

UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE QUÍMICA
Departamento de Físico Química

APLICAÇÃO DE MÉTODOS QUIMIOMÉTRICOS
EM
SEPARAÇÃO DE ESPECTROS
E
RECONHECIMENTO DE PADRÕES

DISSERTAÇÃO DE MESTRADO

MARLON MARTINS DOS REIS

Orientadora -Prof^a. Dr^a. MÁRCIA MIGUEL DE CASTRO FERREIRA

(Outubro-1997)

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

R277a Reis, Marlon Martins dos
Aplicação de métodos quimiométricos em separação de
espectros e reconhecimento de padrões / Marlon Martins
dos Reis. - - Campinas, [SP : s.n.], 1997.

Orientadora: Márcia Miguel de Castro Ferreira.

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

I. Quimiometria. I. Ferreira, Márcia Miguel de Cas-
tro. II. Universidade Estadual de Campinas. Instituto de
Química. III. Título.

AGRADECIMENTOS

Seriam necessárias milhares de páginas para agradecer todo apoio recebido ao longo da execução desta dissertação, mas espero poder retribuir não apenas através de palavras e linhas mas sim com amizade e semelhante apoio.

Aos meus Pais, Irmãos e Irmã, meus Familiares, minha Orientadora e meus Amigos o meu MUITO OBRIGADO.

Também à CAPES pelo auxílio em forma de bolsa de estudo e à UNICAMP os meus agradecimentos.

Marlon Martins dos Reis

Resumo

Duas aplicações quimiométricas foram apresentadas: a primeira em um problema que envolve separação de espectros e a segunda em um caso de reconhecimento de padrões.

O Método Generalizado de Anulação do Posto “GRAM” foi usado como método de separação de espectros. Ele foi aplicado em dados simulados na elaboração de uma metodologia para a solução única de certo problema pelo método. Espectros sobrepostos, de ordem superior, de luminescência total para depósitos de cálculos dentários em animais e humanos foram usados para mostrar a aplicabilidade do GRAM. O método “Self Modelling Curve Resolution”, desenvolvido por Lawton e Sylvestre, foi usado para complementar a resolução do GRAM. Esta aplicação resultou em quatro espectros de espécies porfirínicas para ambas as amostras, felinos e humanos, sendo a resolução melhor que em casos anteriores.

O método “Soft Independent Modelling of Class Analogy-SIMCA”, com Análise de Componentes Principais e Análise por Agrupamento Hierárquico foram usados em um caso de reconhecimento de padrões. Estes métodos foram aplicados em informações espectrais, na região do ultra violeta, para a identificação de estruturas moleculares de 95 flavonóides. O reconhecimento atingiu seu objetivo, pois foi possível fazer a identificação das estruturas moleculares em conjunto com métodos quimiométricos

Abstract

Two chemometric applications were presented, first in a curve resolution problem and second in a pattern recognition case.

The Generalized Rank Annihilation Method “GRAM” was used as a curve resolution method. It was applied to a simulated data provide the general guidelines concerning the condition for uniqueness of a solution for a given method. The unresolved second-order porphyrinic emission-excitation fluorescence spectra of several animal and human dental calculus deposits were used to show the applicability of the GRAM. Self Modelling Curve Resolution method, developed by Lawton and Sylvestre, was used to accomplish the curve resolution of GRAM. This application resulted in four porphyrinic spectral profile for both humans and felines giving a better resolution than prior case.

The Soft Independent Modelling of Class Analogy “SIMCA” , with Principal Component Analysis and Hierarchical Cluster Analysis were used in a pattern recognition case. These methods were applied to UV-spectra of 95 flavonoids to differentiate their molecular structure. The pattern recognition was successful making possible the identification of molecular structure using UV-spectra information and chemometrics methods.

<u>Introdução</u>	1
<u>Objetivos</u>	2
<u>Capítulo 1</u>	3
<u>1.1-Introdução ao Capítulo 1</u>	4
<u>1.2-Os métodos</u>	5
<u>1.2.1 Método Generalizado de Anulação do Posto</u>	5
<u>1.2.2 O Método de Lawton e Sylvestre</u>	18
<u>1.3-Resultados e Discussões</u>	22
<u>1.4-Conclusões</u>	35
<u>Capítulo 2</u>	36
<u>2.1-Introdução ao Capítulo 2</u>	37
<u>2.2-Métodos</u>	38
<u>2.2.1-Notação</u>	38
<u>2.2.2-Análise de Componentes Principais - ACP</u>	38
<u>2.2.3-SIMCA</u>	43
<u>2.2.4-Análise por agrupamento hierárquico</u>	45
<u>2.3-Flavonóides</u>	48
<u>2.4-Os dados</u>	52
<u>2.5-As classes</u>	52
<u>2.6-Validação do modelo</u>	54
<u>2.7-Resultados</u>	57
<u>2.8-Conclusões</u>	81
<u>ANEXO</u>	82
<u>REFERÊNCIAS BIBLIOGRÁFICAS</u>	100

Introdução

A aplicação de métodos quimiométricos tem se difundido nas mais diversas áreas da química e ciências afins, pois o desenvolvimento de novos instrumentais aplicados aos laboratórios de pesquisas, que passaram a fornecer uma grande quantidade de informações, criou a necessidade do desenvolvimento de uma ferramenta estatística sofisticada aplicada à química. Este grande número de informações e tal necessidade, aliados a uma facilidade crescente de acesso a micro computadores, a cada dia mais rápidos, deu origem ao crescimento da quimiometria.

Nesta dissertação, são apresentadas duas aplicações de métodos quimiométricos em áreas distintas da química.

Na primeira aplicação, um problema envolvendo espectros sobrepostos é resolvido quimiometricamente, possibilitando, assim, a identificação de substâncias que participam na constituição de amostras de cálculo dentário. Neste caso, são espectros de luminescência total resolvidos pelo Método Generalizado de Anulação do Posto, complementado pelo método desenvolvido por Lawton e Sylvestre.

Quanto à segunda aplicação, métodos quimiométricos são usados na elucidação de estruturas moleculares de flavonóides a partir de informações espectrais na região do ultra violeta. Aqui, o SIMCA “Soft Independent Modelling of Class Analogy” , Análise de Componentes Principais e Análise por Agrupamento Hierárquico são usados no tratamento do conjunto de dados dos flavonóides.

Objetivos

O principal objetivo desta dissertação é apresentar metodologias de trabalhos em tratamentos de dados através de métodos quimiométricos. Dentre as diversas áreas englobadas pela quimiometria as escolhidas foram: separação de espectros e reconhecimento de padrões. A primeira, por ser de grande importância em química analítica, no que diz respeito a identificação de substâncias. A segunda, para mostrar um melhor aproveitamento dos dados obtidos durante trabalhos de elucidação de estruturas.

Em separação de espectros, os conjuntos dados, tanto simulado quanto experimental, foram escolhidos segundo sua complexidade, isto é, picos bastante sobrepostos, para que a metodologia adotada em seus tratamentos seja o mais ampla possível. Neste caso, foram usados dois métodos quimiométricos em conjunto.

Para reconhecimento de padrões, o conjunto de dados, composto por informações espectrais na região do ultra violeta de flavonóides, foi escolhido para mostrar uma forma alternativa no trabalho de elucidação de estrutura moleculares através da quimiometria. Neste caso, foram usados três métodos quimiométricos em conjunto.

Outro objetivo é mostrar que a quimiometria pode fornecer soluções práticas, rápidas e de fácil aplicação para a solução de problemas experimentais.

O Método Generalizado de Anulação do Posto e o método desenvolvido por Lawton e Sylvestre foram aplicados através de algoritmos escritos a partir de seus fundamentos, utilizando rotinas internas do MATLAB versão para WINDOWS. Quanto aos métodos SIMCA, Análise por Componentes Principais e Análise por Agrupamento Hierárquico foram aplicados através do TOOLBOX para o MATLAB versão para WINDOWS.

Capítulo 1

1.1-Introdução ao Capítulo 1

O desenvolvimento de instrumentos combinados^{F2,F3}, como cromatografia líquida de alta eficiência (“HPLC”) tendo como detector um espectrofotômetro UV-Visível, espectros de emissão-excitação, etc., tem colocado o químico frente a uma grande quantidade de informações numéricas. Nestes casos, suas respostas experimentais são chamadas de dados de ordem superior^{C1}. Para o tratamento de tais informações, a quimiometria tem desenvolvido vários métodos como o GRAM^{C3} do inglês “*Generalized Rank Annihilation Method*”, PARAFAC^{C8} do inglês “*Parallel Factor Analysis*” e TLD^{C4} do inglês “*Trilinear Decomposition*”, os quais têm como aplicação a separação de sinais^{C5}, necessária por exemplo, quando ocorre sobreposição de espectros e também a calibração tensorial^{C2}.

Neste trabalho, é apresentado o Método Generalizado de Anulação do Posto^{C3}, o GRAM, sendo ele a extensão desenvolvida por Wilson, Sanchez e Kowalski aos primeiros métodos de anulação do posto^{B6,B7,B8}. O GRAM se aplica ao caso geral onde não é necessário que as duas amostras analisadas tenham todas as substâncias em comum. Neste caso, a solução é conseguida através da resolução de um problema generalizado de autovalores-autovetores. Além do GRAM, também é aplicado aquele que foi o primeiro método desenvolvido para separação de espectros, sem fazer restrições à forma da curva e sem processos iterativo. Este último, foi desenvolvido por Lawton e Sylvestre^{B3} e se aplica apenas a misturas constituída por duas substâncias químicas.

A primeira aplicação para o GRAM é apresentada em dados simulados. A segunda, tem como objeto de trabalho espectros de luminescência total coletados para amostras de cálculo dentário em felinos e humanos, sendo que, seus resultados são complementados pelo método de Lawton e Sylvestre.

A simulação é importante para a apresentação do método, pois com ela é possível apresentar a diferença entre a solução matemática para dados perfeitos, neste caso aqueles simulados, e dados experimentais.

1.2-Os métodos

1.2.1 Método Generalizado de Anulação do Posto

Para facilitar a compreensão das técnicas algébricas utilizadas no GRAM, foi simulado um conjunto de espectros, representando as respostas de um experimento hipotético. Onde, duas amostras constituídas por misturas de duas substâncias químicas, são submetidas à cromatografia líquida, cujo o detector é um espectrofotômetro, de forma que para cada tempo de eluição é coletado um espectro. Os cromatogramas foram simulados contendo vinte tempos de eluição, de 0 a 20 minutos com 1 minuto de intervalo, e os espectros em cinquenta comprimentos de onda, 400-449nm com intervalo de 1nm. Os espectros e cromatogramas foram normalizados pelo maior pico, isto é, os valores encontrados nos cromatogramas e espectros foram divididos pelos valores de seus respectivos máximos. A Figura 1.1 apresenta as curvas, cromatogramas e espectros, para as duas substâncias hipotéticas.

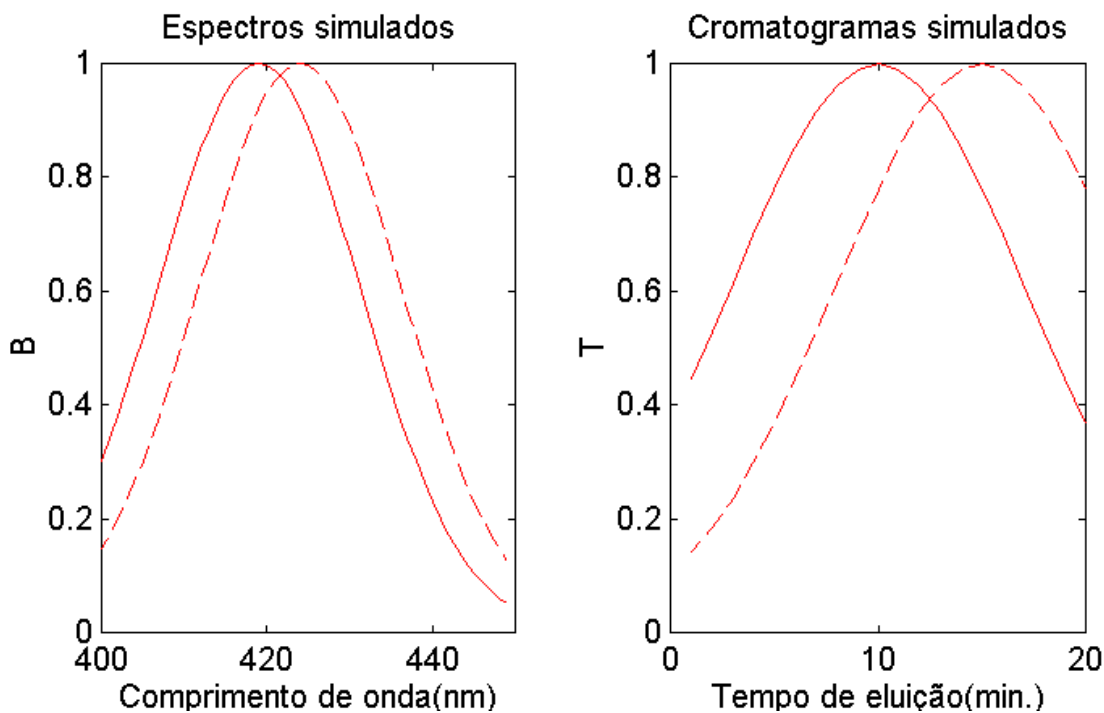


Figura 1.1- Espectros e cromatogramas simulados, normalizados pelo pico de máximo. B - absorvância molar. T constante referente ao tempo de eluição.

Admitindo que a Lei de Beer^{F1} é seguida, considerando o caminho ótico igual a um centímetro, a relação entre absorvância em cada comprimento de onda e o respectivo tempo de eluição, será dada pela equação (1.1).

$$a_{ki} = c_A \cdot t_k \cdot b_i \quad (1.1)$$

onde:

a_{ki} é a absorvância no tempo de eluição k e no comprimento de onda i .

c_A é a concentração da substância A . ($\text{mol} \cdot \text{l}^{-1}$).

t_k é a constante referente ao tempo de eluição k .

b_i é a absorvância molar no comprimento de onda i . ($\text{l} \cdot \text{mol}^{-1} \cdot \text{cm}^{-1}$).

Para k tempos de eluição e i comprimentos de onda tem-se a equação (1.2).

$$\mathbf{A} = c_A \cdot \begin{bmatrix} t_1 \\ t_2 \\ t_3 \\ \vdots \\ t_k \end{bmatrix} \times [b_1 \quad b_2 \quad b_3 \quad \dots \quad b_i] \quad (1.2)$$

onde \mathbf{A} é uma matriz ($k \times i$) na qual as linhas correspondem a espectros de absorvência. Graficamente, esta matriz representa uma superfície de resposta.

Considerando agora uma amostra S constituída pela mistura das substâncias A e B . Os espectros desta amostra S correspondem à soma dos espectros de A e B , em suas respectivas concentrações. Em termos matriciais esta amostra é representada pela equação (1.3).

$$\mathbf{S} = \mathbf{X} \mathbf{C} \mathbf{Y}^t \quad (1.3)$$

onde :

\mathbf{S} é uma matriz contendo os espectros e representa uma superfície de resposta para a amostra S ;

\mathbf{X} é uma matriz na qual as colunas correspondem aos cromatogramas, normalizados para concentração unitária, das substâncias puras ;

\mathbf{C} é uma matriz diagonal contendo concentrações das substâncias puras;

\mathbf{Y} é uma matriz onde as colunas correspondem às absorvências molares das substâncias puras;

A equação (1.3) pode ser estendida ao caso geral para este tipo de dados. Por exemplo, \mathbf{X} poderia ser espectros de intensidade de excitação, \mathbf{Y} espectros de intensidade de emissão e \mathbf{C} as concentrações das espécies luminescentes.

Para se aplicar o método GRAM são necessárias duas amostras, constituídas por misturas de substâncias químicas comuns a estas duas, mas em concentrações diferentes. Desta forma, os cromatogramas e

espectros, simulados para as substâncias A e B , foram colocados em suas formas matriciais e substituídos na equação (1.3) (ver equações (1.3.1) e (1.3.2)), para dar origem a duas matrizes de espectros simulados, correspondentes às amostras hipotéticas M e N . As concentrações usadas para A e B são as seguintes:

-Para M : 2 unidades de concentração de A e 5 unidades de concentração de B ;

- Para N : 7 unidades de concentração de A e 3 unidades de concentração de B .

$$M = \begin{bmatrix} t_{A1} & t_{B1} \\ t_{A2} & t_{B2} \\ t_{A3} & t_{B3} \\ \vdots & \vdots \\ t_{Ak} & t_{Bk} \end{bmatrix} \times \begin{bmatrix} 2 & 0 \\ 0 & 5 \end{bmatrix} \times \begin{bmatrix} b_{A1} & b_{A2} & b_{A3} & \cdots & b_{Ai} \\ b_{B1} & b_{B2} & b_{B3} & \cdots & b_{Bi} \end{bmatrix} \quad (1.3.1)$$

$$N = \begin{bmatrix} t_{A1} & t_{B1} \\ t_{A2} & t_{B2} \\ t_{A3} & t_{B3} \\ \vdots & \vdots \\ t_{Ak} & t_{Bk} \end{bmatrix} \times \begin{bmatrix} 7 & 0 \\ 0 & 3 \end{bmatrix} \times \begin{bmatrix} b_{A1} & b_{A2} & b_{A3} & \cdots & b_{Ai} \\ b_{B1} & b_{B2} & b_{B3} & \cdots & b_{Bi} \end{bmatrix} \quad (1.3.2)$$

onde:

t_{Ak} e t_{Bk} são as constantes dos tempos de eluição para as substâncias A e B , respectivamente, ou seja, os valores correspondentes aos pontos das curvas simuladas para cromatogramas.

b_{Ai} e b_{Bi} são as absorvidades molares para as substâncias A e B , respectivamente, ou seja, os valores correspondentes aos pontos das curvas simuladas para espectros.

A Figura 1.2 apresenta as superfícies das amostras M e N .

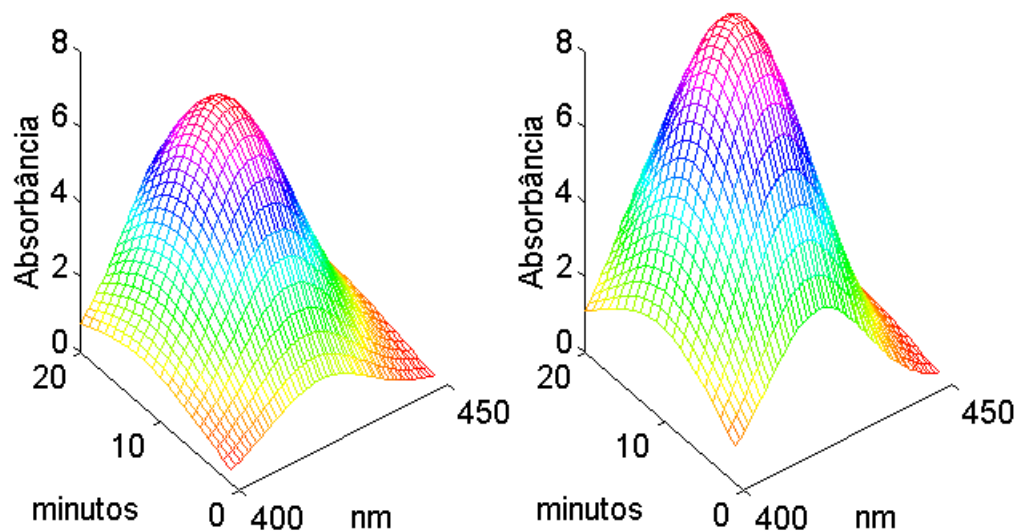


Figura 1.2 - Superfícies de respostas correspondentes às amostras simuladas M (direita) e N (esquerda).

Em termos vetoriais os espectros correspondentes a estas amostras são combinações lineares^{A1,C5} dos espectros das substâncias puras A e B . Como as amostras M e N diferem entre si apenas nas concentrações das substâncias A e B , então suas matrizes correspondentes podem ser representadas pelo mesmo conjunto de autovetores, mas com autovalores^{A2,A3,C3} diferentes (equações (1.4) e (1.5)).

$$\mathbf{M} \Psi = \Psi \lambda_{\mathbf{M}} \quad (1.4)$$

$$\mathbf{N} \Psi = \Psi \lambda_{\mathbf{N}} \quad (1.5)$$

onde :

Ψ - é a matriz de autovetores ;

λ - é a matriz de autovalores ;

Como a matriz de autovalores é uma matriz diagonal e sua inversa é a matriz diagonal cujos elementos são os inversos daqueles de sua diagonal principal, pode se escrever as equações (1.6) e (1.7).

$$\mathbf{M} \Psi \lambda_{\mathbf{M}}^{-1} = \Psi \quad (1.6)$$

$$\mathbf{N} \Psi \lambda_{\mathbf{N}}^{-1} = \Psi \quad (1.7)$$

A igualdade dos autovetores nas equações (1.6) e (1.7) permite que se iguale o lado esquerdo destas equações e como resultado tem-se a equação (1.8) e, por conseguinte, a (1.9).

$$\mathbf{M} \Psi \lambda_M^{-1} = \mathbf{N} \Psi \lambda_N^{-1} \quad (1.8)$$

$$\mathbf{M} \Psi = \mathbf{N} \Psi \lambda_N^{-1} \lambda_M \quad (1.9)$$

O produto dos autovalores, $\lambda_N^{-1} \lambda_M$, resulta em uma matriz diagonal representada pela equação (1.10), onde, Λ é a matriz cujos elementos são dados pela relação entre os autovalores, que representam as concentrações das substâncias A e B , das matrizes correspondentes às amostras M e N .

$$\Lambda = \lambda_N^{-1} \lambda_M \quad (1.10)$$

A substituição da equação (1.10) em (1.9) da origem à equação (1.11):

$$\mathbf{M} \Psi = \mathbf{N} \Psi \Lambda \quad (1.11)$$

A equação (1.11) mostra que para se encontrar os autovetores é necessário resolver um problema de autovetores-autovalores generalizado^{C6,A4}. Para isto, as matrizes devem ser quadradas e não existir correlações entre colunas e entre linhas, para que seus determinantes sejam diferentes de zero. Para evitar correlação, as matrizes \mathbf{M} e \mathbf{N} devem ser projetadas em uma base ortogonal, porque assim as linhas e colunas passam a ser representadas por vetores linearmente independentes. A base pode ser encontrada através de uma decomposição em valores singulares^{A2,A3}.

Após esta decomposição, a matriz original passa a ser representada pelo produto de três matrizes, duas ortonormais e uma diagonal. As ortonormais possuem seus vetores ortogonais entre si e com módulo unitário. A matriz diagonal possui os valores singulares da matriz original.

Como a base do espaço vetorial deve ser a mesma para as duas matrizes correspondentes às amostras M e N , então é necessário usar informações destas duas para efetuar a decomposição em valores singulares. Para tal, são criadas as matrizes justapostas $(M | N)$, M ao lado de N e $\begin{pmatrix} M \\ N \end{pmatrix}$ M acima de N . As equações (1.12) e (1.13) apresentam as decomposições para as matrizes justapostas.

$$(M | N) = P S_1 V^t \quad (1.12)$$

$$\begin{pmatrix} M \\ N \end{pmatrix} = U S_2 Q^t \quad (1.13)$$

onde U , P , V e Q são as matrizes ortonormais e S_1 e S_2 as matrizes diagonais com os valores singulares.

A Tabela 1.1 apresenta os cinco primeiros valores singulares para as decomposições das matrizes correspondentes aos espectros simulados. Destes valores singulares, apenas dois são diferentes de zero porque só existem duas substâncias diferentes nas amostras M e N , ou seja, apenas dois vetores linearmente independentes^{A2,A3} são necessários para formar uma base para estas matrizes.

Os quadrados dos valores singulares estão ligados à proporção da variação nos novos vetores^{B1}, ou seja, quanto cada novo vetor está sendo importante na representação das matrizes originais. Desta forma, os valores singulares são dispostos na matriz diagonal de forma decrescente, isto é, daquele de maior importância para o de menor. Portanto, para M e N onde apenas dois vetores ortogonais são necessários para descrever as matrizes originais, só os dois primeiros valores singulares, correspondentes

aos vetores de maior importância, serão diferentes de zero. O número de valores singulares diferentes de zero é igual ao posto da matriz.

A relação entre valor singular e variação dos dados também explica a grande importância do primeiro valor singular da Tabela 1.1. Tal importância, é devida ao fato de o vetor, correspondente ao primeiro valor singular, ter a mesma direção que o vetor média dos espectros, onde esta presente a maior variação dos dados, resultando assim em um alto valor singular.

Tabela 1.1- Valores singulares das matrizes, correspondentes às amostras simuladas, para os cinco primeiros vetores ortonormais de maior importância. (V.sing./soma total= o valor singular dividido pela soma total dos valores singulares).

Vetor n ^o	(M N)		$\left(\frac{M}{N}\right)$	
	Valores singulares	V.sing./soma total	Valores singulares	V.sing./soma total
1	185,9297	0,9192	186,2927	0,9419
2	16,3425	0,0808	11,4879	0,0581
3	0	0	0	0
4	0	0	0	0
5	0	0	0	0

Em se tratando de dados experimentais, com a presença de ruídos, o número de valores singulares será diferente do posto da matriz, pois um número maior de vetores ortogonais será necessário para descrever tais ruídos^{C5}. Por este motivo, é necessário escolher o número de vetores importantes para a representação da matriz original e eliminar aqueles usados para descrever os ruídos.

Para mostrar tal influência dos ruídos, um novo conjunto de dados foi criado, a partir dos espectros e cromatogramas já simulados. Mas agora, com a adição de interferentes mais ruídos aleatórios, gerados com média igual a zero e variância um. Ainda sobre os ruídos, sua adição à matriz de espectros simulados é feita após a elaboração de uma matriz de ruídos,

que é multiplicada elemento a elemento pela matriz de espectros na qual será somada e por um fator de 0,25.

A Tabela 1.2 apresenta os valores singulares para esta última simulação, onde se pode verificar que a variação, a partir do terceiro valor singular, é menos “significativa”, se comparada com aquela existente entre o primeiro e o segundo valores. Para facilitar a compreensão da significância dos valores singulares, vale lembrar o exemplo dos espectros simulados sem interferentes e ruídos, onde de 0,9192 no primeiro valor singular caiu para 0,0808 no segundo e depois para zero no terceiro valor singular, permanecendo constante nos valores seguintes, ou seja a variação igual a zero. Agora voltando ao caso da simulação com ruídos e interferentes, a variação entre o primeiro e o segundo valores singulares é de 0,8087 para 0,0707 e deste para 0,0117 o terceiro valor singular, de onde a variação passa a ser praticamente constante semelhante ao simulado. Ainda sobre os valores citados nestas comparações, eles correspondem às matrizes justapostas verticalmente para simulações com e sem ruídos e interferentes. Aquela queda na significância da variação entre valores singulares, é uma indicação de que apenas os dois primeiros vetores, correspondentes aos dois primeiros valores singulares, são importantes para descrever as matrizes justapostas.

Tabela 1.2- Valores singulares da matrizes, correspondentes às amostras simuladas com adição de interferentes e ruídos, para os cinco primeiros vetores ortonormais de maior importância. (V.sing./soma total = o valor singular dividido pela soma total dos valores singulares).

Vetor	(M N)		$\left(\frac{M}{N}\right)$	
	Valores singulares	V.sing./soma total	Valores singulares	V.sing./soma total
1	190,4566	0,8070	190,8266	0,7989
2	16,6865	0,0707	11,6497	0,0488
3	2,7715	0,0117	2,7563	0,0115
4	2,5383	0,0108	2,5157	0,0105
5	2,4590	0,0104	2,3320	0,0098
6	2,2540	0,0096	2,1750	0,0091

Conhecido o número de vetores necessários para formar a base para as matrizes \mathbf{M} e \mathbf{N} , são feitas as projeções mostradas nas equações (1.14) e (1.15), onde \mathbf{M}_{pq} e \mathbf{N}_{pq} são as matrizes correspondentes a \mathbf{M} e \mathbf{N} na nova base.

$$\mathbf{M}_{pq} = \bar{\mathbf{P}}' \mathbf{M} \bar{\mathbf{Q}} \quad (1.14)$$

$$\mathbf{N}_{pq} = \bar{\mathbf{P}}' \mathbf{N} \bar{\mathbf{Q}} \quad (1.15)$$

Obs.: O traço acima de \mathbf{P} e \mathbf{Q} indica que estão truncadas, isto é, só os vetores importantes estão sendo usados, tendo sido eliminados aqueles correspondentes aos ruídos experimentais.

Agora o problema de autovetores-autovalores generalizado, equação (1.16), pode ser resolvido, pelo algoritmo QZ^{A4}.

$$\mathbf{M}_{pq} \boldsymbol{\Psi} = \mathbf{N}_{pq} \boldsymbol{\Psi} \boldsymbol{\Lambda} \quad (1.16)$$

A matriz de cromatogramas será dada pela equação (1.17) e a de espectros pela equação(1.18).

$$\mathbf{X} = \mathbf{P} (\mathbf{M}_{pq} + \mathbf{N}_{pq}) \boldsymbol{\Psi} \quad (1.17)$$

$$\mathbf{Y} = \mathbf{Q} (\boldsymbol{\Psi}^{-1})^t \quad (1.18)$$

A Figura 1.3 apresenta o resultado para o conjunto de dados simulado, sem ruído e interferentes. A proporção das concentrações $\boldsymbol{\Lambda}$ é 3,5 e 0,6 que correspondem a 7 unidades de concentração em N dividido por 2 unidades de concentração em M para a substância A e 3 unidades de concentração em N dividido por 5 unidades de concentração em M para a substância B .

Quanto ao segundo conjunto de dados simulados, com interferentes e ruídos, na Figura 1.4 são mostrados os espectros e cromatogramas dos interferentes, as superfícies após a adição de ruídos e interferentes. Para este conjunto, o GRAM foi aplicado para posto das matrizes igual a dois e depois igual a três. Estes dois testes têm por objetivo mostrar que o terceiro vetor esta sendo usado para descrever ruídos e interferentes. Os

resultados destes testes são mostrados na Figura 1.5, onde é verificado, a distribuição aleatória no terceiro vetor.

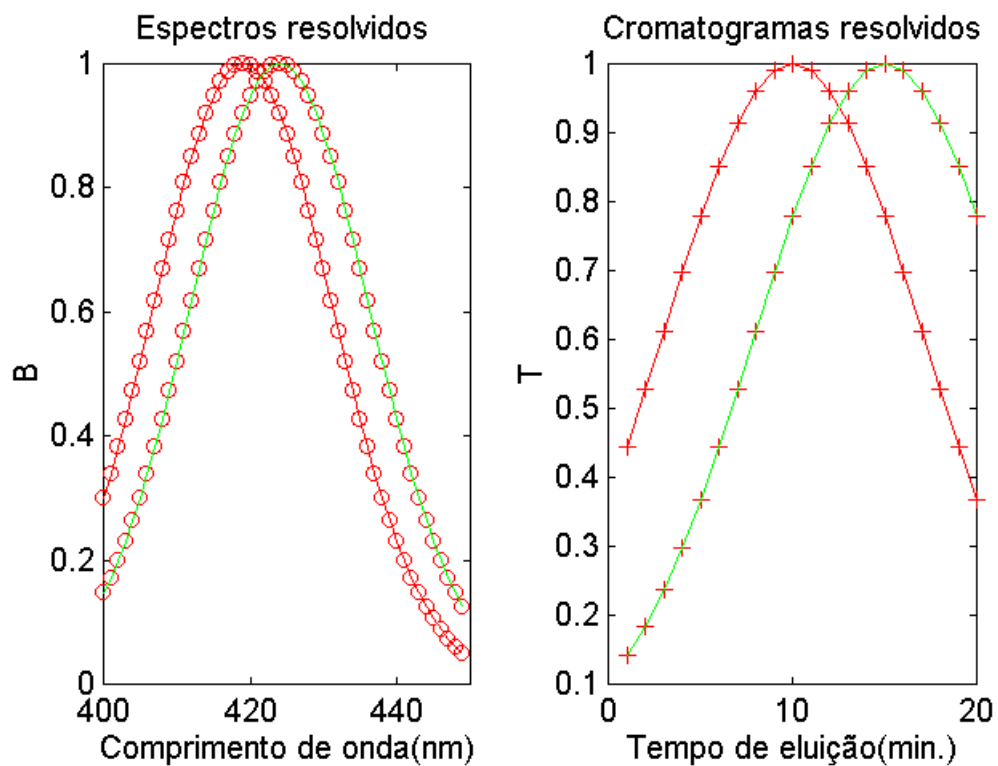


Figura 1.3- Resultado do GRAM para as amostras simuladas M e N . (o) e (+) valores reais (—) resultado da separação dos espectros. B - absorvidade molar; T- constante referente ao tempo de eluição.

