

**UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE QUÍMICA**



UNICAMP

Dissertação de Mestrado

***"MODELAGEM DE DADOS CINÉTICOS ESPECTROFOTOMÉTRICOS
POR MÉTODOS QUIMIOMÉTRICOS DE ORDEM SUPERIOR"***

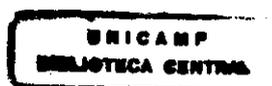
Jacqueline M. Dueñas Cueva

Orientador: Dr. Ronei Jesus Poppi

Co-orientadora: Dr. Adriana Vitorino Rossi

Julho 2000

R1 CTN T. 0000



UNIDADE	I.Q.
N.º CHAMADA:	T/Unicomel
	D.86.8.m
V.	Ex.
TOMBO BC/	43540
PROC.	16-392101
C	<input type="checkbox"/>
D	<input checked="" type="checkbox"/>
PREC.º	R\$11,00
DATA	31/01/01
N.º CPD	

**CHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DO INSTITUTO DE QUÍMICA**

CM-00153693-1

D778m

Dueñas Cueva, Jacqueline M.

Modelagem de dados cinéticos espectrofotométricos por métodos quimiométricos de ordem superior / Jacqueline M. Dueñas Cueva. -- Campinas, SP: [s.n], 2000.

Orientador : Ronei Jesus Poppi.

Co-orientadora: Adriana Vitorino Rossi.

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

1. PARAFAC. 2. N-PLS. 3. Misturas de isômeros. 4. Aminofenol. I. Poppi, Ronei Jesus. II. Rossi, Adriana Vitorino. III. Universidade Estadual de Campinas. Instituto de Química. IV. Título.

*Aos meus pais Jorge e Margarita,
aos meus irmãos,
e ao amor da minha vida.*

AGRADECIMENTOS

- Ao professor Ronei Jesus Poppi e à professora Adriana Vitorino Rossi, pela orientação, o valioso apoio e amizade.
- À Universidade Estadual de Campinas e, em especial, ao Instituto de Química pela oportunidade de estudo oferecido.
- Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico CNPq pelo apoio financeiro.
- A Rasmus Bro do Instituto Food Technology da Dinamarca por disponibilizar os programas em MATLAB utilizados nesta tese.
- Aos colegas do grupo de Quimiometria, que contribuíram no desenvolvimento deste trabalho: Cleidiane, Eduardo, Marcelo, Paulo Augusto, Paulo Fidencio e Silvio.
- Aos professores do Instituto de Química pela contribuição acadêmica.
- A Arnaldo, Elizabeth, Gaspar, Gliseida, José, Karin, Laércio, Maria, Maritza, Miyuki, Percy, Pilar, Rodolfo, Rosangela, Simone, entre outros, pela amizade e apoio.
- E em geral a todos que direta ou indiretamente ajudaram no desenvolvimento deste trabalho.

MODELAGEM DE DADOS CINÉTICOS ESPECTROFOTOMÉTRICOS POR MÉTODOS QUIMIOMÉTRICOS DE ORDEM SUPERIOR

Autor: Jacqueline Marilyn Duenãs Cueva

Orientador: Ronei Jesus Poppi

Co-orientadora: Adriana Vitorino Rossi

RESUMO

Neste trabalho, investigou-se a aplicação de métodos quimiométricos de ordem superior para o modelagem da cinética de misturas de isômeros de posição do aminofenol. Como estes isômeros apresentam espectros muito semelhantes, foi preciso desenvolver uma metodologia experimental para diferenciá-los. Esta metodologia consistiu em monitorar as reações de diazotação entre os isômeros de aminofenol e nitrito de sódio, em $\text{pH} = 4$, na faixa espectral de 276 – 488 nm, em intervalos de tempos de 3 segundos, durante 28 segundos.

A informação obtida para cada amostra contendo os isômeros foi arranjada numa matriz na qual uma dimensão continha os valores de absorbância para diferentes comprimentos de onda e outra dimensão continha os valores de absorbância para diferentes intervalos de tempo. O conjunto de dados de várias amostras formava um arranjo tridimensional de dados de terceira ordem.

Estes dados foram processados por métodos quimiométricos de ordem superior, como PARAFAC1 e PARAFAC2 para sua modelagem e N-PLS para a quantificação. Foram obtidos resultados bastantes satisfatórios na quantificação e na recuperação dos perfis espectrais e cinéticos de cada produto da reação, o que permite identificar a composição da mistura reacional.

MODELAGEM DE DADOS CINÉTICOS ESPECTROFOTOMÉTRICOS POR MÉTODOS QUIMIOMÉTRICOS DE ORDEM SUPERIOR

Autor: Jacqueline Marilyn Duenãs Cueva

Orientador: Ronei Jesus Poppi

Co-orientadora: Adriana Vitorino Rossi

ABSTRACT

In this study, it was investigated the application of multi-way methods for the kinetic spectrophotometric data modelling of aminofenol isomer mixtures. These isomers present similar spectra, then it was necessary to develop an experimental methodology to differ them. This methodology consisted in monitoring the diazotization reaction between aminofenol isomers and sodium nitrite, at pH = 4, in the spectral range from 276 to 488 nm, at time intervals of 3 seconds during 28 seconds.

The information obtained for every sample containing isomers was arranged in a matrix, where one dimension gave absorbance values for different wavelengths and another dimension contained absorbance values for different intervals of time. The data set of various samples formed a three-way data array.

The data were processed by multi-way methods, as PARAFAC1 and PARAFAC2 for the modeling and N-PLS for the quantification. The results obtained in the quantification and in the recovery of spectral and kinetic profiles of each reaction product were very satisfactory, allowing to identify the mixture composition.

CURRICULUM VITAE

A - IDENTIFICAÇÃO:

NOME: Jacqueline Marilyn Dueñas Cueva

FILIAÇÃO: Jorge Dueñas e Margarita Cueva

NATURALIDADE: Lima - Peru

NACIONALIDADE: Peruana

NASCIMENTO: 20/08/73

ESTADO CIVIL: Solteira

B - FORMAÇÃO:

- CURSO 1º GRAU e 2º GRAU : 1979 a 1989

Colegio Fé y Alegría Nº 2 , Lima - Peru

- CURSO SUPERIOR :

Universidad Nacional de Ingeniería (UNI), Lima - Peru

Facultad de Ciencias – Escuela Profesional de Química

Bacharelado em Química: 1991 a 1995

C - CURSOS:

- Espectrometría de emisión por plasma (ICP, ICP - MS) y sus aplicaciones, 1995, Facultad de Minas - Universidad Nacional de Ingeniería, Lima - Peru.

- Taller de galvanía, 1991, Facultad de Ciencias - Universidad Nacional de Ingeniería, Lima - Peru.

- Intoxicaciones, prevención y tratamiento, 1994, Facultad de Farmacia y Bioquímica, Universidad Nacional Mayor de San Marcos, Lima - Peru.

D. TRABALHOS APRESENTADOS:

- " PLS Tridimensional para quantificação de misturas ternárias de isômeros em análise cinética ", 10º Encontro Nacional de Química Analítica, Universidade de Santa Maria, Rio Grande do Sul, Brasil, 1999.

- " Quimiluminiscencia : Montaje y Exposición " , Expociencia 93 y medio ambiente, Facultad de Ciencias, Universidad Nacional de Ingeniería, Lima - Peru.

- " Pigmentos y Pinturas " , Expociencia 91, Facultad de Ciencias, Universidad Nacional de Ingeniería, Lima - Peru.

E. EXPERIÊNCIA PROFISSIONAL:

Asistente no Laboratório de Control de Calidad de la Fábrica Nacional de Acumuladores Etna, Peru, 1997.

F. ESTÁGIO:

- Práticas de Aplicações Analíticas em Recursos Minerais e Produtos Industrias no Laboratório de Espectrometria, Facultad de Ingeniería Geológica, Minera e Metalúrgica, Universidad Nacional de Ingeniería, Lima - Peru, 1995.

SUMÁRIO

Sumário	I
Conceitos fundamentais	IV
Operadores e notação matemática	VI
Capítulo 1 Introdução e objetivos	1
1.1 Introdução	1
1.2 Objetivos	5
Capítulo 2 Dados multidimensionais	6
2.1 Introdução	6
2.2 Arranjo dos dados multidimensionais	8
2.3 Dados químicos multidimensionais	10
Capítulo 3 Pré - processamento de dados multidimensionais	12
3.1 Introdução	12
3.2 Centrar na média	13
3.2.1 Tipos de centragem	14
3.3 Padronização	17
3.3.1 Tipos de padronização	18
3.4 Considerações importantes sobre o escalamento para dados tridimensionais	20

Capítulo 4 Modelos quimiométricos de ordem superior	22
4.1 Introdução	22
4.2 Modelos quimiométricos de ordem superior	23
4.2.1 Estrutura	24
4.2.2 Restrições	25
4.2.3 Solução única	26
4.3 Modelo de PARAFAC1	26
4.3.1 Teoria	26
4.3.2 Estrutura do modelo	27
4.3.3 PARAFAC1 com regressão linear	30
4.4 Modelo de PARAFAC2	31
4.4.1 Teoria	31
4.4.2 Estrutura do modelo	31
4.5 Modelo de N-PLS	35
4.5.1 Teoria	35
4.5.2 Estrutura do modelo	36
4.5.3 Notação para os modelos de N-PLS	40
Capítulo 5 Parte experimental	41
5.1 Obtenção da metodologia cinética espectrofotométrica para a análise dos isômeros de aminofenol	41
5.1.1 Fundamentos para a escolha dos parâmetros para o monitoramento da reação aminofenol e nitrito de sódio	42
5.2 Reação química entre nitrito de sódio e os isômeros aminofenol em meio ácido	43
5.3 Equipamentos utilizados para o monitoramento das reações	43
5.3.1 Espectrofotômetro com arranjo de diodos HP 8452	45
5.3.2 Sistema de injeção simultânea de reagentes	47
5.4 Preparação das soluções reagentes e das amostras	47

5.5	Planejamento experimental	48
5.6	Aquisição de dados a partir do monitoramento das reações cinéticas	50
Capítulo 6 Resultados e discussão		51
6.1	Avaliação preliminar da reação de aminofenol com nitrito em meio ácido	51
6.2	Modelagem dos dados cinéticos com PARAFAC1	57
6.3	Modelagem dos dados cinéticos com PARAFAC2	67
6.4	Comparação dos resultados obtidos pelos métodos de PARAFAC1 e PARAFAC2	77
6.5	Resultados de N-PLS	80
Conclusões		90
Apêndices		92
1.	Algoritmo do ALS	92
2.	Algoritmo do PARAFAC1	94
3.	Algoritmo do PARAFAC2	99
4.	Algoritmo de N-PLS	105
5.	Parâmetros estatísticos	109
Referências bibliográficas		111

CONCEITOS FUNDAMENTAIS

Ajuste	Indica como o modelo construído descreve os dados originais. Pode ser expresso como percentagem da informação explicada.
Componente	Fator que descreve um conjunto de dados.
Dimensão	Usado para denotar o número de variáveis em cada modo.
Fator	Formado por um vetor <i>score</i> e um ou mais vetores <i>loadings</i> .
Modo	Direção de um grupo de dados. Por exemplo, uma matriz de dados tem dois modos. Um arranjo tridimensional de dados ou tensor tem 3 modos.
Ordem	A ordem de um conjunto de dados é o número de modos. Assim, a matriz é de segunda ordem, e o tensor de terceira ordem.
Perfil cinético	Gráfico formado por um vetor <i>loading</i> , pertencente ao modo que contém as variáveis de tempo.
Perfil espectral	Gráfico formado por um vetor <i>loading</i> , pertencente ao modo que contém as variáveis de comprimentos de onda.

-
- Triade** Um fator trilinear formado por um vetor *score* e dois vetores *loadings*.
- Variável latente** Fator que descreve um conjunto de dados levando em consideração outro conjunto de dados.
- Vetor *loading*** Parte de um fator referido a um modo específico de variáveis, por exemplo, o modo ou dimensão dos comprimentos de onda. Nos modelos de PARAFAC1 e PARAFAC2, a parte de um fator referido ao modo das amostras é considerada como *loading*.
- Vetor *score*** Parte de um fator referido ao modo das amostras.

OPERADORES E NOTAÇÃO MATEMÁTICA

Para uma melhor compreensão da álgebra envolvida nos modelos de tratamento de dados, a definição de alguns operadores matemáticos é apresentada, assim como a notação utilizada nesta tese.

x	Escalar, em letra minúscula.
\mathbf{x}	Vetor, em letra minúscula e em negrito.
\mathbf{X}	Matriz, em letra maiúscula e em negrito.
$\underline{\mathbf{X}}$	Tensor, letra maiúscula sublinhada e em negrito.
\mathbf{X}^T	Transposta da matriz \mathbf{X} .
\mathbf{X}^+	Pseudoinversa da matriz \mathbf{X} .
$ \mathbf{x} $	Módulo do vetor \mathbf{x} .
$\mathbf{x} \otimes \mathbf{y}$	Produto <i>Kronecker</i> dos vetores \mathbf{x} e \mathbf{y} . Onde, cada elemento de \mathbf{x} é multiplicado por \mathbf{y} . Por exemplo seja \mathbf{x} de dimensão $I \times 1$ e \mathbf{y} de dimensão $J \times 1$, o produto <i>Kronecker</i> gera um vetor de dimensão $JI \times 1$, dado por:

$$\mathbf{x} \otimes \mathbf{y} = \begin{bmatrix} x_{11}y \\ \vdots \\ x_{I1}y \end{bmatrix}$$

$\mathbf{X} \otimes \mathbf{Y}$	Produto <i>Khatri-Rao</i> entre \mathbf{X} e \mathbf{Y} , onde estas duas matrizes tem o mesmo número de colunas. $\mathbf{X} \otimes \mathbf{Y} = [\mathbf{x}_1 \otimes \mathbf{y}_1 \quad \mathbf{x}_2 \otimes \mathbf{y}_2 \quad \dots \quad \mathbf{x}_t \otimes \mathbf{y}_t]$
$\ \mathbf{X}\ _f^2$	Norma de Frobenius ou Euclidiana de \mathbf{X} , $\ \mathbf{X}\ _f^2 = \text{tr}(\mathbf{X}^T \mathbf{X})$
$\text{tr } \mathbf{X}$	Soma dos elementos da diagonal de \mathbf{X} .
$\text{diag}(\mathbf{X})$	Vetor contendo a diagonal de \mathbf{X} .
$\min(\mathbf{X})$	O elemento mínimo de \mathbf{X} .
$\arg \min_{\mathbf{X}}(f(\mathbf{X}))$	O argumento de \mathbf{X} que minimiza o valor da função $f(\mathbf{X})$.

CAPÍTULO 1

INTRODUÇÃO E OBJETIVOS



INTRODUÇÃO E OBJETIVOS

1.1 – INTRODUÇÃO

Quimiometria¹ é uma área de crescente destaque na Química devido a seu forte potencial para a investigação de sistemas químicos, sendo utilizados para este fim, métodos matemáticos, estatísticos, assim como lógica matemática e planejamentos experimentais. Uma nova tendência da quimiometria, hoje em dia, é o estudo de aplicações e desenvolvimento de métodos matemáticos aplicados à análise de dados químicos de ordem superior (maior a 2). Tais métodos são conhecidos como métodos quimiométricos de ordem superior²⁻³.

Dados químicos de terceira ordem⁴ podem ser gerados pelo agrupamento de informação de várias amostras obtidas por métodos analíticos que fornecem um conjunto de dados de segunda ordem para cada amostra. Esses métodos analíticos são comumente conhecidos nos laboratórios de Química Analítica. Alguns exemplos de tais métodos são fluorescência de emissão e excitação⁵, espectrometria de massa - massa (MS-MS)⁶, Cromatografia líquida de alta resolução com detecção espectrofotométrica (CLAE - UV/VIS)⁷, titulações ácido-base usando espectrofluorimetria a diferentes comprimentos de onda de excitação⁸ e o método cinético espectrofotométrico⁹.

Os métodos cinéticos¹⁰⁻¹¹ de análise baseiam-se em medidas de dados em condições dinâmicas. O método cinético espectrofotométrico é um dos mais utilizado em Química para análises qualitativas e quantitativas. Este método consiste em monitorar reações químicas, medindo valores de absorvância em um ou mais comprimentos de onda, que variam em

função do tempo. A variação de absorvância é produzida pelo desenvolvimento de uma ou mais reações químicas.

Todas as reações químicas têm velocidades que dependem da temperatura, pressão, concentração das espécies reagentes e presença de catalisadores, ativadores ou inibidores. Lembrando que, um catalisador acelera a reação, baixando sua energia de ativação sem modificar seu estado de equilíbrio. Um ativador pode interagir com o catalisador inclusive participando de sua regeneração, o que resulta na aceleração da reação embora não se trate de uma catálise. Um inibidor diminui a velocidade de reação, por interagir com o catalisador formando complexos para diminuir ou impedir sua ação. Em sistemas líquidos a pressão não é um fator relevante, já que a maioria das reações em fase líquida realizam-se à pressão atmosférica. A dependência da velocidade de reação com as concentrações dos reagentes pode ser melhor entendida com o exemplo descrito abaixo:

Seja uma reação genérica:



A velocidade de reação “v” pode ser expressa como:

$$v = \frac{d[C]}{c dt} = -\frac{d[A]}{a dt} = -\frac{d[B]}{b dt} \quad (2)$$

E a equação da cinética é:

$$v = k [A]^n [B]^m \quad (3)$$

Onde:

- k é chamada constante de velocidade de reação.
- m e n representam a ordem da reação com relação a cada reagente.

O valor desta constante depende da temperatura¹⁰ e esta dependência pode ser expressa quantitativamente na seguinte equação:

$$k = Qze^{-E/RT} \quad (4)$$

O “Q” é o fator estérico, “z” é a frequência de colisões, “E” é a energia de ativação, “T” é a temperatura absoluta e “R” é a constante universal dos gases.

Como a velocidade da reação depende de vários fatores, é importante nos métodos cinéticos manter constantes certas condições experimentais, como temperatura, e buscar condições nas quais a velocidade de reação só seja afetada pela concentração do analito. Isso pode permitir identificá-lo e quantificá-lo em procedimentos reprodutíveis que podem ser utilizados em uma metodologia analítica.

O método cinético permite aproveitar a diferença das velocidades de reação das espécies de uma mistura para diferenciar as espécies presentes, viabilizando análises qualitativas e quantitativas.

As principais vantagens dos métodos cinéticos são possibilitar análises a baixo custo e permitir o uso de maior quantidade de reações químicas para fins analíticos. Assim, reações lentas em atingir o equilíbrio ou não suficientemente quantitativas podem ser utilizadas em métodos cinéticos de análise.

O método cinético espectrofotométrico modernizou-se com o desenvolvimento de detetores multicanais de varredura rápida (Arranjo de diodos)¹². Esta melhora permite adquirir espectros completos na região ultravioleta-visível em tempos pequenos. Este feito facilitou o estudo de sistemas cinéticos rápidos aumentando o número de aplicações. Assim, hoje a metodologia cinética de análise representa uma área da química analítica de grande e crescente destaque, com ótimos níveis de sensibilidade e reprodutibilidade, o que viabiliza o estudo de sistemas químicos problemáticos para métodos termodinâmicos de análise.

Esta metodologia cinética espectrofotométrica aliada aos métodos quimiométricos, representa uma forte ferramenta para resolver problemas analíticos como quantificação e separação de isômeros estruturais, especificamente isômeros de posição que apresentam espectros semelhantes, tais como os compostos *o*-aminofenol, *m*-aminofenol e *p*-aminofenol.

As aminas aromáticas primárias como: *o*-aminofenol, *m*-aminofenol e *p*-aminofenol podem reagir com o ácido nitroso mediante reação chamada diazotação¹³, formando sais de areno diazônio, que são muito mais estáveis que os sais de diazônio alifáticos. Graças a esta estabilidade, pode-se monitorar a cinética da reação dos isômeros mencionados com íon nitrito, em meio ácido, numa solução tampão de pH=4. Em valores de pH mais altos a reação torna-se lenta demais podendo ocorrer problemas de acoplamento¹⁴ entre os sais diazônio e os fenóis. Se o pH da solução é inferior a 5 a velocidade de acoplamento da amina é baixa. As reações desenvolvidas no sistema cinético dependem da concentração do isômero e da concentração de nitrito a pH constante. Além disso, também existe a possibilidade que a cinética entre um isômero e o nitrito seja afetada pela presença dos outros isômeros. A figura 1 representa a reação entre o aminofenol e o íon nitrito.

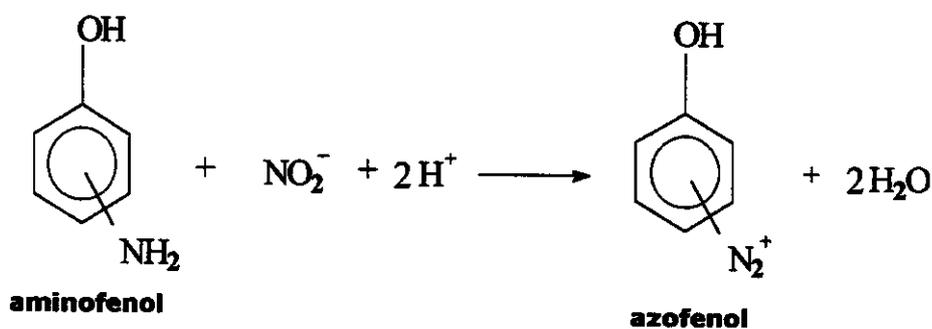


Figura 1: Reação entre *o*-aminofenol e íon nitrito.

As informações obtidas a partir do monitoramento das reações desenvolvidas, podem ser organizadas sob a forma de um tensor de terceira ordem de dados (arranjo tridimensional). Uma representação do tensor de terceira ordem gerada é apresentada na Figura 2.

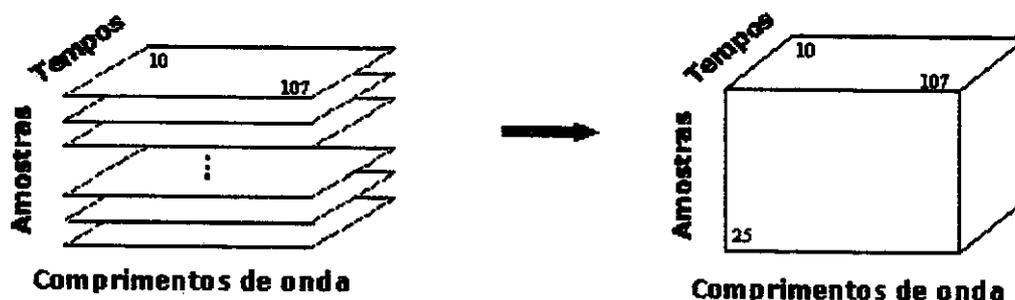


Figura 2: Exemplo de um arranjo tridimensional de dados cinéticos espectrofotométricos para várias amostras.

Estes dados podem ser tratados por métodos quimiométricos de ordem superior para sua modelagem e quantificação dos isômeros nas diferentes amostras. Detalhes sobre as técnicas quimiométricas podem ser vistos no capítulo 4.

1.2 – OBJETIVOS

Os objetivos deste trabalho de pesquisa foram:

- Desenvolver uma metodologia experimental para o monitoramento da reação entre isômeros de *o*-, *m*- e *p*-aminofenol e nitrito de sódio.
- Estudar o tipo de pré-processamento de dados cinéticos mais adequado para ser usado na modelagem com métodos quimiométricos de ordem superior.
- Estudar a estrutura, o algoritmo e restrições de tais métodos com a finalidade de identificar os modelos mais adequados a serem aplicados no arranjo de dados cinéticos espectrofotométricos de terceira ordem.
- Identificar as espécies presentes nas misturas ternárias de isômeros e quantificá-las usando os métodos quimiométricos de ordem superior como PARAFAC1, PARAFAC2 e N-PLS.

CAPÍTULO 2

DADOS MULTIDIMENSIONAIS



DADOS MULTIDIMENSIONAIS

2.1 - INTRODUÇÃO

Um dado multidimensional^{2,4} pode ser descrito da forma:

$$x_{ij} \dots$$

onde: $i = 1, \dots, I$; $j = 1, \dots, J$; ...

O número de índices para este tipo de dados pode ser dois ou mais. O conjunto de dados denotado com um só índice chama-se grupo de dados com uma só dimensão ou de primeira ordem. Um exemplo de dados de primeira ordem seria o espectro digitalizado de uma amostra ou seja um vetor. Com 2 índices, chama-se grupo de dados bidimensionais ou de segunda ordem, isto é, uma matriz. Um exemplo seria o conjunto de dados cinéticos espectrofotométricos, onde para cada tempo monitorado obtém-se um espectro. Com 3 índices, os dados são tridimensionais ou de terceira ordem, e podem ser ordenados em forma de paralelepípedo formando um tensor. Uma representação gráfica deste tensor é mostrado na figura 3. Um exemplo de dados de terceira ordem seria o conjunto de dados cinéticos espectrofotométricos de 2ª ordem de várias amostras.

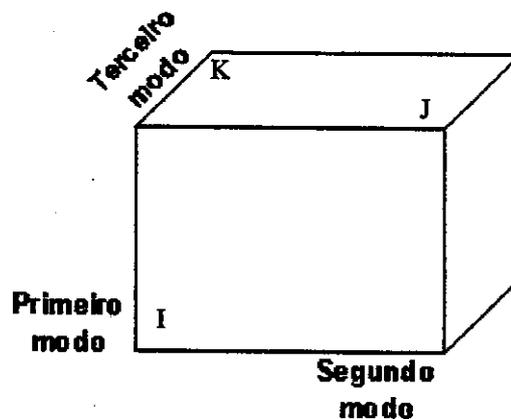


Figura 3: Representação gráfica de tensor de dimensões $I \times J \times K$ formado por um conjunto de dados tridimensionais.

Quando se trabalha com matrizes são usados os termos linha e coluna. Para tensores de dados tridimensionais, além de serem usados os termos linha e coluna, usa-se o termo profundidade¹⁵ para os vetores do terceiro modo. Isto está esquematizado na figura 4:

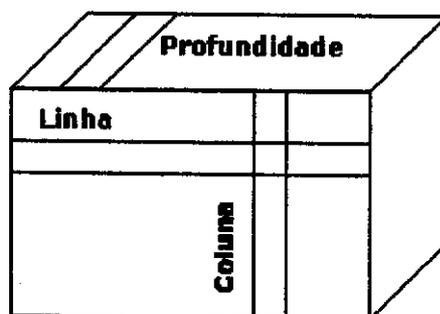


Figura 4: Gráfico demonstrativo da localização de linhas, colunas e profundidades num tensor tridimensional.

