

Este exe.  
corresponde a r  
final da base de dados  
por Paulo Sérgio de Souza  
e aprovada pela comissão julg

UNIVERSIDADE ESTADUAL DE CAMPINAS

INSTITUTO DE QUÍMICA

11/03/86

Orientador: Prof. Dr. Roy Edward Bruns

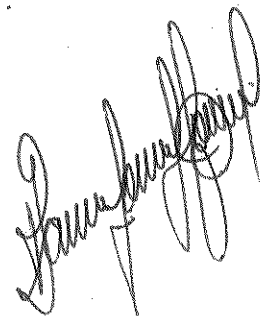


"PRÉ-PROCESSAMENTO DE DADOS ESPECTRAIS NO I.V.  
PARA CLASSIFICAÇÃO DE ALCOÓIS UTILIZANDO  
MÉTODOS DE RECONHECIMENTO DE PADRÕES"

- Paulo Sérgio de Souza -

Tese de Mestrado

Campinas - 1986



UNICAMP  
BIBLIOTECA CENTRAL

## AGRADECIMENTOS

- Ao Prof. Dr. Roy E. Bruns, pela orientação, pela disponibilidade e pela amizade durante todo este trabalho.
- Ao Prof. Dr. J.F.G. Faigle, pelo apoio e discussões sobre o desenvolvimento deste trabalho.
- A Prof<sup>ª</sup>. Dr<sup>ª</sup>. Teresa D.Z. Atvars, por ceder o laboratório para a preparação das amostras.
- A Diretoria do Instituto de Química da Universidade Estadual de Campinas, pelo apoio concedido durante todo este trabalho.
- Ao Centro de Computação da Universidade Estadual de Campinas (CCUEC).
- Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelas bolsas de estudo.
- A Maria Madalena, pela compreensão e incentivo durante todo este trabalho.
- Aos amigos Rogério Custodio e Renata Rodrigues, pelos ensinamentos de computação.
- Aos membros da Comissão Examinadora deste trabalho, de quem aguardo críticas construtivas e sugestões para desenvolvimento futuro.
- A Valda, pela ajuda nas correções de textos.
- A todos aqueles que direta ou indiretamente contribuíram para a realização deste trabalho.
- E, em especial, aos amigos Mário César e Ieda Scarmínio.

A minha família, em especial  
minha mãe Maria Alves,

Dedico

## INDICE

	página
Resumo . . . . .	viii
Abstract . . . . .	ix
<b>Capítulo I - Introdução . . . . .</b>	<b>1</b>
1.1 - Objetivos . . . . .	7
1.2 - Problemas de classificação . . . . .	8
1.3 - Espectros vibracionais no I.V. de álcools	15
<b>Capítulo II - Parte Experimental . . . . .</b>	<b>18</b>
<b>Capítulo III - Métodos de Pré-processamento de dados . . . . .</b>	<b>20</b>
- Atribuições de bandas . . . . .	20
- Semivariograma . . . . .	25
- Função de Autocorrelação . . . . .	29
- Métodos Convencionais de Pré-processamento de dados.	32
<b>Capítulo IV - Métodos de Reconhecimento de Padrões . . . . .</b>	<b>36</b>
IV.1 - Análise de discriminante linear (LDA). . . . .	36
IV.2 - Máquina de aprendizagem linear (LLM) . . . . .	38
IV.3 - Regra do vizinho mais próximo (KNN) . . . . .	39
IV.4 - Modelos independentes de similaridade usando componentes principais (SIMCA). . . . .	42

	página
<u>Capítulo V</u> - Resultados e Discussões . . . . .	55
<u>Capítulo VI</u> - Conclusões . . . . .	81
<u>Referências</u> . . . . .	84
<u>Apêndice A</u> - Programas dos métodos de pré-processamen- to de dados desenvolvidos . . . . .	87
<u>Apêndice B</u> - Modelo dos conjuntos de dados obtidos ex- perimentalmente . . . . .	90

#### FIGURAS

01. Fluxograma de todo desenvolvimento deste trabalho.	11
02. Matriz de dados X, com n objetos e p variáveis . .	12
03. Objeto num espaço p-dimensional . . . . .	13
04. Espectro vibracional com divisões em faixas de frequências . . . . .	22
05. Fluxograma do método de pré-processamento de dados usando as atribuições de bandas. . . . .	23
06. Espectro caracterizado pelo seu semivariograma . .	27
07. Fluxograma do método de pré-processamento de dados (semivariograma) . . . . .	28
08. Função de autocorrelação num espectro vibracional com um determinado retardamento . . . . .	30

## página

09. Fluxograma do método de pré-processamento de dados utilizando a função de autocorrelação) . . . . .	31. ←
10. Separação de categorias em espaço p-dimensional pela análise de discriminante linear (LDA) . . . . .	37
11. Separação de classes pela máquina de aprendizagem linear (LLM) . . . . .	40
12. Classificação das amostras pela regra do vizinho mais próximo (KNN) . . . . .	41
13. Classificação de objetos unando o método SIMCA . . . . .	43
14. Conjunto de dados representado pelo seu ponto médio na análise de componentes principais . . . . .	45
15. Primeiro componente principal definido pelo vetor "loading" $p_1$ . . . . .	46
16. Segundo componente principal perpendicular a $p_1$ . . . . .	48
17. Primeiros componentes A para representar X . . . . .	49
18. Sistema de coordenadas original com as dimensões X e X . . . . .	51

TABELAS

01. Relação dos álcoois estudados com os números da biblioteca SADTLER . . . . .	9
02. Bandas características dos espectros vibracionais no I.V. de álcoois . . . . .	17
03. Pesos obtidos relativo à separação dos álcoois . . . . .	57

	página
04. Resultados do método KNN nas matrizes de dados 24x12 e 13x12 (atribuições de bandas) . . . . .	58
05. Resultados do método SIMCA nas matrizes de dados 24x12 e 13x12 (atribuições de bandas) . . . . .	59
06. Resultados dos métodos KNN e SIMCA na matriz de dados utilizando os valores de absorvâncias, através quatro possíveis alternativas . . . . .	62
07. Resultados do método KNN na matriz de dados utilizando os valores de absorvâncias . . . . .	63
08. Resultados do método SIMCA na matriz de dados utilizando os valores de absorvâncias . . . . .	64
09. Resultados dos métodos KNN e SIMCA na matriz de dados utilizando o semivariograma, através das quatro possíveis alternativas . . . . .	67
10. Resultados do método KNN na matriz de dados utilizando o semivariograma . . . . .	68
11. Resultados do método SIMCA na matriz de dados utilizando o semivariograma . . . . .	69
12. Resultados dos métodos KNN e SIMCA na matriz de dados utilizando a função de autocorrelação, através das quatro possíveis alternativas . . . . .	72
13. Resultados do método KNN na matriz de dados utilizando a função de autocorrelação . . . . .	73
14. Resultados do método SIMCA na matriz de dados utilizando a função de autocorrelação . . . . .	74

## página

15. Resultados dos métodos KNN e SIMCA para o conjunto de treinamento usando as absorvâncias, semivariograma e função de autocorrelação, através das quatro possíveis alternativas . . . . . 78
16. Resultados dos métodos KNN e SIMCA para o conjunto teste (biblioteca SADTLER), usando as absorvâncias, semivariograma e função de autocorrelação, através das quatro possíveis alternativas . . . . . 79
17. Resultados dos métodos KNN e SIMCA nas matrizes de dados utilizando os valores de absorvâncias, semivariograma e a função de autocorrelação . . . . . 80



## RESUMO

Três métodos de pré-processamento de dados foram desenvolvidos e testados, para verificação de suas eficiências, quanto à potencialidade de discriminação dos álcoois primários, secundários e terciários, usando dados espectrais na região de I.V. Os métodos de reconhecimento de padrões contidos no sistema computacional ARTHUR e adaptados para um microcomputador foram aplicados para esta classificação. Estes espectros vibracionais foram obtidos da biblioteca SADTLER e de um espectrofotômetro de I.V., utilizando como medidas as frequências e absorvâncias dos mesmos.

No primeiro método de pré-processamento de dados desenvolvido, foram feitas atribuições de bandas, empregando os valores das frequências e absorvâncias do espectro sob estudo. No segundo método, o espectro foi caracterizado pelo seu semivariograma, um conceito emprestado da geostatística e no terceiro, usamos a função de autocorrelação.

O método KNN forneceu melhores resultados de classificação correta dos álcoois nas categorias primário, secundário e terciário, quando os dados foram pré-processados pela função de autocorrelação, independentes da normalização. O método SIMCA classificou corretamente os álcoois em suas respectivas categorias, utilizando os valores de absorvâncias e a função de autocorrelação, com os dados normalizados ou não, mas autoescalados. Baseado nos espectros da biblioteca SADTLER foi possível classificar corretamente os espectros obtidos experimentalmente.

## ABSTRACT

Three data preprocessing methods were developed and tested, to verify their efficiencies, in the discrimination of primary, secondary and tertiary alcohols, using spectra data of the infrared region. The pattern recognition methods contained in the ARTHUR computational package and adapted for a microcomputer were applied in this classification. The vibrational spectra were obtained from the SADTLER library of infrared spectra and by use of an infrared spectrophotometer, with absorbance and frequency measurements for the ordinate and abscissa.

In the first data preprocessing method developed, band attributions were made, employing the spectrum frequency and absorbance values in the matrix data. In second method, the spectrum was characterized by its semivariogram, a concept borrowed from geostatistics and the third, use was made of the autocorrelation function.

The KNN method provided the best results for the correct classification of alcohols in primary, secondary and tertiary categories, when the data were preprocessed by the autocorrelation function, independent of whether the data were normalized or not. The SIMCA method classified correctly all the alcohols in their respective categories, using either absorbance or autocorrelated values which had been auto-scaled independent of whether the data had been first normalized or not. Based on the spectra selected from the SADTLER library it was possible to correctly classify the spectra obtained experimentally.

## CAPÍTULO I

### INTRODUÇÃO

A maior parte dos problemas em química analítica antes de 1960 eram resolvidos por via úmida através de titulação, precipitação, extração, etc. A determinação das medidas (variáveis) em laboratório era um processo trabalhoso e, portanto, faziam-se poucas medidas em cada objeto (amostra). Conseqüentemente, os métodos matemáticos e estatísticos para tratamentos de dados analíticos eram desenvolvidos para aquela situação, com um número de objetos muito maior do que o número de medidas.

Hoje a situação em algumas áreas como a química, biologia, geologia e medicina é, todavia, diferente, porque as amostras são inseridas em instrumentos como: espectrômetros de infravermelho, U.V.-visível, ressonância magnética nuclear, etc., os quais não apresentam uma informação direta como resultado. Estes instrumentos simplesmente produzem grandes quantidades de sinais em forma de curvas, números etc., em resumo, dados. Estes sistemas multivariados podem, de forma sequencial ou simultaneamente, medir muitas variáveis para cada objeto, com grande eficiência. A aplicação de métodos apropriados de matemática e estatística nessas variáveis, permite obter maiores informações a partir da grande quantidade de sinais produzidos por estes instrumentos modernos. Essa aplicação pertence a um novo campo da química em

desenvolvimento, denominado "quimiometria". A quimiometria já foi definida como a disciplina da química que utiliza métodos matemáticos e estatísticos para definir ou selecionar as condições ótimas de medidas e experiências, e permitir a obtenção do máximo de informações a partir da análise de dados químicos[1].

Entre os diferentes procedimentos matemáticos utilizados na quimiometria, os métodos de reconhecimento de padrões são particularmente importantes. Estes são métodos matemáticos e estatísticos eficientes para analisar dados multivariados e podem ser convenientemente empregados para interpretar e obter maiores informações dos dados químicos. O tipo de problema químico que o reconhecimento de padrões\* (RP) pode solucionar foi descrito por Kowalski e Bender[2]: "dado um conjunto de objetos (amostras) e uma série de medidas (sinais analíticos) obtidas destes objetos, há possibilidades de encontrar e/ou prognosticar uma propriedade destes que não seja diretamente mensurável, mas que está relacionada com essas medidas via alguma relação desconhecida". Por exemplo, o conjunto de medidas características de um objeto (um espectro) é considerado como um modelo abstrato, que contém informação sobre uma propriedade não diretamente mensurável (estrutura molecular, atividade biológica, etc) do objeto. Métodos de RP tentam encontrar relações entre o modelo e a propriedade, sem usar necessariamente conhecimentos químicos.

Nos últimos quinze anos os métodos de RP têm sido aplicados a vários problemas químicos. Podemos citar alguns exemplos

---

\* Do inglês: "pattern recognition"

como: i) A classificação de quarenta amostras de vinhos, onde as uvas foram colhidas em duas regiões demográficas diferentes, analisando suas composições elementar e orgânica por espectrometria de emissão atômica e cromatografia gasosa com coluna capilar, respectivamente. Nestas amostras foram investigadas as correlações entre as medidas químicas e as avaliações sensoriais (côr, sabor, açúcar, etc.). Os compostos que estavam relacionados com a qualidade geral do vinho, foram identificados[3]; ii) A classificação de amostras de trinta e quatro marcas de uísques escoceses, com o objetivo de verificar se os mesmos eram falsos comparando-os com Chivas Regal. Os métodos de RP foram aplicados aos resultados das análises de cromatografia gasosa, sendo que cada amostra apresentou dezessete picos observáveis em comum. A classificação obtida é de tal modo confiável, que está sendo considerada como prova forense de falsificação nos E.U.A.[4]; iii) A identificação de artefatos arqueológicos de quatro regiões diferentes, usando fluorescência de raio-X para dez elementos traços, pois em muitos casos a composição desses elementos é característica das fontes individuais. O material (obsidiana) tem sido o objeto de considerável estudo nos anos recentes porque, como um vidro vulcânico, a composição dentro de um determinado fluxo tende ser homogênea[5]; iv) Em química médica, analisando as concentrações de duas enzimas no sangue de cinquenta e cinco pacientes, para classificação de dois tipos de doença no fígado[6]; v) Outra aplicação desses métodos feita pelo grupo de pesquisa em quimio-metria do IQ-UNICAMP, foi a classificação de águas minerais de três regiões diferentes no estado de São Paulo. Foram analisadas

cento e dezessete amostras, sendo os elementos (Na, K, Ca, Mg, Si e P) analisados por um espectrômetro de emissão atômica com fonte de plasma induzido em argônio[7].

Além dos exemplos citados, os artigos de revisão listados na bibliografia[8-13], cobrem a maioria dos aspectos relacionados às atividades da quimiometria.

Alguns trabalhos de RP também foram publicados com relação à interpretação de espectros vibracionais no I.V. para identificação de vários compostos orgânicos, como por exemplo: i) A interpretação de duzentos e doze espectros de moléculas orgânicas, de modo que cada espectro foi dividido em intervalos de 0,1 micra, gerando portanto, cento e trinta e uma variáveis[14]; ii) A interpretação de quinhentos espectros de compostos orgânicos simples, de modo que cada espectro foi digitalizado com intervalos de 0,1 micra, de 2,0 a 14,7 micra, dando um total de cento e vinte e oito variáveis[15]; iii) A interpretação de quatrocentos e sessenta e dois espectros divididos em três conjuntos: duzentos ácidos, duzentos não ácidos e sessenta e dois tratados como desconhecidos para o trabalho de RP[16]; iv) O estudo de efeito da redução de dados e transformações do pré-processamento na habilidade de reconhecimento para dois mil e trezentos espectros na fase de vapor, com base na presença ou ausência de seis grupos funcionais (carboxílico, hidroxílico, amino, nitro, éster e fenil)[17]; v) A identificação de óleos lubrificantes, devido à dificuldade em comparar suas fontes após intempéries ambientais[18].

Constam ainda da literatura trabalhos quimiométricos re-

lacionados com dados espectrais, para extrair informação de determinados grupos funcionais. Podemos citar por exemplo, a aplicação dos métodos de RP para obter informação sobre a conformação cis e trans de dezessets compostos  $\alpha, \beta$ -insaturados com carbonilas, a partir de espectros de I.V. e U.V., em solução[19].

No presente trabalho, são analisados trinta e sete espectros vibracionais no I.V. de álcoois, nas categorias primário, secundário e terciário, tomados da biblioteca SADTLER[20]. As medidas dos espectros (frequências e absorvâncias) foram obtidas para todos os álcoois puros, no estado líquido e com ramificações de grupos alquilas, exceto no caso de um álcool primário, que contem o grupo alcóxi. Também foram analisados quatro espectros de álcoois obtidos experimentalmente, com o objetivo de comparar seus resultados com os dados da biblioteca SADTLER (conjunto de treinamento). Na tabela 01 estão relacionados os álcoois estudados em seus respectivos conjuntos de dados e os números da biblioteca SADTLER.

Três métodos de pré-processamento de dados foram desenvolvidos, no sentido de aprimorar os métodos de RP para classificação dos espectros vibracionais no I.V. de álcoois em suas respectivas categorias, utilizando como variáveis as frequências e absorvâncias desses espectros.

No primeiro método de pré-processamento de dados desenvolvido, foram utilizados os valores das frequências características e as correspondentes absorvâncias relativas. No segundo método, o espectro foi caracterizado pelo seu semivariograma e no terceiro, foi utilizada a função de autocorrelação. Para comparar

os resultados obtidos com estes dois últimos métodos, também foram utilizadas as absorvâncias originais nos cálculos de RP.

O semivariograma e a função de autocorrelação foram escolhidos devido as suas propriedades matemáticas, que talvez possam ser úteis na discriminação entre as categorias de álcoois.

Primeiro, as duas funções minimizam erros que podem resultar de deslocamentos ocasionais na abcissa e ordenada. Deslocamentos na ordenada preocupam mais devido às dificuldades em definir a quantidade de amostra no caminho ótico e a linha de base.

Teoricamente, o semivariograma aplicado aos espectros no I.V. mede a variância entre valores de absorvâncias de frequências sucessivas no espectro. Nas variáveis independentes, as frequências, não são variáveis aleatórias como são a maioria das variáveis usadas em estatística. No entanto, são variáveis regionalizadas, isto é, o valor de uma variável num ponto no espaço depende do valor desta variável num outro ponto no espaço perto do primeiro. Por exemplo, como todas as frequências perto da absorvância máxima do pico do estiramento O-H, devem ter absorvâncias altas, o semivariograma parece uma boa escolha entre as possíveis funções de representação no infravermelho. O semivariograma não foi aplicado em RP para interpretação de dados químicos, sendo utilizado em geoestatística para estimação de reservas de jazidas geológicas.

Finalmente, a função de autocorrelação é interessante do seguinte ponto de vista teórico. A classificação de álcoois em primário, secundário e terciário não deve depender da substituição de um ou mais átomos de hidrogênio por deutério. A função de



autocorrelação já mostrou sua utilidade em descrever a substituição isotópica em espectrometria de massa com sucesso[21]. Aqui, aplicamos esta função para espectros vibracionais no I.V. de álcoois.

Os métodos de RP foram aplicados aos dados pré-processados pelos três métodos desenvolvidos, para testar suas eficiências quanto às potencialidades de discriminação entre os espectros de álcoois primários, secundários e terciários.

Para o primeiro método de pré-processamento de dados desenvolvido, foram usados os métodos estatísticos e de RP contidos no sistema computacional ARTHUR e executados pelo computador PDP-10 do Centro de Computação da Universidade Estadual de Campinas (CCUEC). Para os outros dois métodos, os cálculos foram executados no microcomputador DICON com 64 Kbytes, adaptando neste os subprogramas principais do ARTHUR[22], devido ao melhor acesso do aluno a este do que ao PDP-10. Todo o procedimento deste trabalho pode ser observado através do fluxograma ilustrado na figura 01.

### 1.1. OBJETIVOS

Os objetivos deste trabalho são:

1. Desenvolver métodos de pré-processamento de dados multivariados, utilizando os espectros vibracionais no I.V. de álcoois, no sentido de aprimorar algumas técnicas de RP.

2. Aplicar os métodos de RP, para testar a eficiência dos métodos de pré-processamento de dados, quanto as suas potencialidades de discriminação entre os espectros de álcoois primários, secundários e terciários.

3. Verificar se o conjunto de treinamento obtido da biblioteca SADTLER, é útil na classificação de espectros obtidos experimentalmente, utilizando os métodos de pré-processamento de dados estudados neste trabalho.

## 1.2. PROBLEMAS DE CLASSIFICAÇÃO

O RP e sua área complementar de conhecimento de padrões são ramos da inteligência artificial. O RP divide-se em duas fases, sendo que a primeira baseia-se na existência de um conjunto de treinamento (ou referência), onde os objetos pertencem a uma classe definida. Nesta fase o conjunto é usado para desenvolver regras matemáticas, que utiliza esta informação sobre as classes e os valores das variáveis, para designar os objetos às respectivas classes baseando-se em seus dados. Na segunda fase, estas regras são usadas para prever as classes desconhecidas de objetos de um conjunto teste.

Os dados químicos normalmente são arranjados numa tabela, na forma de uma matriz de dados  $X_{ik}$ , onde os objetos estão dispostos em linhas e as variáveis em colunas. As variáveis são os valores de um determinado número de medidas feitas em cada ob-

TABELA 01 - Relação dos álcoois estudados em seus respectivos conjuntos de dados e números da biblioteca SADTLER.

Conjunto de treinamento - álcoois alifáticos saturados (1-24).

Conjunto teste - álcoois alifáticos insaturados (25-37).

Nº	Conjunto de treinamento	Nº SADTLER
1	propanol	10961
2	butanol (*)	10934
3	pentanol	205
4	octanol	15016
5	decanol	10965
6	dodecanol	251
7	2-metil-butanol-1	9996
8	3-metil-butanol-1	15289
9	2-etil-butanol-1	10963
10	2-etoxietanol	10992
11	isopropanol	65
12	hexanol-2	15697
13	heptanol-2	15013
14	octanol-2	18118
15	decanol-2	774
16	3-etil-pentanol-2	3610
17	2,2-dimetil-pentanol-3	3672
18	terc-butanol (*)	2

- continuação da tabela 01

19	2-metil-butanol-2	15094
20	2-metil-pentanol-2	8451
21	3-metil-pentanol-3	10274
22	2,3-dimetil-butanol-2	12976
23	2,4-dimetil-pentanol-2	3673
24	2,3,4-trimetil-pentanol-3	8028
<u>Conjunto teste</u>		
25	hexeno-5-ol-1	2878
26	trans-hepteno-3-ol-1	17311
27	buteno-3-ol-2	25305
28	3-metil-buteno-3-ol-2	9026
29	2-metil-penteno-1-ol-3	3619
30	hexeno-5-ol-3	16767
31	2-metil-buteno-3-ol-2	8493
32	2-metil-penteno-4-ol-2	9031
33	3-metil-hexeno-5-ol-3	9029
34	3-etil-hexeno-5-ol-3	9115
35	3,5-dimetil-hexeno-5-ol-3	9816
36	4-metil-hepteno-1-ol-4	9126
37	4-metil-1,6-heptadieno-ol-4	9125

(\*) - Espectros também obtidos experimentalmente em duplicata.

Números de 1 a 10 e de 25 a 26 - álcoois primários.

Números de 11 a 17 e de 27 a 30 - álcoois secundários.

Números de 18 a 24 e de 31 a 37 - álcoois terciários.

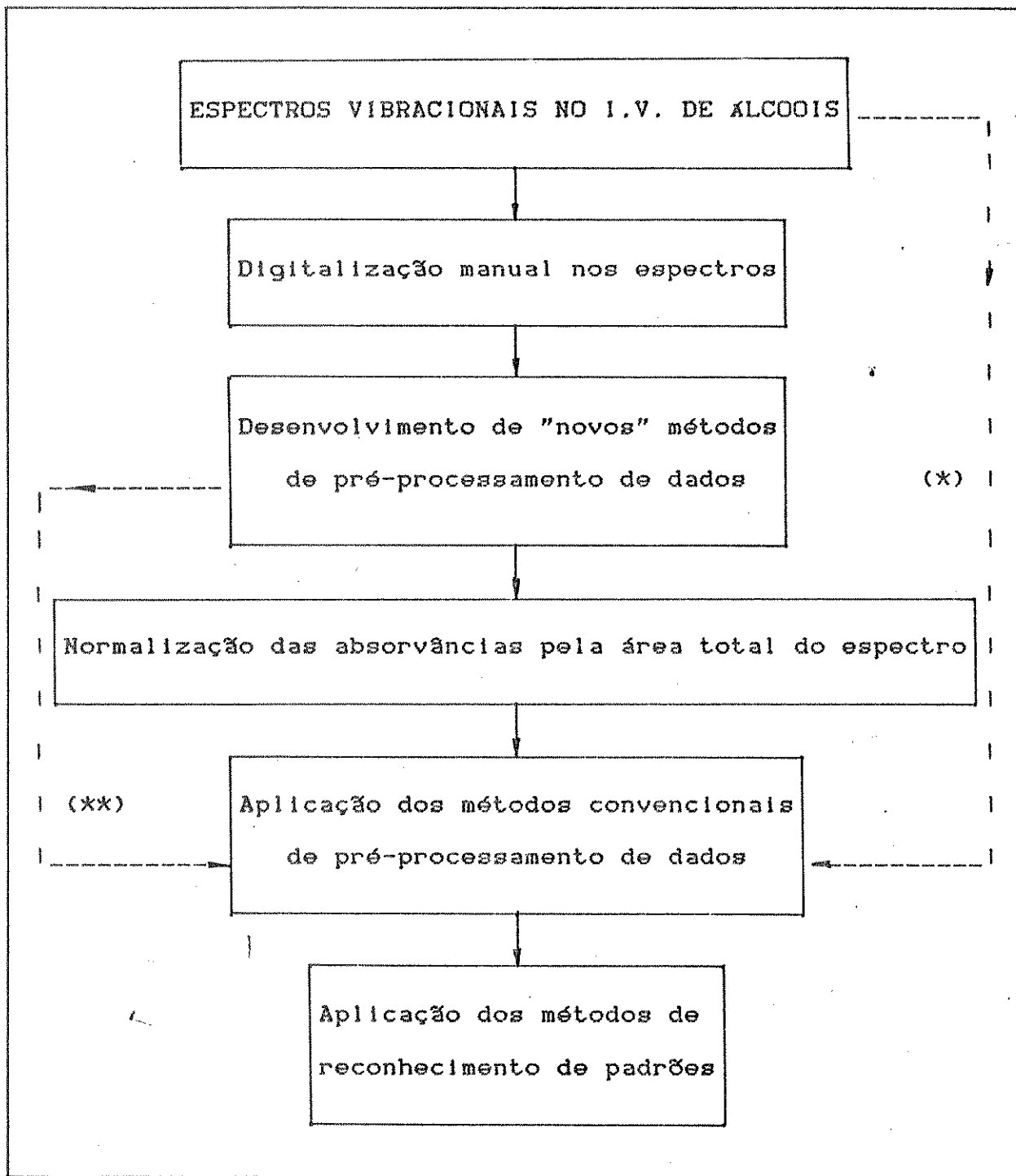
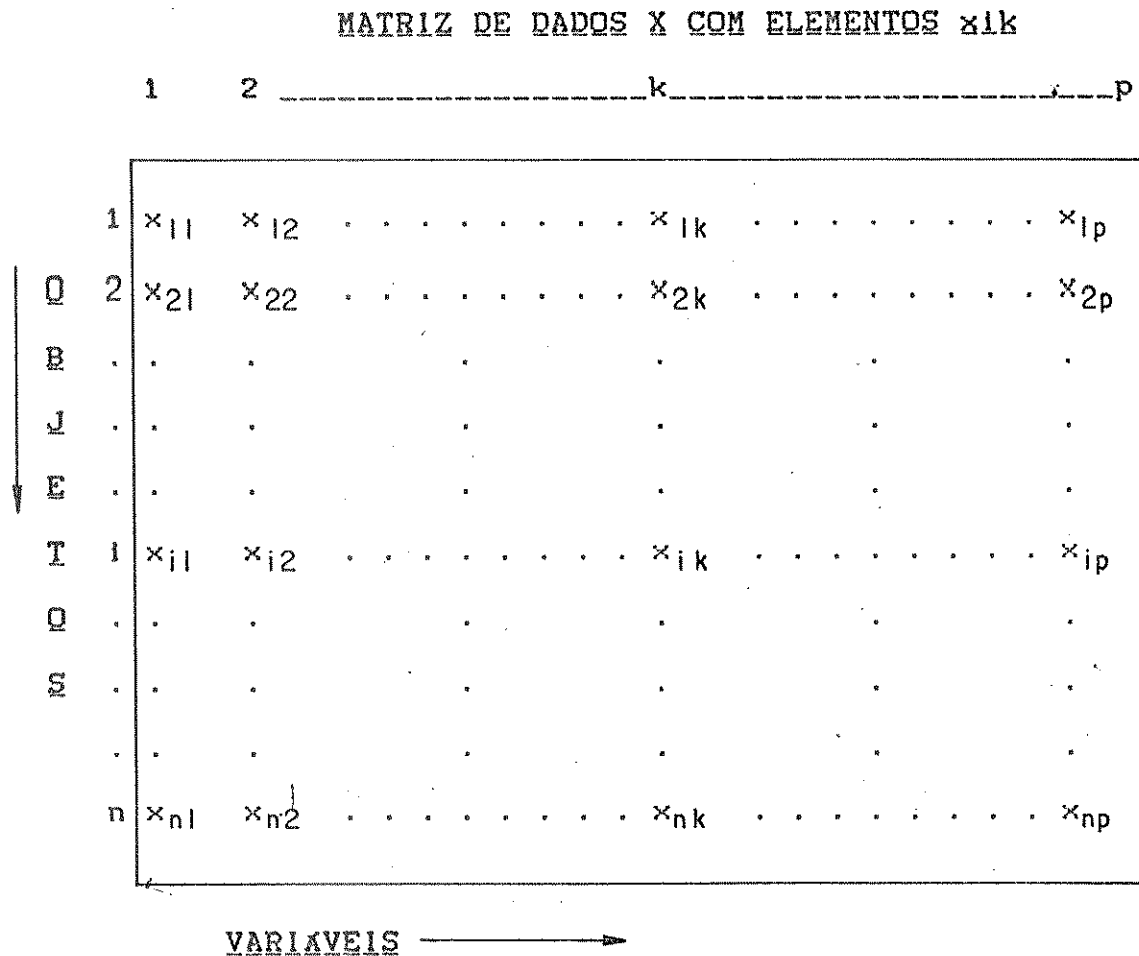


FIGURA 01 - Fluxograma de todo desenvolvimento deste trabalho.