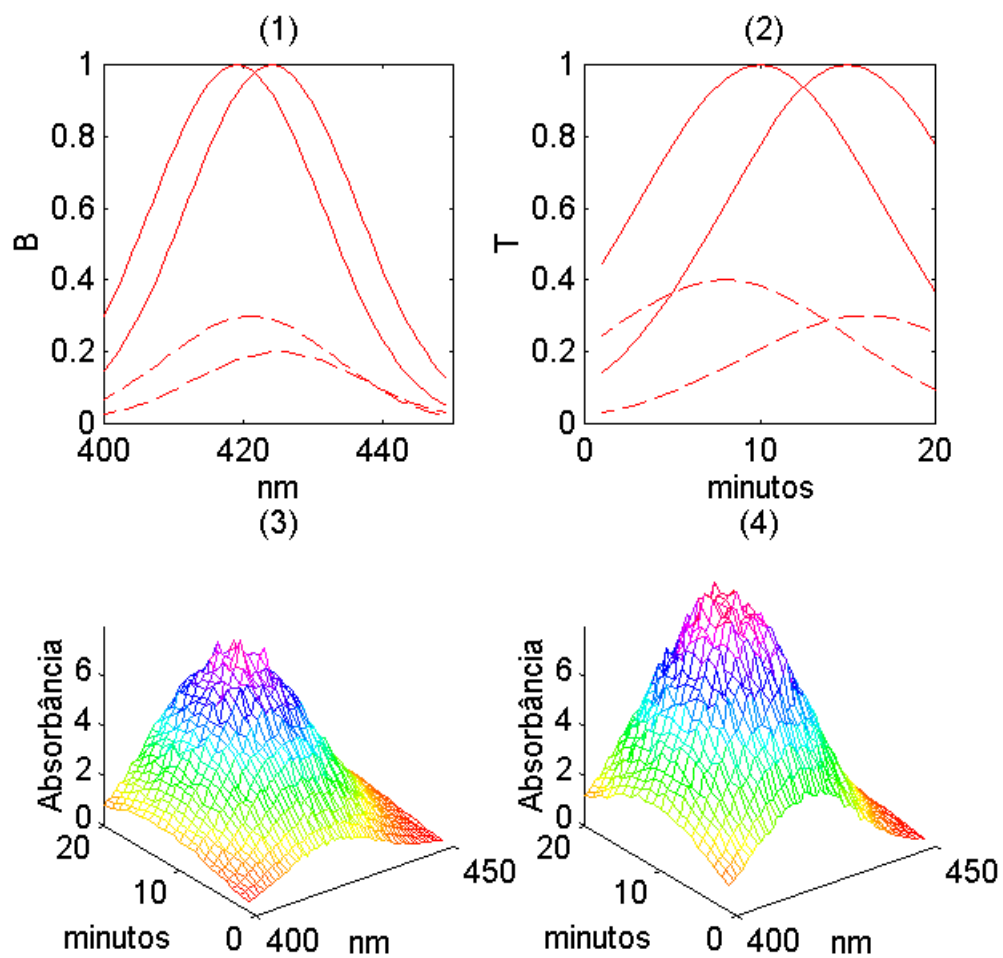


Figura 1.4 - (1) Espectros simulados(—) espectros de interferentes(- - -); (2) Cromatogramas simulados(—) Cromatogramas de interferentes(- - -); (3 e 4) Superfícies correspondentes a M e N após a adição de interferentes e ruído aleatório. B -absortividade molar ; T constante referente ao tempo de eluição.

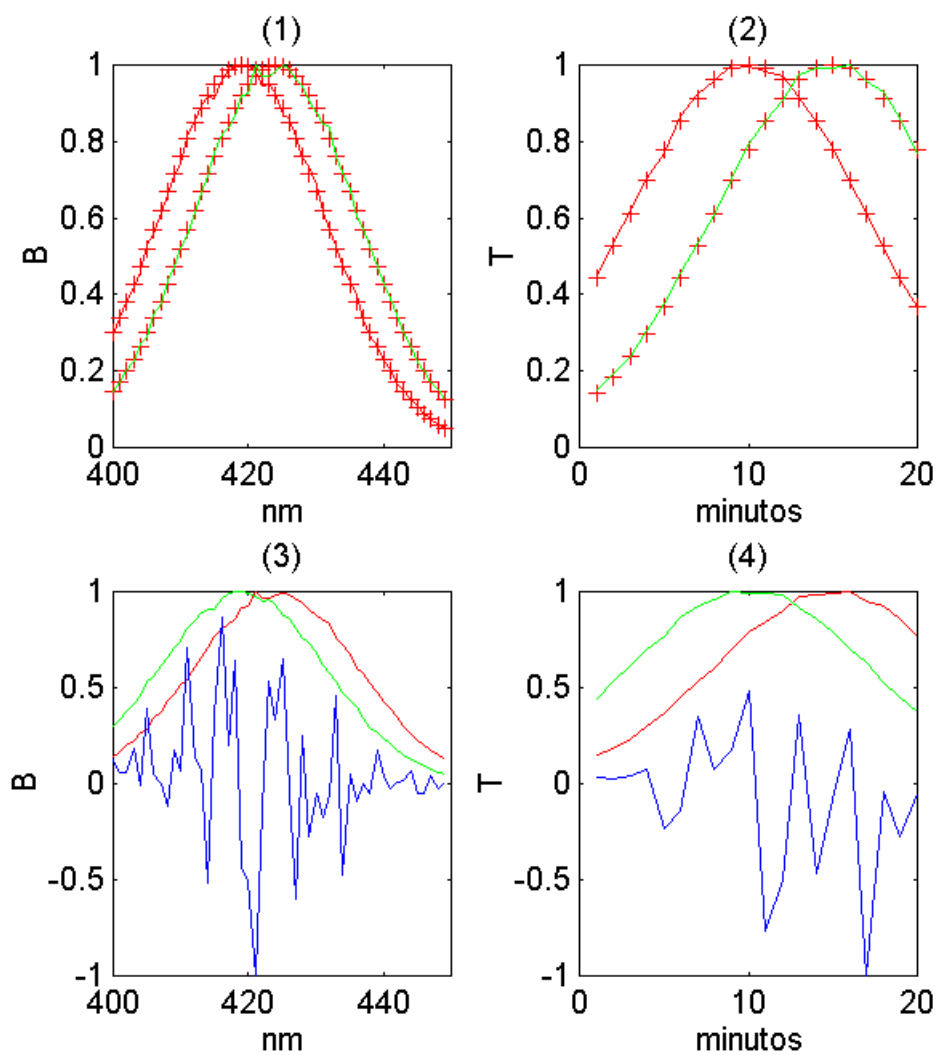


Figura 1.5 - (1) e (2) Resultados do GRAM, (+) curva simulada, (—) curva encontrada pelo GRAM para posto da matriz igual a 2. (3) e (4) espectros e cromatogramas para GRAM com posto igual a 3 vetores. B-absortividade molar; T constante referente ao tempo de eluição.

1.2.2 O Método de Lawton e Sylvestre

O método desenvolvido por Lawton e Sylvestre se aplica a amostras constituídas por misturas de duas substâncias químicas. Para sua utilização, apenas uma amostra é necessária, sendo que em seus espectros, cromatogramas, etc., devem existir pelo menos duas variáveis puras. Estas variáveis são aquelas que possuem informações de apenas uma substância, daquelas constituintes da mistura. Um exemplo deste tipo de variável, é um comprimento de onda, no caso de um espectro de absorção, onde apenas uma substância absorva.

A equação (1.3), $S = X C Y^t$, usada para representar a superfície de dados para a amostra S será transformada na equação (1.19) para a aplicação do método de Lawton e Sylvestre.

$$S = Z_{\text{coef}} Y^t \quad (1.19)$$

Onde Z_{coef} é a matriz resultante da multiplicação entre as matrizes X e C na equação (1.3). Esta matriz Z_{coef} representa os coeficientes das combinações lineares entre os espectros representados em Y . Assim, a matriz S passa a ser descrita em termos da combinação linear dos espectros em Y .

A primeira etapa do método é determinar dois vetores ortogonais entre si para descrever a matriz S . Para isto, pode-se usar a decomposição em valores singulares da matriz S , mostrada na equação (1.20).

$$S = U S_{\text{sing}} V^t \quad (1.20)$$

Onde U e V são matrizes ortonormais e S_{sing} é a matriz diagonal contendo os valores singulares.

Ao efetuar a multiplicação entre a \mathbf{U} e \mathbf{S}_{sing} tem-se a equação (1.21).

$$\mathbf{S} = \mathbf{Z} \mathbf{V}^t \quad (1.21)$$

onde \mathbf{Z} é a matriz contendo os coeficientes das combinações lineares dos vetores ortogonais em \mathbf{V} , para a representação de \mathbf{S} .

O fato de não existirem espectros negativos torna possível a formulação das inequações (1.22) e (1.23).

$$z_{A1} \mathbf{v}_1 + z_{A2} \mathbf{v}_2 = 0 \quad (1.22)$$

$$z_{B1} \mathbf{v}_1 + z_{B2} \mathbf{v}_2 = 0 \quad (1.23)$$

Os valores de z_{A1} , z_{A2} , z_{B1} , z_{B2} , correspondem às coordenadas dos espectros das substâncias puras nos novos vetores ortogonais, pois estes podem descrever qualquer vetor em seu plano. Sendo assim, após a determinação dos valores de z_{A1} , z_{A2} , z_{B1} , z_{B2} basta fazer sua multiplicação, na forma matricial, pela matriz \mathbf{V}^t e encontrar os espectros das substâncias puras.

Os valores de \mathbf{v}_1 e \mathbf{v}_2 , primeira e segunda coluna de \mathbf{V} respectivamente, são conhecidos. Mas é preciso encontrar aqueles que correspondem às variáveis puras, v_{1k} , v_{2k} , v_{1j} e v_{2j} . Para isto, é usado o fato de não existirem espectros negativos e a existência de valores negativos para \mathbf{v}_2 . Estes valores negativos em \mathbf{v}_2 aparecem devido à ortogonalidade entre os vetores de \mathbf{V} , isto é, \mathbf{v}_1 e \mathbf{v}_2 . Assim a razão entre \mathbf{v}_1 e \mathbf{v}_2 permite encontrar os valores para v_{1k} , v_{2k} , v_{1j} e v_{2j} que satisfaçam as inequações (1.22) e (1.23). Em resumo, divide-se os elementos do vetor \mathbf{v}_1 por seus correspondentes em \mathbf{v}_2 , o menor valor positivo e o menor valor negativo em módulo correspondem às variáveis puras.

Para completar a solução do problema mais duas equações são necessárias, pois as inequações (1.22) e (1.23) possuem quatro incógnitas. Sendo assim, mais duas equações envolvendo as mesmas incógnitas podem ser conseguidas pela normalização dos espectros para área sob suas curvas igual a um. Os espectros das substâncias puras podem ser representados pelas equações (1.24) e (1.25).

$$\mathbf{A} = z_{A1} \mathbf{v}_1 + z_{A2} \mathbf{v}_2 \quad (1.24)$$

$$\mathbf{B} = z_{B1} \mathbf{v}_1 + z_{B2} \mathbf{v}_2 \quad (1.25)$$

Para conseguir uma aproximação do valor da área sob os espectros pode-se usar a equação (1.26) e (1.27).

$$S_{\text{sup } A} = \sum_{i=1}^n a_i \cdot \Delta \quad (1.26)$$

$$S_{\text{sup } B} = \sum_{i=1}^n b_i \cdot \Delta \quad (1.27)$$

onde $S_{\text{sup}A}$ e $S_{\text{sup}B}$ são as áreas sob os espectros das substâncias A e B , respectivamente. Δ é o valor do intervalo entre as medidas instrumentais, por exemplo o intervalo entre os comprimentos de onda de um espectro. a_i e b_i são os valores das medidas instrumentais e n é o número de comprimentos de onda por exemplo no caso de um espectro.

A substituição da equação (1.24) na equação (1.26) e da equação (1.25) em (1.27) resulta nas equações (1.28) e (1.29). Considerando que, os espectros normalizados possuem a área sob sua curva igual a um,

pode-se escrever as equações (1.30) e (1.31) , que rearranjadas resultam nas equações (1.32) e (1.33).

$$S_{\text{sup } A} = \sum_{i=1}^n (z_{A1} v_{1i} + z_{A2} v_{2i}) \cdot \Delta \quad (1.28)$$

$$S_{\text{sup } B} = \sum_{i=1}^n (z_{B1} v_{1i} + z_{B2} v_{2i}) \cdot \Delta \quad (1.29)$$

$$\sum_{i=1}^n (z_{A1} v_{1i} + z_{A2} v_{2i}) \cdot \Delta = 1 \quad (1.30)$$

$$\sum_{i=1}^n (z_{B1} v_{1i} + z_{B2} v_{2i}) \cdot \Delta = 1 \quad (1.31)$$

$$z_{A1} \cdot \Delta \cdot \sum_{i=1}^n v_{1i} + z_{A2} \cdot \Delta \cdot \sum_{i=1}^n v_{2i} = 1 \quad (1.32)$$

$$z_{B1} \cdot \Delta \cdot \sum_{i=1}^n v_{1i} + z_{B2} \cdot \Delta \cdot \sum_{i=1}^n v_{2i} = 1 \quad (1.33)$$

Agora com as inequações (1.22) e (1.23), que para variáveis puras v_{1k} , v_{2k} , v_{1j} e v_{2j} se igualam a zero, mais as equações (1.32) e (1.33) é possível determinar os valores de z_{A1} , z_{A2} , z_{B1} , z_{B2} . E assim, com as equações (1.24) e (1.25) obtém-se os espectros para as substâncias A e B

1.3-Resultados e Discussões

O conjunto de dados experimentais é constituído por amostras de cálculo dentário, uma de felino e outra de humano. Seus espectros foram coletados de amostras pesando em torno de 0,01g, dissolvidas em 5ml de ácido clorídrico a 27%, sem filtração, apenas com decantação^{C7}. O instrumento usado foi um espectrofluorímetro Perkin-Elmer LS50, com lâmpada de xenônio para a excitação e cubeta de 1cm. Este instrumento conta com um monocromador para excitação e outro para emissão^{C7}. A região espectral envolvida, no caso da emissão, inicia-se em 460nm indo até 750nm com intervalos de 0,5nm. Em relação à excitação, o comprimento de onda inicial foi fixado em 390nm variando até 450nm, sendo que o intervalo é de 2nm.

As Figuras 1.6 e 1.7 apresentam as superfícies resultantes para estas duas amostras. Os espectros de emissão apresentam espalhamento Raman na região entre 460 a 580nm e também fluorescência residual. Já a amostra humana, também apresenta deslocamento em relação à linha de base, Figura 1.8. Assim, a região a ser trabalhada foi limitada entre 584nm e 709nm, Figura 1.8, pois abaixo de 580nm, ocorre o espalhamento Raman, a fluorescência residual e acima de 709nm a intensidade de emissão é muito baixa, ou seja, da ordem do ruído experimental.

Em se tratando da correção para a linha de base, uma curva foi ajustada ao espectro menos intenso da amostra humana, para que, ao ser subtraída dos espectros desta amostra, a intensidade em 584nm tivesse valor próximo ao valor em 709nm, semelhante ao que ocorre na amostra referente aos felinos. A Figura 1.9 mostra os espectros de emissão e excitação para amostra humana antes e depois da correção da linha de base.

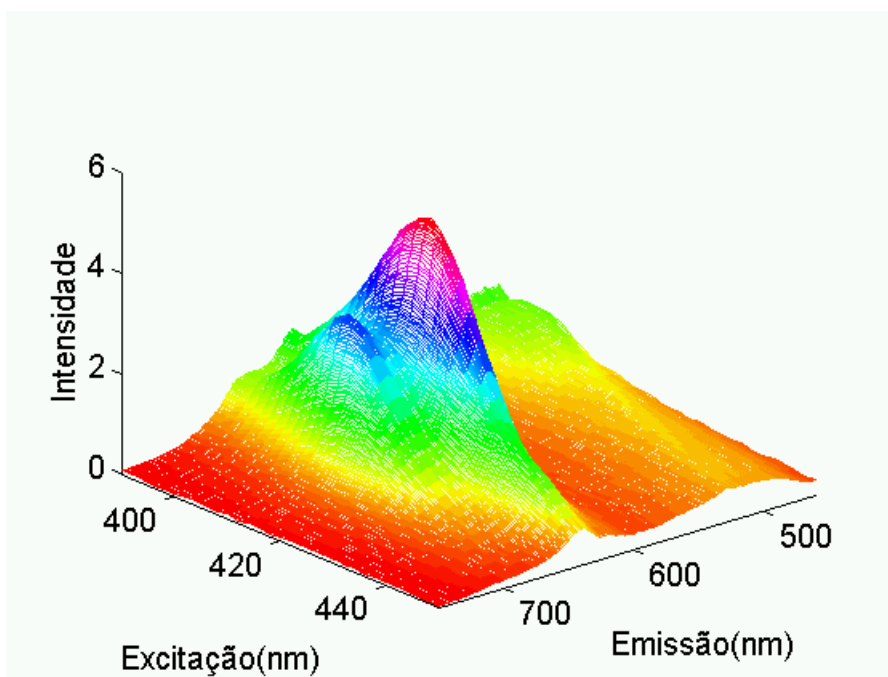


Figura 1.6 Espectros correspondentes à amostra de cálculo dentário de felinos.

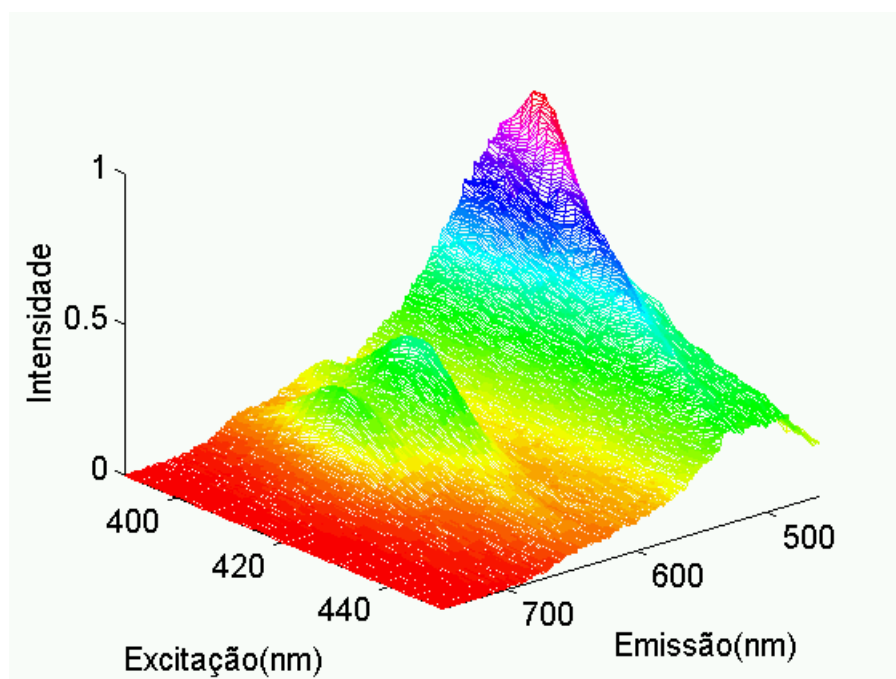


Figura 1.7. Espectros correspondentes à amostra de cálculo dentário de humanos.

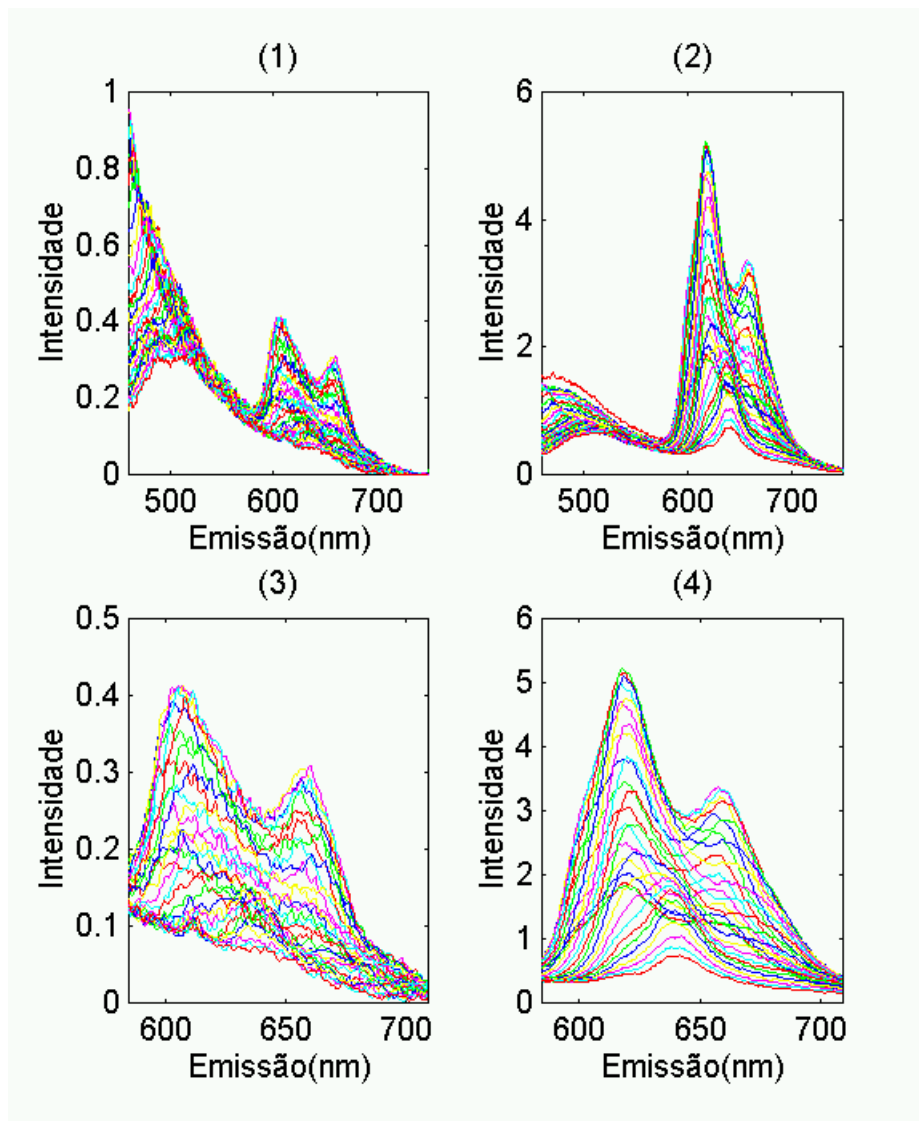


Figura 1.8 . Espectros de emissão das amostras de cálculo dentário. (1) Amostra humana, (2) amostra de felinos. (3) e (4) Espectros na região de trabalho, (3) amostra humana, (4) amostra de felinos.

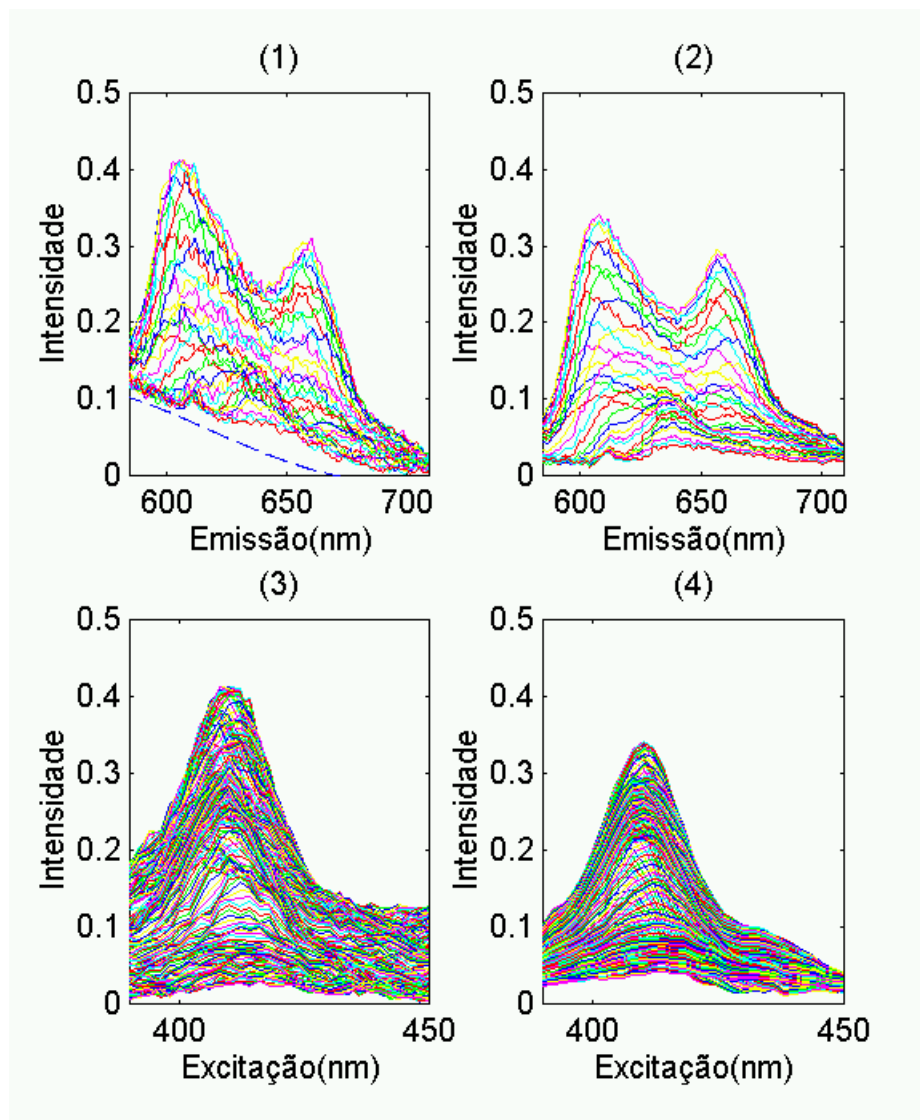


Figura 1.9. Espectros de emissão e excitação, na região de trabalho, da amostra de cálculo dentário em humanos. (1) Espectros de emissão antes da correção de linha de base; (--) linha de base em azul. (2) Espectros de emissão após a correção da linha de base. (3) e (4) Espectros de excitação antes (3) e após (4) a correção de linha de base.

Os valores singulares apresentados na Tabela 1.3, mostram uma queda na variação entre estes valores a partir do quarto valor singular. Para os dados simulados, a queda na variação entre os valores singulares ocorreu quando os vetores, correspondentes a estes valores, passaram a descrever ruídos e interferentes. Sendo assim, uma analogia entre tais valores, simulados e experimentais, sugere que o posto, para as matrizes referentes aos dados experimentais, é três. Este valor para o posto será confirmado após a aplicação do Método Generalizado de Anulação do Posto, ou seja, primeiro aplicado considerando posto igual a três e depois igual a quatro, seguindo a metodologia adotada para os dados simulados.

Tabela 1.3- Valores singulares das matrizes, correspondentes às amostras de cálculo dentário felinos e humanos, correspondentes aos seis primeiros vetores ortonormais de maior importância. (V.sing./ soma total = o valor singular dividido pela soma total de todos os valores singulares.).

Vetor nº	Humanos		Felinos	
	Valores singulares	V.sing./ soma total	Valores singulares	V.sing./ soma total
1	10,6882	0,7106	156,2215	0,7907
2	1,3005	0,0865	17,6101	0,0891
3	0,6992	0,0465	12,9706	0,0656
4	0,2567	0,0171	2,9687	0,0150
5	0,2071	0,0138	1,5169	0,0077
6	0,1426	0,0095	0,7950	0,0040

Outro fato a ser verificado é se as mesmas substâncias estão presentes nas duas amostras. Para isto, os valores singulares das matrizes originais são comparados com aqueles obtidos para as matrizes justapostas. Desta maneira, se o número de vetores linearmente independentes para descrever as matrizes em separado for igual ao das justapostas é uma indicação que as mesmas substâncias são as constituintes das duas amostras. A Tabela 1.4 apresenta os valores singulares obtidos para as amostras de cálculo dentário de humano de felino quando justapostas.

Tabela 1.4- Valores singulares da matrizes justapostas, correspondentes às amostras de cálculo dentário em felinos e humanos, para os seis primeiros vetores ortonormais de maior importância. (V.sing./ soma total = o valor singular dividido pela soma total de todos os valores singulares.).

Vetor	$(A B)$		$\begin{pmatrix} A \\ B \end{pmatrix}$	
	Valores singulares	V.sing./ soma total	Valores singulares	V.sing./ soma total
1	156,5710	0,7894	156,5722	0,7800
2	17,7799	0,0896	17,7251	0,0883
3	13,0030	0,0656	13,0330	0,0659
4	2,9893	0,0151	3,0211	0,0150
5	1,5542	0,0078	1,5862	0,0079
6	0,8306	0,0042	0,9142	0,0046

Para confirmar o posto das matrizes o GRAM foi aplicado, primeiro considerando o posto com sendo três e depois igual a quatro. Os resultados para o primeiro teste são espectros com significado químico, isto é, positivos e com picos distintos, Figuras 1.10 e 1.11. No segundo teste, posto igual a quatro, dois espectros de excitação possuem os seus máximos no mesmo comprimento de onda. Ou seja, destes dois apenas um tem sentido químico, no caso da identificação pelo comprimento de onda do máximo da banda, Figura 1.12.

Na comparação entre os espectros de excitação obtidos para posto igual a três e posto igual a quatro nota-se o aparecimento de um quarto espectro, o espectro D na Figura(1.12). Outro fato relevante, é o deslocamento significativo do máximo do espectro C, comparar Figura(1.10) com Figura(1.12). A conclusão desta comparação é que existem quatro espécies luminescentes, mas o GRAM não consegue separar completamente os espectros B e C, Figura (1.12).

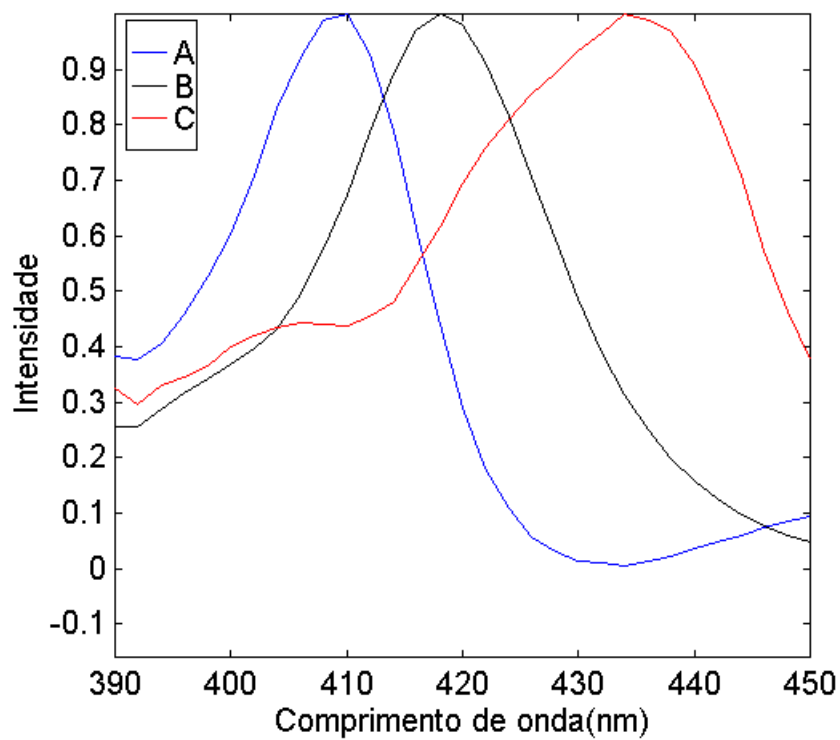


Figura 1.10. Espectros de excitação resolvidos considerando posto igual a três.

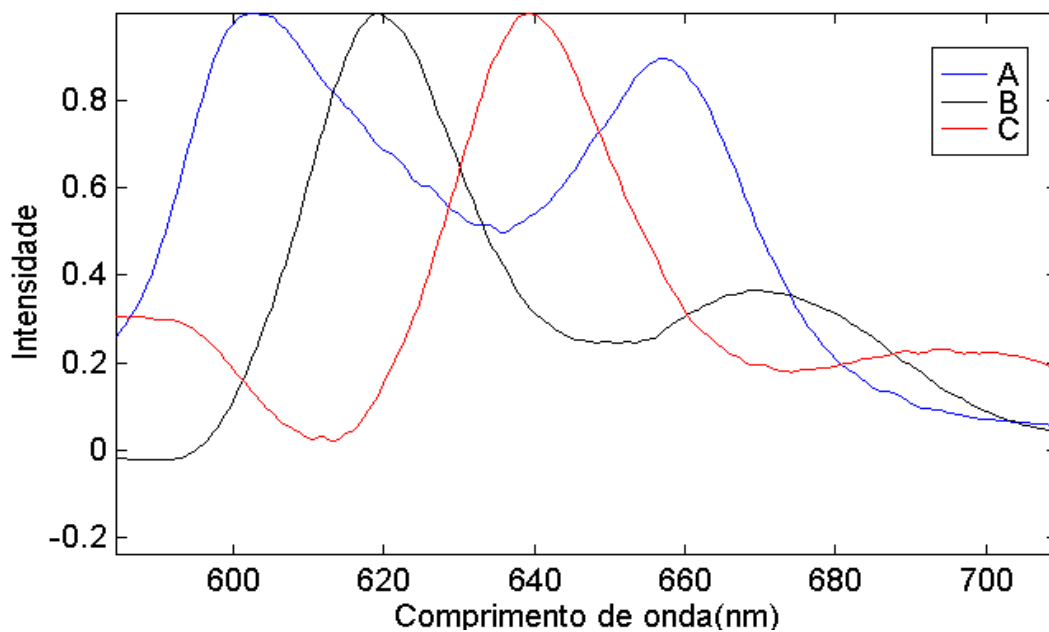


Figura 1.11. Espectros de emissão resolvidos considerando posto igual a três.