Uma submatriz¹⁵ é comumente conhecida como fatia. Uma submatriz de um conjunto de dados tridimensional pode ser obtida fixando, por exemplo, o índice do terceiro modo. Cada direção de um grupo de dados tridimensional é chamado de modo, e o número de níveis ou variáveis desse modo é chamado dimensão. O número de modos é a dimensão geométrica do arranjo. Por exemplo, o tensor apresentado na figura 3, apresenta 3 modos. Neste caso a dimensão do primeiro modo é I, a dimensão do segundo modo é J e a dimensão do terceiro modo é K e a dimensão do tensor é $I \times J \times K$.

Uma matriz de dados é denotada com letra maiúscula em negrito, como X , os caracteres I e J são usado geralmente para indicar a dimensão da matriz. Cada elemento desta matriz pode ser descrito como x_{ij} , onde i e j indicam a localização desse elemento no primeiro e segundo modo respectivamente. A j-ésima coluna de X será denotada como x_j e i-ésima linha de X será descrita como x_i .

O tensor tridimensional é descrito por uma letra maiúscula em negrito e sublinhada, como $\underline{\mathbf{X}}$. A dimensão deste tensor é descrito por três caracteres em maiúsculas, por exemplo I, J, e K. Um caracter em maiúsculo é usado para indicar a dimensão de um modo do tensor. Cada elemento deste tensor é chamado de escalar e pode ser denotado por uma letra minúscula, como x_{ijk} , onde os sub-índices i, j, e k indicam a localização desse escalar no primeiro, segundo e terceiro modo, respectivamente.

As submatrizes obtidas a partir do tensor $\underline{\mathbf{X}}$ podem se denotadas como¹⁵: \mathbf{X}_i para uma matriz com dimensão K x J obtida fixando o primeiro modo à i-ésima variável $\underline{\mathbf{X}}$ (i, :, :); \mathbf{X}_j para uma matriz com dimensão I x K obtida fixando o segundo modo à j-ésima variável $\underline{\mathbf{X}}$ (j, :, :); e \mathbf{X}_k para uma matriz com dimensão I x J obtida fixando o terceiro modo à k-ésima variável $\underline{\mathbf{X}}$ (k, :, :).

2.2 - Arranjo dos dados multidimensionais

Existem várias maneiras de arranjar dados multidimensionais^{7,9}. Por exemplo, para um conjunto de dados cinéticos de "K" amostras, onde cada amostra é representada por uma matriz de dados de dimensões "I" tempos x "J" comprimentos de onda; os dados podem ser arranjos num tensor de terceira ordem de dimensão "I" tempos x "J" comprimentos de onda x "K" amostras, como pode ser mostrado na figura 5a. Uma outra forma é apresentada na figura 5b, no qual o tensor tem dimensão de "I" tempos x "K" amostras x "J" comprimentos de onda.

O conjunto de dados cinéticos de várias amostras (dados de terceira ordem) pode ser arranjos numa matriz. Assim, por exemplo podem ser formadas matrizes de dimensão I x JK, Kx JI, etc. Também podem formar matrizes diferentes de igual dimensão como I x JK e I x KJ. A figura 6, mostra a formação de uma matriz de dimensão I x JK, obtida a partir de dados cinéticos de várias amostras.

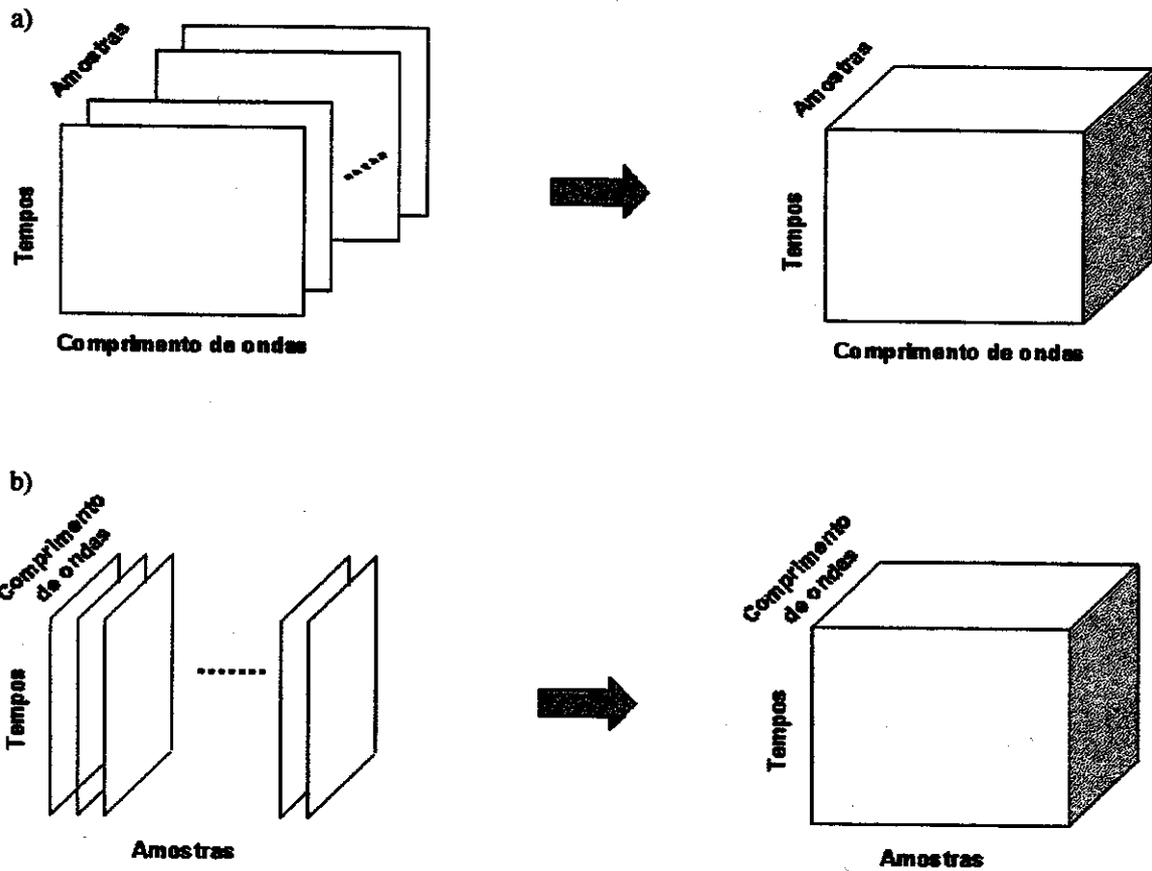


Figura 5: a e b mostram tensores de terceira ordem obtidos por um conjunto de dados cinéticos de várias amostras.

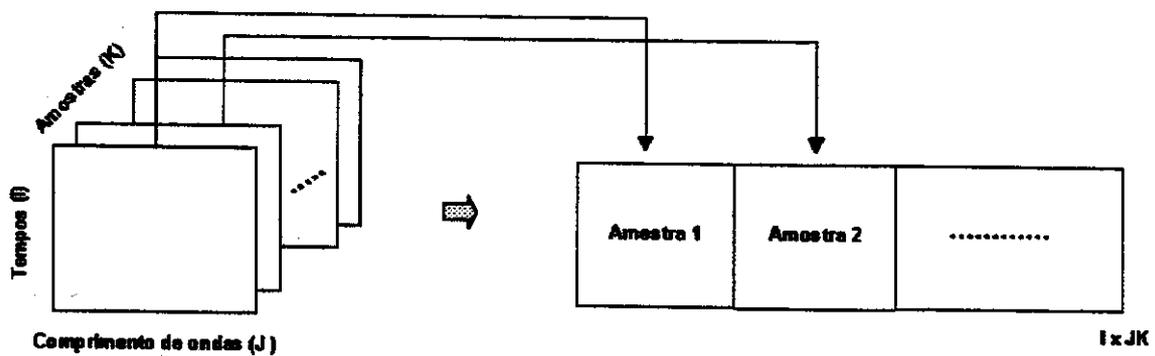


Figura 6: Formação de uma matriz de dimensão I x JK

A figura 7 mostra a formação de uma matriz de dimensão $K \times JI$. Observa-se que é necessário uma etapa adicional para formar esta matriz. Inicialmente cada matriz de dados de cada amostra ($I \times J$) é desdobrada em um vetor linha de dimensão $1 \times JI$. Ou seja, uma linha é colocado ao lado da outra. Após isso, cada vetor linha que representa os dados de uma amostra é colocado um abaixo do outro. A formação da matriz $K \times JI$ é quase semelhante à formação da matriz $K \times JI$, com a diferença que o vetor linha é obtido pelo desdobramento da transposta da matriz de dados de cada amostra.

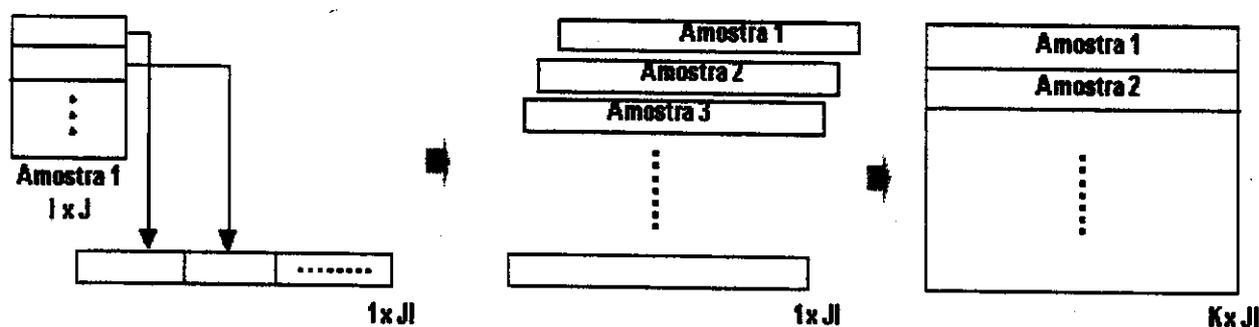


Figure 7: Formação de uma matriz $K \times JI$.

2.3 - DADOS QUÍMICOS MULTIDIMENSIONAIS

O avanço tecnológico fez com que alguns instrumentos analíticos pudessem ser acoplados, dando origem a novos instrumentos, que genericamente são chamados de hifenados.¹⁶⁻¹⁷

Os dados químicos multidimensionais² são gerados por métodos químicos de instrumentos hifenados.

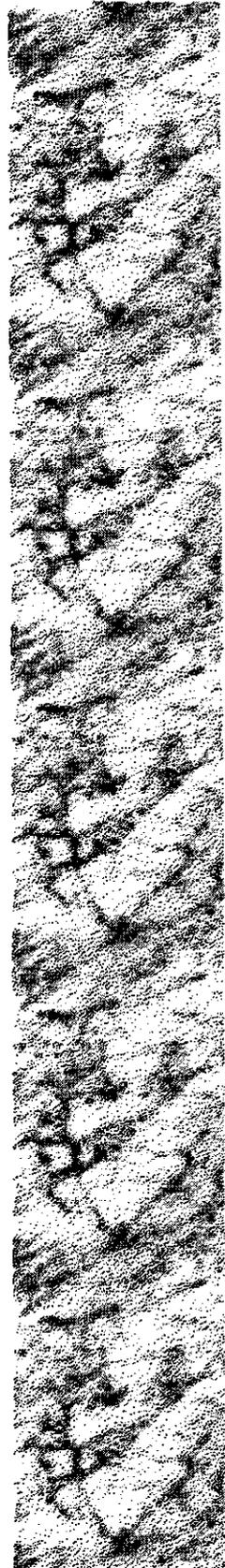
Exemplos destes métodos químicos instrumentais hifenados são:

- Fluorescência de emissão e excitação,⁵ técnica pela qual se obtêm valores de intensidade medidos a vários comprimentos de onda de emissão, para cada comprimento de onda de excitação.

- MS - MS (Espectroscopia de massa / massa),⁶ técnica pela qual se obtêm valores de intensidade medidos a diferentes números massa/carga (fragmentos pais) e medidos a diferentes números de massa/carga dos fragmentos filhos.
- CLAE - UV/VIS (Cromatografia líquida de alta resolução com detecção espectrofotométrica)⁷, técnica pela qual se obtêm espectros na região ultravioleta e visível medidos a diferentes tempos de retenção.
- CG - EM (Cromatografia gasosa hifenada a espectrometria de massa)¹⁸, técnica pela qual se obtêm espectros de massa medidos a diferentes tempos de retenção.
- CG - IR (Cromatografia gasosa hifenada a um infravermelho)¹⁹, técnica pela qual se obtêm espectros na região no infravermelho para diferentes tempos de retenção.

CAPÍTULO 3

PRÉ-PROCESSAMENTO DE DADOS MULTIDIMENSIONAIS



PRÉ-PROCESSAMENTO DE DADOS MULTIDIMENSIONAIS

3.1 - INTRODUÇÃO

O pré-processamento de dados²⁰ para arranjos da ordem superior é um pouco mais complexo que para o caso de dados bidimensionais. É importante compreender as inter-relações⁴ que existem entre os dados dos três modos, já que dessa maneira será mais fácil decidir pelo tipo de pré-processamento que não causará a perda destas inter-relações.

Existem dois tipos básicos de pré-processamento de dados: centrado na média e padronização.

A centragem na média²⁰ é normalmente utilizada para remover os termos constantes dos dados. Estes dados sem serem centrados na média poderiam necessitar de uma variável latente extra para a modelagem. Algumas vezes é necessário centrar os dados na média para torná-los compatíveis com o modelo estrutural.

A padronização²⁰ ou escalamento faz que o conjunto de dados cujas variáveis tem diferentes unidades, como temperatura, pH, etc, tenham a mesma relevância na modelagem.

Outro método de pré-processamento pode ser originado pela combinação dos tipos básicos de escalamento. Este processo consiste em centrar na média e padronizar os dados, tal que os dados resultantes tenham média zero e desvio padrão igual a um.

3.2 - CENTRAR NA MÉDIA

A centragem na média consiste em subtrair um termo constante de cada elemento de um conjunto de dados, de tal modo que os valores resultantes tenham média zero. Esse termo constante é a média dos dados. Assim este pré-processamento, para um conjunto de J dados pode ser expresso matematicamente como:

$$x_j^{\text{cent}} = x_j - \frac{\sum_{j=1}^J x_j}{J} \quad (5)$$

em que:

- x_j^{cent} é o dado centrado na média e x_j é o dado original.

Este tipo de pré-processamento para dados multidimensionais, por exemplo, de três dimensões (ou modos) é complexo. Por exemplo, considere um grupo de dados de três dimensões, onde a primeira dimensão tem "I" variáveis do tipo i, a segunda dimensão têm "J" variáveis do tipo j e a terceira dimensão "K" variáveis do tipo k:

- A centragem na primeira dimensão pode ser feita, arranjando os dados para uma matriz de dimensão I x JK. Logo, calcular a média para cada conjunto de dados, que pertencem às variáveis jk definidas e todas as variáveis de i. Este pré-processamento é chamado de centragem i. Matematicamente pode ser expresso como:

$$x_{ijk}^{\text{cent}} = x_{ijk} - \frac{\sum_{i=1}^I x_{ijk}}{I} \quad (6)$$

Esta centragem é aplicada para cada conjunto de dados com variáveis j e k definidas, onde x_{ijk}^{cent} é o dado já centrado, x_{ijk} é o dado original e "I" é o número de variáveis da primeira dimensão.

- A centragem na segunda dimensão pode ser feita, arranjando os dados para uma matriz de dimensão $J \times KI$ e após isso, calcular a média dos dados que pertencem as variáveis definidas i, k e todas as variáveis j . Este tipo de pré-processamento é chamado de centragem j e o cálculo pode ser expresso pela seguinte equação:

$$X_{ijk}^{cent} = X_{ijk} - \frac{\sum_{j=1}^{j=J} X_{ijk}}{J} \quad (7)$$

Este cálculo é aplicado para cada conjunto de dados com variáveis definidas de i e k . O dado centrado está simbolizado como x_{ijk}^{cent} , o dado original por x_{ijk} e "J" é o número de variáveis da segunda dimensão.

O efeito de centrar sobre o conjunto de dados multidimensionais é descrito por Kruskal²¹ e por Harshman & Lundy²². Centrar, por exemplo, uma matriz em vez de colunas destruirá o comportamento multilinear dos dados, porque mais níveis constantes serão adicionados do que eliminados.

3.2.1 - TIPOS DE CENTRAGEM

Existem vários tipos de centragem²⁰ para um conjunto de dados de terceira ordem e eles são diferenciados de acordo com o sub-conjunto de dados que se tome para o cálculo da média. Estes tipos de centragem serão explicados a seguir para dados de terceira ordem x_{ijk}

que tem "I" variáveis no primeiro modo, "J" variáveis no segundo modo e "K" variáveis no terceiro modo.

1. Centragem sobre todos os dados.

É a centragem na média, sobre todos os dados dos modos. Para facilitar o cálculo, os dados tridimensionais são arranjados num vetor de dimensão $1 \times IJK$. Na figura 8, a flecha indica a direção da centragem.

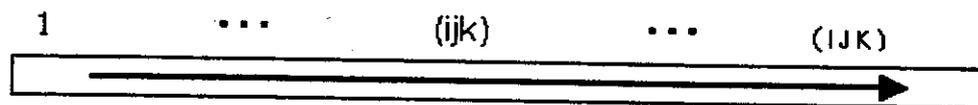


Figura 8: Centragem na média sobre todos os dados tridimensionais.

2. Centragem na média sobre cada conjunto de dados que pertencem a uma variável de um modo e todas as variáveis dos outros dois modos.

Existem três possibilidades: centragem i , centragem j e centragem k . Para facilitar o cálculo os dados são arranjados em uma matriz de $I \times JK$ para centragem- i , $J \times KI$ para centragem- j e $K \times IJ$ para centragem- k .

Por exemplo: a centragem j , é realizada para cada conjunto de dados que pertencem a uma variável j definida. Na figura 9, a flecha indica a direção da centragem j sobre a matriz de dados $J \times IK$:

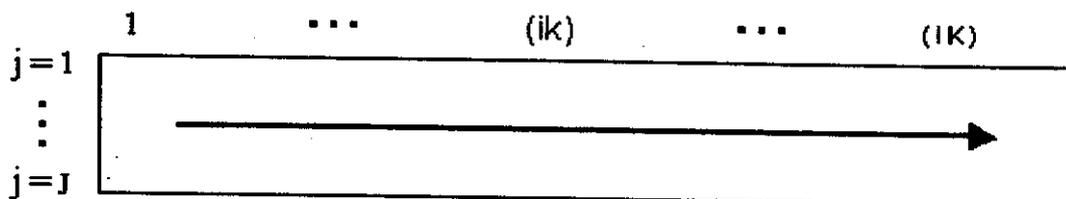


Figura 9: J-centragem sobre o tensor de dados desdobrado numa matriz $J \times I K$.

3. Centragem para cada conjunto de dados, que pertencem a uma combinação de variáveis de dois modos (2 variáveis) e todas as variáveis do modo restante.

Existem 3 possibilidades: centragem ij , centragem jk , e centragem ki . Para facilitar os cálculos os dados são arranjados em uma matriz $K \times I \times J$, $I \times J \times K$, e $J \times K \times I$, respectivamente.

Por exemplo, a centragem jk , é realizada para cada conjunto de dados que pertencem a um par de variáveis definidas j e k .

Na figura 10 é mostrado graficamente a direção da operação para a centragem jk , sobre uma matriz de dados $I \times J \times K$:

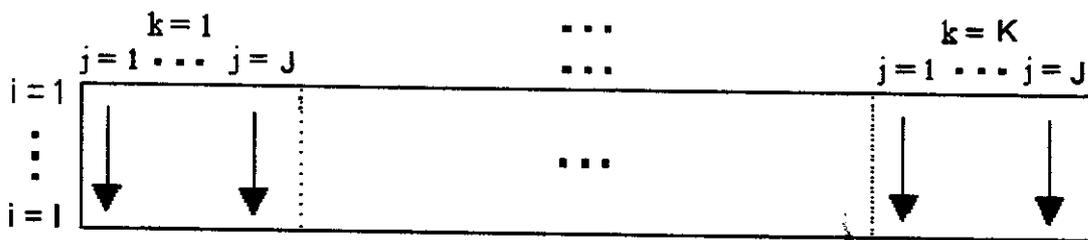


Figura10: Centragem jk sobre os dados tridimensionais arranjados numa matriz $I \times J \times K$.

4. Centragem dupla

É uma combinação das possibilidades de centragem do tipo anterior. Esta centragem têm 3 possibilidades: centragem $ij - jk$, centragem $jk - ki$, centragem $ki - ij$.

Por exemplo, a centragem $jk - ki$, é realizada para um conjunto de dados que pertencem a um par de variáveis jk e para o conjunto de dados que pertencem a um par de variáveis ki definidas. Na figura 11, as flechas indicam as direções da media para o tipo de centragem $jk - ki$, sobre uma matriz de dados de dimensão $I \times J \times K$.

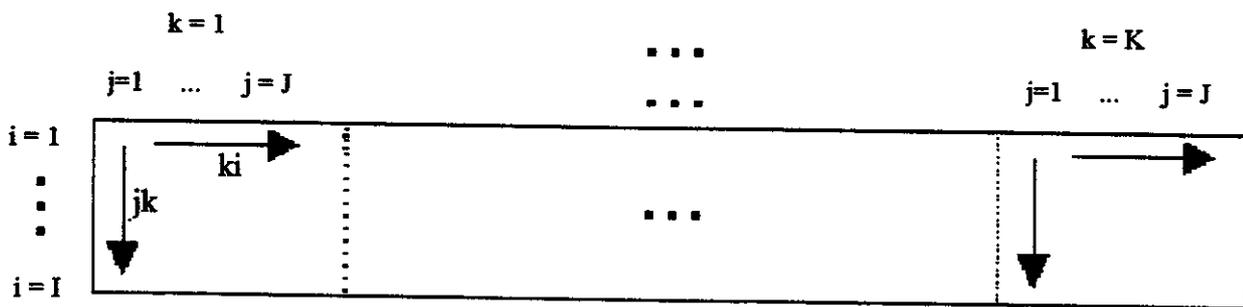


Figura 11: Centrado $jk - ki$ sobre os dados tridimensionais arranjados numa matriz $I \times J \times K$.

5. Centragem tripla

É uma combinação das possibilidades de centrar do item 3. Envolve centrar na média os dados que pertencem a um par de variáveis jk definidas, depois centrar na média os dados que pertencem a um par de variáveis ki definidas e logo em seguida centrar na média os dados que pertencem a um par de variáveis ij definidas. Na figura 12, a direção dos cálculos é mostrado sobre um paralelepípedo de dados de dimensões $I \times J \times K$.

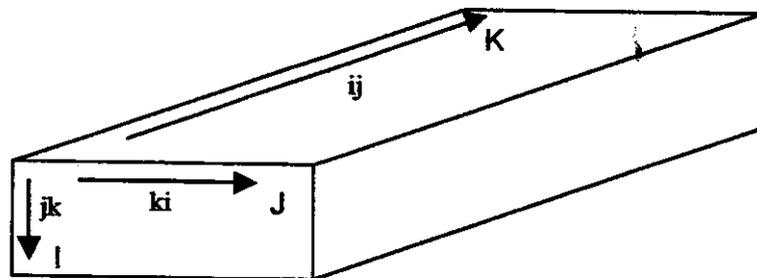


Figura 12: Centragem tripla sobre os dados tridimensionais arranjados num tensor $L \times M \times N$.

3.3 - PADRONIZAÇÃO

A padronização²⁰ consiste em dividir cada elemento por um termo constante. A escala dos dados resultantes tem um valor fixado, freqüentemente igual a 1. Algumas vezes a escala, geralmente está relacionada com alguma medida de variabilidade, freqüentemente desvio padrão.

Para um conjunto de J dados, a padronização pode ser definida como:

$$x_j^{\text{padroniz}} = \frac{x_j}{S} \quad (8)$$

onde:

O sub-índice j varia de 1 até M, x_j é o dado original, x_j^{padroniz} é o dado padronizado, e S é o termo constante que divide a cada dado.

Para um conjunto de dados tridimensionais a padronização em vários modos é complicado porque a padronização de um modo afeta a escala de outros modos.

3.3.1 - TIPOS DE PADRONIZAÇÃO

Os tipos mais conhecidos de padronização^{20,23} são:

1. Dividir o conjunto de dados pela raiz quadrada da somatória dos quadrados dos elementos

O termo constante da divisão mostrada na equação 8, é a raiz quadrada da somatória dos quadrados dos dados do conjunto.

Por exemplo, este tipo de pré-processamento aplicado no primeiro modo para dados tridimensionais x_{ijk} , requer que os dados na direção da variável i, sejam padronizados, como mostrado na equação 9.

$$s_i = \sqrt{\left(\sum_{j=1}^{j=J} \sum_{k=1}^{k=K} x_{ijk}^2 \right)} \quad (9)$$

esta equação calcula o termo constante s_i para a padronização.

$$x_{ijk}^{\text{padrão}} = \frac{x_{ijk}}{s_i} \quad (10)$$

onde:

- $x_{ijk}^{\text{padrão}}$ é o dado padronizado.
- x_{ijk} é o dado original que pertence ao conjunto de dados usado no cálculo de s_i .

2. Dividir o conjunto de dados pelo seu valor máximo

Este tipo de padronização chamado normalização é efetuada dividindo cada dado de um conjunto pelo seu valor máximo.

Por exemplo, uma padronização deste tipo aplicada ao segundo modo para um conjunto de dados tridimensionais utiliza o conjunto de dados na direção da variável j . Os cálculos realizados, são mostrados na equação 11:

$$s_j = \max ([x_{\cdot j \cdot}]) \quad (11)$$

Esta expressão calcula o termo constante para a padronização. A expressão $[x_{\cdot j \cdot}]$ representa um conjunto de dados que pertencem a uma variável j definida. *Max* é uma função matemática que calcula o valor máximo de um conjunto de dados.

$$x_{ijk}^{\text{padroniz}} = \frac{x_{ijk}}{s_j} \quad (12)$$

$x_{ijk}^{\text{padroniz}}$ é o dado padronizado e x_{ijk} é o dado original o qual pertence a um conjunto de dados usado no cálculo de s_j .

3. Dividir um conjunto de dados pelo seu desvio padrão

A padronização é efetuada calculando o desvio padrão de um conjunto de dados e dividindo cada dado desse conjunto, pelo seu desvio padrão.

