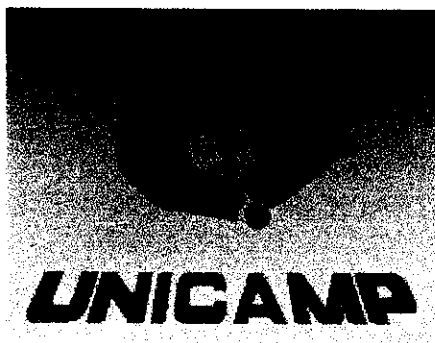


**UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE QUÍMICA**



**“Estudo quimiométrico do complexo *Oenocarpus-
Jessenia da Amazônia*”**

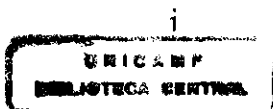
Tese de Doutorado

Juan Manuel Rojas Amasifén

Orientador: *Prof. Dr. Roy Edward Bruns*

Barão Geraldo-Campinas-São Paulo-Brasil

Agosto-2001





Aos Seres que são a razão da minha vida:

Nelly, Isis e Xendor-Roberto-Miguel

Aos meus irmãos César, Julian, Miguel,

Walter, Lina, Yolanda e Nimia.

À Memória dos meus pais que até hoje,

existem no meu coração e morarão na

minha mente pela eternidade: Miguel

Antonio e Maria Alfonsa.

AGRADECIMENTOS

- Ao professor Dr. **Roy Edward Bruns**, pela orientação, apoio, e ensinamentos recebidos durante o curso de doutorado e a execução deste trabalho;
- Aos professores **Harry Serruya, Maria H. da Silva Bentes e Geraldo N. Rocha Filho** da UFPa, pelo importante apoio nesta tese;
- Aos professores **Carol, Ronel, Isabel Jardim e Isabel Bueno** pelas valiosas sugestões no exame de qualificação de área;
- A todos os professores e funcionários do Instituto de Química da Unicamp que ajudaram neste trabalho de tese;
- À Faculdade de Engenharia Química da Universidade Nacional da Amazônia Peruana, pela licença concedida e aos professores **Carlos, Uberlando, Summer e Rafael** que contribuíram com seu apoio;
- Ao Dr. **Milton Motta e Juan Celedonio Ruiz**, pelo grande apoio fornecido na coleta das amostras no Brasil e no Perú respectivamente;
- Aos meus amigos: **Anselmo, Marlon, Roberto, Sayuri, Marisa, Cristiano e Gaspar** que contribuíram, mediante seu apoio neste trabalho de tese;
- Ao Conselho Nacional de Pesquisas (CNPq), pela bolsa de estudos;
- A Deus e todos os seus anjos pela sua infinita bondade por deixar-me morar nesta terra.

RESUMO

Título : “Estudo quimiométrico do complexo *Oenocarpus-Jessenia* da Amazônia”.
Autor : Juan Manuel Rojas Amasifén
Orientador : Prof. Dr. Roy Edward Bruns
Palavras chave: *Oenocarpus*; *Jessenia*; palmeiras; quimiometria, classificação, calibração multivariada, espectroscopia.

O complexo *Oenocarpus-Jessenia* (*bacaba-pataúá*) constitui de palmeiras cuja distribuição geográfica estende-se pela região da Amazônia na América do Sul. O óleo dos frutos destas plantas é semelhante ao azeite de oliva e daí sua importância. Estudos de identificação dos gêneros do complexo com base às características quimiotaxonômicos foram realizados. O estudo das variáveis morfológicas, químicas, físico-químicas e espectrais para fins de exploração, classificação e calibração com fins de melhoramento genético, agrônomo e econômico ainda não foi relatado.

Neste trabalho foram caracterizados duzentos frutos dos cachos, de cada uma, das 25 plantas pertencentes aos gêneros *Oenocarpus* e *Jessenia*, coletados no estado do Pará-Brasil e na região de Loreto-Perú no período de 1993 a 1997.

Através da caracterização morfológica determinou-se a massa e as dimensões do fruto, da casca+polpa e da semente. A caracterização química foi realizada sobre a farinha seca da casca+polpa e sobre os óleos obtidos desta; a determinação da composição dos óleos em ácidos graxos foi realizada mediante cromatografia a gás; os espectros de ultravioleta (UV/Vis) obtiveram-se dos extratos de éter de petróleo da farinha da casca+polpa. Os teores dos elementos quais foram obtidos por fluorescência de raios-X de energia dispersiva (EDXRF). Os espectros de infravermelho médio com transformada de Fourier (FTIR) e espectros de difração de raios-X (XRD) foram obtidos das cinzas da casca+polpa.

Os dados foram organizados em matrizes e aplicados métodos quimiométricos de exploração; análise de componentes principais (PCA) e análise de agrupamento hierárquica (HCA); métodos de classificação, modelos de similaridade com componentes principais (SIMCA) e regra dos K vizinhos mais próximos (KNN), e métodos de calibração; regressão de componentes principais (PCR) e regressão de mínimos quadrados parciais (PLS).

Os resultados da classificação sobre um conjunto de treinamento de 19 amostras e um conjunto teste de cinco amostras de plantas pertencentes aos dois gêneros, mostraram que com dados morfológicos é possível classificar os dois gêneros e talvez até as espécies do gênero *Oenocarpus*. Entretanto, com dados de teor de elementos obtidos a partir das cinzas, desenvolveu-se um método alternativo para classificá-los. Nos dados de UV/Vis foram os flavonóides que serviram para melhor discriminar entre estes gêneros. Com espectros de FTIR foram obtidos modelos para os dois gêneros, sendo as absorções decorrentes dos grupos metal-oxigênio dos óxidos e da carbonila dos carbonatos as mais importantes. Os dados de composição em ácidos graxos não permitiram classificar estes gêneros, sugerindo que poderiam ser empregadas para obter um modelo de controle de qualidade dos óleos deste complexo.

Foram desenvolvidos modelos de calibração multivariada sobre um conjunto de calibração de 19 amostras usando espectros de FTIR das cinzas da casca+polpa, como variáveis independentes, e teor dos elementos químicos obtidos por EDXRF, como variáveis dependentes. Estes modelos aplicados a um conjunto de teste de cinco amostras determinaram aproximadamente os teores dos elementos químicos, sendo os modelos de K e Cu mais exatos, seguido por aqueles de Fe, Mn, P, Ca e Zn.

ABSTRACT

Title : "Chemometric study of *Oenocarpus-Jessenia* complex palms from the Amazon".
Author : Juan Manuel Rojas Amasifén
Advisor : Prof. Dr. Roy Edward Bruns
Keywords: *Oenocarpus*; *Jessenia*; palms; chemometrics, classification, multivariate calibration, spectroscopy.

The *Oenocarpus-Jessenia* complex is a group of palms found over much of the northern half of South America mainly in the Amazon region. The oil extracted from these palms is similar both chemically and in taste to olive oil. Morphological and chemotaxonomic studies in order to discriminate between genera have been made, but fruit morphological, physicochemical and spectroscopic characterization using chemometric methods like exploratory and classification analyses and multivariate calibration for genetic, agronomic and economic improvement have not yet been reported.

In this study 200 fruits for each one of the 25 plants belonging to *Oenocarpus* and *Jessenia* genera, collected between 1993-1997 from the Pará, Brasil and Loreto, Perú regions were characterized.

Morphological characterization consisted of determining weight and dimensions of the fruit and its parts: peel, mesocarp and seed. Dry peel+mesocarp flour and its oils were used for chemical characterization and fatty acid composition was determined by gas chromatography. Ultraviolet-visible (UV/Vis) spectra were obtained using etheric extract of the dry flour. Characterizations were performed on flour peel+mesocarp ashes by energy-dispersive x-ray fluorescence (EDXRF), x-ray diffraction (XRD) and Fourier transform infrared (FTIR) in the mid-infrared region.

Principal component (PCA) and hierarchical cluster analysis (HCA) were used as exploratory analytical methods, K-nearest neighbor (KNN) and soft independent modeling of class analogies (SIMCA) for classification, and principal component regression (PCR) and partial least squares (PLS) regression for multivariate calibration.

Pattern recognition (PR) results obtained on a training set of 19 samples and a test set of 5 samples belonging to *Oenocarpus* and *Jessenia* genera, showed that morphological data were able to classify the two genera as well as the *Oenocarpus*' species. Meanwhile, the metal contents of these plants has been shown to be an alternative method to differentiate the two

genera. Flavonoids discriminated *Oenocarpus* and *Jessenia* using UV/Vis data. The metal-oxygen bond from oxides and carbonyl group from carbonate compounds provided the more important absorptions in FTIR spectra. Oil fatty acid compositions were not able to discriminate between genera but might be used to develop models for oil quality control for this studied complex.

Multivariate calibration models for a calibration set of 19 samples and a test set of 5 samples were developed for predicting metal content from peel+mesocarp ashes using FTIR spectra as independent variables and EDXRF metal contents as dependent variables. Calibration models providing approximate metal contents in ashes were obtained. K and Cu gave the best models, followed by Fe, Mn, P, Ca and Zn in decreasing order of quality.

CURRICULUM VITAE

1.-Dados Pessoais

Nome : Juan Manuel Rojas Amasifén

CIC : 333005502-20

Pasaporte N° 1245625

Nacionalidade : Peruana

email: amasifen@iqm.unicamp.br

e-mail ttilexx@yahoo.com Telf. 005194234405 (Perú)

2.- Formação Acadêmica

- Bacharelado em Engenharia Química pela Faculdade de Engenharia Química da Universidade Nacional da Amazônia Peruana **1978**
- Engenheiro Químico pela Faculdade de Engenharia Química da Universidade Nacional da Amazônia Peruana **1981**.
- Mestre em Química de Produtos Naturais pela Pós-graduação em Química no Centro de Ciências Exatas e Naturais da Universidade Federal do Pará **1993**.

3.-Atividades Docentes

- Docente a D.E. da Faculdade de Engenharia Química da Universidade Nacional da Amazônia Peruana desde **1981**.

4.-Treinamento

- Curso de Óleos Vegetais, outubro-**1991**
- Controle de Processos Químicos, outubro-**1992**
- Curso Métodos Modernos em RMN, novembro-**1993**
- Curso Interdisciplinar em Meio Ambiente, dezembro-**1993**
- Curso de Produtos Naturais: aspectos Químicos e Tecnológicos, 06/**1994**
- Curso Internacional de Compósitos, dezembro-**1994**
- Tratamento de resíduos gerados em laboratórios de ensino, pesquisas e prestação de serviço. Belém Pará do 15 a 18/06/**1999**.
- Introdução à Química Computacional, março-**2001**
- Introdução ao UNIX, abril-**2001**
- Introdução à Tecnologia de Redes, junho-**2001**.

5.-Produção Técnico Científica.

- Rojas, J. M.; Serruya, H.; Bentes, M. H. S. Chemometric Classification of Two Peach Palms (*Bactris gasipaes* H.B.K.). *Journal of the American Oil Chemists Society* 1994, 71, 127.
- Controle de Qualidade da Uña de Gato (*Uncaria tomentosa* e *Uncaria guianensis*) pela quimiometria. (CONCYTEC-Perú), 1997.
- Chemometric classification of *Oenocarpus* and *Jessenia bataua* genera according to their metal content and energy-dispersive x-ray fluorescence spectra. Artigo para ser submetido a publicação.
- Multivariate calibration of metal content using FTIR spectra of flour peel-mesocarp ashes of *Oenocarpus-Jessenia* complex fruits. Artigo para ser submetido a publicação.

6.- Participação em Eventos Científicos.

- XXXII Congresso Brasileiro de Química, Belém Pará do 26-30/10/1992. Organização.
- XIX Congresso Peruano de Química, Lima do 15-19/10/1995
- II Encontro Paraense de Ensino de Química, Belém-Pará do 15 a 18/06/1999. "Constituintes voláteis do óleo essencial de *Ocimum basilicum* L. com ocorrência no Pará (Brasil) e Puerto Almendra Rio Nanay (Perú)".
- XXIV Congresso Latino-americano de Química/XXI Congresso Peruano de Química-Lima-Perú, 15-19 de outubro de 2000.
 - Estudo comparativo da hidrogenação de óleo de ungurahui (*Jessenia bataua*) e óleo de oliva (*Europaea lineu*). Oral.
 - Classificação quimiométrica dos gêneros *Oenocarpus* e *Jessenia bataua* mediante fluorescência de raios x de energia dispersiva (EDXRF). Pannel.

7.- Língua Estrangeira

- Português. Fala, escreve
- Inglês. Fala escreve
- Francês. Fala, escreve pouco.

Campinas, 30 de Agosto de 2001.

ÍNDICE

	Página
Lista de Símbolos e Abreviaturas utilizadas	xvii
Lista de Tabelas	xix
Lista de Figuras	xxiii
 CAPÍTULO I. Introdução	
Introdução	1
Objetivos	5
 CAPÍTULO II. Técnicas analíticas e métodos quimiométricos	
Justificativa do uso das técnicas analíticas	7
Apresentação de métodos quimiométricos	10
 CAPÍTULO III. Materiais e Métodos	
Materiais e Métodos	25
Amostras	25
Procedimentos	28
Tratamento de dados	34
 CAPÍTULO IV. Resultados e discussão de classificação	
Resultados e discussões	37
Dados morfológicos	38
Dados físico-químicos	43
Dados de composição em ácidos graxos dos óleos	48

Dados dos teores dos elementos químicos	52
Dados de espectros de fluorescência de raios-X de energia dispersiva (EDXRF)	56
Dados dos espectros de difração de raios-X (XRD)	60
Dados espectroscópicos de ultravioleta/visível (UV/Vis)	64
Dados dos espectros de infravermelho com transformadas de Fourier (FTIR)	68

CAPÍTULO V.- Resultados e discussão de calibração multivariada

Resultados e discussão de calibração de modelos de calibração multivariada usando espectros de FTIR para determinar teores dos elementos obtidos por EDXRF.	73
---	----

CAPÍTULO VI.- Conclusões

Conclusões	95
Perspectivas futuras	97

Referências bibliográficas	99
-----------------------------------	----

Apêndice	105
-----------------	-----

LISTA DE SÍMBOLOS E ABREVIATURAS UTILIZADAS

ASCII	Formato de texto
GC	Cromatografia gasosa
EDXRF	Fluorescência de raios-X de energia dispersiva
EMBRAPA	Empresa Brasileira de Pesquisa Agropecuária
FTIR	Infravermelho com transformada de Fourier
HCA	Análise hierárquica de agrupamento
FTIR	Infravermelho com transformada de Fourier
keV	Quilo elétron volt
KNN	Regra dos K vizinhos mais próximos
kV	Quilo volt
L	Matriz de loadings
L^t	Matriz transposta dos loadings
MLR	Regressão linear múltipla
MSC	Correção de espalhamento multiplicativo
PC	Componente principal, fator, autovetor ou variável latente
PCA	Análise de componentes principais
PCR	Regressão de componentes principais
PLS	Mínimos quadrados parciais
PRESS	Soma dos quadrados dos erros de previsão
PR	Reconhecimento de padrões
S	Matriz dos valores singulares
SEP	Erro padrão de previsão
SEV	Erro padrão de validação
SIMCA	Modelos independentes de similaridade utilizando componentes principais
SVD	Decomposição em valores singulares
T	Matriz de escores
U	Matriz de escores normalizada
UV	Luz ultravioleta

UV/Vis	Luz ultravioleta-visível
X	Matriz de variáveis independentes
XRD	Difração de raios-X
Y	Matriz de variáveis dependentes
β	Vetor de regressão
λ	Comprimento de onda de luz
ν	Frequência da onda
ν^*	Número de onda
θ	Ângulo de incidência do raio-X

LISTA DE TABELAS

	Página
Tabela 1: Identificação, procedência, gênero e espécie das amostras coletadas no Brasil e no Perú.	37
Tabela 2: Resumo estatístico dos dados morfológicos dos gêneros <i>Oenocarpus</i> e <i>Jessenia</i>	38
Tabela 3: Loadings das variáveis nas duas primeiras componentes principais para dados morfológicos.	40
Tabela 4: Poderes discriminante e de modelagem das variáveis morfológicas	42
Tabela 5: Resultados da classificação das amostras do conjunto de teste usando dados morfológicos através de modelos de SIMCA e KNN.	42
Tabela 6: Resumo estatístico dos dados físico-químicos.	44
Tabela 7: Loadings das variáveis nas duas primeiras componentes principais para dados físico-químicos.	45
Tabela 8: Resultados da classificação das amostras do conjunto de teste usando dados físico-químicos através de SIMCA e KNN.	47
Tabela 9: Resumo estatístico dos dados de composição em ácidos graxos.	48
Tabela 10: Loadings das variáveis nas três primeiras componentes principais para dados de ácidos graxos.	49

Tabela 11: Resultados da classificação das amostras do conjunto de teste usando teores de ácidos graxos com modelos de SIMCA e KNN.	51
Tabela 12: Resumo estatístico das amostras com teores de elementos químicos.	52
Tabela 13: Loadings das variáveis nas três primeiras componentes principais para os dados de teores de elementos químicos.	53
Tabela 14: Poderes discriminante e de modelagem das variáveis de teores de elementos químicos.	55
Tabela 15: Resultados da classificação das amostras do conjunto de teste para os dados de teores de elementos químicos mediante SIMCA e KNN.	56
Tabela 16: Resultados da classificação das amostras do conjunto teste através de SIMCA e KNN usando espectros de EDXRF.	60
Tabela 17: Resultados da classificação das amostras do conjunto teste através de SIMCA e KNN usando difratogramas XRD.	64
Tabela 18: Loadings das variáveis nas três primeiras componentes principais para dados espectroscópicos de UV/Vis.	65
Tabela 19: Resultados da classificação através de SIMCA e KNN usando espectros de UV/Vis.	67
Tabela 20: Resultados da classificação através de SIMCA e KNN das amostras do conjunto teste usando espectros de IR.	72
Tabela 21: Teores (%) de elementos determinados pela técnica analítica EDXRF para o conjunto teste.	77

Tabela 22: Teores (%) de elementos do conjunto teste previstos usando os modelos PCR com 1868 variáveis independentes.	77
Tabela 23: Teores (%) de elementos do conjunto teste previstos usando regressão PLS PCR com 1868 variáveis independentes	78
Tabela 24: Teores (%) de elementos previstos usando os modelos PCR com 860 variáveis independentes.	88
Tabela 25: Teores (%) de elementos previstos usando os modelos PLS com 860 variáveis independentes.	89
Tabela 26: Teores (%) de P para o conjunto teste previstos com modelos de PLS para 1PC, 2PC e 3PC.	91
Tabela A. 1: Massas e dimensões das amostras dos gêneros <i>Oenocarpus</i> e <i>Jessenia</i> .	105
Tabela A. 2: Variáveis físico-químicas dos gêneros <i>Oenocarpus</i> e <i>Jessenia</i> .	106
Tabela A. 3: Composição em ácidos graxos dos óleos da casca+polpa extraídos em natura.	107
Tabela A.4: Conteúdo dos elementos em % encontrados nas cinzas da casca+polpa dos gêneros <i>Oenocarpus</i> e <i>Jessenia</i> através de EDXRF.	108
Tabela A. 5: Dados numéricos dos espectros de EDXRF das amostras dos gêneros <i>Oenocarpus</i> e <i>Jessenia</i> na faixa de energia 0-20 keV com respostas a intervalos de 0.0195 keV.	109

Tabela A.6: Dados numéricos dos difratogramas de XRD das amostras dos gêneros <i>Oenocarpus</i> e <i>Jessenia</i> na faixa de $2\theta = 20-70^\circ$ com respostas a intervalos de $2\theta = 0.02^\circ$.	110
Tabela A. 7: -Dados de absorvâncias na região do UV/Vis dos extratos em éter de petróleo na faixa de 240-440 nm.	111
Tabela A. 8: Dados de absorvâncias das pastilhas de KBr-cinza na faixa de 4000 a 400 cm^{-1} da região do IR médio.	114
Tabela A. 9: Amostras e teores dos elementos químicos que formam o conjunto de calibração e o conjunto teste para obtenção dos modelos de calibração multivariada.	115

LISTA DE FIGURAS

	Pagina
Figura 1: Matriz de dados para as variáveis independentes.	10
Figura 2: Matriz de dados para as variáveis dependentes.	11
Figura 3: Gráfico dos pontos pertencentes à k-ésima e à j-ésima amostras, com $p=3$ variáveis da matriz X.	11
Figura 4: Tipos de conexões para formar um dendrograma.	12
Figura 5: Sequência do método de componentes principais a) gráfico da matrix X, b) modelo de zero PC, c) modelo de 1PC, d) modelo de 2PC e e) projeção das amostras sob duas PC.	15
Figura 6: Classificação de uma amostra com classe desconhecida segundo KNN	16
Figura 7: Modelos de zero, 1 e 2 componentes principais, gerados por SIMCA.	17
Figura 8: Gráfico de SEV de validação cruzada “leave one out” para uma classe versus número de PC. Três PC é ótimo para o modelo da classe.	18
Figura 9: Relação entre as variáveis X e Y através do uso de PCR.	20
Figura 10: Relação entre as variáveis X e y no uso de PLS.	21
Figura 11: Gráfico dos resíduos de Student versus leverage de amostras para identificar outliers.	23

Figura 12: Distribuição geográfica de <i>Oenocarpus bacaba</i> e <i>Oenocarpus macrocalix</i> Burret.	26
Figura 13: Distribuição geográfica de <i>Oenocarpus mapora</i> e <i>Oenocarpus minor</i> .	26
Figura 14: Distribuição geográfica de <i>Oenocarpus circumtextus</i> Martius e <i>Jessenia bataua</i> : subespécie <i>bataua</i> e <i>oligocarpa</i> . Referencias dos locais das coletas Perú (x)) e Brasil (xx)	27
Figura 15: <i>Jessenia bataua</i> (patauí/ungurahui).	27
Figura 16: <i>Oenocarpus mapora</i> .	27
Figura 17: Gráficos dos escores das duas primeiras componentes principais.	40
Figura 18: Dendrograma de amostras com variáveis morfológicas.	41
Figura 19: Gráficos dos escores das duas primeiras componentes principais.	46
Figura 20: Dendrograma de amostras com variáveis físico-químicas.	46
Figura 21: Gráficos dos escores e dos loadings das duas primeiras componentes principais para dados de composição em ácidos graxos.	50
Figura 22: Gráficos dos escores e dos loadings das duas primeiras componentes principais para teores de elementos químicos	54
Figura 23: Gráfico dos espectros de EDXRF de amostras de cinzas.	57
Figura 24: Gráfico dos loadings das PC versus variáveis dos espectros de EDXRF.	58

Figura 25: Gráfico dos escores nas duas primeiras componentes principais..	59
Figura 26: Gráfico dos difratogramas de XRD de amostras de cinzas.	61
Figura 27: Gráfico dos loadings das três primeiras PC dos difratogramas de XRD.	62
Figura 28: Gráfico dos escores nas duas primeiras PC dos difratogramas de XRD.	62
Figura 29: Gráfico das amostras com variáveis de UV/Vis.	65
Figura 30: Gráficos dos escores e dos loadings das duas primeiras componentes principais para dados de UV/Vis.	66
Figura 31.- Gráfico dos espectros no infravermelho médio das amostras de cinzas.	69
Figura 32: Gráficos bidimensionais das absorvâncias em vários números de onda.	69
Figura 33: Gráfico dos loadings das PC versus número de onda.	70
Figura 34: Gráfico dos escores das duas primeiras PC dos espectros de FTIR.	71
Figura 35: Gráfico do erro de validação (SEV) versus o número de componentes principais (PC) para o teor de P.	74
Figura 36: Gráfico dos resíduos de Student versus leverage das amostras para o teor de P.	75
Figura 37: Gráfico do erro de validação (SEV) versus o número de componentes principais (PC) para o teor de K.	76

- Figura 38:** Gráfico dos resíduos de Student versus leverage das amostras para o teor de K 76
- Figura 39:** Gráfico dos valores experimentais versus valores previstos usando PLS de teores de P. 80
- Figura 40:** Gráfico dos valores experimentais versus valores previstos usando PLS de teores de K. 81
- Figura 41:** Gráfico dos valores experimentais versus valores previstos usando PLS de teores de Ca. 82
- Figura 42:** Gráfico dos valores experimentais versus valores previstos usando PLS de teores de Mn. 82
- Figura 43:** Gráfico dos valores experimentais versus valores previstos usando PLS de teores de Fe. 83
- Figura 44:** Gráfico dos valores experimentais versus valores previstos usando PLS de teores de Zn e as amostras 8 e 16. 83
- Figura 45:** Gráfico dos valores experimentais versus valores previstos usando PLS de teores de Zn sem as amostras 8 e 16. 84
- Figura 46:** Gráfico dos valores experimentais versus valores previstos usando PLS de teores de Cu. 84
- Figura 47:** Gráfico dos (a)loadings e (b) dos coeficientes do vetor de regressão versus número de onda para o modelo do teor de P. 86

Figura 48: Gráfico dos (a)loadings e (b) dos coeficientes do vetor de regressão versus número de onda para o modelo do teor de K.	87
Figura 49: Gráfico dos valores experimentais versus valores previstos de teores de P usando PLS com 860 variáveis e uma (01) componente principal	92
Figura 50: Gráfico dos valores experimentais versus valores previstos de teores de P usando PLS com 860 variáveis e duas (02) componentes principais.	93
Figura 51: Gráfico dos valores experimentais versus valores previstos de teores de P usando PLS com 860 variáveis e três (03) componentes principais.	94
Figura A.1: Cromatograma da mistura de padrões de ésteres metílicos.	116
Figura A.2: Espectros de EDXRF das amostras do gênero <i>Oenocarpus</i> .	117
Figura A.3: Espectros de EDXRF das amostras do gênero <i>Jessenia</i> .	117
Figura A.4: Espectro de EDXRF da amostra 8 do gênero <i>Oenocarpus</i> na faixa de energia com respostas relevantes.	118
Figura A.5: Difrátogramas de raios-X das amostras do gênero <i>Oenocarpus</i> .	119
Figura A.6: Difrátogramas de raios-X das amostras do gênero <i>Jessenia</i> .	119
Figura A.7: Espectros de FTIR médio na faixa de 400-4000 cm^{-1} das amostras do gênero <i>Oenocarpus</i> .	120
Figura A.8: Espectros de FTIR médio na faixa de 400-4000 cm^{-1} das amostras do gênero <i>Jessenia</i> .	120

CAPÍTULO I

INTRODUÇÃO

Os patauzeiros (*Jessenia bataua*) e os bacabeiros (*Oenocarpus*) são plantas cujas histórias estão ligadas à própria evolução e sobrevivência das tribos indígenas da Amazônia. A distribuição geográfica destas plantas ocorre em toda a América do Sul, principalmente na região amazônica. Esses dois gêneros denominam-se complexo *Oenocarpus-Jessenia*¹ porque compartilham algumas propriedades.

Devido a grande demanda mundial de óleos e gorduras na atualidade, torna-se necessário usar novas matérias primas e promover o melhoramento genético, agrônômico e econômico das fontes tradicionais e das plantas nativas pouco exploradas.

Os gêneros *Oenocarpus* e *Jessenia* são palmeiras de grande valor econômico na Amazônia e fontes importantes de produtos de subsistência, uma vez que fornecem alimentos, remédios, fibras, material para construção e outros produtos². Suas folhas servem para fazer telhado, produzir fibras e para fins medicinais. Seus troncos são usados para construção, seus espinhos e suas fibras fornecem suplementos para caça e combustível, seus frutos são coletados como fonte de alimentos, bebidas, proteínas e óleos¹. Atualmente são largamente utilizadas pelos camponeses e populações nativas das regiões da Amazônia brasileira e peruana.

A importância econômica está no fato de que o óleo extraído dos frutos destas plantas apresentam propriedades físico-químicas semelhante aos dos azeite de oliva, o que significa que uma produção incentivada, racionalmente, poderia substituir gradualmente a sua importação^{2,3,4}. Portanto, o cultivo destas palmeiras representariam uma nova alternativa para o desenvolvimento dessas comunidades e uma forma de que estas palmeiras entrem no mercado mundial para serem aproveitadas já que estas são utilizadas integralmente.

Martius em 1823 descreveu o gênero *Oenocarpus*. Karsten em 1857, descreveu o gênero *Jessenia* e fez menção da sua afinidade com *Oenocarpus*. Estes dois gêneros

foram mantidos como distintos entre 1848-1889. Burret em 1923 re-avaliou e examinou as relações taxonômicas e foi capaz de separar os gêneros nas bases de material estéril (de plantas). Moore em 1963 eliminou varias espécies de *Oenocarpus* e as chamou *O. bacaba*, e Wessels Bôer em 1965 incluiu *Jessenia* dentro *Oenocarpus*. Outros pesquisadores como Dugand, em 1940, e Moore, em 1973, consideraram que os dois gêneros são distintos¹.

Balick¹ tentou definir as diferenças e delinear o esquema taxonômico válido para os dois gêneros, chamando-o complexo *Oenocarpus-Jessenia*: *Oenocarpus*, constituída de oito espécies e 4 subespécies, e *Jessenia*, constituída de uma espécie, com duas subespécies.

Assim na atualidade a taxonomia do complexo *Oenocarpus-Jessenia* está relativamente estabelecida. *Jessenia* inclui só uma espécie, *Jessenia bataua* Burret (patauá), com duas sub-espécies. O gênero *Oenocarpus*, inclui 8 espécies: *Oenocarpus bacaba* Mart (bacaba), *Oenocarpus disticus* Mart, *Oenocarpus macrocalyx* Burret (bacabão), *Oenocarpus minor* Mart. (bacabi ou bacabunha), *Oenocarpus mapora*, *Oenocarpus circumtextus*, *Oenocarpus discolor* e *Oenocarpus tarampabo* e quatro subespécies.

Os botânicos não chegaram a um consenso sobre a diferenciação dos dois gêneros, pois ainda tem alguns pesquisadores como Herderson⁵ que defendem que o patauá, que é do gênero *Jessenia*, estaria melhor incluído no gênero *Oenocarpus*.

A *Jessenia bataua* (Mart) Burret é uma palmeira (Arecaceae) que está distribuída no Panamá e por mais da metade das planícies do norte da América do Sul, Colômbia, Perú, Guiana Francesa, Equador, Norte do Brasil, etc., principalmente na região amazônica²; é originária das planícies do norte da América do Sul, ou seja, na florestal tropical úmida. O nome de bataua, ditado por Martius, é uma corrupção de nome índio patauá, pela pronuncia alemã⁶.

Uma descrição botânica dos *Oenocarpus* mostra que a altura dos talos varia entre 3-20 m, o diâmetro 9-25 cm e as folhas na faixa de 2,5-8 m de comprimento, dependendo da espécie considerada⁷.

Quando se pretende melhorar as características genéticas, agronômicas e econômica de uma planta é necessário que ela esteja bem identificada. A caracterização dos gêneros *Oenocarpus* e *Jessenia*, através de algumas medidas morfológicas dos frutos, medidas físico-químicas, químicas e espectroscópicas das farinhas, cinzas e óleos dos frutos, podem ajudar neste problema de identificação e classificação.

O avanço tecnológico tem favorecido o emprego combinado de técnicas espectrométricas, cromatográficas e equipamentos com multicanais, permitindo medir muitas variáveis simultaneamente, que com o auxílio de métodos matemáticos e estatísticos multivariados promove a melhor compreensão das características das plantas. O emprego destas técnicas multivariadas para uso no tratamento dos dados químicos levou ao desenvolvimento de uma nova área na química, chamada quimiometria. A existência de computadores cada vez mais velozes e com maiores capacidades e facilidades para armazenar um número muito grande de informações, permitiram que métodos quimiométricos pudessem ser usados para tratar conjuntos extensos de dados químicos. Desta maneira os problemas que não puderam ser tratados antes, agora estão sendo resolvidos rotineiramente.

Segundo vários autores^{8,9} a quimiometria pode ser definida como a disciplina de química que utiliza métodos matemáticos e de estatística multivariada para: a) definir ou selecionar as condições ótimas de medição, e b) extrair o máximo de informações úteis dos dados químicos. Então, a quimiometria seria entendida como a parte da química que utiliza métodos matemáticos, estatísticos, recursos computacionais e a mente humana para selecionar ou definir melhores condições para a execução de medidas experimentais e permitir a obtenção de uma quantidade máxima de informações.

Um dos objetivos da quimiometria é a conversão de dados em informação e finalmente em conhecimento que possa ser utilizado na resolução de problemas. Métodos de reconhecimento de padrões são importantes em quimiometria e permitem a identificação rápida e eficiente de relações básicas, que geralmente existem em uma grande massa de dados. A quimiometria ficou indispensável em muitas áreas da química porque permite a classificação das amostras, de acordo com suas propriedades, bem como a previsão quantitativa de resultados. Outras áreas também utilizam técnicas de análise de dados multivariados: biologia, economia, psicologia, geologia, geoquímica e ecologia; onde campos como biometria, econometria e psicometria vem crescendo rapidamente¹⁰.

Desse modo muitos trabalhos já foram desenvolvidos com o objetivo de identificar e/ou classificar corretamente óleos e graxas^{11,12}, vinhos¹³, mel¹⁴ bactérias¹⁵, gêneros e espécies de plantas^{16,17,18,19,20} entre outros. Em controle de qualidade os dados de espectros de infravermelho próximo de produtos de petróleo serviram para classifica-los em tempo real com análise de componentes principais e o classificador Bayesiano²¹.

A calibração multivariada aplicada aos espectros de NIR foi usada para a previsão por ex: da composição química e o valor energético de alimentos para gado²², e na quantificação de uma droga num medicamento²³, permitindo nenhuma ou uma mínima de preparação de amostra. Os espectros de FTIR também já foram usados na calibração multivariada para o óleo de soja epoxidado²⁴ e na determinação direta de glicose no sangue²⁵. Dados de EDXRF²⁶ foram usados na determinação quantitativa de metais em plantas e solos tratados com métodos de calibração multivariada.

Considerando que os métodos quimiométricos têm sido aplicados com sucesso em diferentes problemas de classificação^{27,28,29} e também que ainda não há um consenso sobre a diferenciação dos dois gêneros do complexo *Oenocarpus-Jessenia*, estabelecemos os seguintes objetivos para este estudo:

Gerais:

- Ajudar a esclarecer dúvidas sobre a identificação dos gêneros *Oenocarpus* e *Jessenia* através de dados morfológicos, espectroscópicos e cromatográficos;
- Caracterizar amostras de frutos, casca+polpa, semente, óleos, farinhas e cinzas dos *Oenocarpus* e *Jessenia*, aplicando métodos quimiométricos aos dados experimentais;
- Usar métodos de classificação e calibração para extrair informação útil sobre os frutos destes gêneros.

Específicos:

- Obter as variáveis morfológicas dos frutos, físico-químicas das farinhas e da composição em ácidos graxos dos óleos;
- Obter os espectros de UV/Vis dos extratos obtidos das farinhas e espectros de FTIR, EDXRF e XRD das cinzas das farinhas;
- Identificar o número mínimo de variáveis que discrimine patauás de bacabas e aplicar calibração multivariada, se necessário.

No capítulo II está a justificativa das técnicas analíticas utilizadas neste trabalho bem como a descrição dos métodos quimiométricos usados para o tratamento dos diferentes tipos de dados obtidos, entre eles métodos exploratórios, de reconhecimento de padrões e métodos de calibração multivariada. O capítulo III apresenta as amostras, as condições de análises para a obtenção dos dados e a forma como estes dados foram tratados através dos métodos quimiométricos. O capítulo IV mostra os resultados e a discussão dos 8 tipos de dados: dados morfológicos, físico-químicos, composição dos óleos em ácidos graxos, teor dos elementos químicos, espectros de fluorescência de raios-X de energia dispersiva (EDXRF), espectros de difração de raios-X (XRD), espectros de ultravioleta/visível (UV/Vis) e espectros de infravermelho na região média (FTIR). O capítulo V mostra os resultados e discussão da calibração multivariada usando os espectros de FTIR das cinzas de polpa+casca dos frutos como variáveis independentes e os teores dos elementos químicos destas cinzas obtidos pela técnica analítica de EDXRF, como variáveis dependentes. Para finalizar, no capítulo VI apresentam-se as conclusões mais importantes deste trabalho.