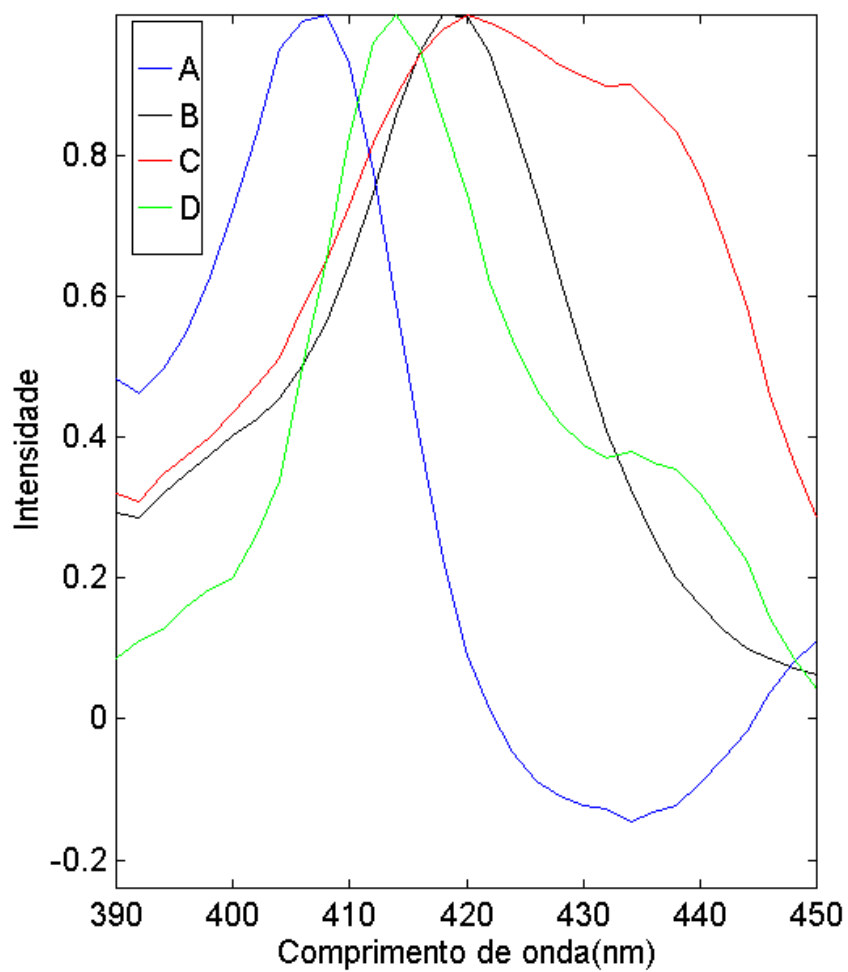


Figura 1.12. Espectros de excitação resolvidos considerando posto igual a quatro.

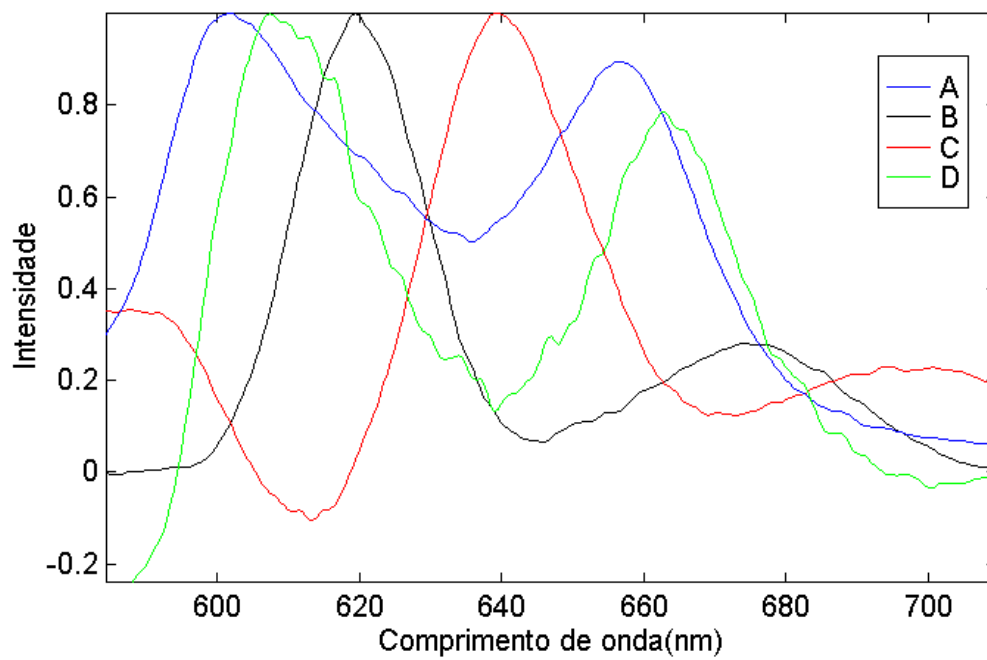


Figura 1.13. Espectros de emissão resolvidos considerando posto igual a quatro.

Para fazer a separação dos espectros B e C por completo foi usado um segundo método, aquele desenvolvido por Lawton e Sylvestre. Para isto, as informações dos espectros A e D foram subtraídas dos espectros originais, pois este método se aplica apenas a espectros de amostras constituídas por mistura de duas substâncias químicas.

Para subtrair os espectros A e D, primeiro foram calculadas suas concentrações nas amostras de cálculo dentário de humanos e felinos, através das equações (1.34) e (1.35), e depois a subtração através das equações (1.36) e (1.37).

$$\mathbf{C}_M = (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \mathbf{M} \mathbf{Y}^t (\mathbf{Y}^t \mathbf{Y})^{-1} \quad (1.34)$$

$$\mathbf{C}_N = (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \mathbf{N} \mathbf{Y}^t (\mathbf{Y}^t \mathbf{Y})^{-1} \quad (1.35)$$

$$\mathbf{M}_{CB} = \mathbf{M} - (\mathbf{X}_{AD} \mathbf{C}_{MAD} \mathbf{Y}_{AD}^t) \quad (1.36)$$

$$\mathbf{N}_{CB} = \mathbf{N} - (\mathbf{X}_{AD} \mathbf{C}_{NAD} \mathbf{Y}_{AD}^t) \quad (1.37)$$

onde \mathbf{C}_M e \mathbf{C}_N são as matrizes de concentrações para as amostras M e N , respectivamente. \mathbf{X} e \mathbf{Y} são as matrizes com os espectros de excitação e emissão, respectivamente, resolvidos pelo GRAM para posto igual a quatro. \mathbf{M}_{CB} e \mathbf{N}_{CB} as matrizes sem as informações dos espectros A e D. \mathbf{X}_{AD} , \mathbf{Y}_{AD} , \mathbf{C}_{NAD} e \mathbf{C}_{MAD} as matrizes correspondentes aos espectros A e D e suas respectivas concentrações.

O resultado para a aplicação do método de Lawton e Sylvestre para a matriz \mathbf{N}_{CB} , é mostrado na Figura(1.14). A matriz \mathbf{N}_{CB} foi escolhida por possuir maiores intensidades em relação a \mathbf{M}_{CB} aumentado assim, a razão sinal/ruído.

Sendo assim, o número de espectros resolvidos é quatro e o posto é confirmado como sendo quatro, isto é, existem quatro espécies luminescentes em cada amostra.

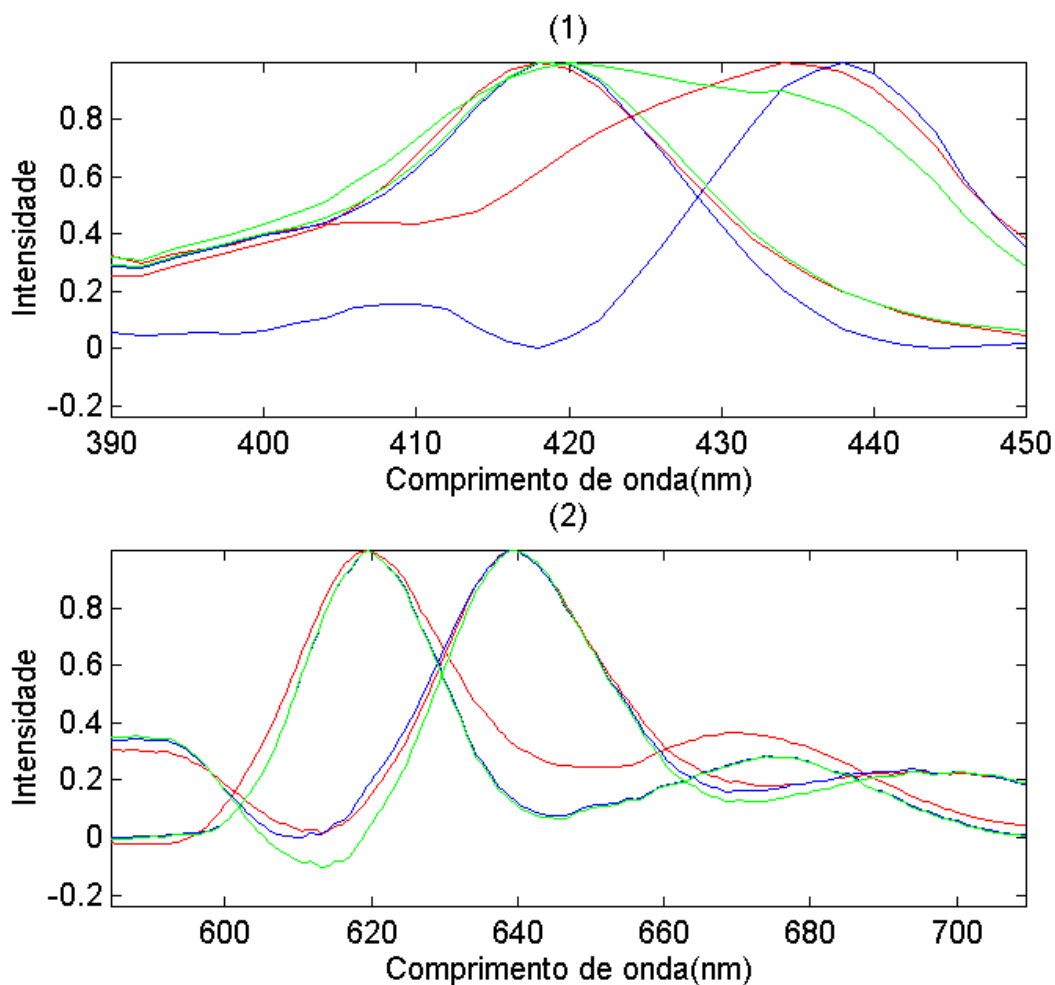


Figura 1.14 Espectros de excitação e emissão, para espécies luminescentes B e C, resolvidos pelo GRAM posto igual a 3, em vermelho, e 4 em verde e pelo método de Lawton e Sylvestre, em azul. (1) excitação e (2) emissão.

Os espectros de excitação destas quatro espécies, mostrados na Figura 1.15, são os responsáveis pela sua caracterização. Eles aparecem na região da banda Soret, que é característica de espécies porfirínicas^{E3}.

Os comprimentos de onda em que aparecem os máximos, nos espectros de excitação resolvidos pelo método, são 408nm(A), 414nm(D), 418nm(B) e 438nm(C).

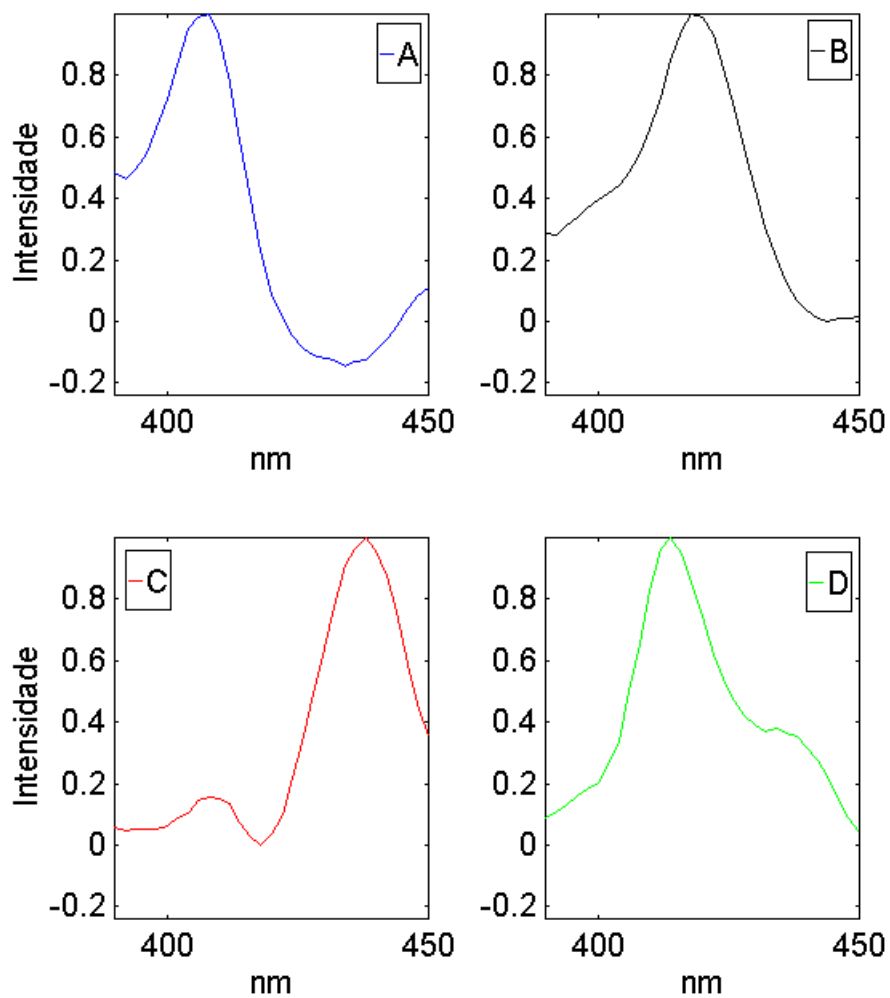


Figura 1.15 Espectros de excitação para as quatro espécies luminescentes. Máximos A- 408nm ; B - 418nm ; C - 438nm ; D - 414nm .

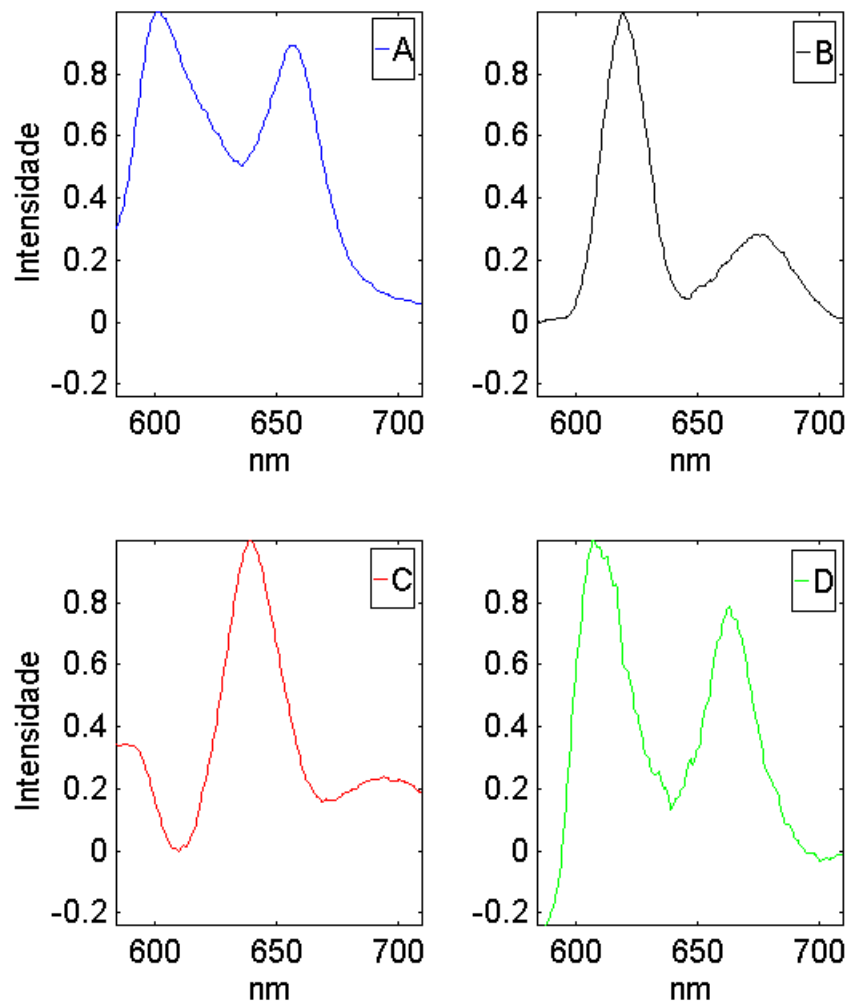


Figura 1.16 Espectros de emissão para as quatro espécies luminescentes. Máximos A- 601 e 656nm ; B - 619 e 674nm ; C - 639 e 694nm ; D - 607 e 663nm .

1.4-Conclusões

A apresentação do Método Generalizado de Anulação do Posto em conjunto com os dados simulados foi importante, principalmente, para a interpretação da decomposição em valores singulares. Isto por mostrar que, na ausência de ruídos e interferentes, o número de valores singulares diferentes de zero é igual ao número de substâncias presentes nas amostras, cujas matrizes foram decompostas. Outro fato que reafirma tal importância, é a comparação dos valores singulares resultantes da decomposição feita após a adição de interferentes e ruídos aos dados simulados com aqueles encontrados para os dados experimentais. Esta comparação mostra que a variação entre os valores singulares, usados para descrever os ruídos e interferentes, não é significativa frente aquela variação encontrada para os valores singulares importantes na descrição da matriz original, fato este, que serve de parâmetro na escolha do posto das matrizes.

Em se tratando desta escolha, também é importante o conhecimento químico, o que é verificado nos dados experimentais, onde a confirmação do posto está diretamente ligada ao fato de serem os espectros de excitação os responsáveis pela identificação das espécies luminescentes presentes nas amostras.

A grande vantagem que o GRAM apresenta, mesmo tendo sido complementado pelo método de Lawton e Sylvestre, é uma solução matemática, de rápida aplicação e com um baixo custo computacional, sendo que, a determinação destas substâncias por um tratamento químico, no que diz respeito a sua separação, poderia ser mais dispendiosa em tempo e custos operacionais, ou até mesmo inviável.

Capítulo 2

2.1-Introdução ao Capítulo 2

Neste capítulo é apresentada uma aplicação quimiométrica em reconhecimento de padrões. O objetivo desta aplicação é mostrar uma metodologia para o tratamento de um conjunto de informações espectrais. Este tipo de dados, que é comum nos laboratórios de química, forma uma população de informações estatisticamente significativa. O seu uso, em elucidação de estruturas moleculares, é de grande importância e normalmente feito através da comparação entre as informações já conhecidas com aquelas obtidas para novas estruturas. Aqui, a quimiometria tem como função fornecer métodos estatísticos que possibilitem maior agilidade nesta comparação.

Para apresentar uma metodologia com tais características, foi escolhido um conjunto de dados, com informações a respeito de flavonóides, e efetuado seu tratamento através do SIMCA “Soft Independent Modelling of Class Analogy”, da Análise de Componentes Principais e Análise por Agrupamento Hierárquico.

2.2-Métodos

2.2.1-Notação

Neste capítulo as linhas de uma matriz são chamadas de amostras, por se tratarem de substâncias diferentes. Já as colunas são ditas variáveis, por serem comuns a todas as substâncias.

2.2.2-Análise de Componentes Principais - ACP

A análise de componentes principais é um método matemático usado para reduzir o número de variáveis da matriz original dos dados^{D1,D2,D4}. A vantagem desta redução é a possibilidade de visualizar a distribuição espacial das amostras, influenciada pelas variáveis originais, em gráficos bi e tridimensionais. Para efetuar tal redução, sem perda de informações, a matriz original passa a ser representada por novas variáveis, agora ortogonais entre si e direcionadas em função da distribuição espacial dos dados.

As novas variáveis são chamadas de componentes principais. A primeira componente principal é aquela situada na direção de maior variação dos dados, a segunda é ortogonal à primeira e em uma direção que lhe permita descrever a maior variação nos dados e assim por diante. Para melhor ilustrar a idéia de variação dos dados no espaço a Figura 2.1 mostra a representação gráfica de uma matriz de dimensão (15 × 2). Nesta Figura 2.1, a variável 1 representa a primeira coluna da matriz e a variável 2 a segunda coluna. Em termos da distribuição espacial dos dados, a reta cp1 esta colocada na direção de maior variação e a reta cp2 na segunda maior variação.

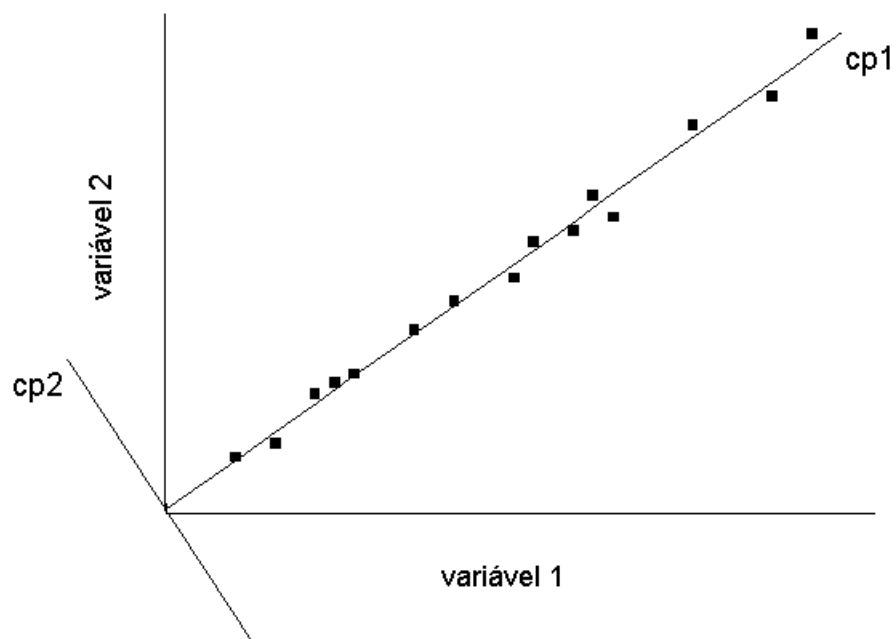


Figura 2.1. Representação de uma matriz de dados (15×2) onde as variáveis 1 e 2 correspondem, respectivamente, às colunas 1 e 2 desta matriz. As linhas cp1 e cp2 são ortogonais e estão direcionadas em função da variação dos dados.

A representação através das componentes principais é conseguida com uma mudança de bases^{A1}, ou seja, da matriz representada pelas variáveis originais para a matriz representada pelas componentes principais. Tal mudança é feita através de uma transformação linear^{A1,B1}, o que em termos matriciais é dado pela equação (2.1).

$$\mathbf{X} \mathbf{P} = \mathbf{T} \quad (2.1)$$

onde \mathbf{X} é matriz de dados, \mathbf{T} a matriz contendo as coordenadas na nova base e \mathbf{P} é a matriz de transformação linear entre as bases. Destas três matrizes, a única conhecida é a matriz de dados. Assim, para encontrar a nova base, dentre outras técnicas de ortogonalização, a indicada para o propósito do método é Decomposição em Valores Singulares(DVS)^{A2,A3}.

Nesta técnica a matriz original é decomposta em duas matrizes ortonormais e uma diagonal. As ortonormais possuem todos seus vetores ortogonais entre si e com módulo igual a um. A decomposição de \mathbf{X} é mostrada na equação (2.2).

$$\mathbf{X} = \mathbf{U} \mathbf{S} \mathbf{V}^t \quad (2.2)$$

onde :

\mathbf{U} , \mathbf{V} - São as matrizes ortonormais.

\mathbf{S} - Matriz diagonal contendo os valores singulares da matriz \mathbf{X} .

O primeiro vetor de \mathbf{V} está na direção de maior variação dos dados e seu valor singular correspondente, o primeiro elemento de \mathbf{S} , representa a variação ao longo deste vetor^{B1}. Isto também é válido para os outros vetores, ou seja, o segundo na direção de maior variação dos dados, sendo que o seu valor singular correspondente é o segundo maior e assim por diante.

A matriz \mathbf{V} possui seus vetores direcionados em função da distribuição espacial dos dados, mas com módulo unitário. Por outro lado, os valores singulares representam a variação em cada vetor de \mathbf{V} . Assim, a matriz obtida pela multiplicação entre \mathbf{U} e \mathbf{S} representa as coordenadas nestes vetores ortogonais, direcionados em função da variação dos dados no espaço e sem perda de informação em relação à matriz original. Para isolar $\mathbf{U} \mathbf{S}$ do lado direito da equação (2.2) e obter as matrizes \mathbf{T} e \mathbf{P} são feitas as operações mostradas nas equações de (2.3) a (2.7).

$$\mathbf{XV} = \mathbf{USV}^t \mathbf{V} \quad (2.3)$$

$$\mathbf{XV} = \mathbf{US} \quad (2.4)$$

$$\mathbf{T} = \mathbf{US} \quad (2.5)$$

$$\mathbf{P} = \mathbf{V} \quad (2.6)$$

$$\mathbf{XP} = \mathbf{T} \text{ ou } \mathbf{X} = \mathbf{TP}^t \quad (2.7)$$

Quanto aos valores singulares, seus quadrados correspondem aos autovalores da matriz $\mathbf{X}^t \mathbf{X}$ ^{A2,A3}. Isto pode ser mostrado a partir da equação (2.8).

$$\mathbf{X}^t \mathbf{X} = \mathbf{V S}^t \mathbf{U}^t \mathbf{U S V}^t \quad (2.8)$$

Como \mathbf{U} é ortonormal, $\mathbf{U}^t \mathbf{U}$ é igual a matriz identidade pois a matriz transposta de uma ortonormal é igual a sua inversa. Tem-se assim, a equação (2.9).

$$\mathbf{X}^t \mathbf{X} = \mathbf{V S}^t \mathbf{S V}^t \quad (2.9)$$

Sendo \mathbf{S} uma matriz diagonal, sua transposta é igual a ela mesma, deste modo; os valores de $\mathbf{S}^t \mathbf{S}$ correspondem aos quadrados de \mathbf{S} . A multiplicação dos dois lados da equação (2.9) por \mathbf{V} resulta na equação (2.10).

$$\mathbf{X}^t \mathbf{X V} = \mathbf{V S}^t \mathbf{S V}^t \mathbf{V} \quad (2.10)$$

De forma semelhante ao dito para $\mathbf{U}^t \mathbf{U}$, $\mathbf{V}^t \mathbf{V}$ resulta também em uma matriz identidade. O resultado destas operações é a equação (2.11).

$$(\mathbf{X}^t \mathbf{X}) \mathbf{V} = \mathbf{V S}^2 \quad (2.11)$$

Estas operações mostram que V corresponde ao conjunto de autovetores da matriz $(X^t X)$ e S^2 , matriz na qual os valores estão elevados ao quadrado, seus autovalores. Ainda sobre a matriz $(X^t X)$, ela é proporcional à matriz de variância covariância, onde o número de graus de liberdade é o fator de proporcionalidade, sendo desta relação que provém a ligação entre direção das componentes principais e variação dos dados. De forma semelhante, é possível mostrar que a matriz U corresponde aos autovetores da matriz XX^t e S^2 seus autovalores.

Uma das principais vantagens da análise de componentes principais é a eliminação de grande parte dos ruídos experimentais. Isto é possível, pois tais ruídos constituem uma população de erros aleatórios, ou seja, não estão correlacionados com as informações contidas na matriz de dados. Devido a esta falta de correlação, um número maior de componentes principais é necessário para representar a matriz de dados. Por exemplo, Na Figura 2.1, a variação na direção da reta cp2 é muito menor em relação à de cp1. Além disto, a variação de cp2 parece ser aleatória, o que não ocorre em cp1. Considerando, assim, a variação em cp2 como ruído experimental, esta matriz de dados pode ser representada por apenas uma componente principal, ou seja, na direção da reta cp1.

Incluído o erro presente na matriz de dados, sua decomposição passa a ser representada pela equação (2.12).

$$X = \bar{T} \bar{P}^t + E \quad (2.12)$$

onde o traço acima das matrizes representa que elas estão "truncadas", isto é, apenas com as componentes principais importantes para descrever os dados, tendo sido eliminadas aquelas usadas para descrever erros.

2.2.3-SIMCA

O SIMCA é um método para classificação^{D1,D2,D3} que considera informações da distribuição da população, estima um grau de confiança da classificação e pode prever novas amostras como pertencentes a uma ou mais classes ou nenhuma classe.

Para fazer a classificação o SIMCA utiliza o espaço das componentes principais de cada classe. Desta forma, a classe n passa a ser representada pela equação (2.13).

$$\mathbf{X}_n = \bar{\mathbf{T}}_n \bar{\mathbf{P}}_n^t + \mathbf{E}_n \quad (2.13)$$

onde \mathbf{X}_n são os dados da classe, \mathbf{T}_n a matriz contendo as coordenadas nas componentes principais da classe n , \mathbf{P}_n a matriz de transformação linear e \mathbf{E}_n a matriz de resíduos .

Na construção do modelo de classificação o SIMCA calcula, para cada classe em separado, o desvio padrão dos resíduos. Para o espaço descrito pelas componentes principais, são calculadas as variâncias das amostras, em cada eixo. Estes dois parâmetros são usados na classificação de novas amostras.

O objetivo do SIMCA é criar um espaço limitado para cada classe. Isto pode ser melhor compreendido para uma classe descrita por duas componentes principais. Em termos geométricos, os resíduos desta classe correspondem às distâncias das amostras ao plano dos componentes principais. Desta forma, o cálculo do desvio padrão dos resíduos dá origem a dois planos paralelos ao destes componentes, isto é, um acima e outro abaixo. Considerando a variância em cada componente principal e os planos, referentes ao desvio padrão dos resíduos, pode-se dizer que a classe está limitada por uma caixa, uma hipercaixa no caso de três ou mais componentes e um cilindro para uma componente principal.

A classificação de uma nova amostra é feita através de sua projeção nas componentes principais de cada classe, onde são

calculadas as variâncias e seu resíduo. Estes dois são comparados por teste F com aqueles já determinados na construção do modelo. Assim, naquelas classes onde o resíduo é menor ou igual, o mesmo é válido para as variâncias, a amostra é classificada positivamente. Com isto, a amostra pode ser colocada em uma ou mais classes. Em caso contrário, desvio ou variâncias maiores, a amostra é classificada como não pertencente à classe.

A Figura 2.2 apresenta um exemplo de uma matriz, com três colunas, dividida em duas classes. A primeira é representada por duas componentes principais, onde v_1 é a variância na primeira componente principal Cp_1 , v_2 a variância na segunda componente Cp_2 e o desvio padrão dos resíduos é representado por e .

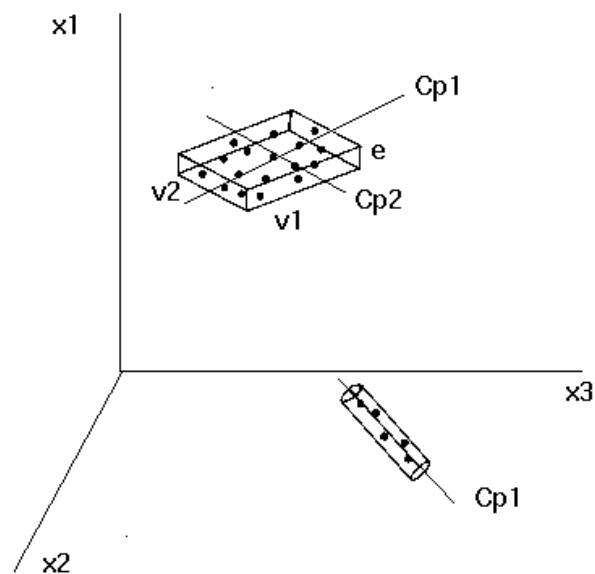


Figura 2.2. Representação gráfica de uma matriz com três colunas e dividida em duas classes. x_1, x_2 e x_3 representam as colunas da matriz; Cp_1 o primeiro componente principal, Cp_2 o segundo componente principal, v_1 a variância em Cp_1 , v_2 a variância em Cp_2 , o desvio padrão dos resíduos é representado por e .