Por exemplo, este tipo de padronização aplicado no terceiro modo de um conjunto de dados tridimensionais, utiliza o conjunto de dados na direção da variável k . Os cálculos realizados, são mostrados na equação 13:

$$s_k = \sqrt{\frac{\sum_{i=1}^I \sum_{j=1}^J (\bar{x}_k - x_{ijk})^2}{J \times I - 1}} \quad (13)$$

Esta equação calcula o desvio padrão para um conjunto de dados que pertencem a uma variável k definida e todas as variáveis i e j . O símbolo s_k é o termo constante para a padronização e \bar{x}_k é a média calculada desses dados.

$$x_{ijk}^{\text{padrão}} = \frac{x_{ijk}}{s_k} \quad (14)$$

$x_{ijk}^{\text{padrão}}$ é o dado padronizado e x_{ijk} é o dado original pertencente ao conjunto de dados usado no cálculo de s_k .

3.4 - CONSIDERAÇÕES IMPORTANTES SOBRE O ESCALAMENTO PARA DADOS TRIDIMENSIONAIS

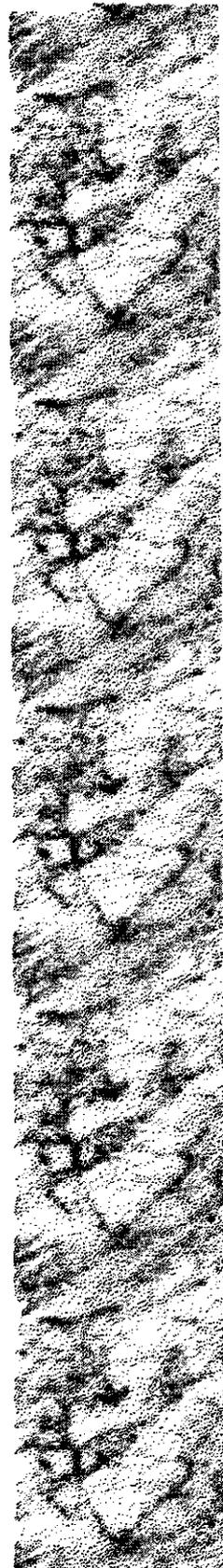
A diferença fundamental entre centrar na média e padronizar para um conjunto de dados tridimensionais, é que as combinações de diferentes centragens não são afetadas entre si, enquanto na padronização é afetada. Por exemplo, centragem dupla $jk - ik$ é uma combinação

das centrages j_k e i_k , eles podem ser realizados em qualquer ordem. Para a padronização a situação é mais complexa, a padronização de um modo destruirá a padronização de outro modo.

Poucos estudos sobre as conseqüências da mistura de pré-processamentos sobre um conjunto tridimensional de dados tem sido realizado. Sabe-se apenas que a padronização num modo destruirá a centragem sobre outros modos, e a centragem num modo perturbará a padronização de todos os outros modos.^{22,24}

CAPÍTULO 4

MODELOS QUIMIOMÉTRICOS DE ORDEN SUPERIOR



MODELOS QUIMIOMÉTRICOS DE ORDEM SUPERIOR

4.1 - INTRODUÇÃO

Um modelo é uma expressão matemática que procura descrever um conjunto de dados experimentais.

Por exemplo, o modelo conhecido como PCA²⁵ (*Principal component analysis*) tem como base estrutural, que a matriz de dados originais ou experimentais \hat{X} seja descrita como um modelo bilinear, calculado como:

$$\mathbf{X} = \mathbf{AB}^T \quad (15)$$

A matriz \mathbf{X} calculada pelo modelo bilinear²⁶ é uma aproximação dos dados originais. No PCA as matrizes \mathbf{A} e \mathbf{B} são calculadas com certas restrições, como:

$$\mathbf{A}^T \mathbf{A} = \mathbf{D}, \text{ onde } \mathbf{D} \text{ é uma matriz diagonal.} \quad (16)$$

$$\mathbf{B}^T \mathbf{B} = \mathbf{I}, \text{ onde } \mathbf{I} \text{ é uma matriz identidade.} \quad (17)$$

E tem-se a seguinte equação:

$$\min_{A,B} \|\hat{X} - \mathbf{AB}^T\|^2 \quad (18)$$

Esta expressão matemática calcula os parâmetros \mathbf{A} e \mathbf{B} para minimizar o erro entre os dados originais e os dados calculados pelo modelo

Observa-se que o modelo PCA, assim como um modelo bilinear sem restrições, terão a mesma base estrutural e a mesma equação (15) para calcular seus parâmetros.

É importante ressaltar, que a escolha adequada da base estrutural e restrições, para o modelamento de um conjunto de dados, deve estar baseada em análise exploratória dos dados e seus objetivos.

Existem três grupos de modelos em Quimiometria: modelos aplicados a conjunto de dados de apenas um modo como a regressão linear²⁷; modelos aplicados a um conjunto de dados de dois modos, como PCA²⁵, PLS²⁸, PCR²⁹ entre outros e modelos quimiométricos de ordem superior²⁵⁻²⁶, aqueles que são aplicados a dados multidimensionais de três ou mais modos.

4.2 - MODELOS QUIMIOMÉTRICOS DE ORDEM SUPERIOR

O desenvolvimento de instrumentos analíticos hifenados como CLAE - UV/VIS⁷, CG - IR¹⁹, MS - MS⁶, possibilitou a obtenção de uma grande quantidade de informação de uma amostra. A utilização desses dados tornou-se mais complexa quando dados de várias amostras são arranjadas em um único conjunto, porque o agrupamento de dados forma um tensor de terceira ordem. A necessidade de manipular esta grande quantidade de informação levou ao desenvolvimento de vários métodos quimiométricos de ordem superior para o tratamento desses dados.

Assim, os modelos quimiométricos para dados multidimensionais têm hoje em dia grande importância, já que usam todas as informações obtidas a partir de instrumentos hifenados para resolução de problemas analíticos, como investigar os componentes presentes numa mistura e realizar sua quantificação.

Foram desenvolvidos modelos quimiométricos de ordem superior, como GRAM (*Generalized rank annihilation method*)³⁰⁻³¹, TLD (*Trilinear decomposition*)³², PARAFAC1

(*Parallel factor analysis 1*)^{7,24,33,34}, PARAFAC2 (*Parallel factor analysis 2*)³⁵⁻³⁷, N-PCA (*Multi-way principal component analysis*)³⁸, TUCKER³⁹, N-PLS (*Multi-way partial least squares regression*)⁴⁰⁻⁴¹, MCR (*Multivariate curve resolution*)⁴², entre outros. O método GRAM é uma generalização do modelo RA (*Rank annihilation*). Este método analisa uma matriz de dados de uma amostra por comparação com uma matriz de dados de uma amostra padrão, onde a amostra padrão e a amostra desconhecida precisam apenas ter uma ou mais espécies em comum. Sua extensão é TLD que analisa os dados de duas ou mais amostras ao mesmo tempo, onde os dados de cada amostra estão arranjados numa matriz. Modelos como N-Tucker, N-PCA, PARAFAC1, PARAFAC2, N-PLS; entre outros, podem analisar simultaneamente dados de várias amostras, onde os dados de uma amostra podem ser de segunda ordem a mais.

4.2.1 - ESTRUTURA

As estruturas dos modelos quimiométricos de ordem superior anteriormente citados são lineares²⁶. Para alguns dados esta linearidade pode estar diretamente relacionada ao processo de aquisição de tais dados. Exemplo de dados lineares podem ser dados espectrais bem comportados e apropriadamente pré-processados. Entretanto, para certos tipos de dados, como dados sensoriais, existe pouca teoria da forma básica como eles são gerados. Neste caso os modelos podem freqüentemente ser usados devido a suas propriedades de aproximação.

No primeiro caso, a decomposição linear em vários modos (multilinear) pode ser vista como a resolução de um espectro, enquanto que, no segundo caso, podem atuar como um método característico de extração ou compressão e que elimina o ruído e informações não relevantes. Ajudado por um ponto de vista numérico e interpretacional.

Notar que quando algumas vezes é mencionado que um modelo estrutural é teoricamente verdadeiro, é uma via simplificada de dizer que existe alguma teoria que explica como os dados são gerados, similar a como os descreve o modelo.

Comumente tal teoria está baseada sobre um número de suposições. Por exemplo, a lei de Lambert Beer, estabelece que a absorvância de um analito é diretamente proporcional a sua concentração e está condicionada para analitos em soluções diluídas e para soluções concentradas são esperados desvios desta lei⁴⁴. Assim, o espectro UV/VIS de um analito não é um simples espectro, ele depende da diluição, temperatura, etc. Na prática, as análises de dados de interesse não envolve sempre incorporar todos os detalhes do problema. Para identificação por exemplo, uma descrição aproximada do perfil do espectro do analito é suficiente para identificá-lo.

4.2.2 - RESTRIÇÕES

As restrições⁴⁵⁻⁴⁷ podem ser aplicadas por várias razões, para identificar o modelo ou para que os resultados dos parâmetros do modelo tenham algum significado físico, ou seja, para que a solução seja interpretável e estável. O ajuste de um modelo restringido será sempre pior ou igual ao ajuste de um modelo sem restrição, mas se o modelo é interpretável e realista, isso pode justificar a perda da qualidade de ajuste.

Assim, restrições de ortogonalidade⁴⁷ em PCA são aplicados para identificação do modelo, isto é para diferenciá-lo de um simples modelo bilinear. Restrições de não negatividade são aplicados, porque sabe-se com antecedência que os parâmetros esperados, não devem ser negativos. Por exemplo, espera-se que *loadings* não sejam negativos quando se trata de dados espectrais que apresentam valores de absorvância positiva.

Não existe uma regra geral para escolher uma determinada estrutura e restrições adequadas para o modelamento. Assim, problemas específicos requerem soluções individuais.

4.2.3 - SOLUÇÃO ÚNICA

A solução única^{24,48} é um termo importante em análises de dados multidimensionais. Para a identificação de um modelo não é preciso adicionar restrições, mas quando se deseja que a solução do modelo obtida seja única, é necessário aplicar certas restrições.

Nos modelos bilineares²⁶ existe um problema conhecido de rotação, onde a decomposição da matriz de dados não é exclusiva, fornecendo uma infinidade de possíveis soluções. Em PCA, a livre rotacionalidade do modelo bilinear exige que certas restrições sejam aplicadas a sua solução, por exemplo, ortogonalidade.

4.3 - MODELO DE PARAFAC 1

4.3.1 - TEORIA

Este tipo de modelo foi independentemente proposto em 1970, pelos pesquisadores Carrol e Chang em Psicologia, que o chamaram de CAMDECOMP³³ (*Canonical Decomposition*), e por Harshman em 1970 que o chamou de PARAFAC³⁴ (*Parallel Factor Analysis*). Este modelo é uma generalização da Análise de Componentes Principais (PCA) bilinear para grupos de dados de ordem superior.

PARAFAC^{7,24,33,34} é um tipo de modelo matemático que serve para a decomposição de um conjunto de dados multidimensionais ordenados. Este modelo está restrito para dados que podem ser descritos com um modelo trilinear, e utiliza mínimos quadrados alternados (ALS - ver apêndice 1) para o cálculo de seus parâmetros (*loadings*). A decomposição trilinear é restrita só para o caso onde os vetores *loadings* são linearmente independentes entre si para dois modos e com pelo menos 2 vetores *loadings* não linearmente dependentes em um terceiro modo.

Assim, este modelo matemático decompõe os dados tridimensionais em tríades ou componentes trilineares. Cada componente consiste em um vetor *score* e dois vetores *loadings*. Deve-se ressaltar que se trata de forma semelhante os vetores *scores* e *loadings*.

O modelo de PARAFAC está intimamente relacionado ao princípio de "*parallel propotional profiles*"⁴⁹ introduzido por Cattell (1944). Este princípio estabelece que um conjunto de vetores *loadings* pertence simultaneamente a várias amostras. Por exemplo, um vetor *loading* que descreve o perfil espectral de um analito em uma amostra, também pertence a esse analito em outra amostra, além de que o perfil espectral multiplicado por uma determinada constante fornece o espectro de esse analito a uma determinada concentração.

4.3.2 - ESTRUTURA DO MODELO

O modelo trilinear de PARAFAC é expresso matematicamente pela seguinte equação:

$$\underline{\mathbf{X}} = \sum_{f=1}^F \mathbf{a}_f \otimes \mathbf{b}_f \otimes \mathbf{c}_f \quad (19)$$

onde:

- $\underline{\mathbf{X}}$ é o tensor tridimensional de dados calculado pelo modelo, cujas dimensões são: $I \times J \times K$.
- Os vetores *loadings* \mathbf{a}_f , \mathbf{b}_f , e \mathbf{c}_f correspondentes as f -ésimas colunas das matrizes *loadings* \mathbf{A} ($I \times F$), \mathbf{B} ($J \times F$) e \mathbf{C} ($K \times F$) respectivamente.
- F é o número de componentes utilizado para cálculo do modelo.

A decomposição de tensor de dados é dado pela seguinte equação:

$$\underline{\mathbf{X}}' = \sum_{f=1}^F \mathbf{a}_f \otimes \mathbf{b}_f \otimes \mathbf{c}_f + \underline{\mathbf{E}} \quad (20)$$

onde:

- $\underline{\mathbf{X}}'$ é o tensor tridimensional de dados originais.

- \underline{E} é o tensor residual obtido pela diferença do tensor de dados originais e do modelo trilinear.

O modelo é encontrado, com a condição de minimizar a soma dos quadrados dos elementos do tensor residual \underline{E} . O algoritmo para o cálculo do modelo de PARAFAC1 pode ser visto no apêndice 2.

A figura 13 mostra o gráfico de decomposição do tensor de dados por PARAFAC1 para três componentes.

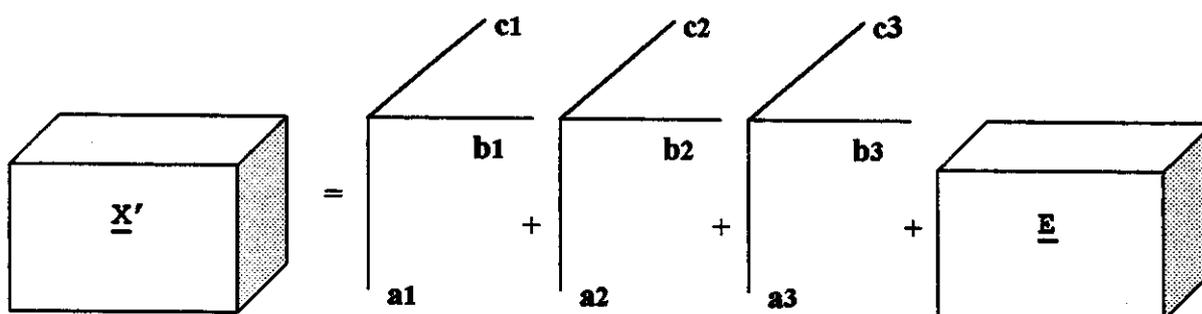


Figura 13: Esquema da decomposição de dados em *loadings* para 3 componentes por o modelo de PARAFAC1

Onde:

- Os vetores *loadings* a_1 , b_1 e c_1 descrevem os perfis do primeiro componente na primeira, segunda e terceira dimensão, ou modo, respectivamente.
- De forma semelhante a_2 , b_2 e c_2 descrevem os perfis do segundo componente na primeira, segunda e terceira dimensão, ou modo, respectivamente.
- E os vetores *loadings* a_3 , b_3 e c_3 descrevem os perfis do terceiro componente da primeira, segunda e terceira dimensão, ou modo, respectivamente.
- Os vetores colunas a_1 , a_2 e a_3 formam a matriz *loading* **A**, os vetores b_1 , b_2 e b_3 formam a matriz *loading* **B** e os vetores c_1 , c_2 e c_3 formam a matriz *loading* **C**.

O tensor de dados $\underline{\mathbf{X}}$ pode ser desdobrado numa matriz e pode ser decomposto pela seguinte somatória:

$$\mathbf{X} = \mathbf{a}_1 (\mathbf{c}_1 \otimes \mathbf{b}_1)^T + \mathbf{a}_2 (\mathbf{c}_2 \otimes \mathbf{b}_2)^T + \dots + \mathbf{a}_f (\mathbf{c}_f \otimes \mathbf{b}_f)^T \quad (21)$$

onde:

- \mathbf{X} é uma matriz de dimensões $I \times JK$, originada pelo desdobramento de $\underline{\mathbf{X}}$.

Esta equação pode ser reduzida à seguinte expressão:

$$\mathbf{X} = \sum_{f=1}^F \mathbf{a}_f (\mathbf{c}_f \otimes \mathbf{b}_f)^T = \mathbf{A} (\mathbf{C} \otimes \mathbf{B})^T \quad (22)$$

Uma fatia \mathbf{X}_k do tensor $\underline{\mathbf{X}}$, contendo informação só do componente f , pode ser calculada como:

$$(\mathbf{X}_k)_f = \mathbf{a}_f \text{diag}(\mathbf{c}_f(k,1)) \mathbf{b}_f^T \quad (23)$$

onde:

- $(\mathbf{X}_k)_f$ é uma fatia de dimensão $I \times J$, que pertence à k variável do terceiro modo do tensor de dados e tem informação apenas do componente f .
- \mathbf{a}_f , e \mathbf{b}_f são os vetores *loadings* do componente "f" para o primeiro e segundo modo de \mathbf{X} .
- A expressão $\text{diag}(\mathbf{c}_f(k,1))$ fornece a diagonal formada pelo elemento da k -ésima linha do vetor *loading* \mathbf{c}_f , correspondente ao componente f .

A equação 23 pode ser generalizada como:

$$\mathbf{X}_k = \mathbf{A} \mathbf{D}_k \mathbf{B}^T \quad (24)$$

onde:

- \mathbf{D}_k é a matriz diagonal formada pelos elementos da k -ésima linha da matriz *loading* \mathbf{C} .
- \mathbf{X}_k é uma fatia para k -ésima variável do terceiro modo e que fornece informação de todos os componentes.

Fazendo:

$$G = AD_k \quad (25)$$

Tem-se:

$$X_k = G B^T \quad (26)$$

Assim, para cada k-ésima variável (do terceiro modo) um modelo bilinear é obtido de X_k . Estes modelos serão proporcionais porque cada componente está pesado diferentemente através da matriz D_k . Deste modo, demonstra-se como o modelo de PARAFAC1 está relacionado ao princípio de "*parallel proportional profiles*".

4.3.3 - PARAFAC1 COM REGRESSÃO LINEAR

Depois de se obter o modelo de PARAFAC1, pode-se utilizar os vetores *loadings* que fornecem informação de concentração relativa de cada analito, para estimar suas respectivas em todas as amostras de calibração. Isto pode ser feito a partir de uma regressão linear entre um vetor coluna contendo as concentrações verdadeiras de um analito nas diferentes amostras (af_e) e o vetor *loading* que contém informação das concentrações relativas deste analito nas diferentes amostras (por exemplo o vetor *loading* af), estabelecendo-se a seguinte equação:

$$af_e = af \beta + ef \quad (27)$$

onde:

- af_e é o vetor que contém as concentrações verdadeiras do analito f.
- af é o vetor *loading* que contém as concentrações relativas do analito f.
- β é o vetor que contém os coeficientes da regressão.
- ef é o vetor resíduo.

O vetor " β " é calculado pelo método dos mínimos quadrados utilizando-se a seguinte equação:

$$\beta = (af^T af)^{-1} (af^T af_e) \quad (28)$$

4.4 - MODELO DE PARAFAC 2

4.4.1 - TEORIA

Em alguns casos, não se pode aplicar o modelo de PARAFAC1 a um conjunto de dados que forma um arranjo tridimensional, por este não ser trilinear. A perda da trilinearidade pode ser provocada por problemas de amostragem, fatores físicos ou químicos. Também não se pode utilizar o PARAFAC1, em casos onde se têm diferentes dimensões nas fatias de um conjunto de dados (seja na linha ou na coluna). Alguns desses problemas podem ser solucionados com o uso do PARAFAC2.³⁵⁻³⁷

Por exemplo, o PARAFAC1 pode ser utilizado para um conjunto de dados cinéticos, com detecção espectral na região ultravioleta- visível, arranjado como comprimento de onda (I) x tempo (J) x amostras (K). Este modelo fornecerá como resposta a matriz *loadings* **A** (perfis espectrais dos analitos), **B** (perfis das curvas cinéticas) e **C** (perfis das concentrações relativas para cada analito). Para tal caso, assume-se que os perfis cinéticos de um analito "f" para várias amostras é constante. Para situações em que o perfil cinético de um analito "f" varia para diferentes amostras, o PARAFAC2 é o modelo matemático mais adequado para o modelamento desses dados.

4.4.2 - ESTRUTURA DO MODELO

O modelo de PARAFAC2 define cada fatia de um tensor de dados tridimensionais como:

$$\mathbf{X}_k = \mathbf{A} \mathbf{D}_k (\mathbf{B}_k)^T \quad (29)$$

onde:

- \mathbf{X}_k é um fatia de dimensão I x J, o qual pertence à k-ésima variável do terceiro modo do tensor tridimensional de dados $\underline{\mathbf{X}}$ do modelo e que fornece informação para todos os "F" componentes.

- A é uma matriz *loading* de dimensão $I \times F$, que contém F vetores *loadings*. Cada vetor *loading* fornece o perfil de cada componente no primeiro modo do tensor tridimensional de dados do modelo \underline{X} .
- D_k é a matriz diagonal formada pelos elementos da k -ésima linha da matriz *loading* C .
- C é uma matriz *loading* de dimensão $K \times F$, que contém F vetores *loadings*. Cada vetor *loading* fornece o perfil de cada componente no terceiro modo do tensor tridimensional de dados do modelo \underline{X} .
- B_k é uma matriz *loading* de dimensão $J \times F$, que contém F vetores *loadings*. Onde, cada vetor *loading* fornece o perfil de cada componente no segundo modo do tensor tridimensional de dados do modelo \underline{X} , para cada variável k .

A matriz B_k é calculada pelo modelo de PARAFAC2 como:

$$B_k = P_k B \quad (30)$$

onde:

- B é uma matriz constante de dimensão $F \times F$.
- P_k tem dimensão $J \times F$.

Substituindo a equação 30 em 29 tem-se:

$$X_k = A D_k (P_k B)^T \quad (31)$$

Definindo:

$$Y_k = X_k P_k \quad (32)$$

Resolvendo o produto de matrizes tem-se:

$$Y_k = A D_k B^T P_k^T P_k \quad (33)$$

O modelo de PARAFAC 2 assume que:

$$P_k^T P_k = I \quad (34)$$

Usando esta condição na equação 33, é obtida a seguinte expressão matemática:

$$Y_k = A D_k B^T \quad (35)$$

O agrupamento de varias matrizes Y_k , forma uma matriz Y de dimensões $I \times FK$, e esta nova matriz pode ser modelada pelo PARAFAC 1.

Desde que o algoritmo geral de PARAFAC 2 (ver apêndice 3) modela diretamente as fatias Y_k e não as fatias X_k , é possível ter matrizes X_k de diferentes dimensões.

Multiplicando por B_k^T na equação 30, e usando a condição da equação (34), temos:

$$B_k^T B_k = B^T B = H \quad (36)$$

Isto quer dizer que as matrizes B_k que contém os perfis do segundo modo do tensor de dados não precisam ser semelhantes, só que H seja constante para qualquer produto $B_k^T B_k$. Esta é a diferença fundamental com o PARAFAC1.

A figura 14 mostra a decomposição de um tensor de dados tridimensional \underline{X} pelo PARAFAC2:

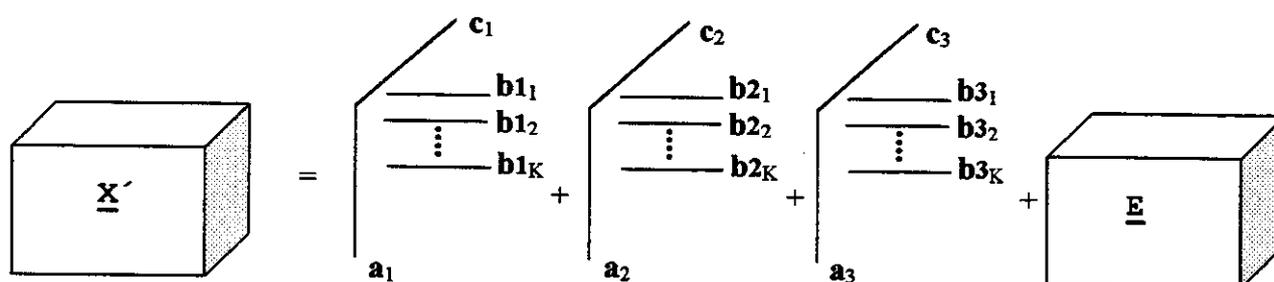


Figura 14: Decomposição do tensor de dados por PARAFAC2 para três componentes.

onde:

- \underline{X} é o tensor tridimensional de dados originais de dimensões $I \times J \times K$.
- Os vetores *loadings* \mathbf{a}_1 e \mathbf{c}_1 descrevem os perfis do primeiro componente na primeira, e terceira dimensão, ou modo, respectivamente.
- Os vetores *loadings* $\mathbf{b}_{11}, \mathbf{b}_{12}, \dots, \mathbf{b}_{1K}$ descrevem os perfis do primeiro componente na segunda dimensão.
- Em forma similar, os vetores *loadings* \mathbf{a}_2 e \mathbf{c}_2 descrevem os perfis do segundo componente na primeira, e terceira dimensão ou modo respectivamente.
- Os vetores *loadings* $\mathbf{b}_{21}, \mathbf{b}_{22}, \dots, \mathbf{b}_{2K}$ descrevem os perfis do segundo componente na segunda dimensão.
- Os vetores *loadings* \mathbf{a}_3 e \mathbf{c}_3 descrevem os perfis do terceiro componente na primeira, e terceira dimensão, ou modo, respectivamente.
- Os vetores *loadings* $\mathbf{b}_{31}, \mathbf{b}_{32}, \dots, \mathbf{b}_{3K}$, fornecem os perfis do terceiro componente na segunda dimensão.
- Os vetores colunas $\mathbf{a}_1, \mathbf{a}_2$ e \mathbf{a}_3 formam a matriz *loading* \mathbf{A} , e os vetores $\mathbf{c}_1, \mathbf{c}_2$ e \mathbf{c}_3 formam a matriz *loading* \mathbf{C} .