CAPITULO II

Justificativa do uso das técnicas analíticas.

As medidas de massas e dimensões dos frutos e sementes representam parte das variáveis morfológicas que, segundo os botânicos, podem incluir também características do tronco, folhas e flores que podem ser utilizadas para classificar as plantas dentro dos diferentes gêneros.

As farinhas, por sua vez, são usadas para obter diferentes variáveis físico-químicas como teores de umidade, óleo, proteína e cinzas. Além disto pode-se medir também o índice de refração do óleo. A quantificação destas variáveis pode fornecer evidências que permitiriam discriminar cada gênero. A análise da farinha para determinação de variáveis físico-químicas pode ser realizada por diversas técnicas analíticas como espectroscopia de infravermelho próximo (NIRS)^{22,30}, fluorescência de raios-X de energia dispersiva, espectrometria de emissão induzida por plasma (ICP-AES)¹⁹. Embora, antes de empregar estas técnicas geralmente se realiza as análises pelos métodos padrões tradicionais reguladas³¹ pelas instituições como British Standards Institution (BSI), International Organization for Standardization (ISO), International Union of Pure and Applied Chemistry (IUPAC).

Nos óleos obtidos diretamente das farinhas secas e frescas, determina-se o teor de ácidos graxos presente. As separações dos ácidos graxos nos óleos podem ser realizadas empregando técnicas como cromatografia de papel, cromatografia de camada delgada (TLC), cromatografia líquida sem pressão por gravidade em coluna, cromatografia gasosa (GC) e cromatografia líquida de alta eficiência (HPLC). A quantificação e identificação destes ácidos graxos separados são realizadas mediante as técnicas de GC^{32,33,34}, HPLC^{35, 36}, cromatografia gasosa acoplada a massa (GC-MS)³⁷, ressonância magnética nuclear (NMR)³⁸, infravermelho^{39,40} e ultravioleta⁴¹. Apesar de que HPLC têm algumas vantagens na análise de alguns tipos de ácidos graxos, como trabalhar na temperatura

ambiente sem ter problemas de decomposição das amostras, a HPLC é mais trabalhosa e a análise demora demais.

A cromatografia gasosa é ainda uma técnica analítica utilizada rotineiramente para separar e analisar qualitativa e quantitativamente os ácidos graxos dos óleos. E, quando manuseada corretamente, esta técnica é rápida e confiável^{42,43,44,45}. Por esse motivo e pela disponibilidade do equipamento, conhecimento das condições de trabalho na separação e pela experiência nesta técnica, foi escolhida a GC para obter dados de composição dos ácidos graxos dos óleos.

Nos extratos, extraídos das farinhas, espera-se encontrar respostas de compostos como clorofilas, que absorvem na faixa de 640-660 nm, antocianinas, 475-550 nm, tocoferóis, 520 nm, carotenóides, 400-500 nm, e flavonóides, 240-270 nm^{46,47}. Em química de produtos naturais a separação e purificação de extratos constituintes das plantas é geralmente realizado por cromatografia de papel, TLC, cromatografia em coluna, cromatografia gás líquido (GLC) e HPLC. A identificação dos constituintes pode ser realizada também pelas técnicas de UV/Vis, infravermelho, espectrometria de massas (MS), NMR, etc. As características das respostas observadas nos extratos, extraídos das farinhas com éter de petróleo, poderiam ser atribuídas à classe de compostos chamados flavonóides. Muitos estudos de determinação de flavonóides foram realizados em plantas para tratamentos medicinais⁴⁸ e em mel, usando eletroforese capilar (EC) como um método alternativo de HPLC⁴⁹, e com fins de estudos de organização e evolução⁵⁰. Dessa forma, sem muito tratamento nas amostras e, como os flavonóides podem ocorrer em diferentes teores nos dois gêneros, empregou-se a técnica espectroscópica de ultravioleta/visível (UV/Vis), que compreende a faixa de 180 a 780 nm^{51,52}.

As cinzas obtidas a partir das farinhas contêm substâncias inorgânicas como óxidos e sais. Estes compostos, ao absorverem energia de determinadas frequências, sofrem excitações vibracionais³⁶ que podem ser registradas no infravermelho (IR)⁵³. Estes espectros podem fornecer impressões digitais, que possibilitam a discriminação de dois

gêneros. Como exemplo, pode-se citar o uso de IR médio para diferenciar variedades de chá²⁰.

Uma outra técnica analítica utilizada para a obtenção de dados é a fluorescência de raios-X com dispersão de energia (EDXRF), que analisa os elementos de forma qualitativa e quantitativa^{54,55,56}. Os teores dos elementos químicos já serviram para diferenciar entre as variedades de café *arábica* e *robusta*⁵⁷.

Nas cinzas encontram-se também compostos com estruturas cristalinas definidas que servem para obter difratogramas. Portanto foi usada a técnica analítica de difração de raios-X (XRD), que baseia-se nas interferências ópticas produzidas quando uma radiação monocromática atravessa uma fenda de espessura comparável ao comprimento de onda da radiação^{58,59}.

A utilização das cinzas nas análises de infravermelho, EDXRF e XRD foi porque estas técnicas fornecem espectros que facilitam o estudo de reconhecimento de padrões dos gêneros *Oenocarpus* e *Jessenia*, além do mais permitem ser usados para fins de calibração. Uma outra justificativa no uso das cinzas é que amostras deste tipo podem ser guardadas por muito tempo sem serem degradadas, permitindo flexibilidade na disponibilidade dos equipamentos.

Em resumo as medidas obtidas para a classificação dos frutos em gêneros veio das regiões de raios-X, ultravioleta e infravermelho médio. Das faixas de energia usadas neste estudo, a única outra faixa aproveitável para obter informações que podem ser relevantes a resolver este problema de classificação é a faixa do infravermelho próximo (NIR). Esta região tem bandas associadas principalmente a transições não harmônicas, chamadas bandas sobretons ou de combinação^{60,61}. Como os estiramentos das ligações CH, NH e OH são vibrações muito anarmônicas e os frutos são ricos em compostos com estas ligações poderíamos esperar informações bastante relevante sobre as características químicas diferentes dos dois gêneros de frutos. Infelizmente na época enquanto as medidas experimentais estavam sendo determinadas não tínhamos acesso ao um

espectrômetro de infravermelho próximo. Tais medidas poderiam ter simplificado a tarefa de classificar os frutos por métodos espectroscópicos além de fornecer informação sobre características químicas diferentes dos dois gêneros.

Apresentação dos métodos quimiométricos.

Os dados obtidos das fontes de trabalho são colocados em uma matriz X , que geralmente contém as variáveis independentes, e uma matriz Y , contendo as variáveis dependentes. Nas matrizes, cada linha representa uma amostra e cada coluna uma variável. Nosso processo consiste principalmente em obter modelos para reconhecimento de padrões usando os dados da matriz X para fins de classificação e, opcionalmente obter modelos para relacionar matematicamente X e Y para fazer modelos de calibração multivariada. Em ambos os casos os modelos têm que ser validados (interna ou externamente), para finalmente serem usados na previsão de outras amostras, que não foram usados para determinar os modelos de classificação e calibração. A Matriz X de dados para p variáveis independentes e n amostras ou objetos é mostrado na Figura 1

$$X = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \vdots \\ \mathbf{x}_n \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & \cdots & x_{np} \end{bmatrix}$$

Figura 1: Matriz de dados para as variáveis independentes.

As variáveis são indicadas com índices de 1 a p e os objetos pelos índices de 1 a n . Cada linha é representada por um vetor ($\mathbf{x}$) que contém os valores das variáveis para um dado objeto, enquanto cada coluna fornece o valor de uma variável para todos os objetos.

Normalmente a matriz X contém resultados de medidas instrumentais como absorvâncias de espectros para diferentes comprimentos de ondas ou alturas de sinal cromatográfico para diferentes tempos após injeção da amostra.

A matriz de dados Y para m variáveis dependentes e n amostras ou objetos está representada na Figura 2. Esta matriz, normalmente contém informações que o químico ou pesquisador quer prever como concentrações de substâncias, valores de propriedades, etc.

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} y_{11} & y_{12} & \cdots & \cdots & y_{1m} \\ y_{21} & y_{22} & \cdots & \cdots & y_{2m} \\ \vdots & \vdots & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ y_{n1} & y_{n2} & \cdots & \cdots & y_{nm} \end{bmatrix}$$

Figura 2: Matriz de dados para as variáveis dependentes.

Cada objeto da matriz X pode ser representado graficamente por um conjunto de variáveis, como se mostra na Figura 3, onde são apresentados os pontos correspondentes às coordenadas de dois objetos num espaço tridimensional. A maior proximidade entre os pontos no espaço, maior é a similaridade existente entre as amostras desde que as variáveis medidas sejam relevantes para resolver o problema que está sendo investigado. Para uma matriz com p colunas teremos um espaço de p dimensões que não pode ser visualizado em três dimensões.

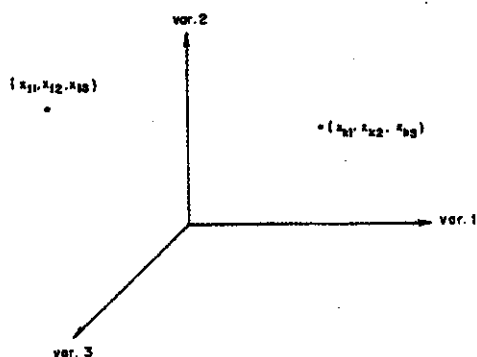


Figura 3: Gráfico dos pontos pertencentes à k -ésima e à j -ésima amostras, com $p=3$ variáveis da matriz X .

Os métodos mais usados em química para classificação de dados são os de análise hierárquica de agrupamento (HCA), análise de componentes principais (PCA), modelos independentes de similaridade utilizando componentes principais (SIMCA) e regra dos K vizinhos mais próximos (KNN). Para calibração, os métodos escolhidos foram regressão de componentes principais (PCR) e mínimos quadrados parciais (PLS).

HCA^{8,9,62,63} está baseada nas medidas das distâncias entre as amostras ou entre as variáveis. Geralmente usa-se a distância Euclidiana (d_{ab}), que para dois pontos a e b é definida pela equação 1.

$$d_{ab} = \left[\sum_{j=1}^p (x_{aj} - x_{bj})^2 \right]^{1/2} \quad (1)$$

Esta distância é calculada mediante conexões diferentes: conexão simples, completa, por meio de medianas, por meio de centróides, incremental, por meio de médias de grupo e conexão flexível. As distâncias são colocadas numa matriz e logo se procede a conexão, sendo o resultado um gráfico, chamado dendrograma, onde se pode visualizar as correlações entre amostras ou variáveis. A Figura 4, mostra alguns tipos de conexões. A conexão simples liga os pontos mais próximos de diferentes classes; a conexão completa liga os pontos mais distantes das classes e a conexão centróide começa ligando os pontos dos centróides das classes.

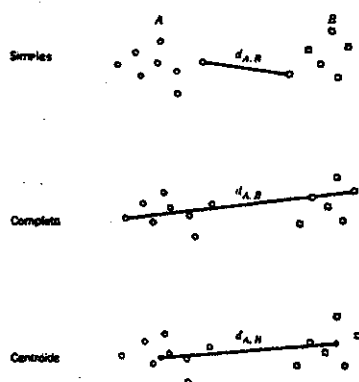


Figura 4: Tipos de conexões para formar um dendrograma⁹.

Nesta tese a conexão completa foi utilizada em todos os cálculos.

O conceito de similaridade entre duas amostras a e b (S_{ab}) é muito usado para analisar o dendrograma e é calculado pela equação 2

$$S_{ab} = 1 - (d_{ab} / d_{\max}) \quad (2)$$

onde $d_{\max}$ é a máxima distância encontrada na matriz de distâncias

Para dois pontos muito distantes $S = 0,0$ e para dois pontos que teriam a máxima similaridade seria 1,0.

PCA^{8,63,64} calcula, a partir das variáveis originais, novas variáveis, chamadas de fatores, componentes principais (PC), variáveis latentes ou autovetores. Estas novas variáveis são combinações lineares das variáveis originais. Estas variáveis tentam agrupar aquelas variáveis originais que são altamente correlacionadas. Estas PC são ortogonais e não apresentam correlação entre si. São calculadas em ordem decrescente da quantidade de variância que descrevem e permitem observar os dados em poucas dimensões. Observa-se que as relações entre as amostras não são alteradas nesta transformação. Na realidade, uma projeção do espaço multidimensional original com número de dimensões igual ao número de colunas da matriz X , é projetado num espaço com número menor de dimensões, número igual ao número de componentes principais significativos obtidos da matriz X .

Obtém-se as componentes principais mediante diferentes procedimentos matemáticos entre eles a decomposição de X em valores singulares (SVD)⁶⁵, pelo algoritmo Nipals⁶⁶ ou através da solução de um problema de autovalores-autovetores⁶².

No procedimento de decomposição de valores singulares^{65,67} a matriz X de dados é decomposta como:

$$X = USV^t \quad (3)$$

onde U é a matriz de escores normalizada para comprimento um e S é a matriz diagonal cujos elementos (valores singulares) contêm informação sobre a quantidade de variância que cada componente principal descreve; V é a matriz dos loadings ($L=V$) e US corresponde à matriz dos escores ($T=US$), ou seja

$$X = TL^t \quad (4)$$

A interpretação geométrica do método PCA tipo-R (estudo da relação entre as variáveis) resume-se nos seguintes passos para obter as componentes principais para um conjunto de dados^{8,66,68}: Considere-se uma matriz X (Figura 5a), com n amostras e p variáveis (supondo que $p=3$ dimensões). Encontra-se seu ponto central $\bar{x}$ (Figura 5b), tirando-se a média de cada variável. Este ponto é considerado o modelo de zero componentes principais. Por fim calcula-se os escores (projeções das amostras nas PC) e os loadings (peso de cada variável nas PC) para as PC até que a variância residual tende a zero. Os gráficos 5c e 5d da mesma figura mostram os modelos de uma e duas componentes principais respectivamente. A primeira componente principal está alinhada na direção da maior variância (em relação o origem localizado no ponto central) dos dados do conjunto que está sendo estudado. Por isto forma o melhor modelo unidimensional possível para os dados no sentido que esta direção tem a maior informação estatística. Se uma dimensão não é suficiente para fornecer um modelo exato, ou seja com pouco erro, adiciona-se uma segunda componente principal. Esta segunda componente é perpendicular à primeira e está alinhada com a direção da maior variância dos pontos representando os dados, descontando aquela que já foi explicada pela primeira componente principal. Finalmente no gráfico 5e é possível visualizar a projeção dos objetos da matriz desse espaço p dimensional num espaço observável das duas PC.

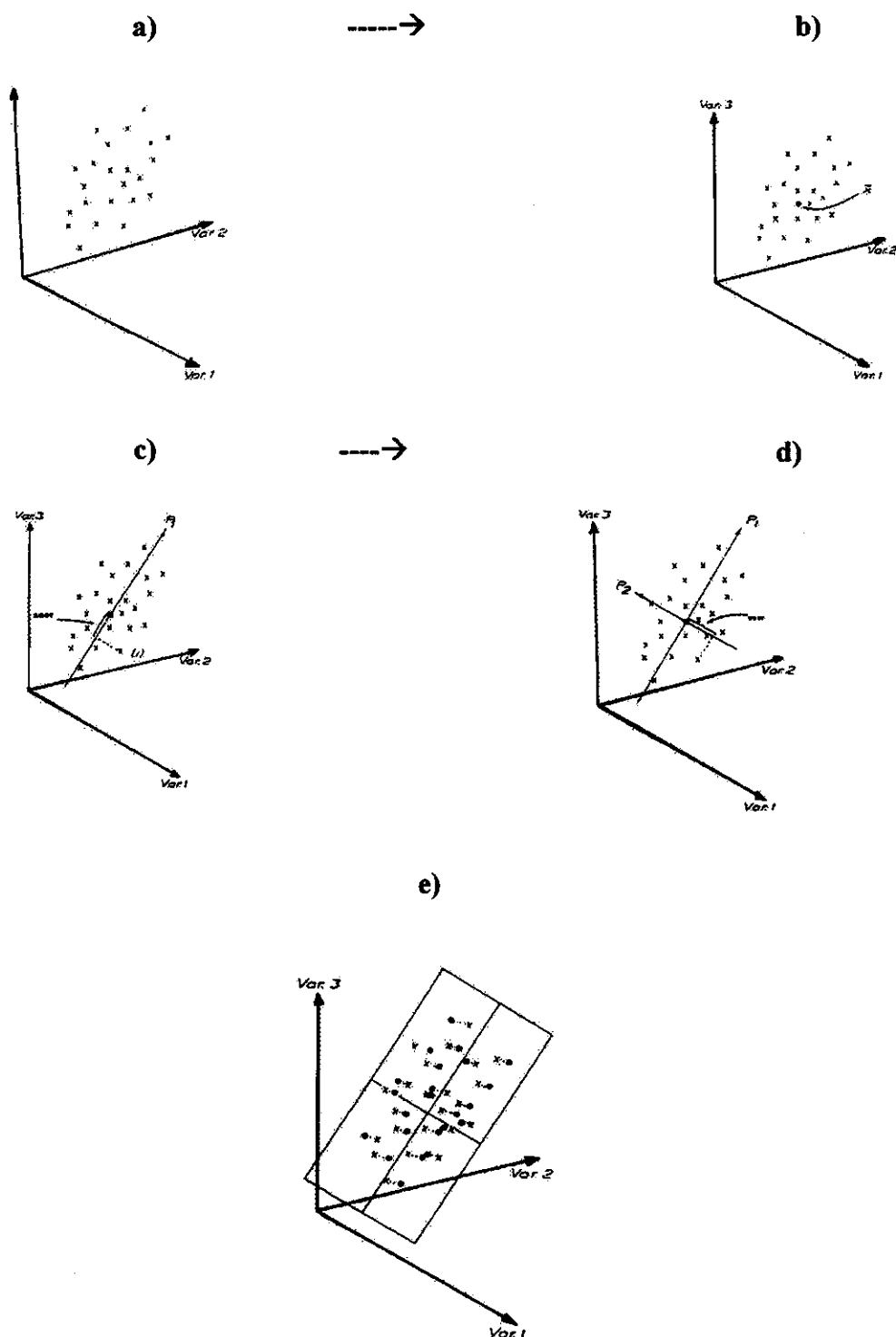


Figura 5: Sequência do método de componentes principais a) gráfico da matrix X , b) modelo de zero CP, c) modelo de 1CP, d) modelo de 2CP e e) projeção das amostras sob duas PC^{8,68}.

KNN^{8,63} (Regra dos K vizinhos mais próximos) é um método muito simples baseado nas similaridades dos objetos. Usa-se a distância euclidiana e conexão simples para medir a proximidade entre amostras. Geralmente os K vizinhos mais próximos são tomados como $K = 1, 2, \dots, 10$. A Figura 6 mostra a classificação de uma amostra de classe desconhecida (classe ou categoria é um dos grupos de objetos estudados com os algoritmos de análises de classificação). Se uma amostra teste representada por * está perto da classe representada pelo Δ do conjunto de treinamento, será colocada naquela classe porque as amostras estão mais próximas nesse espaço multidimensional, caso contrário pertencerá à classe representada por $\bullet$. Entretanto, este método não é capaz de indicar se um objeto pertence a uma classe não definida, porque sempre será colocada na classe de maior número do seus vizinhos mais próximos.

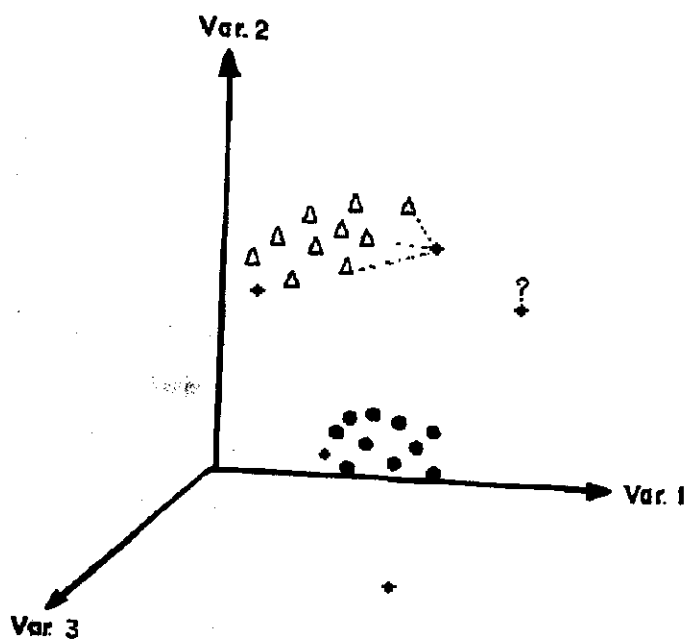


Figura 6: A classificação de uma amostra com classe desconhecida é feita com base das classes dos seus vizinhos mais próximos segundo KNN.

SIMCA^{8,9,63,69} (Modelos independentes de similaridade utilizando componentes principais) é um método de modelagem, paramétrico e probabilístico no qual usa-se as

componentes principais para obter os modelos das classes. Na Figura 7 observamos três classes definidas por um modelo de zero componentes principais (hiperesfera), um modelo de uma componente principal (hipercilindro) e um modelo de duas componentes principais (hipercaixa); neste método aplica-se PCA para cada classe independentemente.

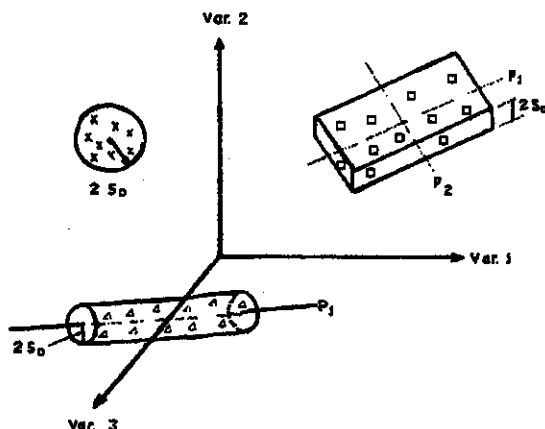


Figura 7: Modelos de zero, 1 e 2 componentes principais, gerados por SIMCA.

A determinação do número de PC em cada modelo ou classe é feita pela variância residual expressa como erro padrão de validação cruzada para PCA (SEV-PCA) versus número de PC. A Figura 8 mostra o gráfico de SEV-PCA versus número de PC, onde vemos que de acordo aos valores de SEV-PCA o número de PC recomendável para o modelo corresponde a 3, porque depois disso a adição de mais PC não diminui significativamente o SEV-PCA.

$$(SEV - PCA)_k = \sqrt{\left\{ \sum_{i=1}^n \sum_{j=1}^p (\text{resíduo}(i,j))^2 \right\} / n} \quad (5)$$

onde n é o número de amostras de validação para um modelo de k PC; $\text{resíduo}(i,j)$ é o resíduo para a i -ésima amostra e a j -ésima variável.

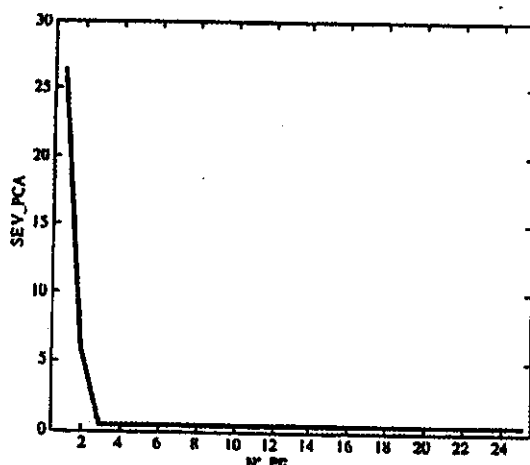


Figura 8: Gráfico de SEV-PCA de validação cruzada “leave one out” para uma classe versus número de PC. Três PC é ótimo para o modelo da classe⁶³.

A seleção de variáveis tem como objetivo minimizar custos e/ou tempo de coleta dos dados e para melhorar a eficiência do modelo⁷⁰:

O objetivo principal da classificação é colocar de forma confiável, novas amostras de classes desconhecidas nas classes existentes delimitadas pelos modelos existentes (Figura 7), bem como escolher as variáveis mais importantes que devem ficar nos modelos ou seja aquelas que têm mais informações sobre a classificação. Neste trabalho as variáveis escolhidas foram aquelas com alto poder discriminatório entre classes e alto poder de modelagem dentro de cada modelo de classe.

Considere uma matriz de dados que contém duas classes p e q

O poder de modelagem^{9,69} MPj da variável “j” numa classe q é definido por:

$$MPj = 1 - \frac{S_j(q)}{S_j} \quad (6)$$

onde $S_j(q)$ é o desvio padrão residual da variável j na classe q e, S_j é o desvio padrão da variável j .

O poder discriminatório^{9,69} DP_{pqj} da variável “j” entre as classes p e q é definido por:

$$DP_{pqj} = \sqrt{\frac{(S_{pqj})^2 + (S_{qpj})^2}{(S_{ppj})^2 + (S_{qqj})^2}} - 1 \quad (7)$$

onde $(S_{pqj})^2$ é a variância residual de pontos na classe p ajustados no modelo da classe q para a variável j , e vice versa para $(S_{qpj})^2$. O denominador contém a variância residual de pontos em p ajustados ao modelo de p para a variável j $(S_{ppj})^2$, e semelhante para a classe q . Um valor do poder de discriminação próximo a zero indica baixo poder de discriminação, e maior de 3 indica um bom poder de discriminação.

Calibração multivariada^{64,71} é o processo de construir um modelo matemático para relacionar respostas múltiplas de um instrumento a uma propriedade ou propriedades de uma amostra. Este processo inclui duas etapas. Primeiro encontra-se uma relação entre dois grupos de variáveis, a obtida do instrumento (X), independente, e aquela que se deseja calcular (Y), dependente,

$$Y = f(X) \text{ ou } Y = X\beta, \quad (8)$$

onde β é a matriz de regressão.

Nesta mesma etapa o modelo é validado interna por validação cruzada ou externa usando outras amostras conhecidas mas que não participaram da construção do modelo.

Segundo, faz-se a previsão, onde obtém-se respostas analíticas do instrumento (X_{prev}), para uma ou mais amostras que servem para prever valores de Y (Y_{prev}) daquelas amostras,

$$Y_{prev} = X_{prev}\beta. \quad (9)$$

PCR^{8,66,64} é um procedimento de dois passos na qual primeiro decompõe-se a matriz X por PCA. Ajusta-se então um modelo através de regressão linear múltipla (MLR), usando os escores dos PC como variáveis independentes. A Figura 9 mostra em forma gráfica a relação entre as variáveis X e Y quando usa-se PCR. No lado esquerdo nota-se que a matriz X foi decomposta em poucas novas variáveis que são as PC obtendo-se a matriz de escores T independente das concentrações dos diferentes analitos, isto é feito na primeira etapa. Entretanto, no lado direito vê-se que estas PC se relacionam com as variáveis dependentes uma por vez, sendo submetidas à análise de regressão linear múltipla para obter a resposta de cada variável dependente y .

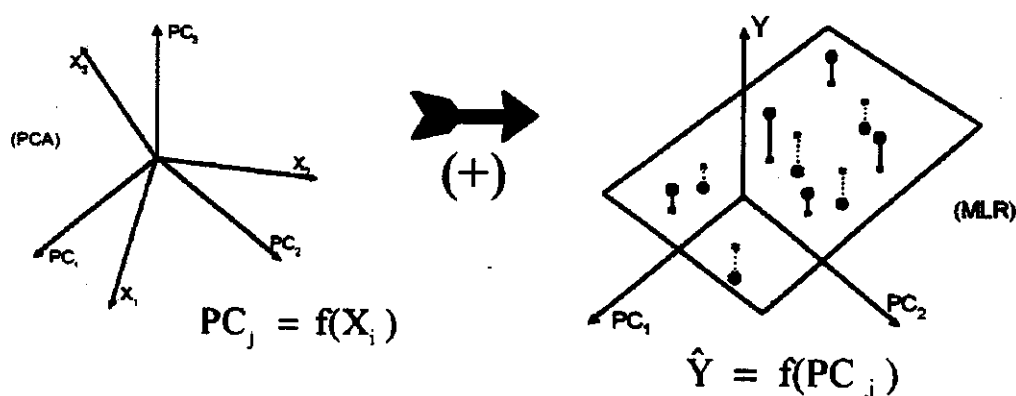


Figura 9: Relação entre as variáveis X e Y através do uso de PCR⁷².

Finalmente, PLS^{8,64,66,71,73} estima as PC na matriz X tendo em consideração a covariância das medidas com as concentrações além da variância gerada em X . O procedimento para desenvolver o modelo de calibração com a aproximação chamada PLS1⁶³ foi usada. Nesta aproximação desenvolve-se modelos separados para cada variável dependente. Na Figura 10, no gráfico do lado esquerdo a matriz X é decomposta para obter a matriz dos escores T^* trocando informação com a variável dependente y para obter os valores mais exatos de y . Nesta primeira etapa calcula-se o vetor de regressão B do modelo. A segunda etapa o gráfico do lado direito mostra o modelo obtido onde pode-se determinar a concentração de outras amostras através da matriz de escores U_{prev} obtida da matriz X_{prev} . $y_{prev} = T^*_{prev} B$. (10)

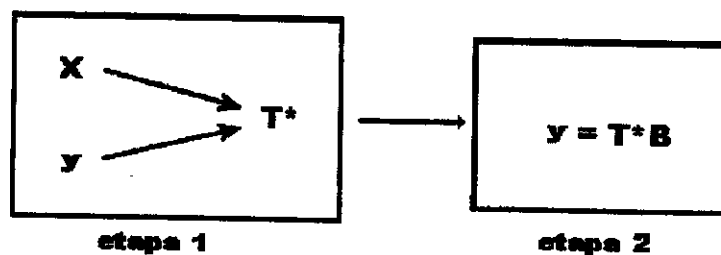


Figura 10: Relação entre as variáveis X e y no uso de PLS⁶³

A escolha do número de PC no modelo faz-se analisando a soma dos quadrados dos erros de previsão (PRESS), erro padrão de validação (SEV) e o erro padrão de previsão (SEP). Estes termos ficam definidos através as seguintes equações^{74,75}:

$$\text{PRESS} = \sum_{i=1}^{i=n} (y_i - \hat{y}_i)^2 \quad (11)$$

onde y_i é a concentração da amostra i já conhecida e $\hat{y}_i$ é a concentração da amostra i prevista pelo modelo,

$$\text{SEV} = \left(\sum_{i=1}^{i=n} (y_i - \hat{y}_i)^2 / n \right)^{1/2} \quad (12)$$

para qual n representa o número de amostras no conjunto de validação,

$$\text{SEP} = \left(\sum_{i=1}^{i=s} (y_i - \hat{y}_i)^2 / s \right)^{1/2} \quad (13)$$

para qual s representa o número de amostras do conjunto teste.

O número mais adequado de PCs nos modelos, será aquele que correspondente ao menor valor de PRESS ou valor mínimo de SEV.

Na construção dos modelos de calibração multivariada é também importante a identificação e remoção de outliers. Isto consegue-se mediante análise do leverage e dos resíduos de Student⁶⁷.

Entende-se geometricamente que leverage (h) é a distância de uma amostra ao centróide do conjunto de dados, e mede a influência de uma amostra no modelo de regressão. Portanto aquelas amostras que têm altos leverages influenciam bastante no modelo. Para uma amostra i

$$h_{ii} = 1/n + (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{x}_i - \bar{\mathbf{x}}) \quad (14)$$

$\mathbf{x}_i$ é o espectro da i -ésima amostra,

$\bar{\mathbf{x}}$ é o espectro médio e

$\mathbf{X}'\mathbf{X}$ é a matriz de variância-covariância.

O leverage crítico (h_{crit}) está relacionado com as componentes principais através da seguinte equação.

$$h_{crit} = 3k/n \quad (15)$$

onde k é o número de componentes principais e n é o número de amostras.

Se para uma amostra i , $h_{ii} > h_{crit}$, então, é bom analisar a amostra porque pode ser anômala.

Os resíduos das concentrações calculados por algum tipo de validação (validação cruzada ou interna e/ou externa) podem ser analisados através do resíduo de Student (RSt), assim, pode-se observar a influencia de cada amostra do conjunto de treinamento. Assim para uma amostra i , o resíduo da concentração corrigido pela leverage ($Resc_i$) é:

$$Resc_i = \sqrt{\frac{(y_i - \hat{y}_i)^2}{(n-1)(1-h_i)}} \quad (16)$$

$$RSt = \frac{(y_i - \hat{y}_i)}{Resc_i \sqrt{1-h_i}} \quad (17)$$

A identificação dos outliers geralmente se faz mediante a análise conjunta dos leverages e dos resíduos de Student. A Figura 11 mostra o gráfico dos resíduos de Student versus os leverages das amostras. Neste gráfico as amostras que têm valores de resíduos de Student e valores de leverages normais ficam dentro do retângulo verde com valores de aproximadamente -2 a 2 , e entre 0 a $0,42$. No eixo de resíduos vê-se uma amostra com resíduo alto ou seja está fora do retângulo. No eixo dos leverages as amostras que têm alto valor ficam fora da linha vertical ($0,42$). Notamos que existe uma amostra com alto leverage ($0,68$). Se uma amostra cai no retângulo do ângulo superior direito do gráfico dizemos que é um outlier. As amostras que ficam fora do retângulo de valores normais de resíduos Student e leverages, devem ser analisadas cuidadosamente.

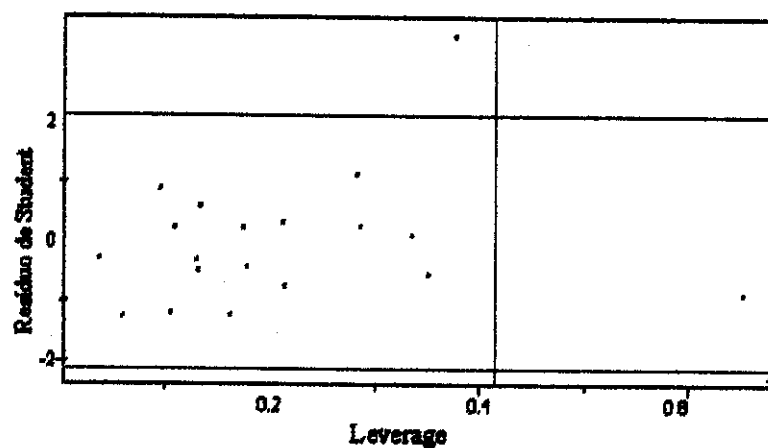


Figura 11: Gráfico dos resíduos de Student versus leverage de amostras para identificar outliers.

CAPITULO III

MATERIAIS E METODOS

Neste capítulo são abordados os aspectos relacionados à procedência das amostras e como estas foram coletadas e preparadas. Também são vistas as técnicas analíticas e as condições das análises para a obtenção dos dados. Finalmente, indica-se como esses dados obtidos foram ordenados para serem tratados mediante os métodos quimiométricos usados.