2.2.4-Análise por agrupamento hierárquico

O objetivo deste método é agrupar as amostras segundo suas semelhanças. O parâmetro usado para tal, é a distância entre elas, pois em termos geométricos, pode-se definir cada amostra como sendo um ponto em um espaço de m variáveis, e a partir disto, calcular a distância entre todas as amostras. Estas distâncias possibilitam verificar como estão distribuídas espacialmente as amostras, isto é, se estão próximas umas das outras, formando ou não grupos.

A equação (2.14) apresenta a fórmula para o cálculo da distâncias entre duas amostras a e b em um espaço de m variáveis.

$$d_{ab} = \sqrt{\sum_{j=1}^m (x_{aj} - x_{bj})^2} \quad (2.14)$$

onde d_{ab} é a distância entre as amostras a e b .

Para facilitar a interpretação deste método a Figura 2.3 apresenta a seqüência de um agrupamento pela menor distância. Neste exemplo, o primeiro grupo é formado pelas duas amostras com menor distância entre si, (amostras 4 e 6). Em seguida são agrupadas aquelas amostras com a segunda menor distância entre si,(amostras 5 com 4) e assim por diante até todos os grupos estarem formados. Conhecidos os grupos, são construídos gráficos chamados dendrogramas, que traduzem a distribuição espacial dos dados. Na Figura 2.4 é apresentado o gráfico referente ao agrupamento da Figura 2.3. Estes gráficos podem estar na mesma escala das distâncias ou escala de similaridade, que dada é pela equação (2.15).

$$s = 1 - \left(\frac{D_{ab}}{D_{max}} \right) \quad (2.15)$$

onde: s é a similaridade, D_{ab} é a distância entre duas amostras e D_{max} é a maior distância entre todas as amostras.

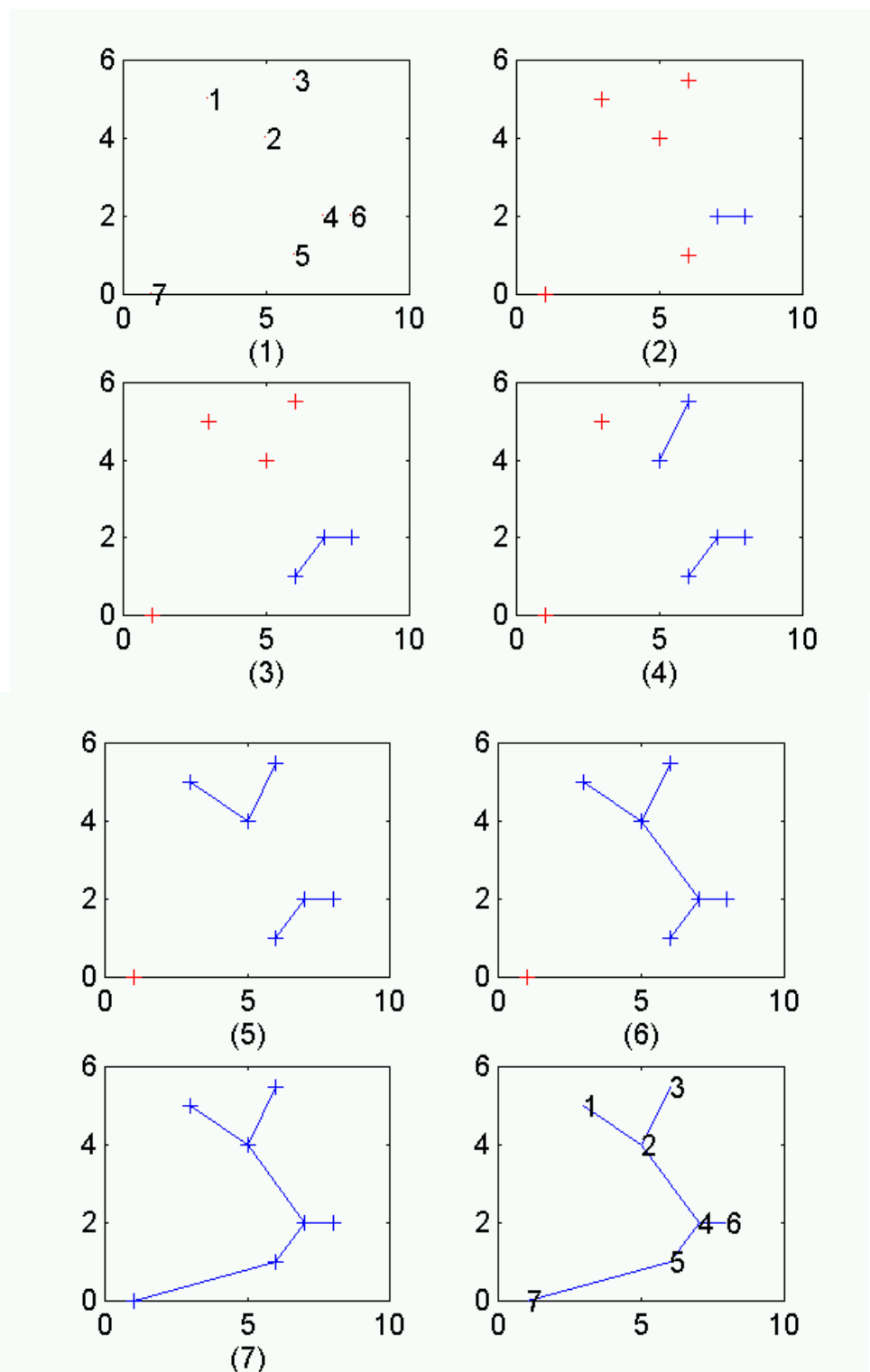


Figura 2.3. Seqüência de um agrupamento hierárquico de uma matriz com duas colunas e sete linhas .

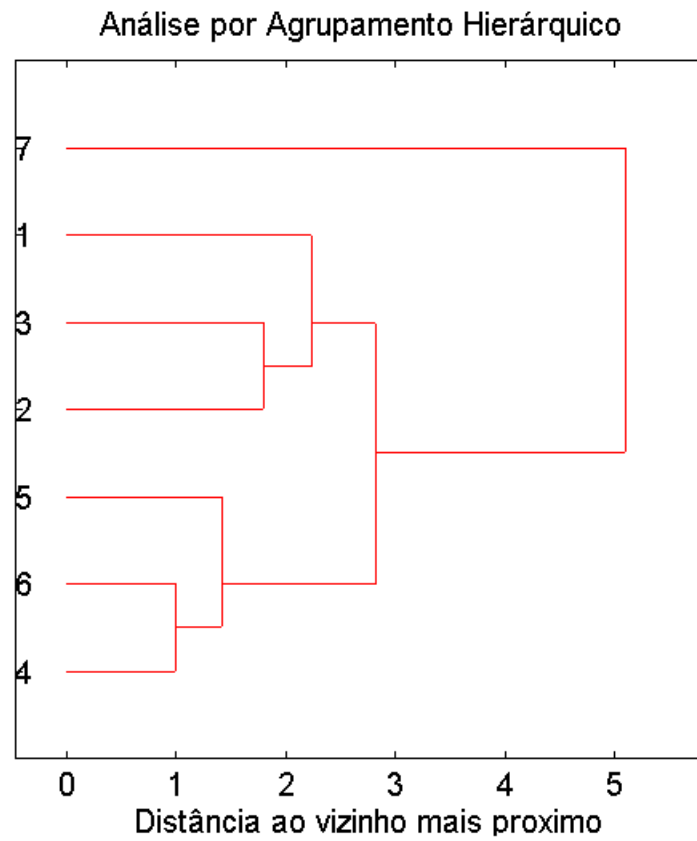


Figura 2.4. Gráfico de dendrogramas para o agrupamento apresentado na Figura 2.3.

2.3-Flavonóides

Este conjunto de dados composto por informações espectrais de noventa e cinco flavonóides, que correspondem a picos, ombros e inflexões de espectros de absorção na região do ultra violeta^{E4}. Nesta região, aparecem duas bandas características destes compostos, a primeira entre 240 e 285nm (Banda II) e a segunda entre 300 e 400nm (Banda I)^{E1}. Na Figura 2.5, estas bandas podem ser visualizadas nos espectros correspondentes à flavona, 4'-hidroxiflavona e 5-hidroxiflavona. A comparação entre as estruturas moleculares destes compostos, Figura 2.6, e o confronto de seus espectros mostra que a substituição no anel A, Figura 2.7, está relacionada com o deslocamento batocrômico da Banda II e a substituição no anel B com o deslocamento da Banda I.

Outra característica destes compostos, é apresentar ou não deslocamentos hipsocrômicos e batocrômicos, em seus espectros de absorção, em função da existência de hidroxilas vicinais no carbonos 3' e 4'. Para identificar estas hidroxilas vicinais, são coletados espectros de soluções do flavonóide com cloreto de alumínio e flavonóide com cloreto de alumínio mais ácido clorídrico para promover as complexações mostradas na Figura 2.8. As diferenças entre os espectros, em tais soluções, podem ser verificadas, por exemplo, nas Figuras 2.9 e 2.10 onde a distinção estrutural entre os dois flavonóides é a presença ou não das citadas hidroxilas vicinais.

Devido a estas características, os espectros de UV dos flavonóides tem sido usados na elucidação de suas estruturas.

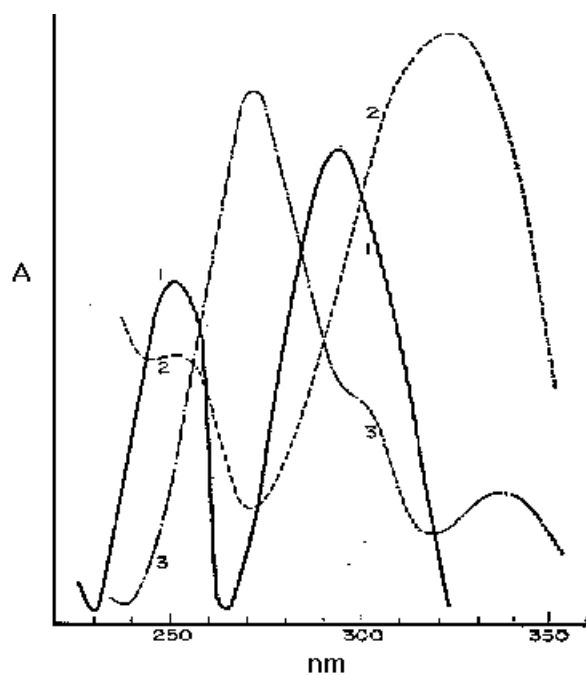


Figura 2.5- Espectros na região do ultra violeta para a flavona (1), 4'-hidroxiflavona(2) e 5-hidroxiflavona(3). Estes espectros foram coletados em metanol. A- densidade óptica.

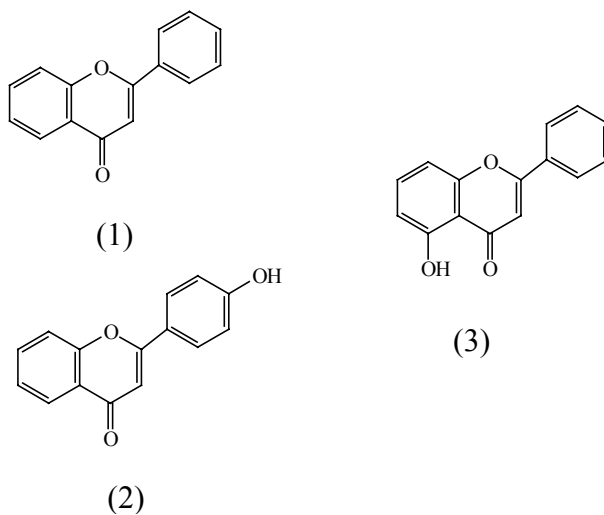


Figura 2.6- Estrutura molecular da flavona (1), 4'-hidroxiflavona(2) e 5-hidroxiflavona(3).

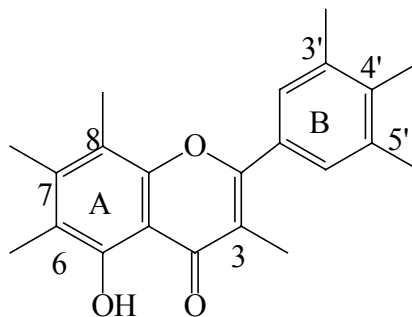


Figura.2.7.Estrutura básica dos flavonóides.

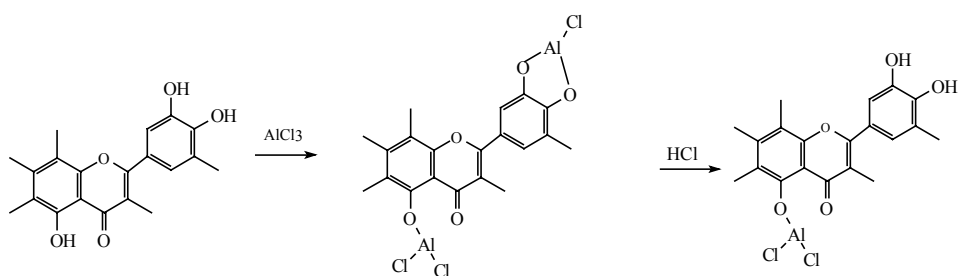


Figura 2.8- Reação de complexação de um flavonóide, com hidroxilas vicinais em 3' e 4', quando em solução de cloreto de alumínio e após a adição de ácido clorídrico.

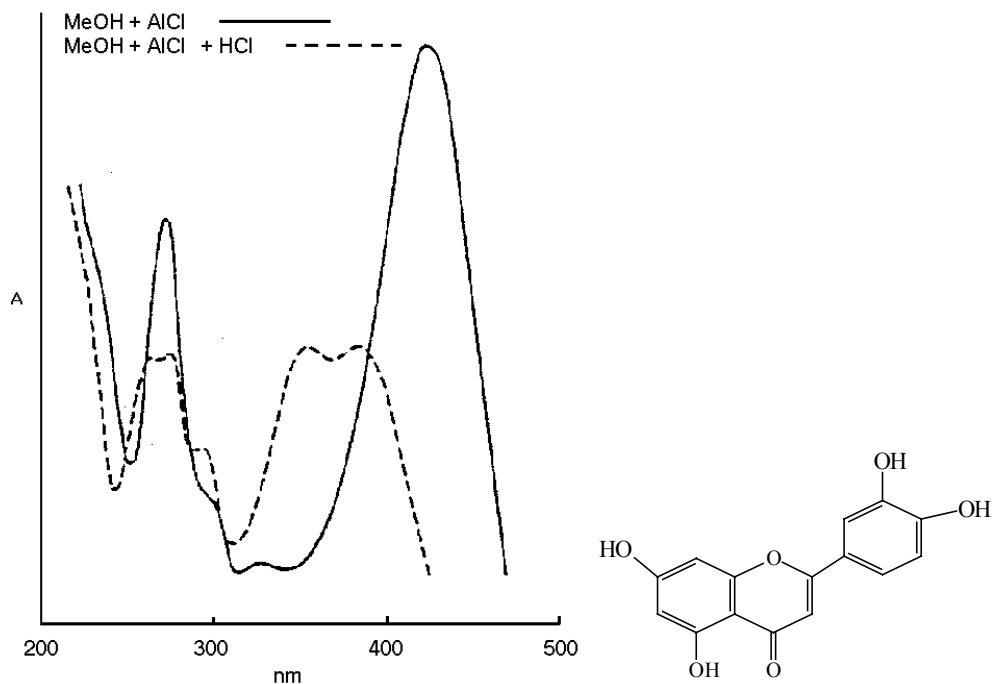


Figura 2.9- Espectros na região do UV coletados em solução de metanol mais cloreto de alumínio e solução de metanol mais cloreto de alumínio mais ácido clorídrico para o flavonóide cuja estrutura molecular está mostrada ao lado. A- densidade óptica.

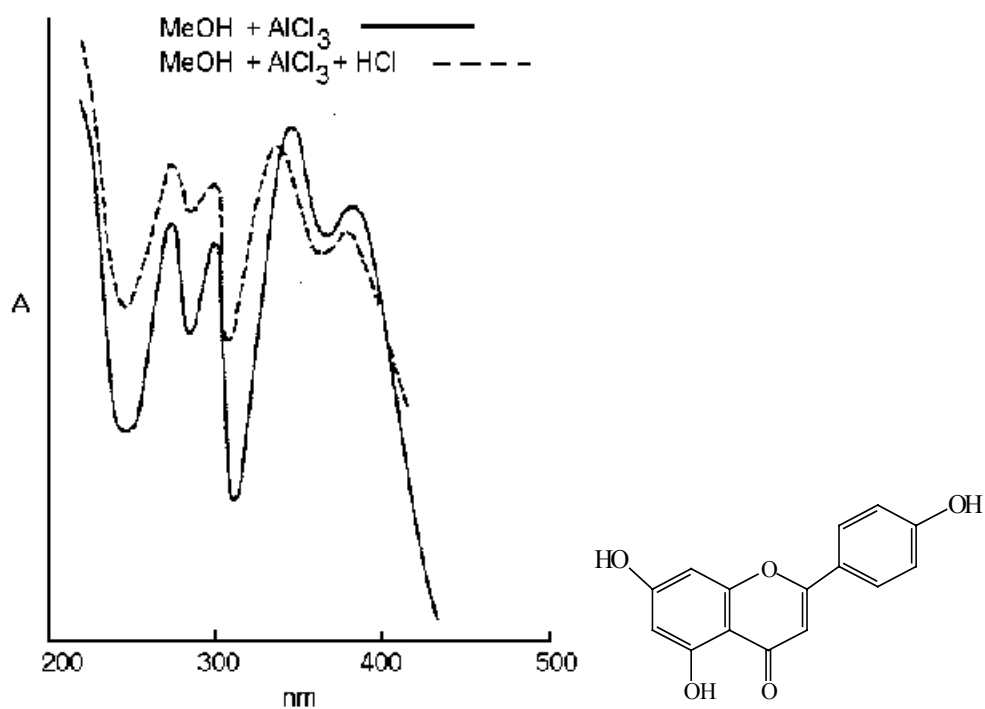


Figura 2.10- Espectros na região do UV coletados em solução de metanol mais cloreto de alumínio e solução de metanol mais cloreto de alumínio mais ácido clorídrico para o flavonóide cuja estrutura molecular está mostrada ao lado. A- densidade óptica.

2.4-Os dados

O conjunto de dados deste trabalho conta com 95 flavonóides distribuídos entre 8 classes. Os possíveis substituintes na estrutura molecular destes 95 compostos são:

-Para os carbonos 3, 6, 8, 3' , 5' os radicais hidroxila(OH), metoxila (OMe) e hidrogênio(H);

-Para os carbonos 7 e 4' o radical hidroxila ou radical metoxila .

Para construção de modelos, que permitam a previsão das estruturas moleculares destes flavonóides, duas etapas foram elaboradas. Na primeira foi construído um modelo de classificação para determinar as substituições nos carbonos 3, 6 e 8. Na segunda, o número de substituições no anel B foi determinado por análise de componentes principais e análise por agrupamento hierárquico. Este número de substituições no anel B varia de 1 a 3, pois o hidrogênio não é considerado uma substituição.

Desta forma, para novas amostras serem elucidadas, primeiro serão classificadas com SIMCA. Uma vez conhecida sua classe, a amostra é projetada no espaço dos componentes principais que descreve sua classe e a semelhança com aquelas amostras, cujas estruturas já são conhecidas, permite determinar o número de substituições no anel B, o que é confirmado pela análise por agrupamento hierárquico.

2.5-As classes

As 8 classes, para a construção do modelo SIMCA, foram escolhidas segundo as substituições nos carbonos 3,6 e 8. As substituições nos carbonos 7, 3', 4' e 5' são responsáveis pela variação dentro de cada classe. Os flavonóides usados são apresentados na Tabela 1 no Anexo 1.

Classes :

Classe 1: 3-H , 6-OMe , 8-H ;

Classe 2: 3-H , 6-OH , 8-H ;

Classe 3: 3-H , 6-H , 8-H ;

Classe 4: 3-OMe , 6-OMe , 8-H ;

Classe 5: 3-H , 6-OMe , 8-OMe ;

Classe 6: 3-OMe , 6-H , 8-H ;

Classe 7: 3-H , 6-H , 8-OMe ;

Classe 8: 3-OMe , 6-H , 8-OMe ;

As informações espectrais destes flavonóides foram organizadas em uma tabela onde picos, ombros e inflexões foram qualificados com os seguintes valores: 1 para inflexão, -1 para ombro e 2 para pico. As Tabelas 2,3 e 4 do Anexo 1, apresentam estes dados para os espectros coletados em metanol, metanol mais cloreto de alumínio e metanol mais cloreto de alumínio mais ácido clorídrico, respectivamente.

Em termos matemáticos as Tabelas 2, 3 e 4 do Anexo 1 foram justapostas lateralmente formando a matriz **F**.

$$\mathbf{F} = \begin{bmatrix} \lambda_{11} T_{11} I_{11} & \lambda_{12} T_{12} I_{12} & \dots & \lambda_{1j} T_{1j} I_{1j} \\ \lambda_{21} T_{21} I_{21} & \lambda_{22} T_{22} I_{22} & \dots & \lambda_{2j} T_{2j} I_{2j} \\ \vdots & \vdots & \vdots & \vdots \\ \lambda_{i1} T_{i1} I_{i1} & \lambda_{i2} T_{i2} I_{i2} & \dots & \lambda_{ij} T_{ij} I_{ij} \end{bmatrix}$$

Onde:

λ - Comprimento de onda do pico, ombro ou inflexão;

T - Tipo :pico=2 , ombro= -1 , inflexão=1;

I - Intensidade;

Nesta matriz **F** são colocados os dados dos espectros em metanol, depois metanol mais AlCl_3 , metanol mais AlCl_3 mais HCl formando uma matriz de 54 colunas($j=54$), as variáveis, e 95 linhas, as amostras($i=95$).

2.5-Pré-tratamento

Os diferentes tipos de variáveis envolvidas, com diferentes grandezas, torna necessária a utilização de um pré-tratamento^{D1,D2}. Isto, para evitar que aquelas variáveis com altos valores tenham maior importância na análise dos dados.

No caso dos flavonóides, é recomendado aplicar um autoescalamiento, pois existem variáveis da ordem de unidades e outras em centenas, como é o caso do tipo do pico e as frequências, respectivamente. Este tratamento transforma todas as variáveis para uma mesma escala, ou seja, em termos de desvio padrão. Isto é feito através da subtração dos valores em cada variável de sua respectiva média, resultados estes que são divididos pelo desvios padrão das respectivas variáveis.

No caso do SIMCA, é interessante autoescalar cada classe separadamente. Isto pode ser feito pois as amostras a serem testadas no modelo serão projetadas em todas as classes.

Para este conjunto de dados, envolvendo os flavonóides, não foi possível autoescalar todas as variáveis de cada classe por existirem classes com variáveis cujos valores eram todos iguais. Assim quando estes valores são subtraídos de suas médias o resultado é zero o que proporciona desvio padrão igual a zero. A solução foi autoescalar as variáveis possíveis e apenas subtrair as restantes de suas médias.

2.6-Validação do modelo

Para verificar o comportamento do modelo para classificação de amostras que não foram utilizadas em sua construção, uma validação cruzada foi elaborada. Nesta validação, o modelo é construído sem uma amostra que é usada para teste, de forma que todas as amostras são testadas. Com relação ao número de componentes principais, estes são

escolhidos para as classes a partir da construção de um modelo SIMCA envolvendo todas as amostras.

Tal validação tem como objetivo identificar amostras "*outliers*", isto é, que estatisticamente sejam diferentes das outras de sua classe. Tal identificação, ocorre quando o modelo é construído sem a amostra "*outlier*", que é prevista em outra classe ou com um resíduo muito maior que as outras de sua classe.

A indicação do alto valor do resíduo é dado pelo valor de Q^2 que representa a soma dos quadrados dos resíduos. Um segundo parâmetro usado é a soma dos quadrados dos "escores", T^2 . Este último parâmetro está diretamente ligado com a variância nas componentes principais usadas para descrever as classes, que é dada pela divisão do somatório das coordenadas nas componentes principais, subtraídas de suas médias, pelo número de graus de liberdade. Como os dados estão autoescalados, o que significa média igual a zero, a diferença entre o valor da variância e de T^2 é o número de graus de liberdades.

O algoritmo do SIMCA usado neste trabalho, quando faz a previsão de uma amostra, verifica em qual ou quais classes ela pertence através dos valores de Q^2 e T^2 . Isto é, na construção do modelo após as amostras terem sido projetadas no espaço dos componentes principais, os limites para os valores de Q^2 e T^2 são calculados, com 95% de confiança, e guardados junto com as outras informações a respeito do modelo. Assim, a previsão de uma nova amostra é feita através de sua projeção nos espaços de componentes principais de cada classe do modelo, onde são calculados os valores de Q^2 e T^2 . Conhecidos Q^2 e T^2 da nova amostra, estes dois são comparados com os limites correspondentes às classes. Então naquela ou naquelas onde estes valores forem menores ela é classificada positivamente, ou seja, pertence a esta classe ou classes. Em caso contrário, Q^2 e T^2 maiores que os valores limites, a amostra é colocada na classe mais próxima ou até mesmo em nenhuma. A comparação entre os valores de Q^2 e T^2 das

classes com os de novas amostras é feita por teste F com 95% de confiança.

Assim, após todas as amostras terem sido testadas na validação do modelo, para cada classe serão conhecidos os valores de Q^2 e T^2 , isto é, daquelas amostras que caíram fora do 95% de confiança mas foram previstas na classe certa. Estes resultados são importantes pois em utilizações futuras do modelo, os valores de Q^2 e T^2 das novas previsões podem ser comparados com aqueles já conhecidos e indicar assim, se as novas amostras foram previstas próximo ao limite de 95% de confiança da classe.

Também foram testadas moléculas com substituições diferentes daquelas usadas na construção do modelo. Isto pode ser feito pois é possível identificar pelos valores de Q^2 e T^2 que estas amostras não pertencem às classes usadas na construção do modelo SIMCA. Este teste fornece uma indicação da capacidade de classificação do modelo.

2.7-Resultados

Os números de componentes principais usados para as oito classes são apresentados na Tabela 2.1. Estes números são escolhidos em separado para cada classe, através da análise de gráficos dos valores de Q^2 versus T^2 , pois eles permitem observar a distância entre as classes e a distribuição dentro de cada classe. As Figuras (2.11), (2.12) e (2.13) mostram os gráficos dos logaritmos de Q^2 e T^2 para a classe 1, calculados para 1, 2, 3, 4 e 5 componentes principais. Nestes gráficos, nota-se que os números de componentes principais que permitem separar a classe 1 por completo são quatro ou cinco. Assim, o número de componentes principais escolhido é quatro, para que o limite da classe fique menor, reduzindo desta maneira, a possibilidade de misturas entre classes.

As Figuras de (2.14) a (2.21) apresentam os gráficos para as oito classes, cujos números de componentes principais estão indicados na Tabela 2.1, onde para cada classe são apresentados dois gráficos, o primeiro usado para identificar as classes e no segundo o número das amostras na Tabela 1 do Anexo 1.

Tabela 2.1- Número de componentes principais e suas variâncias capturadas, para as oito classes.

Classe	Número de componentes principais	Variância capturada (%)
1	4	87,52
2	2	78,92
3	3	69,35
4	3	65,01
5	3	68,69
6	4	76,98
7	1	54,87
8	2	64,51

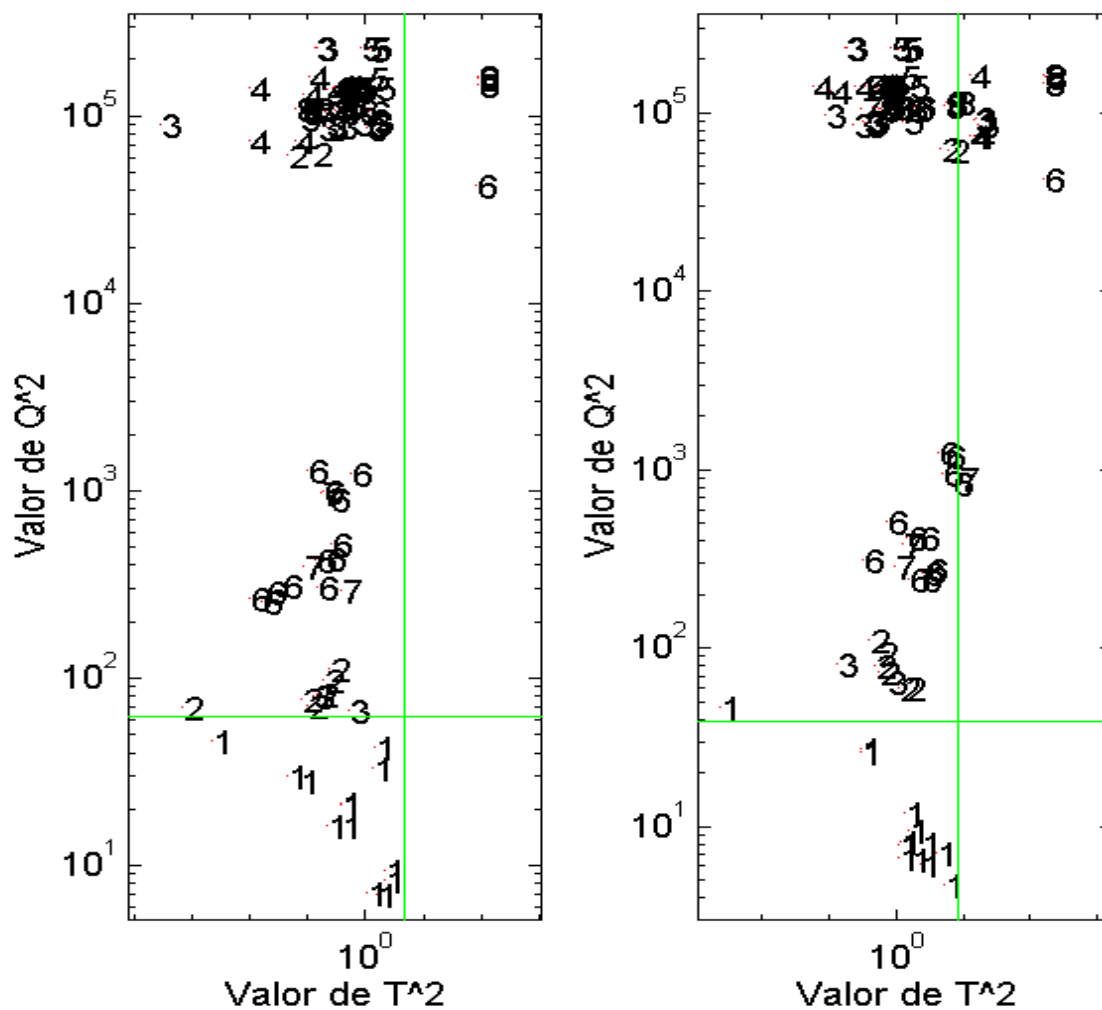


Figura 2.11 -Gráficos para todas as amostras projetadas no(s) componentes principais da classe1. À esquerda, uma componente principal e à direita, duas componentes principais.