A equação 20 pode ser reescrita para o cálculo da fatia \mathbf{X}_k para um componente "f", como:

$$(\mathbf{X}_k)_f = \mathbf{a}_f \text{diag}(\mathbf{c}_f(k,1)) \mathbf{b}_k^T, \quad (37)$$

onde:

- \mathbf{a}_f , e \mathbf{c}_f são os vetores *loadings* do componente f, correspondentes as f-ésimas colunas das matrizes *loadings* \mathbf{A} e \mathbf{C} , respectivamente.
- \mathbf{b}_k é o vetor *loading* do componente f, correspondente à f-ésima coluna da matriz *loading* \mathbf{B}_k .
- A expressão $\text{diag}(\mathbf{c}_f(k,1))$ é a diagonal formada pelo elemento da k-ésima linha do vector *loading* \mathbf{c}_f .

4.5 – MODELO DE N-PLS

4.5.1 - TEORIA

A teoria e vantagens do modelo PLS foram inicialmente apresentadas por Martens e Naes⁵⁰ em 1989 e por Jong e Phatak⁵¹ em 1997. A principal vantagem é que os dados das variáveis independentes são modelados simultaneamente com os dados das variáveis dependentes.

Em 1989 Ståhle⁵² ampliou o modelo de regressão PLS tipicamente conhecido, para dados tridimensionais, mas a otimização do algoritmo proposta não foi comprovada. Em 1996 Bro desenvolveu um modelo de regressão multidimensional, que o chamou de N-PLS⁴⁰ e comprovou a otimização de seu algoritmo, baseado nas teorias fundamentais do PLS.

O N-PLS⁴⁰⁻⁴¹ (*Multiway partial least squares*) é um método para construir um modelo de regressão, isto é, um modelo que faça relação entre o conjunto de dados de variáveis independentes (denotado como X para uma matriz e \underline{X} para um tensor de terceira ordem) e conjunto de dados de variáveis dependentes (denotado como Y para uma matriz e \underline{Y} para um tensor de terceira ordem).

Por exemplo, um conjunto de dados cinéticos espectrofotométricos de várias misturas podem ser arranjados em um tensor de terceira ordem (\underline{X}) e os dados das concentrações reais dos analitos presentes em cada amostra podem ser arranjados em uma matriz (Y).

Com o N-PLS pode-se prever as concentrações dos analitos presentes de novas amostras (dados de variáveis dependentes), com a informação do modelo de regressão construído e os dados das variáveis independentes desse novo conjunto de amostras.

4.5.2 - ESTRUTURA DO MODELO

O seguinte modelo será descrito para um conjunto de dados de variáveis independentes arranjados em um tensor de terceira ordem ($\underline{\mathbf{X}}$) e para um conjunto de dados de variáveis dependentes arranjado numa matriz (\mathbf{Y}), sendo que esta última fornece as concentrações de "P" analitos.

No N-PLS o tensor $\underline{\mathbf{X}}$ e a matriz \mathbf{Y} são decompostos em *scores* e *loadings*, com a restrição de que os vetores *scores* de $\underline{\mathbf{X}}$ e \mathbf{Y} da mesma variável latente são combinados por uma relação interna entre eles. Assim, em N-PLS o tensor $\underline{\mathbf{X}}$ pode ser decomposto na soma de fatores como segue:

$$\mathbf{X} = \mathbf{t}_1 (\mathbf{w}_1^K \otimes \mathbf{w}_1^J)^T + \mathbf{t}_2 (\mathbf{w}_2^K \otimes \mathbf{w}_2^J)^T + \dots + \mathbf{t}_f (\mathbf{w}_f^K \otimes \mathbf{w}_f^J)^T + \mathbf{E} \quad (38)$$

onde:

- \mathbf{X} é uma matriz de dimensões $I \times JK$, obtida pelo desdobramento do tensor de dados tridimensional $\underline{\mathbf{X}}$ de dimensões $I \times J \times K$, que tem I variáveis no primeiro modo, J variáveis no segundo modo e K variáveis no terceiro modo.
- \mathbf{t}_f é o vetor *score* para a f -ésima variável latente, com dimensões $I \times 1$. Por exemplo, \mathbf{t}_1 é o vetor *score* para a primeira variável latente, com dimensões $I \times 1$.
- \mathbf{w}_f^K é um vetor *loading* para a variável latente f . Este vetor tem medida $K \times 1$.
- \mathbf{w}_f^J é um vetor *loading* para a variável f . Este vetor tem medida $J \times 1$.
- \mathbf{E} é uma matriz que contém os resíduos da matriz \mathbf{X} ao ser modelada. Esta matriz tem dimensões $I \times JK$.

A equação 38 pode ser simplificada como:

$$\mathbf{X} = \mathbf{T} (\mathbf{W}^K \otimes \mathbf{W}^J)^T + \mathbf{E} \quad (39)$$

onde:

- T é a matriz contendo os " f " vetores *escores*. Assim, esta matriz tem dimensões $I \times f$.
- W_f^K é uma matriz *loading* contendo os f vetores *loading* do terceiro modo, com dimensão $K \times f$.
- W_f^J é uma matriz *loading* contendo os f vetores *loading* do segundo modo, com dimensão $J \times f$.

Uma matriz *loading* P_f pode ser definida como:

$$P_f = \text{reshape}((w_f^K \otimes w_f^J)^T, 1, J, K) \quad (40)$$

onde:

- P_f é a matriz *loading*, com dimensões $1 \times J \times K$. Por exemplo, P_1 é a matriz *loading* para a primeira variável latente, com dimensões $1 \times J \times K$.
- A função *reshape* arranjou o produto kronecker dos vetores w_f^K e w_f^J em uma matriz de dimensão $1 \times J \times K$.
- O sobrescrito J e K é usado para especificar a qual modo os vetores se referem.

Substituindo a equação 40 na equação 38 tem-se:

$$X = t_1 P_1 + t_2 P_2 + \dots + t_f P_f + E \quad (41)$$

A figura 15 mostra a decomposição do tensor \underline{X} de terceira ordem usando 3 variáveis latentes.

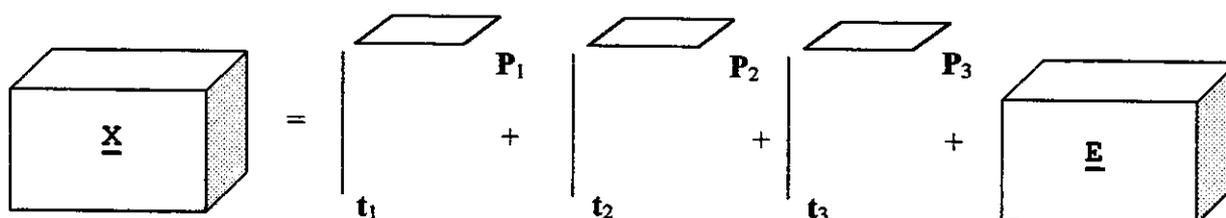


Figura 15: Decomposição do tensor de dados X por N-PLS usando 3 variáveis latentes.

No N-PLS a matriz Y é decomposta na soma dos produtos dos seguintes fatores:

$$Y = u_1 q_1^T + u_2 q_2^T + \dots + u_f q_f^T + F \quad (42)$$

onde:

- Y é a matriz de dados com dimensões $I \times P$, que tem I amostras no primeiro modo e P analitos no segundo modo.
- u_f é o vetor *score* para a f -ésima variável latente, com dimensões $I \times 1$. Por exemplo, u_1 é o vetor *score* para a primeira variável latente, com dimensões $I \times 1$.
- q_f é o vetor *loading*, com dimensões $P \times 1$. Por exemplo, q_1 é o vetor *loading* para a primeira variável latente, com dimensões $P \times 1$.
- F é a matriz que contém os resíduos da matriz Y ao ser modelado, com dimensão $I \times P$.

A equação 42 pode ser re-escrita como:

$$Y = UQ^T + F \quad (43)$$

onde:

- U é a matriz contendo os f vetores *scores*, com dimensão $I \times F$.
- Q é a matriz contendo os f vetores *loadings*, com dimensão $P \times F$.

A figura 16 mostra a decomposição para 3 componentes da matriz Y (dados dependentes) por N-PLS:

$$\boxed{Y} = \begin{matrix} \overline{} \\ | \\ \end{matrix} \begin{matrix} q_1 \\ + \\ \end{matrix} \begin{matrix} \overline{} \\ | \\ \end{matrix} \begin{matrix} q_2 \\ + \\ \end{matrix} \begin{matrix} \overline{} \\ | \\ \end{matrix} \begin{matrix} q_3 \\ + \\ \end{matrix} \boxed{F}$$

Figura 16: Decomposição da matriz de dados Y por N-PLS usando 3 variáveis latentes.

A relação entre a matriz *score* U , com a matriz *score* T , é dada pelo seguinte modelo linear:

$$U = TB \quad (44)$$

onde:

- B é o coeficiente de regressão que relaciona T e U .

Para estimar o valor das variáveis dependentes para um novo conjunto de dados ($Y^{\text{nov}}o$), usa-se o modelo construído e os dados das variáveis independentes desse novo conjunto de amostras ($X^{\text{nov}}o$). Assim, calcula-se primeiro sua matriz *score* $T^{\text{nov}}o$ a partir do arranjo dos f vetores t_f obtida das seguintes equações:

$$t_1 = X^{\text{nov}}o (w_1^K \otimes w_1^J) \quad (45)$$

$$t_2 = (X^{\text{nov}}o - t_1 (w_1^K \otimes w_1^J)^T) (w_2^K \otimes w_2^J) \quad (46)$$

$$t_f = (X^{\text{nov}}o - t_1 (w_1^K \otimes w_1^J)^T \dots - t_{f-1} (w_{f-1}^K \otimes w_{f-1}^J)^T) (w_f^K \otimes w_f^J) \quad (47)$$

Onde f varia de 1 até F , que é o número de variáveis latentes usadas na construção do modelo de calibração e os vetores *loading* w_f são conhecidos do modelo de calibração.

Calcula-se a matriz *score* $U^{\text{nov}}o$ pela equação (48), sendo que B é conhecido do modelo de calibração.

$$U^{\text{nov}}o = T^{\text{nov}}o B \quad (48)$$

Finalmente calcula-se o novo conjunto de dados $Y^{\text{nov}}o$ pela equação (49), utilizando o *loading* Q e a matriz B calculados no modelo de calibração.

$$Y^{\text{nov}}o = T^{\text{nov}}o B Q^T \quad (49)$$

Notação para os modelos de N-PLS

Os modelos multilíneares N-PLS recebem diferentes nomes. A tabela 1, apresenta os nomes destes modelos de acordo com a ordem dos dados das variáveis independentes X e a ordem dos dados das variáveis dependentes Y. Um número escrito depois de PLS define ordem de Y e um prefixo em PLS indica a ordem de X.

Tabela 1: Abreviações para diferentes modelos de N-PLS, de acordo à ordem de X e Y.

Ordem de Y	Ordem de X		
	2	3	4
1	Bi-PLS1	Tri-PLS1	Quadri-PLS1
2	Bi-PLS2	Tri-PLS2	Quadri-PLS2
3	Bi-PLS3	Tri-PLS3	Quadri-PLS3

CAPÍTULO 5

PARTE EXPERIMENTAL



PARTE EXPERIMENTAL

5.1 – OBTENÇÃO DA METODOLOGIA CINÉTICA ESPECTROFOTOMÉTRICA PARA A ANÁLISES DOS ISÔMEROS DE AMINOFENOL

A análise de uma mistura ternária de isômeros *orto*, *meta* e *para* aminofenol foi estudada e como sua separação é difícil, a identificação e quantificação direta é complexa. Usou-se a técnica cinética espectrofotométrica em combinação com os métodos quimiométricos de ordem superior para sua análise.

Para desenvolver a metodologia cinética espectrofotométrica investigaram-se reações entre os isômeros de aminofenol com um reagente em comum que fornecessem produtos de reação estáveis.

Sabe-se que, os isômeros *o*-aminofenol, *m*-aminofenol e *p*-aminofenol são aminas aromáticas primárias. Estas aminas aromáticas primárias têm a característica de reagir com ácido nitroso HNO_2 mediante reação chamada diazotação¹³, formando sais areno diazônios. Como o ácido nitroso é instável, sua preparação é feita no momento da reação, misturando nitrito de sódio com uma solução ácida.

Desenvolveu-se uma metodologia experimental para o monitoramento das reações mencionadas. Preparou-se soluções estoques dos isômeros aminofenol em meio aquoso. Observou-se que as soluções preparadas incolores se decompunham com o tempo adquirindo coloração escura. Esta decomposição era mais lenta, se as soluções estavam protegidas contra a luz.

Procurou-se manter estáveis as soluções estoques preparadas, pelo maior tempo possível. Inicialmente, procurou-se determinar qual era o meio mais adequado para dissolver e manter estáveis os sais. Assim, dissolveram-se os isômeros em solventes preparados em diferentes proporções de água e etanol, mas a estabilidade das soluções dos isômeros não foi observada. Outra tentativa foi preparar soluções dos isômeros em um meio ligeiramente ácido com HCl. Sob estas condições as soluções dos isômeros permaneceram estáveis. Acredita-se que nestas condições forma-se sal de aminofenol $\text{NH}_4^+ \text{C}_6\text{H}_4 \text{OH}$. Por precaução, as soluções continuaram a ser mantidas sob proteção contra a luz.

O passo seguinte foi determinar qual era o pH mais adequado, a concentração ideal de nitrito de sódio e as faixas das concentrações dos isômeros para monitorar as reações.

5.1.1 – FUNDAMENTOS PARA A ESCOLHA DOS PARÂMETROS PARA O MONITORAMENTO DA REAÇÃO ENTRE AMINOFENOL E NITRITO DE SÓDIO

Os parâmetros utilizados para o monitoramento da reação entre aminofenol e nitrito de sódio em meio ácido foram escolhidos de modo a facilitar o monitoramento cinético. Isto é, que a reação não seja muito rápida, de tal modo de ter informação suficiente da cinética desenvolvida pela reação.

Estudos preliminares do efeito do pH na velocidade da reação foram realizados, preparando-se várias amostras contendo um só isômero e fazendo-as reagir com nitrito de sódio $0,013 \text{ mol L}^{-1}$. Estas reações foram repetidas para diferentes pH e os resultados indicaram que a medida que o pH aumentava a reação tornava-se lenta. Observou-se que em pH igual a 1 e 2 a reação foi muito rápida (menos de 1s para ela se completar), inviabilizando o monitoramento. Em pH=5 as soluções apresentavam uma cor amarelo forte. Sabe-se da literatura que o produto formado por uma reação de diazotação reage com o aminofenol em $\text{pH} \geq 5$ por reação chamada de acoplamento diazo, formando um composto azo. De acordo

com isto, o pH mais adequad, e escolhido para ser usado no monitoramento da reação entre uma mistura de isômeros e aminofenol foi 4.

A concentração de nitrito de sódio ($0,013 \text{ mol L}^{-1}$) foi escolhida para que esteja em excesso com respeito à concentração de aminofenol, com a finalidade de que a reação entre aminofenol e nitrito de sódio seja completa, isto é, para garantir que todo o aminofenol reagisse com o nitrito.

As concentrações de cada isômero nas amostras foram dimensionadas para que em todo o momento do monitoramento, a absorbância da mistura não ficasse acima de 1, e que a absorbância do produto formado por cada isômero estivesse relacionado à lei de Lambert Beer, sem problemas de desvios desta lei.

5.2 - Reação Química entre nitrito de sódio e os isômeros aminofenol em medio ácido

Na reação de nitrito de sódio com *o*-, *m*- e *p*-aminofenol, em meio ácido são formados sais areno diazônios . Estes sais apresentam espectros um pouco diferentes na região UV/VIS, o que permite a quantificação dos isômeros reagentes. A figura 17 mostra as reações químicas entre os isômeros aminofenol e nitrito de sódio em meio ácido. Pode-se observar, nesta figura, que as estruturas dos produtos formados apresentam orbitais π conjugados procedentes das duplas ligações $C = C$ no anel aromático, e a tripla ligação $N = N$.

5.3 – Equipamentos utilizados para o monitoramento das reações

Foram utilizados os seguintes equipamentos:

- Um pH metro CORNING pH/ion analyzer 350 com eletrodo combinado de vidro para o ajuste do pH das soluções.
- Um espectrofotômetro HP 8452A com arranjo de diodos e lâmpada de deutério, acoplado a um sistema "Peltier" que termostatiza a cela de reação, a uma temperatura fixa previamente estabelecida.

- Uma cubeta de quartzo de 1cm de caminho ótico para cela de reação, onde se misturam os reagentes.
- Um banho termostatizado para manter a uma temperatura fixa as soluções reagentes. A termostatização é importante para a reprodutibilidade dos resultados.
- Um microcomputador IBM 486 para o controle do espectrofotômetro.
- Um microcomputador IBM Pentium de 200 MHz. para realização dos cálculos.
- Um equipamento para injeção simultânea de reagentes.

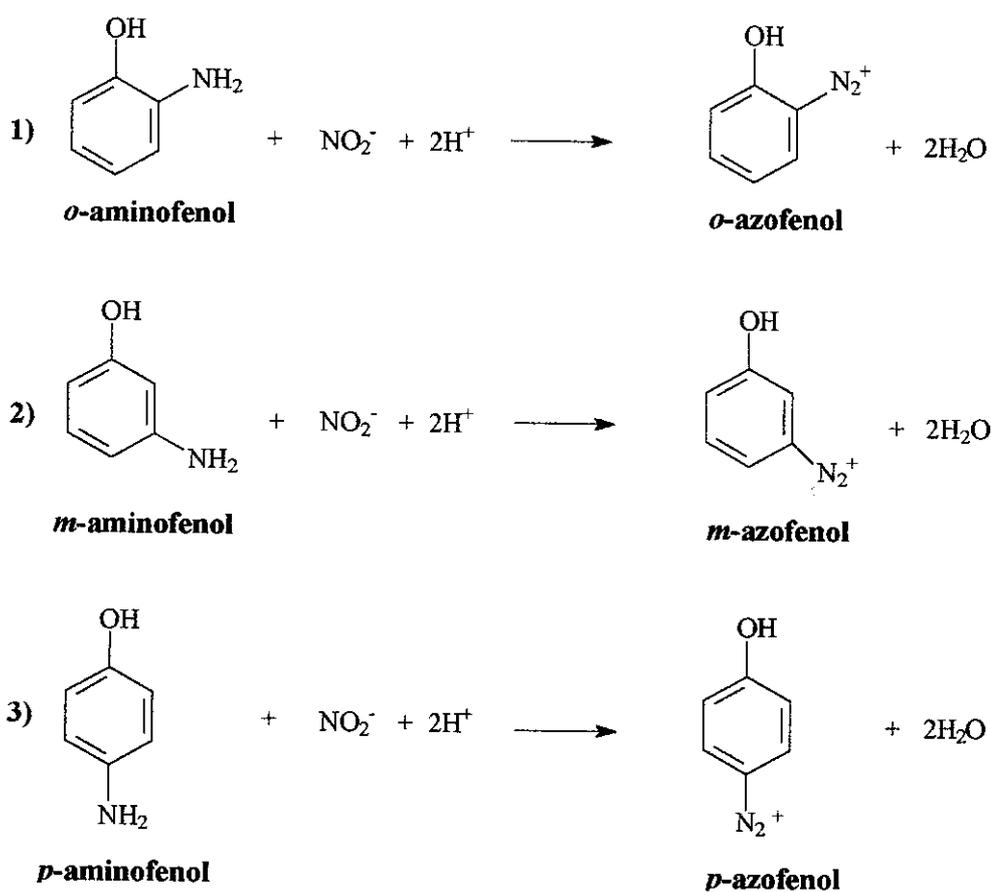


Figura 17: Reações químicas entre nitrito de sódio e os isômeros aminofenol.

5.3.1 - Espectrofotômetro com arranjo de diodos HP 8452A

Este instrumento foi utilizado para fornecer informação espectral na região ultravioleta – visível em função do tempo. A figura 18 mostra um esquema do sistema óptico do espectrofotômetro com lâmpada de deutério.⁵³ A lâmpada de deutério é utilizada para os estudos espectrais na região ultravioleta/visível.

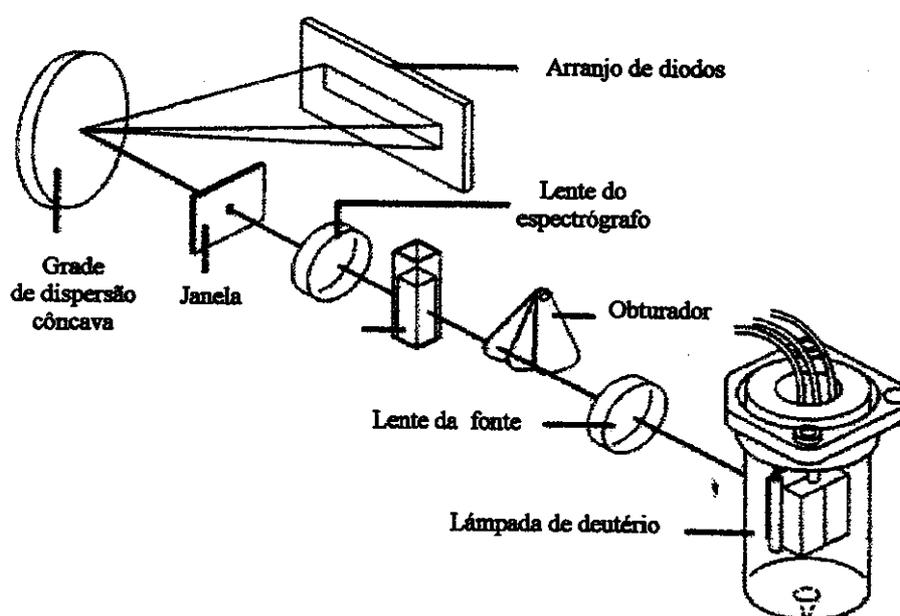


Figura 18: Sistema óptico do espectrofotômetro com lâmpada de deutério.

Partes do sistema óptico:

a) Lâmpada de deutério

Esta lâmpada é uma fonte de luz na região ultravioleta – visível na faixa de 190 a 820 nm. A fonte é gerada pela formação do plasma originado pela descarga elétrica sobre o deutério que se encontra a baixa pressão dentro do bulbo.

b) Lente da fonte

A lente da fonte recebe a luz da lâmpada e a faz convergir de tal modo que passe pela amostra, que se encontra no compartimento apropriado.

c) Obturador

O obturador é eletromecanicamente operado. Quando está aberto, permite a passagem da luz através da amostra e quando está fechado, permite ao espectrofotômetro medir a corrente gerada pelo detector na ausência de radiação.

d) Compartimento da amostra

Este espectrofotômetro tem um compartimento de amostra aberta, o que torna possível o acesso à cela de maneira mais fácil. No compartimento da amostra, coloca-se a cela de reação, que deve ser de quartzo para medidas na região ultravioleta- visível. O compartimento da cela é termostaticado a uma temperatura fixa por um sistema *Peltier*.

e) Lente do espectrógrafo

A lente do espectrógrafo focaliza sobre o detector (arranjo de diodos) o feixe de luz que atravessou a amostra.

f) Janela

Formada por uma chapa metálica que tem um pequeno orifício central. Está localizada no foco da lente do espectrógrafo de tal modo que o raio de luz procedente da lente passe pelo orifício.

g) Grade de dispersão

A dispersão da luz procedente da lente do espectrógrafo é obtida pela grade côncava. A grade separa a luz nos diferentes comprimentos de onda e o reflete sobre o arranjo de diodos.

h) Arranjo de diodos

O arranjo de diodos consiste de uma série de 328 fotodiodos individuais alinhados em série, sendo que uma faixa estreita de comprimentos de onda incide sobre cada fotodiodo.

5.3.2 - Sistema de injeção simultânea de reagentes

Foi utilizado um sistema de injeção simultânea⁵⁴ (figura 19) para a adição e mistura dos reagentes e posterior acompanhamento espectrofotométrico da reação. O sistema de injeção simultânea tem um suporte de nylon (a) onde são fixadas as seringas de vidro (b) para injeção de reagentes. Frascos de vidro (c) servem como reservatórios dos reagentes. Cada frasco está conectado a uma torneira de 3 vias de teflon (d). Os êmbolos das seringas de reagentes são fixos no suporte (e) que desliza sob o eixo (f), impulsionado pela alavanca manual (g) localizada no êmbolo para impulsão (h). A alavanca manual controla o fluxo que preenche a cela de reação (i), através de tubos de polietileno (j).

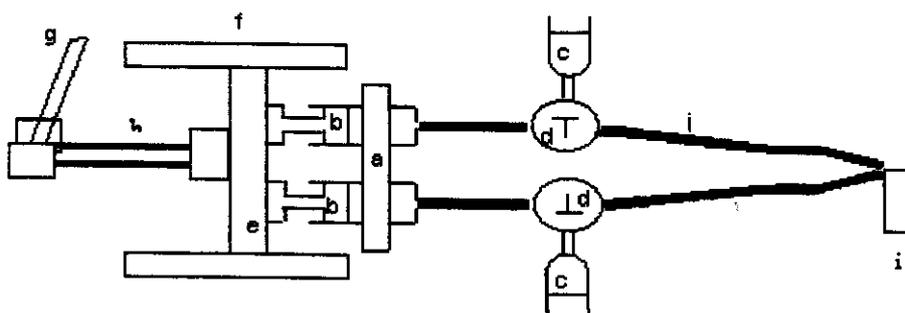


Figura 19: sistema de injeção simultânea de reagentes.

5.4 - Preparação das soluções reagentes e das amostras

Durante todo o processo de preparo dos reagentes foi utilizado água desionizada (Milli Q).

- a) Solução estoque de *o*-aminofenol $4,583 \times 10^{-3} \text{ mol L}^{-1}$ P.A. (Carlo Erba): 0,02500 g dissolvidos em 0,5 mL de HCl 6 mol L^{-1} e completado seu volume até 50,0 mL. Esta solução era preparada imediatamente antes do uso.