3.1. MATERIAIS

3.1.1. Amostras

As matérias primas utilizadas para o desenvolvimento deste trabalho foram selecionadas de acordo com a sua disponibilidade: *Jessenia bataua* Burret e 3 espécies do gênero *Oenocarpus* (*O. bacaba* Mart, *O. minor* e *O. mapora*). No Brasil as coletas foram viabilizadas através da Empresa Brasileira de Produtos Agropecuários (EMBRAPA), nos municípios de Abaetetuba e Igarapé Mirim (Estado do Pará). Foram dez (10) amostras pertencentes ao gênero *Oenocarpus*: oito (8) *Oenocarpus minor*, um (1) *Oenocarpus mapora* e um (1) *Oenocarpus bacaba*. As coletas realizadas em Iquitos (Perú) e interiores (Porto Almendra, Carretera Iquitos Nauta km 20) com o auxílio do Herbarium Amazonense AMAZ da Universidade Nacional da Amazônia Peruana (UNAP) resultaram em 15 plantas: dez (10) *Jessenia bataua* Burret, duas (2) *Oenocarpus minor* e três (3) *Oenocarpus mapora*, obtendo-se um total de vinte e cinco (25) amostras. As coletas dos frutos foram realizadas ao longo dos anos de 1993 a 1997.

A distribuição geográfica do complexo *Oenocarpus-jessenia* segundo Balick¹ é mostrada nas Figuras 12 a 14, sendo que na Figura 14 se referenciam os locais centrais onde as coletas foram realizadas no Perú (x) e no Brasil (xx). Os exemplares das plantas estudadas são apresentadas nas Figuras 15 e 16.

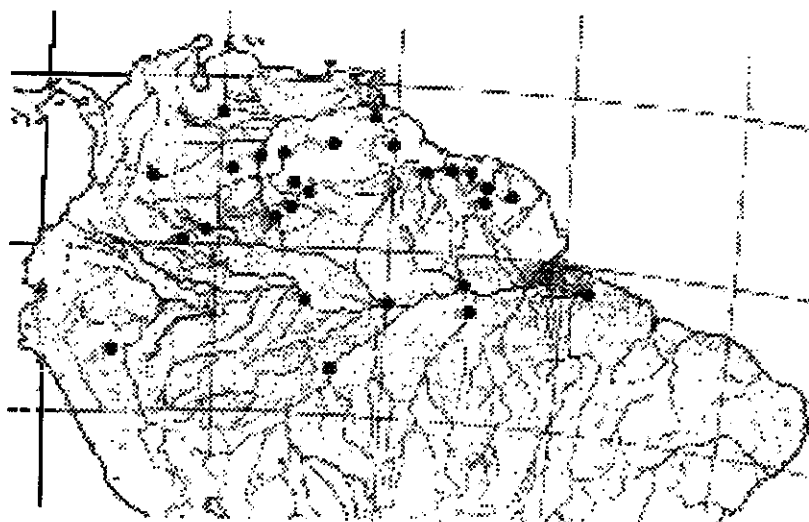


Figura 12: Distribuição geográfica de *Oenocarpus bacaba* e *Oenocarpus macrocalix* Burret.

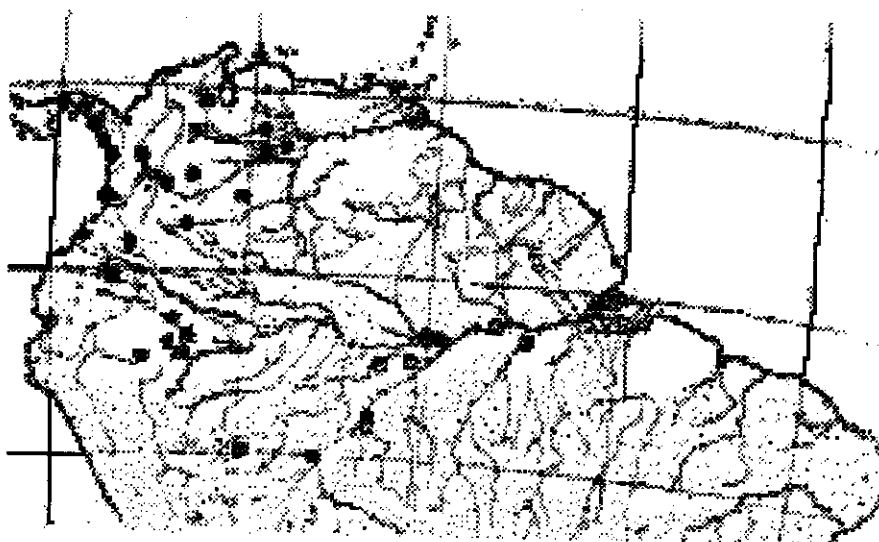


Figura 13: Distribuição geográfica de *Oenocarpus mapora* e *Oenocarpus minor*¹.

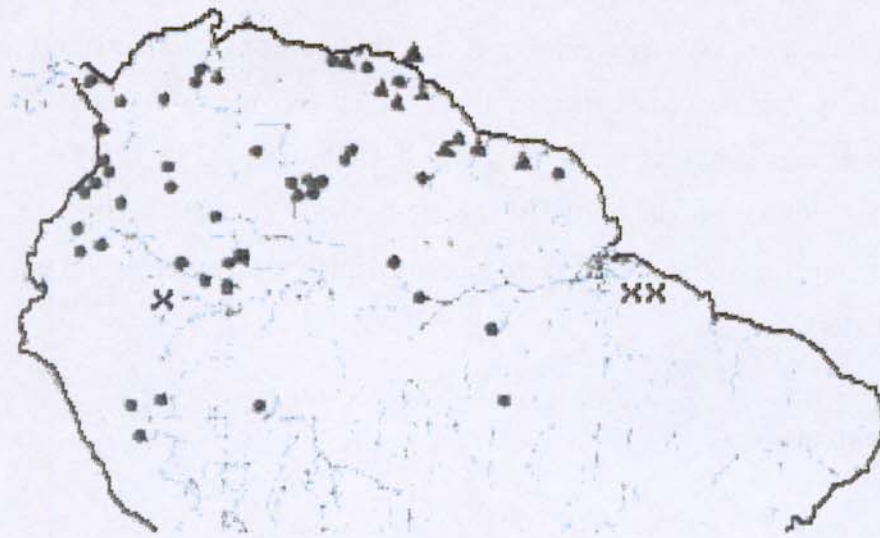


Figura 14: Distribuição geográfica de *Oenocarpus circumtextus* Martius e *Jessenia bataua*: sub-espécie *bataua* e *oligocarpa*. Referências dos locais das coletas Perú (x) e Brasil (xx).



Figura 15: *Jessenia bataua* (pataua)



Figura 16: *Oenocarpus mapora*.

3.1.2. Reagentes

Entre os principais reagentes que foram utilizados mencionamos os seguintes: brometo de potássio, celite, diclorometano, metanol, metóxido de sódio, isooctano, padrões de ésteres metílicos: C:10, C:12, C:14, C:16, C:18, C:18:1, C:18:2 e C:18:3; éter de petróleo, trifluoreto de boro, BF_3 , éter dietílico, hexano, sulfato de cobre, zinco metálico em pó e sulfato de cálcio anidro. Todos os reagentes foram do grau p.a (produtos Merk)

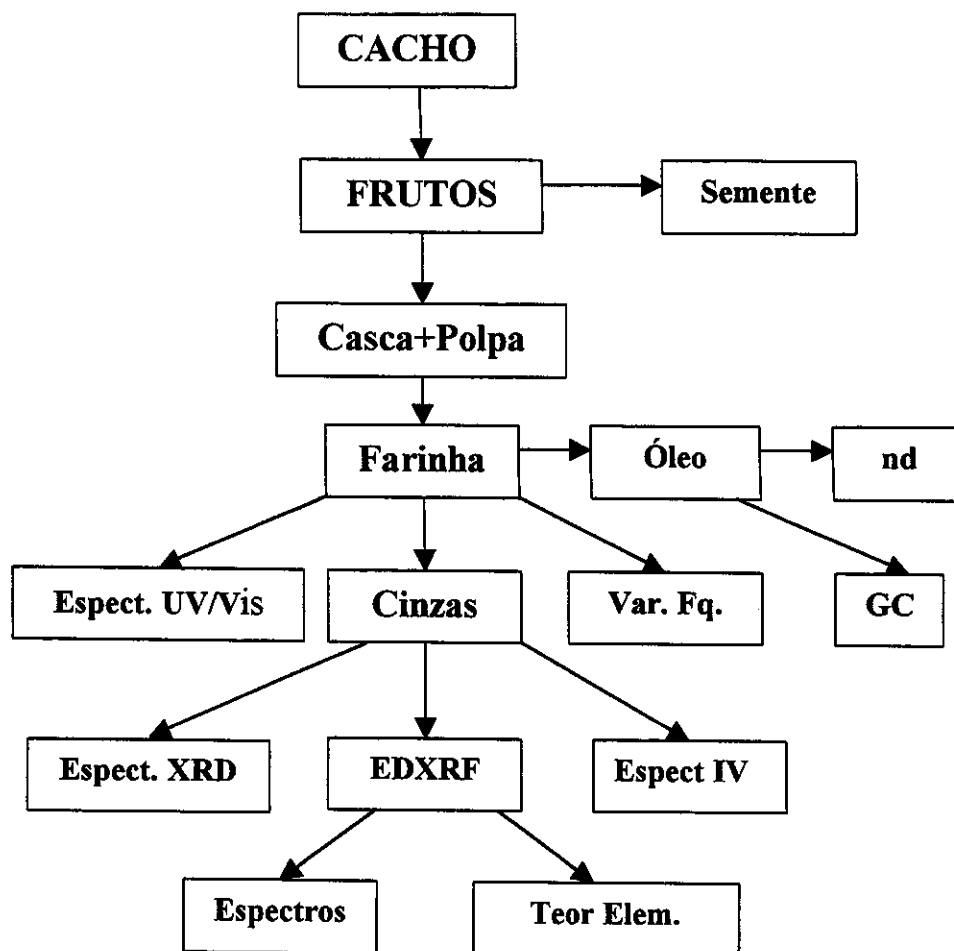
3.1.3. Equipamentos

Os aparelhos e equipamentos usados neste trabalho foram os seguintes: balança Sartorius 2000, paquímetro, estufa, mufla, cromatógrafo CG-300 Shimadzu, espectrômetro UV 160A-Shimadzu, espectrômetro BOMEM MB-Series B-100, espectrômetro Spectrace-mod. 5100 e espectrômetro XRD6000-Shimadzu.

3.2. PROCEDIMENTOS**3.2.1. Coleta das amostras**

De cada uma das 25 árvores foram coletados 200 frutos (para retirada da amostra respectiva) para realizar as análises laboratoriais.

O fluxograma 1 mostra a sequência para a obtenção dos diferentes tipos de dados.



Fluxograma 1. Procedimento seguido na obtenção de dados a partir das amostras coletadas.

3.2.2. Medidas morfológicas

As massas dos frutos, polpas e sementes, assim como o comprimento e o diâmetro dos frutos e das sementes foram obtidos imediatamente após a coleta como uma média de dez (10) frutos escolhidos ao acaso dentre os duzentos (200) frutos disponíveis. Em seguida obteve-se as farinhas da casca+polpa. Para conseguir maior quantidade de farinha utilizou-se os 190 frutos restantes. As farinhas foram secadas ao sol e posteriormente na estufa.

3.2.3. Análises físico-químicas

As análises físico-químicas foram realizadas em duplicata. A umidade⁷⁶ foi determinada nas cascas+polpas frescas; óleo, proteína⁷⁶ e cinzas (Ba-5a—49)⁷⁷ foram determinados nas farinhas da casca+polpa secas, e índice de Refração⁷⁶ no óleo.

A umidade foi determinada na casca+polpa do fruto fresco empregando uma estufa a 103 ± 2 C, segundo o método do Instituto Adolfo Lutz⁷⁶

A extração de óleos se fez misturando a farinha seca com celite em funil de placa porosa, empregando vácuo, e como solvente o diclorometano segundo o método de Schwartz⁷⁸ e Bentes⁷⁹.

O teor de proteína foi realizado pelo método de Kjeldhal, determinando nitrogênio total e utilizando o fator de 6,25 que resulta da consideração de que o conteúdo em percentagens das diversas proteínas nas amostras é de 16%; resultando $100/16 = 6,25$ (fator empírico)⁷⁶.

O teor de cinzas foram feitas nas farinhas secas sem desengordurar, as amostras foram submetidas a incineração em cadinhos de porcelana a 600 ± 12 C segundo o método Ba-5^a-49 da ^oC.S⁷⁷.

O índice de refração foi determinado nos óleos empregando Refratômetro Abbé segundo método do Instituto Adolfo Lutz⁷⁶.

3.2.4. Composição dos ácidos graxos dos óleos

Para determinar a composição dos ácidos graxos nos óleos foram preparados seus ésteres metílicos^{80,81} que foram analisados por cromatografia gasosa, utilizando um cromatógrafo GC-300 Shimadzu. As condições utilizadas foram: Coluna capilar GC-Bore 530 (Carbowax 20 MD) e programação da coluna de temperatura inicial de 110°C, mantida por 1 minuto, seguida por aquecimento em rampa de 10°C/minuto até 170°C, mantida por 2 minutos, seguida por aquecimento em rampa de 10°C/minuto até 190°C, mantida por 6 minutos. O injetor e o detector foram mantidos a 200°C e 220°C, respectivamente. O detector foi de ionização em chama (FID). Os ácidos graxos de cada amostra foram identificados através da comparação do tempo de retenção, com o cromatograma obtido da solução padrão dos ésteres metílicos e a concentração destes ácidos pelo método de normalização de área⁴⁵.

A preparação dos ésteres metílicos^{80,81} se realiza pesando entre 50 a 100 mg de óleo e logo proceder à transesterificação adicionando 3 mL de MeONa/MeOH (3% p/v), aquecer a ebulição por 3 minutos em banho-maria, esfriar e adicionar 3 mL de solução BF₃/MeOH a 10% e novamente aquecer por 3 minutos, esfriar e transferir para um balão de 50 mL, adicionar 2 mL de hexano e 1 mL de éter dietílico e agitar por um minuto para extrair os ésteres, lavar com água destilada até que a solução fique límpida, separar os ésteres contidos na fase orgânica, filtrar e passar para um frasco escuro, e logo completar o volume até aproximadamente 10 mL. Borbulhar nitrogênio e colocar numa geladeira para logo ser analisado em cromatógrafo de gás.

3.2.5. Medidas espectroscópicas na região do ultravioleta visível.

As medidas na região do UV/Vis foram obtidas nos extratos de éter de petróleo das farinhas, preparadas na razão de 10 mL de éter de petróleo por 1 g de farinha seca da casca+polpa submetida a moagem e agitação constante num banho a temperatura ambiente (30 C) durante 30 minutos, filtrada e colocada na cela respectiva de trabalho⁴⁶. As análises foram realizadas submetendo os extratos ao UV/Vis a uma varredura inicial entre 200-1100 nm encontrando-se resposta na faixa de 240 a 440 nm, obtendo-se o respectivo espectro para cada amostra. Nesta etapa foi usado um espectrômetro UV 160A-Shimadzu.

3.2.6. Medidas espectroscópicas na região do infravermelho médio

Os espectros no infravermelho médio foram obtidos das cinzas, empregando pastilhas de KBr- cinza preparadas na relação de 100 mg de KBr e 1 mg de cinza. A região analisada foi de 4000 a 400 cm^{-1} e resolução de 4 cm^{-1} com 20 varreduras/min (em cada varredura se obtém um espectro quando empregado FTIR). Os dados analíticos foram obtidos em transmitância versus número de onda (cm^{-1}) e convertidos à absorbância versus número de onda (cm^{-1}). O espectro de cada amostra foi a média de 16 varreduras. Obteve-se o espectro para cada uma das 24 amostras para formar uma matriz de dados (24 amostras x 1868). O equipamento usado na obtenção dos dados, foi um espectrômetro com transformada de Fourier BOMEM MB- Series B-100.

3.2.7. Medidas de espectros de fluorescência de raios-X de energia dispersiva

As medidas foram realizadas nas cinzas com o Spectrace-mod. 510. As cinzas foram homogeneizadas e colocadas no frasco especial para serem irradiadas pelos raios característicos de energia. As medidas foram efetuadas nas seguintes condições:

- Voltagem do tubo: 15 kV
- Corrente do tubo: 0.02 mA
- Filtro usado : Celulose
- Tempo de vida: 100 s.
- Energia Máxima: 20 keV
- Atmosfera : Vácuo
- Número de canais: 1024

3.2.8. Medidas de espectros de difração de raios-X.

Os difratogramas de raios X foram obtidas das cinzas. As amostras de cinzas obtidas foram homogeneizadas e colocadas na forma de camada sobre um porta amostra, para serem submetidos aos raios-X nas seguintes condições:

- Anodo do tubo : Cu
- Voltagem : 40 kV
- Corrente : 30 mA
- Divergência de abertura: 1.00000 grau
- Abertura de espalhamento: 1.00000 grau
- Abertura de recepção: 0.30000 mm
- Varredura 2θ : na faixa de 20-70°
- Modo de varredura : continuo
- Velocidade de varredura: 2.00000 graus/min
- Tamanho de passo : 0.0200 graus
- Tempo por passo : 0.6 s
- Comprimento de onda $\alpha_1 = 1.54060$ e $\alpha_2 = 1.54439$ Å
- Dados totais obtidos: 2501
- Equipo usado: Shimadzu XRD6000.

3.3. Tratamento dos dados.

Nesta etapa para o tratamento dos dados foram usados os programas para os métodos quimiométricos de análise hierárquica de agrupamento, análise de componentes principais^{63,66}, regra dos K vizinhos mais próximos, modelos independentes de similaridade utilizando componentes principais, regressão de componentes principais e mínimos quadrados parciais; disponíveis nos pacotes de programas de Pirouette 2.02⁸² e Unscrambler⁷².

Os dados morfológicos, físico-químicos, de composição em ácidos graxos, teor de elementos químicos e de ultravioleta visível foram ordenados usando a planilha eletrônica Excel apresentados na forma de matriz e adequados aos formatos que possam interagir com programas acima mencionados.

Os dados numéricos dos espectros de EDXRF, XRD e FTIR foram obtidos em ASCII. Estes dados foram tratados através de Word 2000 e colocados em formatos que possam interagir com os programas já mencionados, e em forma de matriz. As figuras, quando necessárias, foram processadas através de Power Point, Origin 4.1 e Paint.

Dependendo das características individuais das diferentes medidas, as matrizes obtidas foram submetidas a alguns tipos de pré-tratamentos como por exemplo: transformações, que são aplicadas especificamente às amostras, como o alisamento (média móvel e Savitzky e Golay), ponderação, primeira derivada, segunda derivada, log10, correção da linha base correção de espalhamento multiplicativo (MSC), normalização; pré-processamento, aplicado especificamente às variáveis, como centrar na média, escalamento pela variância, auto-escalamento e escalamento pela variação⁶³.

Nos capítulos seguintes serão mostrados as diferentes transformações e pré-processamentos a que foram submetida cada matriz de dados.

Aos dados obtidos foram aplicados métodos de Exploração, Classificação e Calibração. O roteiro seguido no tratamento dos dados foi:

- 1) Verificação da qualidade dos dados e visualização inicial,
- 2) Pré-processamento das amostras e variáveis,
- 3) Exploração dos dados,
- 4) Modelos classificatórios ou de calibração,
- 5) Exame dos resultados e validação dos modelos,
- 6) Uso de modelos para previsões.

CAPÍTULO IV

RESULTADOS E DISCUSSÕES

Na Tabela 1 apresenta-se a identificação dos frutos coletados em termos de procedência, gênero e espécie. Das 25 amostras cinco foram escolhidas aleatoriamente para o conjunto de teste (2, 3, 9, 18 e 22). As demais foram utilizadas para o conjunto de treinamento.

Tabela 1: Identificação, procedência, gênero e espécie das amostras coletadas no Brasil e no Perú.

Amostra	Procedência	Gênero	Data
Conjunto de Treinamento			
1(O.min) ^a	Abaetetuba (Br)	<i>Oenocarpus</i>	set-93
4(O.min)	Abaetetuba (Br)	"	"
5(O.bac) ^b	Abaetetuba (Br)	"	"
6(O.min)	Abaetetuba (Br)	"	"
7(O.min)	Igarapé Mirim (Br)	"	"
8(O.min)	Igarapé Mirim (Br)	"	"
10(O.min)	Igarapé Mirim (Br)	"	set-93
11(O.map)	Carr Quistococha km 13 (Perú)	"	jun-94
12(O.min)	Carr Quistococha km 13 (Perú)	"	"
13(O.min) ^c	Carr Iq-N km 7 F. Pizarro (Perú)	"	"
14(O.map)	Suni Mirano R-Napo (Perú)	"	abril-96
15(O.map)	AlpH Carr Iq-N km 20 (Perú)	"	maio-97
16(J.B) ^d	AlpH Carr Iq-N km 20 (Perú)	<i>Jessenia</i>	jun-94
17(J.B)	AlpH Carr Iq-N km 20 (Perú)	"	jun-94
19(J.B)	AlpH Carr Iq-N km 20 (Perú)	"	abril-96
20(J.B)	AlpH Carr Iq-N km 20 (Perú)	"	abril-96
21(J.B)	AlpH Carr Iq-N km 20 (Perú)	"	abril-96
23(J.B)	AlpH Carr Iq-N km 20 (Perú)	"	maio-97
24(J.B)	AlpH Carr Iq-N km 20 (Perú)	"	maio-97
25(J.B)	AlpH Carr Iq-N km 20 (Perú)	"	maio-97
Conjunto de Teste			
2(O.min)	Abaetetuba (Br)	<i>Oenocarpus</i>	set-93
3(O.map)	Abaetetuba (Br)	"	"
9(O.min)	Igarapé Mirim (Br)	"	set-93
18(J.B)	Carr Iq-N km 7 F. Pizarro (Perú)	<i>Jessenia</i>	abril-96
22(J.B)	AlpH Carr Iq-N km 20 (Perú)	"	maio-97

^a *Oenocarpus minor*;

^b *Oenocarpus bacaba*;

^c *Oenocarpus mapora*;

^d *Jessenia bataua*

Br, Brasil.

4.1.Dados morfológicos.

Os dados estão compostos de medidas nas 25 amostras acima mencionadas e seis variáveis que contêm as massas e as dimensões dos frutos das plantas em estudo. Os dados originais são apresentados no apêndice, Tabela A.1.

Um resumo estatístico destes dados está listado na Tabela 2. Observa-se que as médias dos valores das variáveis de *Oenocarpus* são bem diferentes das *Jessenia*. Esta diferença entre as médias é grande em relação aos valores dos desvios padrões. Também não tem sobreposição das faixas dos valores numéricos mínimos e máximos.

Tabela 2: Resumo estatístico dos dados morfológicos dos gêneros *Oenocarpus* e *Jessenia*

Variável	Total	N ⁿ	Média	Mínimo	Máximo	Desvio Padrão
<i>Oenocarpus</i>						
PF ^a		15	5,34	3,58	7,74	1,29
CF ^b		15	22,88	18,52	26,00	2,10
DF ^c		15	19,53	17,63	21,67	1,32
PS ^d		15	3,21	2,22	4,62	0,75
CS ^e		15	19,52	15,74	23,70	2,30
DS ^f		15	16,07	14,69	17,60	1,03
<i>Jessenia</i>						
PF		10	12,36	11,02	13,68	0,85
CF		10	32,30	29,75	35,58	1,78
DF		10	24,40	23,36	25,40	0,76
PS		10	8,08	7,24	10,01	0,89
CS		10	29,06	26,44	33,06	2,20
DS		10	20,07	17,80	22,50	1,34

ⁿ número de amostras.

^a peso do fruto g;

^b comprimento do fruto mm;

^c diâmetro do fruto mm;

^d peso da semente g;

^e comprimento da semente mm;

^f diâmetro da semente mm.

Análise exploratória dos dados.

Os dados morfológicos foram preprocessados, onde cada variável foi autoescalada subtraindo suas médias e dividindo pelo desvio padrão, para obter variáveis transformadas com média zero e variância um. A finalidade deste tratamento é de evitar ambigüidades que poderiam ocorrer na análise dos dados devido à diferença de unidades de medidas, aqui massas em g e dimensões em mm. Também com esta transformação todas as variáveis são tratadas com o mesmo grau de importância.

PCA foi aplicada a este conjunto de dados. Na Tabela 3 estão listados os loadings (pesos) das variáveis. Podemos ver que as duas primeiras componentes principais explicam 98,7% da variância contida nos dados. A primeira componente principal (PC1) descreve 94,6% da variância e a segunda (PC2) somente 4,1%. Também verifica-se que PC1 é essencialmente uma média de todas as variáveis.

A Figura 17 (PC1xPC2) mostra o gráfico dos escores das duas primeiras componentes principais. Visualiza-se que somente PC1 pode separar os dados em dois grupos, ficando os *Oenocarpus* para a esquerda e as *Jessenia* para a direita. Dentro dos *Oenocarpus*, vê-se um sub-grupo integrado pelas espécies *mapora* e *bacaba*, e outro formado, inteiramente, pela espécie *minor*. Evidentemente os *Oenocarpus* estão mais espalhados no eixo PC1 porque existem dentro deste gênero três espécies. Os valores dos escores maiores em PC1 correspondem às espécies *mapora* e *bacaba*. Em outras palavras, PC1 está correlacionada com o tamanho, onde as amostras de *Jessenia* são bem maiores do que *Oenocarpus*.

HCA com o método de conexão completa resultou em um dendrograma da Figura 18, em que se formam 2 grupos para níveis de similaridades entre 0 e 0,54. Estes grupos correspondem aos dois gêneros vistos no gráfico de componentes principais.

Tabela 3: Loadings das variáveis nas duas primeiras componentes principais para dados morfológicos.

	PC1	PC2
PF	0,418	-0,005
CF	0,408	-0,441
DF	0,409	0,297
PS	0,415	0,091
CS	0,402	-0,561
DS	0,399	0,628
*	94,6	4,1

* percentagem de variância explicada em cada componente.

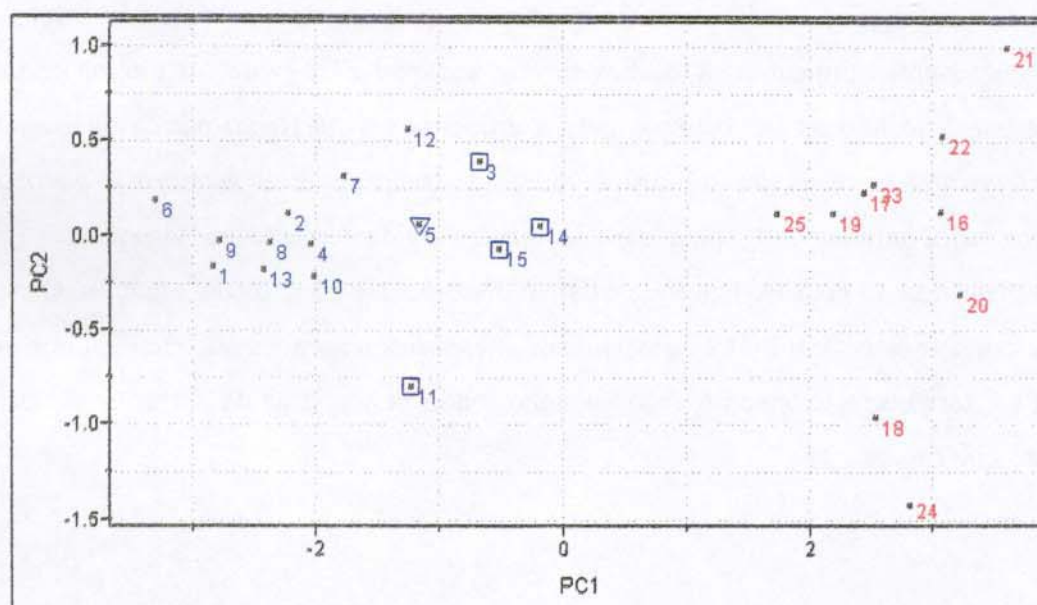


Figura 17: Gráficos dos escores das duas primeiras componentes principais para os dados morfológicos PC1xPC2: *Oenocarpus* (azul): • *O. minor*, ▽ *O. bacaba*, □ *O. mapora*; *Jessenia* (vermelho).

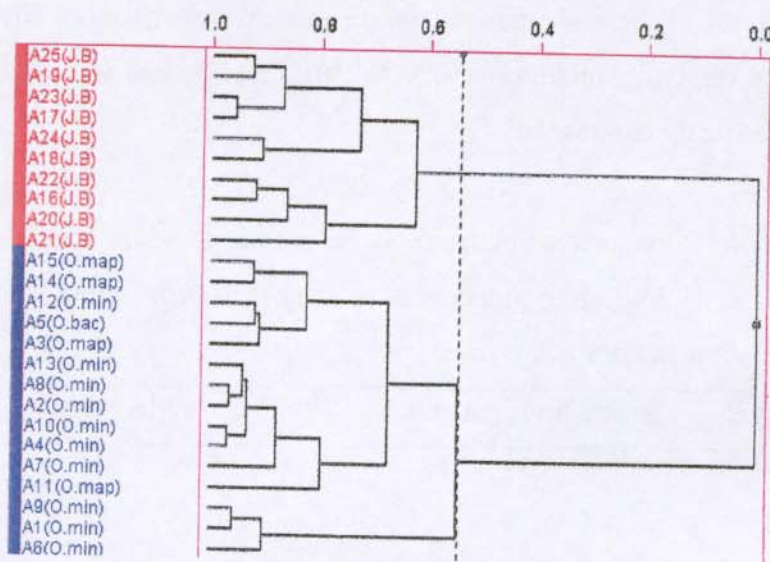


Figura 18: Dendrograma de amostras das variáveis morfológicas: *Oenocarpus* (azul) e *Jessenia* (vermelho).

Portanto, espera-se que as técnicas de classificação como SIMCA e KNN, possam discriminar os dois gêneros corretamente.

Classificação dos dados.

Aplicação do SIMCA gerou os modelos individuais para os gêneros *Oenocarpus* e *Jessenia*. Para a determinação do número ótimo de PCs para a modelagem de cada gênero foi usada a técnica de Validação Cruzada “Full leave one out” bem como critérios de variância contida nos PC, análises dos resíduos versus valor da variável e erro padrão de validação cruzada para PCA (SEV-PCA).

A modelagem foi feita considerando 20 amostras do conjunto de treinamento e a previsão com as cinco amostras do conjunto teste. O modelo *Oenocarpus* para variáveis morfológicas ficou definido por uma hipercaixa com dimensões de 86,2% da variância ao longo da PC1 e 10,7% ao longo da PC2. O modelo *Jessenia* foi descrito por uma

hipercaixa de dimensões de 50,2% da variância ao longo da PC1 e 37,9% da variância ao longo da PC2. A Tabela 4 mostra os resultados dos poderes discriminante e de modelagem das variáveis, onde vê-se que as variáveis PS, CS, DF e CF tem os maiores valores de poder de discriminação.

Tabela 4: Poderes discriminante e de modelagem das variáveis morfológicas pertencentes às amostras dos gêneros de *Oenocarpus* e *Jessenia*.

Variável	Poder discriminante	Poder de modelagem
PF	3,32	0,79
CF	6,89	0,83
DF	11,63	0,66
PS	16,71	0,66
CS	13,87	0,86
DS	6,26	0,75

A Tabela 5 mostra os resultados da previsão das amostras do conjunto teste empregando os modelos obtidos de SIMCA e KNN. Nota-se que todas as cinco amostras do conjunto de teste foram classificadas corretamente.

Tabela 5: Resultados da classificação mediante modelos de SIMCA e KNN das amostras do conjunto de teste pertencentes aos gêneros de *Oenocarpus* e *Jessenia* para dados morfológicos.

SIMCA		
Gênero	Número de amostras	Número de amostras classificadas incorretamente
<i>Oenocarpus</i>	3	0
<i>Jessenia</i>	2	0
KNN		
Gênero	Número de amostras	Número de amostras classificadas incorretamente
<i>Oenocarpus</i>	3	0
<i>Jessenia</i>	2	0

Em resumo, os dois gêneros *Oenocarpus* e *Jessenia* são facilmente discriminados pelos dados morfológicos. Análises multivariada, usando técnicas complexas, não são necessárias para esta discriminação. Análise univariada é suficiente para a classificação correta. Contudo a PCA indicou que dados morfológicos podem ser úteis para discriminar diferentes espécies do mesmo gênero.

Dados morfológicos, físico-químicos e da composição dos ácidos graxos já foram usados para classificação das raças das palmeiras *Bactris gasipaes* H.B.K sob o ponto de vista univariada⁸³ e multivariada⁸⁴. Obteve-se sucesso na classificação usando apenas as variáveis morfológicas⁸⁵.

Segundo Balick¹, dependendo das espécies os frutos variam em tamanho, sendo pequenos para os frutos das espécies do gênero *Oenocarpus* e mais pesados e de maior tamanho para *Jessenia bataua*. No presente trabalho, as medidas das massas dos frutos do gênero *Oenocarpus* estiveram entre 3,58 a 7,74 g, e para os frutos do gênero *Jessenia* entre 11,02 a 13,68 g; enquanto que os comprimentos dos frutos estiveram entre 18,52 a 26,00 mm e 29,75 a 35,58 mm para os *Oenocarpus* e *Jessenia* respectivamente. As variações observadas para a massa e o tamanho dos frutos em ambos os gêneros depende, provavelmente, de fatores como idade, conservação, genótipo, condições do solo, densidade de árvores por hectare, entre outros.

4.2.Dados físico-químicos.

Os dados correspondem a 25 amostras e cinco variáveis físico-químicas que são as medidas de teor de umidade (%Umid), teor de óleo (%Óleo), teor de cinzas (%Cinza), teor de proteínas (%Prot) e índice de refração (nd₄₀) e estão apresentados no apêndice, Tabela A.2. Os teores foram medidos na casca+polpa dos frutos e o índice de refração nos óleos. Um resumo estatístico destes dados está listado na Tabela 6. Observa-se que as médias dos valores das variáveis de *Oenocarpus* diferem pouco das de *Jessenia* comparado com os valores das diferenças observadas dos dados morfológicos.

Análise exploratória dos dados.

Nos dados físico-químicos cada variável também foi autoescalada. PCA foi aplicada a este conjunto de dados. Na Tabela 7 estão listados os loadings das componentes principais 1 e 2. Podemos ver que as duas primeiras componentes principais explicam 61,6% da variância contida nos dados. A primeira componente principal (PC1) descreve 35,6% da variância, e a segunda (PC2) 26,0%.

Tabela 6: Resumo estatístico dos dados físico-químicos

Variável	Total	N ^a	Média	Mínimo	Máximo	Desvio Padrão
<i>Oenocarpus</i>						
%Umid ^a		15	52,24	43,21	58,86	4,73
%Óleo ^b		15	22,91	11,08	36,73	6,74
%Cinza ^c		15	1,81	1,36	2,27	0,32
%Prot ^d		15	7,81	5,28	10,14	1,42
nd ₄₀ c ^e		15	1,4617	1,4600	1,4661	0,0018
<i>Jessenia</i>						
%Umid		10	43,22	39,6	47,35	2,35
%Óleo		10	22,19	17,05	28,20	4,29
%Cinza		10	2,00	1,34	2,70	0,37
%Prot		10	7,09	3,34	10,63	2,50
nd ₄₀ c		10	1,4606	1,4572	1,4625	0,0015

^a número de amostras.^a teor de umidade,^b teor de óleo,^d teor de proteínas,^c teor de cinzas,^e índice de refração

A Figura 19 (PC1xPC2) mostra o gráfico dos escores das duas primeiras componentes principais. Visualiza-se que PC1 separa parcialmente os dados, ficando a maior parte das amostras do gênero *Jessenia* para a esquerda e as de *Oenocarpus* para a direita. No grupo constituído da maioria das amostras de *Jessenia* encontram-se aparentemente duas amostras de *Oenocarpus*, 1 e 14, sendo que a amostra 1 está mais próxima às amostras de *Jessenia*. Por outro lado a amostra 14 não está tão perto das amostras *Jessenia* como aparece na Figura 19. Isto será mostrado logo usando HCA. Como este gráfico é uma projeção bidimensional de cinco dimensões parte da informação estatística perdida são as distancias da amostra 14 às amostras de *Jessenia*.