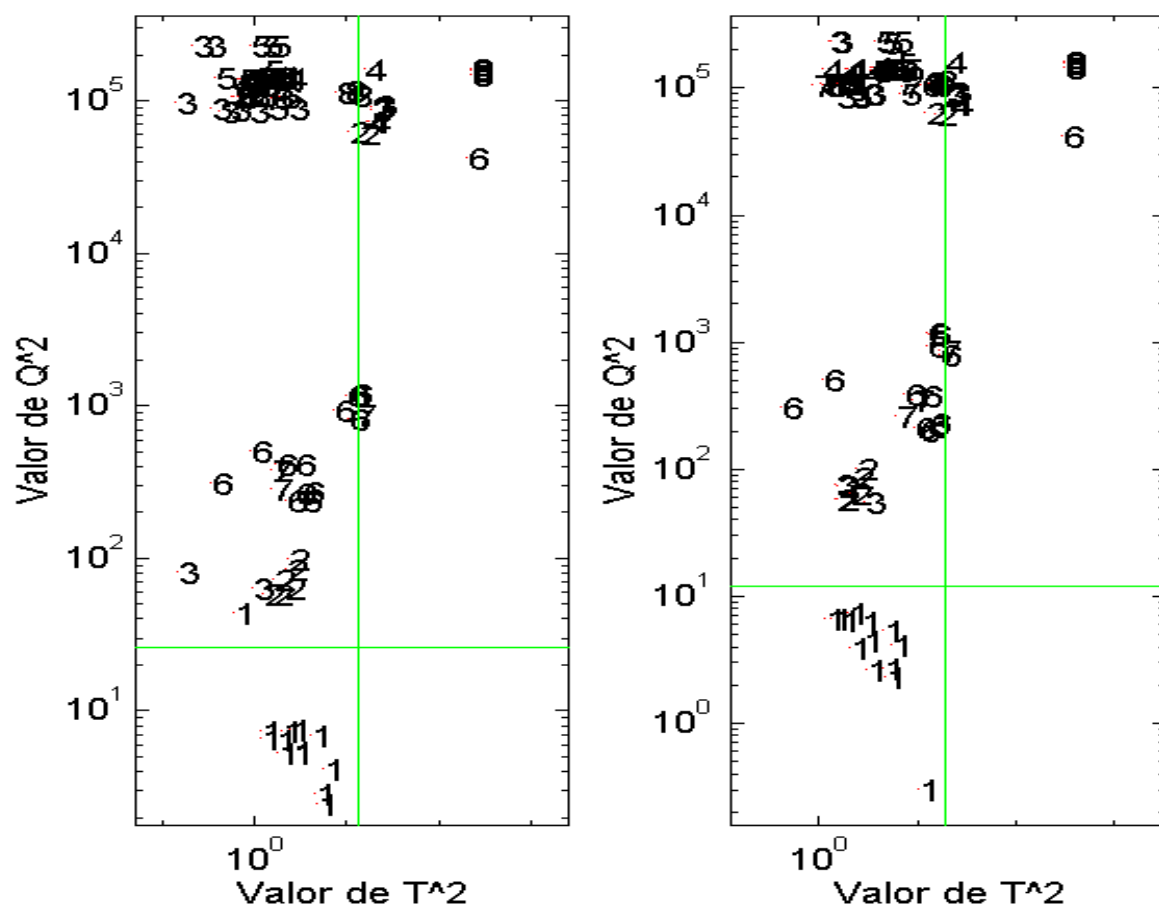


Figura 2.12 -Gráficos para todas as amostras projetadas nos componentes principais da classe1. À esquerda,3 componentes principais e à direita, 4 componentes principais.

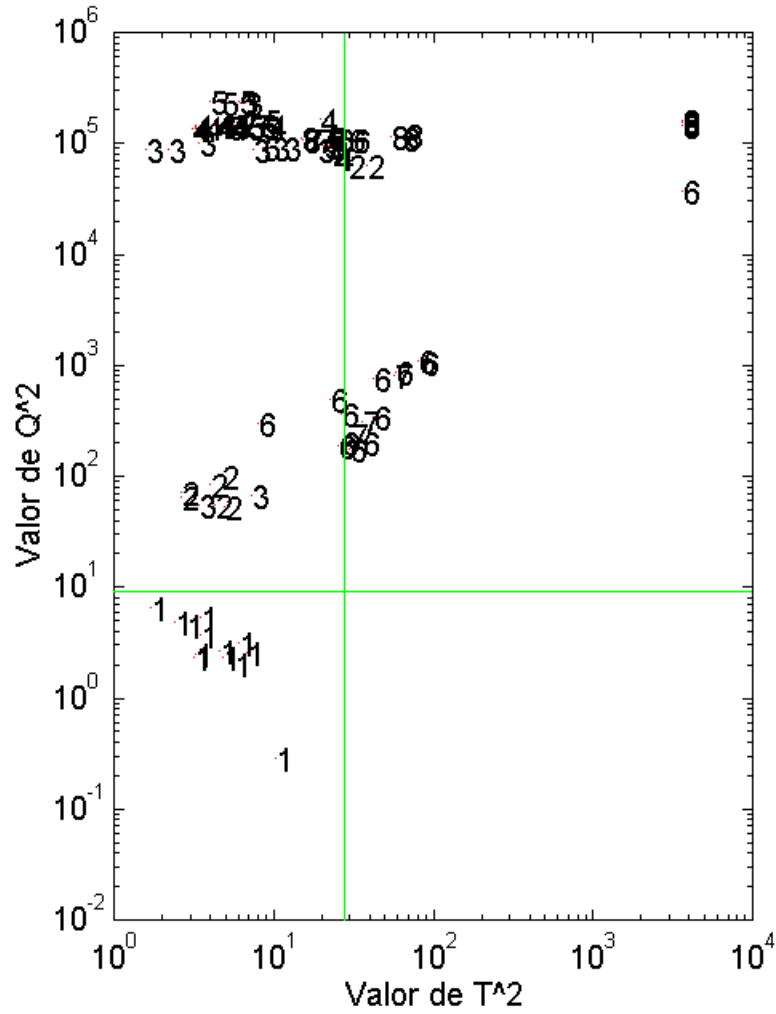


Figura 2.13 -Gráficos para todas as amostras projetadas nos componentes principais da classe1, neste caso 5 componentes.

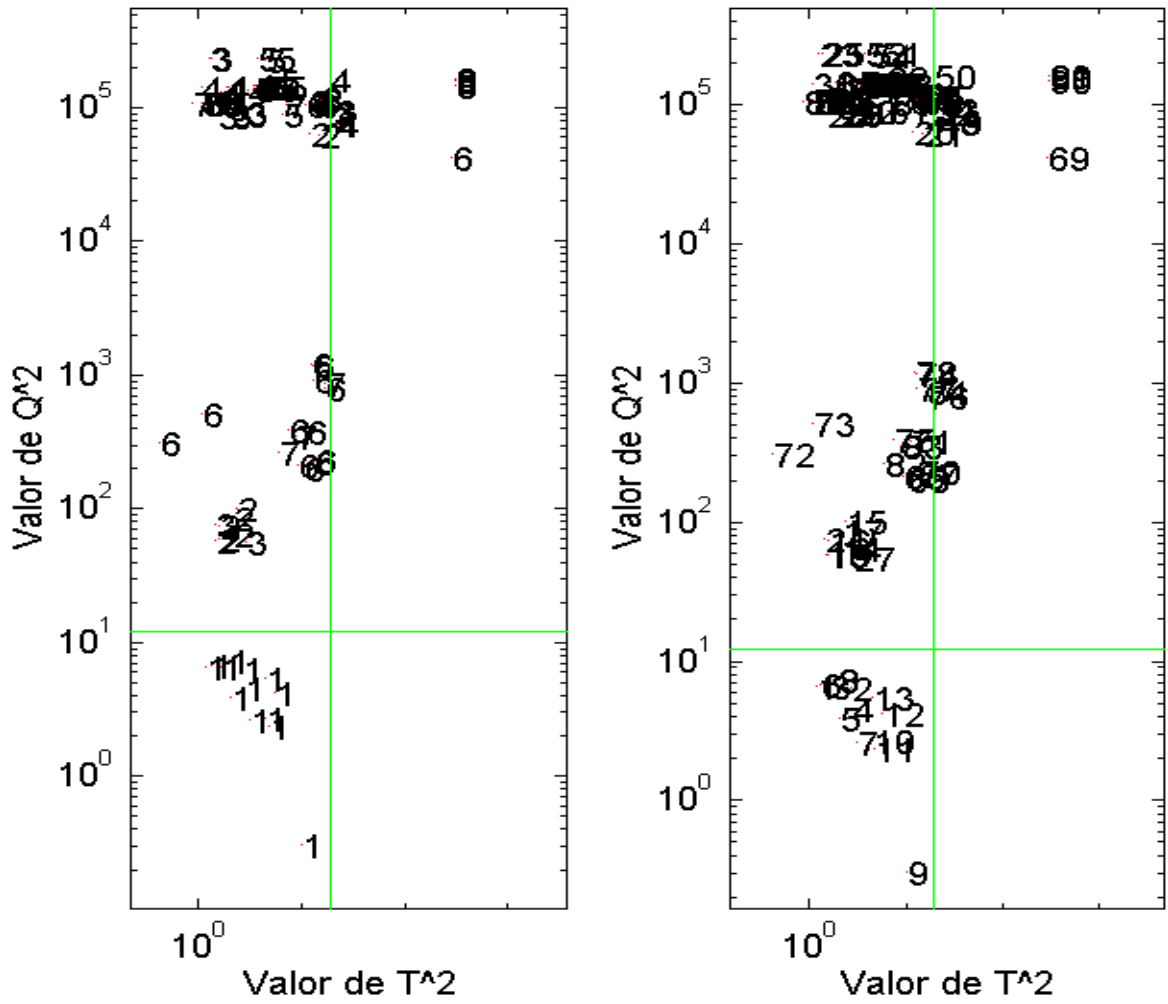


Figura 2.14 -Gráficos para todas as amostras projetadas nos componentes principais da classe1, para o primeiro modelo SIMCA,(4 componentes principais). À esquerda identificação das classes e à direita identificação do número da amostra na Tabela 1 do Anexo 1.

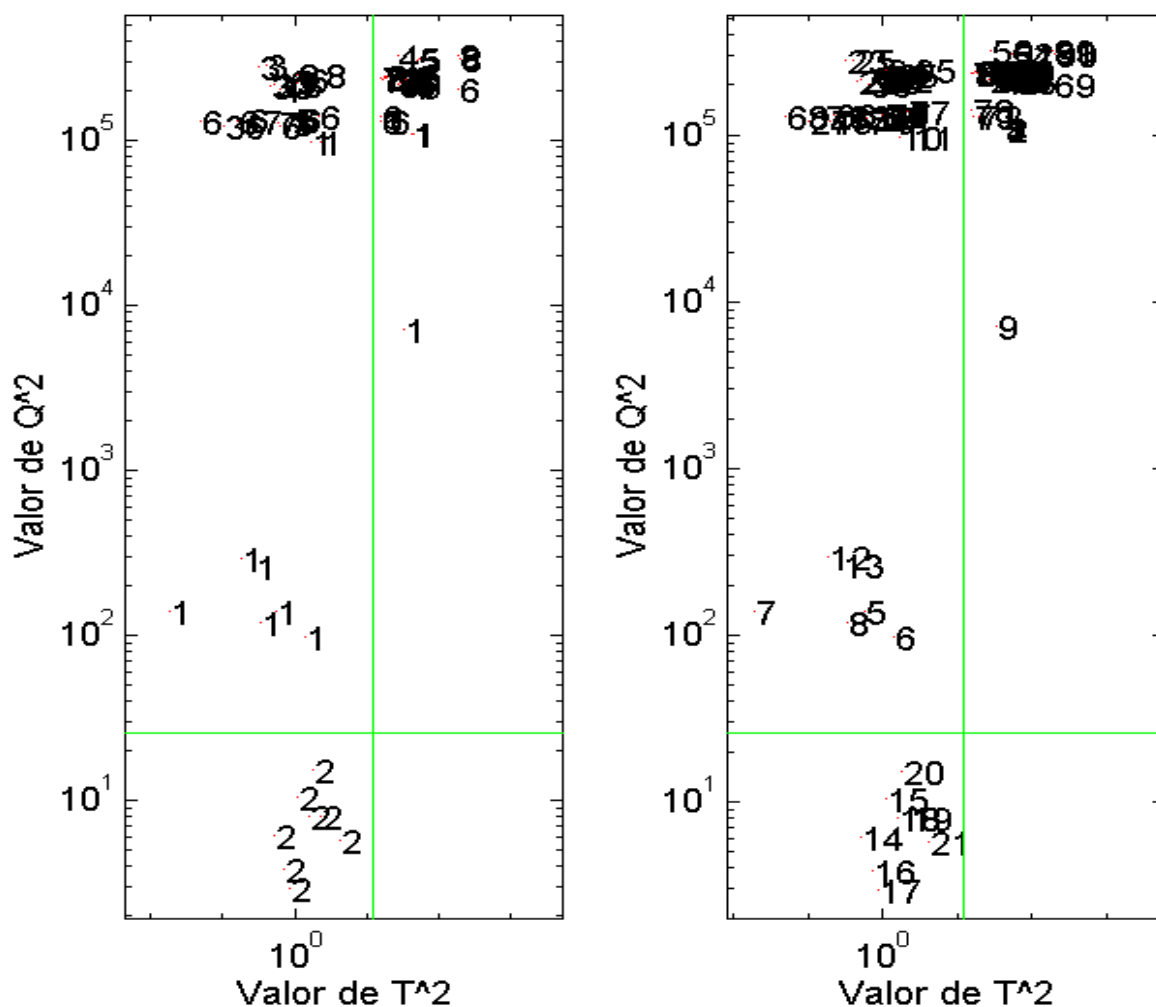


Figura 2.15 -Gráficos para todas as amostras projetadas nos componentes principais da classe2, para o primeiro modelo SIMCA, (2 componentes principais). À esquerda identificação das classes e à direita identificação do número da amostra na Tabela 1 do Anexo 1.

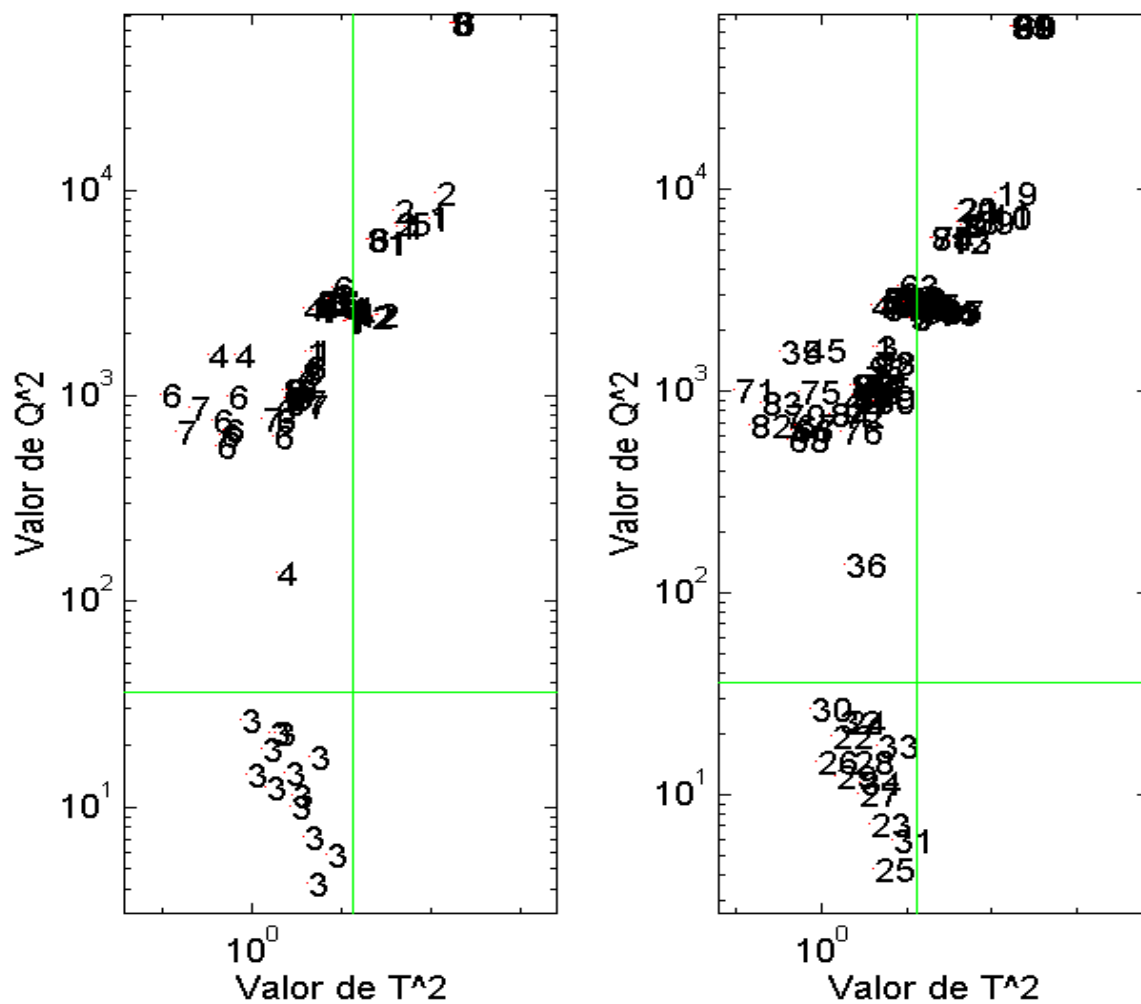


Figura 2.16 -Gráficos para todas as amostras projetadas nos componentes principais da classe 3, para o primeiro modelo SIMCA, (3 componentes principais). À esquerda identificação das classes e à direita identificação do número da amostra na Tabela 1 do Anexo 1.

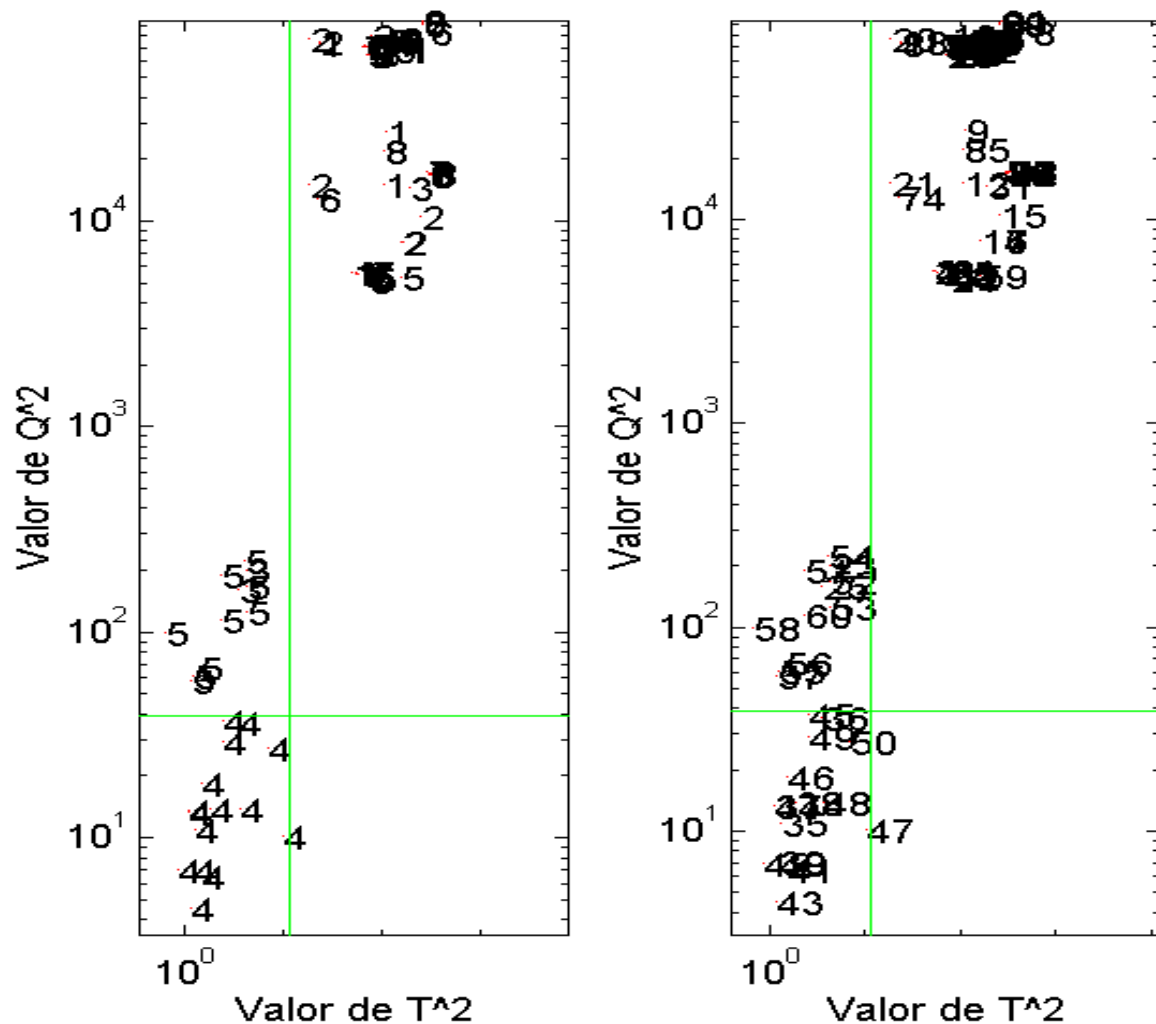


Figura 2.17 -Gráficos para todas as amostras projetadas nos componentes principais da classe 4, para o primeiro modelo SIMCA, (3 componentes principais). À esquerda identificação das classes e à direita identificação do número da amostra na Tabela 1 do Anexo 1.

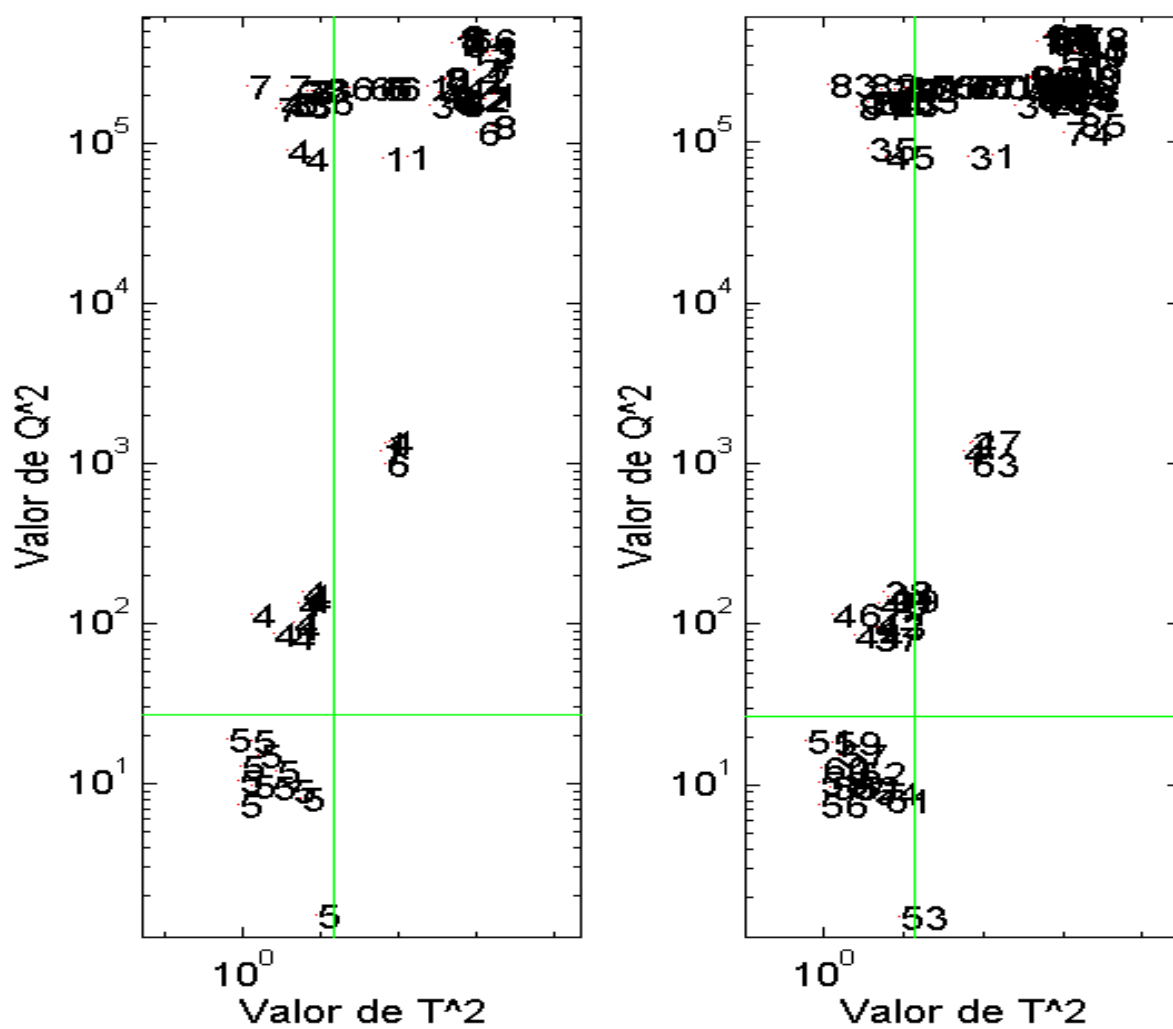


Figura 2.18 -Gráficos para todas as amostras projetadas nos componentes principais da classe 5, para o primeiro modelo SIMCA, (3 componentes principais). À esquerda identificação das classes e à direita identificação do número da amostra na Tabela 1 do Anexo 1.

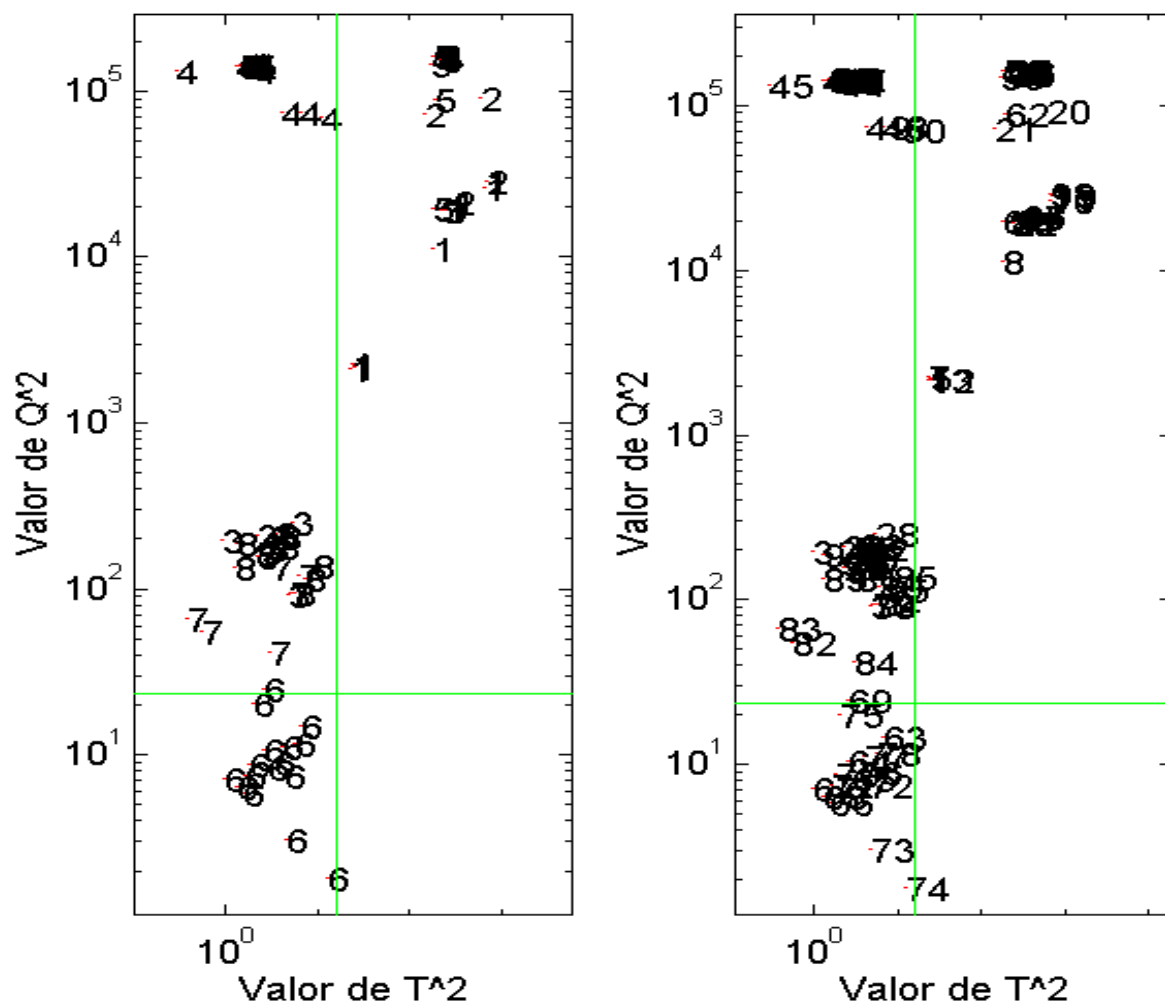


Figura 2.19 -Gráficos para todas as amostras projetadas nos componentes principais da classe 6, para o primeiro modelo SIMCA, (4 componentes principais). À esquerda identificação das classes e à direita identificação do número da amostra na Tabela 1 do Anexo 1.

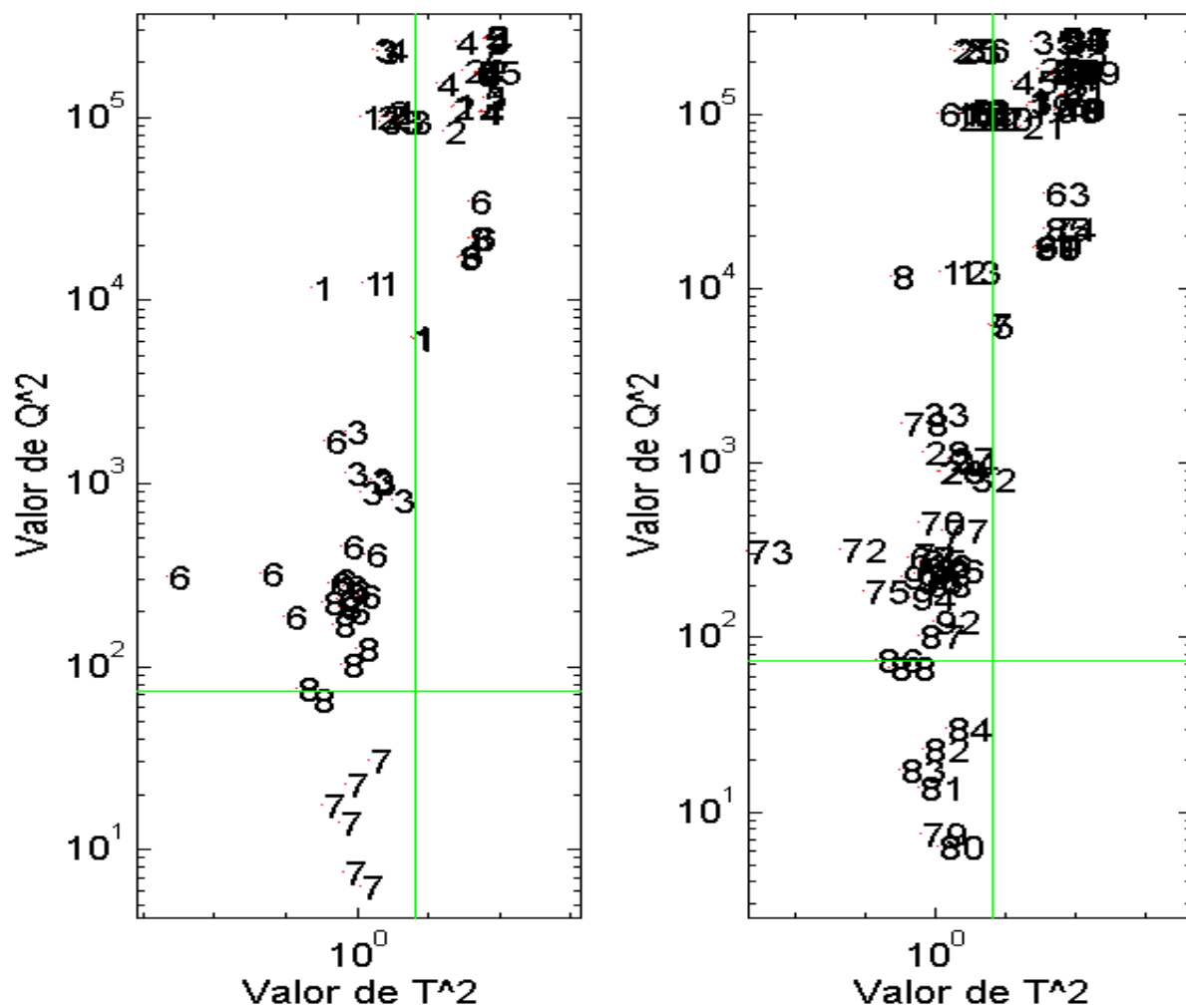


Figura 2.20 -Gráficos para todas as amostras projetadas nos componentes principais da classe 7, para o primeiro modelo SIMCA, (1 componente principal). À esquerda identificação das classes e à direita identificação do número da amostra na Tabela 1 do Anexo 1.