- b) Solução estoque de *m*-aminofenol $4,583 \times 10^{-3} \text{ mol L}^{-1}$ P.A. (British Drug Houses): 0,02500 g dissolvidos em 0,5 mL de HCl 6 mol L^{-1} e completado seu volume até 50,0 mL. Esta solução era preparada imediatamente antes do uso.
- c) Solução estoque de *p*-aminofenol $1,145 \times 10^{-3} \text{ mol L}^{-1}$ P.A. (British Drug Houses): 0,01250 g dissolvidos em 1,0 mL de HCl 6 mol L^{-1} e completado seu volume até 100,0 mL. Esta solução era preparada imediatamente antes do uso.
- d) Solução de nitrito de sódio $0,013 \text{ mol L}^{-1}$ P.A. (Mallinckrodt): Esta solução era preparada imediatamente antes do uso no momento da análise.
- e) Preparação da solução tampão Universal⁵⁵
Solução a: $0,078 \text{ mol L}^{-1}$ de ácido cítrico P.A. (Quimibrás) e $0,310 \text{ mol L}^{-1}$ de ácido bórico P.A. (Merck).
Solução b: $0,045 \text{ mol L}^{-1}$ de fosfato de sódio P.A. (Merck).
Para a preparação do tampão, misturaram-se proporções de 155 mL da solução (a) com 45 mL da solução (b).
- f) **Preparação das amostras:**
As amostras foram preparadas misturando soluções dos isômeros em um balão volumétrico de 25 mL, adicionando 20,0 mL de solução tampão Universal, completando-se o volume com solução $0,06 \text{ mol L}^{-1}$ de HCl.
O branco foi preparado misturando 20 ml da solução tampão com 5 ml de HCl $0,06 \text{ mol L}^{-1}$.

5.5 - Planejamento experimental

Foram preparadas 3 amostras puras, contendo o isômero *o*-, *m*- e *p*-aminofenol respectivamente, e 32 amostras contendo os três isômeros em diferentes concentrações. As concentrações iniciais dos isômeros na cela reação para cada amostra estão apresentadas na tabela 2.

Tabela 2 : Concentrações das misturas dos isômeros

Composição das misturas	<i>o</i> -aminofenol ppm*	<i>m</i> -aminofenol ppm	<i>p</i> -aminofenol ppm
1	10,00	0,00	0,00
2	0,00	16,00	0,00
3	0,00	0,00	2,50
4	5,00	12,00	1,25
5	5,00	12,00	2,50
6	5,00	12,00	3,75
7	5,00	16,00	1,25
8	5,00	16,00	2,50
9	5,00	16,00	3,75
10	5,00	20,00	1,25
11	5,00	20,00	2,50
12	5,00	20,00	3,75
13	10,00	12,00	1,25
14	10,00	12,00	2,50
15	10,00	12,00	3,75
16	10,00	16,00	1,25
17	10,00	16,00	2,50
18	10,00	16,00	3,75
19	10,00	20,00	1,25
20	10,00	20,00	2,50
21	10,00	20,00	3,75
22	15,00	12,00	1,25
23	15,00	12,00	2,50
24	15,00	12,00	3,75
25	15,00	16,00	2,50
26	15,00	16,00	3,75
27	15,00	20,00	1,25
28	15,00	20,00	2,50
29	15,00	20,00	3,75
30	7,00	14,00	1,75
31	12,00	18,00	3,00
32	7,00	18,00	3,00
33	12,00	14,00	1,75
34	7,00	14,00	3,00
35	12,00	18,00	1,75

* Usou-se ppm em vez de mol L⁻¹ para simplificar a quantidade de dígitos e letras usadas para denotar a concentração. Onde, 1 ppm ≈ 9,166 x 10⁻⁶ mol L⁻¹.

5.6 - Aquisição de dados a partir do monitoramento das reações cinéticas

Todas as soluções reagentes foram termostalizadas a $25,0 \pm 0,1^{\circ}\text{C}$ por 1 hora. Foram misturados simultaneamente volumes iguais da solução contendo os isômeros e da solução de NaNO_2 , usando o sistema de injeção simultânea. As reações foram realizadas em uma cela de quartzo de 1cm de caminho ótico colocada em um espectrofotômetro HP 8452A com arranjo de diodos. Um sistema "Peltier" foi utilizado para manter constante a temperatura da cela a 25°C durante o monitoramento. Foram obtidos espectros na região de 276-488 nm com resolução de 2nm e tempo de integração de 0,1 s, em intervalos de 3s com tempo inicial de monitoramento de 1s depois da injeção dos reagentes na cela, até 28 s. A figura 20 mostra o gráfico dos espectros obtidos no monitoramento da cinética de uma amostra ternária de isômeros aminofenol e nitrito de sódio.

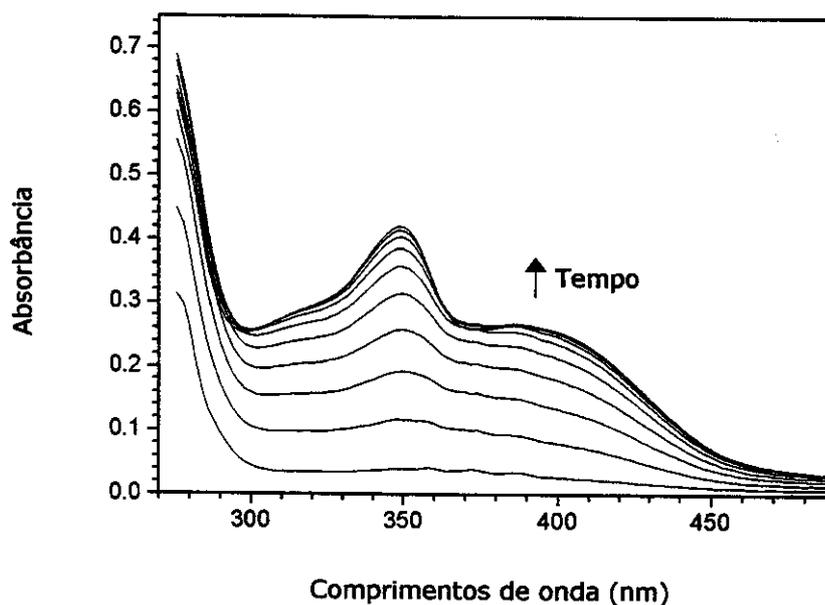
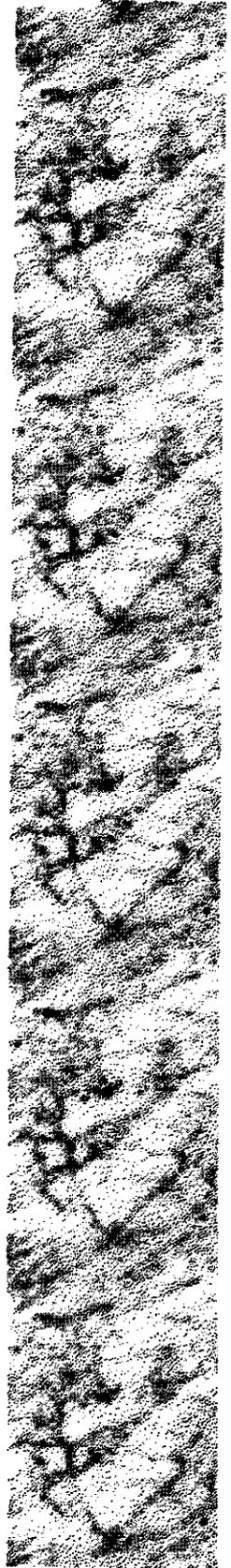


Figura 20: Ilustração de um conjunto de dados tipicamente obtidos no monitoramento da reação de amino fenol com NaNO_2 , nas condições estudadas neste trabalho.

CAPÍTULO 6

RESULTADOS E DISCUSSÃO



RESULTADOS E DISCUSSÃO

6.1 – AVALIAÇÃO PRELIMINAR DA REAÇÃO DE AMINOFENOL COM NITRITO EM MEIO ÁCIDO.

Estudos preliminares da cinética das reações desenvolvidas foram realizadas. Para isto, vários monitoramentos de reações foram efetuados variando as concentrações de nitrito de sódio entre 0,0007 e 0,0355 mol L⁻¹ com uma mistura ternária de isômeros aminofenol (0,0007 mol L⁻¹) em pH=4 (amostra 29 da tabela 2), sendo que as concentrações mencionadas pertencem à solução na cela no início da reação

Cada uma das figuras 21a, 21b, 21c, e 21d mostra 30 espectros que foram coletados a cada 3 segundos da reação entre uma mistura de isômeros aminofenol, e nitrito em diferentes concentrações. Pode-se observar nestes gráficos, que a medida em que aumenta a concentração de nitrito de sódio, a reação atinge mais rapidamente o equilíbrio, o qual pode ser verificado pelo menor número de espectros coletados antes da reação atingir o equilíbrio.

O valor máximo de absorvância a 400 nm na figura 21a é 0,54, enquanto que nas outras figuras seus valores máximos de absorvância a 400 nm são próximos a 0,62. Desta forma pode-se presumir que a reação desenvolvida na figura 21a ainda não atingiu o equilíbrio, no tempo monitorado.

Inicialmente esperava-se que a reação só dependesse da concentração dos isômeros, mas estes resultados demonstraram que a velocidade de reação também depende da concentração de nitrito.

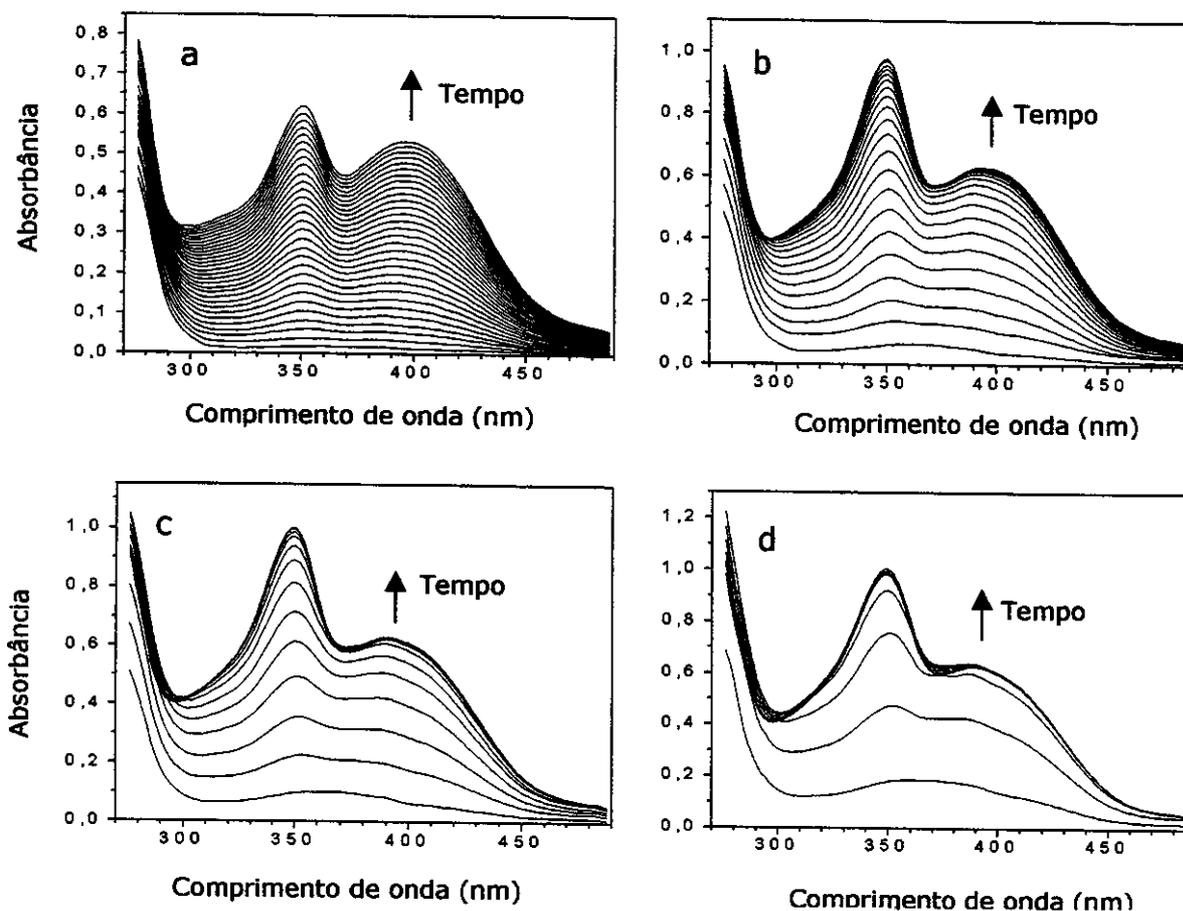


Figura 21: Espectros obtidos pelo monitoramento da reação entre aminofenol e nitrito de sódio, cujas razões de suas concentrações molares [aminofenol]:[nitrito] estão: a)1:10; b)1:20; c)1:30; d)1:40.

Sabe-se que a ordem "n", com respeito a uma espécie participante "A" em uma reação, pode ser definida em termos da concentração de A (C_A) e sua variação com respeito ao tempo pode ser expressa na seguinte equação:

$$-d C_A / d t = k C_A^n \quad (50)$$

Esta expressão é conhecida como a equação cinética de uma reação de ordem "n", onde "k" é a constante cinética ou constante de velocidade.

Quando existem três espécies participantes, envolvidas na reação, a equação cinética da reação fica definida como:

$$-d C_A / d t = k C_A^n C_B^m C_C^p \quad (51)$$

onde "n" é a ordem com respeito à espécie A, "m" é ordem com respeito a B, "p" é a ordem com respeito à espécie C e (m + n + p) é a ordem da reação.

A equação 51 pode ser utilizada para representar a velocidade de reação entre aminofenol e nitrito, sendo que estas duas espécies além da concentração de H⁺ afetam a velocidade de reação, da formação do produto azofenol. Assim, a equação cinética para a reação entre *o*-aminofenol (*o*-amf) e nitrito, pode ser definida como:

$$-d C_{o\text{-amf}} / d t = k [o\text{-amf}]^n [\text{nitrito}]^m [H^+]^p \quad (52)$$

Um estudo aproximado foi realizado para determinar o ordem "n" da reação do *o*-aminofenol, mantendo constante a concentração de nitrito e o pH. Utilizou-se para este fim o método de velocidade inicial que é empregado quando existe mais de uma espécie envolvida na reação. Este método aplicado à reação em estudo, estabelece que quando as concentrações iniciais de nitrito e íon hidrogênio se mantêm constantes numa série de monitoramentos cinéticos, a velocidade inicial $(-d C_{o\text{-amf}} / dt)_o$ está diretamente relacionada com a concentração inicial de *o*-aminofenol $(C_{o\text{-amf}})_o$ como:

$$\log (-d C_{o\text{-amf}} / dt)_o = \text{constante} + n \log (C_{o\text{-amf}})_o \quad (53)$$

Observando a reação de *o*-aminofenol com nitrito de sódio mostrada na figura 17, pode-se estabelecer a relação que existe entre a velocidade de decomposição do reagente aminofenol e a velocidade de formação do produto azofenol na seguinte equação:

$$-d C_{o-amf} / dt = d C_{o-azf} / dt \quad (54)$$

onde, C_{o-azf} é a concentração do *o*-azofenol.

A equação 54 é válida desde que a razão entre os coeficientes estequiométricos do reagente (*o*-amf) e do produto na reação (*o*-azf), seja 1. Substituindo a equação 54 em 53 tem-se:

$$\log (d C_{o-azf} / dt)_0 = \text{constante} + n \log (C_{o-amf})_0 \quad (55)$$

Devido ao fato que, em um monitoramento cinético espectrofotométrico entre *o*-aminofenol e nitrito, o valor da absorvância medida é proporcional à concentração do *o*-azofenol formado, pode-se assumir que a concentração inicial do *o*-aminofenol é proporcional ao valor de absorvância no tempo infinito (A_∞). Também pode-se assumir que a variação da concentração do *o*-azofenol no tempo é proporcional à variação do valor de absorvância no tempo. Assim, a equação anterior fica como:

$$\log (d A_{o-azf} / dt)_0 = \text{constante} + n \log (A_\infty) \quad (56)$$

Esta equação foi utilizada para calcular o valor de "n". Como foi realizado o monitoramento espectrofotométrico simultâneo das reações desenvolvidas dos isômeros *o*-, *m*- e *p*-aminofenol com nitrito, tomou-se informação de absorvância de um comprimento de onda onde apenas o *o*-azofenol absorva, que neste caso era 410nm.

Calculou-se aproximadamente a velocidade inicial a partir dos valores de absorvância coletados nos primeiros intervalos de tempo monitorados, de acordo à seguinte equação:

$$d A_{o-azf} / dt = (A_{t=4s} - A_{t=1})/3 \quad (57)$$

$A_{t=1}$ e $A_{t=4}$ são os valores de absorvância medidos a 1 e 4 segundos respectivamente, depois do início da reação.

O valor de absorvância no tempo infinito (A_{∞}) foi assumido ser semelhante ao valor de absorvância que se mantém constante no tempo (A_t), isto é 28 segundos após do início da reação. Este, pode ser expresso matematicamente como:

$$A_{\infty} = A_{t=28} \quad (58)$$

onde, $A_{t=28}$ é o valor de absorvância medido depois de 28 segundos do início da reação.

Para o cálculo aproximado de "n" pela equação 56, usou-se informação cinética de 22 misturas contendo *o*-aminofenol. De acordo a esta equação, a inclinação da linha reta obtida pelo gráfico do logaritmo da velocidade inicial ($\log v_0$) da reação entre *o*-aminofenol e nitrito contra a absorvância medida a 28 segundos ($A_{t=28}$) fornece o valor de "n". Este gráfico é mostrado na figura 22. Também nesta figura foram graficados os logaritmos das velocidades de uma amostra preparada em triplicata contendo só o isômero *o*-aminofenol.

Na figura 22, os pontos vermelhos correspondem às amostras que contém apenas o isômero *o*-aminofenol, os pontos pretos correspondem às misturas que contem os isômeros *o*-, *m*- e *p*-aminofenol. Os pontos pretos tem tendência a formar uma reta, e o pontos vermelhos estão fora dessa tendência. Este resultado pode ser interpretado como que a cinética de *o*-aminofenol numa amostra pura tem um comportamento diferente da cinética de *o*-aminofenol em uma mistura, mesmo mantendo constantes os outros parâmetros que afetam a cinética da reação como concentração de nitrito e pH.

Presumivelmente, a cinética de reação de *o*-aminofenol com nitrito é afetada pela presença dos outros isômeros, já que eles competem pelo mesmo reagente (nitrito). Este efeito é conhecido como sinergismo. Destacando que o sinergismo pode modificar a velocidade pela presença de outra espécie, já sea alterando a ordem da reação o que leva a um novo mecanismo ou modificando o valor da constante de velocidade o qual é manifestado como uma competição pelo nitrito. A dispersão dos pontos pretos perto à linha reta mostrada na figura 22, pode ser explicada de acordo a este efeito. Os isômeros *m*- e *p*- encontram-se em

diferentes concentrações nas misturas representadas pelos pontos pretos, ocasionando diferentes efeitos sobre a cinética do isômero orto.

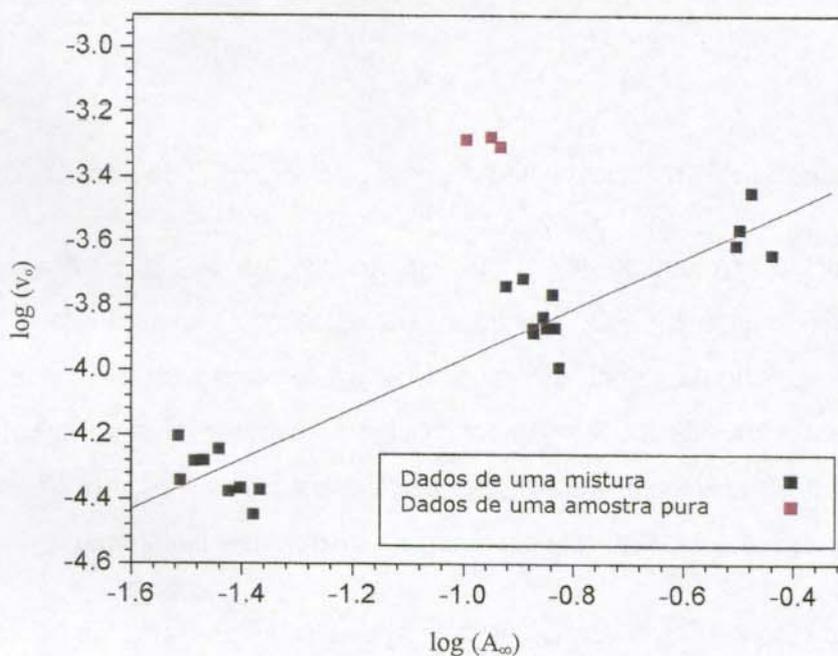


Figura 22: Gráfico do logaritmo da velocidade inicial (log v₀) da reação entre *o*-aminofenol e nitrito contra a absorbância A_{t=28} (medido a 28 segundos da reação).

O valor "n" obtido a partir da reta mostrada na figura 22 foi 0,78. Com este resultado pode-se estabelecer a seguinte equação:

$$v = k_{\text{nov}} [o\text{-amf}]^{0,78} \quad (59)$$

Sendo que $k_{\text{nov}} = k [\text{nitrito}]^m [\text{H}^+]^p$.

Da literatura¹⁰ sabe-se que valores da ordem compreendidos entre 0,9 e 1,1 são considerados de primeira ordem, e devido a isto pode-se presumir que a reação não é de primeira ordem com respeito ao isômero *o*-aminofenol.

6.2 – MODELAGEM DOS DADOS CINÉTICOS COM PARAFAC1

Dados cinéticos espectrofotométricos das primeiras 25 misturas mostradas na tabela 2, foram utilizados para sua modelagem utilizando PARAFAC1. Os dados cinéticos espectrofotométricos de cada amostra, obtidos nos primeiros 10 intervalos de tempo da reação e os 107 comprimentos de ondas pertencente à faixa espectral 276-488nm, foram arranjados numa matriz 10 x 107.

O agrupamento de dados de 25 amostras gera uma terceira dimensão (amostras). Assim, os dados passam a ser tridimensionais porque podem ser descritos por 3 variáveis, cada uma pertencente a uma dimensão ou modo. A figura 23 mostra um tensor de dados de terceira ordem \underline{X} de dimensões 25 x 107 x 10 formado pelos valores de absorvância obtidos pelos monitoramentos cinéticos espectrofotométricos de 25 amostras.

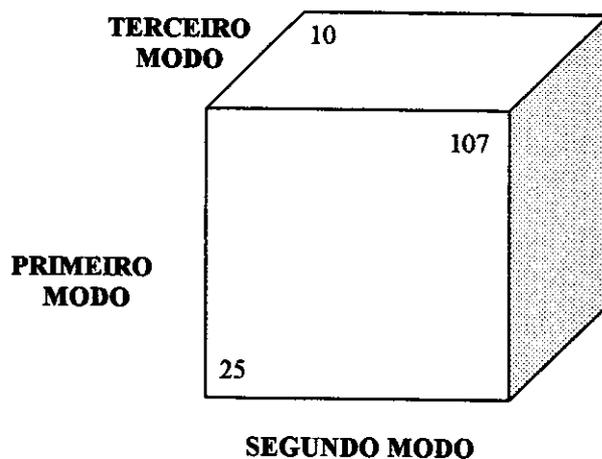


Figura 23: Tensor de dados de terceira ordem obtidos com os dados cinéticos espectrofotométricos de 25 amostras.

Na figura 23, no primeiro modo refere-se aos valores de absorvância medidos para diferentes amostras, no segundo modo encontram-se os valores de absorvância medidos a diferentes comprimentos de onda, e no terceiro modo encontram-se os valores de absorvância medidos para diferentes intervalos de tempo.

Devido ao fato que o programa PARAFAC 1 utilizado só trabalha com dados multidimensionais arranjados em uma matriz, os dados tridimensionais coletados foram arranjados numa matriz de dimensões 25 x 1070. Este arranjo é similar ao mostrado na figura 7, no qual I=10, J=107 e K=25. É importante notar que com este arranjo o programa de PARAFAC 1 identifica o primeiro modo como a dimensão das amostras, o segundo modo como a dimensão dos comprimentos de onda e o terceiro modo como a dimensão dos tempos.

Depois do arranjo dos dados, foi estudado o tipo de pré - processamento de dados mais adequado a ser usado. A tabela 3 indica as porcentagens de variância obtidas pelo PARAFAC1 em dados com diferentes tipos de pré - processamento. No cálculo dos modelos foram usados 3 componentes principais, já que três isômeros estão sendo analisados nas amostras.

Tabela 3 : Porcentagem de variância obtidas pelo PARAFAC 1 para diferentes tipos de pré - processamentos de dados.

Tipo de pré - processamento de dado	Porcentagem de variância obtida pelo PARAFAC 1
Centrado no primeiro modo	98.38
Centrado no segundo modo	99.37
Centrado no terceiro modo	99.29
Padronização* do primeiro modo	99.51
Padronização do segundo modo	99.60
Padronização do terceiro modo	99.48
Sem pré - processamento	99.63

* A padronização foi baseada no cálculo do desvio padrão, de acordo com a equação 14.

As porcentagens de variância mostrada na tabela 3 foram calculadas pela seguinte equação:

$$\% \text{ variância} = \left(1 - \frac{SQR}{SQM}\right) \times 100 \quad (60)$$

O valor de SQR é a somatória dos quadrados dos elementos dos resíduos do modelo, e SQM é a somatória dos quadrados dos elementos do modelo calculado.

Como pode-se observar nesta tabela, sem pré - processamento de dados obtém - se uma maior porcentagem de variância explicada o que significa uma maior extração de informação dos dados. Além disso, com os dados centrados na média e o escalonamento os *loadings* não são facilmente interpretáveis.

Assim, o modelo de PARAFAC 1 foi construído sem pré - processamento nos dados, usando três componentes, e aplicando restrições de não negatividade nos três modos, já que se esperam obter valores positivos de absorvância e o critério de convergência utilizado foi 2×10^{-3} .

Na figura 24, está apresentado o gráfico construído com os valores do SQR calculado para cada modelo de PARAFAC 1 com um número definido de componentes.

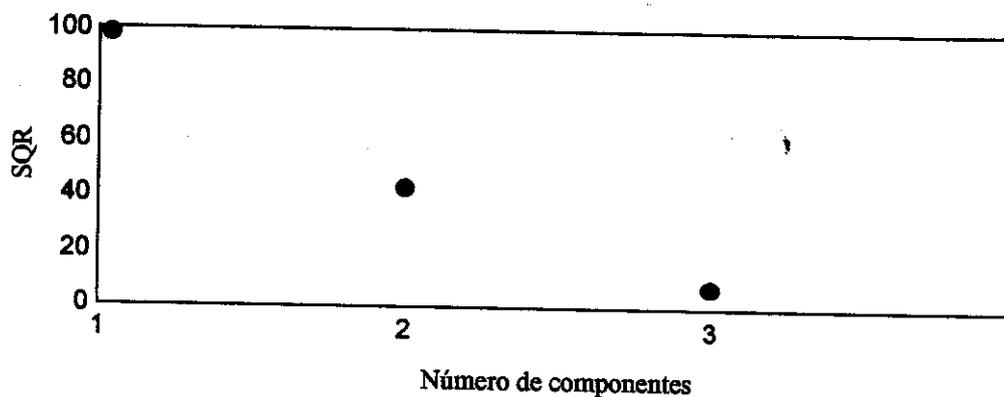


Figura 24: Somatória dos quadrados dos resíduos de \underline{X} contra o número de componentes usadas no cálculo do modelo de PARAFAC1.