(\*) - Método de atribuições de bandas.

(\*\*) - Métodos de semivariograma e f. autocorrelação.

jeto. O índice  $i$  refere-se aos objetos e o índice  $k$ , às variáveis. A matriz contendo os dados é representada por  $X$ . Portanto, o elemento  $X_{ik}$  indica o valor da variável  $k$  no objeto  $i$ , quando medimos  $p$  variáveis em  $n$  objetos, como mostra a figura 02.

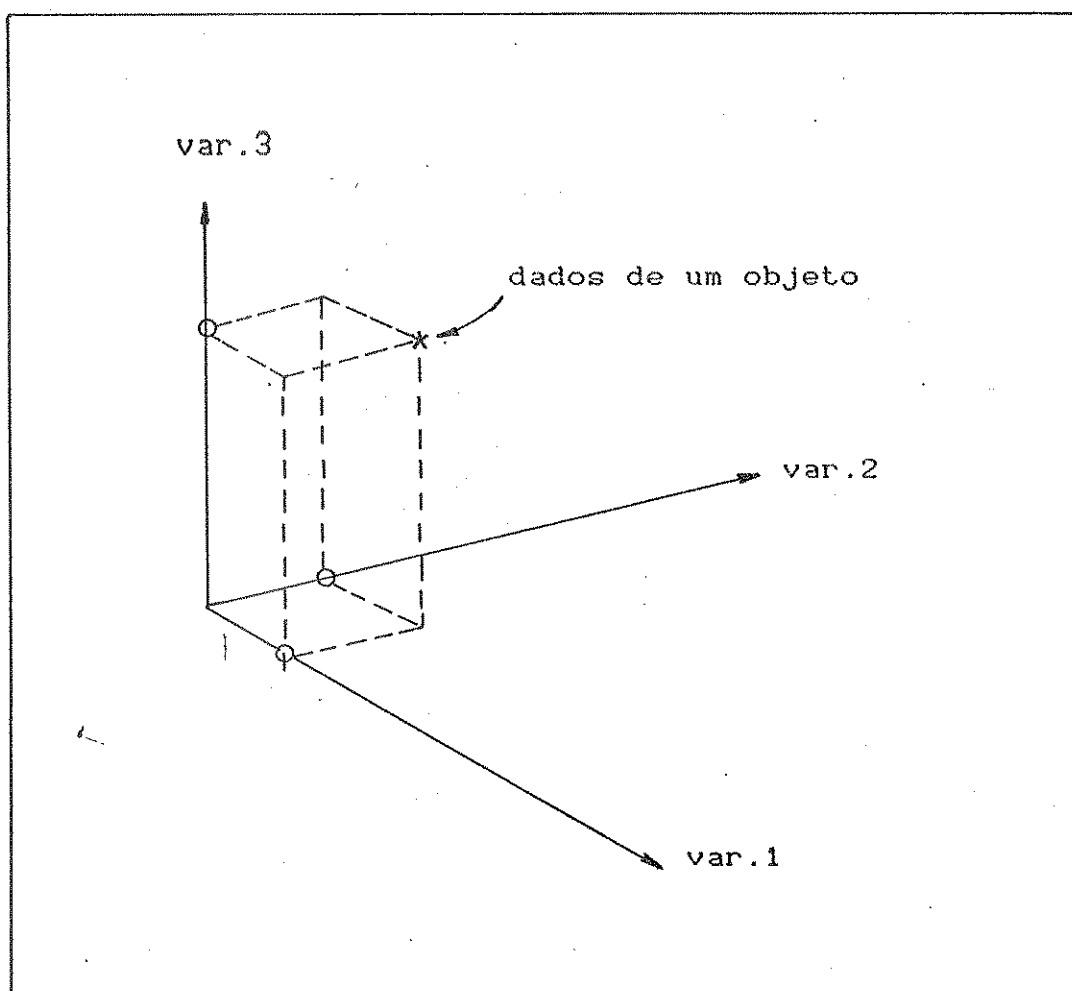


**FIGURA 02** - Matriz de dados  $X$ , com  $n$  objetos e  $p$  variáveis.

A idéia essencial dos métodos de RP pode ser representada em termos de geometria, isso devido a nossa boa habilidade para entender estruturas em espaços bi e tridimensionais e genera-

lizá-los para um espaço de maior dimensão. No espaço  $p$  as construções geométricas como pontos, linhas, planos, ângulos, volumes e distâncias têm definições matemáticas análogas às dos espaços bi e tridimensionais. A propriedade fundamental e prática deste método é que os valores de  $p$  variáveis de um objeto (uma linha na matriz de dados  $X$ ), são representados por um ponto nesse espaço.

Portanto, as medidas são as coordenadas de cada ponto no espaço  $p$ , como ilustra a figura 03.



**FIGURA 03:** Os dados de um objeto são representados pela posição de um ponto no espaço  $p$ -dimensional.

A distância euclidiana entre dois pontos  $i$  e  $j$  de coordenadas  $X_{ik}$  e  $X_{jk}$  é dada pela seguinte equação:

$$D_{ij} = \left[ \sum_{k=1}^p (X_{ik} - X_{jk})^2 \right]^{1/2} \quad (1)$$

onde a soma é feita sobre as  $p$  medidas.

Quanto mais próximo dois pontos neste espaço, maior a similaridade entre os correspondentes objetos. A distância euclidiana é uma medida de similaridade recíproca, porque os objetos são mais semelhantes quando essa distância atinge valores pequenos. A similaridade é definida através da equação:

$$S_{ij} = 1 - D_{ij}/D_{ij}(\text{máx}) \quad (2)$$

onde:

$D_{ij}(\text{máx})$  = maior distância interponto no espaço  $p$

Os tratamentos dos dados em RP podem ser divididos em dois grupos:

1) Métodos paramétricos, onde se assume que as funções de densidades ou distribuições probabilísticas são conhecidas ou podem ser estimadas.

2) Métodos não paramétricos, onde nenhuma suposição é feita quanto à distribuição estatística dos dados. A maior parte dos problemas químicos é relacionada com este procedimento.



Os modos de aplicação para os métodos de RP são:

1) Aprendizagem supervisionada, indica que alguns dos pontos no espaço p-dimensional são nomeados com uma classificação conhecida (conjunto de treinamento ou referência) e o objetivo principal é desenvolver uma regra que classifica estes pontos corretamente e, então, aplicar a mesma regra para a classificação de pontos com classes desconhecidas. Este modo de aplicação é chamado de métodos de reconhecimento de padrões.

2) Aprendizagem não supervisionada, tem como objetivo encontrar densidades ou agrupamentos de pontos reais no espaço p-dimensional, os quais refletem a possível existência de significativas inter-relações. Nesse modo de aplicação não existe um conjunto de treinamento, isto é, o computador não recebe nenhuma informação sobre as possíveis categorias envolvidas, denominado como métodos de conhecimento de padrões.

### 1.3. ESPECTROS VIBRACIONAIS NO I.V. DE ALCÓOIS

As bandas características, observadas nos espectros vibracionais no I.V. de álcoois, resultam das vibrações de estiramento O-H, estiramento C-O e de deformação O-H. Além destas, também são verificadas as bandas de estiramento e de deformação C-H[23]. O estiramento C-O e a deformação O-H não são independentes dos outros modos vibracionais, porque eles acoplam significativamente com as vibrações de grupos adjacentes. O grupo hidroxil-

la, com ligações de hidrogênio intermoleculares, absorve na região de 3550-3200  $\text{cm}^{-1}$  que compreende a região de estiramento O-H. A grande largura desta banda é usualmente atribuída ao fato de que o álcool associa-se em várias formas poliméricas[24]. Todos os álcoois estudados neste trabalho apresentam ligações de hidrogênio intermoleculares, as quais não envolvem estruturas de ressonância.

As vibrações de estiramento C-O em álcoois produz uma banda forte na região de 1260-1000  $\text{cm}^{-1}$ . O estiramento C-O é acoplado com a vibração de estiramento C-C adjacente, de modo que nos álcoois, especialmente nos primários, a vibração seria melhor descrita como uma vibração de estiramento C-C-O assimétrica.

A vibração de deformação O-H no plano ocorre em geral na região de 1420-1330  $\text{cm}^{-1}$ . Nos álcoois primários e secundários a deformação O-H no plano acopla com as vibrações C-H wagging\* para produzir duas bandas: a primeira próxima a 1420  $\text{cm}^{-1}$  e a segunda próxima a 1330  $\text{cm}^{-1}$ . Além dessas bandas, são observadas as vibrações C-H "wagging" para grupos mais afastados do grupo O-H. Os álcoois terciários, nos quais dificilmente podem ocorrer acoplamento, mostram uma única banda essencialmente de deformação O-H além das bandas para C-H "wagging" dos grupos  $\text{CH}_2$  e  $\text{CH}_3$ . As bandas características dos espectros no I.V. de álcoois[25], podem ser verificadas através da tabela 02.

---

\* deformação fora do plano

TABELA Q2: Bandas características dos espectros vibracionais no I.V. de vários tipos de álcoois.

modos vibracionais	tipos de álcoois	faixas de absorção (cm <sup>-1</sup> )
estiramento O-H	O-H livre . . . . .	3650-3590
	ligações de H intermolecular	
	ponte simples . . . . .	3550-3450
	associação polimérica . . . . .	3400-3200
	ligações de H intramolecular	
	ponte simples . . . . .	3570-3450
	compostos quelatados . . . . .	3200-2500
deformação O-H	primário . . . . .	1350-1260
	secundário . . . . .	1350-1260
	terciário . . . . .	1410-1310
estiramento C-O	primário saturado e secundário α-insaturado . . . . .	1085-1050
	secundário saturado . . . . .	1124-1087
	terciário saturado e secundário simétrico . . . . .	1205-1124

## CAPÍTULO II

### PARTE EXPERIMENTAL

Os espectros vibracionais dos álcoois foram obtidos num espectrofotômetro de I.V. Perkin-Elmer, modelo 399-B, colocado no Instituto de Química da Universidade Estadual de Campinas (IQ-UNICAMP). Este instrumento permite a obtenção de espectros em quatro diferentes velocidades de varredura: 1) velocidade lenta com o tempo de sessenta minutos para varrer a faixa espectral de  $4000-400\text{ cm}^{-1}$ , 2) velocidade normal com doze minutos, 3) velocidade rápida com seis minutos e, 4) velocidade muito rápida com três minutos.

Neste trabalho foram analisados quatro espectros vibracionais no I.V. de dois tipos de álcoois alifáticos saturados no estado líquido, pertencentes às categorias primária e terciária, respectivamente: 1) álcool n-butílico de grau analítico (PA) da Merck e, 2) álcool terc-butílico de grau analítico (PA) da Quimis. Estes espectros foram obtidos experimentalmente em temperatura ambiente com duas diferentes velocidades de varredura: normal e rápida.

A janela utilizada nessa experiência foi de NaCl, que foi polida com álcool metílico (PA) da Merck e tomando-se ainda o cuidado de evitar exposição à umidade atmosférica e contato com amostras úmidas. Ambos os álcoois foram usados na forma de pelí-

culas. Uma gota da amostra foi colocada entre as duas janelas de NaCl, com um espaçador de teflon de 0,05 mm, fixadas num suporte e em seguida levado para o espectrofotômetro.

## CAPÍTULO III

### MÉTODOS DE PRÉ-PROCESSAMENTO DE DADOS

Normalmente os métodos de RP são mais eficientes se os dados estão sujeitos a um tratamento chamado pré-processamento. O pré-processamento opera numericamente nos dados originais (medidas), mudando a representação da estrutura dos pontos padrões num espaço modelo, melhorando a eficiência dos métodos de RP.

No primeiro método de pré-processamento de dados foram feitas atribuições de bandas, empregando os valores das frequências características e correspondentes absorvâncias relativas dos espectros vibracionais no I.V. de álcoois. A absorvância é o logaritmo do inverso da transmitância, ou seja,  $A = \log I_0/I = -\log T$  onde,  $I$  é a intensidade da luz transmitida pela amostra e  $I_0$  é a intensidade da luz incidente. As medidas foram obtidas através dos valores dos picos máximos das absorvâncias das regiões espectrais sob estudo. Os trinta e sete espectros vibracionais no I.V. de álcoois, obtidos da biblioteca SADTLER, listados na tabela 01, foram separados em dois conjuntos de dados: 1) o conjunto de treinamento com vinte e quatro amostras de álcoois alifáticos saturados (números 1 a 24 da tabela 01), distribuídos da seguinte maneira: dez primários, sete secundários e sete terciários e; 2) o conjunto teste com treze amostras de álcoois alifáticos insaturados (números 25 a 37 da tabela 01), assim distribuídos: dois

primários, quatro secundários e sete terciários.

Cada espectro analisado foi dividido em faixas de frequências da seguinte maneira: 3500-2900  $\text{cm}^{-1}$ , que corresponde à região das vibrações de estiramento O-H e C-H; de 1500-1300  $\text{cm}^{-1}$ , correspondendo à região de deformação O-H e C-H, normalmente com a ocorrência de dois picos principais nesta região; e de 1200-800  $\text{cm}^{-1}$ , onde ocorrem, entre outras, as vibrações de estiramento C-O. Nessa região normalmente são observados quatro picos relativamente fortes comparados com outros na região. Esse procedimento dado aos espectros pode ser observado através da figura 04.

Com este processo foi possível obter dezesseis variáveis, ou seja, oito frequências características e oito correspondentes absorvâncias. Dada a impossibilidade de se utilizar as absorvâncias absolutas, devido à indefinição da linha base nos espectros e a dificuldade de medir o caminho ótico, foram utilizadas as absorvâncias relativas. Os valores numéricos das absorvâncias foram transformados, dividindo-se o primeiro pelo segundo, o terceiro pelo quarto, e assim sucessivamente. Com esta aplicação foi possível obter doze variáveis, com oito frequências e quatro correspondentes absorvâncias relativas, portanto, gerando uma matriz de dados de dimensão 24x12 para o conjunto de treinamento e 13x12 para o conjunto teste. Na figura 04, que mostra o espectro vibracional no I.V. de um álcool, as oito bandas cujas frequências foram usadas na matriz X são marcadas com uma x. As quatro intensidades relativas, foram calculadas das absorvâncias máximas da banda 1 em relação a esta absorvância de banda 2, a banda 3

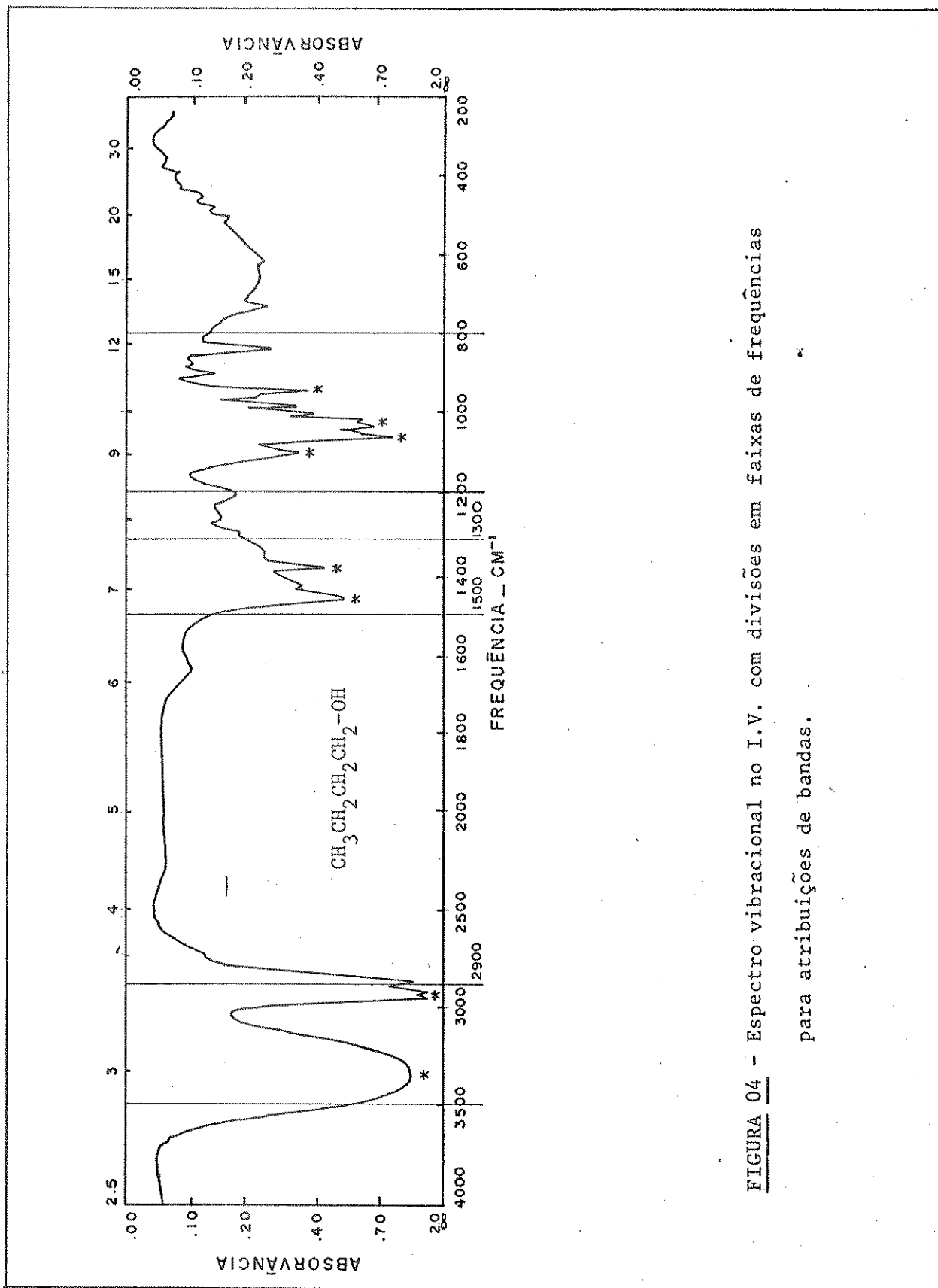
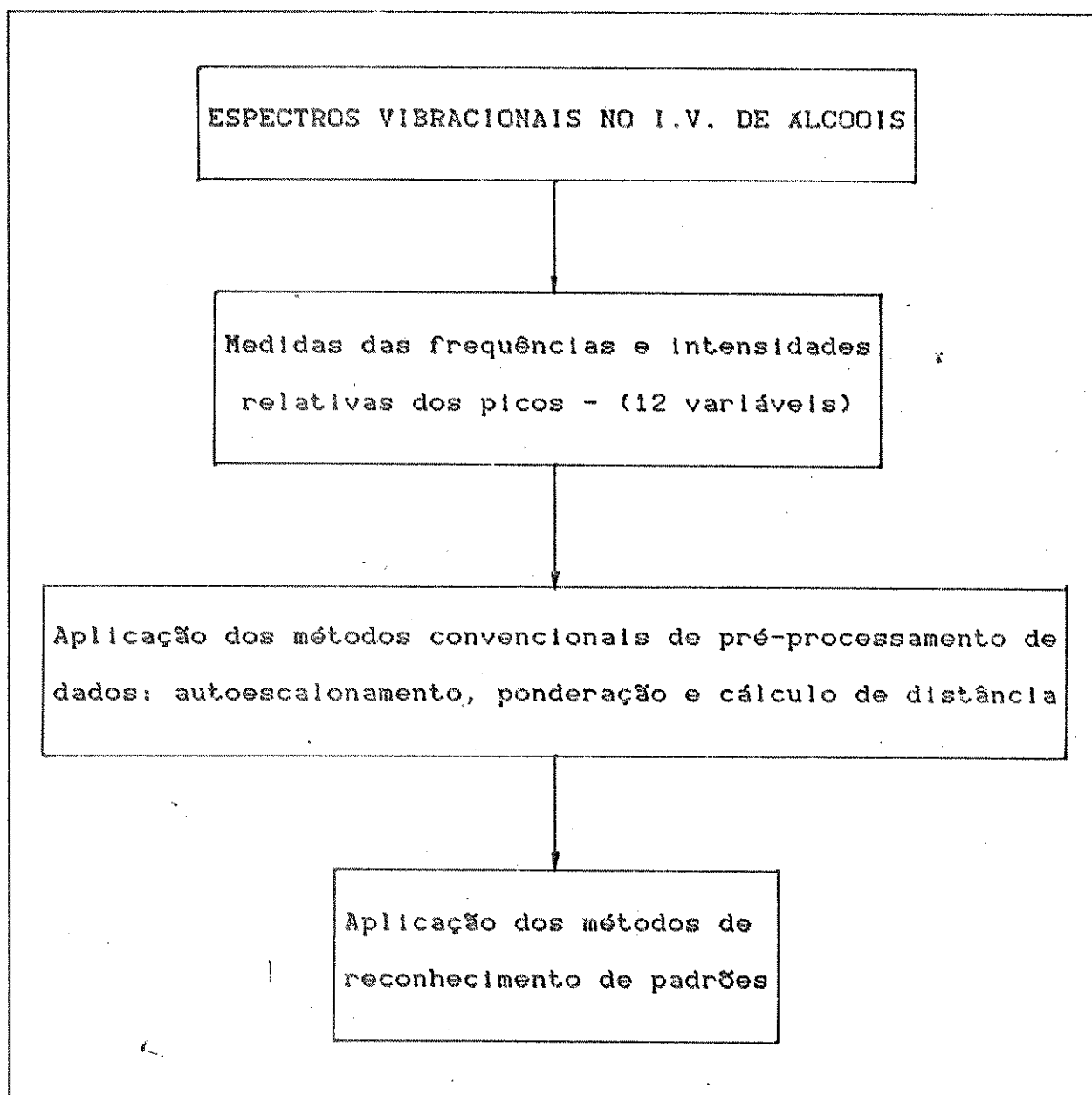


FIGURA 04 - Espectro vibracional no I.V. com divisões em faixas de frequências para atribuições de bandas.





**FIGURA 05:** - Fluxograma do método de pré-processamento de dados usando atribuições de bandas.

com a banda 4 e assim sucessivamente. Todo o procedimento realizado neste método para obtenção dos resultados, pode ser verificado através do fluxograma representado pela figura 05.

No outros métodos de pré-processamento de dados desenvolvidos, foi aplicado um método para se obter maiores informações dos espectros, quanto às suas frequências e absorvâncias. Para isso, os espectros vibracionais foram digitalizados manualmente, com a região de I.V. estendendo-se de  $3700-800\text{ cm}^{-1}$  e com intervalos de frequência de  $20\text{ cm}^{-1}$ , gerando cento e quarenta e seis valores de absorvâncias (variáveis) para cada espectro analisado. A região de  $2180-1560\text{ cm}^{-1}$  não apresenta bandas características para os álcoois saturados, por isso, foi tomado um valor médio para os trinta e dois valores de absorvâncias do respectivo espectro em estudo, para minimizar o trabalho de digitalização. Para os álcoois insaturados a banda correspondente de estiramento C=C é observada na região de  $1680-1640\text{ cm}^{-1}$ , que permite a discriminação dos álcoois entre as categorias saturados e insaturados. Como os objetivos deste trabalho não focalizam esta discriminação, mas a classificação de álcoois primários, secundários e terciários, tentamos eliminar o efeito de insaturação no espectro, representando este intervalo de frequências por uma linha base. Dessa maneira podemos estudar se a ligação C=C perturba as bandas no espectro de I.V., associados com outros grupos característicos na molécula.

Para estes dois métodos foram analisados os mesmos espectros de álcoois utilizados no primeiro método gerando, assim, uma matriz de dados de dimensão  $24 \times 146$  para o conjunto de treina-

mento e 13x146 para o conjunto teste. Esta digitalização foi realizada, também, em quatro espectros vibracionais no I.V. de dois tipos de álcoois alifáticos saturados: n-butanol e terc-butanol, obtidos experimentalmente, em duas diferentes velocidades de varredura e incluídos como um segundo conjunto teste, gerando uma matriz de dados de dimensão 4x146.