O grupo que contém maior número de amostras de *Oenocarpus* está formado de 15 amostras dos quais claramente três pertencem às *Jessenia* (16, 22 e 23). À direita da PC1 estão afastadas as amostras de *Oenocarpus* 11 e 13. A variável responsável pela discriminação parcial é %Umid, devido seu alto loading na PC1 (ver Tabela 7). A maioria das amostras de *Oenocarpus* tem %Umid maior do que as encontradas nas amostras de *Jessenia*. A localização da amostra 11 ao longo da PC1, deve-se ao maior valor de %Umid de todas as amostras. A localização da amostra 14 ao longo da PC2 deve-se ao seu alto índice de refração do óleo.

HCA foi usada com o método de conexão completa, obtendo-se o dendrograma da Figura 20, onde se verifica a formação de 4 grupos com níveis de similaridades entre 0 e 0,4. De cima para baixo: um grupo formado de 6 amostras de *Jessenia* e uma amostra de do gênero *Oenocarpus* (1), o segundo grupo contém 11 amostras de *Oenocarpus* e 4 amostras de *Jessenia* (24, 23, 22 e 16). Além destes dois grupos majoritários se vê a amostra 14 isolada e as amostras 13 e 11 formando um grupo. Estes grupos são semelhantes aos vistos no gráfico de componentes principais.

Tabela 7: Loadings das variáveis nas duas primeiras componentes principais para dados físico-químicos.

	PC1	PC2
%Umid	0,601	0,098
%Óleo	-0,491	0,288
%Cinza	-0,503	-0,115
%Prot	0,277	0,694
nd _{40 C}	0,262	-0,643
*	35,6	26,0

* percentagem de variância explicada em cada componente.

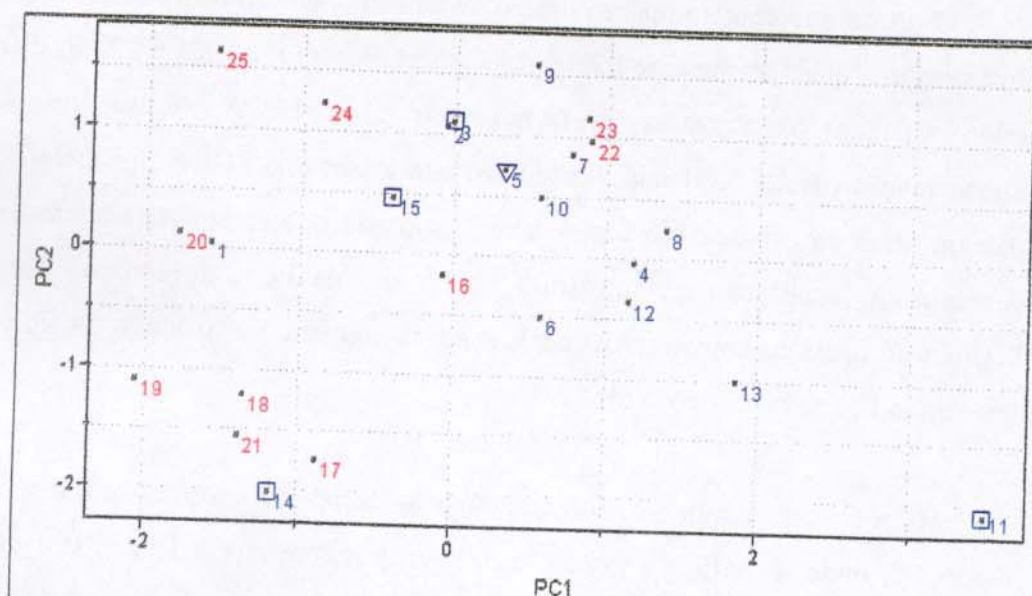


Figura 19: Gráficos dos escores das duas primeiras componentes principais PC1xPC2: *Oenocarpus* (azul): • *O. minor*, ▽ *O. bacaba*, □ *O. mapora*; *Jessenia* (vermelho) para dados físico-químicas.

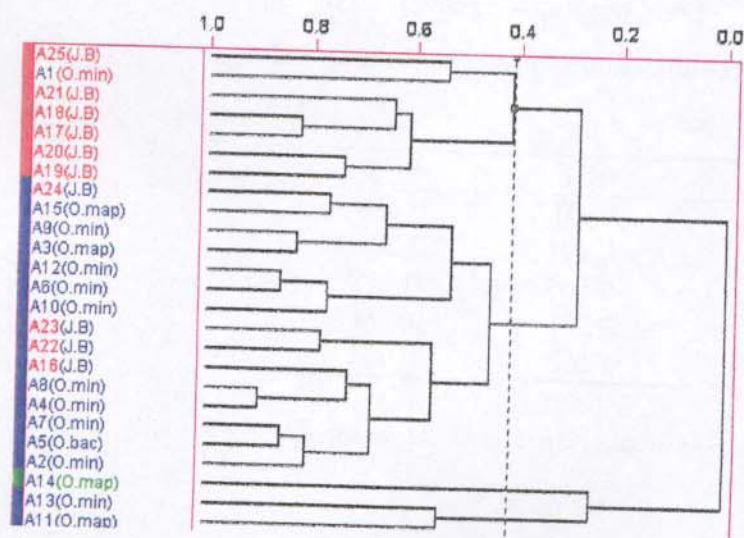


Figura 20: Dendrograma de amostras com variáveis físico-químicas: *Oenocarpus* (azul) e *Jessenia* (vermelho).

Classificação dos dados.

Aplicação do SIMCA gerou os modelos individuais para os gêneros *Oenocarpus* e *Jessenia*. Para a determinação do número ótimo de PCs para a modelagem de cada gênero, foram usadas a técnica de Validação Cruzada "Full leave one out" e critérios de variância contida nos PC, análises dos resíduos versus valor da variável e SEV-PCA.

A partir do conjunto de treinamento obteve-se o modelo *Oenocarpus* para variáveis físico-químicas ficando definido por uma hipercaixa de 64,3% da variância ao longo da PC1 e 27,3% ao longo da PC2, e o modelo *Jessenia* por uma hipercaixa de 46,5% da variância ao longo da PC1 e 32,1% ao longo da PC2.

A Tabela 8 mostra os resultados da aplicação de SIMCA e KNN para a classificação das amostras do conjunto de teste, notando-se que todas essas amostras foram classificadas corretamente. As variáveis que tiveram maiores poderes discriminatório e de modelagem foram %Umid, %Óleo e %Cinza.

Tabela 8: Resultados da classificação mediante SIMCA e KNN das amostras do conjunto de teste pertencentes aos gêneros *Oenocarpus* e *Jessenia*.

SIMCA		
Gênero	Número de amostras	Número de amostras classificadas incorretamente
<i>Oenocarpus</i>	3	0
<i>Jessenia</i>	2	0
KNN		
Gênero	Número de amostras	Número de amostras classificadas incorretamente
<i>Oenocarpus</i>	3	0
<i>Jessenia</i>	2	0

A variável mais importante para classificar os gêneros *Oenocarpus* e *Jessenia* de todas as amostras do conjunto teste foi o teor de umidade, embora os teores de óleo e das cinzas também contribuíram para este fim.

Este resultado poderia ser considerado surpreendente porque o teor de umidade depende de vários outros fatores como, chuva, solo, época de coleta, entre outros. A Tabela 8 mostra que a classificação das amostras do conjunto teste teve sucesso. Supondo que a classificação é completamente aleatória e cada amostra tem 50% de chance de ser classificada corretamente em seu gênero, a probabilidade de classificar corretamente todas as amostras do conjunto teste é somente 1/32.

4.3. Dados de composição em ácidos graxos dos óleos.

Os dados foram obtidos somente de 24 amostras porque a amostra número 15 foi totalmente consumida nas análises anteriores. As seis variáveis usadas foram as percentagens dos seus ácidos graxos nos óleos. Os dados estão apresentados no apêndice, Tabela A.3. O cromatograma dos padrões de ésteres metílicos está apresentado na Figura A.1, no apêndice. Um resumo estatístico destes dados está listado na Tabela 9. Observa-se que as médias dos valores das variáveis de *Oenocarpus* têm pequenas diferenças das *Jessenia*. Estas diferenças são de tamanhos comparáveis aos valores dos desvios padrão.

Tabela 9: Resumo estatístico dos dados de composição em ácidos graxos.

Variável	Total	N ^x	Média	Mínimo	Máximo	Desvio Padrão
<i>Oenocarpus</i>						
C14:0 ^a	14	0,12	0,00	0,38	0,11	
C16:0 ^b	14	16,82*	12,95	29,56	4,42	
C18:0 ^c	14	7,54	3,54	14,87	3,42	
C18:1 ^d	14	73,26	54,11	80,47	7,37	
C18:2 ^e	14	1,21	0,00	3,96	1,20	
C18:3 ^f	14	0,78	0,00	2,04	0,55	
<i>Jessenia</i>						
C14:0	10	0,00	0,00	0,00	0,00	
C16:0	10	14,90	12,55	18,51	1,96	
C18:0	10	5,05	2,82	9,31	2,08	
C18:1	10	76,94	66,36	81,04	5,16	
C18:2	10	2,56	1,50	5,76	1,33	
C18:3	10	0,31	0,00	0,79	0,31	

^x número de amostras.

^a Ácido mirístico,

^d Ácido oléico,

^b Ácido palmítico,

^e Ácido linoléico,

^c Ácido esteárico,

^f Ácido linolênico.

* o número de algarismos significativos é 3

Análise exploratória dos dados.

As amostras foram normalizadas para compensar os diferentes volumes das injeções feitas no cromatógrafo e cada variável foi autoescalada.

PCA foi aplicada a este conjunto de dados e os loadings das variáveis estão listados na Tabela 10. Podemos ver que as três primeiras componentes principais explicam 96,5% da variância contida nos dados. A primeira componente principal (PC1) descreve 60,0% da variância, a PC2, 20.1% e PC3, 16.4%.

Tabela 10: Loadings das variáveis nas três primeiras componentes principais para dados de ácidos graxos.

	PC1	PC2	PC3
C14:0	0,455	0,401	-0,092
C16:0	0,472	-0,370	0,105
C18:0	0,501	-0,097	0,126
C18:1	-0,486	0,316	-0,064
C18:2	-0,099	0,125	0,979
C18:3	0,270	0,761	-0,018
*	60,0	20,1	16,4

* percentagem de variância explicada em cada componente.

A Figura 21a é o gráfico dos escores das duas primeiras componentes principais. Visualiza-se uma separação parcial dos dados em dois grupos, ficando os *Oenocarpus* para a direita e as *Jessenia* para a esquerda e para baixo.

A Figura 21b mostra o gráfico dos loadings das variáveis das duas primeiras componentes principais. Vê-se que os loadings de C:18:1 e C:18:2 são negativos e os de C:14:0, C16:0, C:18:0 e C:18:3 são positivos ao longo de PC1. O espalhamento dos escores na Figura 21a mostra que o gênero *Oenocarpus* está mais espalhado ao longo da PC1. A localização das amostras 3 e 5 devem-se aos maiores valores de C16:0 e C18:0. Ao longo da PC2, a localização das amostras 22 e 23 deve-se aos menores valores da variável C18:3 e as amostras 4 e 6 pelos maiores valores desta variável.

Evidentemente não se nota agrupamentos de acordo com os dois gêneros significando que a composição dos ácidos graxos nos óleos, em conjunto, não ajuda muito nas suas discriminações.

HCA foi usada com o método de conexão completa, resultando na formação de 2 grupos majoritários com níveis de similaridades entre 0 e 0,7. Estes grupos não separam de acordo com os dois gêneros em forma semelhante ao observado em PCA.

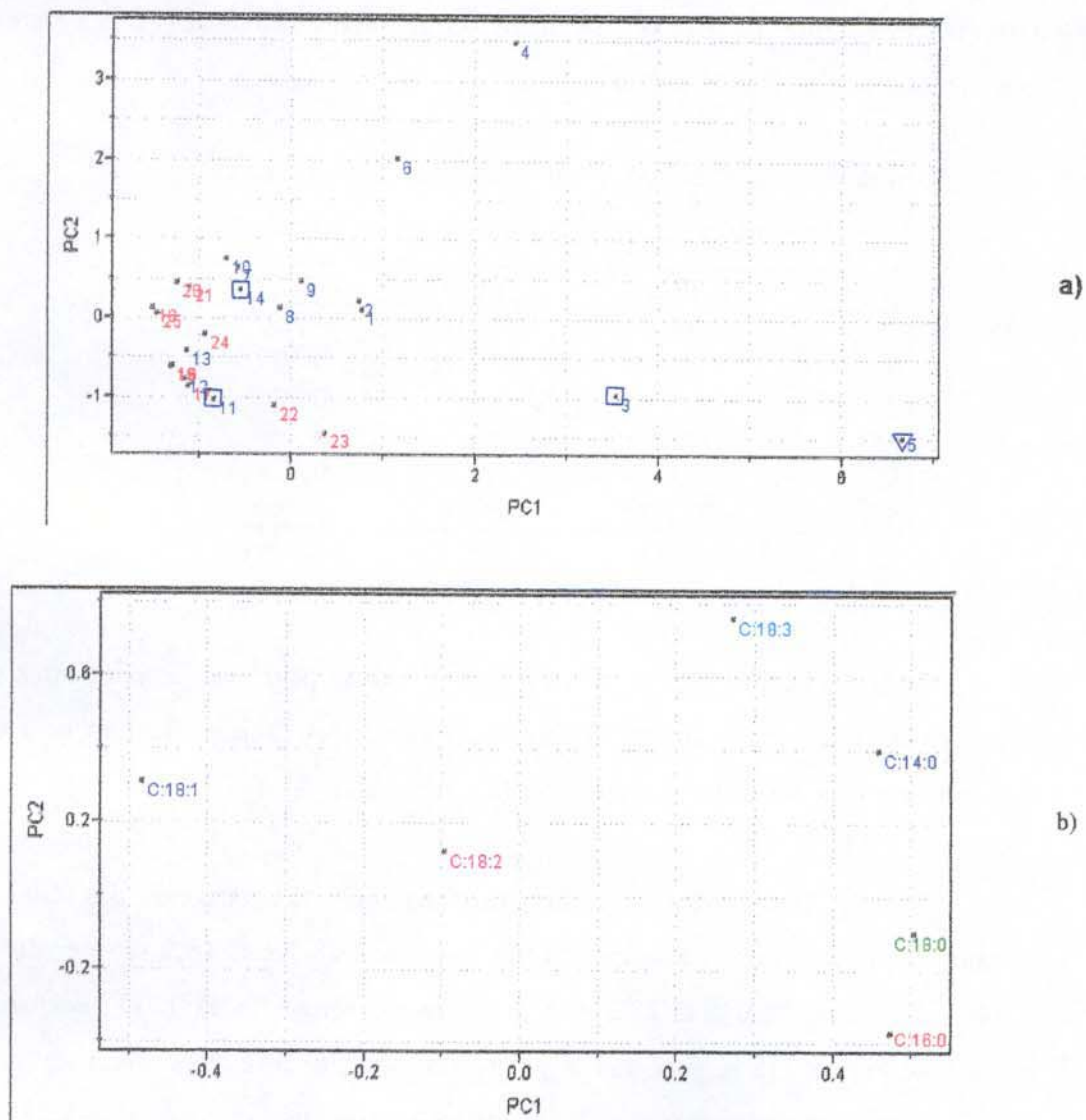


Figura 21: Gráficos das duas primeiras componentes principais para dados de composição em ácidos graxos: a) Escores: *Oenocarpus* (azul), • *O. minor*, ▽ *O. bacaba*, □ *O. mapora*, *Jessenia* (vermelho) b) Loadings das seis variáveis.

Classificação dos dados.

Foi aplicado SIMCA aos dados dos dois gêneros, obtendo-se o modelo *Oenocarpus* com 3 PC e o modelo *Jessenia* com 2 PC. Foram usadas somente 5 variáveis das seis disponíveis, sem considerar a C14:0 devido ao valor de zero que tem para o gênero *Jessenia*. As variáveis que tiveram maiores poderes discriminante na separação foram: 18:1, 18:2 e 18:3. A Tabela 11 mostra os resultados de SIMCA e KNN na previsão das amostras do conjunto de teste.

Tabela 11: Resultados da classificação das amostras do conjunto de teste usando teores de ácidos graxos com modelos de SIMCA e KNN.

SIMCA		
Gênero	Número de amostras	Número de amostras classificadas incorretamente
<i>Oenocarpus</i>	3	0
<i>Jessenia</i>	2	2 ^a
KNN		
Gênero	Número de amostras	Número de amostras classificadas incorretamente
<i>Oenocarpus</i>	3	0
<i>Jessenia</i>	2	0

^a amostras 18 e 22

O perfil dos ácidos graxos em tecidos fotossintéticos serviu para diferenciar duas classes de plantas⁸⁶, como também para classificar e identificar algumas bactérias^{87,15}. No nosso caso, as composições dos ácidos graxos possibilitou discriminar parcialmente entre os gêneros em estudo como visto na Tabela 11. Nota-se que SIMCA só classificou corretamente três amostras, enquanto que com os modelos de KNN não se errou nenhuma das 5 amostras do conjunto teste. O método SIMCA tentou classificar as amostras de *Jessenia* como amostras *Oenocarpus*.

4.4.Dados dos teores dos elementos químicos contidos nas cinzas

Estes dados estão constituídos por 24 amostras, faltando a amostra número 15 por razões explicadas na seção sobre ácidos graxos. As variáveis contêm as medidas dos teores de sete elementos encontrados nas cinzas da casca+polpa dos frutos. Os dados originais são apresentados na Tabela A.4 no apêndice. Um resumo da estatística univariada está listado na Tabela 12. Observa-se que as médias dos valores das variáveis de *Oenocarpus* são pouco diferentes das de *Jessenia* e suas diferenças são comparáveis aos valores dos desvios padrão. Também se nota que os desvios padrão são bem maiores que os erros instrumentais para todos os elementos encontrados, salvo fósforo, indicando que há variações significativas nos teores destes elementos nas diferentes amostras do mesmo gênero.

Tabela 12: Resumo estatístico das amostras com teores de elementos químicos.

Variável	Total	Média ^a	Mínimo ^a	Máximo ^a	Desvio Padrão ^a	Erro ^b
<i>Oenocarpus</i>						
P	14	0,344	0,114	0,602	0,136	0,1046
K	14	6,010	1,627	7,925	1,689	0,0243
Ca	14	0,625	0,156	1,473	0,378	0,0112
Mn	14	0,030	0,005	0,189	0,049	0,0008
Fe	14	0,023	0,009	0,045	0,010	0,0011
Zn	14	0,033	0,000	0,290	0,075	0,0008
Cu	14	0,005	0,001	0,010	0,002	0,0006
<i>Jessenia</i>						
P	10	0,144	0,000	0,283	0,078	0,0834
K	10	4,524	2,861	7,642	1,700	0,0211
Ca	10	1,349	0,577	2,601	0,680	0,0123
Mn	10	0,052	0,008	0,120	0,036	0,0009
Fe	10	0,035	0,014	0,060	0,012	0,0008
Zn	10	0,072	0,001	0,554	0,170	0,0009
Cu	10	0,005	0,003	0,008	0,002	0,0006

^a Estimativas estatísticas das 14 amostras de *Oenocarpus* e das 10 amostras de *Jessenia*;

^b Erros relativos das determinações individuais dos elementos padrões segundo a resposta instrumental.

Análise exploratória dos dados

Os dados foram autoescalados e PCA foi aplicada a este conjunto. Na Tabela 13 estão listados os loadings das variáveis. Podemos verificar que as três primeiras componentes principais explicam 81% da variância contida nos dados. A primeira componente principal descreve 38,5% da variância, a segunda 26,2% e a terceira 16,3%.

A Figura 22a (PC1xPC2) mostra o gráfico dos escores das duas primeiras componentes principais com pontos para as amostras de *Oenocarpus* na parte superior e as *Jessenia* na parte inferior. A Figura 22b mostra o gráfico dos loadings das variáveis das duas primeiras componentes principais. PC2 faz discriminação parcial devido aos teores de P e K que são altos para as amostras *Oenocarpus* e baixo para amostras *Jessenia*. A localização da amostra 13 ao longo de PC2 deve-se aos altos conteúdos P e K. No entanto a localização da amostra 12 deve-se ao menor conteúdo de Ca.

Análise Hierárquica de Agrupamentos com o método de conexão completa aplicado a estes dados, formou grupamentos semelhantes a PCA com níveis de similaridades entre 0 e 0,24.

Tabela 13: Loadings das variáveis nas três primeiras componentes principais para os dados de teor de elementos químicos.

	PC1	PC2	PC3
P	0,088	0,564	-0,406
K	0,178	0,646	0,017
Ca	0,475	-0,341	-0,026
Mn	0,517	-0,158	-0,156
Fe	0,549	-0,094	0,023
Zn	0,019	0,129	0,867
Cu	0,407	0,3148	0,238
*	38,5	26,2	16,3

* percentagem de variância explicada em cada componente.

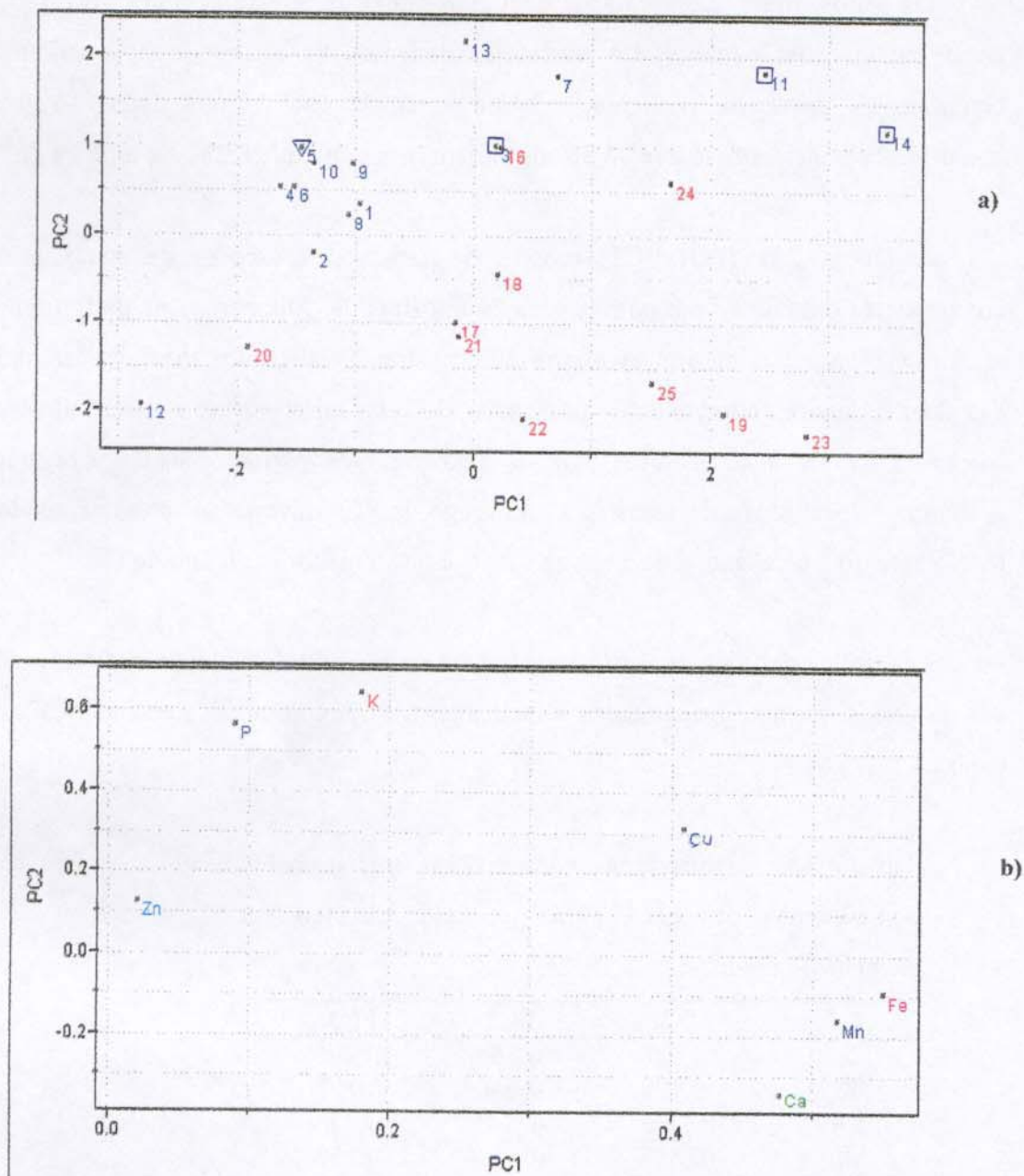


Figura 22: Gráficos das duas primeiras componentes principais para dados de teores de elementos químicos: a) Escores: *Oenocarpus* (azul), • *O. minor*, ▽ *O. bacaba*, □ *O. mapora*; *Jessenia* (vermelho) b) Loadings das sete variáveis.

Classificação dos dados.

A aplicação do SIMCA gerou os modelos individuais para os gêneros *Oenocarpus* e *Jessenia*. Para a determinação do número ótimo de PC na modelagem de cada gênero foram usadas: a técnica de Validação Cruzada “Full leave one out”, critérios de variância contida nos PC, análise dos resíduos versus distância Mahalanobis e SEV-PCA.

Os modelos foram construídos a partir das amostras do conjunto de treinamento. Deve-se mencionar que a amostra 20 pertencente a este conjunto, esta mais próxima aos pontos do gênero *Oenocarpus*. O modelo *Oenocarpus* para variáveis de teor de elementos nas cinzas ficou definido por um cilindro num espaço tridimensional contendo 76,8% da variância ao longo da PC1. O modelo *Jessenia* ficou definido por um cilindro com 55,3% da variância ao longo da PC1. A Tabela 14 mostra os resultados dos poderes discriminante e de modelagem das variáveis, onde nota-se que as variáveis P, K, Ca e Fe têm os maiores valores de poderes de discriminação e modelagem.

Tabela 14: Poderes discriminante e de modelagem das variáveis de teor de elementos químicos pertencentes às amostras dos gêneros de *Oenocarpus* e *Jessenia*.

Variável	Poder discriminante	Poder de modelagem
P	43,60	0,52
K	11,75	0,63
Ca	18,85	0,62
Mn	3,47	0,61
Fe	9,73	0,70
Zn	2,94	0,27
Cu	2,80	0,54

A Tabela 15 mostra os resultados da previsão das amostras do conjunto de teste empregando os modelos obtidos com SIMCA e KNN. Notando-se que SIMCA acertou quatro amostras e errou uma (18), e KNN acertou as cinco.

Tabela 15: Resultados da classificação mediante SIMCA e KNN das amostras do conjunto de teste dos gêneros *Oenocarpus* e *Jessenia* para teor de elementos químicos.

SIMCA		
Gênero	Número de amostras	Número de amostras classificadas incorretamente
<i>Oenocarpus</i>	3	0
<i>Jessenia</i>	2	1 ^a
KNN		
Gênero	Número de amostras	Número de amostras classificadas incorretamente
<i>Oenocarpus</i>	3	0
<i>Jessenia</i>	2	0

^a amostra 18

Aqui, foi determinado o teor dos elementos químicos nas cinzas dos gêneros *Oenocarpus* e *Jessenia* mediante EDXRF, encontrando-se sete elementos: P, K, Ca, Mn, Fe, Zn e Cu. Embora as composições destes gêneros sejam similares, é possível encontrar parâmetros que possam diferenciá-las. Foi assim que P, K, Ca e Fe mostraram maiores poderes discriminatório e de modelagem, sendo úteis para fazer as classificações das amostras. A Tabela 15 mostra o sucesso de SIMCA e KNN na classificação das amostras do conjunto teste.

4.5. Dados de espectros de fluorescência de raios-X de energia dispersiva (EDXRF).

Os dados foram obtidos das cinzas das 24 amostras com um mil e vinte e quatro (1024) variáveis correspondentes às respostas dos raios-X secundários dos elementos químicos em quilo elétron volt (keV) entre 0 a 20 keV em intervalos de 0,0195 keV. A Figura 23 mostra os espectros de EDXRF das 14 amostras de *Oenocarpus* (azul) e dez amostras de *Jessenia* (vermelho). Nota-se que a maioria das respostas para os dois gêneros são semelhantes, embora uma observação mais detalhada da para perceber que existem faixas com pequenas diferenças, as que poderiam ajudar na diferenciação destes dois gêneros. Uma parte dos dados numéricos originais estão apresentados no apêndice,

Tabela A.5. Os espectros de EDXRF a partir de dados numéricos para cada gênero estão apresentados nas Figuras A.2 e A.3, também no apêndice, e o espectro da amostra 8 na faixa de energia das respostas relevantes está na Figura A.4. Esta visualização inicial já mostra a complexidade da estrutura destes dados e a necessidade de empregar análise multivariada para seu tratamento.

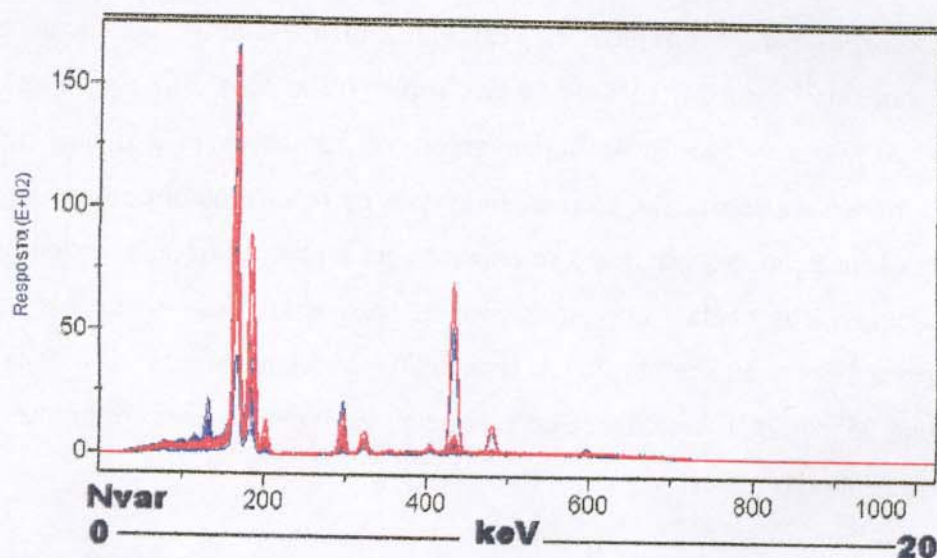


Figura 23: Espectros de EDXRF com cinzas das amostras de *Oenocarpus* (azul) e *Jessenia* (vermelho) mostrando no eixo das abscissas: Número de variáveis (Nvar) e energia em quilo elétron volt (keV).

Análise exploratória.

As variáveis foram centradas na média e análise de componentes principais foi aplicada. A Figura 24 mostra os loadings das variáveis para as duas primeiras PCs (PC1 azul e PC2 vermelho). Estas componentes principais explicam 82.4% da variância contida nos dados. A primeira componente principal descreve 62,5% da variância e a PC2 19,9%.

A Figura 25 mostra o gráfico dos escores nas duas primeiras componentes principais. Visualiza-se que PC1 discrimina parcialmente os dados, ficando os *Oenocarpus* para a direita e as *Jessenia* para a esquerda. As faixas de energia nos espectros da Figura 24: 1,93 a 4,19 keV (Nvar 100 a 216), 5,48 a 6,14 keV (282 a 316) e 8,37 a 8,76 keV (430 a 450) têm grandes loadings na PC1 o que significa que são importantes na separação. Comparando com os resultados obtidos na Figura 22 nota-se que a separação dos gêneros é semelhante em ambas as figuras devido que as respostas espectrais observadas correspondem às respostas dos elementos químicos mais importantes na discriminação. A separação parcial na Figura 22 é feita por PC2 e na Figura 25 por PC1. Isso acontece porque as variáveis em uma foram autoescaladas e na outra centradas na média.

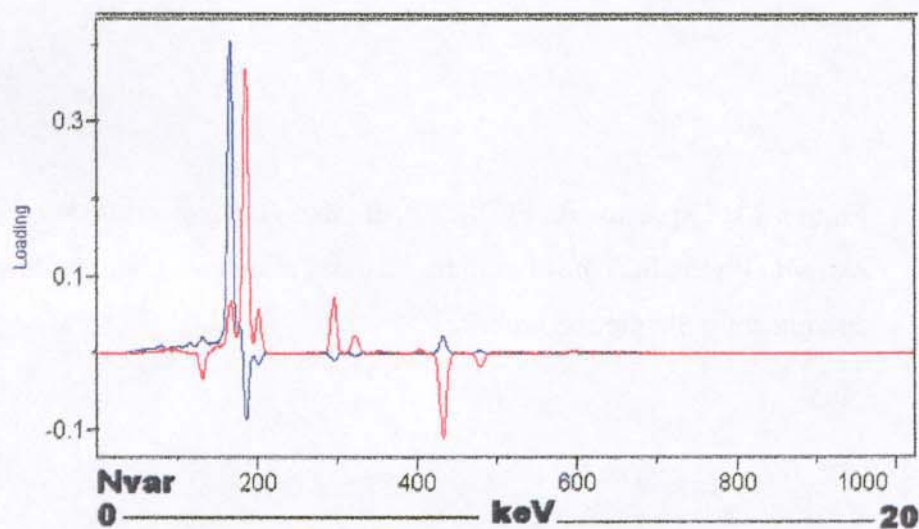


Figura 24: Gráfico dos loadings das PC versus variáveis dos espectros de EDXRF.(PC1 azul, PC2 vermelho).

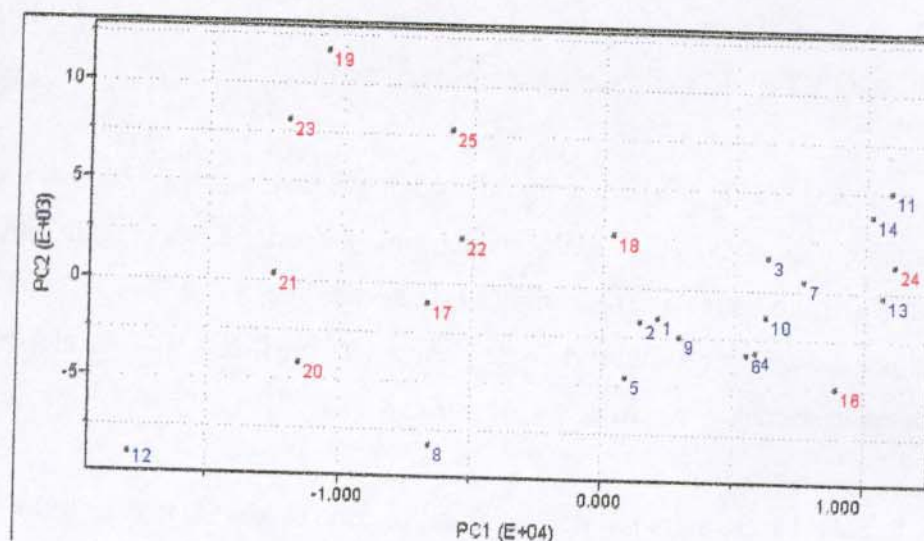


Figura 25: Gráfico dos escores nas duas primeiras componentes principais PC1-PC2 usando espectros de EDXRF: *Oenocarpus* (azul) e *Jessenia* (vermelho).

Análise hierárquica de agrupamento foi usada com o método de conexão completa, formando-se dois grupos com níveis de similaridades entre 0 e 0,26, semelhantes aos observados no gráfico de componentes principais. Os grupos formados são semelhantes em discriminação aos encontrados no dendrograma com variáveis de teor de elementos químicos.

Classificação mediante os dados espectrais de EDXRF.