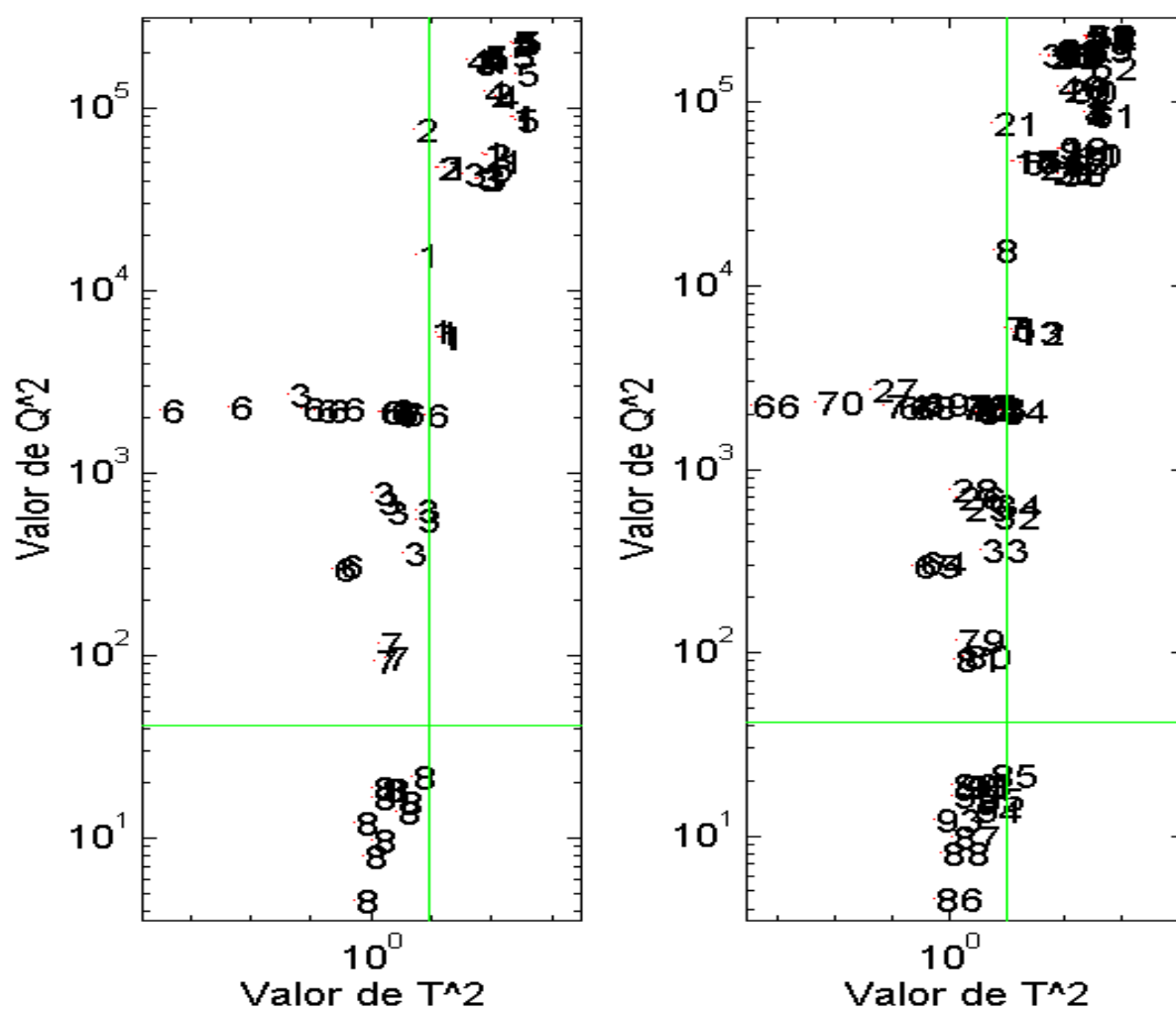


Figura 2.21 -Gráficos para todas as amostras projetadas nos componentes principais da classe 8, para o primeiro modelo SIMCA, (2 componentes principais). À esquerda identificação das classes e à direita identificação do número da amostra na Tabela 1 do Anexo 1.

Após o primeiro teste da validação cruzada no SIMCA, dezessete amostras apresentaram problemas quando previstas. Destas, quatro foram previstas na classe certa, mas com altos valores de resíduos. As treze restantes foram previstas em classes erradas. O resultado deste teste é apresentado na Tabela 5 do Anexo 1, onde as amostras com problemas estão marcadas com um asterisco, são elas:

- 9, 15, 19, 21, 33, 36, 45, 47, 50, 53, 59, 61, 63, 69, 74, 84 e 85.

As amostras 9, 15, 19, 21, 36, 53, 69 foram previstas erradamente. Tais amostras possuem como diferença em seus espectros a presença ou não de um ombro, se comparados aos daquelas amostras de suas classes, que possuem o mesmo número de substituição e semelhança em termos da presença ou não de hidroxilas vicinais em 3' e 4'. Estas, também apresentam altos valores de resíduos e foram retiradas para a construção de um novo modelo.

As amostras 74 e 85 também foram excluídas para a construção de um novo modelo, pois foram previstas erradamente no primeiro teste, pelo fato de seus espectros terem sido coletados em etanol.

As amostras 45, 47, 50 e 84 foram previstas na classe correta, mas apresentaram altos valores de resíduos. Por apresentarem estes alto valores, elas também foram excluídas para a construção de um novo modelo. Isto é necessário, pois o limite da "caixa" criada pelo SIMCA, para a classe de 45, 47 e 50, é bastante maior com a suas presenças.

Quanto ao restante das amostras, 33, 59, 61 e 63 também foram retiradas para a construção de um novo modelo, por terem sido previstas em classes erradas.

Após retirar estas amostras, um novo modelo foi construído. Para este último, a Tabela 2.2 apresenta o número de componentes principais necessários para descrever cada classe e suas variâncias capturadas. Esta Tabela 2.2 mostra que nas classes 1, 2, 5 e 6 houve redução no

número de componentes principais indicando, assim, o uso de componentes principais adicionais, no primeiro modelo, para descrever aquelas amostras que foram retiradas.

As Figuras de (2.22) a (2.29) mostram os gráficos de Q^2 versus T^2 , correspondentes aos dados indicados na Tabela 2.2, onde a distribuição de cada classe passa a ser mais homogênea, se comparada com aquela referente ao primeiro modelo. Isto é resultado de uma melhor definição dos limites de cada classe e maior semelhança entre suas amostras .

Tabela 2.2- Número de componentes principais e suas variâncias capturadas, para as oito classes, para o segundo modelo SIMCA.

Classe	Número de componentes principais	Variância capturada (%)
1	3	85,48
2	1	67,37
3	3	69,79
4	3	75,96
5	2	67,74
6	3	78,70
7	1	69,81
8	2	72,03

O resultado da validação cruzada para este novo modelo apresentou cinco amostras previstas erradamente e uma com alto valor de resíduo. Aquelas previstas erradamente são: 18, 20, 35, 62, 82; e a com alto valor de resíduo foi a 78, Tabela 6, no Anexo 1 .

Estes novos erros de previsão confirmam que as amostras previamente retiradas estavam aumentando os limites das "caixas" de suas classes. Assim, com sua retirada os limites foram reduzidos e amostras próximas a estes limites agora foram previstas erradamente.

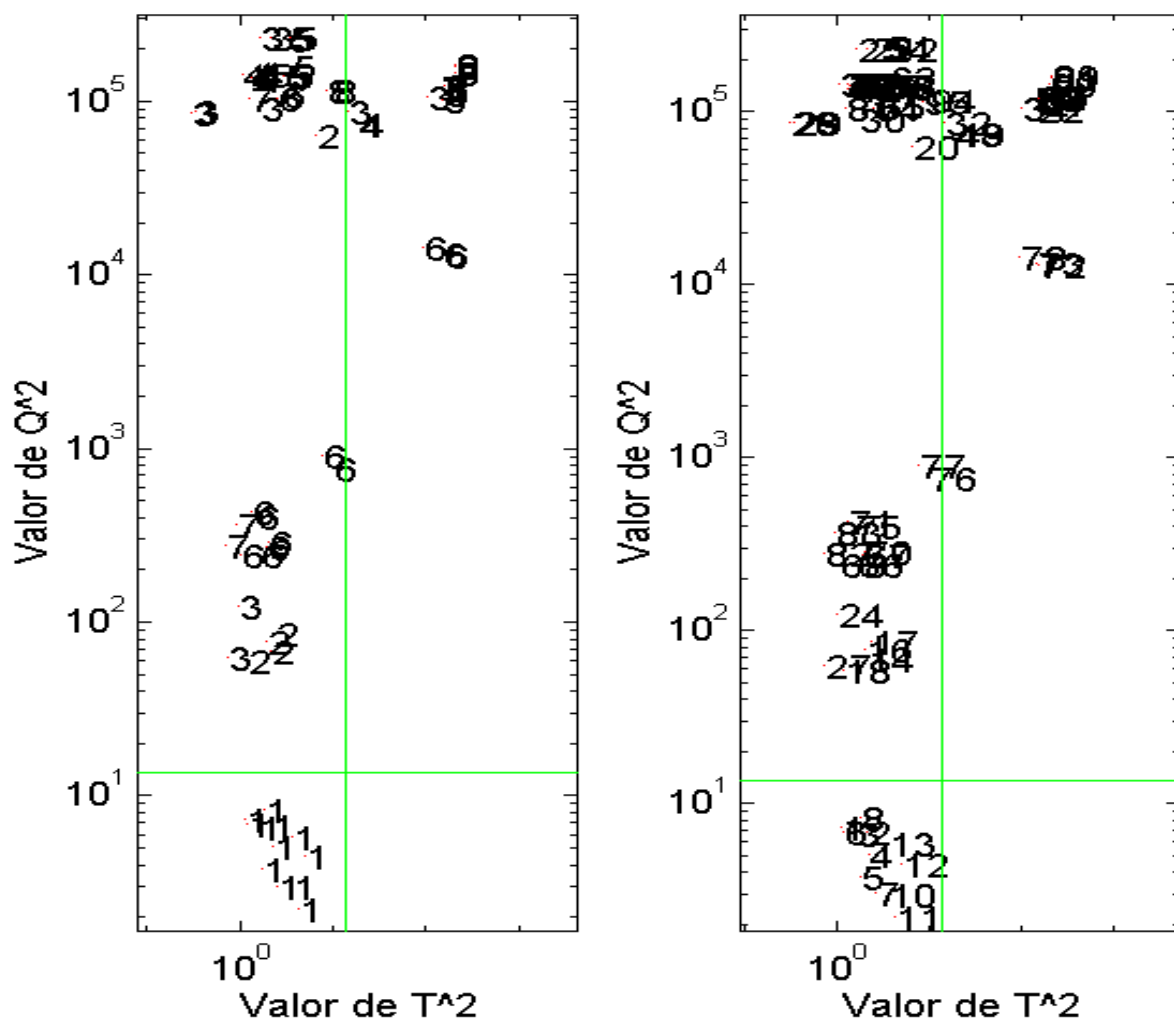


Figura 2.22 -Gráficos para todas as amostras projetadas nos componentes principais da classe 1, para o segundo modelo do SIMCA,(3 componentes principais). À esquerda identificação das classes e à direita identificação do número da amostra na Tabela 1 do Anexo 1.

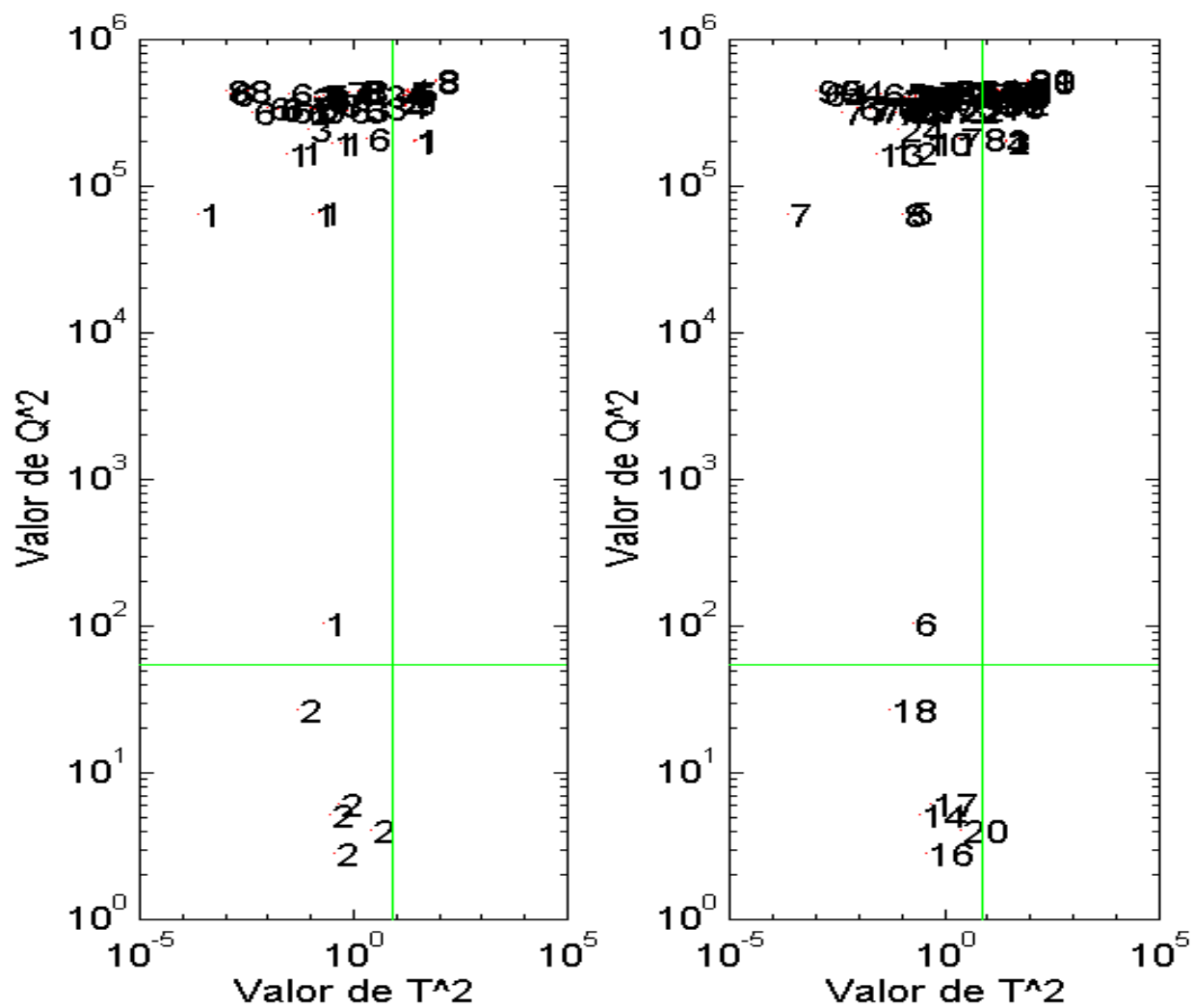


Figura 2.23 -Gráficos para todas as amostras projetadas nos componentes principais da classe 2, para o segundo modelo do SIMCA, (1 componente principal). À esquerda identificação das classes e à direita identificação do número da amostra na Tabela 1 do Anexo 1.

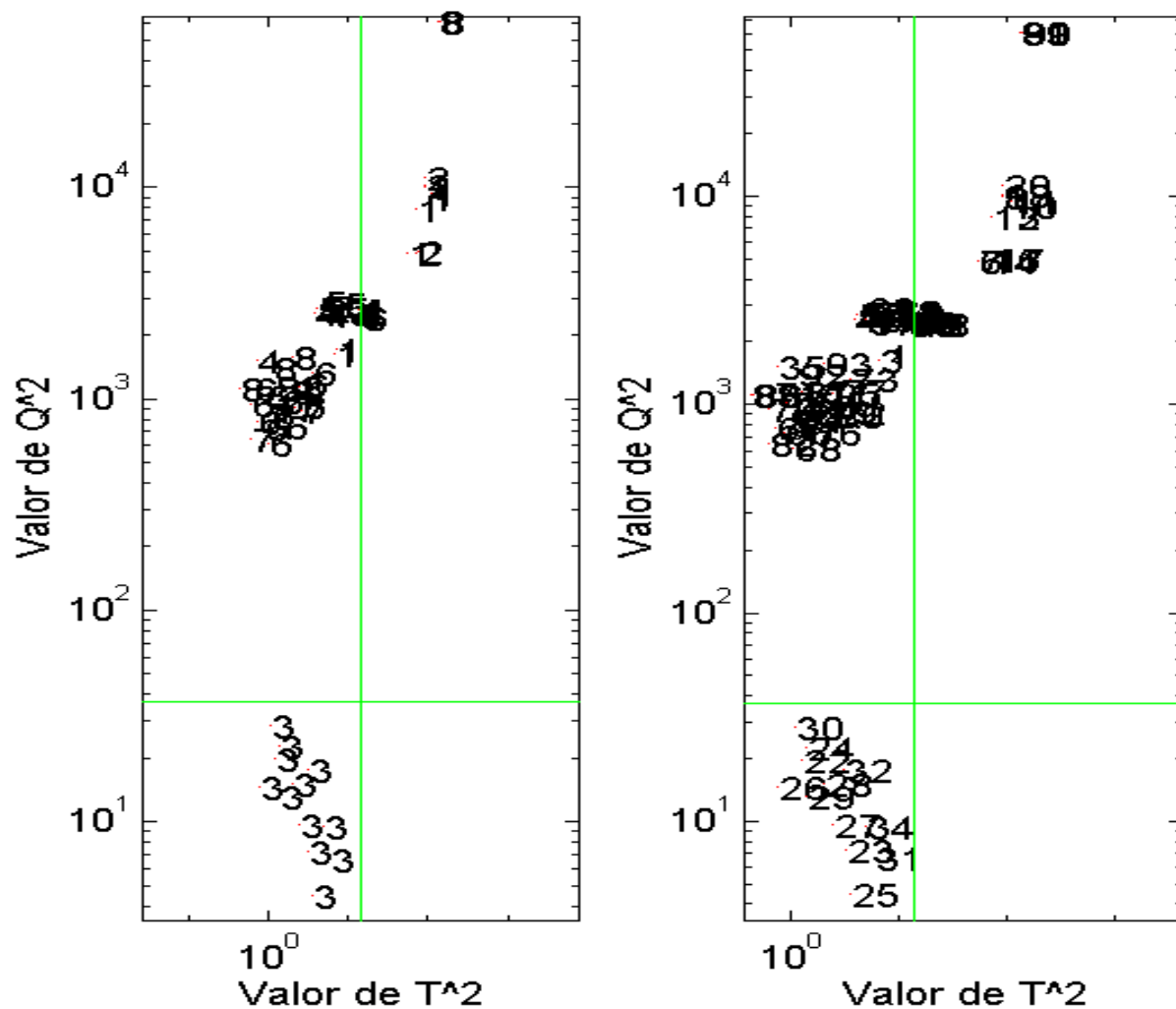


Figura 2.24 -Gráficos para todas as amostras projetadas nos componentes principais da classe 3, para o segundo modelo do SIMCA, (3 componentes principais). À esquerda identificação das classes e à direita identificação do número da amostra na Tabela 1 do Anexo 1.

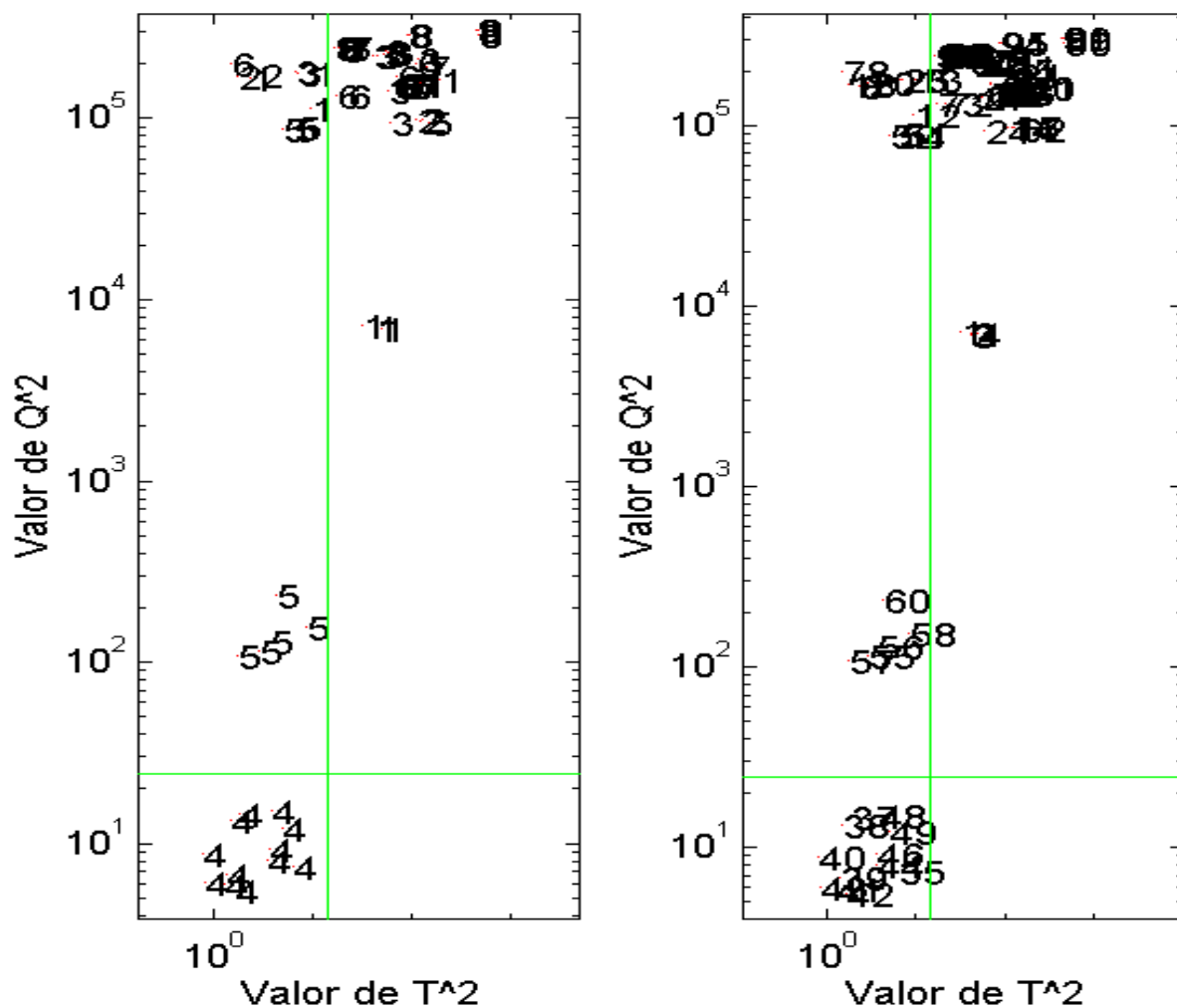


Figura 2.25 -Gráficos para todas as amostras projetadas nos componentes principais da classe 4, para o segundo modelo do SIMCA, (3 componentes principais). À esquerda identificação das classes e à direita identificação do número da amostra na Tabela 1 do Anexo 1.

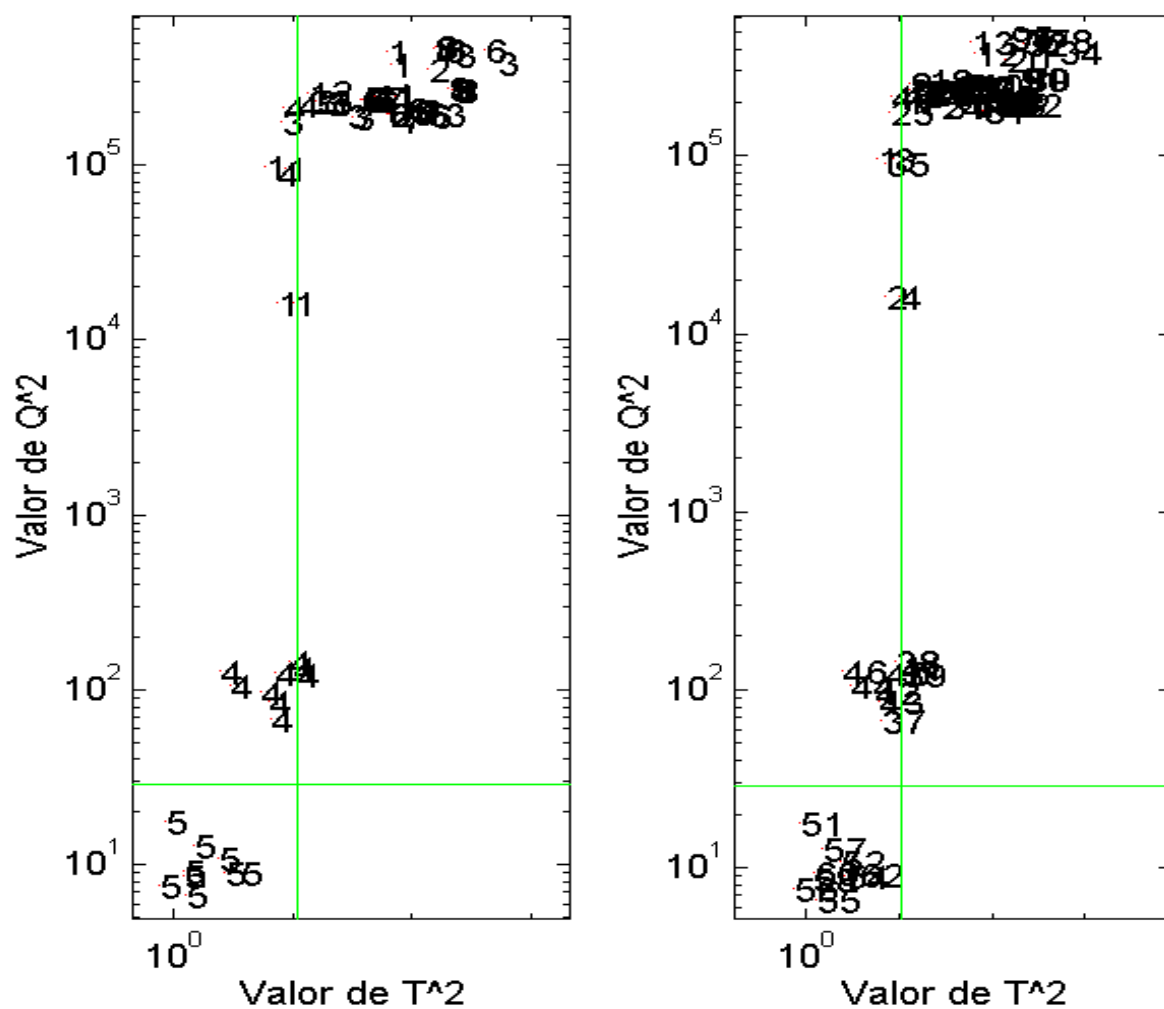


Figura 2.26 -Gráficos para todas as amostras projetadas nos componentes principais da classe 5, para o segundo modelo do SIMCA, (2 componentes principais). À esquerda identificação das classes e à direita identificação do número da amostra na Tabela 1 do Anexo 1.

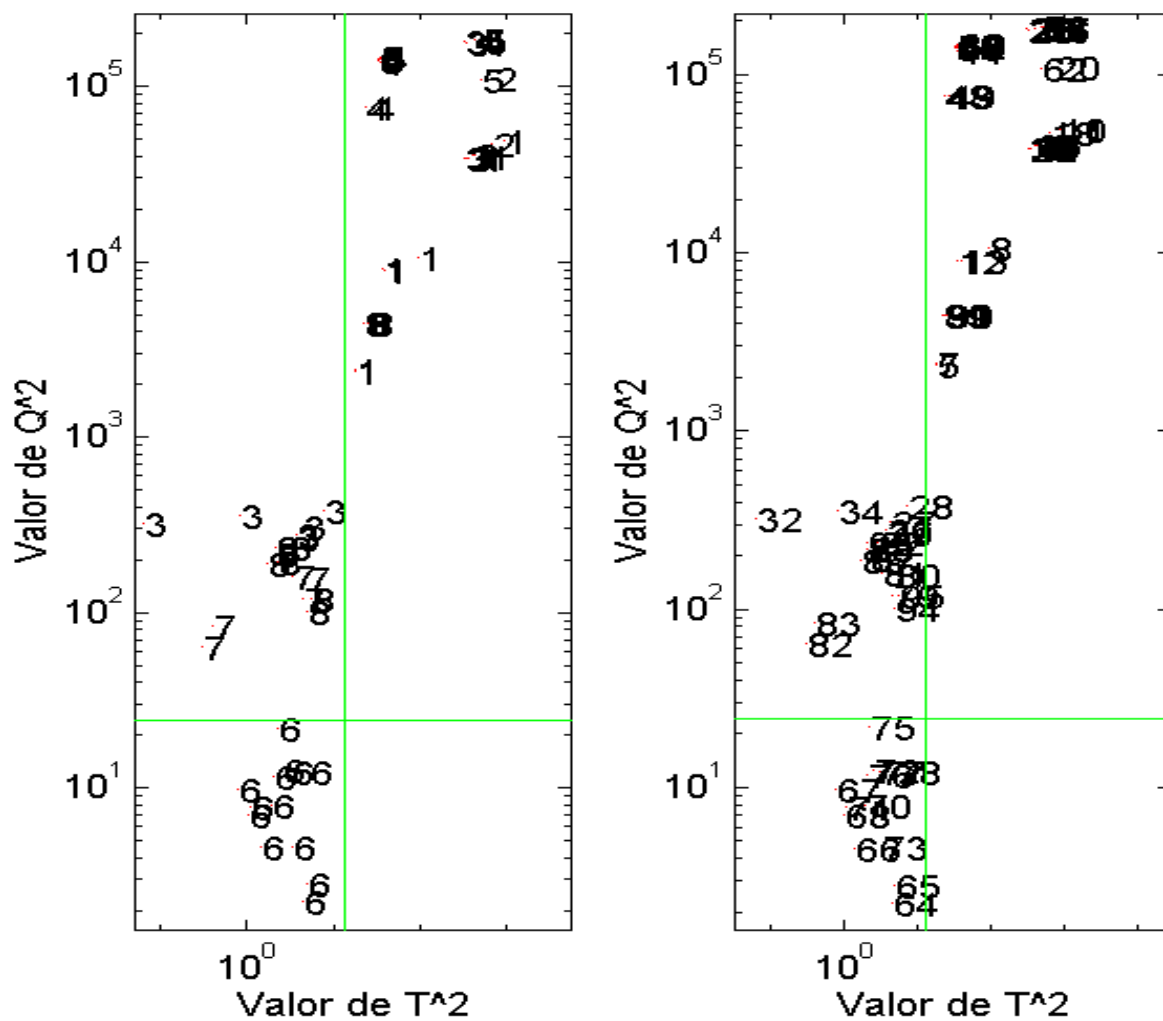


Figura 2.27 -Gráficos para todas as amostras projetadas nos componentes principais da classe 6, para o segundo modelo do SIMCA, (3 componentes principais). À esquerda identificação das classes e à direita identificação do número da amostra na Tabela 1 do Anexo 1.

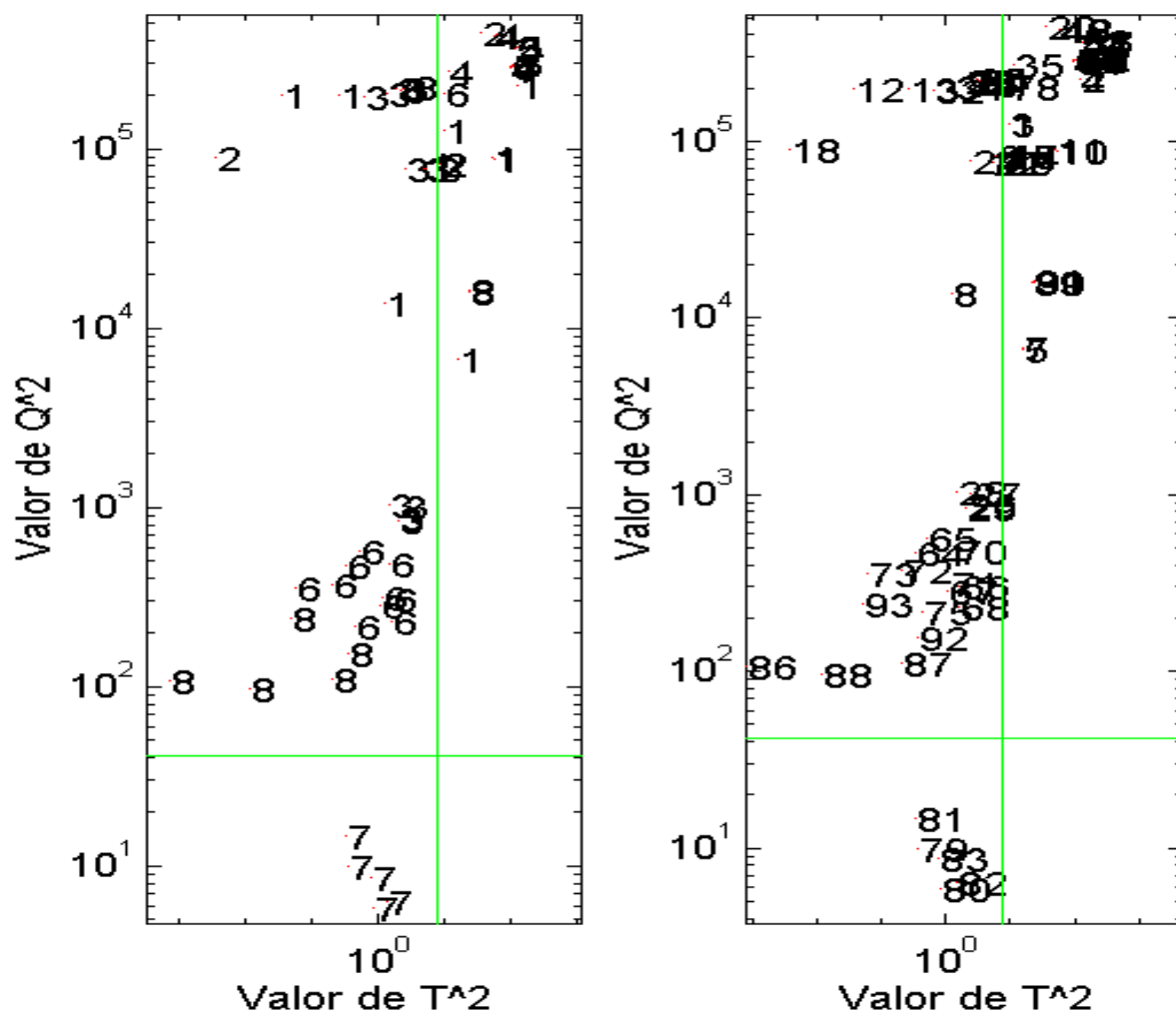


Figura 2.28 -Gráficos para todas as amostras projetadas nos componentes principais da classe 7, para o segundo modelo do SIMCA, (1 componente principal). À esquerda identificação das classes e à direita identificação do número da amostra na Tabela 1 do Anexo 1.