Após a digitalização, os dados originais (absorvâncias) foram normalizados pela área total do respectivo espectro, empregando este feito para compensar as diferentes quantidades das amostras no caminho ótico do espectrofotômetro. Esta normalização é uma medida opcional de pré-processamento de dados e pode ser efetuada através da aproximação trapezoidal[26], definida pela seguinte equação:

$$T = [1/2(Y_1) + Y_2 + Y_3 + \dots + Y_{n-1} + 1/2(Y_n)] \Delta x \quad (3)$$

sendo que:

$$\Delta x = (b-a)/n$$

onde:

b-a = intervalo da região do espectro sob estudo

n = número de variáveis

Y = absorvâncias

Um programa desenvolvido para a execução no microcomputador, desta normalização pela área total do espectro, está listado no apêndice A com a designação de AREANORM.FOR.

O segundo método de pré-processamento de dados empregado foi um conceito emprestado da geoestatística, ou seja, a caracte-

rização do espectro pelo seu semivariograma. Define-se o semivariograma[27], indicado por  $\gamma^*(h)$ , como a metade da variância das diferenças dos valores das absorvâncias do espectro entre pontos separados por uma distância  $h$ . Cada um destes valores pode ser representado por um ponto num gráfico de  $\gamma^*$  vs.  $h$ , como ilustra a figura 06. Obviamente, valores de  $\gamma^*(h)$  podem ser calculados apenas a distâncias 1,2,3 etc., desde que estes sejam múltiplos do intervalo das frequências, conforme a equação geral:

$$\gamma^*(h) = 1/[2(N-h)] * [ \sum_{i=1}^{N-h} (G_i - G_{i+h})^2 ] \quad (4)$$

onde:

$N-h$  = número de pares usados

$G_i - G_{i+h}$  = diferença dos valores das absorvâncias

O cálculo dos valores desta função foi realizado conforme o programa relacionado no apêndice A, designado SEMIV.FOR. O procedimento geral realizado para obtenção dos resultados pode ser verificado por um fluxograma, como mostra a figura 07.

distancia gama(h)

1	.00302
2	.00426
3	.00665
4	.00906
5	.00957
6	.00996
7	.01025
8	.01044
9	.01057
10	.01061
11	.01054
12	.01036
13	.01000
14	.00944
15	.00951
16	.00958
17	.00966
18	.00973
19	.00981
20	.00989
21	.00997
22	.01005
23	.01013
24	.01021
25	.01030
26	.01038
27	.01047
28	.01056
29	.01065
30	.01074
31	.01083
32	.01093
33	.01103
34	.01112
35	.01123
36	.01133
37	.01143
38	.01154
39	.01164
40	.01175
41	.01187
42	.01198
43	.01210
44	.01222
45	.01234
46	.01246
47	.01259
48	.01271
49	.01285
50	.01298

(B)

distancia gama(h)

1	.02665
2	.02687
3	.02714
4	.02768
5	.02760
6	.02772
7	.02777
8	.03001
9	.03023
10	.01575
11	.01525
12	.01593
13	.01602
14	.01611
15	.01618
16	.01624
17	.01636
18	.01649
19	.01662
20	.01675
21	.01689
22	.01702
23	.01716
24	.01730
25	.01744
26	.01759
27	.01774
28	.01789
29	.01804
30	.01820
31	.01835
32	.01852
33	.01868
34	.01885
35	.01902
36	.01919
37	.01937
38	.01954
39	.01973
40	.01991
41	.02010
42	.02030
43	.02049
44	.02069
45	.02090
46	.02111
47	.02132
48	.02154
49	.02176
50	.02199

(A)

FIGURA 06 - Espectro caracterizado pelo seu semivariograma:  
 (A) - Heptanol-2  
 Região de 3700-2720 cm<sup>-1</sup>  
 (B) - n-propanol

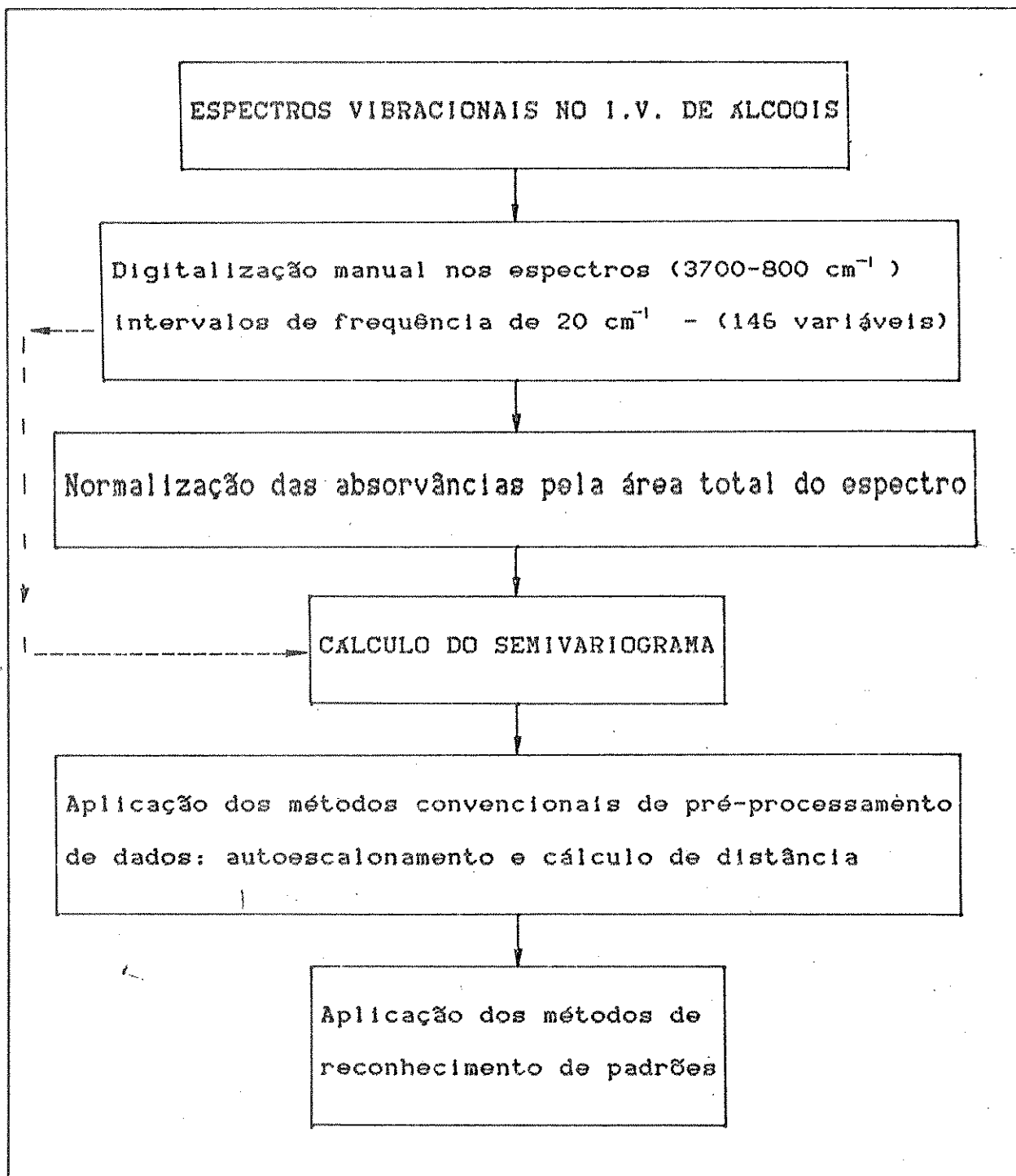


FIGURA 07: Fluxograma do método de pré-processamento de dados usando o semivariograma.

Para o terceiro método de pré-processamento de dados usamos a função de autocorrelação. Esta função é definida como a correlação linear entre uma série de medidas da região do espectro e a mesma série com um determinado retardamento, como ilustra a figura 08. A utilização desta função de autocorrelação tem como objetivo, a determinação dos graus de similaridades ou dissimilaridades, entre diferentes regiões do espectro vibracional em estudo[28].

A autocorrelação de uma série de medidas com um determinado retardamento L é dada pela seguinte equação:

$$r(L) = \frac{[(n-L)(\sum Y_i Y_{i+L}) - (\sum Y_i)(\sum Y_{i+L})] / (n-L)(n-L-1)}{[n(\sum Y_i^2) - (\sum Y_i)^2] / n(n-1)} \quad (5)$$

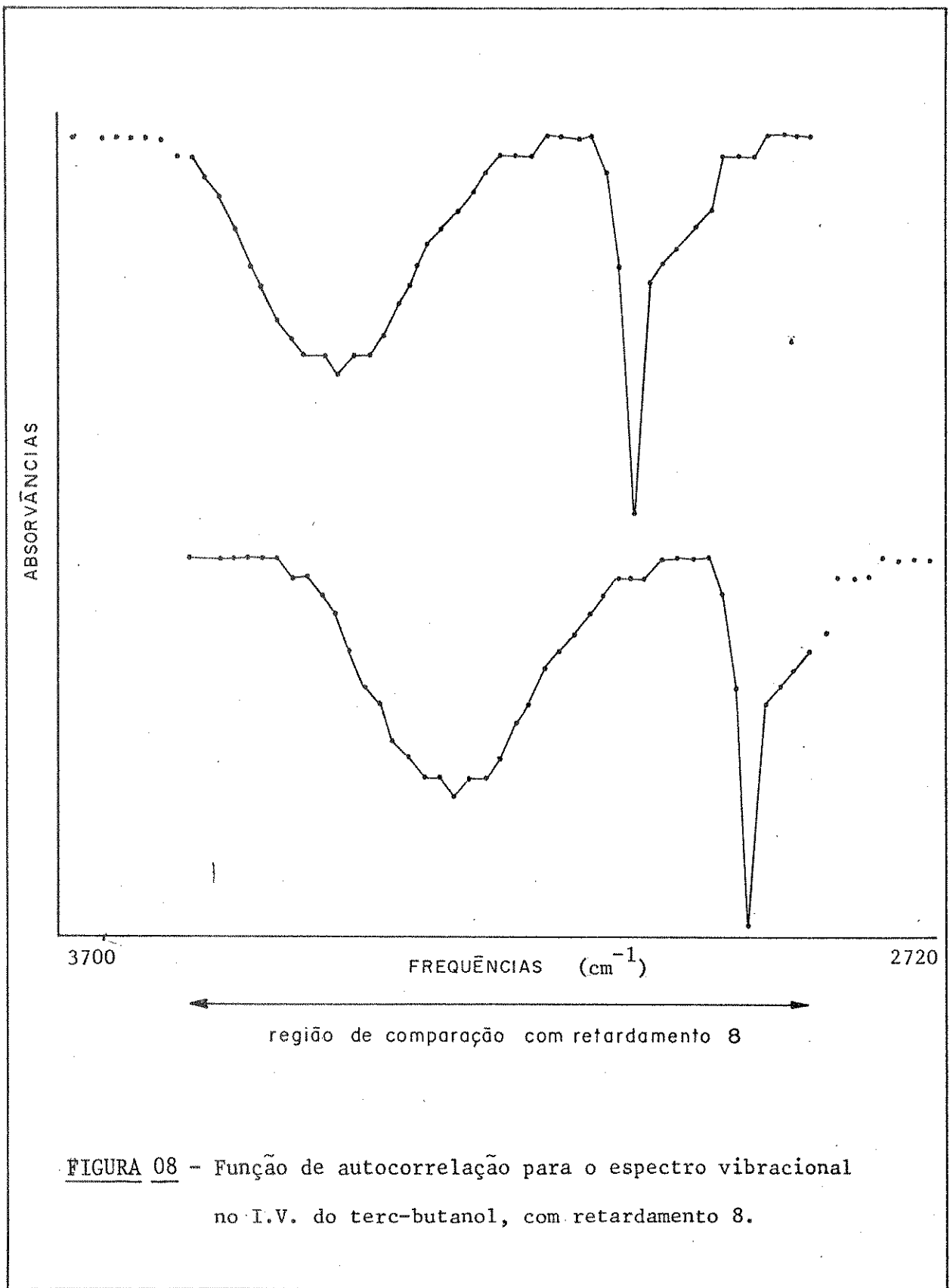
onde:

L = intervalo entre pontos correspondentes no espectro

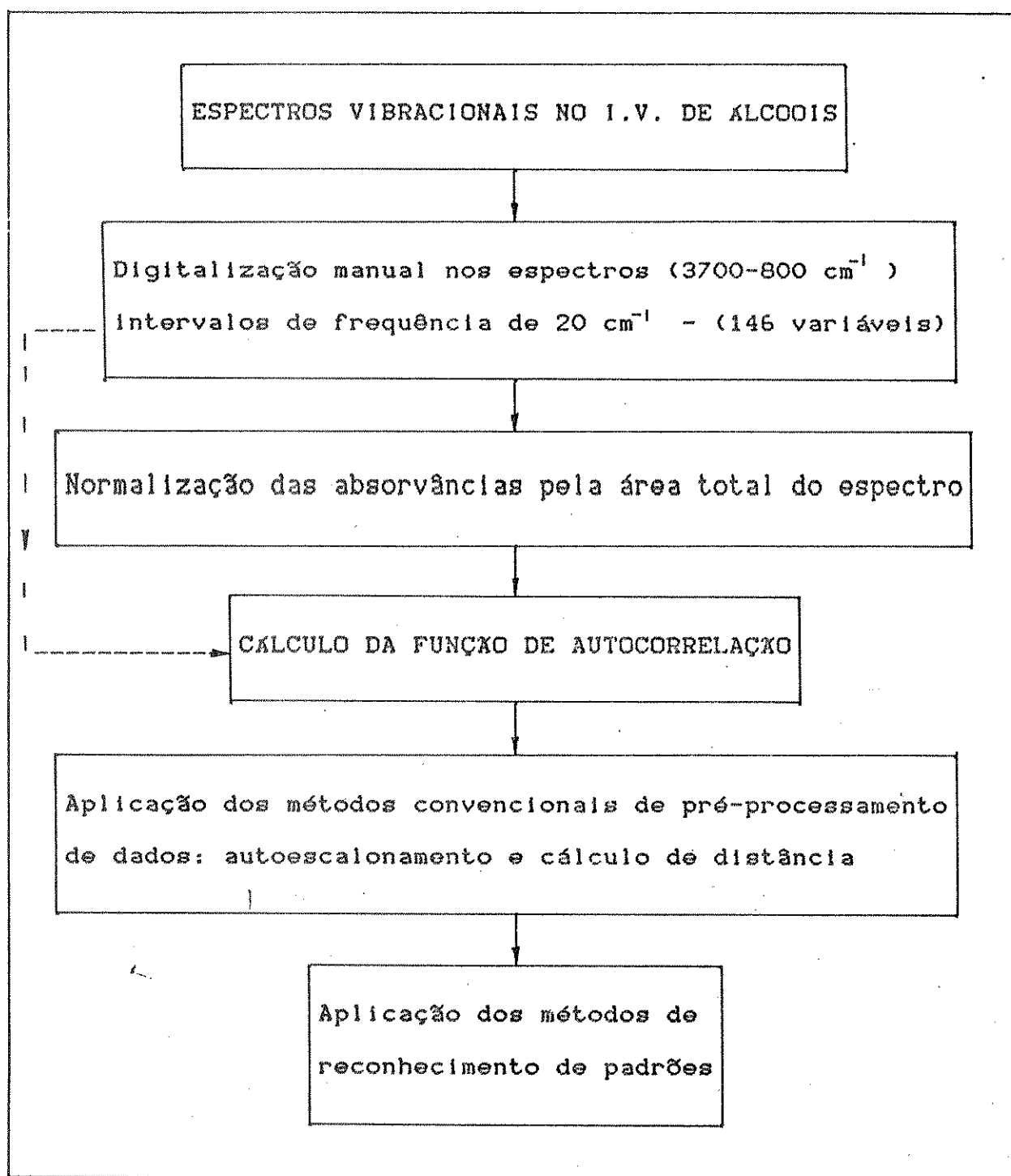
Y<sub>i</sub> = absorvâncias

n = número de variáveis

O cálculo para esta função foi realizado através do programa relacionado no apêndice A, com o nome de AUTCOR.FOR. Todo o procedimento realizado neste pré-processamento para a obtenção dos resultados pode ser melhor observado através de um fluxograma, como mostra a figura 09.







**FIGURA 09:** Fluxograma do método de pré-processamento de dados usando a função de autocorrelação.

No apêndice B, está relacionado o modelo do conjunto de dados em forma de matriz, inserido no microcomputador para posterior aplicação dos métodos de pré-processamento e de RP.

Em seguida discutiremos algumas subrotinas pertencentes ao programa ARTHUR, que também foram aplicadas neste trabalho no pré-processamento de dados. Os métodos executados por estas subrotinas, são comumente usados no pré-processamento de dados em quimiometria:

1) CHANGE: permite uma variedade de mudanças nos conjuntos de dados, envolvendo a escolha de variáveis, redefinição de categorias e mudanças de arquivos de dados. Esta subrotina foi aplicada somente nos dados pré-processados pelo primeiro método desenvolvido, para transformar os valores numéricos das variáveis absolutas (absorvâncias) em variáveis relativas.

2) SCAL: emprega-se o autoescalamento a fim de que cada variável receba o mesmo peso estatístico inicial nos cálculos, isto é, serve para que todas as variáveis sejam iguais em grandeza, sem destruir sua informação de discriminação. As medidas são escalonadas de modo que a média seja igual a zero e a variância é igual a um. Então, as coordenadas k-ésima do ponto i-ésimo tornam-se:

$$\bar{Y}_i = \sum_{k=1}^P Y_{ik} / N \quad (6)$$

$$\sigma_i = \left[ \sum_{k=1}^P (Y_{ik} - \bar{Y}_i)^2 \right]^{1/2} / (N-1)^{1/2} \quad (7)$$

onde:

$\bar{Y}_i$  = a média para a variável  $i$

$N$  = número de amostras no conjunto de treinamento

$\sigma_i$  = variância das amostras

A transformação destas equações pode ser escrita como:

$$Y'_{ik} = (Y_{ik} - \bar{Y}_i) / (N)^{1/2} \times \left[ \sum_{k=1}^p (Y_{ik} - \bar{Y}_i)^2 / N \right]^{1/2} \quad (8)$$

$$\sigma'_i = \left[ \sum_{k=1}^p (Y'_{ik} - \bar{Y}'_i)^2 \right]^{1/2} / (N-1)^{1/2} \quad (9)$$

$$\bar{Y}'_i = \sum_{k=1}^p Y'_{ik} / N$$

onde:

$\bar{Y}'_i$  = média dos dados transformados ( $\bar{Y}_i=0$ )

$\sigma'_i$  = variância dos dados transformados ( $\sigma'_i=1$ )

Esta subrotina foi aplicada aos três métodos de pré-processamento de dados discutidos anteriormente.

3) WEIGHT: o peso das variáveis é um outro tratamento de pré-processamento, aplicado aos dados já pré-processados anteriormente pelo autoescalamento ou nos dados originais.

Com este tratamento, às vezes usado para ponderação, pode-se avaliar a importância individual (peso) de cada variável na discriminação entre cada par de categorias, ou seja, determina quais destas medidas são relevantes na separação das categorias, enfatizando-as e quais poderão ser desprezadas para esta finalidade. Neste trabalho calculamos os pesos para ter uma medida da importância de cada variável (usando somente os dados de frequên-

cias e absorvâncias relativas) na discriminação das categorias. Entretanto, os dados não foram ponderados usando estes resultados. Quando se tem três ou mais categorias, as médias destes pesos são também calculadas para cada variável. Há vários métodos para esta ponderação, mas os que mais se destacam são: a) peso de Fisher e b) peso de variância.

a) Peso de Fisher: estima quantitativamente a importância de uma medida para distinguir entre duas categorias. Para a variável  $i$  e para as amostras nas categorias  $r$  e  $s$ , este pode ser calculado conforme a equação:

$$WFrs(i) = \frac{[\bar{X}_i(r) - \bar{X}_i(s)]^2}{S^2_i(r) + S^2_i(s)} \quad (10)$$

$$S^2_i(r) = \frac{[\sum_{k=1}^p Y^2_{ik}(r)]Nr - [\sum_{k=1}^p Y_{ik}(r)]^2}{(Nr-1)}$$

onde:

$\bar{X}_i$  = média da variável  $i$

$S^2_i(r)$  = variância da variável  $i$  para amostra da categoria  $r$

$r$  e  $s$  = categorias das amostras

$Nr$  = número de amostras na categoria  $r$

O valor do peso de Fisher aumenta com o aumento do numerador da equação (10), ou seja, quanto mais separados os valores médios da variável para as duas classes, mais importante será essa variável para discriminar os objetos das classes  $r$  e  $s$ . O valor do peso de Fisher também aumenta com a diminuição da soma das

variâncias dentro de cada categoria, assim, quanto menor esta soma, maior o peso de Fisher e maior o potencial de discriminação da variável  $i$  para as classes  $r$  e  $s$ .

b) Peso de variância: este peso tem a mesma função que o de Fisher. Para as classes  $r$  e  $s$ , o peso de variância é calculado dividindo-se a variância interclasse pela soma da variância intraclasse, conforme a equação:

$$WVrs(i) = \frac{\sum_{k=1}^p Y^2_{i(r)}/Nr + \sum_{k=1}^p Y^2_{i(s)}/Ns - 2 \sum_{k=1}^p Y_{i(r)}/Ns \sum_{k=1}^p Y_{i(r)}/N}{S^2_{i(r)} + S^2_{i(s)}} \quad (11)$$

onde:

$$S^2_{i(r)} = \frac{\sum_{k=1}^p Y^2_{i(r)}Nr - [\sum_{k=1}^p Y_{i(s)}]^2}{(Nr)^2}$$

4) DISTAN: fornece a matriz de distâncias entre amostras, tendo significado quando supõe-se que cada amostra pode ser considerada como um ponto no espaço  $p$ -dimensional, cujas coordenadas correspondem às variáveis envolvidas. A matriz de distância é guardada num arquivo para ser utilizada por outras subrotinas. Vários tipos de distâncias[2] podem ser calculados como: distância euclidiana, Mahalonobis generalizada, quadra de cidade e distância de Anders. Neste trabalho apenas a primeira foi calculada.

## CAPÍTULO IV

### MÉTODOS DE RECONHECIMENTO DE PADRÕES

A atuação do RP em sua forma mais simples se faz usando dados multivariados, medidos num conjunto de treinamento de objetos pertencentes a um número de classes especificadas, em forma sequencial, de modo a encontrar as amostras representativas de cada classe. Em seguida, discutiremos alguns métodos de RP normalmente aplicados para interpretação de dados químicos.

#### IV.1. ANÁLISE DE DISCRIMINANTE LINEAR (LDA)

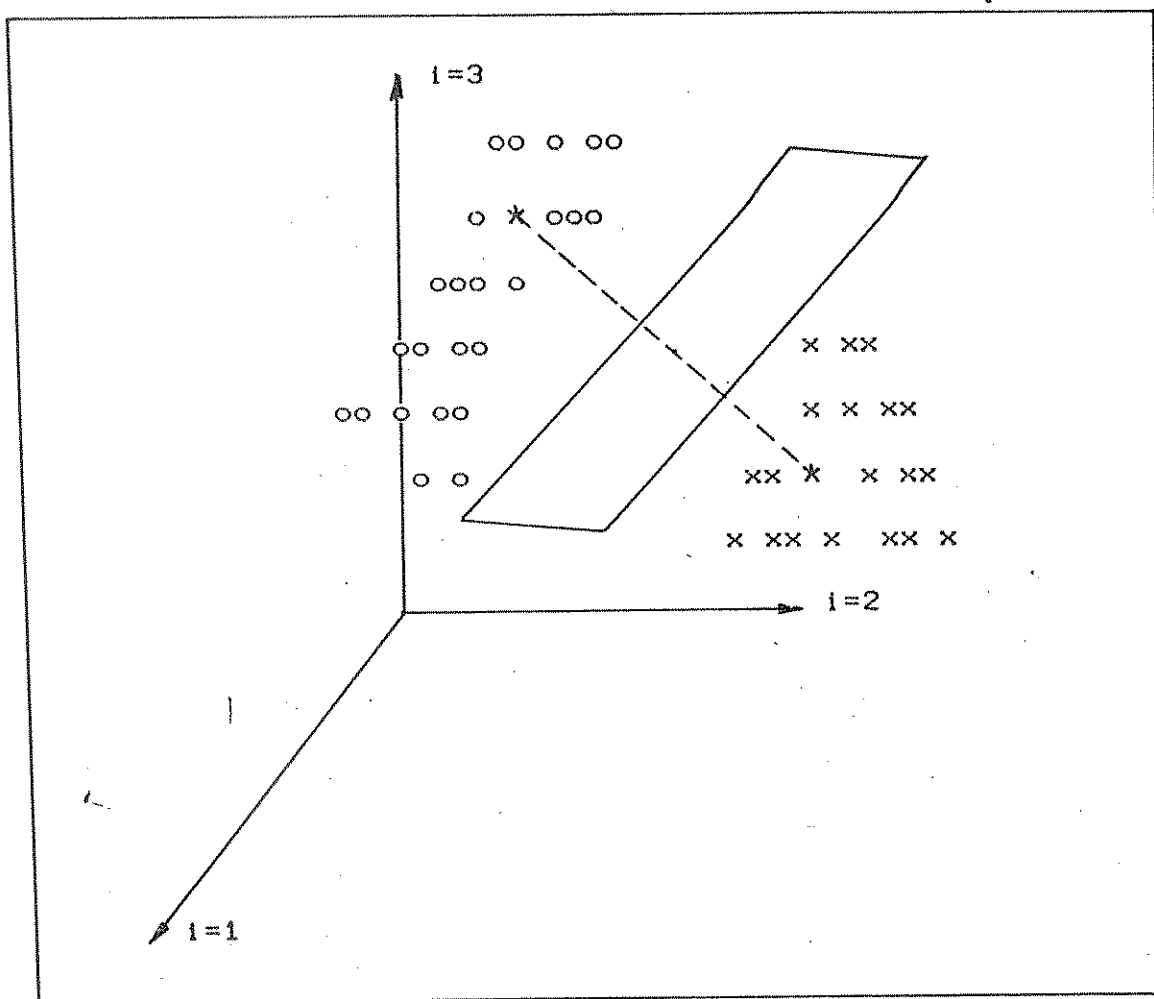
A análise de discriminante linear (LDA)\* baseia-se numa função discriminante ou plano de decisão, determinado por uma combinação linear de variáveis, que permite uma distinção entre dois pontos pertencentes a duas categorias diferentes. Esta análise não é difícil de se estender para  $r$  categorias, utilizando-se  $(r-1)$  funções lineares capazes de separar cada uma das  $r$  categorias das demais. Seus pesos ou coeficientes lineares das variáveis que definem o plano de decisão, são determinados de uma maneira que possam exibir a maior razão das variâncias entre dois grupos relativos àquele dentro dos grupos[29].