Empregando as amostras do conjunto de treinamento foram construídos os modelos de *Oenocarpus* e *Jessenia*. As variáveis foram escolhidas mediante o critério de erro de previsão na classificação, ficando 71 variáveis das um mil e vinte e quatro iniciais. As faixas de energia escolhidas foram: 2,38 a 2,69 keV (Nvar 123 a 139) e 3,02 a 4,06 keV (156 a 209). Observação das figuras 23 e 24 mostra que as características espectrais das amostras dos dois gêneros são bastante diferentes nestas faixas de espectros.

A aplicação do SIMCA gerou os modelos individuais para os gêneros *Oenocarpus* e *Jessenia*. O modelo *Oenocarpus* ficou definido por uma hipercaixa num espaço de 71

dimensões contendo 88,7% da variância ao longo de PC1 e 9,9% da variância ao longo de PC2. O modelo *Jessenia* ficou definido por uma hipercaixa com 84,1% da variância ao longo do seu PC1 e 15,7% da variância ao longo de PC2.

Estes modelos de SIMCA, com duas componentes principais cada um, classificaram corretamente quatro das cinco amostras do conjunto teste. KNN classificou corretamente também quatro amostras e errou uma (18), como se mostra na Tabela 16. Entretanto os resultados de classificação observado na Tabela 15, o SIMCA e KNN classificaram uma amostra incorretamente, sendo esta a amostra 18.

Tabela 16: Resultados da classificação através de SIMCA e KNN das amostras do conjunto teste dos gêneros *Oenocarpus* e *Jessenia* usando espectros de EDXRF.

SIMCA		
Gênero	Número de amostras	Número de amostras classificadas incorretamente
<i>Oenocarpus</i>	3	0
<i>Jessenia</i>	2	1 ^a
KNN		
Gênero	Número de amostras	Número de amostras classificadas incorretamente
<i>Oenocarpus</i>	3	0
<i>Jessenia</i>	2	1 ^a

^a amostra 18

4.6.Dados dos espectros de difração de raios-X (XRD).

Este conjunto também foi obtido das cinzas das amostras. Cada amostra foi caracterizada por duas mil e quinhentos e uma (2501) variáveis que são as respostas da difração entre os ângulos 20 a 70° em intervalos de 0,02°. Os dados numéricos para obter os difratogramas estão apresentados no apêndice, Tabela A.6. A Figura 26 mostra os difratogramas das 24 amostras em estudo. Estes difratogramas foram comparados visualmente e encontrou-se que para os *Oenocarpus* (azul), em geral, as respostas foram

maiores do que para as *Jessenia* (vermelho). Estas diferenças são melhor observadas em determinadas faixas dos difratogramas. Para melhor observação os difratogramas dos dois gêneros separadamente estão apresentados nas Figuras A.5 e A.6 no apêndice.

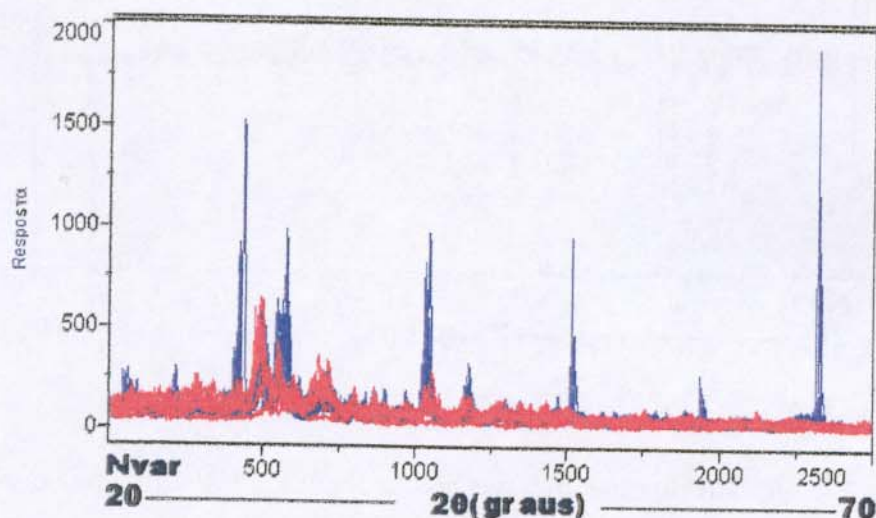


Figura 26: Gráfico dos difratogramas de XRD nas cinzas das amostras de *Oenocarpus* (azul) e *Jessenia* (vermelho).

Análise exploratória.

Com a finalidade de reduzir as variações decorrentes do processo incompleto de homogeneização das amostras, os difratogramas foram normalizados e as variáveis foram centradas na média, e PCA foi aplicada a este conjunto de dados. A Figura 27 mostra os loadings das variáveis para as PC. As três primeiras componentes principais explicaram 57,3% da variância contida nos dados. A primeira componente principal PC1 descreveu 22,8% da variância, a PC2 19,2% e a PC3 15,3%.

A Figura 28 mostra o gráfico dos escores das duas primeiras componentes principais. Visualiza-se que PC1 separa parcialmente, ficando os *Oenocarpus* para a esquerda e as *Jessenia* para a direita. As amostras A4 e A8 ao longo de PC2 têm grandes escores em relação às demais.

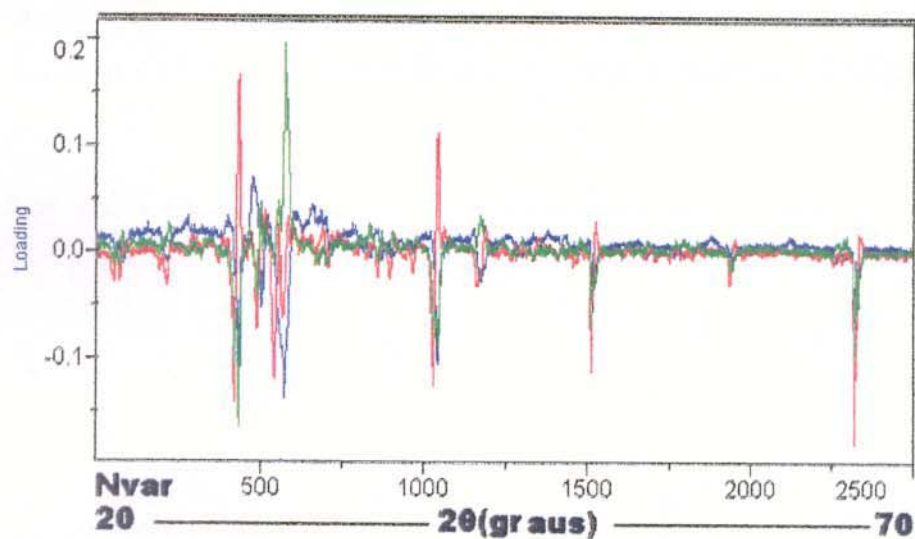


Figura 27: Gráfico dos loadings versus PC1(azul), PC2(vermelho) e PC3(verde) dos difratogramas de XRD.

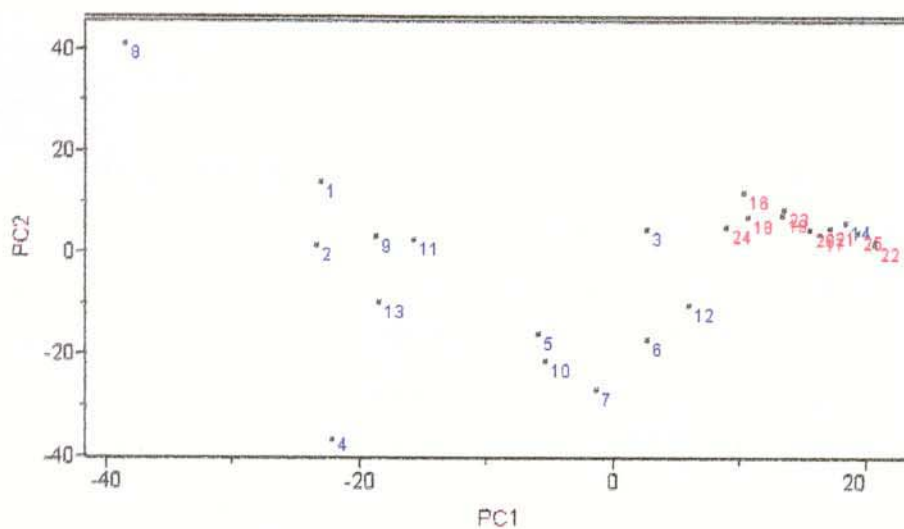


Figura 28: Gráfico dos escores nas duas primeiras PC dos difratogramas de XRD para os gêneros *Oenocarpus* (azul) e *Jessenia* (vermelho).

Os difratogramas da Figura 26 analisados em conjunto com os loadings em PC1, Figura 27, permitem escolher as faixas que têm grandes loadings ou seja aqueles que são importantes na discriminação.

HCA foi aplicada com o método de conexão completa, obtendo-se a formação de 3 grupos com níveis de similaridades entre 0 e 0,44 e novamente as amostras 4 e 8 estão afastadas das demais. As características dos agrupamentos obtidos com HCA são muito similares aqueles obtidos com PCA.

Classificação mediante os difratogramas de XRD.

Para desenvolver os modelos de SIMCA e KNN foram utilizadas amostras do conjunto de treinamento. Ao trabalhar com este conjunto, notou-se que as amostras 4 e 8 tinham alta distância Mahalanobis aos pontos do gênero *Oenocarpus* e a amostra 14 estava mais próxima aos pontos das amostras de *Jessenia*.

As variáveis foram escolhidas mediante o critério de erro de previsão na classificação, restando 288 variáveis, das 2501 iniciais. As faixas escolhidas dos difratogramas em concordância com os “loadings” em função de 2θ em graus ($^{\circ}$) foram: 20,70 a 20,98 $^{\circ}$ (Nvar 36 a 50), 24,30 a 24,56 $^{\circ}$ (216 a 229), 28,38 a 28,78 $^{\circ}$ (420 a 440), 29,28 a 29,48 $^{\circ}$ (465 a 475), 30,58 a 31,78 $^{\circ}$ (530 a 590), 40,20 a 40,98 $^{\circ}$ (1011 a 1050), 43,18 a 43,56 $^{\circ}$ (1160 a 1179), 50,00 a 50,78 $^{\circ}$ (1501 a 1540), 58,48 a 58,98 $^{\circ}$ (1925 a 1950) e 66,20 a 66,98 (2311 a 2350).

O modelo *Oenocarpus* para variáveis de XRD nas cinzas ficou definido por um hiper-volume num espaço de 288 dimensões contendo 34,9% da variância ao longo de PC1, 24,9% da variância ao longo de PC2 e 16,9% da variância ao longo de PC3. O modelo *Jessenia* ficou definido por uma hipercaixa com 39,3% da variância ao longo de PC1 e 21,0% da variância ao longo de PC2. A Tabela 17 mostra que os modelos de SIMCA classificaram corretamente as cinco amostras do conjunto teste e KNN acertou 4 amostras e errou uma (3).

Tabela 17: Resultados da classificação através de SIMCA e KNN das amostras do conjunto teste dos gêneros *Oenocarpus* e *Jessenia* usando difratogramas XRD.

SIMCA		
Gênero	Número de amostras	Número de amostras classificadas incorretamente
<i>Oenocarpus</i>	3	0
<i>Jessenia</i>	2	0
KNN		
Gênero	Número de amostras	Número de amostras classificadas incorretamente
<i>Oenocarpus</i>	3	1 ^a
<i>Jessenia</i>	2	0

^a amostra 3

Da mesma forma que trabalhou-se com os espectros puros de EDXRF, também se empregou difratogramas de raios-X para a obtenção dos modelos, os quais funcionaram com bom acerto na classificação de amostras de *Oenocarpus* e *Jessenia* como mostrada na Tabela 17.

4.7. Dados espectroscópicos de ultravioleta/visível (UV/Vis).

Estes dados contêm as medidas de 21 absorvâncias no UV/Vis dos extratos de éter de petróleo das farinhas da casca+polpa dos frutos. Os comprimentos de onda trabalhados foram de 240 a 440 nm, em intervalos de dez nm. Os dados originais estão apresentados no apêndice, Tabela A.7.

A Figura 29 mostra os espectros de UV/Vis para as amostras. Nesta figura pode-se visualizar as faixas que são diferentes nos espectros de ambos os gêneros e que possivelmente serviriam para a discriminação: de 240 a 290 nm.

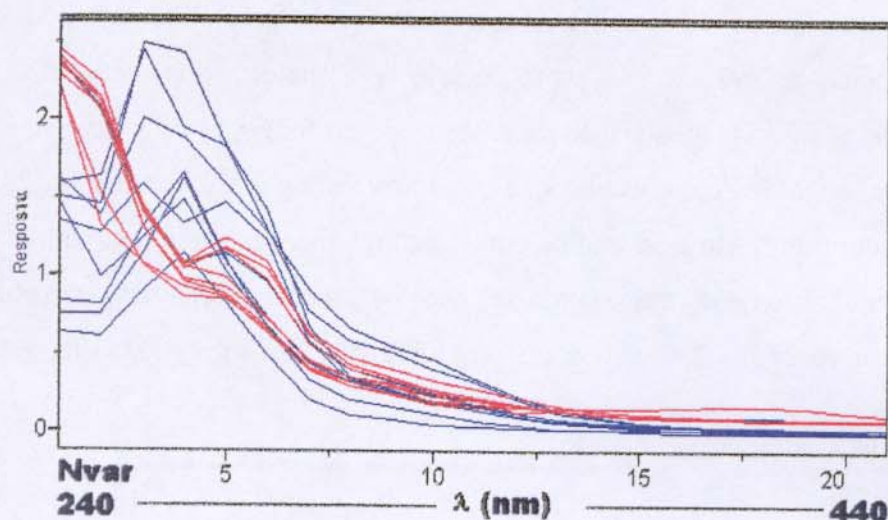


Figura 29: Espectros de UV/Vis para as amostras: *Oenocarpus* (azul) e *Jessenia* (vermelho). Na abscissa temos a escala do número da variável (Nvar) e a escala do comprimento de onda (λ) em nanômetro (nm).

Análise exploratória dos dados

Os dados foram centrados na media e PCA foi aplicada a este conjunto. Na Tabela 18, estão listados os loadings das variáveis relevantes. As três primeiras PCs explicaram 94% da variância contida nos dados. A PC1 descreveu 57,1% da variância; PC2, 28,7%; e PC3, 8%.

Tabela 18: Loadings das variáveis nas três primeiras componentes principais para dados espectroscópicos de UV/Vis de 240 a 290 nm.

nm	PC1	PC2	PC3
240	0,251	0,709	-0,062
250	0,419	0,453	-0,234
260	0,497	-0,173	-0,233
270	0,426	-0,423	-0,227
280	0,434	-0,234	0,008
290	0,317	-0,051	0,541
*	57,1	28,7	8,0

* percentagem de variância explicada em cada componente.

A Figura 30a (PC1xPC2) mostra o gráfico dos escores das duas primeiras componentes principais. Visualiza-se que PC2 separa parcialmente os dados em dois grupos, ficando as *Jessenia* na parte superior e a maioria dos *Oenocarpus* na parte inferior. Nesta figura a localização da amostra 12 ao longo de PC1 deve-se à absorção a 260 nm, ao longo de PC2 a localização das amostras 5 e 7 a 270 nm e 16, 22 e 25 a 240 nm. A Figura 30b mostra os gráficos dos loadings das variáveis para as duas primeiras componentes principais observando-se que os maiores loadings correspondiam às absorções entre 240 a 290 nm, o que significa que são importantes na separação dos gêneros.

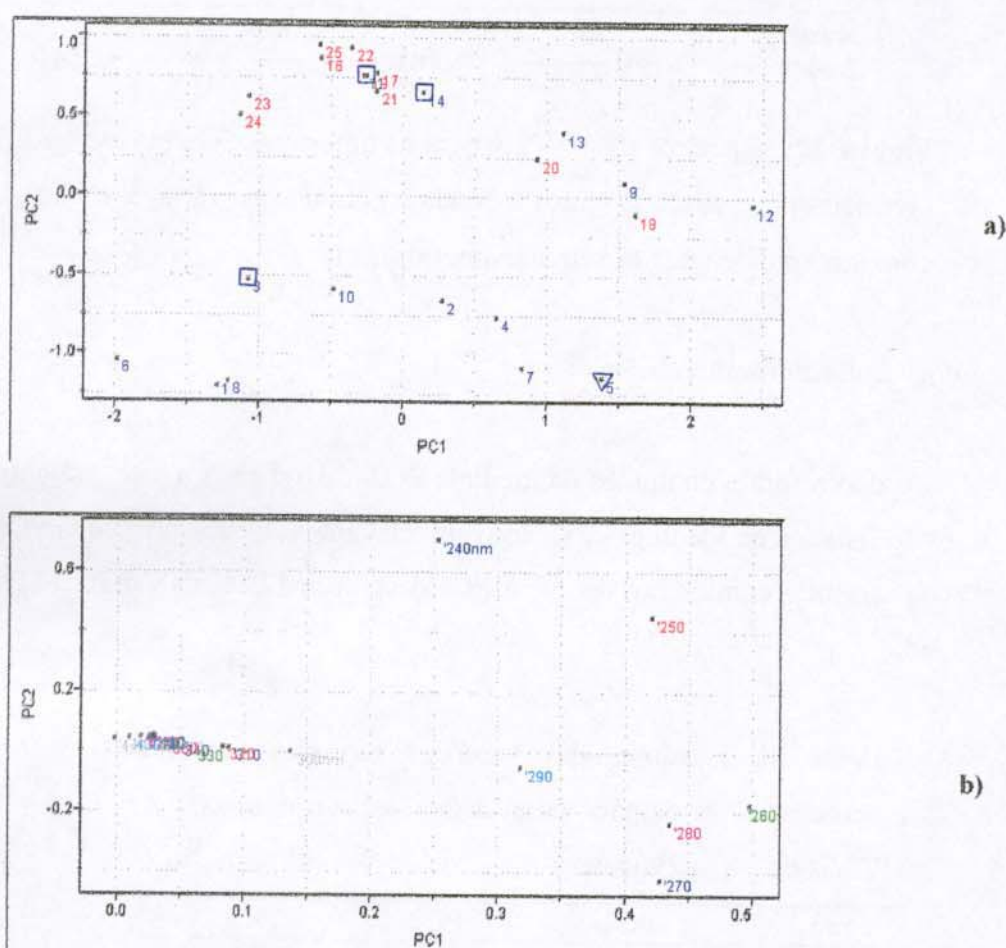


Figura 30: Gráficos das duas primeiras componentes principais dos dados de UV/Vis: a) Escores: *Oenocarpus* (azul), • *O. minor*, ▽ *O. bacaba*, □ *O. mapora*; *Jessenia* (vermelho) b) Loadings.

HCA foi usada com o método de conexão completa e com as variáveis centradas na média, obtendo-se 2 grupos majoritários nos níveis de similaridade entre 0 e 0,5 e que são semelhantes aos observados nos gráficos das componentes principais.

Classificação dos dados.

A modelagem para classificação com SIMCA foi realizada usando as amostras do conjunto de treinamento. As variáveis que tiveram maior poder discriminatório e poder de modelagem estiveram na faixa de 240 a 290 nm, cuja região pertencem as absorções dos flavonóides. SIMCA gerou o modelo *Oenocarpus* ficando definido por uma hipercaixa contendo 69,7% da variância ao longo da PC1 e 21,9% ao longo de PC2, e o modelo *Jessenia* por um hipercilindro de 81,6% da variância ao longo do PC1. SIMCA classificou corretamente as cinco amostras do conjunto teste, enquanto que KNN só classificou corretamente quatro amostras e errou uma (18). A Tabela 19 mostra os resultados da aplicação de SIMCA e KNN na previsão das amostras do conjunto de teste pertencentes aos dois gêneros em estudo.

Tabela 19: Resultados da classificação através de SIMCA e KNN dos gêneros *Oenocarpus* e *Jessenia* usando espectros de UV/Vis.

SIMCA		
Gênero	Número de amostras	Número de amostras classificadas incorretamente
<i>Oenocarpus</i>	3	0
<i>Jessenia</i>	2	0
KNN		
Gênero	Número de amostras	Número de amostras classificadas incorretamente
<i>Oenocarpus</i>	3	0
<i>Jessenia</i>	2	1 ^a

^a amostra 18

As variáveis espectroscópicas obtidas no UV/Vis que mais discriminaram entre *Oenocarpus* e *Jessenia* compreendiam a região de absorção dos flavonóides. Isso está de acordo com o fato que os flavonóides, bem como os terpenóides e os lipídios, são úteis

para resolver problemas quimiotaxonômicos, devido sua diversidade estrutural e sua ampla distribuição na natureza^{46,47} Nossos resultados concordam com os obtidos por Balick¹ que encontrou somente 32% dos flavonóides semelhantes em ambos os gêneros *Oenocarpus* e *Jessenia*, significando que os outros 68% dos flavonóides poderiam servir para fins de diferenciação entre estes gêneros.

4.8.Dados dos espectros de infravermelho com transformada de Fourier (FTIR).

Estes dados foram obtidos das amostras de cinzas dos gêneros *Oenocarpus* e *Jessenia*. As variáveis são as medidas das absorvâncias no IR entre 4000 a 400 cm^{-1} em intervalos de 1,93 cm^{-1} , resultando em um total de um mil oitocentos sessenta e oito (1868) variáveis. A Figura 31 mostra os espectros de FTIR, onde vê-se que poucos espectros das amostras de *Jessenia* (vermelho) tiveram suas absorvâncias maiores que as dos *Oenocarpus* (azul), mas em geral a maioria dos espectros foram semelhantes. As absorções originais para obter os gráficos estão na Tabela A.8 no apêndice. Os espectros para cada gênero podem ser visualizados nas Figuras A.7 e A.8 no apêndice.

A estratégia nesta parte foi escolher as máximas absorções e pontos do espectro com maior discordância para obter gráficos bidimensionais. Desta forma pode-se observar as separações e/ou as correlações existentes. A Figura 32 mostra estes gráficos para as variáveis de IR em determinados números de onda. Nota-se separação parcial, sendo que pontos das amostras do gênero *Oenocarpus* ficam perto do origem e os de *Jessenia* mais afastado do origem, também percebe-se boa correlação.

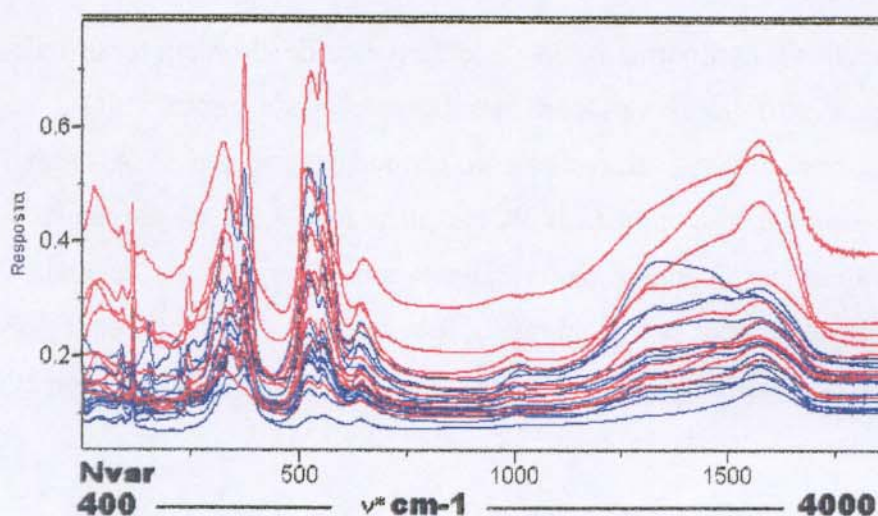


Figura 31: Espectros de FTIR médio das cinzas das amostras de *Oenocarpus* (azul) e *Jessenia* (vermelho).

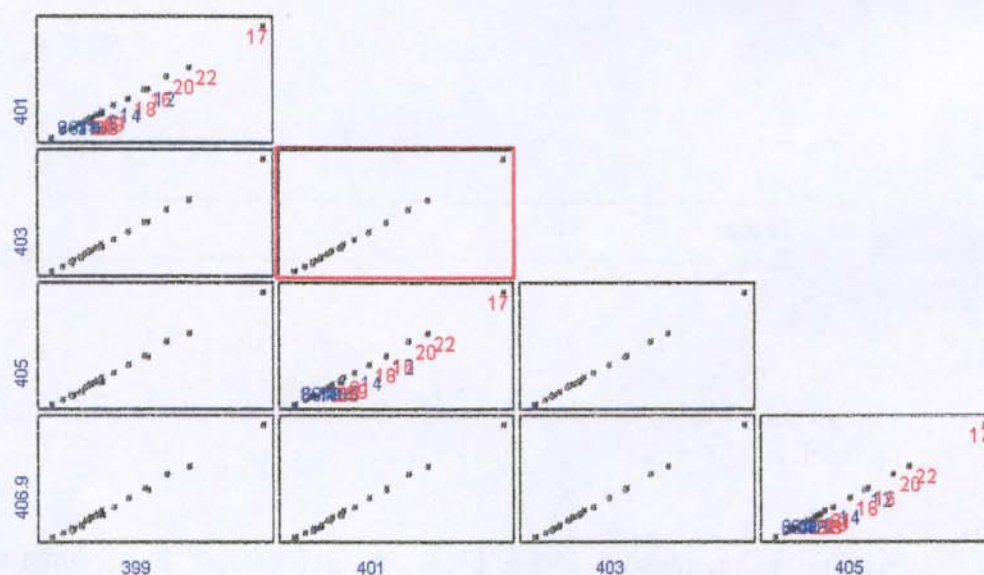


Figura 32: Gráficos bidimensionais das absorções em vários números de onda

Análise exploratória dos dados espectrais de IR.

Os espectros foram normalizados, e as derivadas de 2ª ordem foram calculadas, com a finalidade de corrigir as variações nas proporções de cinzas e KBr, nas pastilhas, espessura das pastilhas. As variáveis foram centradas na média. PCA foi aplicada a este conjunto de dados e as 3 primeiras PCs explicaram 79,5% da variância. A Figura 33 mostra o gráfico dos loadings das variáveis para a primeira e segunda componente principal. Pode-se notar as faixas em que os loadings são importantes em comparação com as diferenças nos espectros na Figura 31. Esta separação é feita principalmente pelas absorções na faixa de variáveis 609 a 638 cm^{-1} (110 a 125).

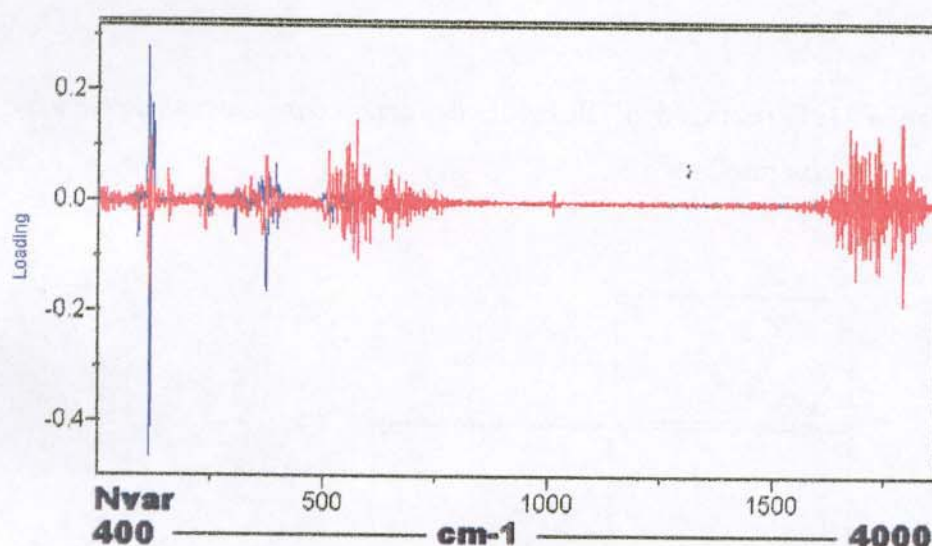


Figura 33: Gráfico dos loadings das PC versus número de onda: PC1 (azul) e PC2 (vermelho).

Na Figura 34, mostra-se o gráfico dos escores nas duas primeiras componentes principais e nota-se que existe uma separação parcial dos dois gêneros. A primeira componente principal PC1 descreve 37,4% da variância e a PC2 29,9%.

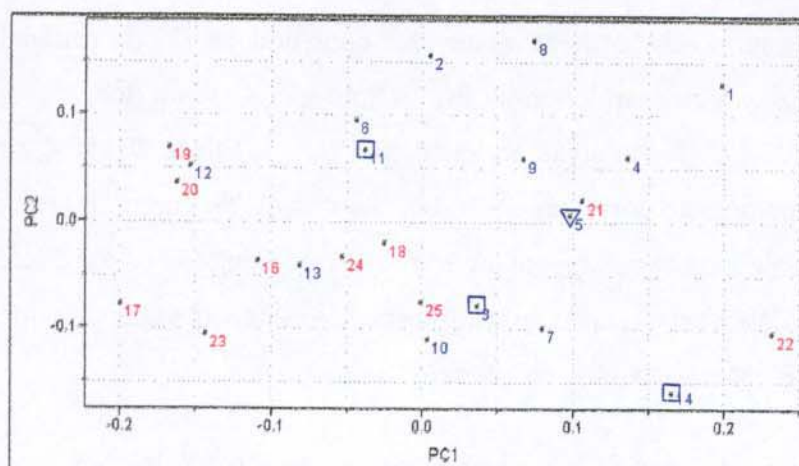


Figura 34: Gráfico dos escores das duas primeiras PC dos espectros de FTIR: *Oenocarpus* (azul): • *O. minor*, ▽ *O. bacaba*, □ *O. mapora*; *Jessenia* (vermelho).

Análise Hierárquica de Agrupamento com o método de conexão completa formou dois agrupamentos com níveis de similaridade entre 0 e 0,26.

Classificação mediante os espectros de FTIR.

A modelagem para a classificação com SIMCA e KNN foi realizada considerando as amostras do conjunto de treinamento. Ao serem tratadas, notou-se que a amostra 12 era mais próxima das *Jessenia* e a amostra 21 mais próxima das *Oenocarpus*. A redução das variáveis foi realizada mediante o critério de erro de previsão na classificação, permanecendo 80 variáveis, das 1868 iniciais. As faixas escolhidas dos espectros em concordância com os loadings foram: 440 a 455 cm^{-1} (Nvar 22 a 30), 565 a 580 (87 a 95), 609 a 627 (111 a 119), 1020 a 1033 (323 a 330), 1110 a 1126 (370 a 378), 1409 a 1425 (525 a 533), 1455 a 1471 (549 a 557), 1642 a 1658 (646 a 654), 2332 a 2348 (1004 a 1012) e 3421 a 3437 cm^{-1} (1569 a 1577). As bandas de absorção da ligação metal-oxigênio nos óxidos dos metais em estudo caem na faixa de 435 a 574 cm^{-1} e para os carbonatos as principais absorções estão 1085 a 1120 cm^{-1} ; 1410 a 1476 cm^{-1} ; enquanto que as absorções entre 1600 a 1640 cm^{-1} e 3200 a 3450 cm^{-1} corresponde à umidade que

poderiam ter pega as amostras de cinzas (altamente higroscópicas) ao serem manipuladas para o análise⁸⁷. O modelo *Oenocarpus* para variáveis de FTIR nas cinzas ficou definido por uma hipercaixa num espaço de 80 dimensões contendo 58,7% da variância ao longo da PC1 e 17,3% da variância ao longo da PC2. O modelo *Jessenia* ficou definido por um hipercilindro com 77,2% da variância ao longo da PC1. A Tabela 20 mostra os modelos de SIMCA que classificaram corretamente as cinco amostras do conjunto teste e o modelo de KNN que classificou corretamente quatro e errou uma (18). As diferenças entre gêneros encontraram-se em certas regiões do espectro, sinais devidas principalmente aos carbonatos dos diferentes elementos encontrados nas cinzas.

Tabela 20: Resultados da classificação através de SIMCA e KNN das amostras do conjunto teste dos gêneros *Oenocarpus* e *Jessenia* usando espectros de FTIR.

SIMCA		
Gênero	Número de amostras	Amostras classificadas incorretamente
<i>Oenocarpus</i>	3	0
<i>Jessenia</i>	2	0
KNN		
Gênero	Número de amostras	Amostras classificadas incorretamente
<i>Oenocarpus</i>	3	0
<i>Jessenia</i>	2	1 ^a

^a amostra (18)

Também para avaliar as diferenças entre os dois gêneros foram escolhidas as absorções a dez números de onda em cm^{-1} : 447 (26), 573 (91), 619 (115), 1026 (326), 1028 (327), 1119 (371), 1418 (529), 1464 (553), 2342 (1008) e 3431 (1573). Estas absorções correspondem às maiores respostas nos espectros. Foi aplicado SIMCA resultando na obtenção do modelo *Oenocarpus* com quatro componentes principais e o modelo *Jessenia* com duas PC. Estes modelos fizeram as previsões corretas de 4 das 5 amostra do conjunto teste, errando a amostra (18).

A Espectroscopia no infravermelho médio foi empregada para análise qualitativa de café verde⁸⁸; da mesma forma esta técnica foi aplicada para discriminar variedades de chá²⁰.

CAPÍTULO V.-RESULTADOS E DISCUSSÃO DE MODELOS DE CALIBRAÇÃO MULTIVARIADA USANDO ESPECTROS DE FTIR PARA DETERMINAR TEORES DOS ELEMENTOS OBTIDOS POR EDXRF

Os espectros foram obtidos das amostras de cinzas dos gêneros *Oenocarpus* e *Jessenia*. As variáveis independentes são as medidas das absorvâncias no infravermelho médio entre 4000 a 400 cm^{-1} em intervalos de 1,93 cm^{-1} , resultando em 1868 variáveis. As variáveis dependentes foram os teores de P, K, Ca, Mn, Fe, Zn e Cu obtidos através da técnica analítica de EDXRF.

A Figura 31 mostra os espectros de FTIR para as 24 amostras. Foram testados vários métodos de pré-tratamento nos dados. Os espectros foram normalizados. Foram calculadas as derivadas de 2ª ordem com a intenção de remover variação sistemática não requerida, tais como, efeitos da preparação da pastilha KBr-cinza. Várias tentativas de redução de variáveis foram feitas. Também os espectros foram testados com a transformação chamada “Correção de espalhamento multiplicativo” (MSC) que é usado para compensar os efeitos aditivos e multiplicativos de espalhamento nos dados espectrais. Mas esta transformação não se mostrou útil no desenvolvimento de modelagem para calibração multivariada.

Os dados foram analisados aplicando regressão de componentes principais e regressão de mínimos quadrados parciais. Na construção dos modelos foram usados os espectros de 19 amostras do conjunto de calibração e cinco amostras para previsão do conjunto de teste.

No primeiro estagio dos cálculos todas as 1868 variáveis independentes foram utilizadas assim como as sete variáveis dependentes. Na Tabela A.9, no apêndice, estão mostrados os dados dos teores dos elementos químicos que formam o conjunto de calibração e do conjunto teste. As variáveis foram centradas na média, as amostras normalizadas e as derivadas de 2ª ordem aplicadas. Antes de fazer a modelagem foi necessário escolher o número de componentes principais a ser usado no modelo, tanto

para PCR quanto para PLS. A escolha correta é necessária para ter modelos com precisão adequada para fazer previsões confiáveis.

As Figuras 35 a 38 mostram a maneira de escolha do número das componentes principais para P e K. A mesma metodologia foi aplicada para os outros elementos, Ca, Mn, Fe, Zn e Cu e para todos os modelos de PCR e PLS.