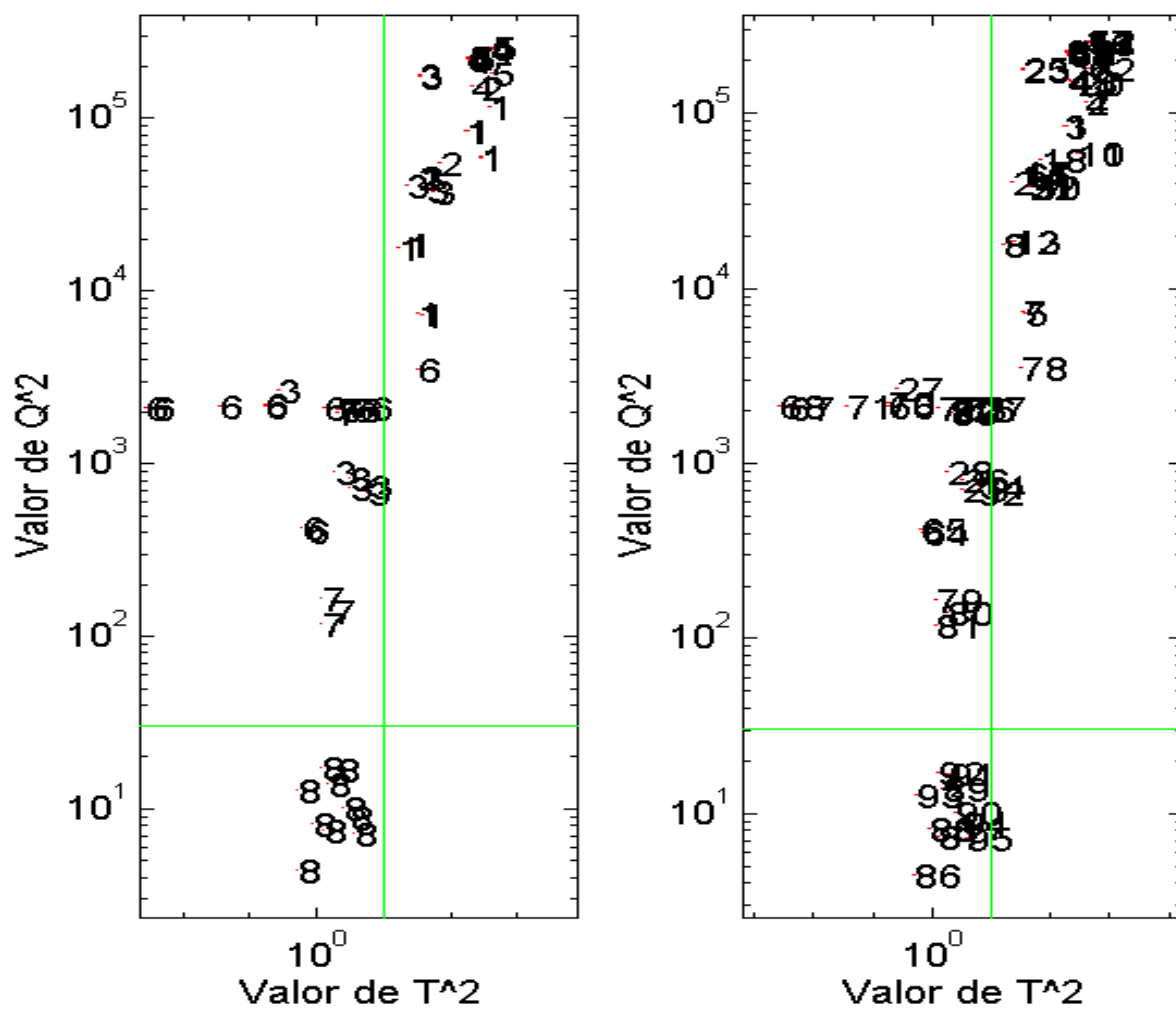


Figura 2.29 -Gráficos para todas as amostras projetadas nos componentes principais da classe 8, para o segundo modelo do SIMCA, (2 componentes principais). À esquerda identificação das classes e à direita identificação do número da amostra na Tabela 1 do Anexo 1.

As amostras com substituições, na estrutura molecular da flavona, diferentes daquelas usadas para a construção deste último modelo também foram testadas. O resultado mostra que elas foram colocadas nas classes com altos valores de resíduos, o que permite identificá-las.

Dentre as 24 testadas, Tabela 7 no Anexo 1, 7 apresentaram problemas quanto aos valores dos resíduos ou seja, possuem resíduo próximo aos encontrados para as amostras do modelo. Destas 7, três possuem como diferença em sua estrutura molecular, se comparada com as das classes em que foram previstas, um radical metila a mais na cadeia. Na Tabela 7 no Anexo 1 são as amostras 7, 8 e 9.

Quanto ao restante, as amostras 14 e 15 possuem um radical metoxila a mais do que a classe em que foi prevista, para 21 e 22 a diferença está no carbono 8, primeiro substituído por uma hidroxila e depois por um radical metila, pois na classe em que foram previstas o carbono 8 possui uma metoxila.

Os gráficos de componentes principais de cada classe mostram que é possível identificar o número de substituições no anel B, o que também pode ser feito através de uma análise por agrupamento hierárquico. Nas Figuras 2.30 e 2.31 são apresentados dois exemplos destes gráficos para a classe 1, o primeiro para componentes principais e o segundo para dendrogramas. É interessante notar, também, a possibilidade de identificar as duas hidroxilas vicinais em 3' e 4'. A identificação do número de substituições e existência de hidroxilas vicinais é feita pela verificação da formação de grupos nos gráficos dos componentes e dendrogramas. No caso das Figuras 2.30 e 2.31 nota-se a formação de grupos para amostras com uma substituição, identificadas por mono, com duas substituições, identificadas por di, com três substituições, identificadas por tri e com hidroxilas vicinais identificadas por diOH. Desta forma, na identificação do número de substituições do anel B para uma nova amostra basta verificar de qual grupo ela se

aproxima mais após a projeção, no caso dos gráficos dos componentes principais, ou em que grupo do gráfico de dendrograma ela é colocada.

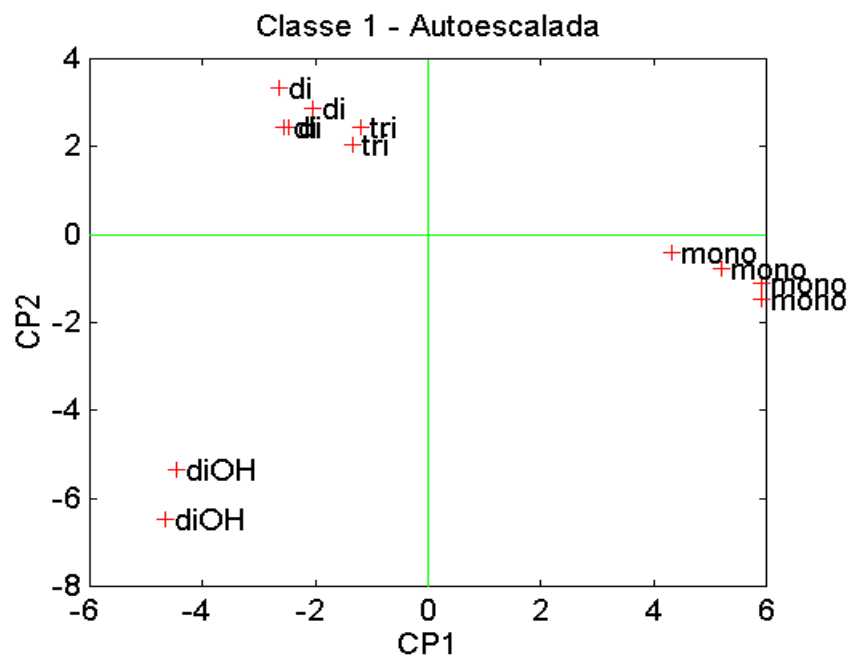


Figura 2.30. CP1 e CP2 -Componentes principais para os flavonóides da classe 1.

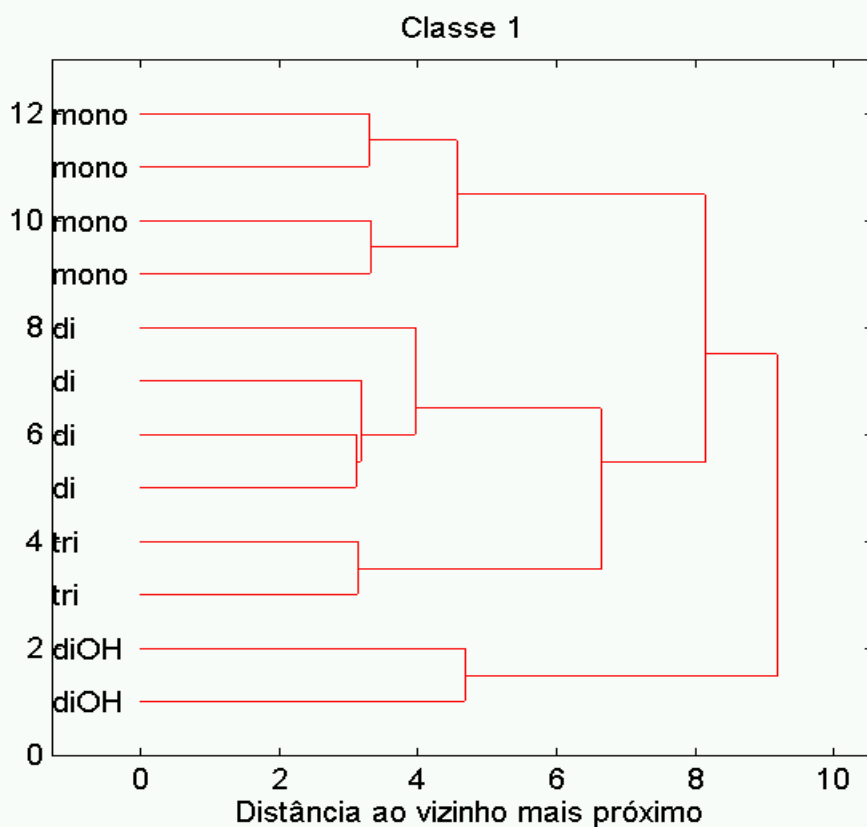


Figura.2.31 Análise por agrupamento hierárquico para os flavonóides da classe1.

2.8-Conclusões

A solução do problema, ou seja, elucidação de todos os substituintes, não é completa por serem as informações contidas no modelo insuficientes para separar matematicamente todas as classes. Ou seja, separar aquelas moléculas com substituições diferentes em 4', 3', 5' e 7. Por outro lado os resultados para as posições 3, 6, 8 são bastante confiáveis, tendo visto que não houve mistura entre classes, o que é comprovado pelo baixo número de amostras previstas erradamente, 7 em 78 no segundo modelo e por nenhuma amostra ter sido prevista em duas ou mais classes. Para o anel B ainda é possível identificar se existem hidroxilas vicinais em 3' e 4'.

Estes resultados mostram que os objetivos foram alcançados, uma vez que foi possível a construção de modelos, com baixo índice de erros e possíveis de serem detectados através dos resíduos. Também mostra que a quimiometria pode facilitar o trabalho de elucidação de estruturas de forma prática, pois após a construção dos modelos novas amostras podem ser testadas rapidamente. Sendo que, para respostas positivas a maior parte da estrutura molecular estará elucidada.

O avanço em tal área pode ser conseguido com um número maior de informações, como o uso do espectro completo, pois assim, os erros experimentais terão uma menor importância frente a estas informações usadas na construção de modelos. Este avanço se traduz em um maior índice de acertos e redução no número de tipos de espectros a serem usados, ou seja, seria necessário espectros em apenas uma solução. Estas são hipóteses que não puderam ser verificadas neste trabalho por ele contar apenas com informações publicadas.

ANEXO

Tabela 1- Substituições nos anéis A e B da estrutura da flavona para os 95 flavonóides usados na classificação.

n ^o	OH	OMe	Classe
1	5,7,4'	6	1
2	5,7	6,4'	1
3	5,4'	6,7	1
4	5	6,7,4'	1
5	5,7,3'	6,4'	1
6	5,7,4'	6,3'	1
7	5,3'	6,7,4'	1
8	5	6,7,3',4'	1
9	5,7,3'	6,4',5'	1
10	5,7,4'	6,3',5'	1
11	5,4'	6,7,3',5'	1
12	5,7,3',4'	6	1
13	5,3',4'	6,7	1
14	5,6,7,4'	-	2
15	5,6,7	4'	2
16	5,6,4'	7	2
17	5,6	7,4'	2
18	5,6,7,3'	4'	2
19	5,6,3'	7,4'	2
20	5,6,7,3',4'	-	2
21	5,6,3',4'	7	2
22	5,7,4'	-	3
23	5,7	4'	3
24	5,4'	7	3
25	5	7,4'	3
26	5,7,3'	4'	3
27	5,7,4'	3'	3
28	5,7	3',4'	3
29	5,3'	7,4'	3
30	5,7,4'	3',5'	3
31	5,7	3',4',5'	3
32	5,7,3',4'	-	3
33	5,7,3',4',5'	-	3
34	5,7,3',4'	5'	3
35	5,7,4'	3,6	4

Continuação: Tabela 1

n ^o	OH	OMe	Classe
36	5,7	3,6,4'	4
37	5,4'	3,6,7	4
38	5,7,3'	3,6,4'	4
39	5,7,4'	3,6,3'	4
40	5,3'	3,6,7,4'	4
41	5,4'	3,6,7,3'	4
42	5,7	3,6,3',4'	4
43	5	3,6,7,3',4'	4
44	5,7,3',5'	3,6,4'	4
45	5,7,3'	3,6,4',5'	4
46	5,3',5'	3,6,7,4'	4
47	5,4'	3,6,7,3',5'	4
48	5,7,3',4'	3,6	4
49	5,3',4'	3,6,7	4
50	5,7,3',4'	3,6,5'	4
51	5,7,4'	6,8	5
52	5,4'	6,7,8,	5
53	5,7	6,8,4'	5
54	5	6,7,8,4'	5
55	5,7,3'	6,8,4'	5
56	5,7,4'	6,8,3'	5
57	5,4'	6,7,8,3'	5
58	5,3'	6,7,8,4'	5
59	5,7	6,8,3',4'	5
60	5	6,7,8,3',4'	5
61	5,3',5'	6,7,8,4'	5
62	5	6,7,8,3',4',5'	5
63	5,7,4'	3	6
64	5,7	3,4'	6
65	5	3,7,4'	6
66	5,7,4'	3,3'	6
67	5,7,3'	3,4'	6
68	5,7	3,3',4'	6
69	5,4'	3,7,3'	6
70	5,3'	3,7,4'	6
71	5	3,7,3',4'	6
72	5,7,3',5'	3,4'	6
73	5,3',5'	3,7,4'	6
74	5,7	3,3',4',5'	6
75	5	3,7,3',4',5'	6

Continuação: Tabela 1

n ^o	OH	OMe	Classe
76	5,7,3',4'	3	6
77	5,3',4'	3,7	6
78	5,7,3',4',5'	3	6
79	5,7,4'	8	7
80	5,7	8,4'	7
81	5	7,8,4'	7
82	5,7,4'	8,3'	7
83	5	7,8,3',4'	7
84	5,7,3',4'	8	7
85	5,7,4'	3,8	8
86	5,4'	3,7,8	8
87	5,7	3,8,4'	8
88	5	3,7,8,4'	8
89	5,7,4'	3,8,3'	8
90	5,4'	3,7,8,3'	8
91	5,3'	3,7,8,4'	8
92	5,7,3'	3,8,4',5'	8
93	5,3',5'	3,7,8,4'	8
94	5,7,3',4'	3,8	8
95	5,3',4'	3,7,8	8

Fonte: Referência E4

Tabela 2 - Informações espectrais na região do ultra violeta para os 95 flavonóides da Tabela 1. Estes dados se referem aos espectros coletados para os flavonóides em metanol.

n ^o	Banda II									Banda I								
	P	T	Int.	P	T	Int.	P	T	Int.	P	T	Int.	P	T	Int.	P	T	Int.
1	0	0	0	0	0	0	275	2	0,65	0	0	0	336	2	1	0	0	0
2	0	0	0	0	0	0	276	2	0,78	0	0	0	330	2	1	0	0	0
3	0	0	0	0	0	0	277	2	0,73	0	0	0	331	2	1	0	0	0
4	0	0	0	0	0	0	276	2	0,81	0	0	0	329	2	1	0	0	0
5	244	-1	0,71	254	-1	0,7	274	2	0,7	0	0	0	342	2	1	0	0	0
6	245	-1	0,71	0	0	0	275	2	0,67	0	0	0	344	2	1	0	0	0
7	244	2	0,7	252	-1	0,68	276	2	0,7	0	0	0	342	2	1	0	0	0
8	240	2	0,8	252	-1	0,69	276	2	0,72	0	0	0	340	2	1	0	0	0
9	0	0	0	0	0	0	274	2	0,9	0	0	0	332	2	1	0	0	0
10	242	-1	0,8	0	0	0	275	2	0,66	0	0	0	350	2	1	0	0	0
11	240	-1	0,96	0	0	0	274	2	0,61	0	0	0	348	2	1	0	0	0
12	0	0	0	257	2	0,9	272	2	0,9	0	0	0	348	2	1	0	0	0
13	243	2	0,8	256	2	0,79	274	2	0,79	0	0	0	346	2	1	0	0	0
14	0	0	0	0	0	0	284	2	0,84	0	0	0	336	2	1	0	0	0
15	0	0	0	0	0	0	286	2	0,86	0	0	0	335	2	1	0	0	0
16	0	0	0	0	0	0	285	2	0,83	0	0	0	335	2	1	0	0	0
17	0	0	0	0	0	0	286	2	0,89	0	0	0	331	2	1	0	0	0
18	242	-1	0,8	0	0	0	284	2	0,78	0	0	0	342	2	1	0	0	0
19	246	-1	0,69	0	0	0	284	2	0,75	0	0	0	341	2	1	0	0	0
20	248	-1	0,82	0	0	0	283	2	0,77	0	0	0	346	2	1	248	-1	0,99
21	0	0	0	256	-1	0,69	284	2	0,65	0	0	0	345	2	1	246	-1	0,56
22	0	0	0	0	0	0	267	2	0,92	296	-1	0,68	336	2	1	0	0	0
23	0	0	0	0	0	0	269	2	0,98	303	-1	0,84	327	2	1	259	-1	0,52
24	0	0	0	0	0	0	269	2	0,86	0	0	0	336	2	1	0	0	0
25	0	0	0	0	0	0	270	2	0,93	298	-1	0,7	329	2	1	262	-1	0,53
26	240	-1	0,7	252	2	0,78	267	2	0,8	291	-1	0,49	344	2	1	0	0	0
27	241	2	0,85	249	-1	0,82	269	2	0,81	0	0	0	347	2	1	0	0	0
28	240	2	0,92	248	-1	0,87	269	2	0,87	291	2	0,58	340	2	1	0	0	0
29	244	2	0,84	252	2	0,92	268	2	0,84	293	-1	0,62	344	2	1	0	0	0
30	244	-1	0,87	0	0	0	268	2	0,69	298	-1	0,49	350	2	1	0	0	0
31	0	0	0	0	0	0	270	2	0,95	310	-1	0,9	331	2	1	0	0	0
32	242	-1	0,76	253	2	0,86	267	2	0,77	291	-1	0,54	348	2	1	0	0	0

Continuação: Tabela 2

n ^o	Banda II									Banda I								
	P	T	Int.	P	T	Int.	P	T	Int.	P	T	Int.	P	T	Int.	P	T	Int.
33	0	0	0	260	2	0,81	269	2	0,8	302	-1	0,5	354	2	1	0	0	0
34	0	0	0	250	-1	0,79	268	2	0,69	300	-1	0,48	350	2	1	0	0	0
35	0	0	0	0	0	0	271	2	0,84	0	0	0	342	2	1	268	-1	0,62
36	0	0	0	0	0	0	272	2	0,92	0	0	0	338	2	1	262	-1	0,53
37	0	0	0	0	0	0	271	2	0,92	0	0	0	340	2	1	264	-1	0,58
38	0	0	0	258	2	0,87	269	2	0,77	0	0	0	346	2	1	267	2	0,92
39	0	0	0	254	2	0,81	270	2	0,75	0	0	0	350	2	1	267	2	0,86
40	0	0	0	257	2	0,92	270	2	0,84	0	0	0	348	2	1	267	2	0,95
41	0	0	0	256	2	0,86	270	2	0,77	0	0	0	350	2	1	268	2	0,85
42	0	0	0	255	2	0,88	272	2	0,78	0	0	0	348	2	1	265	2	0,99
43	0	0	0	256	2	0,85	271	2	0,78	0	0	0	346	2	1	266	2	0,9
44	0	0	0	254	-1	0,77	273	2	0,89	0	0	0	341	2	1	254	-1	0,79
45	0	0	0	255	-1	0,71	272	2	0,85	0	0	0	338	2	1	254	-1	0,65
46	0	0	0	0	0	0	274	2	0,91	0	0	0	339	2	1	258	-1	0,79
47	0	0	0	255	2	0,95	267	2	0,91	0	0	0	360	2	1	263	2	1
48	0	0	0	260	2	0,97	272	-1	0,88	0	0	0	354	2	1	0	0	0
49	0	0	0	259	2	1	273	2	0,87	0	0	0	354	2	1	0	0	0
50	0	0	0	255	2	0,8	268	-1	0,75	300	-1	0,56	357	2	1	0	0	0
51	0	0	0	0	0	0	282	2	0,9	296	1	0,87	332	2	1	266	2	0,46
52	0	0	0	0	0	0	281	2	0,9	294	-1	0,86	332	2	1	267	-1	0,35
53	0	0	0	0	0	0	284	2	1	0	0	0	329	2	0,75	265	-1	0,38
54	0	0	0	0	0	0	282	1	0,9	293	2	0,92	328	2	1	263	-1	0,28
55	0	0	0	254	-1	0,62	281	2	0,79	0	0	0	344	2	1	263	2	0,59
56	0	0	0	256	-1	0,66	281	2	0,8	0	0	0	344	2	1	262	2	0,61
57	0	0	0	255	2	0,76	281	2	0,81	0	0	0	346	2	1	265	-1	0,62
58	0	0	0	255	2	0,85	283	2	0,96	0	0	0	344	2	1	265	2	0,71
59	0	0	0	255	2	0,66	282	2	0,89	0	0	0	339	2	1	260	2	0,76
60	0	0	0	254	2	0,67	284	2	0,85	0	0	0	342	2	1	264	2	0,59
61	0	0	0	0	0	0	285	2	1	296	1	0,95	332	2	0,97	0	0	0
62	0	0	0	0	0	0	286	2	1	297	1	0,94	331	2	1	260	-1	0,5
63	0	0	0	269	2	1	298	2	0,68	322	-1	0,71	350	2	0,82	0	0	0
64	0	0	0	268	2	1	298	-1	0,65	320	-1	0,74	347	2	0,82	0	0	0
65	0	0	0	269	2	1	300	-1	0,6	324	-1	0,76	348	2	0,85	0	0	0

Continuação: Tabela 2

n ^o	Banda II									Banda I								
	P	T	Int.	P	T	Int.	P	T	Int.	P	T	Int.	P	T	Int.	P	T	Int.
66	256	2	1	268	-1	0,84	292	-1	0,46	0	0	0	356	2	0,88	0	0	0
67	253	2	1	266	-1	0,87	293	-1	0,53	0	0	0	353	2	0,91	0	0	0
68	253	2	1	267	2	0,88	291	-1	0,5	0	0	0	353	2	0,94	0	0	0
69	254	2	1	262	-1	0,83	0	0	0	0	0	0	354	2	0,99	0	0	0
70	256	2	1	265	-1	0,81	293	-1	0,4	0	0	0	353	2	0,86	0	0	0
71	255	2	1	268	2	0,84	299	-1	0,46	0	0	0	351	2	0,88	0	0	0
72	0	0	0	267	2	1	294	-1	0,49	0	0	0	350	2	0,74	0	0	0
73	0	0	0	267	2	1	302	-1	0,65	0	0	0	344	2	0,83	0	0	0
74	0	0	0	268	2	1	305	-1	0,69	0	0	0	346	2	0,92	0	0	0
75	252	-1	0,85	266	2	1	300	2	0,72	0	0	0	346	2	0,94	0	0	0
76	256	2	1	268	-1	0,87	292	-1	0,55	0	0	0	359	2	0,96	0	0	0
77	258	2	1	270	-1	0,77	296	-1	0,37	0	0	0	357	2	0,9	0	0	0
78	254	2	0,94	268	1	0,88	305	-1	0,67	0	0	0	361	2	1	0	0	0
79	0	0	0	274	2	1	300	2	0,82	324	2	0,92	350	1	0,72	0	0	0
80	0	0	0	275	2	1	299	2	0,85	322	2	0,83	358	1	0,51	0	0	0
81	0	0	0	273	2	1	299	2	0,86	319	2	0,88	356	1	0,51	0	0	0
82	253	2	0,89	273	2	1	294	-1	0,65	0	0	0	337	2	0,85	0	0	0
83	254	2	0,88	275	2	1	298	1	0,66	0	0	0	340	2	0,95	0	0	0
84	258	2	0,88	274	2	1	296	-1	0,64	0	0	0	349	2	0,95	0	0	0
85	0	0	0	274	2	1	305	-1	0,75	327	2	0,82	359	2	0,75	0	0	0
86	0	0	0	273	2	1	300	-1	0,58	328	2	0,68	362	2	0,65	0	0	0
87	0	0	0	273	2	1	300	-1	0,64	322	2	0,65	356	-1	0,55	0	0	0
88	0	0	0	273	2	1	300	-1	0,7	322	2	0,77	361	2	0,69	0	0	0
89	256	2	0,93	273	2	1	0	0	0	344	1	0,75	360	2	0,86	0	0	0
90	256	2	1	273	2	0,99	0	0	0	322	1	0,82	366	2	0,91	0	0	0
91	259	2	1	273	2	0,93	0	0	0	340	1	0,64	364	3	0,73	0	0	0
92	0	0	0	277	2	1	304	1	0,72	326	2	0,74	362	-1	0,67	0	0	0
93	0	0	0	275	2	1	304	-1	0,62	326	2	0,65	360	2	0,64	0	0	0
94	261	2	1	273	2	0,99	298	-1	0,6	336	1	0,73	363	2	0,84	0	0	0
95	260	2	1	272	-1	0,94	298	-1	0,44	334	1	0,65	367	2	0,81	0	0	0

Fonte: Referência E4

Legenda: P-Comprimento de onda, em nanômetros, do pico ou ombro ou inflexão; T- tipo, 2 para pico -1 para ombro e 1 para inflexão; Int.- Absorbância relativa ao pico de maior intensidade no espectro, pico este que possui absorbância igual a um.