Na figura 10, verifica-se uma das maneiras mais simples para se obter uma função linear discriminatória. Determinam-se os

---

\* Do inglês: "Linear Discriminant Analysis"

centros de gravidade para as categorias, unindo-os posteriormente por uma linha reta. O plano de decisão ou função de discriminante é o perpendicular a essa reta, equidistante dos centros de gravidade. Portanto, o espaço fica dividido em duas regiões,  $R_1$  e  $R_2$ . Assim, um ponto, referente a uma categoria desconhecida, que aparece na região  $R_1$ , é classificado como pertencente à categoria 1 e se aparece na região  $R_2$ , à categoria 2.



**FIGURA 10** - Análise de discriminante linear, (o) objetos pertencentes à categoria 1; (x) objetos pertencentes à categoria 2 e (\*) pontos desconhecidos.

## IV.2. MÁQUINA DE APRENDIZAGEM LINEAR (LLM)

A máquina de aprendizagem linear (LLM)\* é um processo iterativo no qual se começa com um plano arbitrário. Durante um determinado número de iterações, o plano está deslocado para colocar todos os pontos pertencentes a uma categoria num lado do plano e todos da outra categoria do outro lado. Este processo pode convergir, caso os dados sejam linearmente separáveis, encontrando um plano de decisão que separa completamente as categorias[30].

Observa-se na figura 11 os pontos relativos a duas classes, separados linearmente por um hiperplano, cuja posição é determinada pelo processo iterativo. Este método é computacionalmente simples, mas pode requerer um grande tempo computacional, por causa das muitas iterações que podem ser necessárias. Uma outra desvantagem é que pode ser encontrado mais do que um plano que separa as duas categorias, dependendo das estimativas do primeiro plano.

Este trabalho envolveu grandes conjuntos de dados mantidos em forma de matrizes, com um número de objetos bem menor do que o número de variáveis. Os métodos de LDA e LLM não foram utilizados neste estudo, pois requerem exatamente o contrário, ou seja, que a relação entre o número de objetos e o número de variáveis sejam maior ou igual a cinco ( $n/p \geq 5$ ).

---

\* Do inglês: "Linear Learning Machine"



Os métodos de RP eficientes para nossos propósitos são: a) regra do vizinho mais próximo (KNN)\* e b) modelos independentes de similaridade utilizando componentes principais (SIMCA)\*\*. Esses métodos foram utilizados para classificar os espectros vibracionais no I.V. de álcoois em suas respectivas categorias, após cada aplicação dos métodos de pré-processamento de dados citados no capítulo anterior.

### IV.3. REGRA DO VIZINHO MAIS PRÓXIMO (KNN)

A regra do vizinho mais próximo (KNN) é um método conceitualmente simples e bastante comum, com melhores propriedades do que a LDA e a LLM. Esse método baseia-se na existência de uma distância sendo inversamente proporcional à similaridade dos objetos. O KNN classifica um objeto com categoria desconhecida, de acordo com sua distância aos objetos do conjunto de treinamento[31].

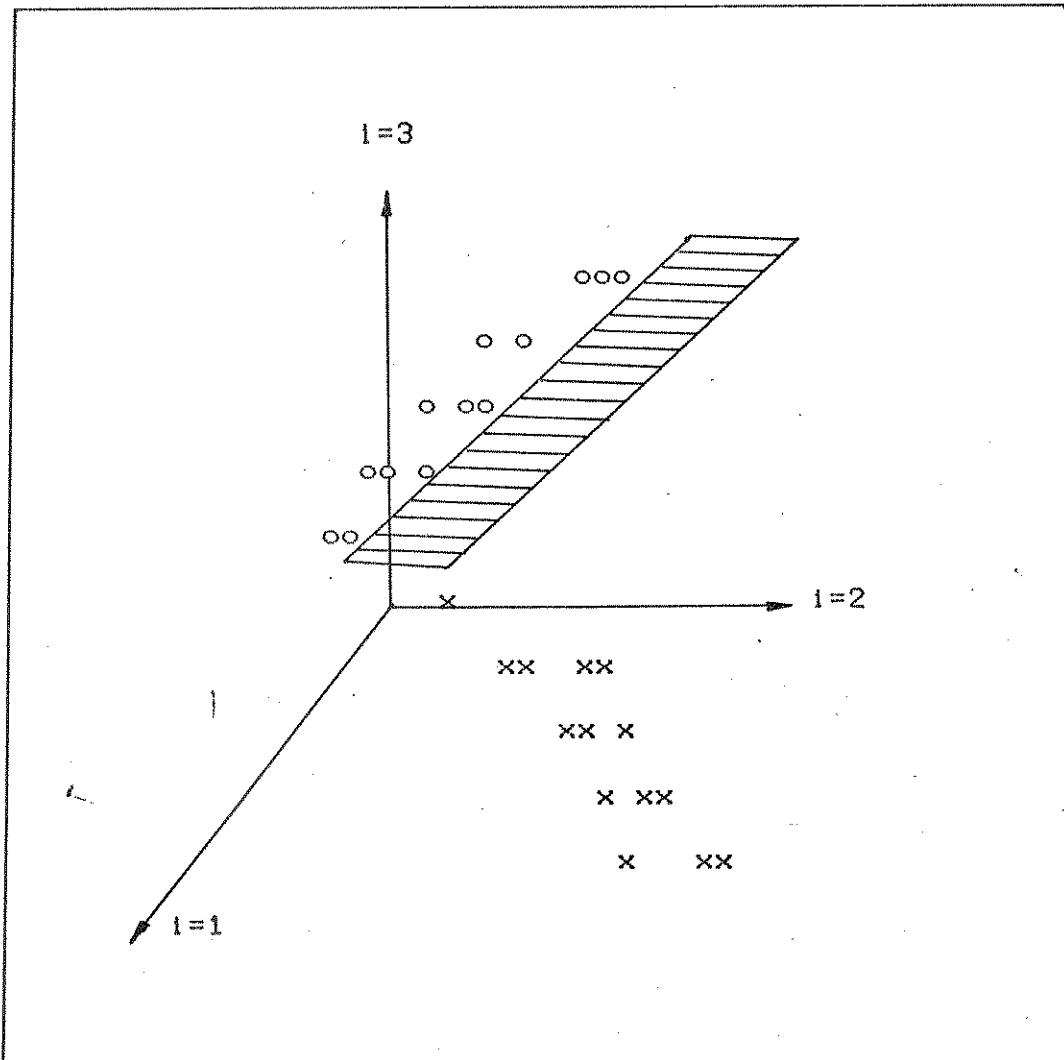
Nesse método não se determina nenhuma função matemática usando o conjunto de treinamento, como se faz nos métodos de LDA e LLM, o que ocorre é a classificação de um ponto na mesma categoria a qual pertence o seu vizinho mais próximo, como ilustra a figura 12. Normalmente, é interessante utilizar mais que um vizinho mais próximo. As rotinas de classificação dos programas de

---

\* Do inglês: "K-Nearest Neighbour"

\*\* Do inglês: "Soft Independent Modelling by Class Analogy"

computação determinam normalmente  $K$  vizinhos mais próximos, sendo  $K=1,3,\dots,10$ . A filosofia do KNN também é diferente da LDA e da LLM. Nestes dois métodos, o que se procura é uma separação entre categorias. O KNN, por outro lado, é um exemplo de modelo de similaridade, onde as amostras do conjunto de treinamento, mais similares à amostra em estudo, servem como referência para sua classificação.



**FIGURA 11** - Máquina de aprendizagem linear, onde as classes são separadas por um hiperplano dimensional  $p-1$ .

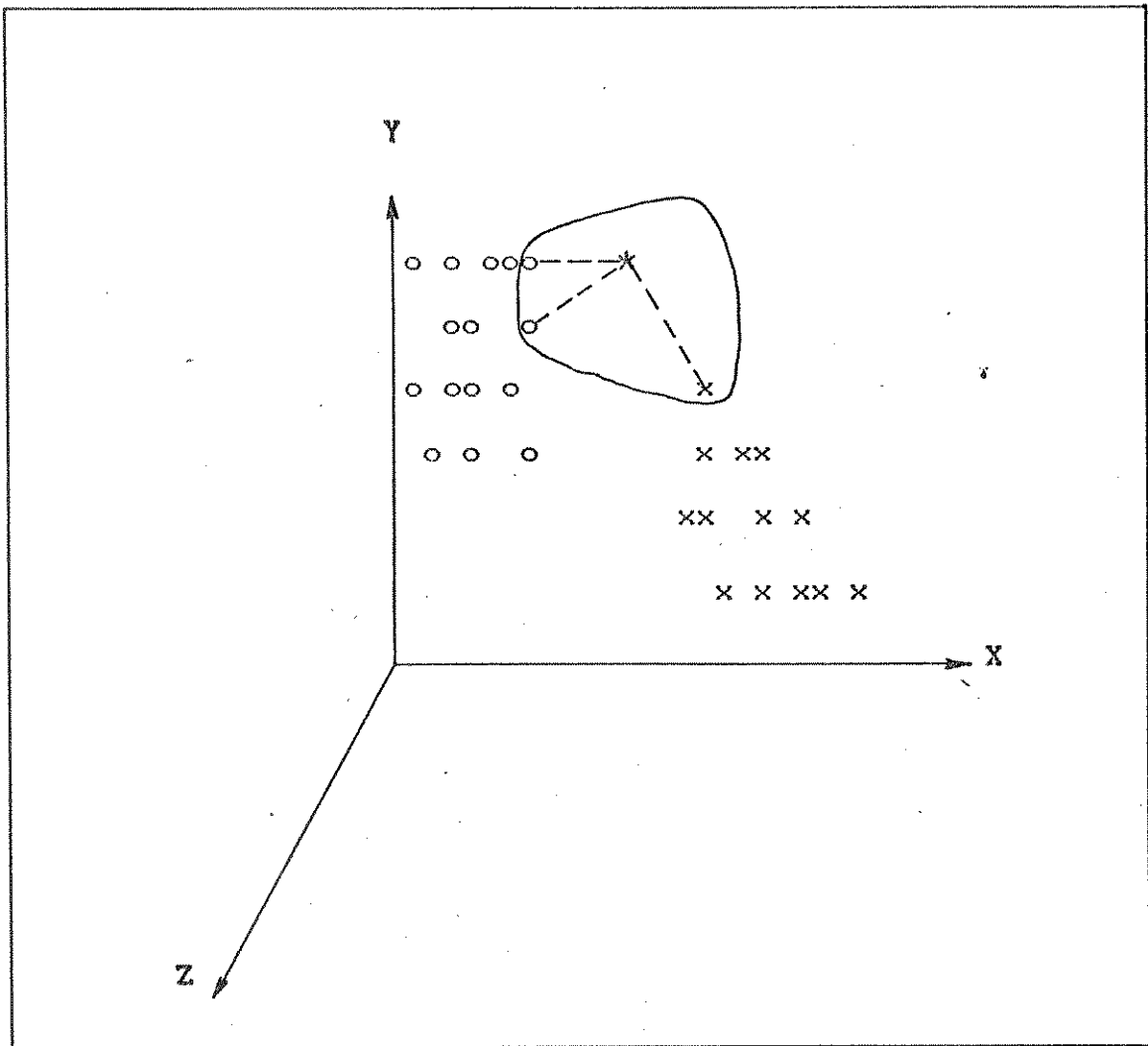


FIGURA 12 - Regra do vizinho mais próximo num espaço tridimensional: (o) objetos do conjunto de treinamento da categoria 1; (x) objetos do conjunto de treinamento da categoria 2; (\*) objeto do conjunto teste.

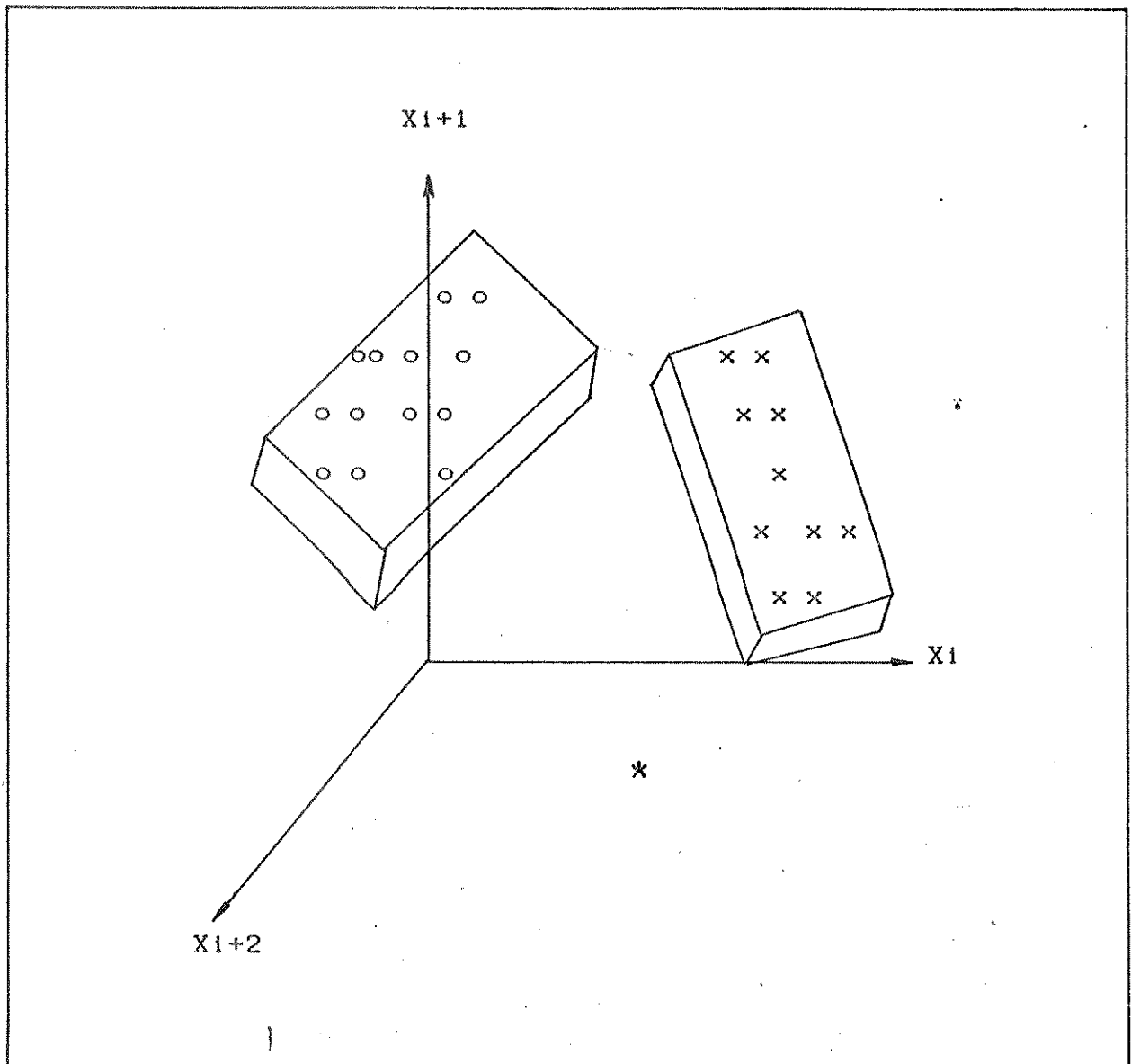
Existem algumas vantagens no método KNN em relação a outros métodos de RP, como por exemplo, a não necessidade em manter a relação entre o número de objetos e o número de variáveis menor

ou igual a cinco. Todavia, esse método nesta forma simples, não é capaz de sugerir que uma amostra possa pertencer a uma categoria ainda não definida.

Esta propriedade é acoplada com o fato que funções matemáticas que discriminam entre categorias, não são produzidas para KNN. Estas funções facilitam a classificação de amostras do conjunto teste.

#### IV.4. MODELOS INDEPENDENTES DE SIMILARIDADE UTILIZANDO COMPONENTES PRINCIPAIS (SIMCA)

O método de RP que desperta muito interesse em estudos quimiométricos é o SIMCA[32]. Este método assume que, todas as classes do conjunto de treinamento estão contidas dentro de envelopes fechados[33]. Os modelos são construídos em envelopes, de modo que cada objeto dentro desses envelopes é considerado um membro normal dessas classes e os objetos fora dos mesmos são considerados "outliers" (pontos deslocados) para todas as classes, como mostra a figura 13. Estes "outliers" podem ser dados anômalos, isto é, resultantes de erros experimentais ou membros de categorias ainda não definidas.



**FIGURA 13** - No método SIMCA todos os pontos situados dentro dos envelopes, são classificados como pertencentes à categoria definida. O ponto (\*) é acusado como um ponto deslocado (outlier) e não pertencente a nenhuma categoria definida.

Os envelopes do método SIMCA são construídos utilizando o modelo de análise de componentes principais (PCA)\*. Em seguida, descreveremos as etapas desse modelo[34]:

a) Representar a aglomeração dos dados pelo seu "ponto médio". As coordenadas deste ponto são as médias das variáveis  $\bar{x}_K$ , as quais juntas formam o vetor linha  $\bar{x}$  da matriz de dados definida anteriormente. Esta etapa está representada na figura 14;

b) Subtrair estas médias dos dados para determinar os resíduos  $e_{ik}$  (elementos da matriz E), que corresponde a um movimento na origem do sistema de coordenada, estando agora centrada no ponto  $\bar{x}$ ;

c) Denominar os resíduos E para X. Agora, temos os dados X escalonados e centrados;

d) Adaptar uma linha reta que passa pela nova origem, aos n pontos no espaço p-dimensional, de modo que os desvios perpendiculares sejam tão pequenos quanto possíveis, no sentido dos mínimos quadrados, como mostra a figura 15. Os coeficientes direcionais desta linha são chamados de "loadings", sendo um para cada variável K e são incluídos no vetor linha  $p_1$  (o primeiro vetor "loading"). Quando um ponto é projetado nesta linha determinamos os "scores"  $t_{i1}$ , isto é, a coordenada de ponto i ao longo do eixo  $p_1$  (figura 15). Subtraindo o produto  $t_i p_k$  de  $X_{ik}$  determinamos os novos resíduos  $e_{ik}$ ;

---

\* Do inglês: "Principal Component Analysis"

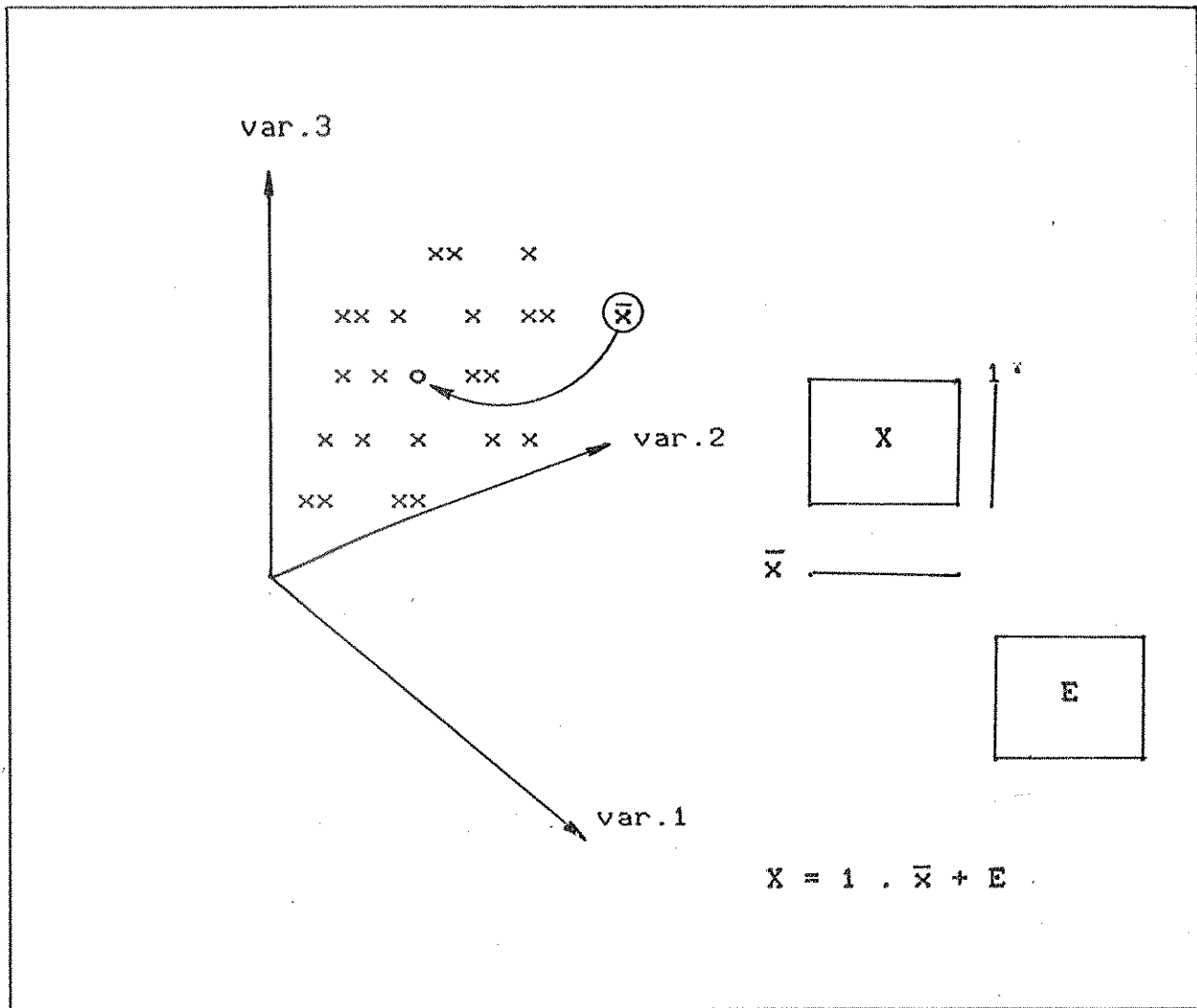
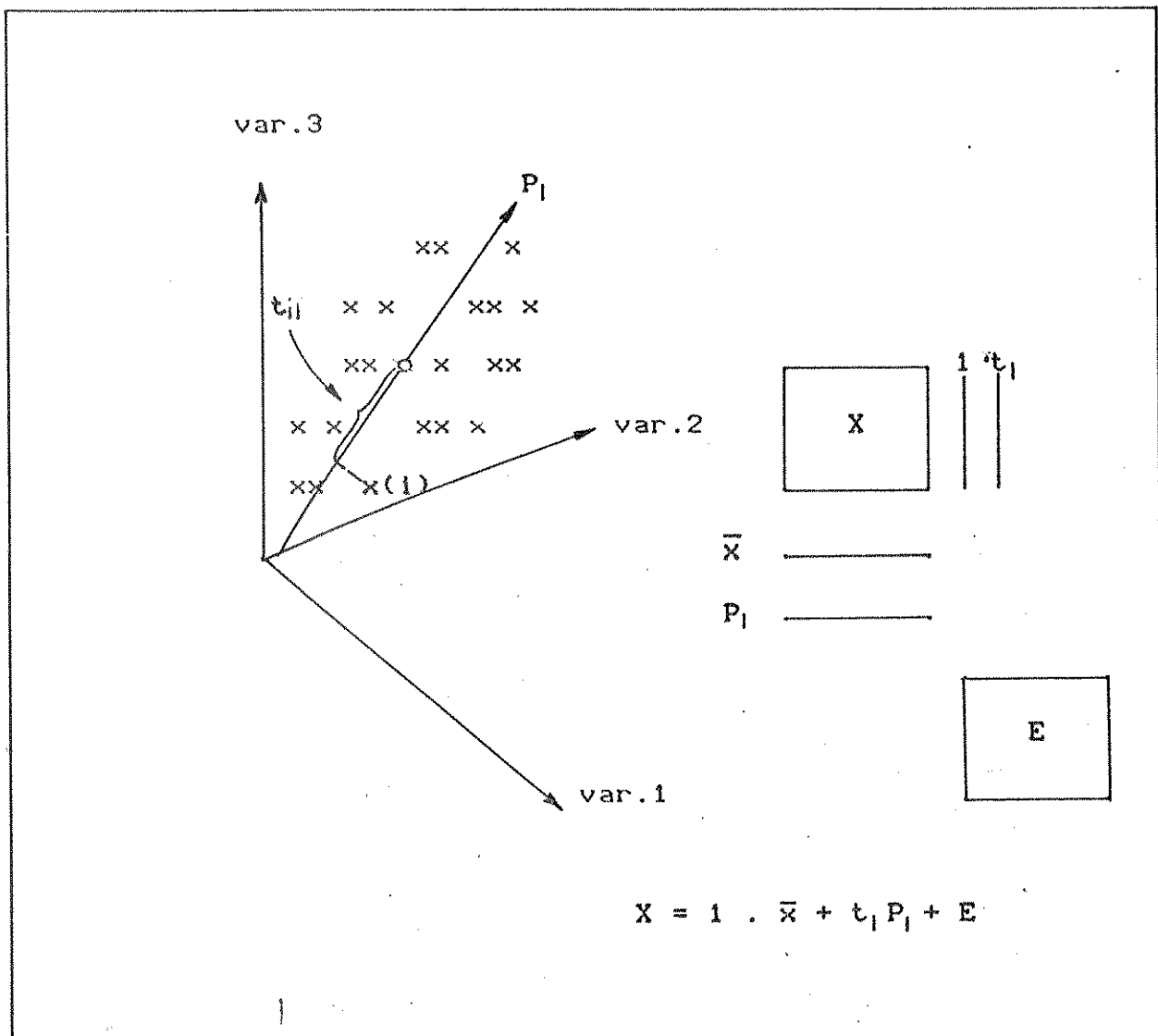


FIGURA 14 - O modelo mais simples é representar os dados do conjunto pelo seu ponto médio,  $\bar{x}$ . Este tem as coordenadas definidas pelas médias das variáveis. A matriz  $E$  contém os resíduos  $e_{ik}$ .