A Figura 35 mostra o gráfico dos valores de SEV versus o número de componentes principais (PC) para o modelo de P. Observamos que o menor valor de SEV corresponde a duas PC portanto foram escolhidas este número de PC para este modelo. O comportamento da curva observada é um pouco atípica mas pode acontecer dependendo da estrutura dos dados analisados. Isto causa as vezes problemas na escolha do número de PC para o modelo. Uma análise dos resultados como função do número de PC está relatado mais adiante. A Figura 36 apresenta o gráfico do resíduo de Student versus leverage. Tem-se três amostras (1, 8 e 14) com altas distâncias ao centróide do conjunto de dados. Estas amostras influenciam muito no modelo porque estão nos extremos superiores dos teores de P e dos escores em relação às outras amostras do conjunto de calibração. Por isso, tentam puxar a linha da regressão em sua direção. A amostra 13, embora tenha o maior teor de P, não influencia no modelo porque tem valores dos escores mais próximo aos escores médios das amostras.

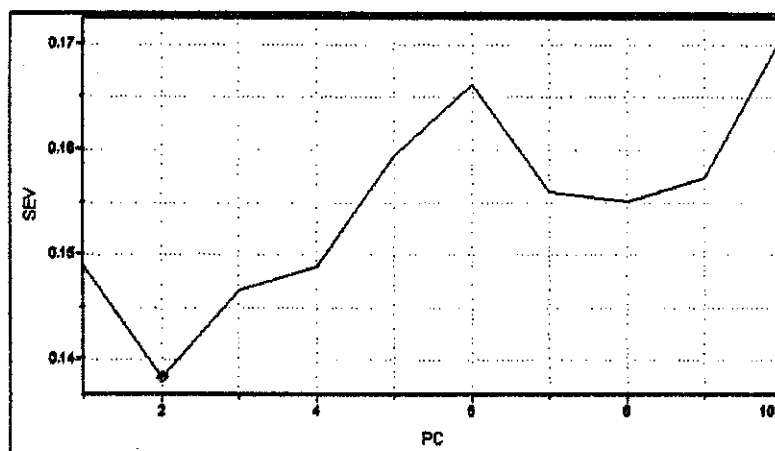


Figura 35: Gráfico do erro de validação (SEV) versus número de componentes principais (PC) para o teor de P.

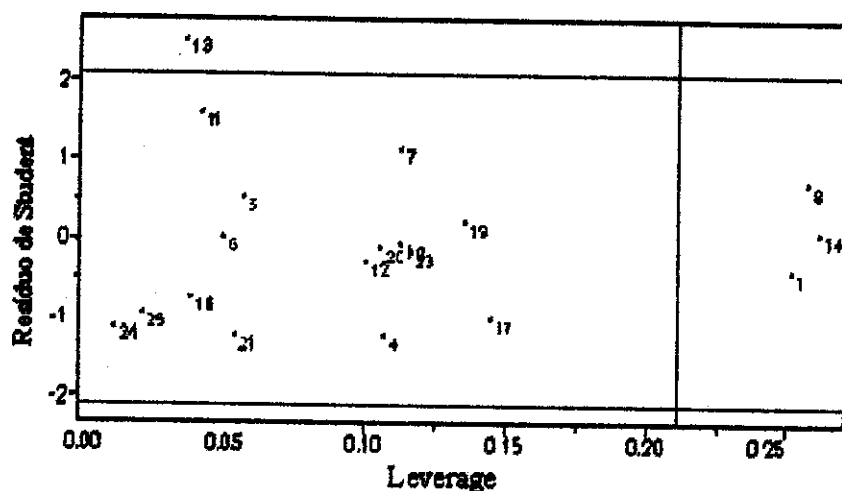


Figura 36: Gráfico do resíduo de Student versus leverage das amostras para o teor de P.

Ao analisar a Figura 37 que mostra o gráfico dos valores de SEV versus o número de PC para o K, nota-se que o menor valor de SEV corresponde a oito PC. Por isso foi escolhido este número de PC para o modelo de K, apesar que a razão número de amostras dividido por número de componentes fica perigosamente baixa. Por outro lado, na Figura 38 as amostras 12, 25 e 8 têm distâncias um pouco altas ao centróide do conjunto de dados. Estas amostras influenciam no modelo em menor proporção do que as observadas para o modelo do teor de P. Outra vantagem vista é que não existem amostras com alto residual mas precisamos lembrar que o número de graus de liberdade dos resíduos é baixo. Nota-se este exemplo também é atípico porque é o número máximo de PC escolhido por qualquer modelo investigado nesta tese. Quase sempre números bem menores de PC foram escolhidos, tipicamente 2 a 4.

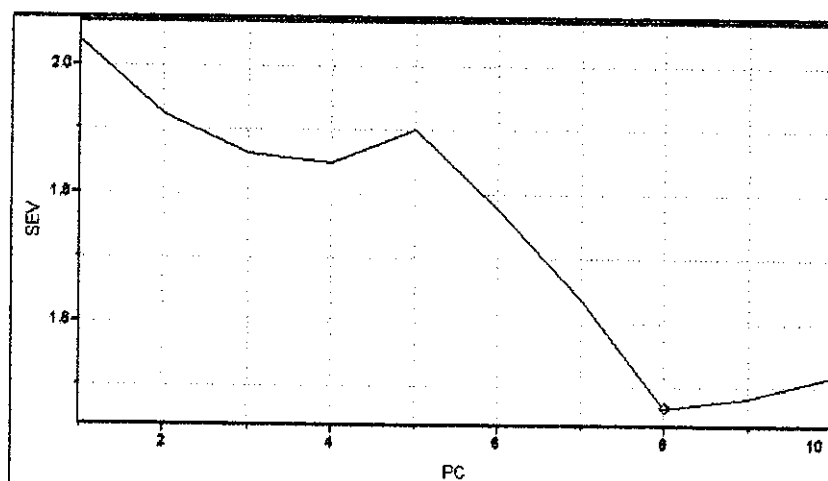


Figura 37: Gráfico do erro de validação (SEV) versus número de componentes principais (PC) para o teor de K.

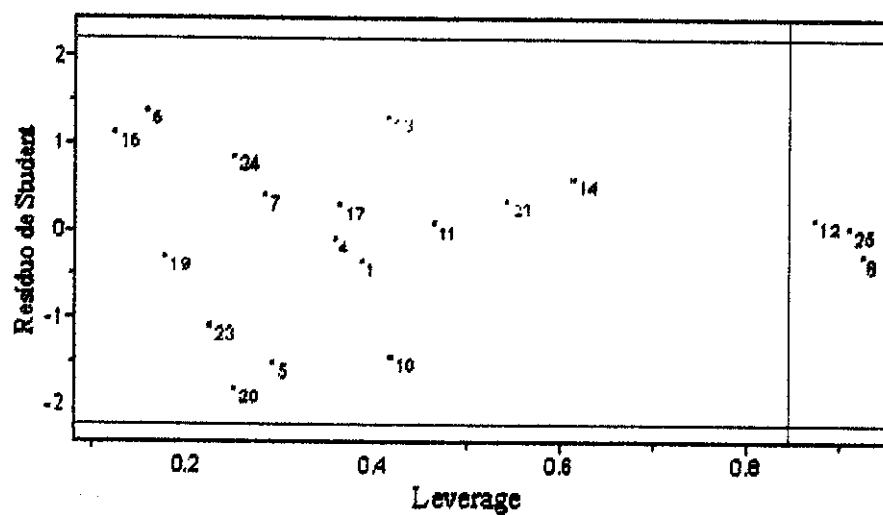


Figura 38: Gráfico do resíduo de Student versus leverage das amostras para o teor de K.

Depois de escolhida as componentes principais para cada modelo, fez-se as previsões necessárias.

A Tabela 21 mostra os valores experimentais (valores EDXRF) para os teores de elementos das cinco amostras do conjunto teste com fins de referencia. A Tabela 22 mostra os resultados das previsões dos teores dos elementos nas amostras do conjunto teste, bem como os valores do erro de validação (SEV, validação interna através de validação cruzada "Full leave one out"), soma de quadrados dos erros de previsão na validação (PRESS), valor do coeficiente de correlação (r), valores dos erros padrão de previsão (SEP), o número de componentes principais para cada modelo (PC) e a percentagem da variância (%Var) contida nas PC de cada modelo, no emprego de PCR.

Tabela 21: Teores de elementos determinados pela técnica analítica EDXRF (%) para o conjunto teste

Amostra\yi	P	K	Ca	Mn	Fe	Zn	Cu
2	0.197	5.504	0.636	0.008	0.019	0.006	0.003
3	0.350	6.727	1.041	0.015	0.026	0.023	0.006
9	0.372	5.863	0.470	0.014	0.02	0.008	0.004
18	0.283	5.524	1.291	0.042	0.033	0.016	0.003
22	0.000	4.220	1.275	0.075	0.036	0.029	0.003

Tabela 22: Teores de elementos do conjunto teste previstos usando os modelos PCR (%).

Amostra\yi	P	K	Ca	Mn	Fe	Zn*	Cu
2	0.192	5.190	0.727	0.014	0.016	0.055	0.005
3	0.327	6.146	1.162	0.087	0.036	0.055	0.007
9	0.288	6.886	0.662	0.032	0.021	0.051	0.006
18	0.251	5.712	1.276	0.073	0.034	0.060	0.007
22	0.424	7.463	1.448	0.127	0.043	0.042	0.009
SEV	0.138	1.466	0.667	0.050	0.011	0.145	0.002
PRESS	0.481	26.652	5.654	0.030	0.002	0.341	9.000E-05
r	0.57	0.89	0.59	0.71	0.69	0.073	0.78
SEP	0.194	1.552	0.134	0.043	0.006	0.038	0.0034
PC	2	8	3	4	3	1	6
%Var	59.0	93.0	70.0	77.0	70.0	39	87

* tem outliers.

A Tabela 23 mostra os resultados da previsão das amostras do conjunto teste, os valores de SEV, PRESS, r , os valores de SEP, o número dos PC e %Var contida para cada modelo, empregando PLS.

Tabela 23: Teores de elementos do conjunto teste previstos usando regressão PLS (%).

Amostra \ $\hat{y}_i$	P	K	Ca	Mn	Fe	Zn	Cu
2	0.226	5.166	0.665	0.010	0.013	0.093	0.005
3	0.303	6.144	1.188	0.094	0.038	0.043	0.006
9	0.321	6.735	0.689	0.033	0.022	0.039	0.006
18	0.234	5.577	1.403	0.078	0.035	0.088	0.007
22	0.403	6.897	1.841	0.143	0.048	0.042	0.009
SEV	0.140	1.610	0.700	0.051	0.012	0.162	0.002
PRESS	0.427	22.648	3.838	0.030	0.002	0.283	8.000E-05
r	0.64	0.89	0.77	0.77	0.76	0.45	0.84
SEP	0.185	1.295	0.284	0.050	0.008	0.053	0.003
PC	1	3	2	2	2	1	3
%Var	30.0	62.0	46.0	31.0	46.0	11.0	54.0

De acordo aos resultados observados nas Tabelas 22 e 23, nota-se que os modelos de PLS normalmente usam menos componentes principais que os modelos de PCR. Isto acontece porque PCR usa os escores de PCA obtido somente da matriz de dados das variáveis independentes X . Estes valores são usados na regressão com os valores de Y . O PLS também decompõe a matriz X para a obtenção dos escores de PCA, mas, com a variação que durante esta etapa de decomposição, passa-se informação extraída da matriz das variáveis dependentes Y e vice versa. Estes valores dos escores das matrizes X e Y são usados na regressão para obter valores de Y com maior exatidão. Em outras palavras as variáveis de PLS contem informações sobre as correlações entre os dados das matrizes X e Y ou correlação entre os dados da matriz X e vetor y (como em PLS1) e a variância da matriz X . Também observa-se que todos os valores de SEV de PLS são levemente maiores que os valores de SEV de PCR mas todos os valores de SEP de PLS são um pouco menores do que estes valores obtidos com PCR.

Em forma complementar as Figuras 39 a 46 mostram os gráficos dos valores experimentais versus os valores previstos dos teores de elementos químicos, para as amostras de calibração e do conjunto teste empregando modelos obtidos por PLS.

A Figura 39 mostra o gráfico dos valores dos teores de P experimentais versus os valores previstos. Nota-se que os pontos das amostras do conjunto de calibração têm valores de 0,075% a 0,602% de P para teores experimentais (EDXRF) e valores de 0,122% a 0,424%P para teores previstos. A maioria dos pontos pertencentes às amostras do conjunto teste têm valores entre 0,197% a 0,372% de P para teores experimentais e de 0,226% a 0,321% para teores previstos. Uma das amostras do conjunto de calibração, por ter o maior teor de P (0,602%), está um pouco afastada da linha de regressão. Somente o ponto pertencente à amostra 22 deste conjunto teste está mais afastada das demais amostras, pois seus valores são 0% de P valor obtido de EDXRF e 0,403% de P valor previsto. A resposta experimental de 0% para esta amostra foi possivelmente devido ao erro do detector do equipamento de EDXRF mas uma segunda determinação deu um resultado quase igual. Outra razão para que a amostra 22 esteja tão fora é que não houve uma amostra com concentração próxima a zero no modelo de calibração e por isso esta amostra caiu fora da faixa de linearidade do modelo e foi prevista com muito erro. Este ponto faz uma contribuição grande ao valor do SEP, 0,185, que é maior o valor de SEV, 0,140 (ver Tabela 23). Nota-se que o resto dos pontos correspondendo às amostras do conjunto teste têm valores previstos em concordância com os valores EDXRF.

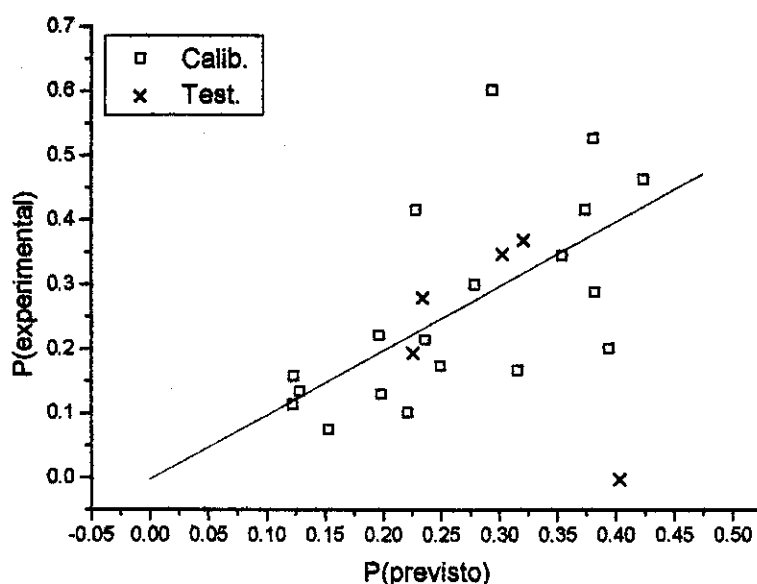


Figura 39: Gráfico dos valores experimentais versus valores previstos de teores de P das amostras do conjunto de calibração ($\square$) e conjunto teste ($\times$) usando PLS.

A Figura 40 mostra o gráfico dos valores de teores de K experimentais versus os valores previstos. Neste gráfico os pontos pertencentes às amostras do conjunto de calibração estão entre os valores de 1,627% a 7,925% de K no eixo de teores experimentais e entre os valores de 1,497% a 7,6745%K no eixo de teores previstos. Enquanto, que os pontos pertencentes às amostras do conjunto teste têm valores entre 4,22% a 6,727% de K para teores experimentais, e valores entre 5,166% a 6,897% de K no eixo de teores previstos, com um SEP de 1,295%. Novamente a amostra 22 deste conjunto teste está mais afastada das demais amostras porque foi prevista sobreestimada no teor de K.

Também nesta figura podemos observar o espalhamento dos pontos ao longo da linha de resultados exatos é menor do que aquele observado para P. Apesar que o valor de SEV (1,610) é maior do que seu valor análogo para P precisamos lembrar que a variação dos valores EDXRF para K é muito maior do que para P. Também o valor do SEP (1,295) de K é maior que o valor do SEP (0,185) de P.

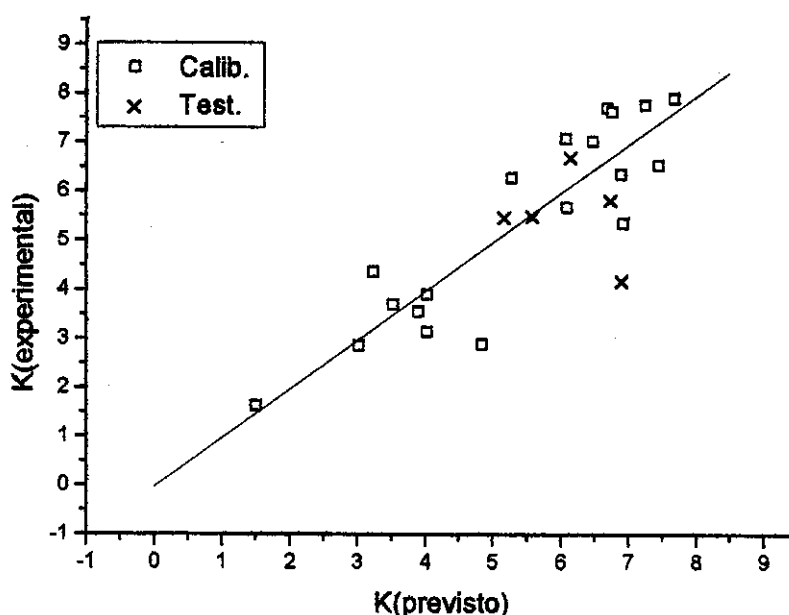


Figura 40: Gráfico dos valores experimentais versus valores previstos de teores de K das amostras do conjunto de calibração (□) e conjunto teste (x) usando PLS.

As Figuras 41 a 46 analisadas em forma semelhante às Figuras 39 e 40 nos permitiram notar que os melhores modelos de calibração multivariada foram os de K e Cu; seguido na ordem descendente os modelos de Mn, Fe, Ca, P e finalmente de Zn. O modelo de Zn (Figura 44) construído com as amostras 8 e 16 que têm teores de Zn maiores do que aqueles do resto do grupo apresentam esse espalhamento no espaço ($r=0,45$). Estas amostras apresentaram altas distâncias ao centróide do conjunto de dados (leverage) e/ou altos residuais, que influenciam muito no modelo e podem ser considerados como outliers. Com fins comparativos o modelo de Zn (Figura 45) foi determinado sem considerar as amostras 8 e 16, o que mostra que sim houve melhora no modelo de Zn ($r=0,90$).

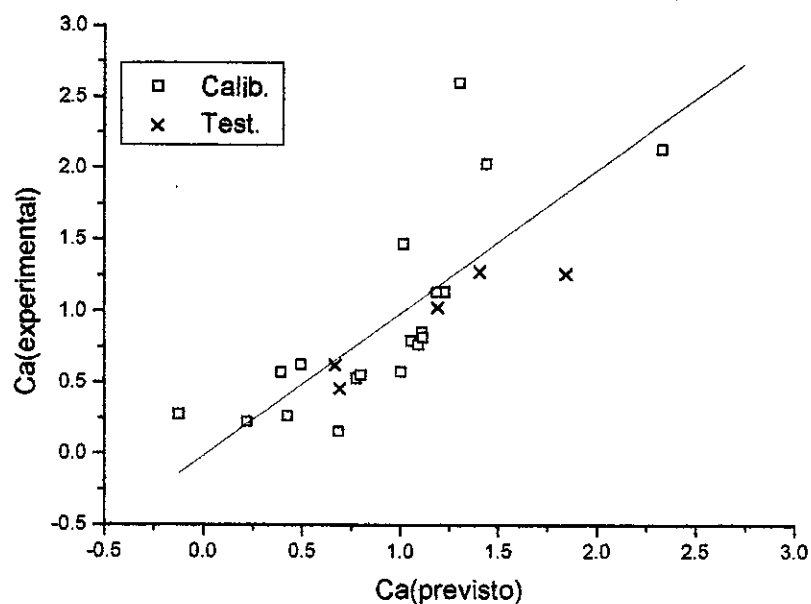


Figura 41: Gráfico dos valores experimentais versus valores previstos de teores de Ca das amostras do conjunto de calibração (□) e conjunto teste (x) usando PLS.

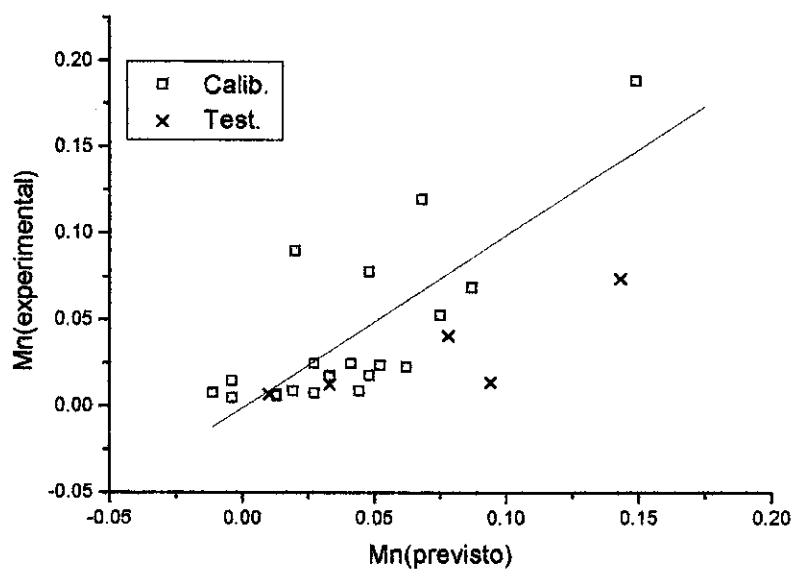


Figura 42: Gráfico dos valores experimentais versus valores previstos de teores de Mn das amostras do conjunto de calibração (□) e conjunto teste (x) usando PLS.

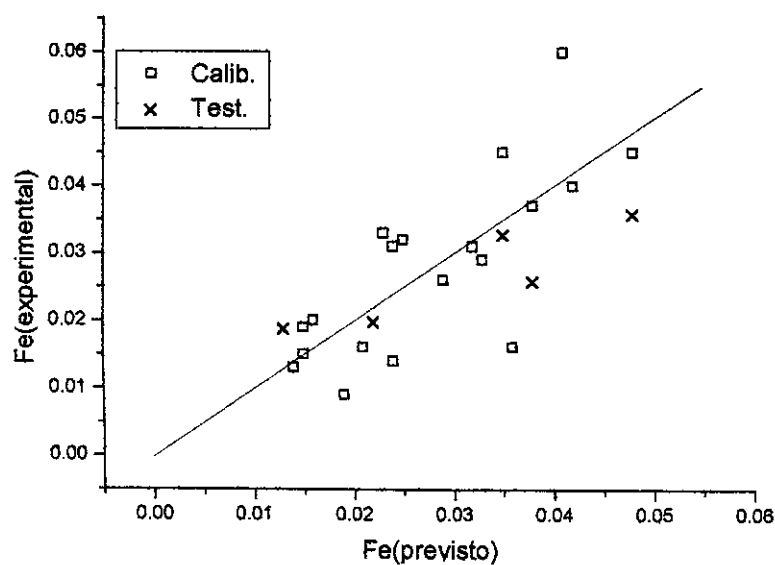


Figura 43: Gráfico dos valores experimentais versus valores previstos de teores de Fe das amostras do conjunto de calibração (□) e conjunto teste (x) usando PLS.

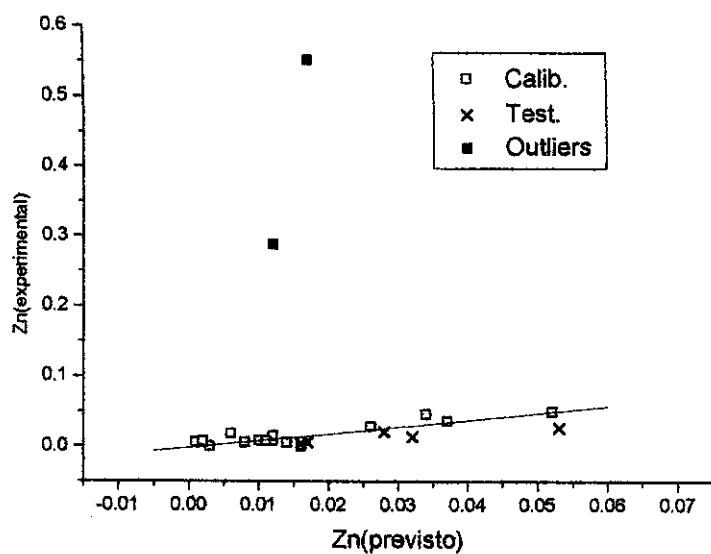


Figura 44: Gráfico dos valores experimentais versus valores previstos de teores de Zn das amostras do conjunto de calibração (□) e conjunto teste (x) usando PLS.

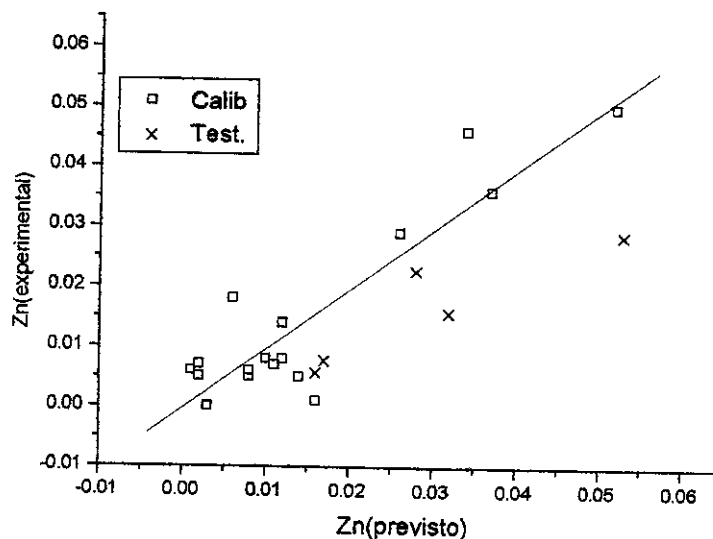


Figura 45: Gráfico dos valores experimentais versus valores previstos de teores de Zn das amostras do conjunto de calibração (□), conjunto teste (x). Nota-se que os outliers do gráfico na Figura 44 (amostras 8 e 16) foram eliminados na determinação deste modelo de calibração.

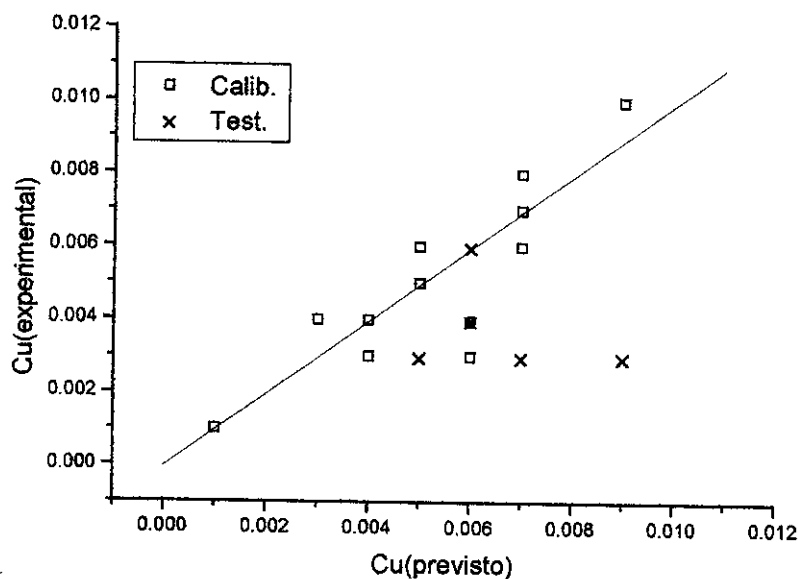


Figura 46: Gráfico dos valores experimentais versus valores previstos de teores de Cu das amostras do conjunto de calibração (□) e conjunto teste (x) usando PLS.

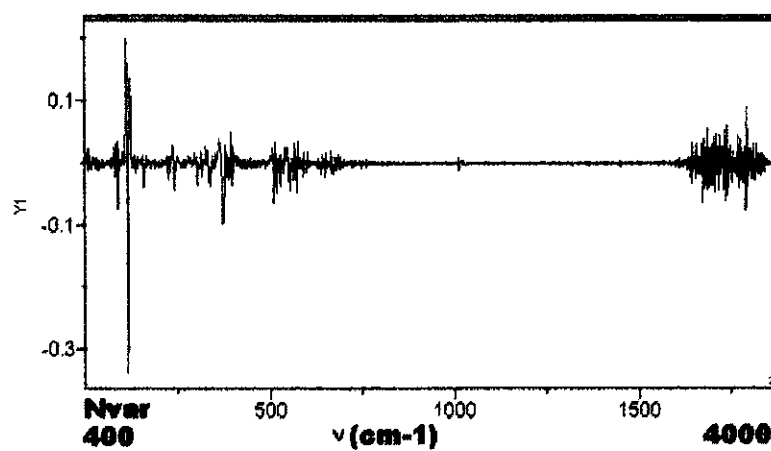
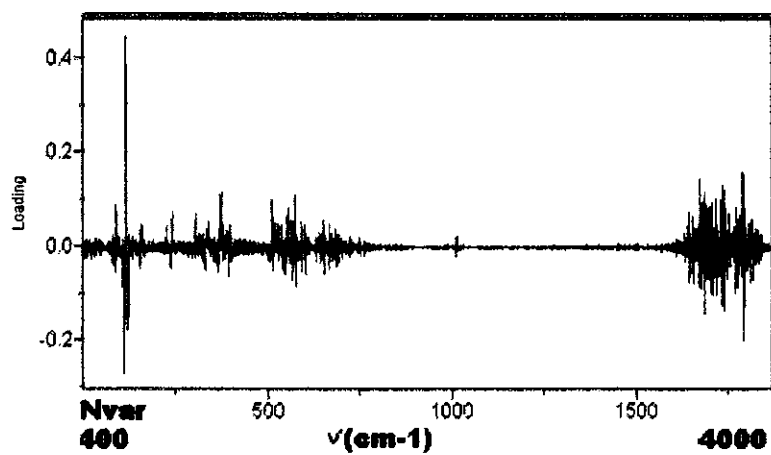
A próxima etapa da calibração teve a finalidade de reduzir as variáveis independentes com a intenção de melhorar os modelos de calibração. Examinando os espectros no IR na Figura 31 observamos que tem varias regiões espectrais onde não existem absorções significativas e por isso não contem informação útil para a calibração. Foram analisados os gráficos dos loadings e do vetor de regressão nos modelos para ver melhor as regiões espectrais importantes na calibração. Na tabela a seguir está resumido as condições necessárias para excluir regiões espectrais do análise.

<i>Caso</i>	<i>Loadings</i>	<i>Vetor Regressão</i>	<i>Excluir</i>
1	Importante	Importante	Não
2	Não importante	Importante	Não
3	Importante	Não importante	Não
4	Não importante	Não importante	Sim

Para eliminar uma variável temos que observar que o valor tanto do seu loading quanto do seu coeficiente no vetor regressão não devem ser importantes.

As Figuras 47 e 48 mostram os gráficos dos loadings e do vetor de regressão para os modelos de calibração dos teores de elementos de P e K.

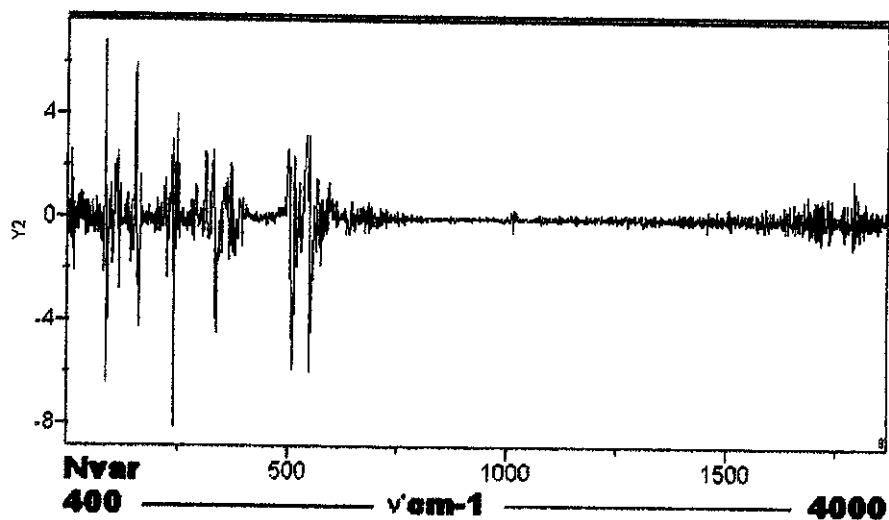
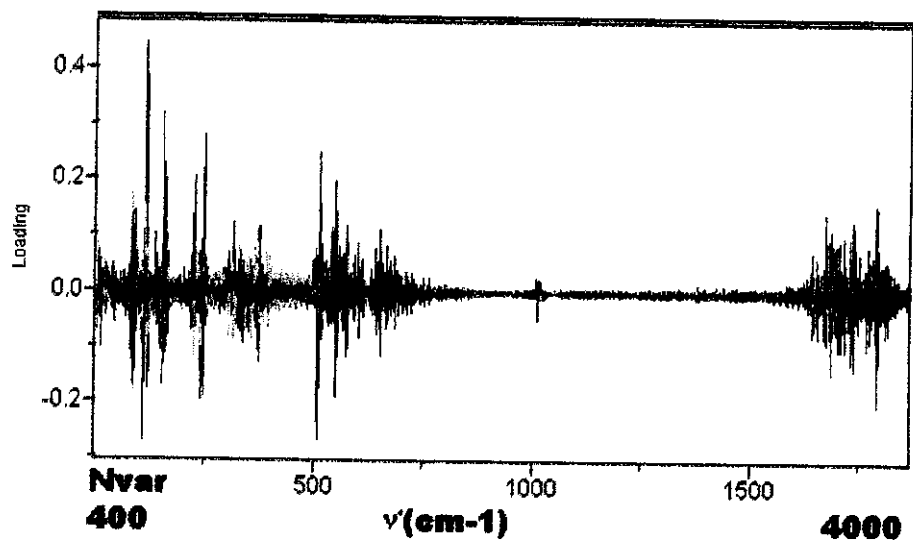
a)



b)

Figura 47: Gráfico dos (a) loadings e (b) dos coeficientes do vetor de regressão versus número de onda para o modelo do teor de P.

a)



b)

Figura 48: Gráfico dos (a) loadings e (b) dos coeficientes do vetor de regressão versus número de onda para o modelo do teor de K.

A Figura 47 na parte (a) mostra o gráfico dos loadings para duas componentes principais versus o número de variáveis (Nvar) de 1 a 1868 equivalente o intervalo espectral entre 4000 a 400 cm^{-1} . Nesta mesma figura parte (b) vemos o gráfico dos coeficientes do vetor de regressão para duas componentes principais versus o número de onda na faixa de 4000 a 400 cm^{-1} . Aqui notamos as faixas em que o caso quatro é obedecido para este modelo. Da mesma forma na Figura 48 vê-se estas mesmas faixas. Em geral estas faixas observadas são semelhantes nos outros elementos em estudo o que permite fazer a redução do número das variáveis originais, permanecendo as faixas seguintes para todos os elementos analisados: 400 a 1227 cm^{-1} (Nvar 1 a 430), 1364 a 1767 cm^{-1} (501 a 710), 3466 a 3888 cm^{-1} (1591 a 1810).

Desta redução ficaram 860 variáveis das 1868 iniciais. Os resultados das previsões dos teores de elementos das amostras do conjunto teste obtidos através dos modelos de PCR e PLS mostram-se nas Tabelas 24 e 25.

Tabela 24: Teores de elementos previstos usando os modelos PCR (%) com 860 variáveis independentes.

