrochemical signals. A comparative study including other statistical methods. *Microchimica Acta*, **142: 27, 2003**.

PAPADOPOULOS K. et al., Investigations of the adulteration of extra virgin olive oils with seed oils using their weak chemiluminescence. *Analytica Chimica Acta*, **464: 135, 2002**.

PIMENTEL, M. F.; SALDANHA, T. C. B.; ARAÚJO, M. C. U., Effects of experimental design on calibration curve precision in routine analysis, *Journal of Automatic Chemistry*, **20: 9, 1998**.

PONTES, M. J. C. et al., Classification of distilled alcoholic beverages and verification of adulteration by near infrared spectrometry. *Food Research International*, **39: 182, 2006**.

PONTES, M. J. C. et al., The Successive Projections Algorithm for spectral variable selection in classification problems. *Chemometrics And Intelligent Laboratory Systems*, **78: 11, 2005**.

POULLI, K. I.; MOUSDIS, G A.; GEORGIU, C. A., Classification of edible and lampante virgin olive oil based on synchronous fluorescence and total luminescence spectroscopy. *Analytica Chimica Acta*, **542: 151, 2005**.

POULLI, K. I.; MOUSDIS, G. A; GEORGIU, C. A., Synchronous fluorescence spectroscopy for quantitative determination of virgin olive oil adulteration with sunflower. *Analytical and Bioanalytical Chemistry*, **386: 1571, 2006.**

RAMALEY, L.; KRAUSE Jr., M. S., Theory of Square Wave Voltammetry. *Analytical Chemistry*, **41: 1362, 1969.**

REGINA R.; SOLFERINI O. Produção de cultivares de ingredientes de alto valor nutricional: Características e benefícios. Disponível em: <http://www.gtconsult.com.br/ingles/artigos/producao_cultivares.pdf>. Acessado em: 18/10/07.

RODRIGUES, J. N.; GIOIELLI L. A.; ANTON C., Propriedades físicas de lipídios estruturados obtidos de misturas de gordura do leite e óleo de milho. *Ciência e Tecnologia de Alimentos*, **23: 226, 2003.**

RUBINSON, K. A.; RUBINSON, J. F. *Análisis Instrumental*. Madrid: Pearson Educación, S.A., 2001.

SANZ, M. B. et al., Capability of discrimination: application to soft calibration methods. *Analytica Chimica Acta*, **446: 297, 2001.**

SAURINA, J. et al., Cyclic voltammetric simultaneous determination of oxidizable amino acids using multivariate calibration methods. *Analytica Chimica Acta*, **405: 153, 2000.**

SCHREYER, S. K.; MIKKELSEN, S. R., Chemometric analysis of square wave voltammograms for classification and quantitation of untreated beverage samples. *Sensors and Actuators B*, **71: 147, 2000.**

SIKORSKA, E. et al., Classification of edible oils using synchronous scanning fluorescence spectroscopy. *Food Chemistry*, **89: 217, 2005**.

SKOOG A. D.; HOLLER, F. J.; Nieman, T. A., *Princípios de Análise Instrumental*. 5 ed. Porto Alegre: Bookman, 2002.

SMITH, A.K.; CIRCLE, S.J. *Soybeans: Chemistry and Technology*. Westport : The AVI Publishing Company, INC., 1972, V. 1.

SOUZA, D. et al., Voltametria de Onda Quadrada. Segunda Parte. Aplicações. *Química Nova*, **27: 790, 2004**.

SOUZA, D.; MACHADO, S. A. S.; AVACA, L. A., Voltametria de Onda Quadrada. Primeira Parte: Aspéctos Teóricos. *Química Nova*, **26: 81, 2003**.

STEEL, S. J.; LOUW, N.; ROUX, N. J. L., A comparison of the post selection error rate behavior of the normal linear and quadratic discriminant rules. *Journal of Statistical Computation And Simulation*, **65: 157, 2000**.

STEPANYAN, V. et al., Chemiluminescent evaluation of peroxide value in olive oil. *Talanta*, **65: 1056, 2005**.

TAKAMURAT, K.; FUSE, T.; KUSU, F., Determination of the free fatty acid contents in fats and oils by Flow Injection Analysis by Electrochemical Detection. *Analytical Sciences*, **11: 979, 1995**.

TURNER, J. A. et al., Square Wave Voltammetry at the Dropping Mercury Electrode: Experimental. *Analytical Chemistry*, **49: 1904, 1977**.

VONCINA, D. B.; KODBA, Z. C.; NOVIC, M., Multivariate data analysis in classification of vegetable oils characterized by the content of fatty acids. *Chemometrics and Intelligent Laboratory Systems*, **75: 31, 2005**.

WANG, J. *Analytical Electrochemistry*. 3 ed. New Jersey: John Wiley & Sons, Inc., 2006.

Wang, L. et al., Feasibility study of quantifying and discriminating soybean oil adulteration in camellia oils by attenuated total reflectance MIR and fiber optic diffuse reflectance NIR. *Food Chemistry*, **95:529, 2006**.

WEISS, T. J. *Food Oils and Their Uses*. London: The Avi Publishing Company, Inc., 1970.

WU, W.; MASSART, D. L., Regularized nearest neighbor classification method for pattern recognition of near infrared spectra. *Analytica Chimica Acta*, **349: 253, 1997**.

YANG, H.; IRUDAYARAJ, J.; PARADKAR, M. M., Discriminant analysis of edible oils and fats by FTIR, FT-NIR and FT-Raman spectroscopy. *Food chemistry*, **93: 25, 2005**.

ZHANG, G. et al., Authentication of vegetable oils on the basis of their physico-chemical properties with the aid of chemometrics. *Talanta*, **70: 293, 2006**.