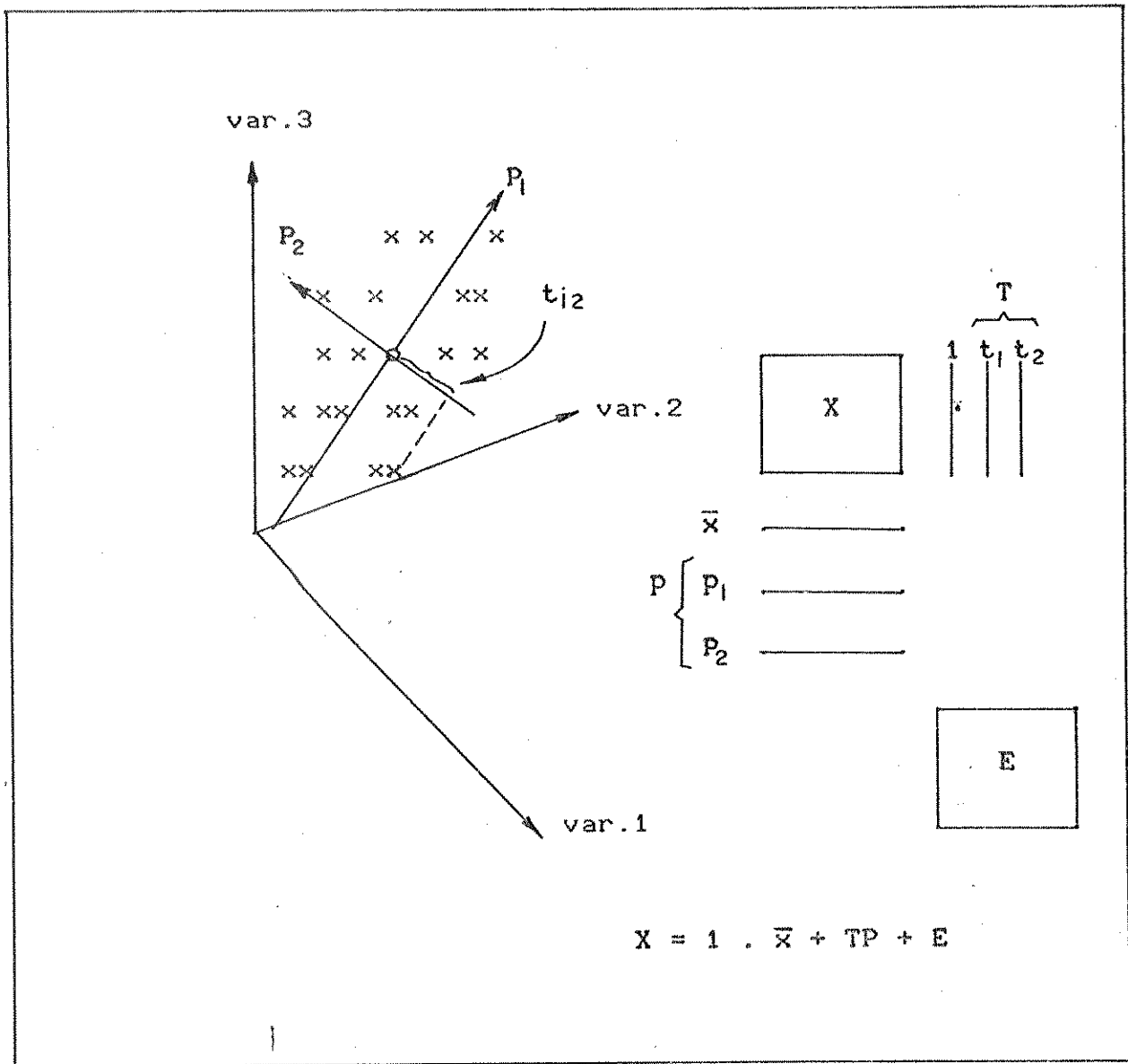
**FIGURA 15** - Os dados são modelados pelos mínimos quadrados por uma linha direta a  $\bar{x}$ . Esta linha é o primeiro componente principal (CP) e sua equação é definida pelo vetor loading  $p_1$ . A projeção do ponto da linha existe na distância  $t_{ik}$  de  $\bar{x}$ .



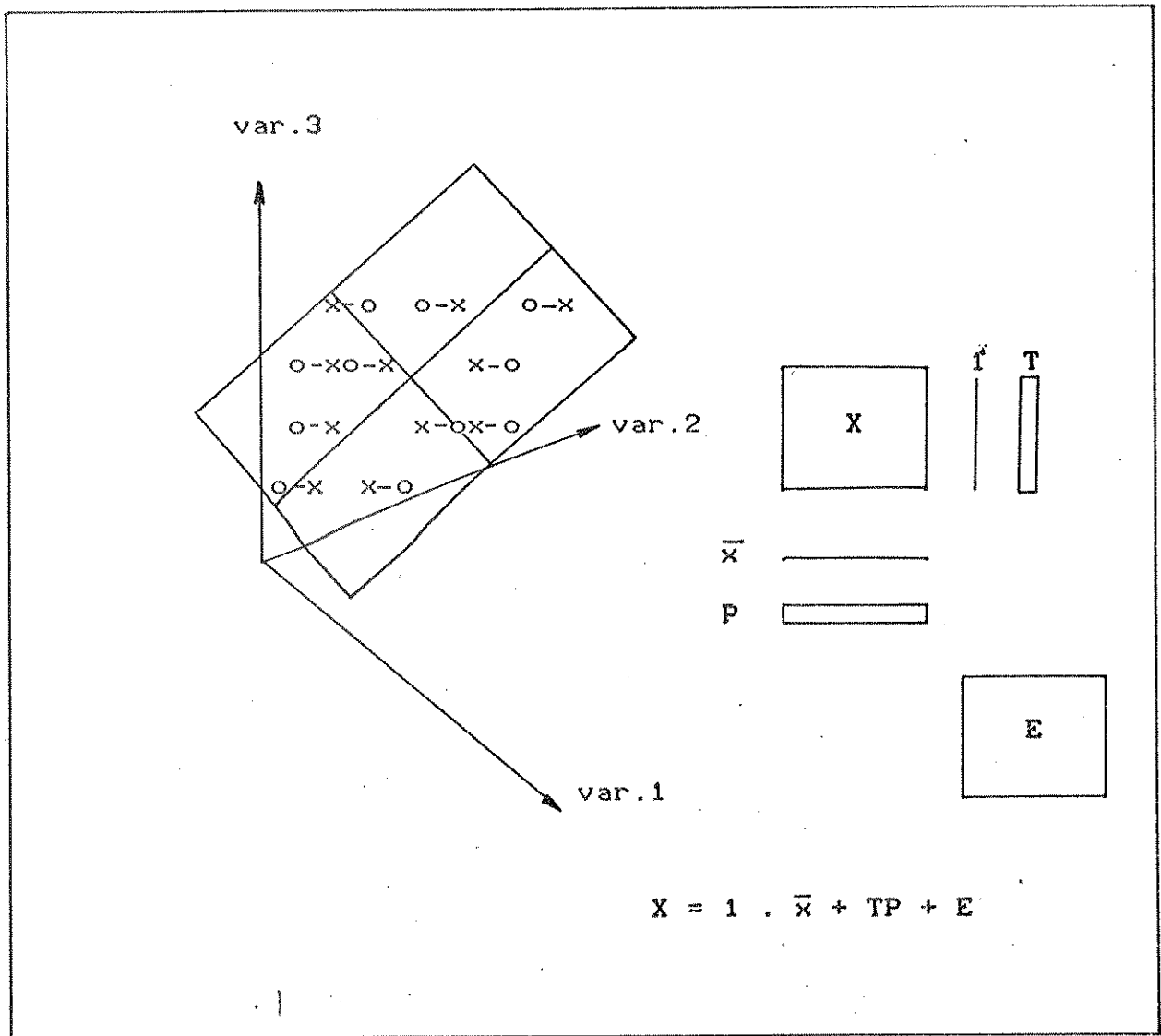
e) Novamente denominar os resíduos E para X, que corresponde em remover a direção  $p_1$  dos dados. Quando esse X é usado para adaptar outra linha reta para seus objetos, estar-se-á adaptando uma segunda linha direta a  $\bar{x}$ , ortogonal à primeira linha, e novamente fazendo os resíduos tão pequenos quanto possíveis (mínimos quadrados). Esta etapa é mostrada na figura 16;

f) O ítem acima pode ser repetido novamente até que, após p ou n fatores (o que vier primeiro), os resíduos sejam identicamente zero. Assim, tem-se construído um novo sistema de coordenadas com eixos H (H é o menor de n e p) e representa os n objetos como pontos neste espaço H;

g) Todavia, a idéia útil com análise de componentes principais, é que podemos usar apenas os primeiros componentes A para representar X. Isso é visto numa projeção de X sob o espaço A-dimensional, representado na figura 17, com  $A=2$ . Agora, o plano nesta figura tem o caráter de uma janela de A-dimensões no espaço-p.



**FIGURA 16** - O segundo componente principal é uma linha ( $p_2$ ) direta a  $\bar{x}$ , perpendicular a  $p_1$ .



**FIGURA 17** - Juntos, o primeiro e o segundo CP definem um plano. Assim, tem-se uma janela que conduz a olhar no espaço p e ver uma figura bidimensional da configuração do ponto objeto.

O método SIMCA classifica as amostras baseado em suas similaridades ao modelo de componente principal de cada categoria, de modo que a maior parte das variáveis expresse essa similaridade[35]. Calcula-se para cada categoria os componentes principais (CP), que servem como modelos para várias categorias.

A redução do número de variáveis  $p$  para  $A$  componentes principais pode ser visualizada como descreveremos em seguida. De acordo com a dispersão dos pontos no sistema de coordenadas original, como ilustra a figura 18, tanto a dimensão  $X_1$  como a  $X_2$  são importantes, visto que, parte significativa da variância das amostras está representada em cada eixo segundo uma relação linear. Como um componente principal vai definir esta linha de um componente principal, este modelo é capaz de descrever adequadamente a estrutura dos dados.

O cálculo dos componentes principais é convenientemente feito através da multiplicação da matriz de dados original  $\underline{X}$  pela sua transposta  $\underline{X}^t$ , obtendo-se a matriz de covariância  $\underline{C}$ , que é posteriormente diagonalizada fornecendo a posição do novo sistema de coordenadas original, conforme as equações:

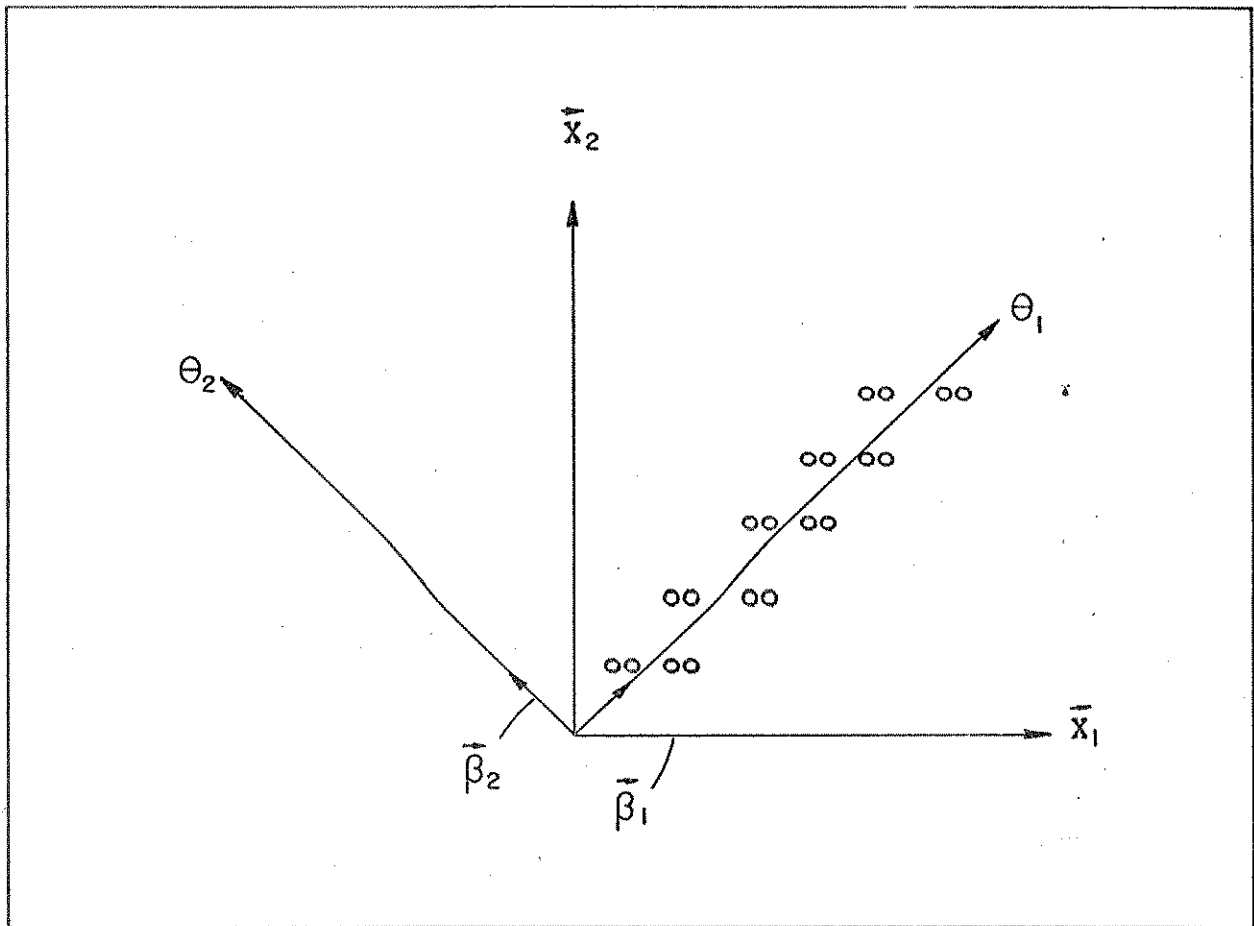
$$\underline{C} = \underline{X}^t \cdot \underline{X} \quad (12)$$

$$\underline{C} \cdot \vec{\beta}^t = \lambda_j \cdot \vec{\beta}^t \quad (13)$$

onde:

$\vec{\beta}^t$  = autovetor correspondente ao  $j$ -ésimo autovalor compondo a  $j$ -ésima coluna da matriz.

$\lambda_j$  = autovalor



**FIGURA 1B** - Sistema de coordenadas original com as dimensões  $X_1$  e  $X_2$ , representando em cada eixo parte significativa da variância das amostras.

O  $j$ -ésimo autovalor traduz a fração da variância original expressa pelo  $j$ -ésimo autovetor. Esta diagonalização da matriz de covariância permite a projeção de um espaço de  $p$  variáveis num outro de ordem inferior, em geral de duas ou três dimensões, a fim de facilitar a visualização, que é de grande importância em problemas classificatórios.

Para o caso bidimensional (2 CP) apresentado na figura 18, o autovalor  $\lambda_1$ , expresso pelo eixo do CP  $\vec{\beta}_1$ , representa cerca de 90% da variância total e os 10% restantes atribuídos a  $\vec{\beta}_2$ . Como  $\vec{\beta}_1$  e  $\vec{\beta}_2$  são autovetores, devem ser mutuamente ortogonais e os eixos devem guardar entre si um ângulo de 90°. Se um componente principal fosse usado para descrever os dados da figura 18, somente 10% de informação estatística seria desprezada. Muitas vezes, esta quantidade de variância corresponde ao ruído no problema que está sendo estudado.

Em geral, para estruturas de dados muito complexas, utiliza-se um método com A componentes, conforme a equação:

$$X_{ik} = \bar{x}_k + \sum_{a=1}^A \theta_{ia} \beta_{ak} + e_{ik} \quad (14)$$

onde:

A = número de componentes principais

$\bar{x}_k$  = valor médio da k-ésima variável

$e_{ik}$  = desvios devido aos erros aleatórios nas medidas

Essa equação representa uma hipercaixa no espaço p. O valor do objeto i projetado no eixo do componente principal a é dado por  $\theta_{ia}$ , enquanto  $\beta_{ak}$  representa a velocidade de variação da variável k devida a variações unitárias nos valores do a-ésimo componente principal.

Os parâmetros  $\bar{x}$ ,  $\beta$  e  $\theta$  descrevem a parte sistemática de X. Nota-se que o modelo de CP separa as p variáveis dos n objetos dentro de parâmetros separados, com  $\bar{x}$  e  $\beta$  para os primeiros e  $\theta$  para o último.

Os resíduos descrevem a parte aleatória de X. Esta parte consiste de: 1) erros de medidas e outra imprecisão e 2) erros de modelagem, isto é, imperfeição no modelo calculado. O desvio padrão residual mede a distância típica entre os modelos de classes e um objeto pertencente à classe. Os resíduos conforme a equação:

$$S_0^2 = \sum_{k=1}^n \sum_{k=1}^p e^2_{ik} / (n-A)(p-A-1) \quad (15)$$

determinam o tamanho dos envelopes, onde o denominador representa o número de graus de liberdade, também os resíduos podem ser usados para calcular a importância de cada variável k no modelo e detectar "outliers", objetos anômalos no conjunto de treinamento.

Os dados do conjunto de treinamento são utilizados para determinar o modelo de cada categoria, que são totalmente independentes das outras categorias. Para a determinação do número apropriado de componentes principais, pode-se utilizar técnicas de "cross-validation". Concluída a modelagem, resta classificar os pontos correspondentes a amostras desconhecidas. A segunda idéia deste método é que os objetos no conjunto teste podem ser classificados de acordo com seu grau de similaridade aos modelos de classes.

Em resumo, o método SIMCA apresenta várias vantagens em relação a outros métodos. Por exemplo, de classificar um ponto referente a uma amostra, quando este não pertence a nenhuma das categorias classificadas, indicando-o como um ponto deslocado (outlier) ou membro de uma categoria ainda não definida. Outra vantagem é a de que pode ser aplicado aos casos onde o número de

amostras  $n$  é bem menor do que as variáveis  $p$ , uma vez que o número de graus de liberdade dado por  $(n-A)(p-A-1)$  aumenta tanto com o número de amostras quanto com o número de variáveis. Finalmente, o método fornece equações matemáticas, que podem ser facilmente utilizadas para descrever as diferentes categorias.



## CAPÍTULO V

### RESULTADOS E DISCUSSÕES

No método de pré-processamento de dados usando atribuições de bandas, foram analisados trinta e sete espectros vibracionais no I.V. de álcoois, tomados da biblioteca SADTLER e separados em dois conjuntos: 1) o conjunto de treinamento com vinte e quatro amostras de álcoois alifáticos saturados, distribuídos da seguinte maneira: dez primários, sete secundários e sete terciários, 2) o conjunto teste com treze amostras de álcoois alifáticos insaturados, com apenas uma ligação dupla, exceto um álcool terciário com duas ligações duplas, assim distribuídos: dois primários, quatro secundários e sete terciários.

Como não temos certeza que os modelos de classificação, calculados com base nos dados de álcoois saturados, são apropriados para a classificação de álcoois insaturados, com desprezo da banda de estiramento C=C, os relativos méritos dos diferentes procedimentos de pré-processamento serão julgados utilizando principalmente os resultados obtidos para o conjunto de treinamento. Os resultados do conjunto teste serão usados para confirmar as conclusões deduzidas dos dados do conjunto de treinamento.

Para esse método de atribuições de bandas foram utilizadas duas matrizes de dados de dimensões: 24x12 para o conjunto de treinamento e 13x12 para o conjunto teste. Nessas matrizes apli-

camos as equações para calcular os pesos de Fisher e de variância, sendo os resultados apresentados na tabela 03. Através deste cálculo, observamos que os valores das frequências têm pesos maiores do que aqueles para absorvâncias relativas, portanto, são mais importantes na discriminação dos álcoois, destacando-se principalmente nos terciários das outras duas categorias. As frequências que mais pesam nessa separação estão na região de estiramento O-H ( $3550-3200\text{ cm}^{-1}$ ) e de estiramento C-O ( $1260-1000\text{ cm}^{-1}$ ). Isto é surpreendente, considerando a largura da banda de estiramento O-H que pode causar dificuldades na definição das frequências usadas. Quanto às absorvâncias relativas, destacaram-se mais discriminando os álcoois primários dos outros na região de deformação O-H ( $1420-1330\text{ cm}^{-1}$ ) e de estiramento C-O.

O método KNN com até cinco vizinhos mais próximos, classificou corretamente com um índice médio de aproximadamente 72% no conjunto de treinamento, todavia, no conjunto teste o mesmo apresentou uma discriminação considerada baixa, com um índice médio de classificação correta de aproximadamente 56%, como mostra em mais detalhes a tabela 04.

Na mesma matriz de dados foi aplicado o método SIMCA, utilizando dois componentes principais (2 CP) para cada categoria. Esse método forneceu um índice de 100% de classificação correta nas categorias primária, secundária e terciária para o conjunto de treinamento. Entretanto, o conjunto teste apresentou um índice de aproximadamente 70%, classificando incorretamente um álcool primário (categoria 1), um álcool secundário (categoria 2) e dois álcoois terciários (categoria 3). Os resultados obtidos

com esse método podem ser observados na tabela 05.

**TABELA 03** - Resultados dos pesos de Fisher e variância nas matrizes 24x12 para conjunto de treinamento e 13x12 para o conjunto teste (atribuições de bandas).

variável	categ. x categ.	PESOS	
		Fisher	Variância
1. frequência de estiramento O-H	1 x 2	0,0155	1,156
	1 x 3	0,1895	2,740
	2 x 3	0,1566	2,279
5. frequência de estiramento C-O	1 x 2	0,0132	1,135
	1 x 3	0,1893	2,822
	2 x 3	0,1807	2,476
10. absorvância relativa na região de deformação O-H	1 x 2	0,0570	1,551
	1 x 3	0,2085	3,104
	2 x 3	0,0463	1,378
12. absorvância relativa na região de estiramento C-O	1 x 2	0,0165	1,173
	1 x 3	0,0715	1,789
	2 x 3	0,0710	1,580

Variáveis: 1 e 5 - utilizando as frequências

10 e 12 - utilizando as absorvâncias relativas

TABELA 04 - Resultados do método KNN nas matrizes 24x12 conjunto de treinamento (4a) e 13x12 conjunto teste (4b) - atribuições de bandas.

Tabela 4a - dados autoescalados

categorias	Nº de amostras	Nº de pontos classificados incorretamente			
		1-NN	3-NN	4-NN	5-NN
1	10	1	2	2	3
2	7	4	3	3	3
3	7	1	2	1	2
Total:	24	6	7	6	8
% classificação correta:		75,0	70,8	75,0	66,7

Tabela 4b - dados autoescalados

categorias	Nº de amostras	Nº de pontos classificados incorretamente			
		1-NN	3-NN	4-NN	5-NN
1	2	0	1	0	1
2	4	3	4	4	4
3	7	2	2	2	0
Total:	13	5	7	6	5
% classificação correta:		61,5	46,2	53,8	61,5

**TABELA 05** - Resultados do método SIMCA utilizando 2 CP para cada categoria, nas matrizes 24x12 para o conjunto de treinamento (5a) e 13x12 para o conjunto teste (5b) - atribuições de bandas.

Tabela 5a - dados autoescalados

categorias	Nº de amostras	Nº de pontos classificados
		incorretamente
1	10	0
2	7	0
3	7	0
<b>Total:</b>	<b>24</b>	<b>0</b>
<b>% classificação correta: 100,00</b>		

Tabela 5b - dados autoescalados

categorias	Nº de amostras	Nº de pontos classificados
		incorretamente
1	2	1
2	4	1
3	7	2
<b>Total:</b>	<b>13</b>	<b>4</b>
<b>% classificação correta: 69,23</b>		

Após a digitalização dos espectros, os dados originais (absorvâncias) foram inseridos no microcomputador DICON e posteriormente aplicados os cálculos de normalização, semivariograma, função de autocorrelação, métodos de pré-processamento de dados convencionais e os métodos de RP, com o objetivo de analisar os resultados das seguintes alternativas: A) normalizar e autoescalonar, B) normalizar e não autoescalonar, C) não normalizar e autoescalonar e, D) não normalizar e não autoescalonar.

Na tabela 06 estão relacionados os resultados obtidos com a matriz de dados de dimensão 24x146 para o conjunto de treinamento e 13x146 para o conjunto teste, utilizando as absorvâncias, através da aplicação dos métodos KNN com até cinco vizinhos mais próximos (5-NN) e o SIMCA utilizando 2 CP para cada categoria.

Através desta tabela, verificamos que o método KNN forneceu a melhor classificação entre os álcoois primários, secundários e terciários, com um índice médio superior a 78% no conjunto de treinamento e de aproximadamente 77% no conjunto teste, quando os dados foram normalizados mas não autoescaloados (alternativa B). Esses resultados estão relacionados com mais detalhes na tabela 07, especificando o número de álcoois classificados incorretamente para cada categoria.

Devido ao grande número de variáveis para cada objeto nessas matrizes de dados, o procedimento de "cross-validation" torna-se lento, por isso, inicialmente aplicamos o método SIMCA utilizando 2 CP para cada categoria. Os melhores resultados foram obtidos quando os dados foram normalizados e autoescaloados (al-

ternativa A), apresentando um índice de 100% de classificação correta no conjunto treinamento e de 77% no conjunto teste, classificando incorretamente três álcoois secundários (categoria 2). Deve ser notado que o resultado para o conjunto de treinamento, com os dados não normalizados mas autoescalados (alternativa C), é ligeiramente inferior com relação aos obtidos com a alternativa A, sendo que para o conjunto teste os mesmos são bem inferiores. Também aplicamos esse método utilizando apenas 1 CP para cada categoria, o qual classificou incorretamente dois álcoois no conjunto de treinamento e quatro no conjunto teste. Isto significa que o modelo com 2 CP para a classificação dos álcoois em suas respectivas categorias, talvez seja um pouco melhor do que o uso de modelo com 1 CP. Na tabela 08, estão comparados os resultados de classificação para modelos de 1 e 2 CP para cada categoria, especificando o número de álcoois classificados incorretamente em cada categoria.

Considerando que o autoescalamento é um pré-requisito para modelagem usando o método SIMCA, é de se esperar melhores índices de classificação correta, obtidos através das alternativas A e C. Por outro lado, isto não é o caso do método KNN, onde os resultados para o conjunto de treinamento são melhores utilizando a alternativa B. De acordo com os resultados observados nas tabelas 06, 07 e 08, verificamos que a normalização das absorvâncias pela área total do respectivo espectro em estudo, revela-se ser um importante procedimento de pré-processamento de dados antes de aplicar o método KNN, mas não necessariamente para o método SIMCA, cujos dados devem ser principalmente autoescalados.

**TABELA 06** - Resultados dos métodos KNN e SIMCA usando 2 CP para cada categoria, nas matrizes 24x146 e 13x146 (absorvâncias), através das seguintes alternativas:.

- A) dados normalizados e autoescalonados
- B) dados normalizados e não autoescalonados
- C) dados não normalizados e autoescalonados
- D) dados não normalizados e não autoescalonados

DADOS	CONJUNTOS	Nº de pontos classificados incorretamente				
		1-NN	3-NN	4-NN	5-NN	SIMCA
A	treinamento	9	8	11	12	0
	% correta	62,5	66,7	54,2	50,0	100,00
	teste	3	3	3	6	3
	% correta	76,9	76,9	76,9	53,8	76,92
B	treinamento	6	6	5	4	9
	% correta	75,0	75,0	79,2	83,3	62,50
	teste	3	3	4	3	3
	% correta	76,9	76,9	69,2	76,9	76,92
C	treinamento	12	16	15	13	1
	% correta	50,0	33,3	37,5	45,8	95,83
	teste	5	6	6	6	6
	% correta	61,5	53,8	53,8	53,8	53,85
D	treinamento	9	9	9	9	8
	% correta	62,5	62,5	62,5	62,5	66,67
	teste	4	4	4	5	9
	% correta	69,2	69,2	69,2	61,5	30,77



TABELA 07 - Resultados do método KNN nas matrizes 24x146 para o conjunto de treinamento (7a) e 13x146 para o conjunto teste (7b), utilizando as absorvâncias.