Amostra\yi	P	K	Ca	Mn	Fe	Zn	Cu
2	0.188	5.240	0.747	0.010	0.016	0.056	0.005
3	0.333	6.183	1.170	0.076	0.037	0.054	0.007
9	0.294	7.069	0.666	0.026	0.022	0.049	0.006
18	0.247	5.663	1.269	0.065	0.034	0.060	0.007
22	0.410	7.265	1.404	0.110	0.042	0.041	0.009
SEV	0.136	1.432	0.659	0.050	0.011	0.141	0.002
PRESS	0.460	24.890	5.776	0.030	0.002	0.340	3.000E-05
r	0.59	0.90	0.57	0.65	0.69	0.08	0.78
SEP	0.188	1.491	0.130	0.034	0.006	0.038	0.003
PC	2	8	3	3	3	1	6
%Var	60.0	93.0	71.0	71.0	71.0	40.0	88.0

Segundo os resultados observados nas Tabelas 24 e 25, nota-se que os modelos ficaram bem iguais aos determinados com 1868 variáveis. As únicas diferenças significativas podemos ver no modelo PCR de Mn que tem um CP menos com 860 variáveis do que com 1868, e no modelo PLS (860) de Ca que tem 1 variável latente a

menos do que PLS (1868). Também observa-se que os valores previstos dos teores de elementos para as amostras do conjunto teste são quase iguais usando modelos com 860 ou 1868 variáveis. Nota-se que a variância explicada com modelo PLS de Ca com duas PC foi de 46% enquanto que com só um PC foi de 25%. Deve mencionar-se que a escolha do número de componentes principais para cada modelo foi realizada como já descrita no primeiro estágio. Os melhores modelos de calibração multivariada foram os de K e Cu; seguido na ordem descendente os modelos de Fe, Mn, P, Ca e de Zn.

Tabela 25: Teores de elementos previstos usando os modelos PLS (%) com 860 variáveis independentes.

Amostra \ $\hat{y}_i$	P	K	Ca	Mn	Fe	Zn	Cu
2	0.222	5.062	0.764	0.008	0.013	0.090	0.005
3	0.308	6.235	0.939	0.094	0.039	0.044	0.007
9	0.327	6.830	0.614	0.033	0.022	0.037	0.006
18	0.225	5.411	1.157	0.073	0.034	0.091	0.006
22	0.379	6.579	0.845	0.129	0.046	0.055	0.008
SEV	0.140	1.541	0.711	0.051	0.012	0.162	0.002
PRESS	0.399	19.979	5.961	0.028	0.002	0.286	2.000E-05
r	0.66	0.91	0.57	0.76	0.77	0.43	0.83
SEP	0.174	1.179	0.224	0.046	0.008	0.054	0.003
PCs	1	3	1	2	2	1	3
%Var	31.0	63.0	25.0	33.0	46.0	11.0	55.0

Podemos dizer que o número ótimo de componentes principais pode ser escolhido, baseado no valor de SEV, do PRESS ou na curva de variância residual, no entretanto, a observação visual destas quantidades em função do número de PC para usar na calibração permitem uma melhor escolha do número de componentes principais. Mas nem sempre estes critérios indicam o mesmo número de componentes principais para incluir no modelo de calibração. Em resumo, o problema de escolher o número correto de componentes para ser incluídos em modelos de calibração não foi resolvido em geral. Isto pode causar erros nos modelos de calibração.

Também foram realizadas calibrações com os números de onda correspondentes as dez maiores absorções ou, seja, 447, 573, 619, 1026, 1028, 1119, 1418, 1464, 2341 e 3431 cm^{-1} , correspondentes aos números de variáveis de 26, 91, 115, 326, 327, 374, 529,

553, 1008 e 1573 respectivamente. Os resultados das previsões dos teores dos elementos do conjunto teste obtidos com os métodos de PCR e PLS mostraram que não foram melhores aos obtidos nas duas primeiras etapas. Da mesma forma foram testados os dados para calibração com PCR e PLS depois de que as variáveis foram centradas na media e os espectros de FTIR corrigidos através da transformação MSC. Os resultados das previsões dos teores dos elementos do conjunto teste não mostraram melhora.

Uma análise geral dos resultados acima mostrados sugere que a proposta desenvolvida permite somente a determinação aproximada dos teores dos elementos nas cinzas das plantas em estudo.

Quando os dados com tendências não lineares são tratados por métodos analíticos lineares podem ocorrer problemas, embora os métodos quimiométricos de PCR e PLS permitem as vezes a resolução de não linearidades encontradas na calibração de espectros, bem como problemas de sobreposição de linhas espectrais multielementares⁸⁸.

Adicionando mais componentes no análise em nosso caso não resolve totalmente este problema. A Tabela 26 mostra os valores das previsões dos teores de P para modelos com uma, duas e três variáveis latentes. Nota-se que quando aumenta o número de variáveis latentes do modelo além do ótimo, o valor do SEV também aumenta da mesma forma que o valor do coeficiente de correlação. Entretanto os valores do SEP e do PRESS diminuíram a medida que aumenta o número de PC ao contrário do esperado. Isto é atípico e foi devido à resposta observada da amostra 22 do conjunto de teste.

Tabela 26: Teores de elementos do P do conjunto teste previstos com modelos de PLS para 1 PC, 2PC e 3PC.

Amostra	Pexp	PC1	PC2	PC3
2	0.197	0.222	0.140	0.205
3	0.350	0.308	0.298	0.226
9	0.372	0.327	0.308	0.372
18	0.283	0.225	0.202	0.185
22	0.000	0.379	0.306	0.226
SEV		0.140	0.151	0.162
PRESS		0.399	0.282	0.179
r		0.660	0.770	0.870
SEP		0.174	0.148	0.123
%Var		31,0	48,2	61,5

As Figuras 49 a 51 mostram os gráficos dos valores experimentais versus os valores previstos dos teores de P, para as amostras de calibração e do conjunto teste empregando modelos obtidos por PLS com 860 variáveis independentes e com uma, duas e três componentes.

A Figura 49 mostra o gráfico dos valores dos teores experimentais de P versus os valores previstos para o modelo com uma componente principal. Nota-se que os pontos das amostras do conjunto de calibração e maioria dos pontos pertencentes às amostras do conjunto teste estão espalhados similarmente ao observado na Figura 39. Somente o ponto pertencente à amostra 22 deste conjunto teste está afastada das demais amostras e tem 0,379% de P como o valor previsto.

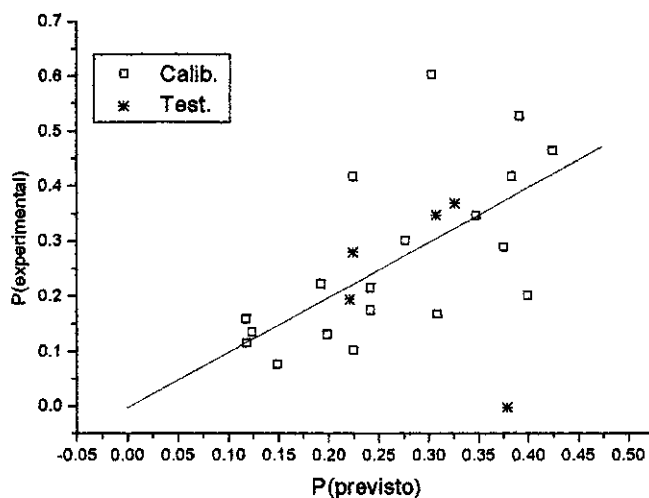


Figura 49: Gráfico dos valores experimentais versus valores previstos de teores de P das amostras do conjunto de calibração (□) e conjunto teste (*) usando PLS com 860 variáveis e uma (01) componente principal.

A Figura 50 mostra o gráfico dos valores de teores experimentais de P versus os valores previstos para o modelo de duas componentes principais. Neste gráfico os pontos pertencentes às amostras do conjunto de calibração e os pontos das amostras do conjunto teste estão espalhados ao longo da linha de valores exatos e menos afastados dela. Apesar que o coeficiente de correlação aumenta de 0,66 para 0,77 a concordância entre os resultados previstos e experimentais melhora pouco.

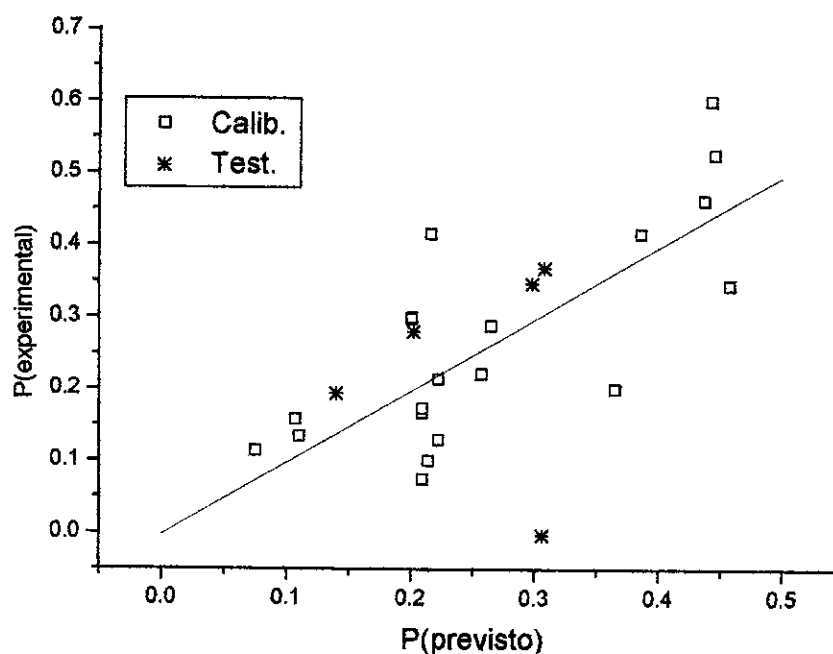
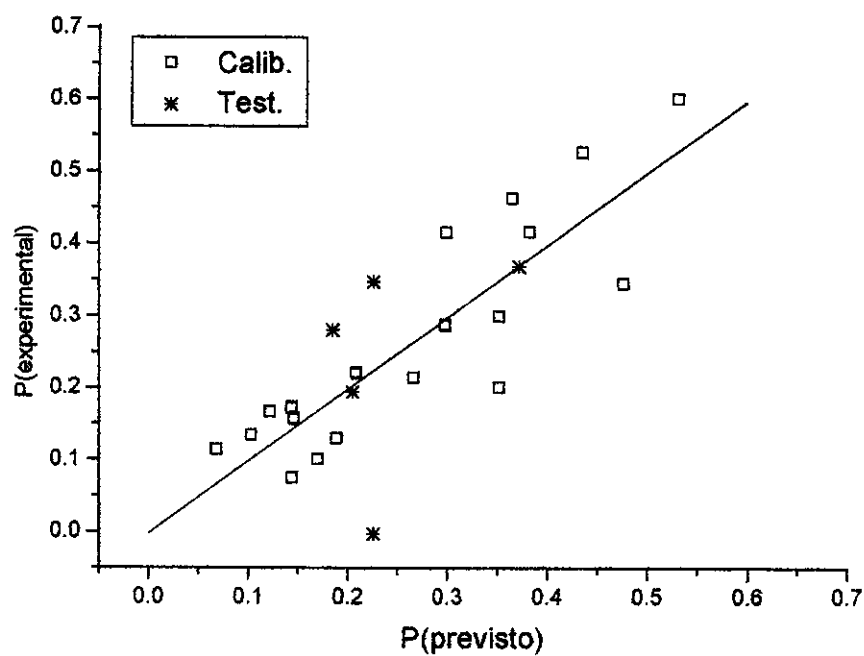


Figura 50: Gráfico dos valores experimentais versus valores previstos de teores de P das amostras do conjunto de calibração (□) e conjunto teste (*) usando PLS com 860 variáveis e duas (02) componentes principais.

A Figura 51 mostra o gráfico dos valores de teores experimentais de P versus os valores previstos para o modelo de três componentes principais. Neste gráfico o espalhamento dos pontos pertencentes às amostras do conjunto de calibração e do conjunto teste não está tão grande como aqueles nas Figuras 49 e 50. O aumento no coeficiente de correlação de 0,66 para 0,87 mostra uma melhora na concordância entre os resultados previstos e experimentais, comparativamente aos observados nas duas figuras anteriores.

As vezes a adição de PC para melhorar os modelos de calibração pode criar problemas quando são adicionados mais do que o número necessário de componentes porque podemos simplesmente estar adicionando ruído ao sinal.



CAPÍTULO VI

CONCLUSÕES

A seguir está apresentado um quadro mostrando as identidades das amostras que foram classificadas incorretamente para cada conjunto de dados investigados nesta tese.

Conjunto de dados	Amostras classificadas incorretamente
Morfológicos	-
Físico-químicos	-
Composição de ácidos graxos	18,22
Teor de elementos químicos	18
EDXRF	18
UV/Vis	18
Infravermelho	18
XRD	3

Os dados morfológicos referidos ao tamanho e dimensões dos frutos mostraram que a total discriminação feita nos gêneros estudados foram principalmente pelas variáveis PS, CS, DF e CF. Porém, com estes dados tanto a análise univariada quanto a análise multivariada podem ser usadas para classificação de *Oenocarpus* e *Jessenia*. Talvez com isto seja possível para classificar as espécies do gênero *Oenocarpus* como já observado nas amostras disponíveis das espécies *mapora* e *bacaba*.

Os dados físico-químicos, como mostrado acima, também classificaram corretamente todas as amostras do conjunto teste. Embora, os modelos dos gêneros *Oenocarpus* e *Jessenia* estiveram mais próximos um do outro no hiper-espço do que os modelos obtidos com dados morfológicos. Isto mostrou a maior complexidade dos dados físico-químicos em relação aos dados morfológicos. As variáveis mais importantes na discriminação foram %Umid e %Óleo.

No quadro anterior a amostra número 18 foi classificada errada quando foram usados dados de UV/Vis, infravermelho, teor de elementos químicos, dados de espectros de EDXRF e dados de composição de ácidos graxos. Isto pode ser explicado por razão do estado em que essa amostra foi encontrada, já que foi coletada de um árvore sofrendo com praga. Apesar que a amostra 18 não apresentava sinais de ataque com a praga, poderia ser que suas propriedades químicas foram afetadas.

Os dados de ultravioleta/visível mostraram que os flavonóides serviram para discriminar entre os gêneros *Oenocarpus* e *Jessenia*, o que está em concordância com os dados da literatura.

Com os espectros no infravermelho das cinzas foram possíveis obter modelos para os gêneros *Oenocarpus* e *Jessenia*, sendo as absorções decorrentes da ligação metal-oxigênio dos óxidos e do grupo carbonila C-O dos carbonatos as mais importantes (respostas no IR médio dos óxidos e sais inorgânicos dos metais em estudo 447, 573, 619, 1026, 1028, 1119, 1418, 1464 cm^{-1}).

As composições dos ácidos graxos não classificaram corretamente entre os gêneros *Oenocarpus* e *Jessenia*, sugerindo que estas composições podem ser empregadas para obter um modelo com fins de controle da qualidade dos óleos destes gêneros.

Os dados morfológicos neste trabalho resultaram em ser muito mais diferentes para os gêneros *Oenocarpus* e *Jessenia* do que qualquer dos conjuntos dos dados químicos, espectrais ou cromatográficos. Isto concorda com o Balick¹, quem encontrou que os frutos tanto de *Oenocarpus* quanto de *Jessenia*, tinham a mesma aparência e que os frutos, foram menores para as espécies do gênero *Oenocarpus* e de maior tamanho e peso para os frutos do gênero *Jessenia*. Também encontrou que na mesma população de *Jessenia bataua* em Pebas no Perú houve grande variação no tamanho e peso de duas coletas. Ao estudar os dois gêneros com UV encontrou que *Oenocarpus* e *Jessenia* compartilhavam 32% dos flavonóides o que optou por chamar-los complexo

Oenocarpus-Jessenia. Por outro lado, dados espectroscópicos de UV, IR, EDXRF e XRD, nesta tese, mostraram que verdadeiramente estes gêneros compartilham alguns fatores pelo qual a discriminação das amostras do conjunto teste não foi totalmente correto. Isto esteve em concordância com os estudos preliminares realizados por Balick.

Com relação aos resultados de calibração multivariada, os modelos dos teores dos elementos P, K, Ca, Mn, Fe, Zn e Cu obtidos serviram somente para fazer a previsão semiquantitativa dos valores dos teores de elementos nas cinzas das amostras do conjunto teste. Resultaram melhores modelos os de K e Cu; seguido na ordem descendente os modelos de Fe e Mn, e finalmente os de P, Ca e Zn.

A maior crítica circunstancial que poderia ser feita nesta tese, tanto para efeitos de classificação quanto para calibração, foi a utilização de um número limitado de amostras (25 amostras) nas diversas análises. Esta situação foi produto de uma série de fatores que complicaram a coleta. Estes fatores críticos, como época de frutificação e maturação, condições ambientais, facilidade de deslocamento aos locais de coleta, comportamento dos camponeses em relação ao aproveitamento dos frutos e, por último, a disponibilidade, foram os que dificultaram a coleta, pelo qual somente deu para obter o que estivera disponível. De fato o ideal seria coletar maior número de amostras na mesma época.

Perspectivas futuras

Desde o descobrimento da região espectral do Infravermelho Próximo por Frederick William Herschel em meados de 1800 até esta época, teve que passar muito tempo para que a técnica de análise no Infravermelho Próximo (NIR), tornasse uma das mais promissórias no campo das técnicas não destrutivas, podendo ser aplicada a diversos tipos de amostras, desde produtos agrícolas, polímeros, na indústria de fármacos⁸⁹. O uso de técnicas analíticas como fluorescência de raios-X e infravermelho médio²⁰ já foram usadas para classificar plantas como cafés ou chás, então fica a possibilidade de usar NIR para discriminar entre gêneros de plantas. Isto pode acontecer porque as bandas de

absorção na região NIR são sobretons ou combinação das bandas fundamentais. Os grupos químicos que existem nas plantas contêm as ligações C-H, O-H e N-H, que são os que apresentam ocorrências espectrais usadas no desenvolvimento de metodologias analíticas³⁰. Portanto, a espectroscopia no NIR pode permitir quantificar parâmetros químicos e físicos dos frutos dos gêneros *Oenocarpus* e *Jessenia* através de calibração multivariada.

O estilo da pesquisa seguida nesta tese abre a possibilidade de estudar extratos de plantas diversas e relacioná-las com a concentração de algum princípio ativo. Significa que a quimiometria, seja mediante reconhecimento de padrões ou calibração multivariada, poderia entrar na pesquisa de produtos naturais, desta maneira facilitando o trabalho no laboratório químico.

Referências bibliográficas

-
- ¹ Balick, M. *Advances in Economic Botany*; New York, U.S.A, 1986, 3.
- ² Prance, G. T. *Botânica Econômica de algumas espécies amazônicas*; INPA/FUA; Manaus, 1987.
- ³ Chaves, J.; Pechnick, M. *Pesquisa sobre a constituição química dos óleos de Pataú e Bacaba*; Instituto de Nutrição, 1941.
- ⁴ Serruya, H.; Bentes, M.H.S. *Anais CRQVI*, Belém-PA, 1985, 13.
- ⁵ Henderson, A.; Galeano, G.; Bernal, R. *Palms of the Americas*; Princeton University Press; New Jersey 08340, 1995.
- ⁶ Pesce, C. *Oleaginosas da Amazônia. Belém*; Revista da Veterinária, 1941.
- ⁷ FAO y CATTIE. *Palmeras poco utilizadas de América Tropical*; Informe de la Reunión de consulta organizada por FAO y CATTIE; San José Litografía e imprenta L.K. S.A, 1984.
- ⁸ Scarminio, I. S. Tese de Doutorado, Universidade Estadual de Campinas, 1989.
- ⁹ Sharaf, M. A.; Ilman, D. L.; Kowalski, B. R. *Chemometrics*; John Wiley & Sons; New York, 1986.
- ¹⁰ Batchelor, B.G. *Pattern Recognition*; Plenum Press, New York, 1978.
- ¹¹ Dupuy, N.; Duponchet, L.; Huvenne, J.P.; Sombret, B.; Legrand, P. *Food Chem.* **1996**, *57*, 245.
- ¹² González, Y.; Pérez, J.L.; Moreno, B.; García, C. *Anal. Chim. Acta* **1999**, *384*, 83.
- ¹³ Rebolo, S.; Peña, R. M.; Latorre, M.J.; García, S.; Botana, A. M.; Herrero, C. *Anal. Chim. Acta* **2000**, *417*, 211.
- ¹⁴ Latorre, M.J.; Peña, R.; Pita, C.; Botana, A.; Herrero, C. *Food Chem.* **1999**, *66*, 263.
- ¹⁵ Busse, H. J.; Denner, E. B.M.; Lubitz, W. *J. Biotechnol.* **1996**, *47*, 3.
- ¹⁶ Simonton, W.; Graham, D. *Appl. Eng. Agric.* **1996**, *12*, 89.
- ¹⁷ Valera, P.; Pablos, F.; González, A. G. *Talanta* **1996**, *43*, 415.
- ¹⁸ Martín, M. J.; Pablos, F.; González, A. G. *Anal. Chim. Acta* **1996**, *320*, 191.

-
- ¹⁹ Martín, M.J.; Pablos, F.; González, A. G. *Anal. Chim. Acta* **1998**, 358, 177.
- ²⁰ Budínová, G.; Vlácil, D.; Mestek, O.; Volka, K. *Talanta* **1998**, 47, 255.
- ²¹ Minjin, K.; Younh-Hak, L.; Chonghun, H. *Computers and Chemical Engineering* **2000**, 24, 513.
- ²² De Boever, J. I.; Cottyn, B. G.; Vanacker, J. M.; Boucqué, Ch. V. *Animal Feed Sci. Technol.* **1995**, 51, 243.
- ²³ Forina, M.; Casolino, M. C.; Pezuela M. C. *J. Pharm. Biomed. Anal.* **1998**, 18, 21.
- ²⁴ Parreira, T. F.; Ferreira, M. C.; Sales, H. J.; de Almeida, W. B. Livro de Resumos do 10º Encontro Nacional de Química Analítica, Santa Maria, Brasil, 1999, QM-9.
- ²⁵ Vonach, R.; Buschmaann, J.; Falkowski, R.; Schinder, R.; Lendl, B.; Kellner, R. *Appl. Spectrosc.* **1998**, 52, 820.
- ²⁶ Schmidt, F.; Bueno, M. I.; De Abreu, M. F. Livro de Resumos do 10º Encontro Nacional de Química Analítica, Santa Maria, Brasil, 1999, QM-3.
- ²⁷ Malmierca, M. S.; Seip, K. L.; Osen, K. K. *Anat. Embriol.* **1995**, 19, 343.
- ²⁸ Dubouzet, J. G.; Murata, N.; Shinoda, K. *Scientia Horticulturae* **1998**, 73, 37.
- ²⁹ Rozendaal, E. L. M.; Kurstjens, S. J. L.; Beek, T. A.; Van der Berg, R. G. *Phytochemistry* **1999**, 52, 427.
- ³⁰ Watson, C. A. Near-Infrared Reflectance Spectrophotometric Analysis of Agricultural Products, *Anal. Chem.* **1977**, 49, 835A.
- ³¹ Hamilton, R. J.; Rossell, J. B. *Analysis of Oils and Fats*. Elsevier, Elsevier Publishing Co INC; New York, 1987.
- ³² Tangerman A; Nagengast, FM., *Anal. Biochem.* **1996**, 236, 1.
- ³³ BorchJensen, C.; Mollerup, J. *American Chemical Society symposium series* **1997**, (670), 154.
- ³⁴ Brown, D. J.; Boyd, I. L.; Cripps, G. C.; Butler, P. J. *Mar. Ecol. Progr. Series* **1999**, 187, 251.
- ³⁵ Bailey, AL.; Southon S. *Anal. Chem.* **1998**, 70, 415.
- ³⁶ Kotani, A.; Fuse, T.; Kusu, F. *Anal. Biochem.* **2000**, 284, 65.

-
- ³⁷ Yamashiro, H.; Oku, H.; Higa, H. Chinen, I.; Sakai, K. *Compar. Biochem. Physiol. Part B: Biochem. Mol. Biol.* **1999**, 122, (4), 397.
- ³⁸ Attaelmannan, MA.; Dahl, AA.; Reid, RS. *Can. J. Aanimal Sci.* **1999**, 79, (3), 401.
- ³⁹ Adam, M.; Mossoba, MM.; Dawson, T.; Chew, M.; Wasserman, S. *J.Am. Oil Chem. Soc.* **1999**, 76, (3), 375.
- ⁴⁰ Ma, KM.; van der Voof, FR.; Sedman, J.; Ismail, A. *J. Am. Oil Chem. Soc.* **1999**, 76, (12), 139.
- ⁴¹ Cert, A.; Moreda, W.; Leon, C. M.; Perez, MC. *Grasas y Aceites* **1996**, 47, (6), 401.
- ⁴² Guiochon, G.; Pommier, C. *La chromatographie en phase gazeuse en chimie inorganique*; Gauthier-Villares Editeur; Paris, 1971.
- ⁴³ McNair, H. M. *Cromatografia de Gases*; Copyright by The General Secretariat of the Organization of American States; Washington D.C, 1981.
- ⁴⁴ Ambrose, D. *Gas Chromatography*; Butterworths; London, 1971.
- ⁴⁵ Collins, C. H.; Braga, G. L.; Bonato, P. S. *Introdução a Métodos Cromatográficos*; 4^a. ed. Editora Unicamp; Campinas, SP, 1990.
- ⁴⁶ Harborne, J. B. *Phytochemical methods*; Chapman and Hall; London, 1991.
- ⁴⁷ Ikan, R. *Natural products: a laboratory guide*; Academic Press; San Diego, 1991.
- ⁴⁸ Liu, YM.; Sheu, SJ. *J. High Resolut. Chromatogr.* **1994**, 17, (7), 559.
- ⁴⁹ Delgado, C.; Tomasbarberan, FA.; Talou, T.; Gaset, A. *Chromatographia* **1994**, 38, (1-2), 71.
- ⁵⁰ Campeol, E.; Catalano, S.; Cremonini, R.; Morelli, I. *Caryologia* **2000**, 53, (1), 63.
- ⁵¹ Dyer, J. R. *Aplicações da Espectroscopia de Absorção aos compostos Orgânicos*; Editora Edgard Blucher Ltda; São Paulo, 1977.
- ⁵² Laidlaw, W. G. *Introduction to Quantum Concepts in Spectroscopy*; McGraw Hill Inc; New York, 1970.
- ⁵³ Ishida, H. *Fourier transform infrared characterization of polymers*; Hatsuo Ishida; New York, 1987.
- ⁵⁴ Whiston, C. *X-Ray Methods*; John Wiley & Sons; Chichester, 1987.

-
- ⁵⁵ Rosales, M. A.; Chao, C.; Cesareo, R.; Gigante, G. E. *Análisis por Fluorescencia de Rayos X mediante Equipos portátiles*; UNAM; México, 1999.
- ⁵⁶ Leyden, D. E. *Spectroscopy* **1987**, 2 (6), 28.
- ⁵⁷ Martin, M. J.; Pablos, F.; González, A G. *Food Chem.* **1999**, 66, 365.
- ⁵⁸ Ladd M. F. C; Palmer, R. *Structure determination by X-Ray Crystallography*; New York, 1985.
- ⁵⁹ Lyalikov, Y. *Physicochemistry Analysis*; Mir Publishers ed.; Moscow, 1968.
- ⁶⁰ Skoog, D. A. and Leary, J. J. *Principles of Instrumental Analysis*; Ed. Saunders College, 4^a ed., EUA, 1992.
- ⁶¹ Bokobza, L. *J. Near Infrared Spectrosc.* **1998**, 6, 3.
- ⁶² Massart, D. L.; Vandeginste, B. G. M.; Deming, S. N.; Michotte, Y.; Kaufman, L. *Chemometrics: A textbook*; Elsevier, 1988.
- ⁶³ Beebe, K. R; Pell, R. J.; Seasholtz, M. B. *Chemometrics: A Practical Guide*; John Wiley; New York, 1998.
- ⁶⁴ Mardia, K. V.; Kent, J. T.; Bibby, J. M. *Multivariate Analysis*; Academic Press Inc.; London, 1982.
- ⁶⁵ Press, W. H.; Flannery, B. P.; Teukolsky, S. A. ; Vetterling, W. T.; *Numerical Recipes* (Cambridge: Cambridge University Press, 1986), p.p 52-64.
- ⁶⁶ Geladi, P.; Kowalski, B. R. *Anal. Chim. Acta* **1986**, 185, 1.
- ⁶⁷ Ferreira, M. M. C; Antunes, A M.; Melgo, M. S.; Volpe, P. L. *Química Nova* **1999**, 22, 724.
- ⁶⁸ Kowalski, B. R. *Chemometrics: Mathematics and statistics in chemistry Series C. Mathematical and Physical Sciences*; Vol 138, Reidel Publishers, Holland, 1984.
- ⁶⁹ Wold, S; Sjostrom, M. *SIMCA: A Method for Analyzing Chemical Data in Terms of Similarity and Analogy in*; Kowalski, 1977.
- ⁷⁰ Eklöv, T.; Martenson, P.; Lundström, I. *Anal. Chim. Acta* **1999**, 381, 221.
- ⁷¹ Martens, H.; Naes, T. *Multivariate Calibration*; John Wiley & Sons Ltd; Chichester, 1996.

-
- ⁷² CAMO-Computer Aided Modelling A/S. *The Unscrambler 6.0*; Trondheim, Norway, 1986-1996.
- ⁷³ Kalivas, J. H.; Roberts, N.; Sutter, J. M. *Anal. Chem.* **1989**, 61, 2024.
- ⁷⁴ Xie, Y. L.; Kalivas, J. H. *Anal. Chim Acta* **1997**, 348, 19.
- ⁷⁵ Xie, Y. L.; Kalivas, J. H. *Anal. Chim Acta* **1997**, 348, 29.
- ⁷⁶ Instituto Adolfo Lutz. *Normas Analíticas do Instituto Adolfo Lutz. Métodos físicos e químicos para análise de alimentos*; São Paulo, Vol 1, 1985.
- ⁷⁷ A.O.C.S. *Official Methods of Analysis*. American Oil Chemists Society, Volume 1 e 2, 1989.
- ⁷⁸ Schwartz, N. O. A.; Maxwell, R. J. *J. Am. Oil Chem. Soc.* **1979**, 56, 634.
- ⁷⁹ Bentes, M. H. *Lipídios em Vegetais-Novo esquema de extração*. Manaus-Brasil, 1988.
- ⁸⁰ Khan, W. O.; Scheinmann, F. *Prog. Chem Fats other Lipids* **1978**, 15, 347.
- ⁸¹ I.U.P.A.C. *International Union of Pure and Applied Chemistry*, 1987.
- ⁸² Infometrix, Inc. *Pirouette Version 2.02*; Woodinville; Washington, 1996.
- ⁸³ Garcia, T. D. E. *Tese Mestrado*, Instituto Nacional de Pesquisas da Amazônia, Amazonas, 1989.
- ⁸⁴ Rojas, J. M.; Serruya, H.; Bentes, M. H. S. *J. Am. Oil Chem. Soc.* **1994**, 71, 127.
- ⁸⁵ Clement, C. R. *Tese Mestrado*, Costa Rica, 1986.
- ⁸⁶ Mongrand, S.; Bessoule, J. J.; Cabantous, F.; Cassagne, C. *Phytochemistry* **1998**, 49, 1049.
- ⁸⁷ Nakamoto, K. *Infrared Spectra of Inorganic and Coordination Compounds*; John Wiley & Sons, Inc., New York, 1970.
- ⁸⁸ Suchánek, M.; Filipová, H.; Volka, K.; Delgadillo, L.; Fresenius, A. N. D. *J. Anal. Chem.* **1996**, 354, 327.
- ⁸⁹ Da Silva, B.H. E. *Tese de doutorado*, Universidade Estadual de Campinas, 1999.

APÊNDICE

Tabela A. 1: Massas e dimensões das amostras dos gêneros *Oenocarpus* e *Jessenia* (n=10)*

Amostra	PF ^a (g)	CF ^b (mm)	DF ^c (mm)	PS ^d (g)	CS ^e (mm)	DS ^f (mm)
A1(O.min)	4,09	20,70	17,99	2,49	17,60	14,69
A2(O.min)	4,59	21,71	18,89	3,07	18,15	15,84
A3(O.map)	6,79	24,61	21,67	3,99	20,81	17,55
A4(O.min)	4,76	22,49	19,13	2,94	19,37	15,87
A5(O.bac)	5,93	24,89	20,75	3,49	20,34	16,67
A6(O.min)	3,58	18,52	17,63	2,22	15,74	14,76
A7(O.min)	5,21	22,33	20,01	3,12	18,68	16,50
A8(O.min)	4,30	21,70	18,48	2,73	18,67	15,67
A9(O.min)	3,97	21,35	18,47	2,31	16,73	14,90
A10(O.min)	4,82	23,39	19,15	2,88	19,38	15,55
A11(O.map)	6,03	25,50	18,80	3,84	23,70	15,80
A12(O.min)	6,10	22,80	21,40	3,37	19,00	17,10
A13(O.min)	4,62	21,60	18,80	2,62	18,70	15,00
A14(O.map)	7,74	26,00	21,60	4,62	23,10	17,60
A15(O.map)	7,52	25,54	20,20	4,50	22,90	17,50
A16(J.B)	12,27	32,75	24,78	8,24	30,02	21,00
A17(J.B)	12,14	31,80	24,80	7,24	27,20	20,10
A18(J.B)	12,29	34,00	23,50	7,38	31,50	18,60
A19(J.B)	11,25	30,90	23,60	7,52	27,90	20,10
A20(J.B)	13,55	34,10	24,70	8,18	30,90	20,20
A21(J.B)	13,68	31,60	25,40	10,01	27,90	22,50
A22(J.B)	12,81	31,14	24,81	9,08	28,59	21,30
A23(J.B)	12,40	31,33	25,25	7,71	27,07	19,80
A24(J.B)	12,25	35,58	23,81	8,16	33,06	17,80
A25(J.B)	11,02	29,75	23,36	7,27	26,44	19,30

* número de ensaios para obter cada média;

^a peso do fruto;

^b comprimento do fruto;

^c diâmetro do fruto;

^d peso da semente;

^e comprimento da semente;

^f diâmetro da semente.

Tabela A. 2: Variáveis físico-químicas dos gêneros *Oenocarpus* e *Jessenia* (casca+polpa, n=2)*

Amostras	%Umid ^a	%Oleo ^b	%Cinza ^c	%Prot ^d	nd(40C) ^e
A1(O.min)	44,50	36,73	1,67	5,61	1,4611
A2(O.min)	49,27	29,39	1,50	7,92	1,4600
A3(O.map)	54,11	27,94	2,19	9,72	1,4611
A4(O.min)	52,23	18,77	1,55	7,47	1,4612
A5(O.bac)	53,05	27,82	1,77	8,45	1,4612
A6(O.min)	55,05	18,49	2,14	6,83	1,4614
A7(O.min)	51,06	25,32	1,51	8,93	1,4611
A8(O.min)	55,17	19,82	1,63	8,12	1,4613
A9(O.min)	54,29	24,98	1,93	10,14	1,4600
A10(O.min)	56,82	20,97	2,27	8,74	1,4609
A11(O.map)	58,86	11,08	1,36	7,37	1,4661
A12(O.min)	54,61	18,17	1,88	7,68	1,4620
A13(O.min)	56,26	14,97	1,40	5,28	1,4615
A14(O.map)	43,21	30,12	2,22	6,06	1,4657
A15(O.map)	45,13	19,02	2,18	8,78	1,4603
A16(J.B)	45,21	20,91	1,74	7,08	1,4610
A17(J.B)	39,60	18,86	1,96	4,93	1,4624
A18(J.B)	39,73	25,22	2,01	5,65	1,4625
A19(J.B)	44,51	23,75	2,70	5,03	1,4610
A20(J.B)	44,09	28,20	2,28	6,56	1,4601
A21(J.B)	42,55	17,65	2,19	3,34	1,4601
A22(J.B)	43,50	17,05	1,34	9,54	1,4601
A23(J.B)	47,35	17,53	1,74	10,60	1,4605
A24(J.B)	43,10	25,51	2,21	10,63	1,4608
A25(J.B)	42,56	27,24	1,87	7,58	1,4572

* número de ensaios,

^a teor de umidade,

^b teor de óleo,

^c teor de cinzas,

^d teor de proteínas,

^e índice de refração

Tabela A. 3: Composição em ácidos graxos dos óleos da casca+polpa dos frutos dos gêneros *Oenocarpus* e *Jessenia* extraídos em natura (n=2)*.