Na figura 24 percebe-se que o número adequado de componentes a ser usado na construção do modelo de PARAFAC é 3, pois com esse número de componentes praticamente não existe mais resíduo em \underline{X} , concordando com o valor esperado uma vez que existem 3 espécies que estão sendo analisadas.

Como resultado da modelagem dos dados cinéticos de \underline{X} por PARAFAC1 obtém-se uma matriz *loading* **A** correspondente à dimensão das amostras, uma matriz *loading* **B** correspondente à dimensão dos comprimentos de onda, e uma matriz *loading* **C** correspondente à dimensão dos tempos.

Os vetores *loadings* **a1**, **a2** e **a3** mostradas na figura 25, formam a matriz *loading* **A** obtida pelo modelo de PARAFAC1, correspondente à dimensão das amostras do tensor \underline{X} . Estes vetores fornecem as concentrações relativas dos isômeros azofenol formados. Como 1 mol de aminofenol forma 1 mol de azo fenol, pode se assumir que estas concentrações relativas são as mesmas para as concentrações do aminofenol no início da reação.

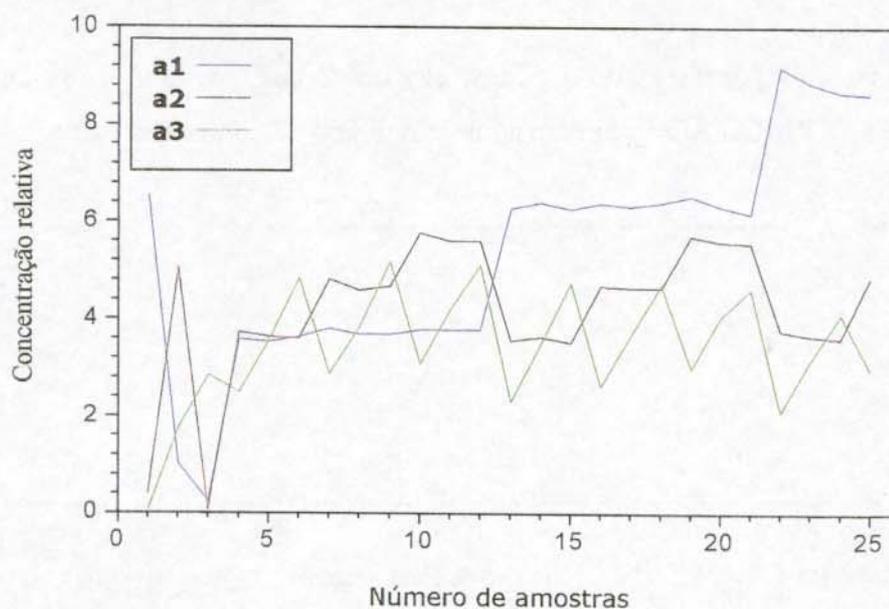


Figura 25: Matriz loading A correspondente à dimensão da amostra obtida pelo modelo de PARAFAC1.

Assim, o vetor *loading* **a1** fornece informação das concentrações relativas iniciais de *o*-aminofenol entre as amostras. O vetor **a2** nos fornece as concentrações relativas inicial de *m*-aminofenol entre as amostras. Similarmente o vetor **a3** fornece as concentrações relativas inicial de *p*-aminofenol entre as 25 amostras usadas no modelo de PARAFAC.

Os vetores loadings **b1**, **b2** e **b3** mostrados na figura 26 são componentes da matriz loading **B** correspondente à dimensão dos comprimentos de onda do tensor \underline{X} . O vetor loading **b1** fornece informação do perfil espectral do *o*-azofenol formado a partir do isômero *o*-aminofenol. O vetor loading **b2** nos fornece informação do perfil espectral do *m*-azofenol formado a partir do *m*-aminofenol. Similarmente o vetor loading **b3** fornece informação do perfil espectral do *p*-azofenol formado a partir do *p*-aminofenol, na faixa espectral de 276 a 488nm.

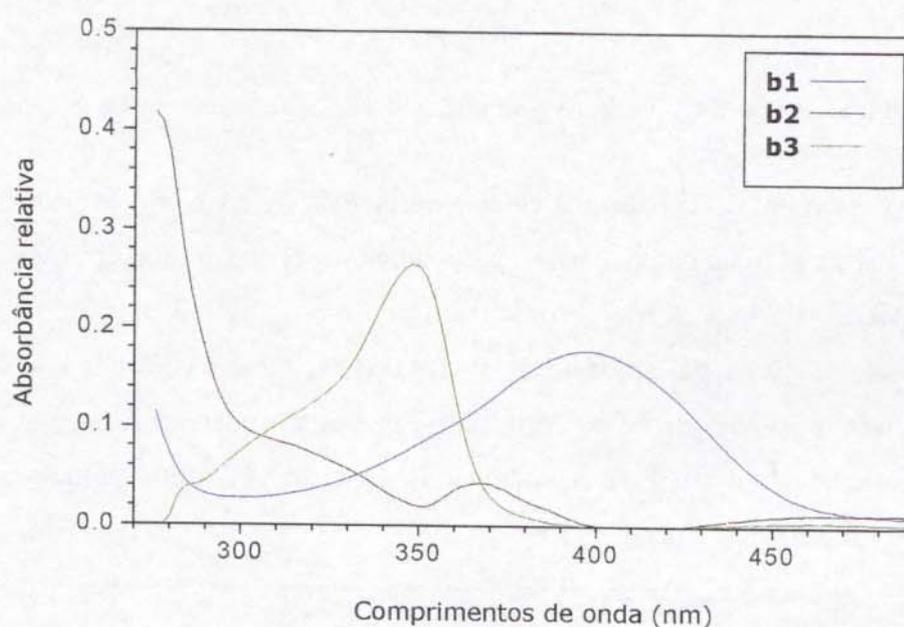


Figura 26: Matriz loading **B** correspondente à dimensão do comprimento de onda obtido pelo modelo de PARAFAC 1

Os vetores loadings **c1**, **c2** e **c3** mostrados na figura 27 são componentes da matriz loading **C** correspondente à dimensão dos tempos do tensor \underline{X} . O vetor loading **c1** fornece informação do perfil cinético do isômero *o*-aminofenol. O vetor loading **c2** fornece informação do perfil cinético do *m*-aminofenol. Similarmente o vetor loading **c3** fornece informação do perfil cinético do *p*-aminofenol no intervalo de tempo de 1 até 28 segundos.

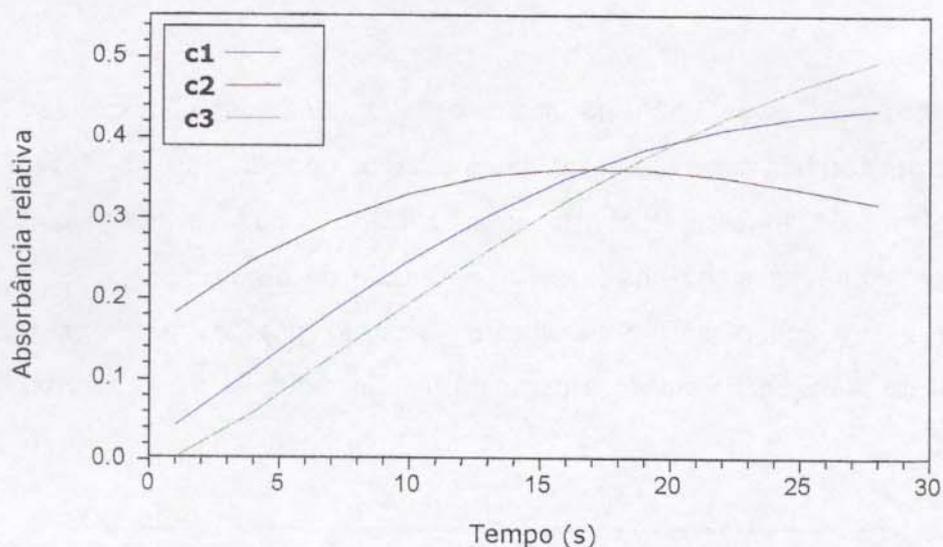


Figura 27 : Matriz *loading* **C** correspondentes à dimensão do tempo obtida pelo modelo de PARAFAC 1.

Uma regressão linear foi realizada entre o vetor *loading* **a1** e um vetor que contém as concentrações reais do *o*-aminofenol **c1_e**, para obter as concentrações previstas **c1_p** pelo modelo de PARAFAC 1 para este isômero. Assim, a figura 28 mostra o gráfico da concentração real e prevista pelo modelo de PARAFAC 1. Observa-se uma boa semelhança entre elas. Calculou-se o coeficiente de correlação entre estas concentrações, o qual forneceu um resultado aceitável igual a 0,9976. O valor de RMSEP foi 0,25 ppm, o que indica que o erro da previsão foi baixo e aceitável.

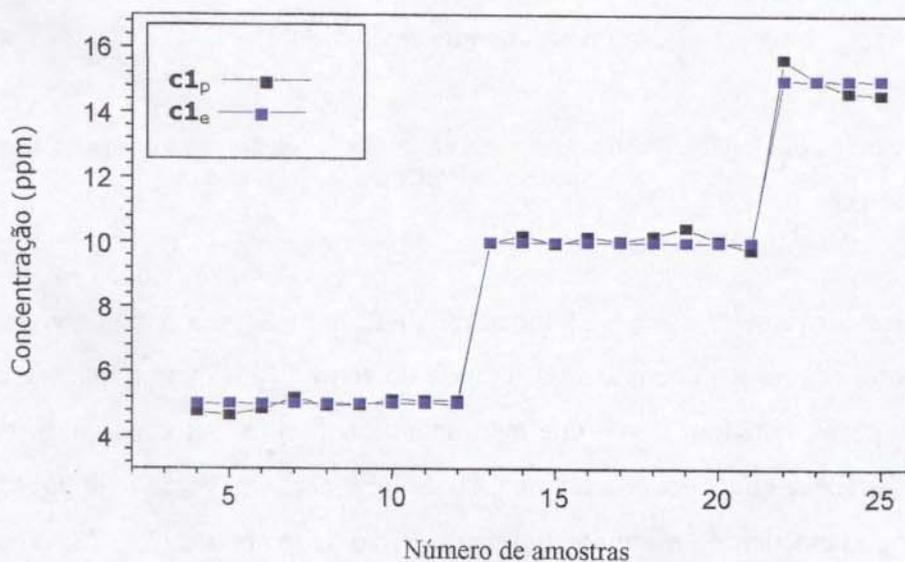


Figura 28: Concentração real ■ e prevista ■ pelo modelo de PARAFAC 1 para *o*-aminofenol.

Com uma regressão linear entre o vetor *loading* a_2 e um vetor que contem as concentrações reais do *m*-aminofenol c_{2e} , podemos obter a concentração prevista c_{2p} pelo modelo de PARAFAC 1 para este isômero. A figura 29 mostra o gráfico da concentração real e prevista pelo modelo de PARAFAC 1 para o isômero *m*-aminofenol. Observa-se uma boa correlação entre elas. A relação linear entre estas concentrações foi avaliada pelo coeficiente de correlação que forneceu um resultado satisfatório igual a 0,9946. O valor de RMSEP foi de 0,34 ppm, o que indica que o erro de previsão foi baixo e aceitável.

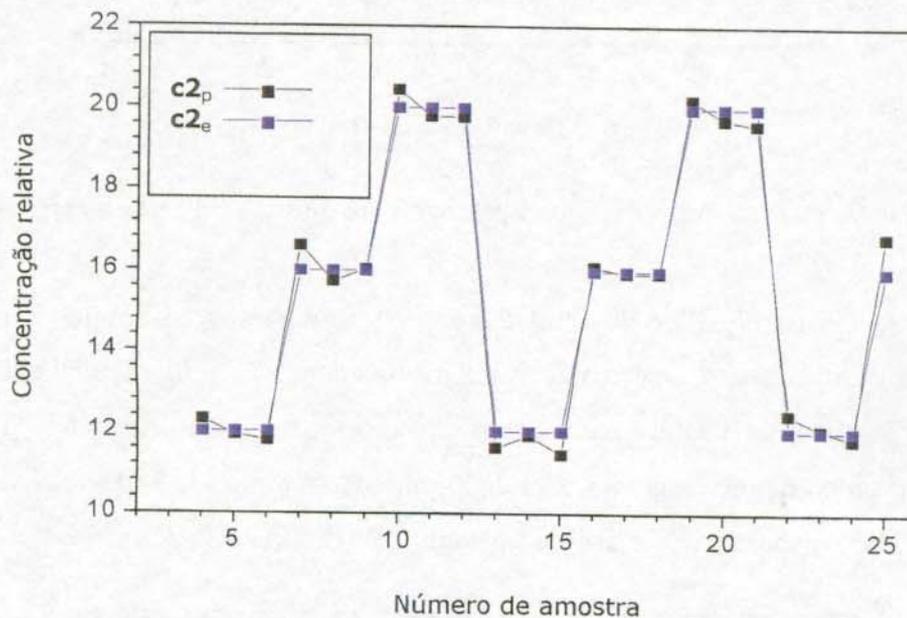


Figura 29: Concentração real ■ e prevista ■ pelo modelo de PARAFAC 1 para o *m*-aminofenol.

Podemos obter as concentrações previstas c_{3p} pelo modelo de PARAFAC 1 para *p*-aminofenol, fazendo uma regressão linear entre o vetor *loading* a_3 e um vetor que contém as concentrações reais deste isômero c_{3e} . A figura 30 mostra o gráfico da concentração real e prevista pelo modelo de PARAFAC 1 para o isômero *p*-aminofenol. Observa-se uma aceitável correlação entre elas. A relação linear entre estas concentrações foi avaliada pelo coeficiente de correlação que forneceu um resultado satisfatório igual a 0,9210. O erro de previsão foi baixo e aceitável de acordo ao valor de RMSEP calculado, que foi 0,40 ppm.

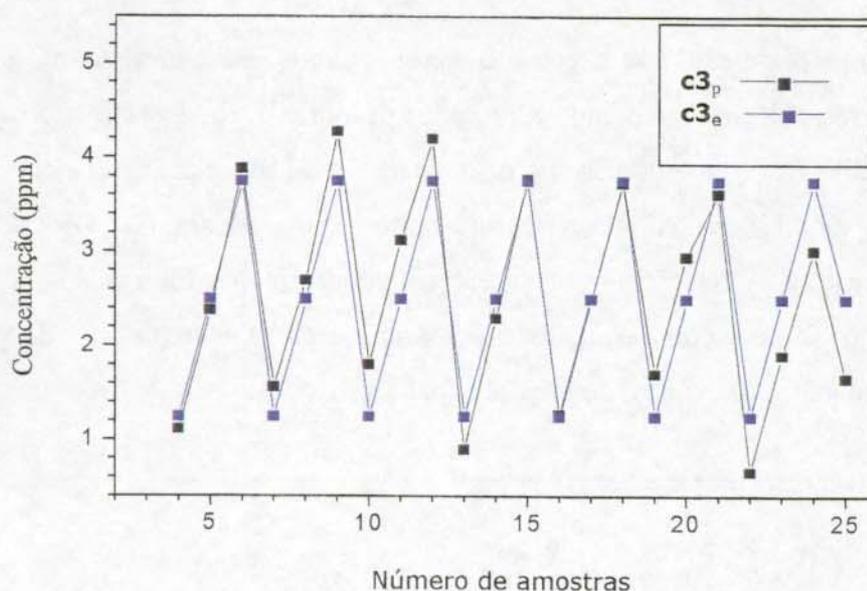


Figura 30: Concentração real ■ e prevista ■ pelo modelo de PARAFAC 1 para *p*-aminofenol.

A figura 31 mostra o gráfico do vetor *loading* $\mathbf{b1}$ normalizado $\mathbf{b1}_N$ obtido pelo modelo de PARAFAC 1, que fornece o espectro do isômero *o*-azofenol. Esta figura também mostra o gráfico do espectro experimental normalizado do isômero *o*-azofenol $\mathbf{e1}_N$ formado a partir do monitoramento cinético de uma amostra pura de *o*-aminofenol e nitrito. Pode-se observar uma boa correlação entre o espectro real e previsto pelo modelo de PARAFAC1.

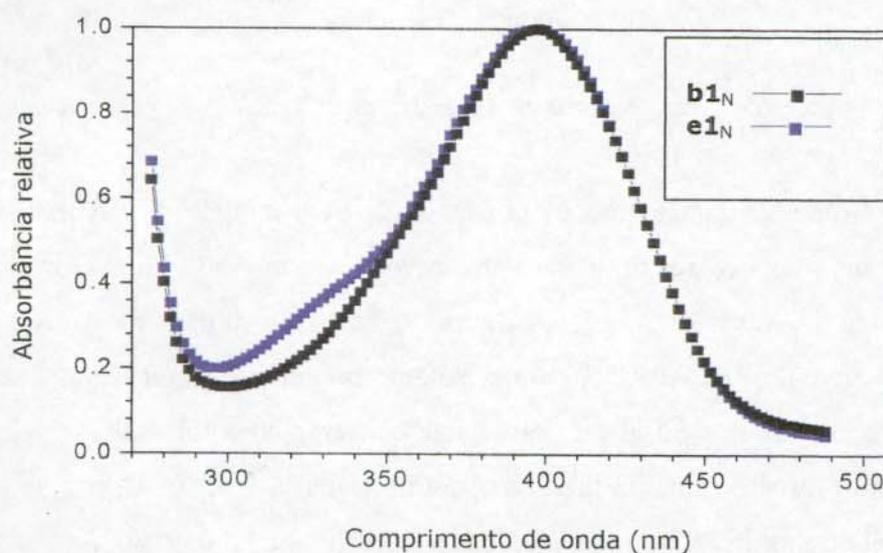


Figura 31: Espectros normalizados real ■ e previsto ■ pelo modelo de PARAFAC1 para o produto formado a partir de *o*-aminofenol.

A figura 32 mostra o gráfico do vetor *loading* $\mathbf{b2}$ normalizado $\mathbf{b2}_N$ obtido pelo modelo de PARAFAC 1, que fornece o espectro do isômero *m*-azofenol. Também esta figura mostra o gráfico do espectro experimental normalizado do *m*-azofenol $\mathbf{e2}_N$ formado a partir do monitoramento cinético de uma amostra pura de *m*-aminofenol. Pode-se observar nesta figura que existiu certa semelhança entre o espectro normalizado real e previsto pelo modelo de PARAFAC1.

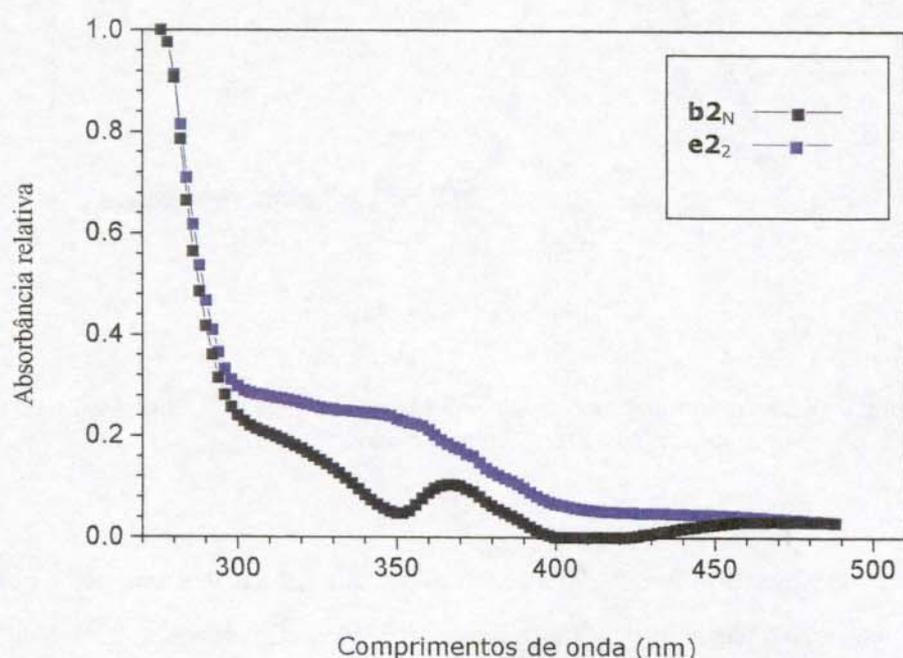


Figura 32: Espectros normalizados real ■ e previsto ■ pelo modelo de PARAFAC 1 para o produto formado pelo isômero *m*-aminofenol

A figura 33 mostra o gráfico do vetor *loading* $\mathbf{b3}$ normalizado $\mathbf{b3}_N$ obtido pelo modelo de PARAFAC 1, que fornece o espectro do isômero *p*-azofenol. Também esta figura mostra o gráfico do espectro experimental normalizado do isômero *p*-azofenol $\mathbf{e3}_N$ formado a partir do monitoramento espectrofotométrico de uma amostra pura de *p*-aminofenol. Pode-se observar nesta figura que existe uma aceitável semelhança entre os espectro normalizado real e previsto pelo modelo de PARAFAC1.

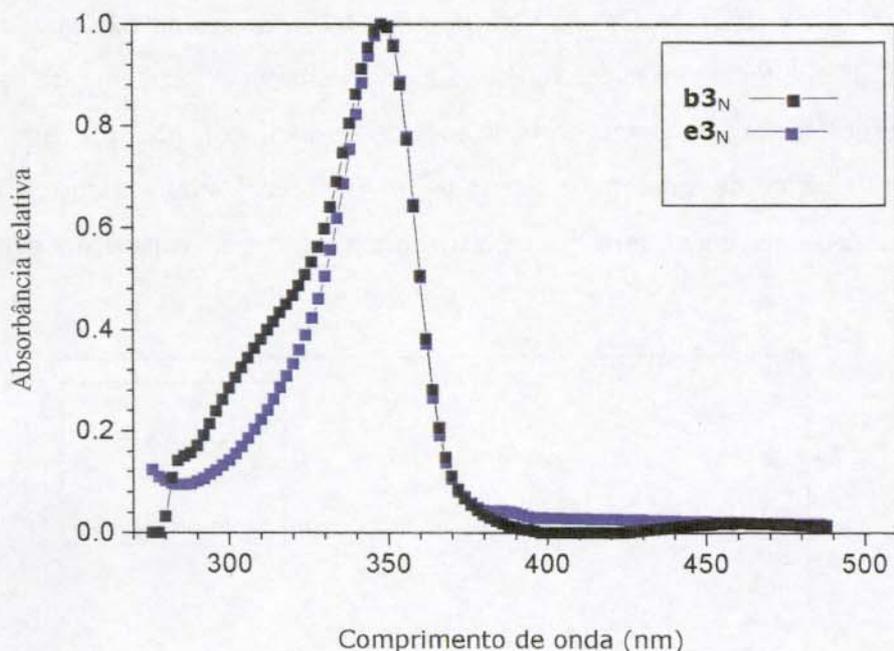


Figura 33: Espectros normalizados real e previsto pelo modelo de PARAFAC 1 para o produto formado pelo p-aminofenol.

Na sessão 6.1, mencionou-se e demonstrou-se que as curvas cinéticas obtidas pela reação de uma mistura ternária dos isômeros e nitrito de sódio depende das concentrações iniciais dos isômeros. Observou-se também que a medida que aumentava a concentração de nitrito aumentava a velocidade de reação. Em outras palavras, diminuindo a razão [aminofenol]:[nitrito], aumenta a velocidade de reação entre os isômeros e nitrito, além de existir provavelmente efeitos de sinergismo, no qual a cinética de um isômero é afetada pela presença dos outros isômeros.

De acordo com estas observações o modelo de PARAFAC 1 provavelmente não é adequado para prever as curvas cinéticas, uma vez que elas variam.

6.3 – MODELAGEM DOS DADOS CINÉTICOS COM PARAFAC 2

O PARAFAC 2 foi utilizado para a modelagem dos dados cinéticos espectrofotométricos, isto é, resolver a mistura de espectros, assim como obter as curvas cinéticas e as concentrações de cada isômero.

Um conjunto formado pelas primeiras 25 misturas de isômeros de *o*-, *m*- e *p*-aminofenol mostradas na tabela 2 foi usado para construir o modelo. Os dados cinéticos coletados foram arranjados num tensor tridimensional com dimensões 107 x 10 x 25, (comprimentos de onda, tempos e amostras, respectivamente). A figura 34 mostra o gráfico deste tensor, onde se pode observar que o primeiro modo está formado pelas absorvâncias medidas em 107 variáveis de comprimentos de onda, o segundo modo em 10 variáveis de tempo, e o terceiro modo em 25 variáveis de amostras.

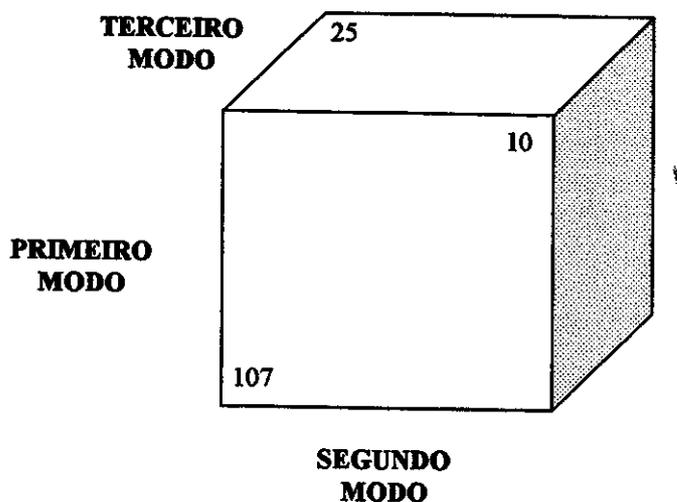


Figura 34: Dados cinéticos arranjados no tensor de dados tridimensionais de absorvância com dimensão 107 x 10 x 25 para ser modelado por PARAFAC 2

O modelo de PARAFAC 2 foi construído sem pré - processamento nos dados, usando três componentes, e aplicando restrições de não negatividade no primeiro modo (comprimentos de onda) e no terceiro modo (amostras) já que se espera obter valores

positivos de absorvância. O critério de convergência utilizado foi 10^{-5} e a porcentagem de variância capturada foi de 99,86%.