Tabela 7a - dados normalizados mas não autoescalados

categorias	Nº de amostras	Nº de pontos classificados incorretamente			
		1-NN	3-NN	4-NN	5-NN
1	10	2	3	2	1
2	7	2	1	1	1
3	7	2	2	2	2
<b>Total:</b>	<b>24</b>	<b>6</b>	<b>6</b>	<b>5</b>	<b>4</b>
<b>% classificação correta:</b>		<b>75,0</b>	<b>75,0</b>	<b>79,2</b>	<b>83,3</b>

Tabela 7b - dados normalizados mas não autoescalados

categorias	Nº de amostras	Nº de pontos classificados incorretamente			
		1-NN	3-NN	4-NN	5-NN
1	2	0	0	0	0
2	4	3	3	4	3
3	7	0	0	0	0
<b>Total:</b>	<b>13</b>	<b>3</b>	<b>3</b>	<b>4</b>	<b>3</b>
<b>% classificação correta:</b>		<b>76,9</b>	<b>76,9</b>	<b>69,2</b>	<b>76,9</b>

TABELA 08 - Resultados do método SIMCA nas matrizes 24x146 para o conjunto de treinamento (8a) e 13x146 para o conjunto teste (8b), utilizando as absorvâncias.

Tabela 8a - dados normalizados e autoescalonados

categorias	Nº de amostras	Nº de pontos classificados incorretamente	
		2 CP	1 CP
1	10	0	1
2	7	0	0
3	7	0	1
Total:	24	0	2
% classificação correta:		100,00	91,67

Tabela 8b - dados normalizados e autoescalonados

categorias	Nº de amostras	Nº de pontos classificados incorretamente	
		2 CP	1 CP
1	2	0	0
2	4	3	3
3	7	0	1
Total:	13	3	4
% classificação correta:		76,92	69,23

Os dados pré-processados pelo semivariograma apresentam o último valor de  $\gamma^*(h)$  indeterminado, quando a diferença entre o número de variáveis (N) e a distância (h) é igual a zero (equação 04). Portanto, as matrizes de dados consistiram de cento e quarenta e cinco valores da função de semivariograma, gerando assim, uma matriz de dados de dimensão 24x145 para o conjunto de treinamento e 13x145 para o conjunto teste.

Com os resultados obtidos aplicando os dois métodos de RP, através das quatro possíveis alternativas (A, B, C e D) nessas matrizes de dados, utilizando os valores de semivariograma, verificamos que os mesmos foram inferiores quanto às classificações dos álcoois em suas devidas categorias, em relação à matriz das absorvâncias, como podem ser observados comparando os valores da tabela 09 com os da tabela 06.

O método KNN com até cinco vizinhos mais próximos (5-NN) forneceu a melhor classificação correta, com um índice médio acima de 60% no conjunto de treinamento, quando os dados foram normalizados mas não autoescalados (alternativa B). Entretanto, o conjunto teste apresentou um índice de aproximadamente 56%, com os mesmos procedimentos. A tabela 10 relaciona esses resultados, especificando o número de álcoois classificados incorretamente em cada categoria.

Na tabela 11 estão relacionados os resultados obtidos com o método SIMCA, utilizando 2 CP para cada categoria, o qual foi superior em classificação do que o modelo com 1 CP, principalmente quando os dados foram normalizados e autoescalados

(alternativa A). Esse método apresentou uma classificação correta com um índice de 79% no conjunto de treinamento, classificando incorretamente um álcool primário, três álcoois secundários e um álcool terciário. Todavia, o conjunto teste nessa mesma condição, forneceu um índice de aproximadamente 54%, classificando incorretamente dois álcoois primários e quatro álcoois secundários.

**TABELA 09** - Resultados dos métodos KNN e SIMCA usando 2 CP para cada categoria, nas matrizes 24x145 e 13x145 (semivariograma), através das seguintes alternativas:

- A) dados normalizados e autoescalonados
- B) dados normalizados e não autoescalonados
- C) dados não normalizados e autoescalonados
- D) dados não normalizados e não autoescalonados

DADOS	CONJUNTOS	Nº de pontos classificados incorretamente				
		1-NN	3-NN	4-NN	5-NN	SIMCA
A	treinamento	12	13	10	9	5
	% correta	50,0	45,8	58,3	62,5	79,17
	teste	6	6	6	4	6
	% correta	53,8	53,8	53,8	69,2	53,85
B	treinamento	9	10	9	10	16
	% correta	62,5	58,3	62,5	58,3	33,33
	teste	6	6	6	5	9
	% correta	53,8	53,8	53,8	61,5	30,77
C	treinamento	14	15	15	15	11
	% correta	41,7	37,5	37,5	37,5	54,17
	teste	6	6	5	5	9
	% correta	53,8	53,8	61,5	61,5	30,77
D	treinamento	16	13	15	12	17
	% correta	33,3	45,8	37,5	50,0	29,17
	teste	6	7	6	6	6
	% correta	53,8	46,2	53,8	53,8	53,85

TABELA 10 - Resultados do método KNN nas matrizes 24x145 para o conjunto de treinamento (10a) e 13x146 para o conjunto teste (10b) - semivariograma.

Tabela 10a - dados normalizados mas não autoescalonados

categorias	Nº de amostras	Nº de pontos classificados incorretamente			
		1-NN	3-NN	4-NN	5-NN
1	10	4	4	4	4
2	7	2	3	1	2
3	7	3	3	4	4
Total:	24	9	10	9	10
% classificação correta:		62,5	58,3	62,5	58,3

Tabela 10b - dados normalizados mas não autoescalonados

categorias	Nº de amostras	Nº de pontos classificados incorretamente			
		1-NN	3-NN	4-NN	5-NN
1	2	2	2	2	1
2	4	4	4	4	4
3	7	0	0	0	0
Total:	13	6	6	6	5
% classificação correta:		53,8	53,8	53,8	61,5

**TABELA 11** - Resultados do método SIMCA usando 2 CP para cada categoria, nas matrizes 24x145 para o conjunto de treinamento (11a) e 13x145 para o conjunto teste (11b) - semivariograma.

Tabela 11a - dados normalizados e autoescalonados

categorias	Nº de amostras	Nº de pontos classificados	
		Incorretamente	
1	10	1	
2	7	3	
3	7	1	
<b>Total:</b>	<b>24</b>	<b>5</b>	
<b>% classificação correta: 79,17</b>			

Tabela 11b - dados normalizados e autoescalonados

categorias	Nº de amostras	Nº de pontos classificados	
		Incorretamente	
1	2	2	
2	4	4	
3	7	0	
<b>Total:</b>	<b>13</b>	<b>6</b>	
<b>% classificação correta: 53,85</b>			

Os dados pré-processados pela função de autocorrelação apresentam os dois últimos valores indeterminados, quando a diferença entre o número de variáveis ( $n$ ) e o intervalo entre os pontos correspondentes no espectro ( $L$ ) é igual a zero ou um, de acordo com a equação 05. Portanto, foi utilizada uma matriz de dados de dimensão  $24 \times 144$  para o conjunto de treinamento e uma de  $13 \times 144$  para o conjunto teste.

A tabela 12 relaciona os resultados com essas matrizes de dados autocorrelacionados, obtidos com a aplicação dos métodos KNN com até cinco vizinhos mais próximos (5-NN) e o SIMCA utilizando 2 CP para cada categoria, empregando as quatro possíveis alternativas (A, B, C e D). Através dessa tabela, verificamos que os resultados são mais ou menos iguais para o método KNN, usando as alternativas A e C ou B e D, sendo as alternativas B e D superiores às outras. Isto indica que o autoescalamento, não é necessário para um bom procedimento nos dados pré-processados pela função de autocorrelação, quando aplicado este método. Também, para o método SIMCA as semelhanças destacaram-se nas alternativas A e C, B e D. Neste caso, a normalização das variáveis não é mais importante, mas o autoescalamento é, como foi para as absorvâncias e semivariograma. De acordo com essa tabela, verificamos que o método KNN com até cinco vizinhos mais próximos (5-NN), forneceu uma classificação correta ligeiramente melhor entre os álcoois primários, secundários e terciários, quando os dados não foram normalizados e nem autoescalados, isto é, com as absorvâncias somente autocorrelacionadas (alternativa D), apresentando um índice médio de 87% no conjunto de treinamento. Nas mesmas condi-



ções, o correspondente conjunto teste teve um índice superior a 70%. A tabela 13 relaciona esses resultados com mais detalhes, especificando para cada categoria o número de álcoois classificados incorretamente.

Com estes dados autocorrelacionados, aplicamos o método SIMCA utilizando 2 CP para cada categoria, o qual apresentou uma melhor discriminação dos álcoois nas três categorias, com um índice de 100% no conjunto de treinamento, quando os dados foram, ou não, normalizados mas principalmente autoescalonados (alternativas A e C). Entretanto, o conjunto teste nessas mesmas condições, forneceu um baixo índice de classificação correta com 46%, classificando incorretamente três álcoois secundários e quatro álcoois terciários. Também, aplicamos este método utilizando apenas 1 CP para cada categoria, com os dados normalizados e autoescalonados (alternativa A). Nessa aplicação, o conjunto teste apresentou uma melhor classificação do que a obtida com o modelo de 2 CP. Uma comparação mais específica dos resultados através desse método, utilizando 1 e 2 CP para cada categoria, pode ser observadas na tabela 14.

Se a única diferença entre os espectros de álcoois alifáticos saturados e insaturados, acontece na região das frequências de estiramento C=C ( $1680-1640\text{ cm}^{-1}$ ) e não incluída neste estudo, este resultado pode indicar que um modelo com 1 CP seja mais preciso do que com 2 CP para a classificação de álcoois. De fato, no conjunto de treinamento os dois modelos de SIMCA são semelhantes em termos de suas capacidades de classificar os vinte e quatro álcoois alifáticos saturados.

**TABELA 12** - Resultados dos métodos KNN e SIMCA (2 CP) para cada categoria, nas matrizes 24x144 e 13x144 (função de autocorrelação), através das seguintes alternativas:

- A) dados normalizados e autoescalonados
- B) dados normalizados e não autoescalonados
- C) dados não normalizados e autoescalonados
- D) dados não normalizados e não autoescalonados

DADOS	CONJUNTOS	Nº de pontos classificados incorretamente				
		1-NN	3-NN	4-NN	5-NN	SIMCA
A	treinamento	4	6	4	11	0
	% correta	83,3	75,0	83,3	54,2	100,00
	teste	4	3	3	3	7
	% correta	69,2	76,9	76,9	76,9	46,15
B	treinamento	2	3	2	6	9
	% correta	91,7	87,5	91,7	75,0	62,50
	teste	4	4	3	4	4
	% correta	69,2	69,2	76,9	69,2	69,23
C	treinamento	4	4	4	10	0
	% correta	83,3	83,3	83,3	58,3	100,00
	teste	6	4	4	4	7
	% correta	53,8	69,2	69,2	69,2	46,15
D	treinamento	2	3	2	5	9
	% correta	91,7	87,5	91,7	79,2	62,50
	teste	4	4	3	4	4
	% correta	69,2	69,2	76,9	69,2	69,23

TABELA 13 - Resultados do método KNN nas matrizes 24x144 para o conjunto de treinamento (13a) e 13x144 para o conjunto teste (13b) - função de autocorrelação.

Tabela 13a - dados não normalizados e nem autoescalonados

categorias	Nº de amostras	Nº de pontos classificados incorretamente			
		1-NN	3-NN	4-NN	5-NN
1	10	1	2	1	5
2	7	1	1	1	0
3	7	0	0	0	1
Total:	24	2	3	2	6
% classificação correta:		91,7	87,5	91,7	75,0

Tabela 13b - dados não normalizados e nem autoescalonados

categorias	amostras	Nº de pontos classificados incorretamente			
		1-NN	3-NN	4-NN	5-NN
1	2	0	0	0	0
2	4	3	4	3	4
3	7	1	0	0	0
Total:	13	4	4	3	4
% classificação correta:		69,2	69,2	76,9	69,2

**TABELA 14** - Resultados do método SIMCA nas matrizes 24x144 para o conjunto de treinamento (14a) e 13x144 para o conjunto teste (14b) - função de autocorrelação.

Tabela 14a - dados normalizados e autoescalados

categorias	Nº de amostras	Nº de pontos classificados incorretamente	
		2 CP	1 CP
1	10	0	2
2	7	0	0
3	7	0	0
Total:	24	0	2
% classificação correta:		100,00	91,67

Tabela 14b - dados normalizados e autoescalados

categorias	Nº de amostras	Nº de pontos classificados incorretamente	
		2 CP	1 CP
1	2	0	0
2	4	3	3
3	7	4	0
Total:	13	7	3
% classificação correta:		46,15	76,92

A tabela 15 apresenta um quadro geral dos resultados obtidos no conjunto de treinamento, aplicando os dois métodos de RP nas matrizes de absorvâncias, semivariograma e função de autocorrelação, através das alternativas A, B, C e D, discutidas anteriormente. Com essa tabela podemos discutir os aspectos mais relevantes para cada método de RP:

a) Para o método KNN

1. Os melhores índices de classificações corretas foram obtidos utilizando a matriz de autocorrelação, sendo que os mesmos diminuíram para as absorvâncias e para o semivariograma.

2. Os resultados obtidos com os dados somente autocorrelacionados indicam que, para a função de autocorrelação, não há necessidade da normalização e autoescalamento (alternativa D).

3. As alternativas B e D confirmam que o autoescalamento não é um pré-requisito para este método de RP.

4. A matriz de absorvâncias mostra que a normalização das mesmas é aparentemente importante, de acordo com os resultados obtidos com a alternativa B. Quando utilizamos a alternativa A os índices de classificação correta diminuem, mostrando que nem sempre o autoescalamento de dados resulta em melhores índices de classificação para o método KNN.

b) Para o método SIMCA

1. As matrizes de absorvâncias e função de autocorrelação foram as que apresentaram melhores índices de classificação correta, com os dados autoescalados mas, seja com os dados normalizados ou não (alternativas A e C).

2. Teoricamente este resultado pode ser esperado, porque

a estrutura do método SIMCA é mais propícia para dados autoescalonados.

3. Pelo menos para este conjunto de dados, normalização não parece ser necessária para o método SIMCA. Este resultado é surpreendente para os dados com absorvâncias.

Diante desse quadro recomendamos que, para diferentes métodos de RP, há necessidade de um pré-processamento de dados específico. Por exemplo, se usamos a LDA ou a LLM, talvez precisaríamos de outros procedimentos de pré-processamento.

A tabela 16 relaciona um quadro geral dos resultados obtidos com as amostras do conjunto teste (biblioteca SADTLER) nas matrizes de absorvâncias, semivariograma e função de autocorrelação, aplicando os dois métodos de RP, através das alternativas A, B, C e D, já discutidas. Nessa tabela pode-se observar que, os resultados obtidos com a matriz de semivariograma permaneceram ligeiramente inferiores em relação às outras, para ambos os métodos KNN e SIMCA.

Também, observamos que o método KNN apresentou resultados ligeiramente superiores, quando os dados da função de autocorrelação e das absorvâncias não são autoescalonados, independentes de serem normalizados ou não (alternativas B e D). O método SIMCA apresentou melhores índices de classificação para os dados das absorvâncias normalizadas. Neste caso, surpreendentemente os resultados com dados não autoescalonados são tão bons quanto aqueles com dados autoescalonados.

Os quatro espectros vibracionais de álcoois obtidos experimentalmente foram incluídos num segundo conjunto teste, de

maneira que, também usamos os procedimentos de pré-processamento que foram julgados superiores para este conjunto. Portanto, foram utilizadas as seguintes matrizes de dados de dimensões: 4x146 (absorvâncias), 4x145 (semivariograma) e 4x144 (função de autocorrelação).

Na tabela 17 estão relacionados os resultados obtidos com esse segundo conjunto teste nas três matrizes de dados, aplicando os métodos de RP através das quatro possíveis alternativas (A, B, C e D). Os resultados apresentados nessa tabela indicam que o conjunto de treinamento é útil para a classificação dos álcoois obtidos experimentalmente, em diferentes velocidades de varredura, utilizando estes procedimentos.

Como era de se esperar, baseado nos resultados do conjunto de treinamento, na matriz de dados utilizando os valores de semivariograma, os métodos KNN e SIMCA deram índices de classificações inferiores àqueles obtidos com as absorvâncias e função de autocorrelação. Também o método KNN, usando tanto as absorvâncias como a função de autocorrelação, dá resultados excelentes. Todavia, para o método SIMCA estes resultados mostram a necessidade de autoescalamento dos dados.

**TABELA 15** - Resultados dos métodos KNN e SIMCA usando 2 CP para cada categoria, no conjunto treinamento com as matrizes de dados 24x146 (absorvâncias), 24x145 (semivariograma) e 24x144 (função de autocorrelação), através das seguintes alternativas:

- A) dados normalizados e autoescalonados
- B) dados normalizados e não autoescalonados
- C) dados não normalizados e autoescalonados
- D) dados não normalizados e não autoescalonados

DADOS	MATRIZES	Nº de pontos classificados incorretamente				
		1-NN	3-NN	4-NN	5-NN	SIMCA
A	Absorvâncias (24x146)	9	8	11	12	0
B		6	6	5	4	9
C		12	16	15	13	1
D		9	9	9	9	8
A	Semivariograma (24x145)	12	13	10	9	5
B		9	10	9	10	16
C		14	15	15	15	11
D		16	13	15	12	17
A	F. Autocorrelação (24x144)	4	6	4	11	0
B		2	3	2	6	9
C		4	4	4	10	0
D		2	3	2	5	9



**TABELA 16** - Resultados dos métodos KNN e SIMCA usando 2 CP para cada categoria, no conjunto teste (SADTLER) com as matrizes de dados 13x146 (absorvâncias), 13x145 (semivariograma) e 13x144 (função de autocorrelação), através das seguintes alternativas:

- A) dados normalizados e autoescalados
- B) dados normalizados e não autoescalados
- C) dados não normalizados e autoescalados
- D) dados não normalizados e não autoescalados

DADOS	MATRIZES	Nº de pontos classificados incorretamente				
		1-NN	3-NN	4-NN	5-NN	SIMCA
A	Absorvâncias (13x146)	3	3	3	6	3
B		3	3	4	3	3
C		5	6	6	6	6
D		4	4	4	5	9
A	Semivariograma (13x145)	6	6	6	4	6
B		6	6	6	5	9
C		6	6	5	5	9
D		6	7	6	6	6
A	F. Autocorrelação (13x144)	4	3	3	3	7
B		4	4	3	4	4
C		6	4	4	4	7
D		4	4	3	4	4

**TABELA 17** - Resultados dos métodos KNN e SIMCA (2 CP) para cada categoria, no conjunto teste (espectros experimentais), com as matrizes 4x146 (absorvâncias), 4x145 (semivariograma) e 4x144 (função de autocorrelação), através das seguintes alternativas:

- A) dados normalizados e autoescalonados
- B) dados normalizados e não autoescalonados
- C) dados não normalizados e autoescalonados
- D) dados não normalizados e não autoescalonados

DADOS	MATRIZES	Nº de pontos classificados incorretamente				
		1-NN	3-NN	4-NN	5-NN	SIMCA
A	Absorvâncias (4x146)	0	0	0	1	0
B		0	0	0	0	2
C		0	0	0	0	0
D		0	0	0	0	2
A	Semivariograma (4x145)	2	0	0	1	1
B		2	2	2	2	4
C		1	1	1	2	4
D		1	2	1	2	2
A	F. Autocorrelação (4x144)	0	0	0	0	0
B		0	0	0	0	1
C		0	0	0	0	0
D		0	0	0	0	1

## CAPÍTULO VI

### CONCLUSÕES

Através da aplicação dos pesos de Fisher e variância, na matriz de dados utilizando as atribuições de bandas, verificamos que os valores das frequências são mais importantes do que as absorvâncias relativas, para a discriminação dos álcoois, principalmente os terciários das outras duas categorias. As frequências que mais se destacaram nessa discriminação estão na região de estiramento O-H e de estiramento C-O. Os dados referentes às absorvâncias relativas para as bandas da região de deformação O-H e de estiramento C-O, contribuem na discriminação dos álcoois primários dos outros álcoois.

Nessa mesma matriz de dados, observamos que o método KNN não fez uma boa classificação dos álcoois nos conjuntos de treinamento e teste da biblioteca SADTLER. Todavia, o método SIMCA classificou corretamente os álcoois em suas devidas categorias para o conjunto de treinamento, mas aproximadamente 70% para o conjunto teste.

Apesar de argumentos teóricos a seu favor, o uso do semivariograma para representar os espectros vibracionais no I.V. de álcoois, forneceu resultados inferiores em relação ao uso das absorvâncias e função de autocorrelação aplicando os métodos de RP, através das quatro possíveis alternativas citadas no capítulo

anterior. O método de RP que apresentou o resultado mais razoável com o uso do semivariograma foi o SIMCA com 2 CP para cada categoria. Este classificou corretamente os álcoois em primários, secundários e terciários com um índice de aproximadamente 80% para o conjunto de treinamento, quando os dados foram normalizados e autoescalonados.

Nas matrizes de absorvâncias, semivariograma e função de autocorrelação, os resultados obtidos com o método KNN mostraram-se ser independentes da normalização e não sendo necessário do autoescalonamento. Porém, o destaque para o método SIMCA foi principalmente o autoescalonamento dos dados.

Entre os três métodos de pré-processamento de dados desenvolvidos, recomendamos o uso da função de autocorrelação. Para o conjunto de treinamento, o método KNN com até cinco vizinhos mais próximos, apresentou uma classificação correta dos álcoois em suas devidas categorias, com um índice médio de aproximadamente 88%, quando os dados foram normalizados, mas não autoescalonados. Entretanto, o método SIMCA utilizando 2 CP para cada categoria, forneceu uma completa separação dos álcoois em primários, secundários e terciários, quando os dados foram ou não normalizados, mas autoescalonados. Na matriz dos valores contendo as absorvâncias, os métodos KNN e SIMCA também deram resultados satisfatórios para a classificação dos álcoois em suas respectivas categorias, apesar de serem ligeiramente inferiores aos da função de autocorrelação.

Estas conclusões também são aparentemente válidas para o conjunto teste dos espectros obtidos experimentalmente.

Finalmente, podemos afirmar que, embora as regras de classificações desenvolvidas para os álcoois saturados contenham informação útil para a classificação dos álcoois insaturados, a substituição da ligação C-C por C=C perturba as coordenadas normais, correspondendo às bandas nas regiões dos espectros estudados ( $3700-2200\text{ cm}^{-1}$  e  $1540-800\text{ cm}^{-1}$ ). Este fato pode explicar porque os índices de classificações para o conjunto teste da biblioteca SADTLER são bem menores do que para o conjunto de treinamento, para os métodos de pré-processamento de dados preferidos neste trabalho.

## REFERÊNCIAS

1. "Chemometrics", John Wiley, New York, 1986. Shraff, M.; Illman, D; Kowalski, B.R.
2. Kowalski, B.R.; Bender, C.F. J.Am.Chem.Soc. 96 5632 (1972).
3. Kwan, W-O; Kowalski, B.R. Anal.Chim.Acta 122 215 (1980).
4. Saxberg, B.E.H; Duewer, D.L.; Booker, J.L.; Kowalski, B.R. Anal.Chim.Acta 103 201 (1978).
5. Kowalski, B.R.; Schatzki, T.F.; Stross, F.H. Anal.Chem. 44 217 (1972).
6. Kowalski, B.R. Anal.Chem. 47 1152A (1975).
7. Scarmínio, I.S. "Análises Espectrométricas de Águas Minerais e sua Classificação por Meios de Reconhecimento de Padrões". Tese de Mestrado, I.Q. Unicamp (1979).
8. Kowalski, B.R. Anal.Chem. 52 112R (1980).
9. Frank I.E. e Kowalski, B.R. Anal.Chem. 54 232R (1982).
10. Kateman, G.; Pipers, F.W. "Quality Control in Analytical Chemistry", Wiley-Interscience, New York, 1981.
11. Taylor, J.K. Anal.Chem. 53 1655A (1981).
12. Eckschlager, K; Stepanek, V. "Information Theory as Applied to Chemical Analysis", Wiley-Interscience, New York, 1979.
13. Malinowski, E.R.; Howery, D.G. "Factor Analysis in Chemistry", Wiley-Interscience, New York, 1980.
14. Lidell, R.W.; Jurss, P.C. Anal.Chem. 46 2126 (1974).
15. Preuss, D.R.; Jurss, P.C. Anal.Chem. 520 46 (1974).
16. Woodruff, H.B.; Munk, M.E. Anal.Chim.Acta 95 113 (1977).

17. Domokos, L.; Frank, I.; Matolcsy, G. Anal.Chim.Acta 154 181 (1983).
18. Anderson, C.P.; Killeem, T.J.; Taft, J.B.; Bentz, A.P. Environ.Sci.Technol.14 1230 (1980).
19. Mecke, R und Noack, K. Strukturbestimmungen von ungesättigten Ketonen mit Hilfe von Infrarot - und Ultraviolet - Spektren (1960).
20. Biblioteca SADTLER Research Laboratories Inc.
21. Kowalski, B.R. ed. Chemometrics; Mathematics and Statistics in Chemistry. Dordrecht, D. Reidel, 1984.
22. Bruns, R.E. "Manual do usuário" - Programas, Convenio Docegeo - Funcamp (1984).
23. "Spectrometric Identification of Organic Compounds" - Second Edition Silverstain and Bassler 84 (1968).
24. "The Infrared Spectra Of Complex Molecules" Bellamy, L.J. 95 (1966).
25. Dyer, J.R. "Aplicações da Espectroscopia de Absorção aos Compostos Orgânicos" Ed. Edgard Blucher Ltda, Série de Textos Básicos de Química Orgânica.
26. "Calculus and Analytic Geometry" George B. Thomas, Jr. Addison-Wesley Publishing Company, Inc. Third Edition, 1962.
27. Geostatistics, part 2 of a series 90, July 1979.
28. "Statistics and Data Analysis in Geology" John C. Davis, Kansas Geological Survey, 1973, John Wiley & Sons, Inc.
29. Sjostrom, M.; Kowalski, B.R. Anal.Chim.Acta 112 11 (1979).
30. Varmuza, K. Anal.Chim.Acta 122 227 (1980).
31. Coomans, D.; Massart, D.L. Anal.Chim.Acta 136 15 (1982).

32. Wold, S. "Pattern Recognition" - Pergamon Press 8 127 (1976)  
Printed in Great Britain.
33. Albano, C; Dunn III, W.; Edlund, U.; Johansson, E.; Norden  
Bo; Sjostron, M. e Wold, S. Anal.Chim.Acta 103 429 (1978).
34. "Proceedings Nato Adv.Study Inst. on Chemometrics", Cosenza,  
Italy, September (1983) Ed. B.R.Kowalski -Reidel Publ.Co.,  
Dordrecht, Holand, (1984).
35. Albano, C.; Blomqvist, G.; Coomans, D.; Dunn III, W.J.;  
Eliasson, B.; Hellber, S.; Johansson, E.; Nordén, B.; Joh-  
nels, D.; Sjostrom, M.; Soderstrom, B.; Wold, H.; Wold,S.  
Pattern Recognition by Means Of Disjoint Principal Components  
Models (SIMCA). Philosophy and Methods.