Tabela 3 - Informações espectrais na região do ultra violeta para os 95 flavonóides da Tabela 1. Estes dados se referem aos espectros coletados para os flavonóides em metanol mais cloreto de alumínio.

n ^o	Banda II									Banda I								
	P	T	Int.	P	T	Int.	P	T	Int.	P	T	Int.	P	T	Int.	P	T	Int.
1	262	-1	0,35	282	1	0,55	294	-1	0,6	303	2	0,64	362	2	1	0	0	0
2	263	-1	0,41	0	0	0	280	1	0,66	303	2	0,77	352	2	1	0	0	0
3	264	-1	0,39	280	1	0,65	296	-1	0,76	303	2	0,77	356	2	1	0	0	0
4	264	-1	0,5	0	0	0	293	-1	0,74	301	2	0,79	356	2	1	0	0	0
5	262	2	0,66	284	2	0,7	298	1	0,69	0	0	0	370	2	1	0	0	0
6	261	2	0,68	284	2	0,68	298	1	0,62	0	0	0	370	2	1	0	0	0
7	262	2	0,72	287	2	0,81	297	1	0,78	0	0	0	370	2	1	0	0	0
8	260	2	0,65	287	2	0,72	297	1	0,67	0	0	0	370	2	1	0	0	0
9	0	0	0	284	2	0,7	298	-1	0,6	0	0	0	358	2	1	0	0	0
10	258	-1	0,62	282	2	0,67	0	0	0	310	-1	0,57	382	2	1	0	0	0
11	256	-1	0,62	283	2	0,62	0	0	0	310	-1	0,37	386	2	1	0	0	0
12	0	0	0	274	2	0,78	0	0	0	306	-1	0,39	338	2	0,23	428	2	1
13	0	0	0	275	2	0,7	0	0	0	302	-1	0,63	338	2	0,52	428	2	1
14	262	2	0,12	292	1	0,64	304	2	0,86	0	0	0	375	2	1	0	0	0
15	0	0	0	276	-1	0,5	302	2	1	0	0	0	380	2	0,9	0	0	0
16	266	-1	0,26	292	1	0,66	303	2	0,78	0	0	0	367	2	1	0	0	0
17	265	-1	0,29	292	1	0,73	303	2	0,86	0	0	0	362	2	1	0	0	0
18	255	-1	0,49	296	2	0,64	315	-1	0,57	0	0	0	380	2	1	0	0	0
19	261	2	0,59	297	2	0,77	0	0	0	0	0	0	373	2	1	0	0	0
20	270	2	0,78	0	0	0	302	2	0,63	0	0	0	0	0	0	420	2	1
21	274	2	0,55	0	0	0	302	2	0,61	343	-1	0,32	0	0	0	425	2	1
22	0	0	0	276	2	0,81	301	2	0,78	348	2	1	384	2	0,85	0	0	0
23	277	2	0,85	292	-1	0,78	302	2	0,87	344	2	1	382	2	0,71	0	0	0
24	277	2	0,87	294	-1	0,6	303	2	0,69	348	2	1	384	2	0,85	0	0	0
25	280	2	0,89	294	-1	0,77	303	2	0,83	344	2	1	382	2	0,64	0	0	0
26	267	2	0,87	273	2	0,83	296	2	0,6	362	2	0,95	390	2	1	0	0	0
27	262	2	0,85	274	2	0,85	296	2	0,53	366	-1	0,91	390	2	1	0	0	0
28	261	2	0,91	276	2	0,92	295	2	0,59	359	2	1	387	2	0,97	0	0	0
29	266	2	0,96	275	2	1	296	2	0,7	360	2	0,93	386	2	0,93	0	0	0
30	256	-1	0,75	277	2	0,82	307	2	0,82	365	-1	0,9	393	2	1	0	0	0
31	253	-1	0,66	278	2	0,97	300	2	0,76	348	2	1	385	-1	0,72	0	0	0

Continuação: Tabela 3

n ^o	Banda II									Banda I								
	P	T	Int.	P	T	Int.	P	T	Int.	P	T	Int.	P	T	Int.	P	T	Int.
32	0	0	0	272	2	0,81	300	-1	0,32	330	-1	0,21	0	0	0	423	2	1
33	0	0	0	272	2	0,75	310	2	0,3	0	0	0	0	0	0	424	2	1
34	0	0	0	270	2	0,71	310	-1	0,26	344	-1	0,18	0	0	0	430	2	1
35	280	2	0,79	298	-1	0,63	305	-1	0,62	367	2	1	404	-1	0,78	0	0	0
36	282	2	0,75	297	-1	0,67	305	-1	0,65	360	2	1	404	-1	0,54	0	0	0
37	282	2	0,71	0	0	0	302	-1	0,57	367	2	1	404	-1	0,65	0	0	0
38	279	-1	0,83	0	0	0	300	-1	0,51	375	2	1	406	1	0,88	0	0	0
39	279	-1	0,78	0	0	0	302	-1	0,47	380	2	1	408	1	0,88	0	0	0
40	280	-1	0,88	0	0	0	298	-1	0,68	380	2	1	406	1	0,84	0	0	0
41	280	-1	0,85	0	0	0	300	-1	0,58	386	2	1	408	1	0,87	0	0	0
42	280	-1	0,89	0	0	0	302	-1	0,56	376	2	1	408	1	0,75	0	0	0
43	280	-1	0,84	0	0	0	299	-1	0,55	377	2	1	408	1	0,75	0	0	0
44	278	2	0,93	0	0	0	311	-1	0,61	362	2	1	407	-1	0,61	0	0	0
45	261	-1	0,69	282	2	0,87	309	-1	0,54	363	2	1	406	-1	0,59	0	0	0
46	282	2	0,9	0	0	0	310	2	0,68	362	2	1	406	-1	0,7	0	0	0
47	275	2	0,99	0	0	0	310	2	0,34	360	2	0,7	420	2	0,98	0	0	0
48	278	2	0,92	0	0	0	306	-1	0,49	342	-1	0,39	0	0	0	440	2	1
49	274	2	1	0	0	0	306	2	0,38	342	-1	0,34	0	0	0	434	2	0,89
50	275	2	0,93	0	0	0	316	-1	0,41	0	0	0	0	0	0	446	2	1
51	290	-1	0,71	0	0	0	310	2	0,92	360	2	1	412	1	0,36	0	0	0
52	288	2	0,45	0	0	0	311	2	0,68	361	2	1	407	-1	0,32	0	0	0
53	290	-1	0,76	0	0	0	310	2	1	356	2	0,97	413	-1	0,32	0	0	0
54	287	-1	0,55	0	0	0	311	2	0,78	354	2	1	411	-1	0,25	0	0	0
55	284	2	0,71	0	0	0	304	-1	0,65	368	2	1	412	-1	0,53	0	0	0
56	290	2	0,73	0	0	0	303	1	0,69	372	2	1	412	-1	0,41	0	0	0
57	285	2	0,72	0	0	0	304	-1	0,6	374	2	1	418	-1	0,48	0	0	0
58	290	2	0,81	0	0	0	304	2	0,8	370	2	1	420	-1	0,36	0	0	0
59	290	2	0,78	0	0	0	302	1	0,75	362	2	1	415	-1	0,42	0	0	0
60	293	2	0,71	0	0	0	302	-1	0,7	362	2	1	420	-1	0,34	0	0	0
61	291	2	0,91	0	0	0	313	2	0,93	358	2	1	420	-1	0,49	0	0	0
62	292	2	0,7	0	0	0	312	2	0,79	357	2	1	420	-1	0,28	0	0	0
63	277	2	1	0	0	0	304	2	0,7	350	2	0,74	399	2	0,67	0	0	0
64	278	2	1	296	-1	0,65	304	2	0,72	350	2	0,92	400	2	0,77	0	0	0

Continuação: Tabela 3

n ^o	Banda II									Banda I								
	P	T	Int.	P	T	Int.	P	T	Int.	P	T	Int.	P	T	Int.	P	T	Int.
65	276	2	1	298	-1	0,57	304	2	0,61	348	2	0,87	399	2	0,73	0	0	0
66	267	2	1	275	1	0,95	300	-1	0,44	364	2	0,72	404	2	0,77	0	0	0
67	268	2	1	278	1	0,95	302	2	0,59	366	2	0,69	406	2	0,77	0	0	0
68	266	2	1	276	-1	0,96	300	2	0,47	362	2	0,75	402	2	0,81	0	0	0
69	269	2	1	280	2	0,87	300	2	0,45	368	2	0,75	404	2	0,92	0	0	0
70	270	2	1	275	-1	0,9	300	2	0,45	366	2	0,68	402	2	0,79	0	0	0
71	271	2	1	277	-1	0,95	299	-1	0,44	364	2	0,72	402	2	0,77	0	0	0
72	0	0	0	275	2	1	304	2	0,48	349	2	0,64	399	2	0,57	0	0	0
73	0	0	0	275	2	1	308	2	0,52	349	2	0,69	399	2	0,65	0	0	0
74	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
75	252	-1	0,71	276	2	1	306	2	0,63	355	2	0,88	399	2	0,72	0	0	0
76	0	0	0	277	2	0,93	303	-1	0,37	339	2	0,32	0	0	0	440	2	1
77	0	0	0	276	2	1	302	-1	0,28	320	-1	0,2	0	0	0	440	2	0,96
78	0	0	0	271	2	1	312	2	0,28	0	0	0	0	0	0	428	2	1
79	0	0	0	282	2	0,86	310	2	0,87	349	2	1	398	2	0,54	0	0	0
80	0	0	0	284	2	0,84	310	2	0,97	346	2	1	398	2	0,43	0	0	0
81	272	1	0,71	283	2	0,83	310	2	0,93	343	2	1	404	2	0,37	0	0	0
82	262	-1	0,84	282	2	1	304	-1	0,67	362	2	1	401	2	0,72	0	0	0
83	263	-1	0,8	282	2	0,98	302	2	0,72	357	2	1	406	2	0,62	0	0	0
84	0	0	0	277	2	0,82	306	-1	0,31	336	-1	0,14	0	0	0	433	2	1
85	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
86	0	0	0	282	2	1	312	2	0,64	354	2	0,9	423	2	0,56	0	0	0
87	0	0	0	283	2	1	312	2	0,76	350	2	0,84	416	2	0,49	0	0	0
88	0	0	0	284	2	1	313	2	0,8	350	2	0,98	422	2	0,64	0	0	0
89	266	-1	0,9	283	2	1	308	2	0,64	362	2	0,84	418	2	0,67	0	0	0
90	272	1	0,97	282	2	1	306	2	0,57	366	2	0,85	424	2	0,77	0	0	0
91	274	1	0,97	282	2	1	307	2	0,51	365	2	0,7	424	2	0,61	0	0	0
92	0	0	0	286	2	1	314	2	0,71	354	2	0,82	420	2	0,59	0	0	0
93	0	0	0	282	2	1	312	2	0,57	348	2	0,7	424	2	0,49	0	0	0
94	0	0	0	281	2	1	312	-1	0,39	342	-1	0,35	0	0	0	451	2	0,92
95	0	0	0	281	2	1	312	-1	0,33	344	-1	0,24	0	0	0	454	2	0,98

Fonte: Referência E4. (seguir legenda Tabela 2)

Tabela 4 - Informações espectrais na região do ultra violeta para os 95 flavonóides da Tabela 1. Estes dados se referem aos espectros coletados para os flavonóides em metanol mais cloreto de alumínio mais ácido clorídrico.

n ^o	Banda II									Banda I								
	P	T	Int.	P	T	Int.	P	T	Int.	P	T	Int.	P	T	Int.	P	T	Int.
1	0	0	0	260	-1	0,31	0	0	0	287	1	0,65	303	2	0,74	356	2	1
2	0	0	0	262	-1	0,42	0	0	0	293	-1	0,85	301	2	0,86	350	2	1
3	0	0	0	263	-1	0,4	0	0	0	295	-1	0,85	303	2	0,86	353	2	1
4	0	0	0	262	-1	0,54	0	0	0	292	1	0,82	301	2	0,87	348	2	1
5	0	0	0	260	2	0,66	286	2	0,77	296	1	0,76	0	0	0	364	2	1
6	0	0	0	258	2	0,69	286	2	0,73	298	1	0,69	0	0	0	366	2	1
7	0	0	0	261	2	0,63	284	2	0,74	292	2	0,75	0	0	0	363	2	1
8	0	0	0	258	2	0,66	290	2	0,76	0	0	0	0	0	0	362	2	1
9	0	0	0	0	0	0	289	2	0,8	0	0	0	0	0	0	353	2	1
10	0	0	0	258	-1	0,67	283	2	0,72	0	0	0	308	-1	0,67	372	2	1
11	0	0	0	256	-1	0,64	286	2	0,64	0	0	0	308	2	0,44	374	2	1
12	0	0	0	262	2	0,78	280	2	0,82	298	2	0,63	0	0	0	367	2	1
13	0	0	0	262	-1	0,8	286	2	0,85	297	1	0,84	0	0	0	372	2	1
14	0	0	0	262	-1	0,16	292	1	0,68	302	2	0,84	0	0	0	364	2	1
15	0	0	0	262	-1	0,28	292	1	0,86	302	2	0,94	0	0	0	358	2	1
16	0	0	0	266	-1	0,29	290	1	0,69	302	2	0,82	0	0	0	362	2	1
17	0	0	0	264	-1	0,32	292	1	0,78	302	2	0,89	0	0	0	358	2	1
18	0	0	0	257	-1	0,71	294	2	0,69	0	0	0	0	0	0	365	2	1
19	0	0	0	259	2	0,61	296	2	0,81	0	0	0	0	0	0	367	2	1
20	0	0	0	258	2	0,76	296	2	0,76	0	0	0	0	0	0	372	2	1
21	0	0	0	260	2	0,47	296	2	0,7	0	0	0	0	0	0	372	2	1
22	0	0	0	0	0	0	276	2	0,88	299	2	0,85	340	2	1	381	2	0,83
23	260	-1	0,63	279	2	0,92	294	-1	0,9	300	2	0,93	338	2	1	379	2	0,64
24	0	0	0	276	2	0,89	294	-1	0,74	301	2	0,79	343	2	1	382	2	0,75
25	264	-1	0,54	278	2	0,85	294	-1	0,71	304	2	0,8	342	2	1	384	2	0,69
26	0	0	0	264	-1	0,84	276	2	0,85	295	2	0,66	351	2	1	383	2	0,92
27	0	0	0	259	2	0,95	276	2	0,96	294	2	0,67	353	2	1	386	2	0,94
28	0	0	0	259	2	0,77	279	2	0,9	293	-1	0,69	348	2	1	381	-1	0,72
29	0	0	0	262	2	0,96	276	2	1	296	-1	0,76	355	2	0,96	384	2	0,88
30	0	0	0	257	2	0,79	278	2	0,9	306	2	0,54	362	2	1	390	2	0,98
31	0	0	0	0	0	0	280	2	1	298	-1	0,75	340	2	0,96	382	-1	0,47

Continuação: Tabela 4.

n ^o	Banda II									Banda I								
	P	T	Int.	P	T	Int.	P	T	Int.	P	T	Int.	P	T	Int.	P	T	Int.
32	0	0	0	264	-1	0,95	276	2	1	296	-1	0,66	358	2	1	385	2	0,98
33	0	0	0	0	0	0	277	2	0,94	306	2	0,52	364	-1	0,89	393	2	1
34	0	0	0	0	0	0	275	2	0,88	303	2	0,51	363	2	0,95	386	2	1
35	262	-1	0,63	282	2	0,86	0	0	0	305	-1	0,68	364	2	1	404	-1	0,67
36	262	2	0,53	282	2	0,8	296	-1	0,73	306	-1	0,67	360	2	1	404	-1	0,51
37	262	-1	0,58	282	2	0,78	0	0	0	302	-1	0,63	361	2	1	402	-1	0,55
38	264	2	0,88	280	-1	0,83	0	0	0	300	-1	0,56	364	2	1	400	-1	0,75
39	264	2	0,83	280	2	0,79	0	0	0	301	-1	0,49	372	2	1	408	-1	0,78
40	266	2	0,94	282	2	0,92	0	0	0	300	-1	0,71	370	2	1	406	-1	0,72
41	265	2	0,81	281	2	0,86	0	0	0	300	-1	0,52	382	2	1	406	-1	0,8
42	260	2	0,96	281	2	0,91	0	0	0	302	1	0,6	364	2	1	406	-1	0,58
43	263	2	0,85	283	2	0,85	0	0	0	300	1	0,63	366	2	1	406	-1	0,58
44	259	-1	0,75	283	2	0,96	0	0	0	308	-1	0,61	360	2	1	406	-1	0,53
45	255	-1	0,71	285	2	0,97	0	0	0	308	-1	0,69	357	2	1	406	-1	0,51
46	260	-1	0,78	281	2	0,93	0	0	0	308	-1	0,76	360	2	1	406	-1	0,78
47	263	2	1	275	2	0,99	0	0	0	310	2	0,35	367	2	0,74	420	2	0,97
48	269	2	0,98	282	1	0,93	0	0	0	302	-1	0,71	372	2	1	410	1	0,84
49	269	2	1	286	1	0,87	0	0	0	300	1	0,53	373	2	0,91	408	1	0,68
50	262	1	0,81	275	2	0,87	0	0	0	310	-1	0,62	375	2	1	414	1	0,91
51	266	-1	0,44	290	-1	0,77	0	0	0	310	2	0,98	356	2	1	412	-1	0,27
52	265	-1	0,31	290	2	0,55	0	0	0	311	2	0,72	354	2	1	408	-1	0,25
53	262	-1	0,37	289	-1	0,75	0	0	0	309	2	1	351	2	0,93	404	-1	0,25
54	264	-1	0,29	286	2	0,62	0	0	0	311	2	0,89	350	2	1	408	-1	0,23
55	261	2	0,61	292	2	0,75	0	0	0	304	1	0,71	363	2	1	414	-1	0,42
56	260	2	0,64	292	2	0,77	0	0	0	302	1	0,81	366	2	1	412	-1	0,41
57	264	-1	0,63	288	2	0,72	0	0	0	304	-1	0,66	366	2	1	418	-1	0,33
58	262	2	0,63	290	2	0,82	0	0	0	303	2	0,83	363	2	1	420	-1	0,32
59	260	2	0,62	293	2	0,81	0	0	0	0	0	0	361	2	1	412	-1	0,41
60	261	2	0,61	294	-1	0,74	0	0	0	303	2	0,75	362	2	1	420	-1	0,31
61	0	0	0	294	-1	0,87	0	0	0	313	2	0,98	352	2	1	420	-1	0,45
62	0	0	0	290	-1	0,83	0	0	0	312	2	0,92	356	2	1	420	-1	0,25
63	0	0	0	278	2	1	0	0	0	304	2	0,69	348	2	0,75	398	2	0,64
64	0	0	0	278	2	1	297	2	0,71	304	2	0,72	345	2	0,9	399	2	0,7

Continuação: Tabela 4.

n ^o	Banda II									Banda I								
	P	T	Int.	P	T	Int.	P	T	Int.	P	T	Int.	P	T	Int.	P	T	Int.
65	0	0	0	277	2	1	296	-1	0,61	304	2	0,63	346	2	0,85	398	2	0,67
66	0	0	0	264	2	1	276	2	0,98	298	-1	0,5	359	2	0,75	402	2	0,74
67	0	0	0	268	2	1	277	-1	0,98	300	2	0,57	360	2	0,74	402	2	0,74
68	0	0	0	265	1	0,99	277	2	1	299	-1	0,51	355	2	0,81	401	2	0,73
69	0	0	0	268	2	1	278	2	0,94	300	2	0,48	362	2	0,74	402	2	0,79
70	0	0	0	268	2	1	274	-1	0,96	298	-1	0,43	360	2	0,64	401	2	0,7
71	0	0	0	270	2	1	276	2	0,98	298	-1	0,5	356	2	0,7	401	2	0,74
72	0	0	0	0	0	0	275	2	1	302	2	0,45	346	2	0,66	398	2	0,5
73	0	0	0	0	0	0	276	2	1	306	2	0,52	344	2	0,66	398	2	0,55
74	0	0	0	256	-1	0,57	280	2	1	305	-1	0,66	343	2	0,83	399	2	0,57
75	0	0	0	253	-1	0,66	278	2	1	304	2	0,62	345	2	0,79	398	2	0,6
76	0	0	0	268	2	1	277	1	0,94	303	2	0,57	360	2	0,75	402	2	0,85
77	0	0	0	272	2	1	278	1	0,96	302	-1	0,36	368	2	0,6	402	2	0,76
78	0	0	0	0	0	0	273	2	1	308	2	0,43	360	2	0,7	405	2	0,84
79	0	0	0	0	0	0	282	2	0,89	309	2	0,98	344	2	1	397	2	0,49
80	0	0	0	0	0	0	285	2	0,88	309	2	1	342	2	0,94	398	2	0,4
81	0	0	0	272	-1	0,79	285	2	0,96	311	2	1	342	2	0,98	404	2	0,35
82	0	0	0	262	-1	0,82	284	-1	0,98	304	-1	0,71	356	2	1	400	2	0,61
83	0	0	0	262	-1	0,77	283	2	0,98	302	2	0,75	353	2	1	406	2	0,57
84	0	0	0	264	-1	0,82	280	2	1	301	2	0,71	358	2	0,98	400	2	0,66
85	0	0	0	0	0	0	284	2	1	313	2	0,86	351	2	0,95	415	2	0,65
86	0	0	0	0	0	0	283	2	1	311	2	0,69	350	2	0,89	420	2	0,52
87	0	0	0	0	0	0	284	2	1	312	2	0,77	348	2	0,83	418	2	0,47
88	0	0	0	0	0	0	283	2	1	312	2	0,82	346	2	0,94	422	2	0,61
89	0	0	0	262	1	0,86	282	2	1	306	2	0,55	360	2	0,86	418	2	0,65
90	0	0	0	268	1	0,92	282	2	1	306	2	0,55	361	2	0,88	422	2	0,75
91	0	0	0	274	-1	0,97	282	2	1	307	2	0,53	361	2	0,69	423	2	0,51
92	0	0	0	0	0	0	286	2	1	313	2	0,77	350	2	0,83	420	2	0,59
93	0	0	0	0	0	0	282	2	1	312	2	0,59	348	2	0,7	424	2	0,47
94	0	0	0	272	1	0,99	282	2	1	306	-1	0,62	361	2	0,78	418	2	0,69
95	0	0	0	273	1	0,99	282	2	1	306	2	0,5	362	2	0,77	422	2	0,67

Fonte: Referência E4.(seguir legenda Tabela 2)

Tabela 5- Resultados para o teste de validação cruzada para o SIMCA.

n ^o	Classe real	Classe prevista	Classe mais próxima	Q ²	T ²
1	1	1	0	0	0
2	1	0	1	1,2629	0,1774
3	1	1	0	0	0
4	1	1	0	0	0
5	1	1	0	0	0
6	1	1	0	0	0
7	1	1	0	0	0
8	1	0	1	1,2415	0,1024
9*	1	0	3	183,0914	3,9995
10	1	1	0	0	0
11	1	1	0	0	0
12	1	0	1	2,048	0,5212
13	1	0	1	1,6595	0,3279
14	2	2	0	0	0
15*	2	0	1	8,1866	0,1266
16	2	2	0	0	0
17	2	2	0	0	0
18	2	2	0	0	0
19*	2	0	1	4,7701	0,0842
20	2	0	2	2,4849	0,1954
21*	2	0	3	220,0743	3,014
22	3	3	0	0	0
23	3	3	0	0	0
24	3	0	3	1,3387	0,1085
25	3	3	0	0	0
26	3	3	0	0	0
27	3	3	0	0	0
28	3	0	3	1,8497	0,1818
29	3	3	0	0	0
30	3	0	3	1,1056	0,0185
31	3	0	3	1,7659	0,0513
32	3	0	3	1,1601	0,1282
33*	3	0	6	3,8146	0,3046
34	3	3	0	0	0
35	4	4	0	0	0
36*	4	0	3	3,7806	0,1488
37	4	4	0	0	0
38	4	0	4	1,7875	0,2078
39	4	4	0	0	0
40	4	4	0	0	0
41	4	4	0	0	0
42	4	4	0	0	0
43	4	4	0	0	0
44	4	4	0	0	0
45*	4	0	4	3,719	0,0587

Continuação da Tabela 5

n ^o	Classe real	Classe prevista	Classe mais próxima	Q ²	T ²
46	4	4	0	0	0
47*	4	0	4	5,6166	0,2561
48	4	4	0	0	0
49	4	0	4	1,5092	0,189
50*	4	0	3	67,9479	1,1176
51	5	5	0	0	0
52	5	0	5	1,1337	0,1792
53*	5	0	4	3,1778	0,3847
54	5	5	0	0	0
55	5	5	0	0	0
56	5	5	0	0	0
57	5	0	5	1,1367	0,116
58	5	5	0	0	0
59*	5	0	4	134,9967	14,5149
60	5	5	0	0	0
61*	5	0	3	78,9317	0,6968
62	5	5	0	0	0
63*	6	0	5	35,989	5,1084
64	6	6	0	0	0
65	6	6	0	0	0
66	6	6	0	0	0
67	6	6	0	0	0
68	6	6	0	0	0
69*	6	0	8	53,8017	0,037
70	6	6	0	0	0
71	6	6	0	0	0
72	6	0	6	1,0451	0,2342
73	6	6	0	0	0
74*	6	0	8	50,0781	0,8641
75	6	0	6	1,6173	0,075
76	6	6	0	0	0
77	6	0	6	1,2588	0,293
78	6	0	6	1,8339	0,3371
79	7	7	0	0	0
80	7	7	0	0	0
81	7	7	0	0	0
82	7	7	0	0	0
83	7	7	0	0	0
84*	7	0	6	1,7431	0,1946
85*	8	0	6	5,6989	0,5676
86	8	8	0	0	0
87	8	8	0	0	0
88	8	0	0	0	0
89	8	8	0	0	0
90	8	8	0	0	0
91	8	0	8	1,0375	0,1628
92	8	0	8	1,2475	0,1402
93	8	8	0	0	0
94	8	8	0	0	0

95	8	0	8	1,108	0,3562
----	---	---	---	-------	--------

Tabela 6- Resultados para o teste de validação cruzada para o SIMCA sem amostras consideradas "outliers" .

n ^o	Classe real	Classe prevista	Classe mais próxima	Q ²	T ²
1	1	1	0	0	0
2	1	0	1	1,1212	0,1999
3	1	1	0	0	0
4	1	1	0	0	0
5	1	1	0	0	0
6	1	1	0	0	0
7	1	1	0	0	0
8	1	0	1	1,2846	0,1256
10	1	1	0	0	0
11	1	1	0	0	0
12	1	0	1	1,9359	0,6594
13	1	0	1	1,4243	0,3943
14	2	2	0	0	0
16	2	2	0	0	0
17	2	2	0	0	0
18*	2	0	1	4,2595	0,0881
20*	2	0	3	228,1625	0,2318
22	3	3	0	0	0
23	3	3	0	0	0
24	3	3	0	0	0
25	3	3	0	0	0
26	3	3	0	0	0
27	3	3	0	0	0
28	3	0	3	1,2614	0,2048
29	3	3	0	0	0
30	3	3	0	0	0
31	3	0	3	1,5337	0,0041
32	3	3	0	0	0
34	3	3	0	0	0
35*	4	0	3	29,3949	0,0433
37	4	4	0	0	0
38	4	0	4	2,1526	0,0816
39	4	4	0	0	0
40	4	4	0	0	0
41	4	4	0	0	0
42	4	4	0	0	0
43	4	4	0	0	0
44	4	4	0	0	0
46	4	0	4	1,0387	0,2739
48	4	0	4	1,4655	0,5497
49	4	0	4	1,8795	0,8278
51	5	0	5	1,0303	0,0791
52	5	0	5	1,191	0,2047
54	5	5	0	0	0
55	5	5	0	0	0

Continuação da Tabela 6

n ^o	Classe real	Classe prevista	Classe mais próxima	Q ²	T ²
56	5	5	0	0	0
57	5	0	5	1,5482	0,1665
58	5	5	0	0	0
60	5	5	0	0	0
62*	5	0	3	56,2973	0,4044
64	6	6	0	0	0
65	6	0	6	1,3679	0,3586
66	6	6	0	0	0
67	6	6	0	0	0
68	6	6	0	0	0
70	6	6	0	0	0
71	6	6	0	0	0
72	6	6	0	0	0
73	6	6	0	0	0
75	6	6	0	0	0
76	6	6	0	0	0
77	6	6	0	0	0
78	6	0	6	12,6797	1,4744
79	7	7	0	0	0
80	7	7	0	0	0
81	7	7	0	0	0
82*	7	0	6	1,3343	0,0362
83	7	7	0	0	0
86	8	8	0	0	0
87	8	8	0	0	0
88	8	8	0	0	0
89	8	0	8	1,1397	0,1001
90	8	8	0	0	0
91	8	0	8	1,7752	0,1203
92	8	0	8	2,2371	0,167
93	8	8	0	0	0
94	8	8	0	0	0
95	8	8	0	0	0

Tabela 7- Resultados para o teste com amostras cujas substituições não pertencem às classes usadas na construção do modelo SIMCA.

	OH	MeO	Me	classe prevista	Q ²	T ²
1	5,6,7,4'	3,3'	-	3	110,292	0,0394
2	5,6,4'	3,7,3'	-	3	150,029	1,0114
3	5,6,3'	3,7,4',5'	-	3	150,763	1,0923
4	5,6,3',4'	3,7	-	1	3,2214	0,2857
5	5,6,4'	7,8,3'	-	5	26,9352	0,939
6	5,7,4'	-	8	3	29,0336	0,0053
7	5,4'	7	6	3	1,4098	0,0192
8	5	7,4'	6	3	0	0
9	5,7,4'	3	6	4	1,2387	0,4551
10	5,7,3',5'	3,4'	6	5	6,1358	0,2116
11	5,7,3',4',5'	3	6	3	47,8835	1,3479
12	5,7,4'	3	6,8	5	2,9912	0,0201
13	5,7,3',5'	3,4'	6,8	3	29,8247	0,0562
14	5,7	3,6,8,4'	-	5	1,1084	0,0005
15	5	3,6,7,8,3',4'	-	5	1,0718	0,2659
16	5,3'	3,6,7,8,4',5'	-	1	25,6877	0,2487
17	5,4'	7	6,8	3	50,0997	0,6552
18	5	7,4'	6,8	3	51,9321	0,5224
19	5,7,8,4'	-	-	3	25,0164	0,5681
20	5,7,8,3',4'	-	-	3	29,1682	0,292
21	5,8,4'	7	-	8	2,5514	0,2456
22	5,7,3',4'	3	8	8	1,0124	0,2897
23	5,8,4'	3,7	-	8	3,1658	0,1551
24	5,8,3',4'	3,7	-	8	6,0724	0,0363

REFERÊNCIAS BIBLIOGRÁFICAS

A

ÁLGEBRA

A1- Boldrini, J. L.; Costa, S.I.R.; Figueiredo, V. L.; Wetzler H. G.; *Álgebra Linear*, **1986**, ed. HARBRA Ltda. 3ª ed.

A2- Noble, B.; Daniel, J. W.; Carvalho, J. P. Carvalho; *Álgebra Linear Aplicada*; **1986**, ed. Prentice-Hall do Brasil Ltda.

A3- Strang, G.; *Linear Algebra and Its Application*, **1976**, ed. Harcourt Brace Jovanovich, Publisher.

A4- Golub, G. H.; Van Loan, C.F.; *Matrix Computations*, **1989**, ed. The Johns Hopkins University Press.

B

SEPARAÇÃO DE ESPECTROS

B1- Malinowski, E. D.; *Factor Analysis in Chemistry*; **1991**, 2ª ed., ed. John Wiley & Sons, Inc.

B2- Pelicán, P.; Ceppan, M.; Liska, M.; *Applications of Numerical Methods in Molecular Spectroscopy*; **1994**, ed. CRC Press, Inc.

B3- Lawton, W. H.; Sylvestre, E. A.; "Self Modeling Curve Resolution", *Technometrics*, **1971**, 13, 617.

B4- Sylvestre, E. A.; Lawton, W. H.; Maggio, M. S.; " *Curve Resolution Using a Postulated Chemical Reaction* ", *Technometrics*, **1974**, 16, 353.

B5- Hamilton, J. G.; Gemperline, P. J.; " *Mixture Analysis Using Factor Analysis II : Self-Modeling Curve Resolution* ", *J. Chemometrics*, **1990**, 4, 1.

B6-Ho, C. N.; Christian, G. D.; Davidson, E. R.; " *Application of the Method of Rank Annihilation to Quantitative Analysis of Multicomponent Fluorescence Data from Video Fluorimeter*", *Anal. Chem.*, **1978**, 50, 1108.

B7- Lorber, A.; " *Quantifying Chemical Composition from Two-Dimensional Data Arrays* ", *Anal. Chim. Acta*, **1984**, 164, 293.

B8- Lorber, A.; " *Features of Quantifying Chemical Composition from Two-Dimensional Data Arrays by the Rank Annihilation Factor Analysis Method*", *Anal. Chem.*, **1985**, 57, 2395.

C

DADOS DE ORDEM SUPERIOR

C1- Sanchez, E.; Kowalski, B.R.; " *Tensorial Calibration : I. First-Order Calibration* ", *J. Chemometrics*, **1988**, 2, 247.

C2- Sanchez, E.; Kowalski, B.R. ; " *Tensorial Calibration : II. Second-Order Calibration* ", *J. Chemometrics*, **1988**, 2, 265.

C3- Wilson, B. E.; Sanchez, E.; Kowalski, B.R. ; " *An Improved Algorithm for The Generalized Rank Annihilation Method* ", *J. Chemometrics*, **1989**, 3, 493.

C4-Sanchez, E.; Kowalski, B.R.; " *Tensorial Resolution : A Direct Trilinear Decomposition* ", *J. Chemometrics*, **1990**, 4, 29.

C5-Tauler, R.; Smilde, A.; Kowalski, B.; " *Selectivity, Local Rank, Three-Way Data Analysis and Ambiguity in Multivariate Curve Resolution*", *J. Chemometrics*, **1995**, 9, 31.

C6- Xie, Y. L.; Baeza-Baeza, J. J. ; Ramis-Ramos, G.; " *Second-Order Tensorial Calibration for Kinetic Spectrophotometric Determination*", *Chem. Intell. Lab. Sys.*, **1996**, 32, 215.

C7-Ferreira, M.M.C.; Brandes, M.L.; Ferreira I.M.C.; Kowalski, B. R.; " *Chemometric Study of the Fluorescence of Dental Calculus by Trilinear Decomposition* ", *A. Spectroscopy*, **1995**, 49, 1317.

C8- Geladi, P.; " *Analysis of Multi-Way (Multi-Mode) Data*, *Chem.Intell.Lab. Sys*, **1989**, 7, 11.

D

MÉTODOS DE CLASSIFICAÇÃO E ANÁLISE EXPLORATÓRIA

D1- Sharaf, M.A.; Illman, D. L.; Kowalski, B. R.; *Chemometrics*, **1986**, ed. John Wiley & Sons, Inc.

D2- Massart, D. L.; Vandeginste, B. G. M.; Deming, S.N.; Michotte, Y.; Kaufman, L; *Chemometrics : A Textbook*, **1988**, ed Elsevier.

D3- Mellinger, M.; " *Multivariate Data Analysis: Its Methods*", *Chem.Intell.Lab. Sys*, **1987**, 2, 29.

D4- Wold, S.; Esbensen, K.; Geladi, P. ; " *Principal Component Analysis*", *Chem.Intell.Lab. Sys*, **1987**, 2, 37.

D5- Kowalski, B. R.; Bender, C. F.; " *A Powerful Approach to Interpreting Chemical Data*", *J. Am. Chem. Soc.* , **1972**, 9, 5632.

D6- Bratchell, N.; " *Cluster Analysis* ", *Chem.Intell.Lab. Sys*, **1989**, 6, 105.

E

FLAVONÓIDES E PORFIRINAS

E1- Geissman, T. A.; *The Chemistry of Flavonoid Compounds*; **1962**, ed. Pergamon Press, Inc.

E2- Harborne, J. B.; Marby, T. J.; Marby, H.; *The Flavonoids*; **1975**, ed. Academic Press, Inc.

E3- Falk, J. E.; *Porphyrins and Metalloporphyrins*, **1964**, ed. Elsevier.

E4- Viorin, B.; " *UV Spectral Differentiation of 5-Hydroxy- and 5-Hidroxy-3-Methoxyflavones with mono-(4'), di-(3',4') or tri-(3',4',5')-Substituted B Rings*", *Phytochemistry*, **1983**, 22, 2107.

F

QUÍMICA ANALÍTICA

F1- Christian, G.D.; *Anal. Chem.* , **1986**, ed. John Wiley & Sons, Inc.

F2-Jones, D.G.; " *Photodiode Array Detectors in UV-Vis Spectroscopy: Part II*", *Anal. chem.*, **1985**, 57, 1207A.

F3- Hirschfeld,T.; " *The Hy-phen-ated Methods*", *Anal. chem.*, **1980**, 52, 297A.