Na figura 35, mostra-se o gráfico construído com os valores do SQR (somatória dos quadrados dos resíduos) calculado para cada modelo de PARAFAC 2 com um número definido de componentes.

Observando a figura 35 percebe-se que o número adequado de componentes principais a ser usado na construção do modelo de PARAFAC 2 é 3, já que não existe grande diferença no valor residual SQR usando um número maior de componentes para descrever o modelo. Estes resultados concordam com o valor esperado uma vez que existem 3 espécies sendo analisadas.

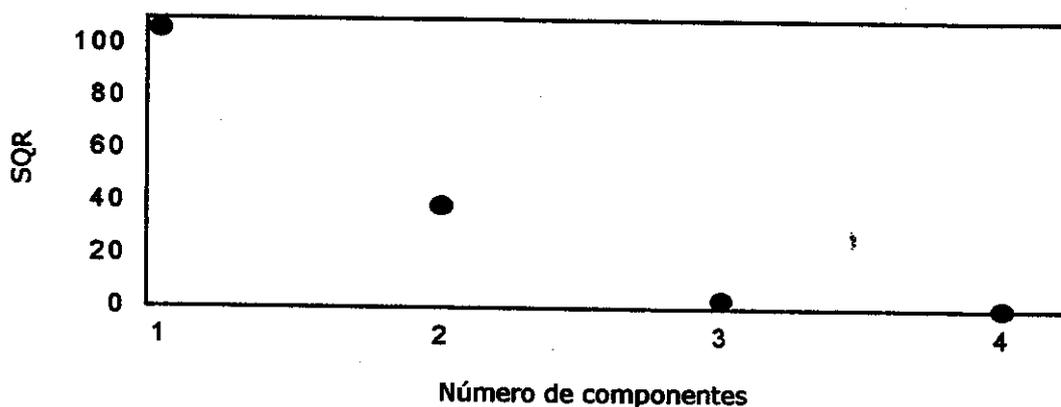


Figura 35: Somatória dos quadrados dos resíduos de \underline{X} contra o número de componentes usados no cálculo do modelo de PARAFAC 2

Como resultado do modelamento dos dados cinéticos \underline{X} por PARAFAC 2 usando 3 componentes obtêm-se, uma matriz *loading* A, correspondente à dimensão dos comprimentos de onda, 25 matrizes *loadings* $B_{k=1, \dots, 25}$ correspondentes à dimensão dos tempos (onde k indica o número da amostra ao qual pertence essa matriz), e uma matriz *loading* C correspondente à dimensão das amostras.

Os vetores a_1 , a_2 e a_3 mostrados na figura 36, são componentes da matriz *loading* A correspondentes à dimensão dos comprimentos de onda do tensor \underline{X} . O vetor *loading* a_1 fornece informação do perfil espectral do *o*-azofenol formado a partir do *o*-aminofenol. O vetor *loading* a_2 , fornece informação do perfil espectral do *m*-azofenol formado a partir do *m*-aminofenol. De maneira similar o vetor *loading* a_3 fornece informação do perfil espectral do *p*-azofenol formado a partir do *p*-aminofenol, na faixa espectral de 276 a 488nm.

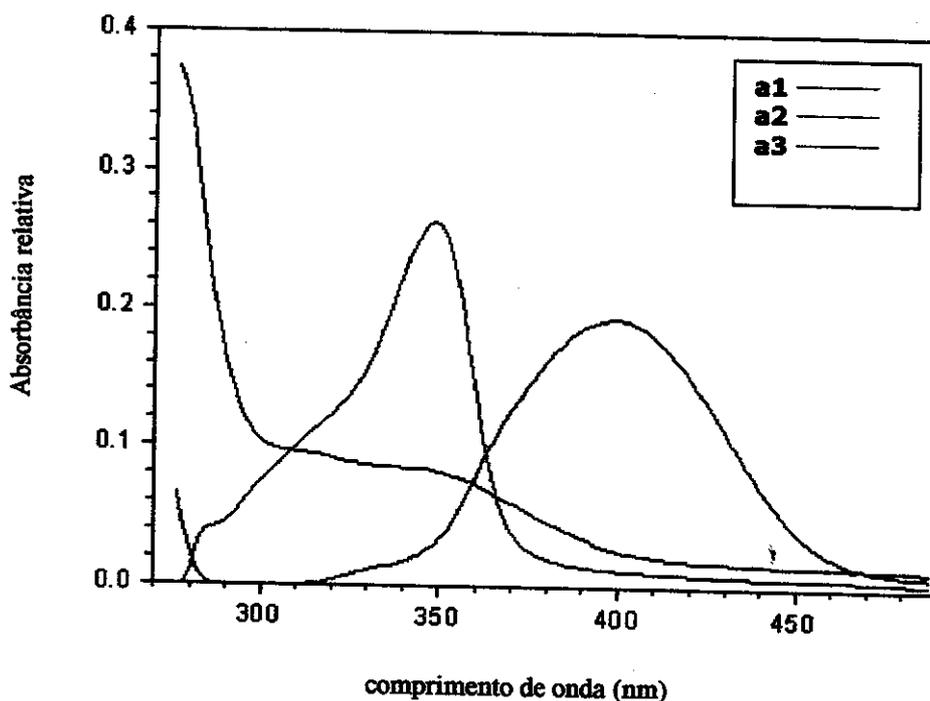


Figura 36: Matriz *loading* A obtida pela modelagem do tensor \underline{X} pelo PARAFAC 2.

A figura 37 mostra um gráfico ilustrativo de uma das 25 matrizes *loadings* obtidas por PARAFAC 2 correspondentes à dimensão dos tempos. Os vetores *loadings* b_{1_8} , b_{2_8} , e b_{3_8} , mostradas nesta figura são componentes da matriz *loading* B_8 correspondentes à dimensão dos tempos para a 8a. amostra. O vetor *loading* b_{1_8} fornece informação do perfil cinético do *o*-azofenol (formado pelo *o*-aminofenol) na oitava amostra. O vetor *loading* b_{2_8} fornece o perfil cinético do *m*-azofenol (formado pelo *m*-aminofenol) na oitava amostra. E o vetor *loading* b_{3_8} fornece o perfil cinético do *p*-azofenol (formado pelo *p*-aminofenol) na oitava amostra.

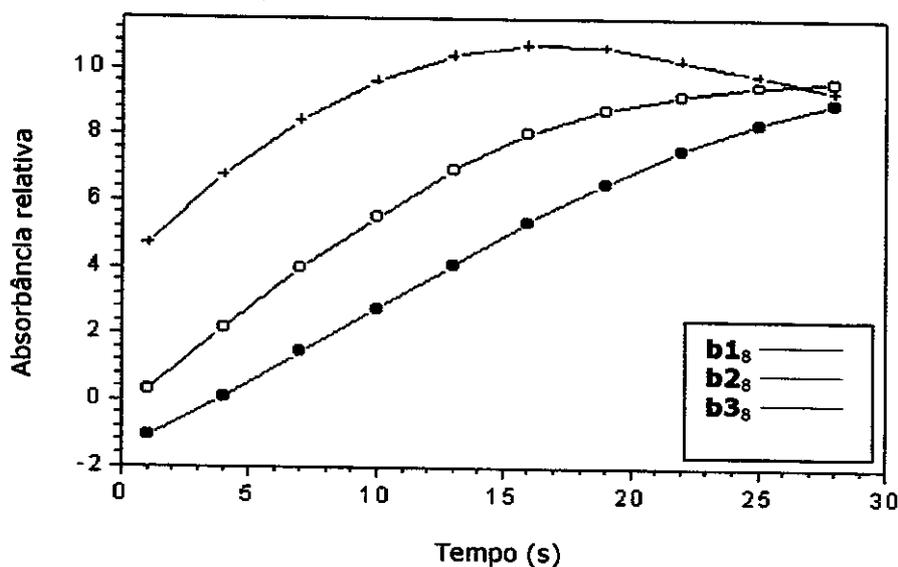


Figura 37: Gráfico da matriz *loading* B_8 obtida pelo modelamento dos dados cinéticos por PARAFAC 2

Os vetores *loadings* c_1 , c_2 e c_3 mostradas na figura 38, formam a matriz *loading* C . O vetor *loading* c_1 fornece informação das concentrações relativas iniciais do *o*-aminofenol entre as amostras. O vetor c_2 fornece as concentrações relativas iniciais do *m*-aminofenol entre as amostras. Similarmente, o vetor c_3 fornece as concentrações relativas iniciais de *p*-aminofenol entre as amostras usadas no modelo de PARAFAC 2.

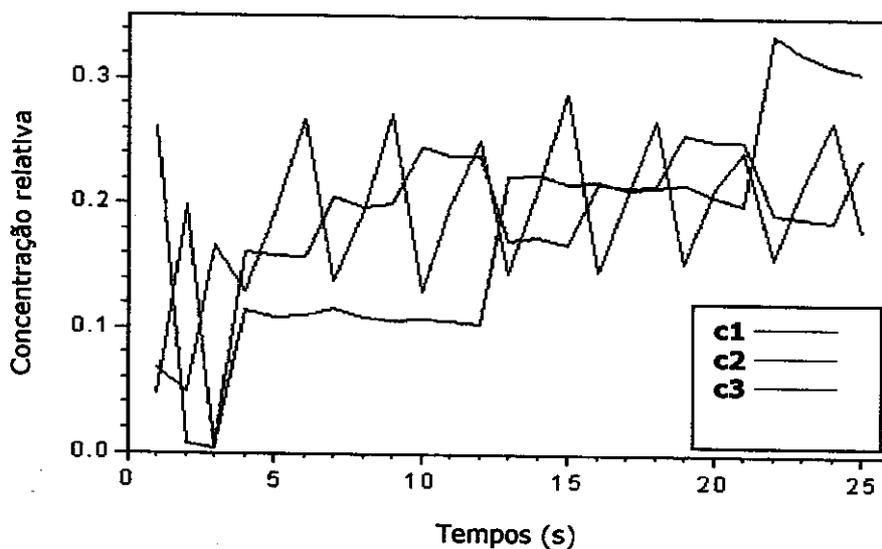


Figura 38: Matriz *loading* C obtida pelo modelamento dos dados cinéticos por PARAFAC 2

A figura 39 mostra o gráfico do vetor *loading* a_1 normalizado a_{1N} obtido pelo modelo de PARAFAC 2, que fornece o espectro do *o*-azofenol. Esta figura também mostra o gráfico normalizado do espectro do isômero *o*-azofenol e_{1N} formado a partir do *o*-aminofenol. Pode-se observar uma boa semelhança entre o espectro experimental e o espectro previsto pelo modelo de PARAFAC 2. Observa-se que os máximos de absorção coincidem muito bem.

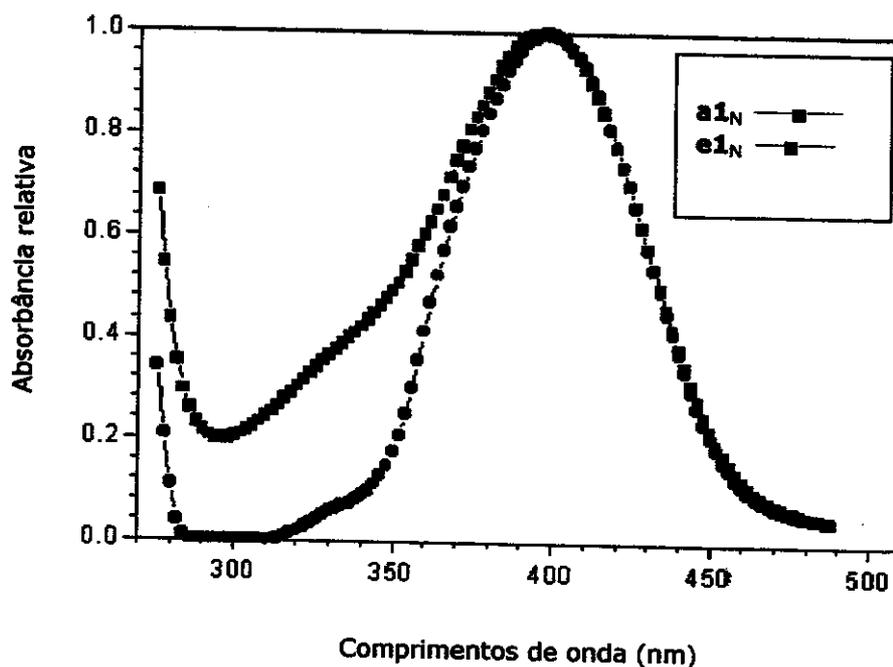


Figura 39: Espectros normalizados experimental ■ e previsto ■ pelo modelo de PARAFAC 2 para *o*-azofenol formado a partir de *o*-aminofenol.

A figura 40 mostra o gráfico do vetor *loading* a_2 normalizado a_{2N} obtido pelo modelo de PARAFAC 1, que fornece o espectro do *m*-azofenol. Nesta figura também é mostrado o gráfico do espectro experimental normalizado do *m*-aminofenol e_{2N} . Observa-se uma boa correlação entre os espectros normalizados experimental e previsto pelo modelo de PARAFAC 2.

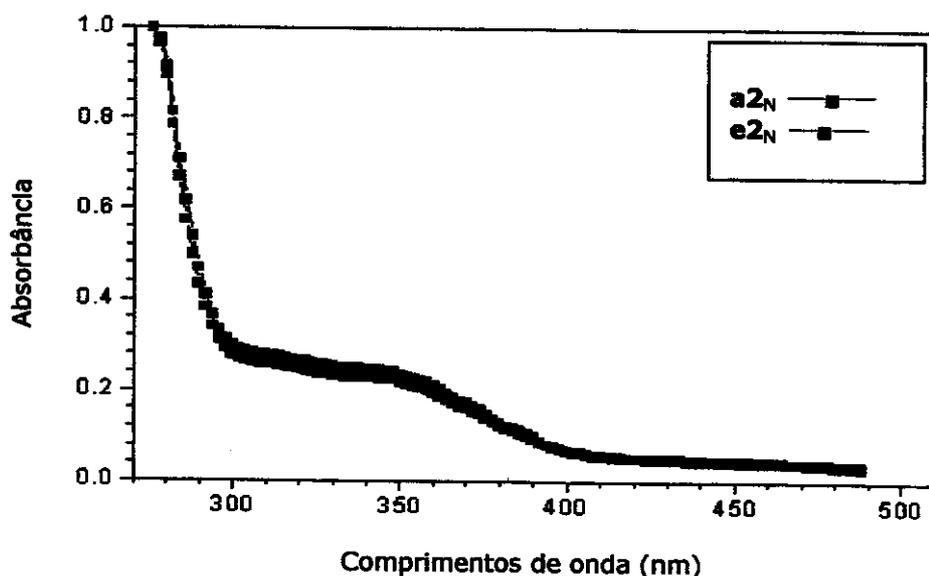


Figura 40: Espectros normalizados experimental (■) e previsto (—) pelo modelo de PARAFAC 2 para *m*-azofenol formado a partir de *m*-aminofenol.

A figura 41 apresenta o gráfico do vetor *loading* a_3 normalizado a_{3N} obtido pelo modelo de PARAFAC 2, que fornece o espectro previsto para *p*-azofenol (formado a partir de *p*-aminofenol). Também nesta figura apresenta-se o gráfico normalizado do espectro experimental do *p*-azofenol e_{3N} . Pode-se apreciar nesta figura uma boa semelhança entre os espectros previsto e experimental do *p*-azofenol.

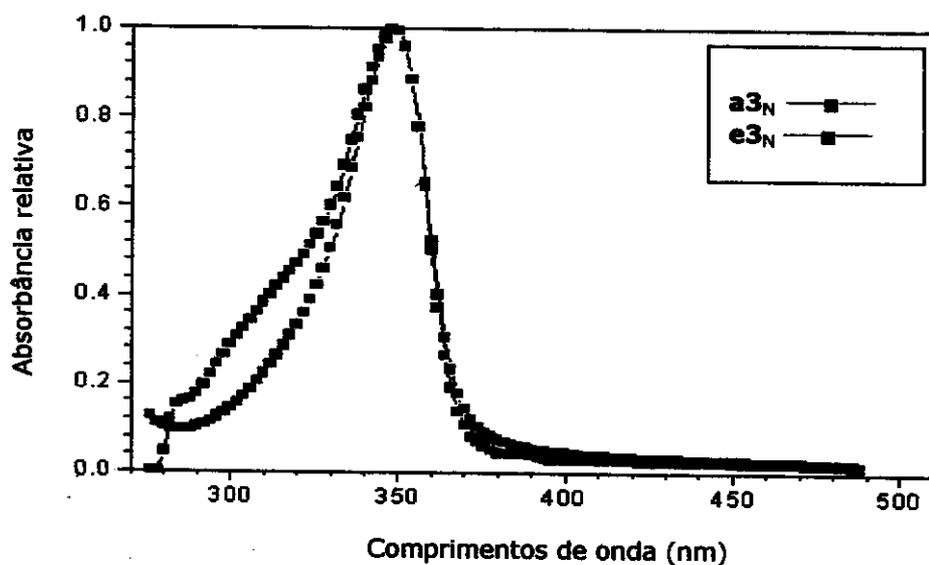


Figura 41: Espectros normalizados experimental (■) e previsto (—) pelo modelo de PARAFAC 2 para *p*-azofenol formado a partir de *p*-aminofenol.

Como a curva cinética de um isômero em uma mistura é diferente da curva quando apenas esse isômero está presente, a comparação das curvas cinéticas previstas pelo modelo com as curvas cinéticas reais de cada isômero na mistura é mais difícil de se realizar. Observando os espectros reais dos isômeros nas figuras 39, 40 e 41, pode-se observar que em comprimentos de onda maiores que 400nm, apenas o isômero *orto* absorve. Devido a isto pode-se obter a curva cinética experimental deste isômero nas misturas, e assim fazer sua comparação com as curvas cinéticas previstas pelo modelo.

Como os outros isômeros não apresentam uma região espectral sem interferência, não é possível a comparação das curvas cinéticas desses isômeros na mistura com as previstas pelo modelo.

A figura 42 mostra os gráficos comparativos entre as curvas cinéticas normalizadas real e prevista pelo modelo de PARAFAC 2 para *o*-azofenol em três misturas ternárias diferentes.

Na figura 42a, b_{1N1} é um vetor que fornece a curva cinética normalizada prevista pelo modelo de PARAFAC 2 para *o*-azofenol na amostra pura 1 e e_{1N1} é a curva cinética normalizada experimental deste isômero na mesma amostra. Similarmente nas figuras 42b e 42c, os símbolos b_{1N13} e b_{1N22} são vetores que fornecem as curvas cinéticas normalizadas previstas pelo modelo de PARAFAC 2 para as amostras 13 e 22 respectivamente. Os símbolos e_{1N13} e e_{1N22} são vetores que fornecem as curvas cinéticas experimentais normalizadas para *o*-azofenol nas amostras 13 e 22 respectivamente. Podemos observar em cada gráfico, a boa semelhança existente entre as curvas reais e as previstas pelo modelo. Também pode-se observar uma grande diferença entre a curva cinética do isômero puro apresentada na figura 42a e a do mesmo isômero na mistura apresentada nas figuras 42b e 42c.

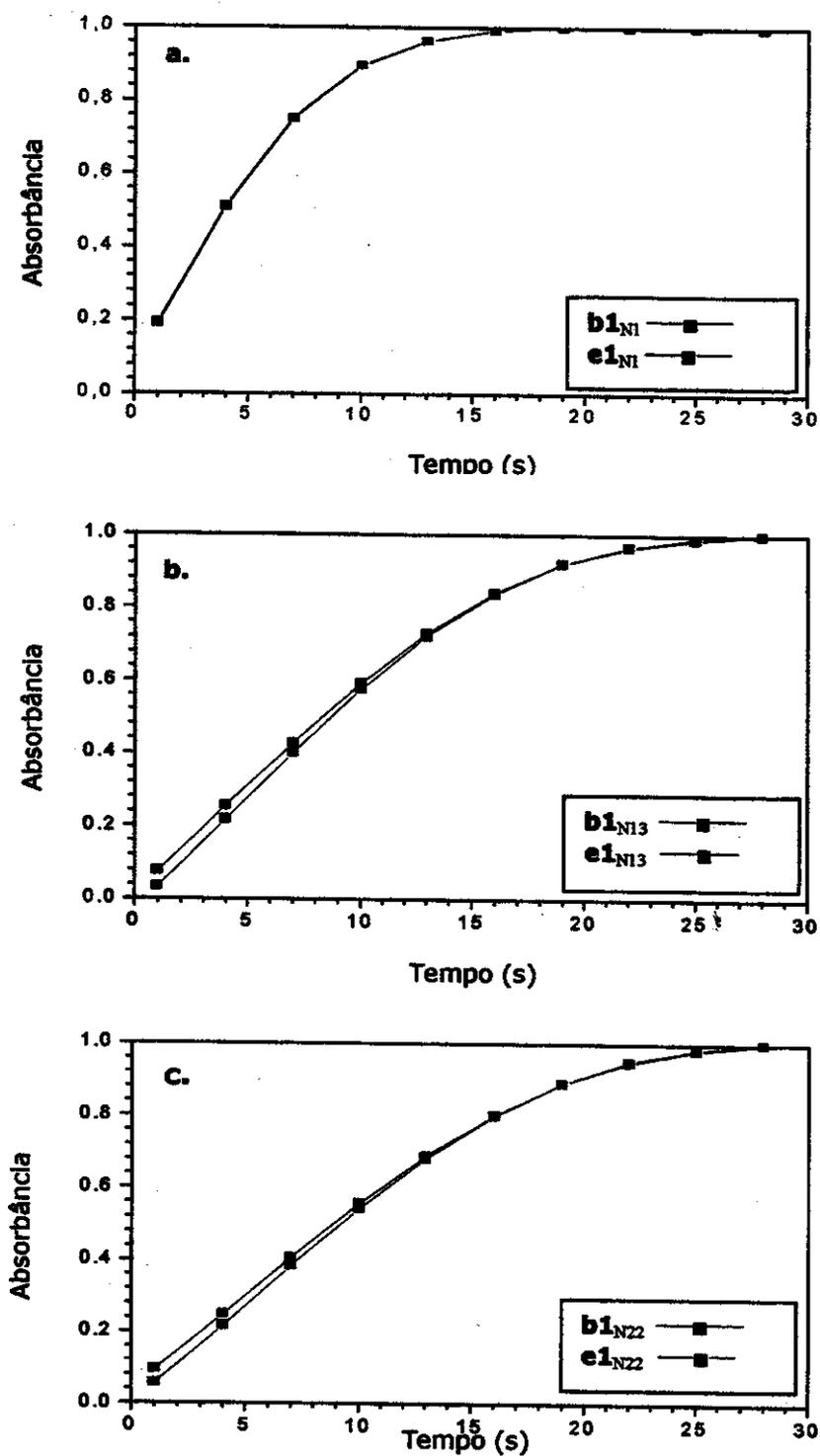


Figura 42: Curva cinética normalizada experimental (■) e prevista pelo modelo de PARAFAC2 (●) para a formação do *o*-azofenol a 410nm na: a) amostra 1, b) amostra 13 e c) amostra 22.

Fazendo uma regressão linear entre o vetor *loading* c_1 e um vetor que contém as concentrações reais do *o*-aminofenol c_{1e} , pode-se obter as concentrações previstas c_{1p} pelo modelo de PARAFAC 2 para este isômero. A figura 43 mostra o gráfico da concentração real e prevista pelo modelo de PARAFAC 2 para o isômero *o*-aminofenol. Observa-se neste gráfico a boa semelhança entre estas concentrações. Calculou-se o coeficiente de correlação linear entre as concentrações experimentais e previstas pelo modelo de PARAFAC2, o qual forneceu um resultado de 0,996, que indica a boa relação linear entre estas concentrações. Os erros foram baixos e aceitáveis, com um resultado do erro de previsão RMSEP igual 0,33 ppm.

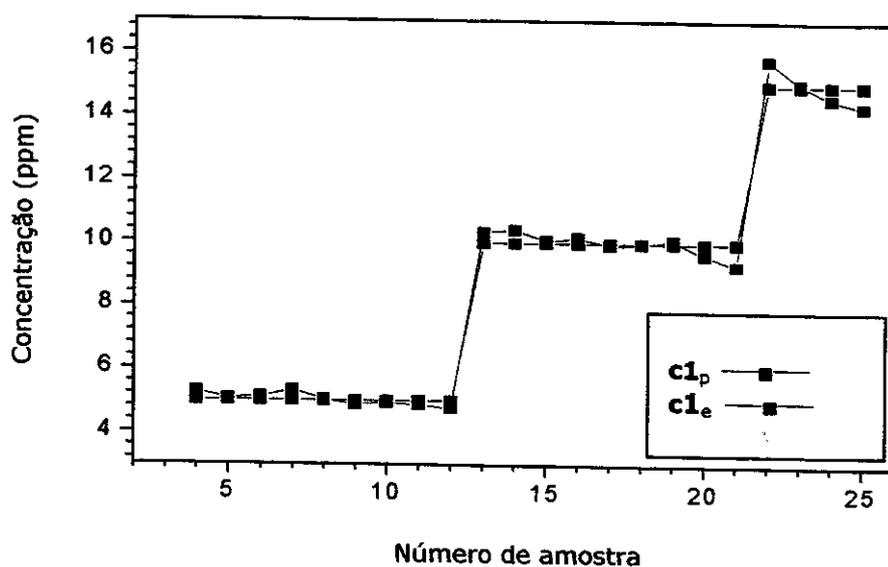


Figura 43: Concentração experimental (■) e prevista (■) pelo modelo de PARAFAC2 para *o*-aminofenol.

Pode-se obter as concentrações c_{2p} pelo modelo de PARAFAC 2 para *m*-aminofenol, fazendo uma regressão linear entre o vetor *loading* c_2 e o vetor que contém as concentrações reais deste isômero c_{2e} . A figura 44 mostra o gráfico da concentração real e prevista pelo modelo de PARAFAC 2 para *m*-aminofenol. Observa-se uma aceitável semelhança entre elas. A relação linear entre as concentrações experimentais e previstas para este isômero, foi avaliada com o cálculo do coeficiente de correlação, o qual foi 0,932. Este resultado indica que existe uma relação linear, mais com um erro não muito baixo (RMSEP foi de 1,18 ppm).