## APENDICE A

Programa: AREANORM.FOR

Normalização das absorvâncias pela área total do espectro, através da aproximação trapezoidal, conforme a equação 03.

```
C.  NORMALIZAR AS ABSORVANCIAS PELA AREA DO ESPECTRO
    DIMENSION Y(146)
    N=146
    NN=N-1
    READ(7,10) ID, XNAME1, XNAME2, CN, (Y(K), K=1, N)
10  FORMAT(14, 2X, 2A4, 2X, F2.0, 2X, 6F10.5/17(8F10.5/)4F10.5/)
    SOMA=0.0
    DO 3 I=2, NN
    SOMA=SOMA+Y(K)
3   CONTINUE
    SUBT=SOMA+(Y(1)+Y(N))*0.5
    TOTAL=SOMA+SUBT
    DELTA=FLOAT(3700-800)/N
    AREA=TOTAL*DELTA
    DO 7 K=1, N
    Y(K)=Y(K)/AREA
7   CONTINUE
    WRITE(6,15) ID, XNAME1, XNAME2, CN, (Y(K), K=1, N)
15  FORMAT(1X, 14, 2X, 2A4, 2X, F2.0, 2X, 6F10.5/17(1X, 8F10.5/)1X, 4F10.5/)
    END
```

Continuação: APÊNDICE A

Programa: SEMIV.FOR

Cálculo do semivariograma nos dados utilizando as absorvâncias, conforme a equação 04.

```
C.  VALORES DE GAMA
      DIMENSION Y(146),GAMA(146)
      REAL NPAR
      N=146
      NN=N-1
      READ(7,5) ID,XNAME1,XNAME2,CN,(Y(K),K=1,N)
      FORMAT(14,2X,2A4,2X,F2.0,2X,6F10.5/17(8F10.5/)4F10.5/)
      DO 1 I=1,NN
      SUMAT=0.0
      NPAR=FLOAT(1/2*(N-1))
      NFIM=N-I
      DO 3 M=1,NFIM
      K=M+I
      SUM=Y(M)-(Y(K))
      DIST=SUM*SUM
      SUMAT=SUMAT+DIST
      VALOR=NPAR*SUMAT
3    CONTINUE
      GAMA(I)=VALOR*(10**6)
1    CONTINUE
      WRITE(6,30) ID,XNAME1,XNAME2,CN,(GAMA(I),I=1,NN)
30   FORMAT(1X,14,2X,2A4,2X,F2.0,2X,6F10.5/17(1X,8F10.5/)1X,4F10.5/)
      END
```

## APÊNDICE B

Conjunto de dados originais (absorvâncias) em forma de matriz, dos espectros vibracionais no I.V. de álcoois, obtidos da biblioteca SADTLER.

- Conjunto de treinamento (1 a 24) e Conjunto teste (25 a 37).

- primeira coluna: indica o número da amostra considerado pelo programa.
- segunda coluna: se refere ao nome da amostra para identificação.
- terceira coluna: se refere à classe da amostra e as demais colunas correspondem aos valores das absorvâncias (variáveis).

- FORMAT(I4,2X,2A4,2X,F2.0,2X,6F10.5/17(8F10.5/14F10.5/)

YOUR INDEX	NAME	CAT.	*****INPUT FEATURES*****							
1	PPROP	1.	.06500	.07500	.07500	.07500	.08000	.09500		
	.11000	.14000	.18000	.23000	.29000	.35000	.41500	.49000		
	.56000	.63000	.69000	.73000	.76000	.74000	.70000	.66000		
	.60000	.53000	.46000	.38000	.32000	.27000	.23000	.19500		
	.15500	.15000	.13500	.13000	.13000	.17000	.30000	.80000		
	.65000	.71000	.47500	.50000	.55000	.32000	.21000	.15500		
	.12500	.11500	.11000	.10500	.09000	.08500	.08500	.07500		
	.07000	.06500	.06500	.06500	.06500	.06500	.06000	.05500		
	.05500	.05500	.05000	.05000	.05000	.04500	.04500	.04500		
	.04500	.04500	.04500	.04500	.04500	.04500	.04000	.04000		
	.04000	.04000	.04000	.04000	.04000	.04000	.04000	.04000		
	.04000	.04000	.04000	.04000	.04000	.04000	.04000	.04000		
	.04000	.04000	.04000	.04000	.04000	.04000	.04000	.04000		
	.04000	.04000	.04000	.04000	.04000	.04000	.04000	.04000		
	.04000	.04000	.04000	.04000	.04000	.04000	.04000	.04000		
	.04000	.04000	.04000	.04000	.04000	.04000	.04000	.04000		
	.07500	.11000	.21500	.23000	.18500	.17500	.17500	.16000		
	.15000	.14000	.10000	.10000	.09500	.09000	.12000	.07700		
	.06500	.06500	.06500	.09000	.12000	.19500	.43000	.58000		
	.12500	.25000	.10000	.52000	.07500	.05700	.08000	.11000		
	.06000	.06000	.05500	.05500						
2	PBUTAN	1.	.04700	.06500	.06300	.08500	.09500	.12500		
	.16000	.20000	.29000	.38000	.47000	.56000	.66000	.76000		
	.84000	.92000	.98000	1.00000	1.00000	1.00000	.95000	.90000		
	.84000	.76000	.67000	.57000	.50000	.42000	.35000	.29000		
	.24000	.21000	.19000	.17700	.17700	.20500	.30000	1.00000		
1.	1.00000	1.20000	.84000	.74000	.95000	.53000	.33500	.21500		
	.16000	.13000	.12500	.12000	.09000	.08000	.07500	.06300		
	.05500	.04500	.04500	.04500	.04000	.04000	.04000	.04000		
	.04000	.04000	.04500	.05000	.05000	.05500	.05500	.05500		
	.05500	.05500	.05500	.05500	.05500	.05500	.05000	.05000		
	.05000	.05000	.05000	.05000	.05000	.05000	.05000	.05000		
	.05000	.05000	.05000	.05000	.05000	.05000	.05000	.05000		
	.05000	.05000	.05000	.05000	.05000	.05000	.05000	.05000		
	.05000	.05000	.05000	.05000	.05000	.05000	.05000	.05000		
	.05000	.05000	.05000	.05000	.05000	.05000	.05000	.05000		
	.05000	.05000	.05000	.05000	.05000	.05000	.05000	.05000		
	.05000	.05000	.05000	.05000	.05000	.05000	.05000	.05000		
	.11500	.17500	.39000	.41000	.34500	.30000	.28000	.30000		
	.24000	.24000	.20000	.19000	.13000	.15500	.13700	.18000		
	.11500	.09500	.11500	.16500	.32000	.25000	.76000	.53000		
	.60000	.37000	.30500	.19000	.30000	.07500	.09000	.09000		
	.09300	.26000	.12500	.13000						



.21500	.14000	.09000	.04000	.02500	.03000	.01500	.02000
.02000	.00700	.00500	.00500				
6 PDDDEC	1.	.03300	.03500	.03500	.04000	.04500	.04500
.05700	.07000	.09000	.12000	.15000	.19000	.23000	.27000
.30500	.35000	.38000	.40000	.42000	.42000	.41000	.39000
.37000	.33000	.28000	.24000	.20500	.17500	.14000	.13500
.11000	.09500	.09000	.09000	.09000	.10500	.15000	.37000
.72000	1.25000	1.25000	.64000	.62000	1.00000	.25000	.15000
.10000	.08000	.07500	.07000	.06000	.06000	.05500	.04700
.04300	.04000	.03500	.03000	.02500	.02500	.02000	.01500
.01300	.01500	.01500	.01500	.01500	.01500	.01500	.01500
.01500	.01500	.01500	.01500	.01500	.01500	.02000	.02000
.02000	.02000	.02000	.02000	.02000	.02000	.02000	.02000
.02000	.02000	.02000	.02000	.02000	.02000	.02000	.02000
.02000	.02000	.02000	.02000	.02000	.02000	.02000	.02000
.02000	.02000	.02000	.02000	.02000	.02000	.02000	.02000
.02000	.02000	.02000	.02000	.02000	.02000	.02000	.02000
.02000	.02000	.02000	.02000	.02000	.02000	.03500	.04000
.05000	.07000	.41500	.24000	.16000	.12500	.12500	.16000
.13000	.11000	.10500	.09000	.08000	.07000	.07000	.06000
.06000	.06000	.06000	.08500	.05500	.15000	.27000	.29000
.20500	.15500	.10000	.07500	.06000	.05500	.06000	.06000
.05500	.05000	.05000	.06000				
7 P2MEBUT	1.	.05500	.06000	.06500	.07000	.08500	.08500
.09500	.13000	.16000	.21000	.26000	.31000	.37000	.42000
.49000	.53000	.56000	.58000	.58000	.58000	.56000	.53000
.50000	.45000	.41000	.35000	.31000	.25500	.22000	.18500
.15500	.13500	.12500	.12000	.12000	.14000	.20000	.68000
.58000	.60000	.54000	.47000	.54000	.33000	.20000	.15000
.11000	.09000	.08500	.08500	.07000	.06000	.05500	.05000
.04500	.04000	.04000	.04000	.03500	.03000	.03000	.03000
.02500	.02000	.01500	.01500	.01500	.01500	.01500	.01500
.01500	.01500	.02500	.02500	.02500	.02500	.02500	.02500
.02500	.02500	.02500	.02500	.02500	.02500	.02500	.02500
.02500	.02500	.02500	.02500	.02500	.02500	.02500	.02500
.02500	.02500	.02500	.02500	.02500	.02500	.02500	.02500
.02500	.02500	.02500	.02500	.02500	.02500	.02500	.02500
.02500	.02500	.02500	.02500	.02500	.02500	.04000	.05000
.06700	.17500	.19000	.11500	.11000	.11000	.13500	.09500
.08000	.07500	.06500	.06000	.05700	.06500	.05500	.03500
.03500	.03500	.06000	.07500	.05000	.07000	.22000	.18000
.19000	.10000	.08000	.07000	.04500	.04000	.05000	.02500
.02500	.03000	.02500	.03000				
8 P3MEBUT	1.	.12000	.12500	.12800	.13000	.14300	.14500
.16000	.18000	.22500	.26000	.31500	.37000	.43000	.49000
.54000	.60000	.64000	.66000	.67000	.68000	.68000	.65000
.60000	.56000	.48000	.44000	.38000	.32500	.28500	.27700
.23500	.20500	.19500	.18500	.18500	.20000	.27000	.53000
.90000	.80000	.69000	.59000	.64000	.38000	.27500	.21500
.18000	.16000	.15000	.14500	.13000	.12000	.11500	.11500
.11000	.11500	.12000	.12000	.12300	.12300	.12000	.13000
.10500	.10300	.08700	.08500	.07500	.07500	.07500	.07300
.07300	.06000	.06000	.05500	.05000	.05000	.05000	.05000
.05000	.05000	.05000	.05000	.05000	.05000	.05000	.05000
.05000	.05000	.05000	.05000	.05000	.05000	.05000	.05000
.05000	.05000	.05000	.05000	.05000	.05000	.05000	.05000
.05000	.05000	.05000	.05000	.05000	.05000	.05000	.05000
.05000	.05000	.05000	.05000	.05000	.05000	.06000	.06300
.07700	.17000	.32000	.18000	.18000	.16000	.24500	.24500

.13500	.12800	.10000	.08000	.07300	.07000	.07000	.08500
.05700	.07500	.06500	.13000	.07500	.09500	.49000	.21500
.11500	.19500	.07500	.12000	.04000	.03500	.03000	.02000
.02500	.06000	.03000	.03000				
9 P2ETBUT	1.	.05000	.05000	.05500	.05500	.07000	.07500
.08000	.10500	.14000	.20000	.26000	.33000	.43000	.52000
.62000	.73000	.82000	.90000	.96000	.98000	.96000	.90000
.82000	.71000	.62000	.52000	.44000	.35000	.30000	.24000
.19000	.18000	.15500	.15000	.14500	.15000	.20000	.31000
1.90000	1.20000	1.30000	1.10000	.84000	1.25000	.50000	.28500
.21000	.15500	.14000	.14000	.13500	.11000	.10000	.09500
.09300	.09000	.09000	.09000	.08000	.07000	.06500	.06000
.05500	.05500	.05500	.05500	.05500	.05300	.05000	.05000
.05000	.05000	.04500	.04500	.04500	.04500	.05000	.05000
.05000	.05000	.05000	.05000	.05000	.05000	.05000	.05000
.05000	.05000	.05000	.05000	.05000	.05000	.05000	.05000
.05000	.05000	.05000	.05000	.05000	.05000	.05000	.05000
.05000	.05000	.05000	.05000	.05000	.05000	.05000	.05000
.05000	.05000	.05000	.05000	.05000	.05000	.05000	.05000
.05000	.05000	.05000	.05000	.05000	.05000	.06500	.07300
.07700	.11000	.23000	.68000	.29000	.23000	.22500	.47000
.20500	.18000	.18000	.17500	.14000	.12500	.13000	.16500
.11500	.08700	.11500	.13000	.18000	.13000	.11000	.45000
.70000	.52000	.26000	.10500	.20500	.13000	.07500	.14500
.07000	.06000	.10000	.11000				
10 PETDET	1.	.04500	.04700	.05000	.06500	.08300	.13500
.17000	.23000	.31500	.43000	.57000	.65000	.72000	.76000
.77000	.76000	.73000	.68000	.64000	.60000	.53000	.47000
.40000	.34000	.28500	.23000	.20000	.16500	.13500	.11500
.10000	.09300	.09000	.09000	.10000	.14500	.30000	.80000
.61000	.78000	.66000	.76000	1.00000	.84000	.43000	.29000
.23000	.18000	.13500	.13000	.11500	.10300	.07500	.06300
.06300	.06700	.06300	.05700	.05700	.05500	.05500	.05500
.05500	.05000	.05000	.05000	.05000	.05000	.04700	.04700
.05000	.05500	.05700	.05500	.05500	.05500	.04500	.04500
.04500	.04500	.04500	.04500	.04500	.04500	.04500	.04500
.04500	.04500	.04500	.04500	.04500	.04500	.04500	.04500
.04500	.04500	.04500	.04500	.04500	.04500	.04500	.04500
.04500	.04500	.04500	.04500	.04500	.04500	.04500	.04500
.07700	.13500	.20000	.32000	.24000	.29000	.34500	.42500
.36000	.19500	.19700	.21000	.21000	.19000	.22000	.21500
.15000	.29000	.34000	.96000	1.25000	.58000	.90000	.90000
.44000	.26000	.24000	.07700	.08000	.23000	.18500	.22500
.16000	.07700	.16000	.11500				
11 SPRDP	2.	.06000	.06000	.06300	.06500	.06700	.07000
.08000	.10500	.13500	.17500	.22000	.30000	.39500	.45500
.58000	.66000	.77000	.82000	.85000	.82000	.77000	.72000
.64000	.55000	.44500	.36500	.30000	.24500	.19500	.17500
.13500	.11500	.11000	.10000	.10000	.13500	.34500	1.30000
.33000	.41000	.34000	.37500	.31000	.20000	.14000	.12000
.11000	.09000	.08500	.08500	.08500	.08500	.07500	.07500
.08000	.06500	.06300	.06300	.06500	.05500	.05500	.05500
.05300	.04500	.04000	.04000	.04000	.03700	.03700	.03700
.03500	.03300	.03300	.03300	.03300	.03500	.03500	.03500
.03500	.03500	.03500	.03500	.03500	.03500	.03500	.03500
.03500	.03500	.03500	.03500	.03500	.03500	.03500	.03500
.03500	.03500	.03500	.03500	.03500	.03500	.03500	.03500

.03500	.03500	.03500	.03500	.03500	.03500	.05500	.06300
.07500	.13000	.31000	.15500	.20000	.25000	.26000	.36500
.21500	.22000	.30000	.17500	.09700	.08700	.07000	.07000
.09700	.57000	.33500	.57000	.42000	.13000	.06500	.05500
.05700	.05000	.05500	.19000	.44500	.07000	.04000	.03500
.03500	.05000	.09700	.07000				
12 SHEX	2.	.07000	.07000	.07500	.07500	.09000	.09500
.09500	.11500	.13500	.17500	.22000	.26500	.32500	.38000
.42000	.48000	.51000	.53000	.53000	.51000	.47000	.41000
.35000	.29500	.24500	.20500	.17500	.14500	.12500	.10000
.10000	.09500	.09500	.09500	.09700	.12000	.23500	.70000
.71000	.98000	.56000	.43000	.56000	.28500	.16000	.11500
.09300	.08300	.08300	.08300	.07500	.07000	.07000	.06500
.06500	.06000	.06000	.06000	.06000	.06000	.06000	.05000
.05000	.04500	.04500	.04500	.04700	.04700	.03700	.03500
.03300	.03000	.03000	.03000	.03000	.03000	.03500	.03500
.03500	.03500	.03500	.03500	.03500	.03500	.03500	.03500
.03500	.03500	.03500	.03500	.03500	.03500	.03500	.03500
.03500	.03500	.03500	.03500	.03500	.03500	.03500	.03500
.03500	.03500	.03500	.03500	.03500	.03500	.03700	.04500
.05700	.10500	.38000	.17000	.16000	.14500	.21000	.14000
.13500	.14500	.12000	.11000	.11500	.11000	.10000	.06300
.05300	.08300	.21500	.21000	.11000	.07500	.08500	.11500
.08000	.23000	.14500	.30000	.03500	.03000	.07500	.04000
.04000	.05000	.02500	.02000				
13 SHEP	2.	.08000	.09300	.09000	.09500	.12000	.13000
.13000	.16500	.21500	.26000	.33000	.40000	.49000	.55000
.62000	.68000	.72000	.74000	.72000	.68000	.64000	.58000
.51000	.42000	.35000	.31000	.26000	.21500	.18000	.16000
.15000	.14000	.13000	.13500	.14500	.19000	.35000	1.30000
1.00000	1.35000	.88000	.62000	.88000	.47000	.24000	.17500
.14000	.12000	.12000	.11500	.10000	.10000	.10500	.09500
.09000	.09000	.08500	.08000	.07500	.07500	.07000	.05700
.05700	.05700	.05700	.05700	.05700	.05700	.05700	.06000
.05500	.05000	.05000	.05000	.05000	.04500	.04500	.04500
.04500	.04500	.04500	.04500	.04500	.04500	.04500	.04500
.04500	.04500	.04500	.04500	.04500	.04500	.04500	.04500
.04500	.04500	.04500	.04500	.04500	.04500	.04500	.04500
.04500	.04500	.04500	.04500	.04500	.04500	.04500	.04500
.04500	.04500	.04500	.04500	.04500	.04500	.05500	.06300
.07500	.15000	.42000	.21000	.21500	.20000	.31000	.17000
.19000	.20000	.20500	.16500	.14500	.11500	.09000	.10500
.07000	.11000	.30000	.29000	.26000	.15000	.26000	.19000
.18000	.09500	.06500	.15000	.18000	.06500	.07500	.05000
.07000	.05500	.06500	.04500				
14 SOCT	2.	.10000	.10500	.10500	.11000	.12500	.13000
.13500	.15500	.18500	.23500	.29000	.35500	.42000	.49000
.56000	.62000	.66000	.67000	.69000	.68000	.62000	.54000
.49000	.42000	.33500	.28000	.25000	.21000	.18000	.18000
.15500	.14500	.14000	.14000	.15000	.18000	.30000	.92000
1.00000	1.50000	1.25000	.68000	.76000	.90000	.25000	.17000
.14000	.12000	.11500	.12000	.11000	.11000	.10800	.10000
.09500	.09000	.08500	.08000	.08000	.08000	.08000	.07500
.07500	.07000	.06800	.06500	.06500	.06500	.06500	.06500
.05300	.04500	.04500	.04500	.04500	.04500	.04500	.04500
.04500	.04500	.04500	.04500	.04500	.04500	.04500	.04500

.04500	.04500	.04500	.04500	.04500	.04500	.04500	.04500
.04500	.04500	.04500	.04500	.04500	.04500	.04500	.04500
.04500	.04500	.04500	.04500	.04500	.04500	.05500	.06000
.07000	.10000	.45000	.28000	.20500	.21000	.20000	.30000
.17000	.19500	.21000	.18500	.14000	.13000	.11000	.09000
.11000	.08000	.25000	.27500	.34000	.16000	.26500	.19000
.16500	.11500	.08500	.10000	.13000	.12500	.05000	.04500
.04500	.08000	.03500	.03500				
15 SDEC	2.	.01700	.02000	.02500	.02700	.05500	.05300
.05500	.07500	.11500	.15500	.22500	.29000	.36000	.42000
.49000	.56000	.60000	.62000	.62000	.60000	.57000	.51000
.45000	.39000	.32000	.27000	.23500	.19500	.16500	.13500
.11500	.10000	.08500	.10500	.12000	.13500	.23000	.73000
.82000	1.05000	1.00000	.74000	.76000	.98000	.33000	.18000
.11000	.07500	.06500	.07000	.06000	.05000	.05500	.04500
.04000	.03000	.02500	.02000	.01300	.01000	.00500	.00500
.00500	.00500	.00500	.00500	.00500	.00500	.00500	.01500
.02000	.01500	.01500	.01500	.01500	.01300	.02000	.02000
.02000	.02000	.02000	.02000	.02000	.02000	.02000	.02000
.02000	.02000	.02000	.02000	.02000	.02000	.02000	.02000
.02000	.02000	.02000	.02000	.02000	.02000	.02000	.02000
.02000	.02000	.02000	.02000	.02000	.02000	.03000	.04000
.05500	.10000	.54000	.31000	.23500	.25500	.21000	.28000
.18500	.20000	.21500	.18000	.14000	.11500	.09500	.09000
.09500	.10000	.26500	.28000	.29000	.24000	.23000	.18000
.15000	.14500	.10000	.07000	.22000	.11000	.09500	.04000
.05500	.08500	.04500	.05000				
16 S3ETPEN	2.	.02000	.02500	.03500	.03500	.06700	.07500
.08000	.12000	.17000	.24000	.32000	.40000	.48000	.56000
.62000	.68000	.72000	.72000	.72000	.68000	.62000	.56000
.47000	.38000	.31500	.25000	.21000	.16000	.13500	.10000
.09500	.08500	.08000	.08000	.09500	.15500	.43000	1.00000
.79000	.86000	.69000	.62000	.76000	.30000	.17000	.11500
.08500	.07000	.06700	.07000	.06000	.05000	.05000	.04300
.04300	.04300	.04000	.03300	.02700	.02000	.02000	.01700
.00700	.00500	.00700	.01300	.01500	.01500	.02300	.02500
.02000	.01500	.01700	.02000	.02000	.02000	.03000	.03000
.03000	.03000	.03000	.03000	.03000	.03000	.03000	.03000
.03000	.03000	.03000	.03000	.03000	.03000	.03000	.03000
.03000	.03000	.03000	.03000	.03000	.03000	.03000	.03000
.03000	.03000	.03000	.03000	.03000	.03000	.06500	.07000
.08700	.17000	.57000	.30000	.22500	.25000	.45000	.23000
.16000	.20000	.24000	.22500	.19500	.17000	.10500	.07700
.07500	.23000	.22000	.21000	.45500	.23000	.21500	.18500
.30000	.22000	.05000	.05500	.32000	.10500	.08000	.10000
.13500	.06500	.09000	.07000				
17 S22DHPEN	2.	.02000	.03500	.04000	.06500	.06700	.07300
.10000	.14000	.19500	.25500	.31500	.37000	.42000	.45000
.48500	.50000	.50000	.48000	.41000	.35000	.29000	.23500
.19500	.16000	.13500	.11500	.09700	.08500	.07500	.06300
.05300	.05000	.04500	.04500	.06000	.09000	.29500	.82000
.86000	.65000	.44000	.38000	.56000	.24500	.13000	.08500
.06000	.04700	.04500	.04500	.04500	.04000	.04000	.04000
.04000	.03500	.03500	.03300	.03000	.02500	.02500	.02000
.02000	.02000	.02300	.02300	.02500	.02500	.02500	.02000



.02000	.02000	.02000	.02000	.02000	.02000	.02000	.02000
.02000	.02000	.02000	.02000	.02000	.02000	.02000	.02000
.02000	.02000	.02000	.02000	.02000	.02000	.02000	.02000
.02000	.02000	.02000	.02000	.02000	.02000	.02000	.02000
.02000	.02000	.02000	.02000	.02000	.02000	.03000	.03500
.05000	.26500	.36500	.14500	.11000	.16000	.21000	.41000
.09000	.14000	.19500	.10500	.10000	.17000	.12500	.11500
.10500	.06500	.06000	.09500	.42500	.18000	.15500	.10500
.09500	.09700	.25000	.43000	.20000	.08500	.05000	.04300
.11000	.04000	.04300	.04500				
18 TBUT	3.	.03000	.04500	.03700	.04300	.06000	.07500
.08500	.11500	.16000	.22000	.29000	.37000	.45000	.53000
.59000	.64000	.66000	.67000	.66000	.62000	.57000	.49000
.42000	.35500	.29500	.24000	.20500	.16500	.12500	.10500
.08500	.07000	.06300	.06300	.07000	.14000	.41000	1.00000
.44000	.39000	.33000	.28000	.25000	.12000	.09500	.08500
.07500	.06300	.05500	.04500	.04500	.03300	.03300	.03000
.02700	.03500	.04000	.04500	.04500	.03300	.02000	.02000
.01500	.01500	.02000	.02300	.02500	.02700	.02000	.01300
.01500	.02000	.02500	.02300	.01500	.02000	.03000	.03000
.03000	.03000	.03000	.03000	.03000	.03000	.03000	.03000
.03000	.03000	.03000	.03000	.03000	.03000	.03000	.03000
.03000	.03000	.03000	.03000	.03000	.03000	.03000	.03000
.03000	.03000	.03000	.03000	.03000	.03000	.05000	.06500
.09000	.21000	.19500	.14000	.15500	.22500	.57000	.65000
.16500	.10500	.09000	.08500	.10000	.34000	.31000	.68000
.56000	.34000	.16500	.08000	.06000	.05000	.04700	.06000
.17500	.05000	.04500	.05000	.08500	.52000	.34500	.06500
.05000	.05000	.05500	.06300				
19 T2HBU2	3.	.07300	.09500	.10500	.10500	.13000	.18500
.19000	.25000	.32500	.41500	.55000	.68000	.79000	.90000
1.00000	1.10000	1.15000	1.15000	1.05000	1.00000	.84000	.70000
.59000	.48500	.41500	.34000	.28500	.23500	.19500	.15500
.15000	.14000	.13000	.13000	.14000	.22000	.64000	1.50000
.82000	.98000	.67000	.65000	.44500	.29000	.17500	.14000
.11500	.10000	.09500	.09000	.07700	.07500	.07500	.07500
.07500	.07300	.07000	.07000	.07000	.06500	.06000	.05500
.04500	.04500	.04500	.04500	.04500	.04500	.04500	.04500
.04000	.04000	.03700	.03500	.03500	.03300	.03000	.03000
.03000	.03000	.03000	.03000	.03000	.03000	.03000	.03000
.03000	.03000	.03000	.03000	.03000	.03000	.03000	.03000
.03000	.03000	.03000	.03000	.03000	.03000	.03000	.03000
.03000	.03000	.03000	.03000	.03000	.03000	.04500	.05000
.06500	.17000	.40000	.16000	.16500	.23500	.43000	.36000
.16000	.15500	.13500	.13500	.22000	.11000	.16000	.20000
.46000	.60000	.36000	.15500	.12000	.08500	.20000	.16000
.11000	.13500	.08000	.07500	.60000	.15000	.07000	.48000
.04700	.03500	.04000	.05000				
20 T2MPEN2	3.	.09300	.09500	.11500	.12000	.12300	.14500
.18000	.22500	.27500	.33500	.39000	.45000	.48000	.50500
.52000	.51500	.50000	.45000	.39500	.33000	.28500	.24500
.21000	.18500	.16500	.14500	.13500	.12000	.11500	.11000
.12000	.12000	.12300	.13500	.16000	.38000	1.20000	.58000
.64000	.50000	.37000	.53000	.25000	.17500	.14500	.12500
.11500	.11000	.11500	.11000	.10000	.09700	.10000	.10000