Amostras	^a C14:0	^b C16:0	^c C18:0	^d C18:1	^e C18:2	^f C18:3
A1(O.mn)	0,16	16,50	8,73	73,30	0,50	0,60
A2(O.min)	0,11	16,57	8,51	70,85	1,17	0,77
A3(O.map)	0,19	23,06	13,42	62,57	0,00	0,64
A4(O.min)	0,38	14,56	9,93	68,84	3,96	2,04
A5(O.bac)	0,30	29,56	14,87	54,11	0,00	0,87
A6(O.min)	0,18	14,23	9,35	72,34	2,06	1,66
A7(O.min)	0,12	12,95	5,16	80,28	0,54	0,73
A8(O.min)	0,12	16,20	5,28	77,19	0,46	0,58
A9(O.min)	0,10	15,99	6,25	76,36	0,29	0,90
A10(O.min)	0,08	13,11	4,70	80,47	0,59	0,90
A11(O.map)	0,00	16,10	5,60	77,80	0,50	0,00
A12(O.mini)	0,00	14,67	5,20	78,50	1,61	0,02
A13(O.min)	0,00	14,56	4,95	77,95	2,30	0,24
A14(O.map)	0,00	17,41	3,54	75,10	2,92	0,96
A16(J.B)	0,00	15,35	3,45	79,50	1,57	0,13
A17(J.B)	0,00	15,40	4,80	78,20	1,60	0,00
A18(J.B)	0,00	13,51	2,82	80,43	2,70	0,49
A19(J.B)	0,00	13,93	4,82	79,68	1,50	0,07
A20(J.B)	0,00	13,55	3,35	79,82	1,96	0,79
A21(J.B)	0,00	14,10	4,05	78,40	2,79	0,75
A22(J.B)	0,00	18,00	7,78	68,44	5,76	0,00
A23(J.B)	0,00	18,51	9,31	66,36	3,82	0,00
A24(J.B)	0,00	14,05	6,02	77,52	2,01	0,40
A25(J.B)	0,00	12,55	4,05	81,04	1,92	0,44

* número de ensaios, três Algarismos são significativos.

^a Ácido mirístico,

^b Ácido palmítico,

^c Ácido esteárico,

^d Ácido oléico,

^e Ácido linolêico,

^f Ácido linolênico.

Tabela A.4: Conteúdo dos elementos em % encontrados nas cinzas da casca+polpa dos gêneros *Oenocarpus* e *Jessenia* através de EDXRF.

Amostra	P	K	Ca	Mn	Fe	Zn	Cu
A1(O.min)	0,289	5,681	0,626	0,018	0,019	0,008	0,004
A2(O.min)	0,197	5,504	0,636	0,008	0,019	0,006	0,003
A3(O.map)	0,350	6,727	1,041	0,015	0,026	0,023	0,006
A4(O.min)	0,201	6,364	0,275	0,005	0,013	0,005	0,004
A5(O.bac)	0,417	5,370	0,224	0,009	0,016	0,005	0,004
A6(O.min)	0,215	6,282	0,266	0,008	0,015	0,006	0,004
A7(O.min)	0,527	7,039	0,795	0,023	0,037	0,014	0,006
A8(O.min)	0,300	3,708	0,529	0,007	0,020	0,290	0,005
A9(O.min)	0,372	5,863	0,470	0,014	0,020	0,008	0,004
A10(O.min)	0,346	6,553	0,553	0,009	0,016	0,008	0,003
A11(O.map)	0,416	7,925	1,473	0,078	0,031	0,029	0,010
A12(O.min)	0,114	1,627	0,156	0,025	0,009	0,000	0,001
A13(O.min)	0,602	7,715	0,571	0,015	0,032	0,018	0,004
A14(O.map)	0,463	7,784	1,137	0,189	0,045	0,036	0,007
A16(J.B)	0,130	7,093	0,767	0,018	0,031	0,554	0,006
A17(J.B)	0,075	3,919	0,857	0,025	0,029	0,007	0,006
A18(J.B)	0,283	5,524	1,291	0,042	0,033	0,016	0,003
A19(J.B)	0,158	3,564	2,601	0,090	0,033	0,007	0,006
A20(J.B)	0,134	2,899	0,577	0,008	0,014	0,001	0,003
A21(J.B)	0,167	2,861	1,136	0,024	0,026	0,005	0,006
A22(J.B)	0,000	4,220	1,275	0,075	0,036	0,029	0,003
A23(J.B)	0,222	3,150	2,031	0,120	0,060	0,006	0,004
A24(J.B)	0,101	7,642	0,820	0,053	0,045	0,046	0,008
A25(J.B)	0,174	4,372	2,135	0,069	0,040	0,050	0,004

Tabela A. 5: Dados numéricos dos espectros de EDXRF das amostras dos gêneros *Oenocarpus* e *Jessenia* na faixa de energia 0-20 keV com respostas a intervalos de 0.0195 keV (algumas respostas mais relevantes).

Nvar*	1	115	185	202	...	322	432	840	1024
keV**	0,0000	2,2230	3,5880	3,9195	...	6,2595	8,4045	9,3405	19,9485
A1(O.min)	0	685	2432	308	...	388	285	191	0
A2(O.min)	0	496	2430	311	...	359	288	171	0
A3(O.map)	0	534	3509	494	...	418	505	222	0
A4(O.min)	0	698	1445	149	...	267	221	149	0
A5(O.bac)	0	460	1273	141	...	332	240	163	0
A6(O.min)	0	612	1419	153	...	288	289	202	0
A7(O.min)	0	761	2821	328	...	518	367	170	0
A8(O.min)	0	540	2220	298	...	391	5167	953	0
A9(O.min)	0	575	1940	252	...	360	287	196	0
A10(O.min)	0	577	2210	262	...	283	253	183	0
A11(O.map)	0	573	4316	628	...	430	547	224	0
A12(O.min)	0	94	835	111	...	265	138	93	0
A13(O.min)	0	492	2237	250	...	444	410	186	0
A14(O.map)	0	609	3662	477	...	696	565	232	0
A16(J.B)	0	483	2683	364	...	465	7051	1285	0
A17(J.B)	0	241	3332	504	...	523	298	187	0
A18(J.B)	0	430	4348	645	...	465	415	206	0
A19(J.B)	0	329	8921	1375	...	550	290	197	0
A20(J.B)	0	200	2413	354	...	333	196	210	0
A21(J.B)	0	493	4427	617	...	546	264	159	0
A22(J.B)	0	441	4798	653	...	575	602	220	0
A23(J.B)	0	287	7496	1108	...	967	318	208	0
A24(J.B)	0	455	2818	366	...	661	786	240	0
A25(J.B)	0	444	7121	1070	...	641	902	299	0

*número de variável

** energia em quilo elétron volt

Tabela A.6: Dados numéricos dos difratogramas de XRD das amostras dos gêneros *Oenocarpus* e *Jessenia* na faixa de $2\theta = 20-70^\circ$ com respostas (intensidade em u.a.) a intervalos de $2\theta = 0.02^\circ$ (algumas respostas mais relevantes).

Nvar*	1	434	492	681	...	2123	2499	2500	2501
$2\theta^{**}$	20.00	28,66	29,82	33,60	...	62.44	69.96	69.98	70
A1(O.min)	80	206	272	132	...	34	52	36	40
A2(O.min)	58	130	360	158	...	42	38	50	34
A3(O.map)	90	146	312	150	...	58	42	48	50
A4(O.min)	118	100	524	132	...	44	38	36	38
A5(O.bac)	90	96	486	166	...	48	40	36	54
A6(O.min)	102	126	346	150	...	50	50	42	34
A7(O.min)	88	82	530	150	...	50	46	40	36
A8(O.min)	54	1534	208	144	...	40	38	42	38
A9(O.min)	78	128	386	192	...	40	40	40	38
A10(O.min)	84	88	540	180	...	50	44	46	44
A11(O.map)	72	108	238	100	...	40	38	46	32
A12(O.min)	96	94	528	158	...	60	50	58	60
A13(O.min)	92	118	306	86	...	58	48	44	46
A14(O.map)	108	154	232	206	...	68	56	50	38
A16(J.B)	66	108	270	160	...	54	54	42	44
A17(J.B)	110	138	360	212	...	70	54	40	58
A18(J.B)	44	40	68	52	...	30	20	18	30
A19(J.B)	82	98	194	166	...	120	34	44	42
A20(J.B)	86	130	546	200	...	52	46	58	66
A21(J.B)	142	166	400	148	...	72	54	50	46
A22(J.B)	104	110	408	180	...	80	62	46	42
A23(J.B)	94	142	514	206	...	54	60	38	50
A24(J.B)	110	130	652	370	...	54	54	42	52
A25(J.B)	104	162	500	194	...	70	58	44	52

*número de variável

** ângulo em graus

Tabela A. 7: Dados de absorvâncias na região do UV/Vis dos extratos em éter de petróleo das amostras dos gêneros *Oenocarpus* e *Jessenia* na faixa de 240-440 nm.

Nvar*	1	2	3	4	5	6	7
λ (nm)**	240	250	260	270	280	290	300
A1(O.min)	0.754	0.756	1.149	1.624	1.094	0.744	0.445
A2(O.min)	1.624	1.389	2.012	1.881	1.588	1.287	0.910
A3(O.map)	1.463	0.993	1.245	1.491	1.031	0.689	0.396
A4(O.min)	1.594	1.641	2.452	1.948	1.829	1.429	0.679
A5(O.bac)	1.054	2.500	2.500	2.500	2.500	1.254	0.612
A6(O.min)	0.643	0.616	0.903	1.145	0.828	0.498	0.206
A7(O.min)	1.510	1.462	2.500	2.438	1.834	1.295	0.774
A8(O.min)	0.835	0.820	1.317	1.659	1.098	0.681	0.332
A9(O.min)	2.500	2.500	2.500	2.450	1.890	1.650	0.151
A10(O.min)	1.353	1.282	1.584	1.335	1.470	1.225	0.574
A11(O.map)	2.384	2.071	1.417	1.084	1.144	0.961	0.574
A12(O.min)	2.500	2.500	2.500	2.500	2.500	2.499	1.505
A13(O.min)	2.500	2.500	2.500	1.872	1.665	1.207	0.615
A14(O.map)	2.314	2.137	1.528	1.131	1.357	1.374	0.941
A16(J.B)	2.377	2.057	1.319	0.981	0.927	0.731	0.461
A17(J.B)	2.344	2.149	1.444	1.043	1.184	1.051	0.641
A18(J.B)	2.500	2.500	2.500	2.500	2.500	0.951	0.469
A19(J.B)	2.359	2.116	1.416	1.074	1.171	0.973	0.566
A20(J.B)	2.500	2.330	2.500	2.133	1.464	0.930	0.475
A21(J.B)	2.294	2.085	1.452	1.087	1.248	1.090	0.626
A22(J.B)	2.454	2.248	1.411	1.066	0.992	0.741	0.451
A23(J.B)	2.183	1.537	1.076	0.865	0.857	0.660	0.407
A24(J.B)	2.197	1.324	1.081	0.932	0.862	0.639	0.394
A25(J.B)	2.413	2.150	1.320	0.965	0.883	0.675	0.417

Tabela A.7 Continuação..

Nvar	8	9	10	11	12	12	14
λ (nm)	310	320	330	340	350	360	370
A1(O.min)	0.327	0.266	0.200	0.146	0.098	0.061	0.041
A2(O.min)	0.664	0.553	0.473	0.345	0.231	0.143	0.097
A3(O.map)	0.297	0.251	0.191	0.146	0.103	0.070	0.054
A4(O.min)	0.352	0.268	0.201	0.168	0.112	0.061	0.045
A5(O.bac)	0.452	0.363	0.259	0.186	0.130	0.071	0.046
A6(O.min)	0.111	0.082	0.049	0.037	0.027	0.019	0.017
A7(O.min)	0.587	0.511	0.410	0.327	0.221	0.123	0.086
A8(O.min)	0.211	0.163	0.115	0.076	0.045	0.022	0.015
A9(O.min)	0.123	0.423	0.237	0.225	0.183	0.182	0.182
A10(O.min)	0.337	0.285	0.245	0.178	0.126	0.080	0.065
A11(O.map)	0.406	0.325	0.254	0.211	0.176	0.143	0.122
A12(O.min)	0.987	0.761	0.579	0.478	0.399	0.328	0.287
A13(O.min)	0.394	0.302	0.207	0.169	0.155	0.150	0.153
A14(O.map)	0.679	0.512	0.381	0.305	0.247	0.197	0.164
A16(J.B)	0.353	0.303	0.251	0.213	0.178	0.143	0.121
A17(J.B)	0.474	0.403	0.334	0.281	0.227	0.175	0.140
A18(J.B)	0.330	0.262	0.200	0.156	0.120	0.075	0.051
A19(J.B)	0.403	0.333	0.267	0.223	0.183	0.146	0.121
A20(J.B)	0.319	0.255	0.174	0.141	0.121	0.108	0.101
A21(J.B)	0.433	0.348	0.274	0.228	0.126	0.146	0.121
A22(J.B)	0.353	0.312	0.185	0.224	0.185	0.145	0.118
A23(J.B)	0.302	0.254	0.198	0.177	0.167	0.158	0.156
A24(J.B)	0.299	0.245	0.184	0.166	0.159	0.152	0.153
A25(J.B)	0.325	0.286	0.239	0.206	0.172	0.138	0.116

Tabela A.7 Continuação..

Nvar	15	16	17	18	19	20	21
λ (nm)	380	390	400	410	420	430	440
A1(O.min)	0.027	0.017	0.012	0.009	0.007	0.006	0.005
A2(O.min)	0.062	0.041	0.031	0.027	0.022	0.015	0.010
A3(O.map)	0.043	0.033	0.026	0.021	0.016	0.012	0.010
A4(O.min)	0.037	0.035	0.035	0.034	0.026	0.018	0.012
A5(O.bac)	0.029	0.017	0.011	0.009	0.007	0.005	0.005
A6(O.min)	0.016	0.016	0.016	0.016	0.014	0.012	0.012
A7(O.min)	0.063	0.048	0.040	0.035	0.031	0.028	0.027
A8(O.min)	0.009	0.004	0.002	0.000	-0.001	-0.001	-0.001
A9(O.min)	0.200	0.202	0.220	0.240	0.180	0.125	0.000
A10(O.min)	0.052	0.042	0.037	0.033	0.031	0.028	0.026
A11(O.map)	0.110	0.103	0.100	0.099	0.091	0.082	0.078
A12(O.min)	0.261	0.257	0.265	0.260	0.204	0.149	0.106
A13(O.min)	0.155	0.175	0.201	0.213	0.145	0.096	0.053
A14(O.map)	0.141	0.127	0.120	0.115	0.107	0.101	0.096
A16(J.B)	0.106	0.097	0.091	0.088	0.086	0.083	0.083
A17(J.B)	0.118	0.105	0.098	0.094	0.092	0.089	0.088
A18(J.B)	0.037	0.030	0.026	0.022	0.016	0.010	0.004
A19(J.B)	0.106	0.096	0.091	0.087	0.086	0.082	0.082
A20(J.B)	0.099	0.101	0.107	0.108	0.092	0.075	0.054
A21(J.B)	0.106	0.095	0.087	0.082	0.081	0.078	0.077
A22(J.B)	0.101	0.091	0.086	0.082	0.081	0.078	0.077
A23(J.B)	0.155	0.161	0.169	0.172	0.159	0.136	0.117
A24(J.B)	0.152	0.160	0.169	0.173	0.164	0.139	0.116
A25(J.B)	0.101	0.091	0.086	0.082	0.081	0.079	0.078

* número de variável

** comprimento de onda em nanômetros

Tabela A. 8: Dados de absorvâncias na região do IR médio das pastilhas KBr-cinza das amostras dos gêneros *Oenocarpus* e *Jessenia* na faixa de 4000-400 cm^{-1} com respostas a intervalos de 1,93 cm^{-1} (respostas mais relevantes).

Nvar*	1	26	115	327	...	553	1573	1868
cm-1**	399.23	447.5	619.1	1027.9	...	1463.8	3431.1	4000.00
A1(O.min)	0.124	0.118	0.197	0.187	...	0.168	0.186	0.104
A2(O.min)	0.161	0.154	0.184	0.187	...	0.192	0.196	0.138
A3(O.map)	0.144	0.141	0.204	0.258	...	0.345	0.271	0.194
A4(O.min)	0.135	0.137	0.239	0.247	...	0.233	0.229	0.157
A5(O.bac)	0.167	0.178	0.256	0.314	...	0.306	0.323	0.201
A6(O.min)	0.088	0.090	0.123	0.138	...	0.135	0.169	0.124
A7(O.min)	0.118	0.119	0.189	0.210	...	0.220	0.225	0.161
A8(O.min)	0.087	0.091	0.120	0.153	...	0.097	0.147	0.110
A9(O.min)	0.106	0.111	0.156	0.192	...	0.197	0.189	0.132
A10(O.min)	0.132	0.128	0.163	0.170	...	0.223	0.167	0.111
A11(O.map)	0.118	0.120	0.143	0.207	...	0.247	0.207	0.110
A12(O.min)	0.239	0.206	0.208	0.290	...	0.395	0.308	0.184
A13(O.min)	0.120	0.120	0.129	0.217	...	0.179	0.261	0.174
A14(O.map)	0.185	0.178	0.302	0.427	...	0.539	0.309	0.159
A16(J.B)	0.233	0.276	0.301	0.375	...	0.495	0.582	0.241
A17(J.B)	0.420	0.498	0.405	0.576	...	0.467	0.556	0.384
A18(J.B)	0.207	0.207	0.227	0.260	...	0.429	0.238	0.169
A19(J.B)	0.159	0.156	0.140	0.177	...	0.154	0.198	0.133
A20(J.B)	0.268	0.333	0.264	0.423	...	0.406	0.475	0.223
A21(J.B)	0.165	0.201	0.264	0.285	...	0.220	0.265	0.179
A22(J.B)	0.303	0.331	0.468	0.424	...	0.750	0.341	0.263
A23(J.B)	0.151	0.157	0.149	0.208	...	0.267	0.235	0.167
A24(J.B)	0.139	0.135	0.163	0.178	...	0.368	0.301	0.209
A25(J.B)	0.148	0.151	0.167	0.216	...	0.340	0.222	0.132

* número de variável

** número de onda em cm^{-1}

Tabela A. 9: Amostras e teores dos elementos químicos que formam o conjunto de calibração e o conjunto teste para obtenção dos modelos de calibração multivariada.

Variáveis dependentes (teor de elementos químicos nas cinzas de amostras dos gêneros *Oenocarpus* e *Jessenia*).

Amostra	P	K	Ca	Mn	Fe	Zn	Cu
Conjunto de Calibração							
1(O.min)	0,289	5,681	0,626	0,018	0,019	0,008	0,004
4(O.min)	0,201	6,364	0,275	0,005	0,013	0,005	0,004
5(O.bac)	0,417	5,370	0,224	0,009	0,016	0,005	0,004
6(O.min)	0,215	6,282	0,266	0,008	0,015	0,006	0,004
7(O.min)	0,527	7,039	0,795	0,023	0,037	0,014	0,006
8(O.min)	0,300	3,708	0,529	0,007	0,020	0,290	0,005
10(O.min)	0,346	6,553	0,553	0,009	0,016	0,008	0,003
11(O.map)	0,416	7,925	1,473	0,078	0,031	0,029	0,010
12(O.min)	0,114	1,627	0,156	0,025	0,009	0,000	0,001
13(O.min)	0,602	7,715	0,571	0,015	0,032	0,018	0,004
14(O.map)	0,463	7,784	1,137	0,189	0,045	0,036	0,007
16(J.B)	0,130	7,093	0,767	0,018	0,031	0,554	0,006
17(J.B)	0,075	3,919	0,857	0,025	0,029	0,007	0,006
19(J.B)	0,158	3,564	2,601	0,090	0,033	0,007	0,006
20(J.B)	0,134	2,899	0,577	0,008	0,014	0,001	0,003
21(J.B)	0,167	2,861	1,136	0,024	0,026	0,005	0,006
23(J.B)	0,222	3,150	2,031	0,120	0,060	0,006	0,004
24(J.B)	0,101	7,642	0,820	0,053	0,045	0,046	0,008
25(J.B)	0,174	4,372	2,135	0,069	0,040	0,050	0,004

Conjunto de Teste

2(O.min)	0,197	5,504	0,636	0,008	0,019	0,006	0,003
3(O.map)	0,350	6,727	1,041	0,015	0,026	0,023	0,006
9(O.min)	0,372	5,863	0,470	0,014	0,020	0,008	0,004
18(J.B)	0,283	5,524	1,291	0,042	0,033	0,016	0,003
22(J.B)	0,000	4,220	1,275	0,075	0,036	0,029	0,003

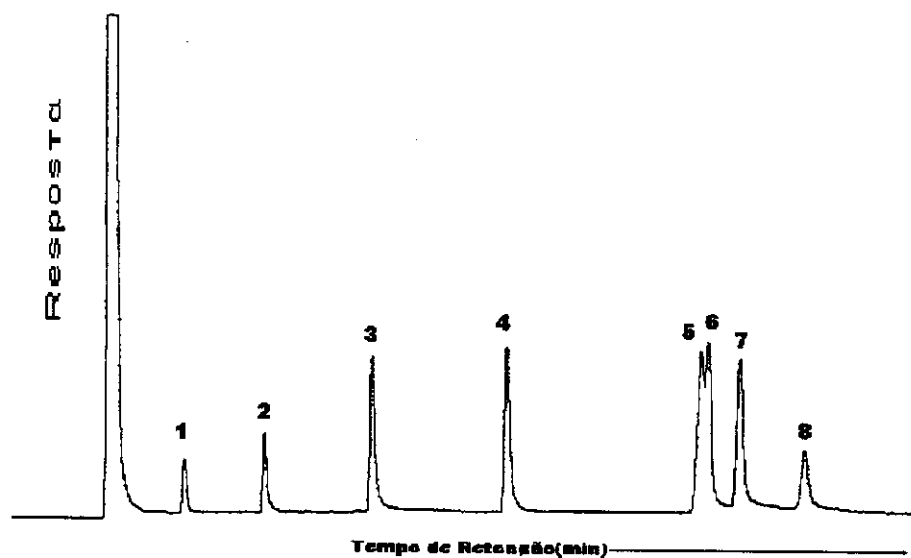


Figura A.1: Cromatograma da mistura de padrões de ésteres metílicos.

No Identificação	Nome	Simbologia	Tempo (min)
1.	Ácido Cáprico	10:0	4,91
2.	Ácido Láurico	12:0	5,91
3.	Ácido Mirístico	14:0	7,24
4.	Ácido Palmítico	16:0	8,93
5.	Ácido Estearico	18:0	11,35
6.	Ácido Oléico	18:1	11,44
7.	Ácido Linoléico	18:2	11,85
8.	Ácido Linolênico	18:3	12,64

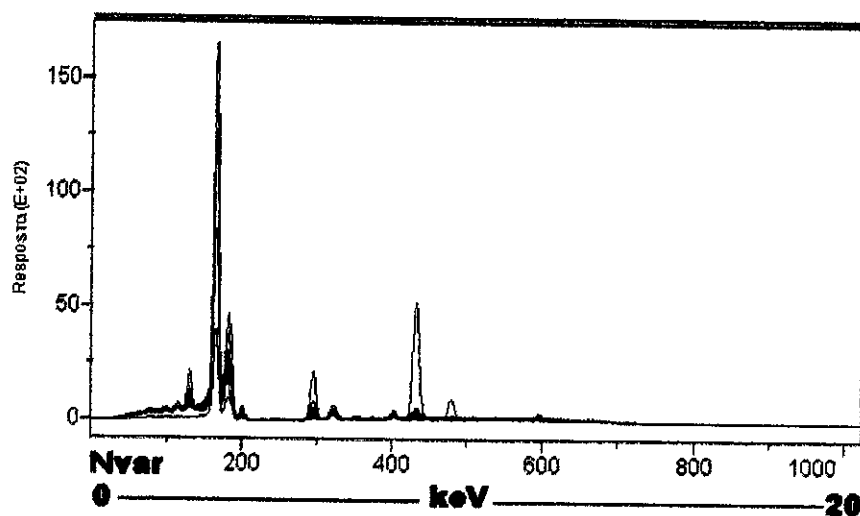


Figura A.2: Espectros de EDXRF das amostras de do gênero *Oenocarpus*

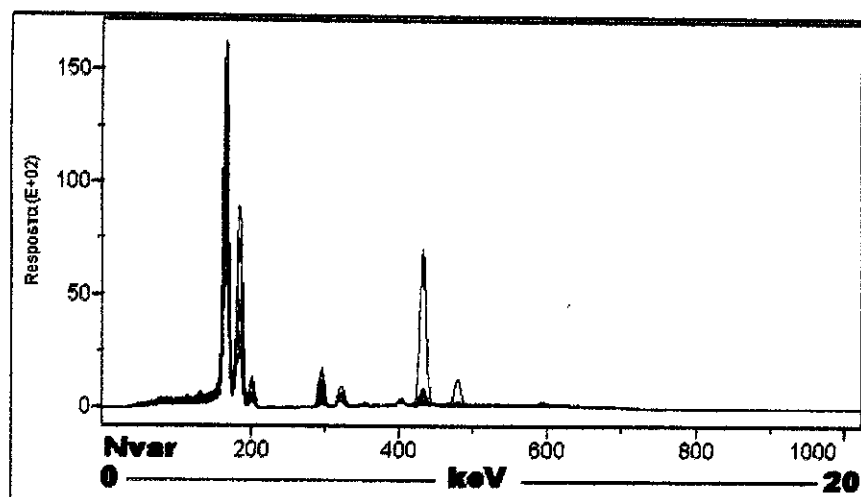


Figura A.3: Espectros de EDXRF das amostras de do gênero *Jessenia*

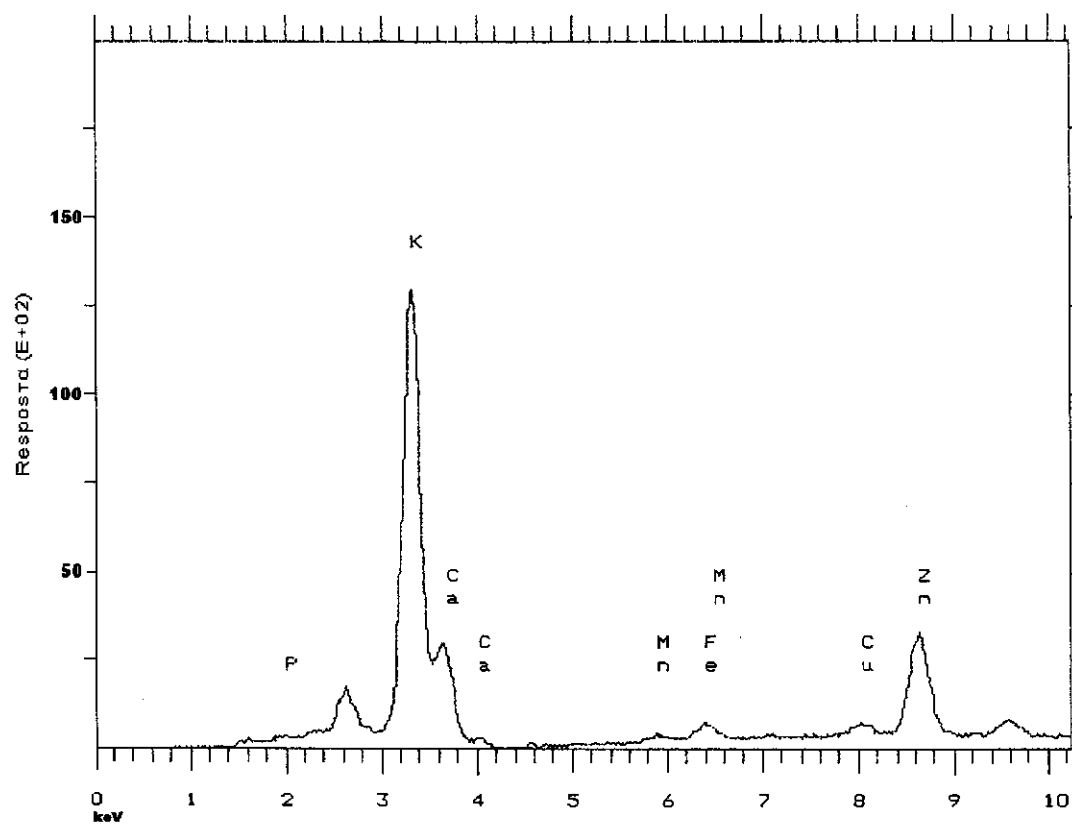


Figura A.4: Espectro de EDXRF da amostra 8 do gênero *Oenocarpus* na faixa de energia com respostas relevantes.

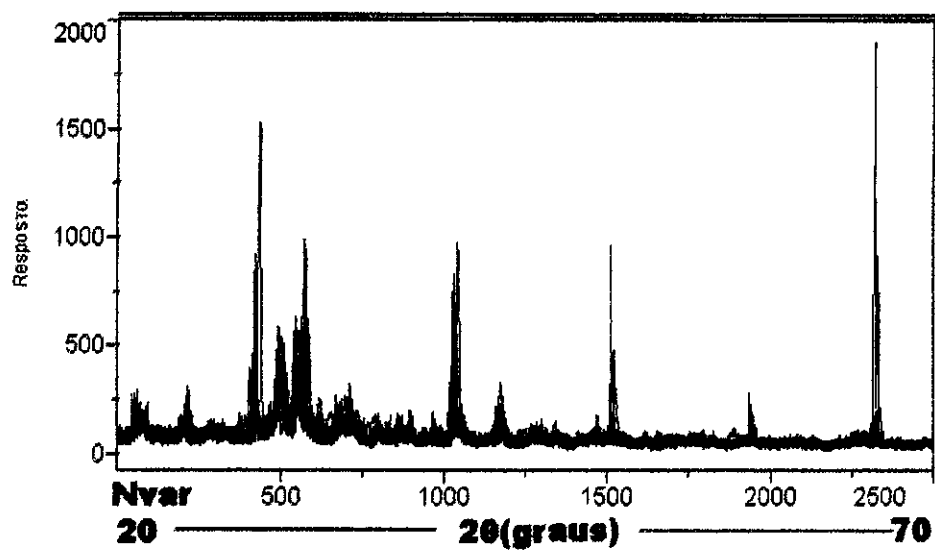


Figura A.5: Difrátogramas de raios-X das amostras do gênero *Oenocarpus*

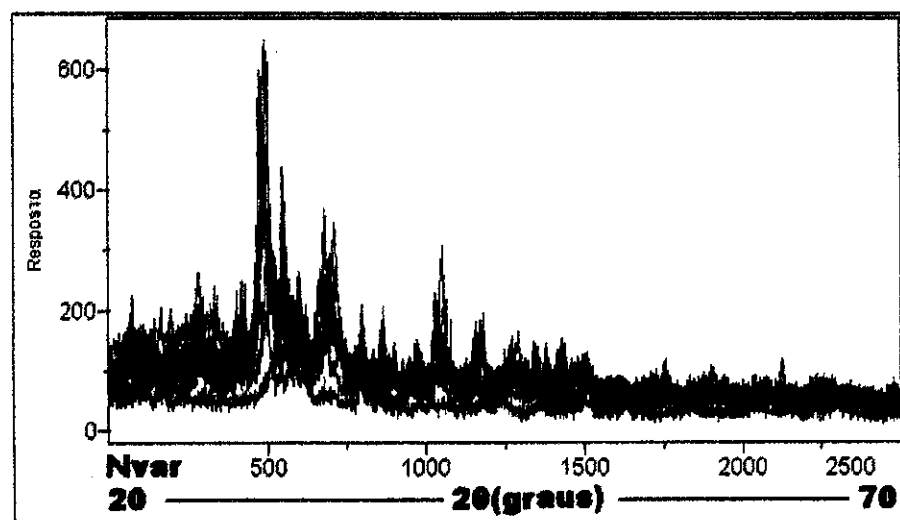


Figura A.6: Difrátogramas de raios-X das amostras do gênero *Jessenia*.

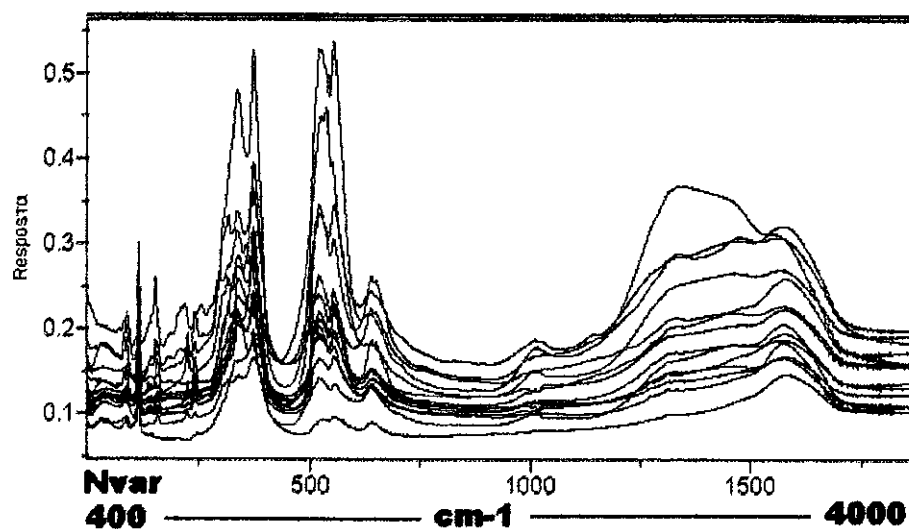


Figura A.7: Espectros de FTIR médio na faixa de 4000 a 400 cm^{-1} das amostras do gênero *Oenocarpus*.

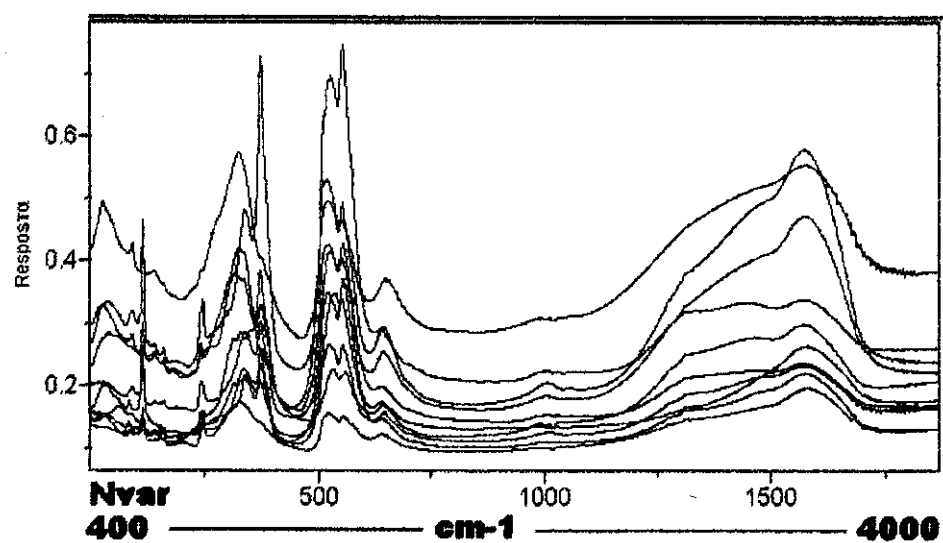


Figura A.8: Espectros de FTIR médio na faixa de 4000 a 400 cm^{-1} das amostras do gênero *Jessenia*.