.10000	.09500	.09500	.09500	.09500	.09000	.08500	.08000
.08500	.08500	.08500	.08300	.08300	.08000	.07500	.07500
.07500	.07700	.07500	.07500	.07500	.07500	.05000	.05000
.05000	.05000	.05000	.05000	.05000	.05000	.05000	.05000
.05000	.05000	.05000	.05000	.05000	.05000	.05000	.05000
.05000	.05000	.05000	.05000	.05000	.05000	.05000	.05000
.05000	.05000	.05000	.05000	.05000	.05000	.05700	.06000
.06500	.09000	.27500	.18000	.11500	.14000	.20000	.31000
.18500	.11000	.10000	.14000	.10000	.18000	.15500	.16000
.27000	.33000	.29000	.14000	.10500	.16500	.13000	.10000
.08700	.08000	.08700	.11500	.11500	.24000	.11000	.18000
.09500	.08000	.07500	.07500				
21 T3MPEN3	3.	.07300	.08000	.07500	.08300	.11500	.11000
.11500	.15000	.20000	.27000	.34000	.42000	.48000	.52000
.56000	.58000	.57000	.53000	.45000	.38000	.31000	.25500
.21500	.18500	.16000	.14500	.13000	.12000	.10500	.10000
.10500	.10500	.10500	.11000	.12500	.19000	.82000	1.25000
.86000	.88000	.42000	.60000	.31500	.21000	.13500	.12000
.10500	.09300	.09300	.08000	.07500	.08000	.08000	.07500
.07500	.07500	.07000	.07000	.07000	.07000	.07000	.06500
.06500	.06500	.06500	.06300	.06300	.06300	.06300	.06000
.05700	.06000	.06000	.06000	.06000	.05700	.06000	.06000
.06000	.06000	.06000	.06000	.06000	.06000	.06000	.06000
.06000	.06000	.06000	.06000	.06000	.06000	.06000	.06000
.06000	.06000	.06000	.06000	.06000	.06000	.06000	.06000
.06000	.06000	.06000	.06000	.06000	.06000	.06500	.06700
.08000	.10500	.30000	.27000	.15000	.19000	.25000	.36000
.16500	.20000	.15000	.18000	.21000	.15500	.12500	.14500
.22500	.37000	.48000	.24500	.13500	.10500	.18500	.14500
.13500	.16000	.19000	.37000	.08300	.09000	.66000	.20000
.06500	.05000	.04700	.05000				
22 T23DMBU2	3.	.06000	.05500	.06000	.08000	.08000	.08500
.10500	.15500	.20000	.27000	.35000	.41000	.46000	.50000
.52000	.50000	.49000	.42000	.35000	.27000	.21000	.18000
.15000	.12500	.11000	.09500	.08500	.07500	.07000	.06500
.06500	.07000	.07000	.07500	.09500	.30000	1.50000	.76000
.50000	.35000	.32000	.45000	.14000	.08500	.07500	.06500
.06500	.06300	.05700	.06000	.05300	.04700	.04500	.04500
.04700	.05500	.04500	.04500	.04500	.04500	.04500	.04300
.04000	.04000	.04000	.04000	.03700	.03500	.03000	.03000
.03500	.03500	.03500	.03000	.03000	.03000	.03000	.03000
.03000	.03000	.03000	.03000	.03000	.03000	.03000	.03000
.03000	.03000	.03000	.03000	.03000	.03000	.03000	.03000
.03000	.03000	.03000	.03000	.03000	.03000	.03000	.03000
.03000	.03000	.03000	.03000	.03000	.03000	.03500	.04000
.05000	.09500	.29000	.15000	.09000	.13000	.31000	.36000
.12500	.11000	.12000	.06500	.05300	.06000	.10500	.11000
.19000	.34000	.33000	.15000	.09500	.14500	.12000	.10000
.04000	.03000	.04000	.07000	.29000	.07000	.22000	.02500
.02500	.11000	.03000	.02500				
23 T24DMPE2	3.	.03000	.03500	.03500	.03500	.05500	.07000
.07000	.08300	.11000	.15500	.21500	.25500	.30500	.35000
.37500	.41000	.41000	.40000	.37000	.31500	.26000	.21000
.18000	.16000	.13000	.11500	.10500	.08000	.06500	.05500
.05000	.05000	.04500	.04500	.04500	.06500	.23000	.72000

.74000	.51000	.49000	.32000	.33000	.13000	.08500	.06000
.04000	.03500	.03000	.03000	.03000	.02700	.02700	.02700
.02500	.02000	.02000	.02000	.02000	.02000	.02000	.02000
.01500	.01500	.01500	.01500	.01500	.02000	.01300	.01700
.01000	.01000	.01000	.01000	.01000	.01000	.02000	.02000
.02000	.02000	.02000	.02000	.02000	.02000	.02000	.02000
.02000	.02000	.02000	.02000	.02000	.02000	.02000	.02000
.02000	.02000	.02000	.02000	.02000	.02000	.02000	.02000
.02000	.02000	.02000	.02000	.02000	.02000	.02000	.02000
.02000	.02000	.02000	.02000	.02000	.02000	.03500	.03000
.03000	.10000	.23000	.10500	.08500	.11500	.25000	.27000
.13000	.08500	.06500	.07500	.11000	.12000	.13000	.12500
.23000	.33000	.20000	.08500	.14000	.12000	.10000	.06500
.05500	.05000	.06000	.05500	.10500	.07000	.24000	.06000
.06500	.03000	.03500	.03500				
24 T234TMP3 3.		.07500	.08000	.09000	.07500	.15000	.17000
.16000	.22000	.30000	.38000	.46000	.50000	.49000	.46000
.41000	.35000	.31000	.26000	.22000	.19000	.16000	.14500
.13000	.12000	.11500	.11500	.11500	.11000	.10000	.08500
.09000	.09000	.09000	.10000	.12000	.20000	.52000	1.10000
.94000	.70000	.62000	.52000	.52000	.20000	.12000	.10000
.08000	.08500	.08000	.07700	.06700	.06000	.06000	.06000
.06500	.06500	.06000	.05500	.05000	.05000	.05000	.05000
.04000	.04000	.03500	.03500	.03700	.04000	.04000	.04000
.03700	.03500	.03000	.03000	.03000	.03000	.08500	.08500
.08500	.08500	.08500	.08500	.08500	.08500	.08500	.08500
.08500	.08500	.08500	.08500	.08500	.08500	.08500	.08500
.08500	.08500	.08500	.08500	.08500	.08500	.08500	.08500
.08500	.08500	.08500	.08500	.08500	.08500	.08500	.08500
.08500	.08500	.08500	.08500	.08500	.08500	.04700	.05000
.07000	.18000	.50000	.29000	.14000	.17000	.50000	.38000
.18000	.19000	.23000	.20500	.12000	.08500	.09000	.17000
.23000	.34000	.29000	.37000	.28500	.16000	.23000	.32000
.13000	.28000	.15000	.09500	.09000	.19000	.15000	.28000
.04500	.02500	.04000	.02000				
25 PHEX50L1 1.		.05500	.06500	.08000	.07500	.12000	.15500
.18500	.22500	.29000	.37000	.46000	.56000	.68000	.76000
.86000	1.00000	1.05000	1.10000	1.15000	1.15000	1.15000	1.05000
.99000	.90000	.80000	.69000	.62000	.52000	.43500	.35500
.32000	.38000	.45000	.27500	.25000	.34000	.42500	.60000
1.00000	1.25000	1.00000	.85000	1.00000	.74000	.44500	.30000
.22500	.19500	.19000	.17500	.15000	.14000	.13000	.12000
.11500	.11000	.10500	.10000	.10000	.09000	.08500	.08000
.08000	.07000	.07000	.07000	.07000	.07300	.07500	.07500
.06500	.06500	.06500	.06000	.06000	.06000	.08000	.08000
.08000	.08000	.08000	.08000	.08000	.08000	.08000	.08000
.08000	.08000	.08000	.08000	.08000	.08000	.08000	.08000
.08000	.08000	.08000	.08000	.08000	.08000	.08000	.08000
.08000	.08000	.08000	.08000	.08000	.08000	.08000	.08000
.08000	.08000	.08000	.08000	.08000	.08000	.08000	.08000
.08000	.08000	.08000	.08000	.08000	.08000	.12000	.13500
.16000	.29000	.39000	.56000	.44000	.39000	.35500	.34000
.33000	.30500	.27000	.24000	.24000	.20000	.19500	.19500
.22500	.19500	.19000	.24000	.23000	.54000	.86000	.78000
.62000	.51000	.37500	.21500	.32000	.59000	.70000	.18000
.18000	.14000	.19500	.15000				
26 PHEP30L1 1.		.09500	.10000	.10000	.10500	.11500	.13000
.14500	.17000	.19000	.22500	.27000	.31000	.35000	.39000
.43500	.47000	.48500	.51000	.51000	.50000	.47500	.45000

.41000	.36000	.33000	.29500	.25000	.22500	.20000	.19000
.17500	.16000	.16000	.16000	.22500	.24000	.31000	.78000
.62000	.74000	.54000	.52000	.51000	.39000	.23000	.16000
.14000	.12500	.12000	.12000	.10500	.10000	.09700	.09500
.09000	.08700	.08500	.08300	.08000	.08000	.08000	.07700
.07300	.07000	.07000	.07000	.06300	.06000	.05500	.05500
.05500	.05500	.05500	.05500	.05500	.05500	.05000	.05000
.05000	.05000	.05000	.05000	.05000	.05000	.05000	.05000
.05000	.05000	.05000	.05000	.05000	.05000	.05000	.05000
.05000	.05000	.05000	.05000	.05000	.05000	.05000	.05000
.05000	.05000	.05000	.05000	.05000	.05000	.05000	.05000
.05000	.05000	.05000	.05000	.05000	.05000	.05000	.05000
.05000	.05000	.05000	.05000	.05000	.05000	.05000	.05000
.05000	.05000	.05000	.05000	.05000	.05000	.05000	.05000
.05000	.05000	.05000	.05000	.05000	.05000	.05500	.06000
.07000	.10000	.25000	.23000	.23000	.17000	.17000	.16000
.15500	.14000	.12000	.11000	.09700	.09000	.08500	.08500
.10000	.08300	.07000	.10000	.09500	.19000	.30000	.61000
.36000	.21000	.16000	.51000	.10500	.07000	.07000	.08500
.09000	.06500	.06500	.06500				
27 SBUT30L2 2.		.04500	.06000	.06500	.08000	.11500	.15000
.19500	.21500	.25000	.43000	.55000	.70000	.89000	1.00000
1.15000	1.35000	1.40000	1.40000	1.40000	1.35000	1.26000	1.10000
.96000	.77000	.62000	.48000	.41000	.33000	.27500	.24500
.31000	.31000	.19500	.18000	.34000	.35500	1.15500	.32000
.45500	.36500	.41500	.50000	.36000	.23500	.19000	.15500
.13000	.12000	.11500	.11500	.11500	.11000	.09700	.09000
.08500	.08000	.07300	.07300	.07000	.06500	.06000	.06000
.05500	.05500	.04500	.04500	.04000	.04300	.04300	.04000
.04000	.03000	.03000	.03000	.03000	.03000	.06000	.06000
.06000	.06000	.06000	.06000	.06000	.06000	.06000	.06000
.06000	.06000	.06000	.06000	.06000	.06000	.06000	.06000
.06000	.06000	.06000	.06000	.06000	.06000	.06000	.06000
.06000	.06000	.06000	.06000	.06000	.06000	.06000	.06000
.06000	.06000	.06000	.06000	.06000	.06000	.06300	.07500
.12000	.26500	.34500	.62000	.43000	.36500	.37500	.23500
.33500	.38000	.40500	.31500	.20000	.14500	.12500	.16000
.34000	.44000	.27500	.84000	.56000	.96000	.45000	.29000
.62000	.43000	.15000	.34500	1.60000	.22000	.10500	.08000
.08500	.15000	.12500	.11000				
28 S3MBU302 2.		.14000	.14000	.14500	.15500	.18000	.19000
.21000	.24000	.28000	.33000	.44000	.52000	.58000	.65000
.70000	.70000	.72000	.73000	.69000	.64000	.57000	.58000
.45000	.38000	.33000	.29000	.23000	.20500	.19000	.19000
.24000	.21000	.18000	.18000	.20000	.20500	.49000	.47000
.34000	.37000	.32000	.33500	.32500	.26500	.20000	.17500
.16000	.15000	.14500	.15000	.14500	.14000	.14000	.13500
.13500	.13000	.12500	.12500	.12000	.12000	.12000	.11500
.11500	.12000	.11500	.11500	.11000	.11500	.12000	.12000
.11000	.11000	.10700	.10700	.10300	.10000	.10000	.10000
.10000	.10000	.10000	.10000	.10000	.10000	.10000	.10000
.10000	.10000	.10000	.10000	.10000	.10000	.10000	.10000
.10000	.10000	.10000	.10000	.10000	.10000	.10000	.10000
.10000	.10000	.10000	.10000	.10000	.10000	.10000	.10000
.10000	.10000	.10000	.10000	.10000	.10000	.09500	.10000
.11500	.12500	.19000	.32000	.22500	.21000	.20500	.28000
.17000	.15500	.17500	.20000	.18000	.17000	.14000	.12000
.10500	.10000	.11000	.13500	.33500	.28000	.25500	.27000
.20000	.13500	.15500	.10000	.21500	.11000	.26000	.31000
.09500	.08500	.08500	.08700				
29 S2MPE103 2.		.02500	.03300	.04500	.05500	.07500	.12500

.14000	.18000	.22000	.29000	.37000	.43500	.51000	.58000
.63000	.66000	.68000	.68000	.67000	.64000	.61000	.55000
.49000	.41000	.35000	.29000	.23500	.20000	.16500	.13000
.13000	.17000	.29500	.13000	.12000	.13500	.31000	.74000
.58000	.68000	.45000	.38000	.56000	.38000	.20500	.14000
.11000	.09700	.08500	.09500	.08300	.07500	.07500	.07000
.06500	.06000	.05500	.05300	.05000	.05000	.05000	.04500
.05000	.04300	.04500	.04500	.04000	.04000	.04500	.05000
.04500	.03500	.03500	.03500	.03500	.04000	.02000	.03000
.08000	.08000	.08000	.08000	.08000	.08000	.08000	.08000
.08000	.08000	.08000	.08000	.08000	.08000	.08000	.08000
.08000	.08000	.08000	.08000	.08000	.08000	.08000	.08000
.08000	.08000	.08000	.08000	.08000	.08000	.08000	.08000
.08000	.08000	.08000	.08000	.08000	.08000	.09000	.09500
.11000	.15500	.44000	.46000	.30500	.31000	.31000	.26500
.23500	.29000	.25000	.23000	.26000	.25000	.22500	.14500
.12000	.11500	.12000	.26000	.35000	.31500	.24500	.33000
.37000	.40500	.49000	.49000	.15500	.15500	.56000	.26000
.10500	.14000	.14000	.09500				
30 SHEXS0L3 2.		.15000	.18000	.23000	.27500	.37000	.50000
.64000	.80000	1.00000	1.15000	1.30000	1.45000	1.50000	1.40000
1.20000	.90000	.74000	.58000	.46500	.37500	.29500	.24000
.20000	.16000	.11000	.12500	.20000	.33500	.15500	.15000
.25000	.41000	2.00000	.86000	2.00000	1.00000	.80000	.66000
.39000	.19000	.16000	.13000	.11000	.10000	.09500	.09000
.09000	.08500	.07700	.07300	.07000	.07000	.07000	.07000
.06500	.06500	.06000	.05500	.05000	.05000	.05000	.05500
.05500	.05500	.05500	.04500	.05000	.05500	.05500	.05500
.04500	.04300	.04000	.04000	.04500	.04000	.07500	.07500
.07500	.07500	.07500	.07500	.07500	.07500	.07500	.07500
.07500	.07500	.07500	.07500	.07500	.07500	.07500	.07500
.07500	.07500	.07500	.07500	.07500	.07500	.07500	.07500
.07500	.07500	.07500	.07500	.07500	.07500	.07500	.07500
.07500	.07500	.07500	.07500	.07500	.07500	.08500	.08500
.09000	.10000	.18000	.27000	.29000	.26500	.21500	.20500
.21500	.22000	.20000	.19500	.18000	.19000	.18000	.17500
.13000	.11500	.13000	.16000	.24000	.19000	.17000	.14000
.17000	.27000	.29000	.31000	.14500	.15000	.32000	.14000
.10000	.11500	.08500	.07700				
31 T2MBU302 3.		.04000	.05500	.07000	.08000	.12000	.27000
.30000	.38000	.50000	.68000	.90000	1.15000	1.30000	1.50000
1.65000	1.75000	1.75000	1.75000	1.50000	1.50000	1.20000	1.00000
.80000	.63000	.52000	.42000	.31000	.25000	.19500	.18000
.19000	.52000	.27500	.18000	.24000	.58000	1.90000	1.75000
.74000	1.00000	.59000	.56000	.50000	.24000	.17500	.15500
.13000	.12000	.11500	.10000	.09000	.07000	.06000	.05500
.06000	.06000	.06000	.06000	.06000	.06500	.06300	.05500
.04300	.04000	.03000	.03000	.03000	.03000	.02500	.02500
.02500	.02500	.02500	.02000	.01500	.01500	.02000	.02000
.02000	.02000	.02000	.02000	.02000	.02000	.02000	.02000
.02000	.02000	.02000	.02000	.02000	.02000	.02000	.02000
.02000	.02000	.02000	.02000	.02000	.02000	.02000	.02000
.02000	.02000	.02000	.02000	.02000	.02000	.02000	.02000
.02000	.02000	.02000	.02000	.02000	.02000	.02700	.03700
.05500	.15000	.27000	.18000	.27000	.28500	.37000	.40000
.25000	.19500	.14000	.11500	.13000	.22500	.25000	.23000
.32000	.35000	.34000	.29000	.15500	.07500	.08500	.11500
.10500	.23000	.11000	.23500	.26000	.32000	.16500	.15500

.03500	.02000	.01500	.01500				
32 T2MPE402 3.	.12000	.13000	.13000	.13000	.13000	.15000	.16500
.18500	.20500	.23000	.27000	.33000	.38000	.42500	.47000
.48000	.50000	.48000	.46000	.41500	.35000	.33000	.28000
.25000	.22000	.20000	.18000	.16000	.15000	.14000	.13000
.13500	.21000	.17000	.14000	.14000	.19500	.42000	.52000
.28000	.38000	.29000	.24000	.20000	.17000	.15000	.13000
.12000	.11000	.11000	.11000	.11000	.10500	.10500	.10500
.10000	.10000	.10000	.10000	.10000	.10000	.10000	.09500
.09500	.09000	.09000	.09000	.09000	.09300	.09000	.09500
.09000	.08500	.08500	.08500	.08500	.08500	.08500	.08500
.08500	.08500	.08500	.08500	.08500	.08500	.08500	.08500
.08500	.08500	.08500	.08500	.08500	.08500	.08500	.08500
.08500	.08500	.08500	.08500	.08500	.08500	.08500	.08500
.08500	.08500	.08500	.08500	.08500	.08500	.08500	.08500
.08500	.08500	.08500	.08500	.08500	.08500	.08500	.08500
.09000	.10000	.19000	.15000	.14500	.15500	.20000	.26000
.16000	.11000	.12500	.14500	.14500	.17000	.15000	.16500
.17500	.23500	.28000	.18000	.11500	.14000	.11500	.08500
.08700	.14500	.13500	.13000	.19500	.16000	.26500	.11000
.07500	.06500	.06700	.06500				
33 T3MHX503 3.	.11000	.12000	.12000	.12000	.12000	.15500	.16500
.18500	.20000	.24500	.29500	.37000	.42500	.46000	.48000
.49000	.48000	.45000	.41000	.35000	.29500	.25000	.21000
.19000	.17000	.15500	.14000	.13500	.12500	.12000	.12000
.12500	.25000	.17500	.13500	.11500	.22000	1.00000	.76000
.50000	.60000	.36000	.37000	.25000	.19500	.14500	.12500
.11500	.11000	.11000	.10500	.10000	.10000	.09700	.10000
.09700	.09500	.09500	.09500	.09500	.09500	.09000	.09000
.09000	.09000	.09000	.08500	.09500	.09000	.08500	.09000
.08500	.08500	.08300	.08000	.08000	.07700	.08000	.08000
.08000	.08000	.08000	.08000	.08000	.08000	.08000	.08000
.08000	.08000	.08000	.08000	.08000	.08000	.08000	.08000
.08000	.08000	.08000	.08000	.08000	.08000	.08000	.08000
.08000	.08000	.08000	.08000	.08000	.08000	.08000	.08000
.08000	.08000	.08000	.08000	.08000	.08000	.08500	.08500
.09500	.10500	.29000	.20500	.17500	.18500	.25000	.25000
.16000	.16500	.18500	.19000	.18500	.16000	.15000	.13000
.19500	.29500	.33500	.26000	.14000	.16500	.13000	.18000
.15500	.25000	.20000	.11000	.13500	.43000	.31000	.10000
.16000	.07000	.06500	.06500				
34 T3EHX503 3.	.06500	.07000	.07000	.07000	.07000	.11000	.12000
.14000	.16000	.17500	.24500	.30000	.33000	.34000	.33000
.31500	.29000	.26000	.22500	.18500	.16000	.14000	.11500
.10000	.09000	.08500	.08000	.07500	.07000	.07000	.07000
.07500	.11500	.14500	.09300	.09500	.16500	.40000	.88000
.55000	.60000	.34000	.35000	.22000	.16000	.11000	.08000
.07000	.06500	.06500	.06500	.06000	.06000	.05500	.05500
.05500	.05500	.05500	.05300	.05000	.05000	.05000	.05000
.04700	.04500	.04500	.04700	.04500	.04500	.05000	.04500
.04500	.04300	.04000	.04000	.04000	.04000	.04000	.04000
.04000	.04000	.04000	.04000	.04000	.04000	.04000	.04000
.04000	.04000	.04000	.04000	.04000	.04000	.04000	.04000
.04000	.04000	.04000	.04000	.04000	.04000	.04000	.04000
.04000	.04000	.04000	.04000	.04000	.04000	.04000	.04000
.04000	.04000	.04000	.04000	.04000	.04000	.04000	.04000
.04000	.04000	.04000	.04000	.04000	.04000	.04500	.04500
.05000	.06000	.25500	.26000	.12500	.13500	.17000	.15500
.12500	.13500	.14000	.12000	.14500	.11500	.11500	.07500

.08300	.17500	.23500	.21500	.10000	.10000	.08000	.08500
.11000	.13000	.19000	.37000	.31500	.21500	.33000	.06500
.12000	.04500	.04000	.04300				
35 TDHXS03 3.		.04700	.05500	.05500	.08000	.12500	.15000
.22000	.28000	.35000	.44000	.53000	.57000	.57000	.56000
.54000	.50000	.43500	.36000	.30000	.24000	.19500	.16500
.13500	.12000	.10500	.09300	.08500	.07500	.07300	.07000
.09000	.21000	.13500	.10000	.12000	.23000	.98000	.86000
.84000	.78000	.54000	.56000	.33500	.24000	.13500	.09500
.07000	.06500	.06500	.07000	.05500	.05000	.05000	.05000
.05000	.05000	.04500	.04500	.04500	.05000	.04700	.04500
.04500	.04500	.04500	.04500	.04500	.04000	.03700	.03000
.03500	.04000	.03500	.03500	.03500	.03500	.05000	.05000
.05000	.05000	.05000	.05000	.05000	.05000	.05000	.05000
.05000	.05000	.05000	.05000	.05000	.05000	.05000	.05000
.05000	.05000	.05000	.05000	.05000	.05000	.05000	.05000
.05000	.05000	.05000	.05000	.05000	.05000	.05000	.05000
.05000	.05000	.05000	.05000	.05000	.05000	.05700	.05700
.06500	.08500	.28000	.52000	.29000	.22500	.31000	.54000
.25000	.23000	.25500	.20000	.23000	.16500	.19000	.25000
.19000	.25000	.32500	.40000	.42000	.11500	.07000	.27000
.20000	.14000	.25000	.11000	.06500	.37000	.31000	.52000
.31000	.07000	.06700	.06000				
36 T4MHP104 3.		.07000	.08000	.08300	.09700	.15500	.17500
.22000	.25500	.32500	.43000	.55000	.65000	.70000	.74000
.76000	.75000	.73000	.66000	.57000	.46000	.37000	.29500
.24500	.20000	.17500	.15500	.13000	.11500	.11000	.10500
.11500	.37000	.23000	.14000	.16500	.31000	.72000	1.25000
.90000	1.00000	.82000	.61000	.58000	.38000	.22000	.14500
.10500	.09500	.09500	.10000	.08300	.08000	.07300	.07500
.07500	.07500	.07500	.07000	.07000	.06500	.06500	.06000
.06000	.06000	.05500	.06000	.05500	.05500	.05500	.05700
.05500	.05000	.05000	.05000	.04700	.04500	.05000	.05000
.05000	.05000	.05000	.05000	.05000	.05000	.05000	.05000
.05000	.05000	.05000	.05000	.05000	.05000	.05000	.05000
.05000	.05000	.05000	.05000	.05000	.05000	.05000	.05000
.05000	.05000	.05000	.05000	.05000	.05000	.06500	.07000
.07700	.10500	.53000	.37000	.28000	.30000	.44000	.47000
.30000	.21000	.28500	.34000	.29500	.28500	.27000	.17000
.32500	.54000	.63000	.45000	.21500	.27000	.27500	.19500
.32000	.32000	.21000	.15000	.37500	.54000	.66000	.28000
.19500	.09500	.07000	.07300				
37 T4MHPD04 3.		.09000	.09700	.10500	.10500	.12000	.16000
.20500	.22000	.23000	.28000	.34000	.39000	.42000	.42500
.42500	.41000	.38500	.34500	.30000	.25500	.21500	.18500
.17000	.15000	.13500	.12500	.11500	.10500	.10000	.10000
.10500	.24000	.24000	.12500	.13000	.23500	.33000	.44000
.23500	.42000	.35000	.24000	.18000	.17000	.14000	.10500
.08700	.08500	.08500	.08300	.08300	.08000	.07700	.07500
.07500	.07500	.07700	.07700	.07500	.07500	.07300	.07000
.07000	.07000	.07000	.07000	.06700	.06700	.07000	.07000
.06500	.06500	.06500	.06500	.06500	.06000	.06000	.06000
.06000	.06000	.06000	.06000	.06000	.06000	.06000	.06000
.06000	.06000	.06000	.06000	.06000	.06000	.06000	.06000
.06000	.06000	.06000	.06000	.06000	.06000	.06000	.06000
.06000	.06000	.06000	.06000	.06000	.06000	.06000	.06000
.06000	.06000	.06000	.06000	.06000	.06000	.06000	.06300