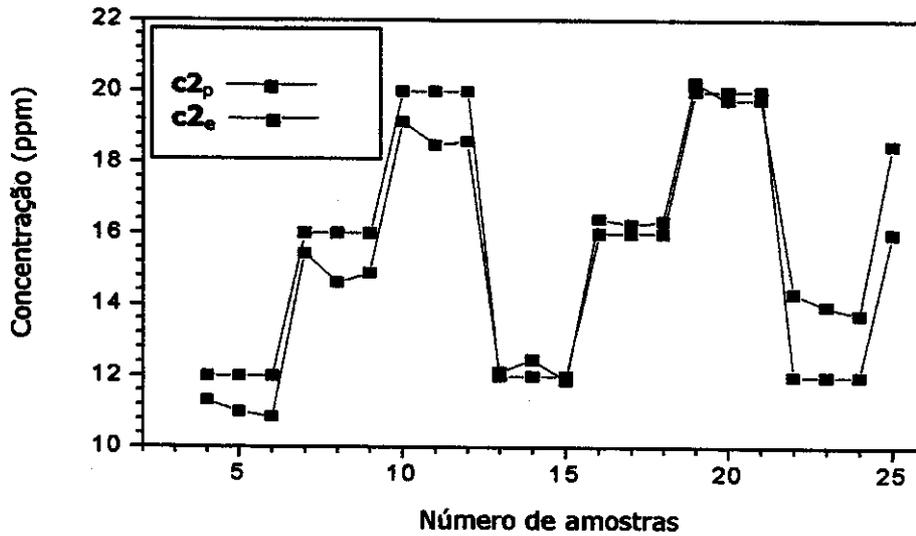


Figura 44: Concentração experimental (■) e prevista (■) pelo modelo de PARAFAC2 para *m*-aminofenol.

Com uma regressão linear entre o vetor *loading* c_3 e um vetor que contém as concentrações reais de *p*-aminofenol c_{3e} , pode-se calcular as concentrações previstas c_{3p} pelo modelo de PARAFAC 2 para este isômero. A figura 45 mostra a boa correlação que existe entre a concentração real e prevista pelo modelo para *p*-aminofenol. A relação linear entre as concentrações experimentais e previstas para este isômero, foi avaliada com o cálculo do coeficiente de correlação, o qual foi 0,971. Este resultado indica uma aceitável relação linear, com erros de previsão baixos, confirmado com o cálculo do RMSEP que foi 0,24.

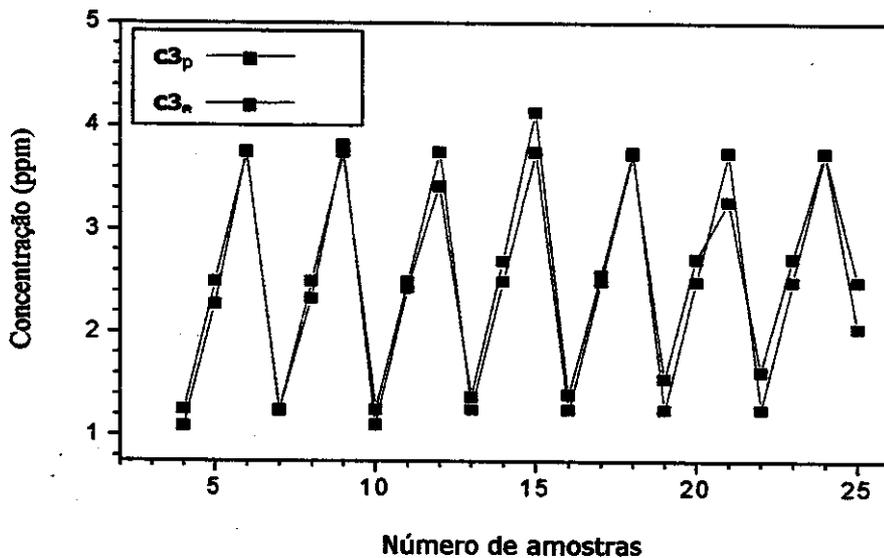


Figura 45: Concentração experimental (■) e prevista (■) pelo modelo de PARAFAC2 para *p*-aminofenol.

6.4 - COMPARAÇÃO DOS RESULTADOS OBTIDOS PELOS MÉTODOS DE PARAFAC 1 E PARAFAC 2

Uma comparação entre os resultados obtidos na modelagem dos dados obtidos pelos métodos de PARAFAC1 e PARAFAC2 foi realizada.

a) Comparação das concentrações

De acordo com os resultados apresentados anteriormente, as concentrações previstas de *o*-aminofenol por PARAFAC 1, foram mais próximas dos valores reais que nas concentrações previstas por PARAFAC 2, o qual pode ser verificado comparando os valores dos erros de previsão RMSEP. Usando o teste-F, pode-se determinar se existe uma diferença estatisticamente significativa entre as previsões com PARAFAC1 e PARAFAC2. O teste estatístico foi calculado pela equação 61 colocando no numerador o maior valor de RMSEP dos métodos e no denominador o menor valor de RMSEP.

$$\text{Teste - F} = \frac{(0,33)^2_{(\text{PARAFAC 2})}}{(0,25)^2_{(\text{PARAFAC 1})}} = 1,74 \quad (61)$$

O valor obtido foi comparado com o valor crítico do teste-F que é 2,12 a 95% de confiança com 21 graus de liberdade, já que os valores de RMSEP foram calculados para 22 amostras. De acordo ao resultado, não existe melhora significativa na previsão por PARAFAC 1 com respeito ao PARAFAC 2.

No caso de *m*-aminofenol, o PARAFAC1 apresenta uma melhora bastante significativa com respeito a PARAFAC 2, para a previsão das concentrações deste isômero nas misturas. Isso pode ser comprovado pelo valor do teste-F, calculado pela equação (62) e o valor do teste-F crítico 2,12.

$$\text{Teste - F} = \frac{(1,17)^2_{(\text{PARAFAC 2})}}{(0,34)^2_{(\text{PARAFAC 1})}} = 11,84 \quad (62)$$

Para o isômero *p*-aminofenol, os resultados previstos por PARAFAC 2 foram mais próximos aos reais que com os valores previstos por PARAFAC 1. O teste estatístico foi calculado pela equação 63 e comparado com o valor crítico do teste-F para avaliar se houve uma melhora significativa na previsão por PARAFAC 2.

$$\text{Teste - F} = \frac{(0,40)^2_{(\text{PARAFAC 1})}}{(0,24)^2_{(\text{PARAFAC 2})}} = 2,78 \quad (63)$$

Como o valor 2,80 calculado é maior que o valor crítico do teste-F 2,12, existe melhora significativa dos resultados de previsão pelo PARAFAC 2 com respeito ao outro método estudado.

b) Comparação dos espectros

Também o teste-F pode ser utilizado para se realizar uma comparação entre os espectros. Neste caso para o cálculo do RMSEP foram tomados como valores reais as absorvâncias dos 107 comprimentos de onda do espectro normalizado obtido do monitoramento cinético do isômero puro, e como valores previstos as absorvâncias dos mesmos comprimentos de onda do espectro normalizado obtido pelo modelo de PARAFAC1 ou PARAFAC2.

O espectro previsto de *o*-azofenol por PARAFAC1 foi mais parecido ao isômero puro que no caso de PARAFAC2. Um teste-F foi usado para avaliar se existe uma melhora significativa de um dos métodos sobre o outro. O teste estatístico foi calculado usando os valores de RMSEP dos métodos, segundo equação 64.

$$\text{Teste - F} = \frac{(0,18)^2_{(\text{PARAFAC 2})}}{(0,04)^2_{(\text{PARAFAC 1})}} = 20,25 \quad (64)$$

O valor obtido desta equação foi comparado com o valor crítico do teste-F, que é 1,35 com 106 graus de liberdade (já que usou-se 107 comprimentos de onda para calcular os

valores de RMSEP) e 95% de confiança. Como o valor do teste estatístico encontra-se acima do valor crítico, podemos dizer que existe uma melhora significativa do espectro resolvido de *o*-azofenol pelo PARAFAC1 com relação ao PARAFAC2.

No caso de *m*-azofenol, o modelo de PARAFAC 2 apresenta uma melhora bastante significativa para a previsão do espectro que PARAFAC 1. Isto pode ser comprovado usando o teste F. A equação 65 mostra o cálculo do teste estatístico:

$$\text{Teste - F} = \frac{(0,08)_{(\text{PARAFAC 1})}^2}{(0,01)_{(\text{PARAFAC 2})}^2} = 64,00 \quad (65)$$

Como o valor 29,19 é maior ao valor crítico do teste - F que é 1,35 com 106 graus de liberdade e 95% confiança, pode-se dizer que existe melhora muito significativa na previsão do espectro de *m*-azofenol por PARAFAC 2 com relação ao outro método estudado.

Os espectros previsto de *p*-azofenol por PARAFAC 1 e 2, são muito semelhantes entre si, e eles são bastantes similares ao espectro real deste isômero. O valor do teste estatístico para comparação destes métodos foi calculado de acordo à seguinte equação:

$$\text{Teste - F} = \frac{(0,07)_{(\text{PARAFAC 1})}^2}{(0,06)_{(\text{PARAFAC 2})}^2} = 1,36 \quad (66)$$

Observa-se que o valor calculado é menor ao valor crítico do teste - F que é 1,35, isto significa que existe melhora significativa do espectro previsto por PARAFAC 1 com respeito ao outro método estudado.

c) Comparação das curvas cinéticas

As curvas cinéticas geradas pelos 2 modelos estudados também podem ser comparadas, mas só que neste caso, como já mostrado anteriormente, apenas para o isômero orto pode ser realizado a comparação.

A figura 46 mostra as curvas cinéticas previstas pelo modelo de PARAFAC1 e PARAFAC2 para a formação do *o*-azofenol nas misturas. Observa-se, nestes gráficos uma maior semelhança entre as curvas reais e previstas pelo modelo de PARAFAC2 do que com as curvas previstas pelo PARAFAC1. Este resultado já era esperado uma vez que o PARAFAC 1 assume que as curvas cinéticas normalizadas de *o*-azofenol nas diferentes misturas são iguais, o que não ocorre para este sistema cinético.

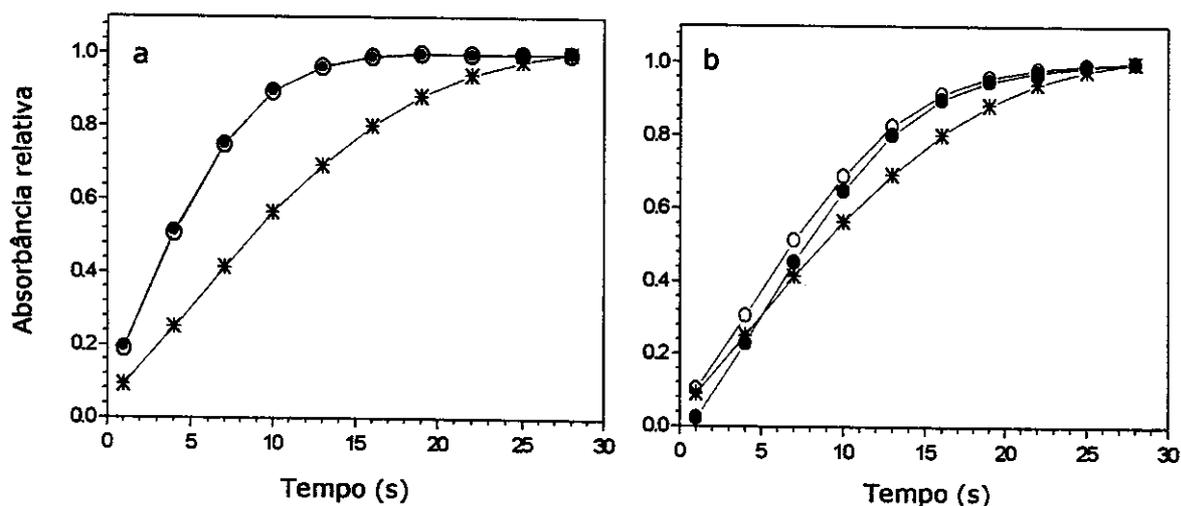


Figura 46: Curvas cinéticas normalizadas a 410 nm pertencente a *o*-azofenol usando (o) dados experimentais, (●) dados do modelo de PARAFAC2, (*) dados do modelo de PARAFAC1, na : (a) amostra 1 e (b) amostra 4.

6.5 - RESULTADOS DE N-PLS

As informações obtidas pelo monitoramento cinético na faixa espectral 288-488nm das primeiras 29 amostras apresentadas na tabela 2 foram usadas para construir o modelo de calibração, e as outras sete amostras foram usadas para validação do modelo obtido.

Os dados das amostras de calibração formam um tensor de dados tridimensional \underline{X}_{cal} . Este tensor de dados foi desdobrado numa matriz X_{cal} de dimensão $I \times JK$, I (número de amostras), J (número de comprimento de ondas), K (número de tempos) para poder ser utilizado no programa de cálculo do modelo N-PLS. Os dados foram pré-processados,

centrando na média no primeiro modo e segundo modo de \underline{X}_{cal} . Este tratamento melhora a habilidade de previsão do método, diminuindo os erros de previsão.

O critério utilizado para a escolha do número ideal de variáveis latentes foi baseado no valor do SQRX (soma dos quadrados dos resíduos de \underline{X}), calculado a partir da diferença entre o tensor \underline{X} obtido pelo modelo de calibração PLS tridimensional para uma determinada quantidade de variáveis latentes, e o tensor \underline{X} original). O número de variáveis latentes escolhido foi aquele a partir do qual não havia uma diminuição significativa no SQRX.

O gráfico mostrado na figura 47, apresenta os erros calculados para o tensor \underline{X} para diferentes quantidades de variáveis latentes usadas na construção do modelo de calibração PLS tridimensional. Observa-se que a partir da terceira variável latente não existe mais uma diferença significativa no valor de SQRX. Assim, o número de variáveis latentes pode ser 3 ou 4.

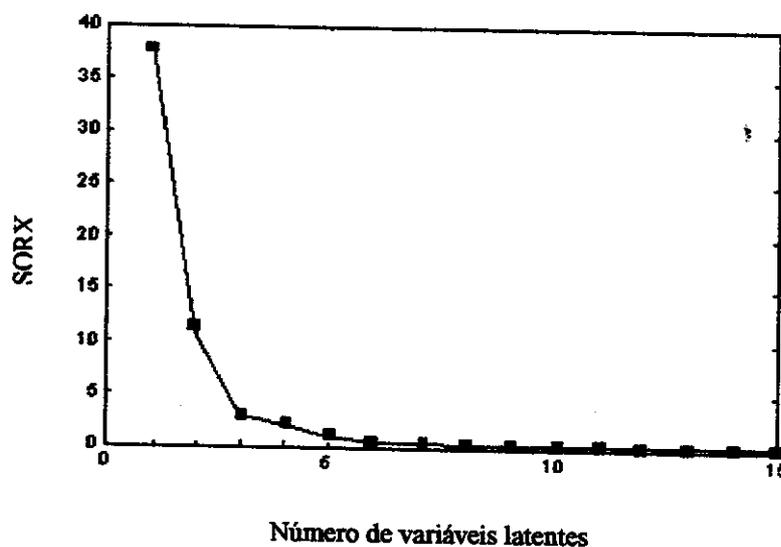


Figura 47: SQRX (somatória dos quadrados dos resíduos de \underline{X}) contra o número de variáveis latentes

Outro critério foi utilizado para decidir qual é o número de variáveis latentes mais apropriadas para se utilizar no modelo. Este critério baseia-se nos métodos de validação

interna MSEE (erro médio quadrático da estimativa) e externa MSEP (erro médio quadrático da previsão). O valor de MSEE e o valor de MSEP são obtidos a partir dos dados de concentração dos isômeros *o*-, *m*- e *p*-aminofenol reais e previstos pelo modelo de calibração para um número definido de variáveis latentes, respectivamente.

O número de variáveis latentes escolhido foi aquele a partir do qual os valores de MSEE e MSEP são muito parecidos, ou a diferença entre eles não é significativa. Além disso, deve-se ter a condição simultaneamente na qual os valores de MSEE e MSEP sejam os mais baixos possíveis, usando o menor número de variáveis latentes.

A figura 48 mostra o gráfico dos valores de MSEE e MSEP calculado para as concentrações reais e previstas pelo modelo de NPLS, para *o*-aminofenol em função do número de variáveis latentes utilizadas na construção do modelo de NPLS. Este gráfico mostra que o número de variáveis latentes ótimo para a calibração do *o*-aminofenol é 4, já que é a menor quantidade de variáveis latentes onde o valor de MSEE e o valor de MSEP são pequenos e muito parecidos entre si. Também pode-se observar neste gráfico que a partir de 7 variáveis latentes o erro de MSEP proveniente dos dados de teste aumenta consideravelmente. Isto é explicado por um sobre-ajuste do modelo, construído com um número maior de variáveis latentes que a adequada.

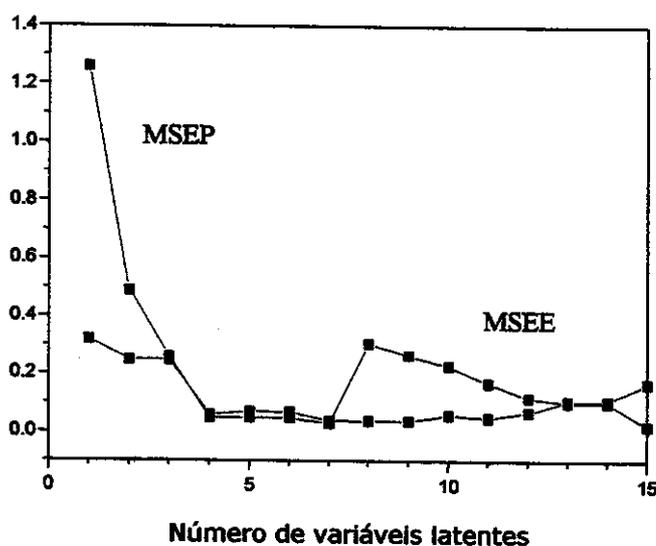


Figura 48: Erro médio quadrático de previsão (MSEP) e da estimativa (MSEE) em função do número de variáveis latentes para as concentrações reais e previstas para *o*-aminofenol

O gráfico da figura 49 mostra os valores de MSEE e os valores de MSEP calculadas para as concentrações reais e previstas pelo modelo de NPLS para o *m*-aminofenol contra o número de variáveis latentes usados na construção do modelo de calibração NPLS tridimensional. Observa-se neste gráfico que 3 variáveis latentes são adequadas para serem utilizadas no cálculo do modelo, por ser a menor quantidade de variáveis latentes que apresenta pequenos valores de MSEE e MSEP além de serem parecidos entre si.

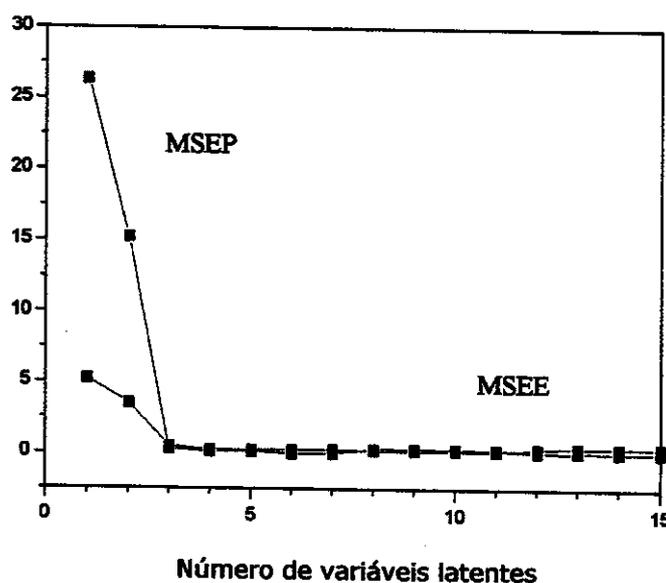


Figura 49: Erro médio quadrático de previsão (MSEP) e da estimativa (MSEE) em função do número de variáveis latentes para as concentrações reais e previstas para *m*-aminofenol.

Na figura 50 observa-se o gráfico dos valores de MSEE e os valores de MSEP calculados para as concentrações reais e previstas pelo modelo de N-PLS para o isômero *p*-aminofenol contra o número de variáveis latentes usados na construção do modelo de calibração N-PLS. A partir deste gráfico pode-se dizer que 4 variáveis latentes são as mais adequadas para serem usadas no cálculo do modelo, por ser a menor quantidade de variáveis latentes que não apresenta diferença significativa entre os valores de MSEE e MSEP.

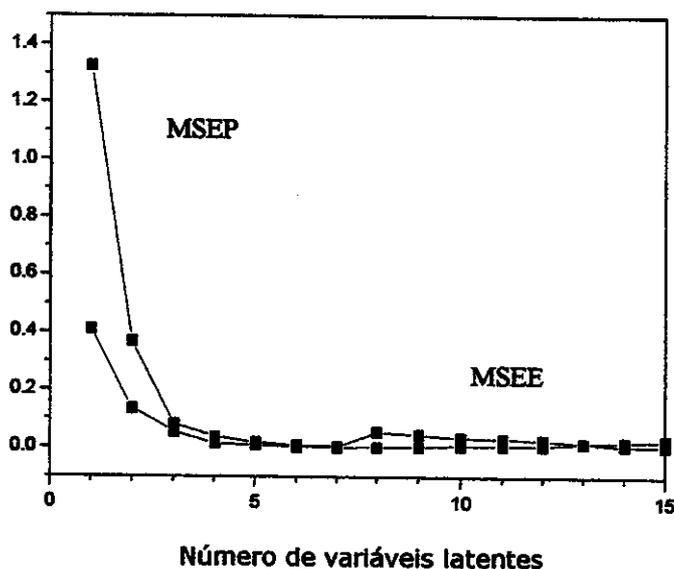


Figura 50: Erro médio quadrático de previsão (MSEP) e da estimativa (MSEE) em função do número de variáveis latentes para as concentrações reais e previstas para *p*-aminofenol.

Pelo exposto anteriormente 4, 3 e 4 variáveis latentes seriam as mais adequadas para serem usadas nas construções dos modelos de N-PLS, para as análises de *o*-aminofenol, *m*-aminofenol e *p*-aminofenol respectivamente. Teria-se assim 2 modelos: um para a análise simultânea de *o*-, e *p*-aminofenol, e outro modelo para *m*-aminofenol. Desta forma, decidiu-se usar 4 variáveis latentes para ter um modelo único de análise simultânea dos três isômeros, uma vez que o aumento de uma variável latente para construir o modelo de N-PLS (usando informação das concentrações dos três isômeros), para *m*-aminofenol, não afeta de maneira significativa nos resultados de previsão.

Assim foi construído o modelo de N-PLS com 4 variáveis latentes obtendo-se 97,69% de variância explicada para \underline{X} (tensor de dados de variáveis independentes) e 98,80% para \underline{Y} (matriz de dados de variáveis dependentes).

Depois de obtido o modelo, estudou-se a correlação existente entre as concentrações reais e previstas pelo modelo para os isômeros no conjunto de amostras usadas tanto na calibração como na previsão do modelo.

Existe uma correlação aceitável entre os valores reais e previstos pelo modelo de PLS tridimensional, para a quantificação do isômero *o*-aminofenol, como é mostrado na figura 51. Os resultados apresentam coeficiente de correlação 0.9984.

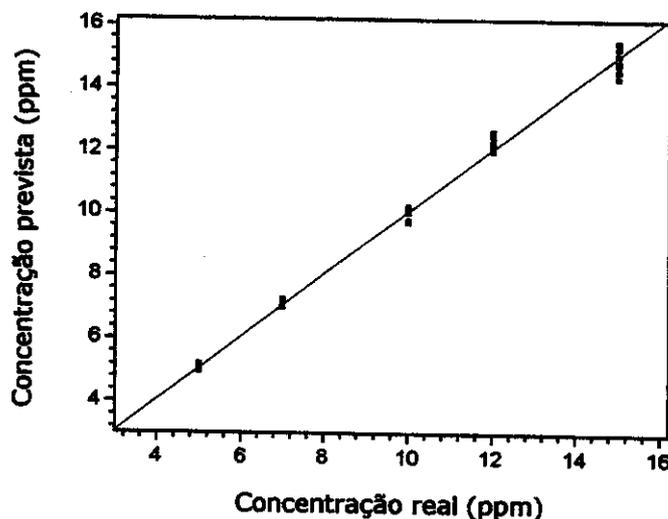


Figura 51: Concentração prevista pelo modelo de PLS tridimensional contra concentração real para *o*-aminofenol

Na figura 52, encontra-se o gráfico dos valores reais e previstos pelo modelo de PLS tridimensional para as concentrações do isômero *m*-aminofenol.

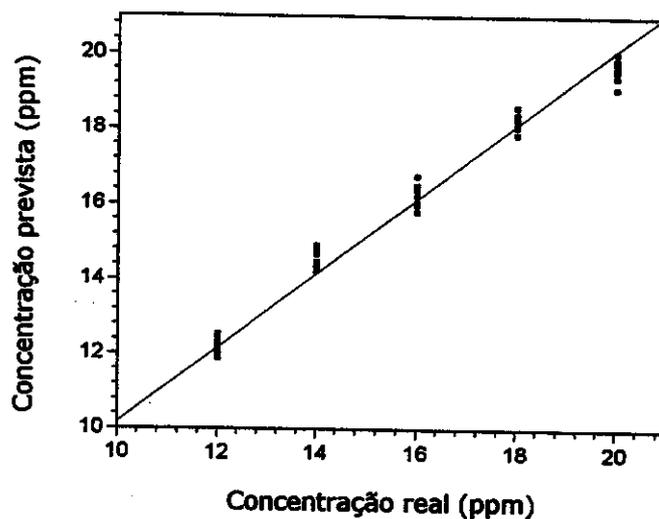


Figura 52: Concentração prevista pelo modelo de PLS tridimensional contra a concentração real para *m*-aminofenol

Novamente na figura 52 uma boa correlação é obtida entre as concentrações reais e previstas pelo modelo de PLS tridimensional, para a quantificação do isômero *m*-aminofenol, obtendo-se coeficiente de correlação de 0.9955.

Da mesma forma na figura 53, observa-se que existe boa correlação entre as concentrações reais e previstas pelo modelo de N-PLS para o isômero *p*-aminofenol. Esta boa correlação é verificada pelo valor do coeficiente de correlação que é 0.9946.

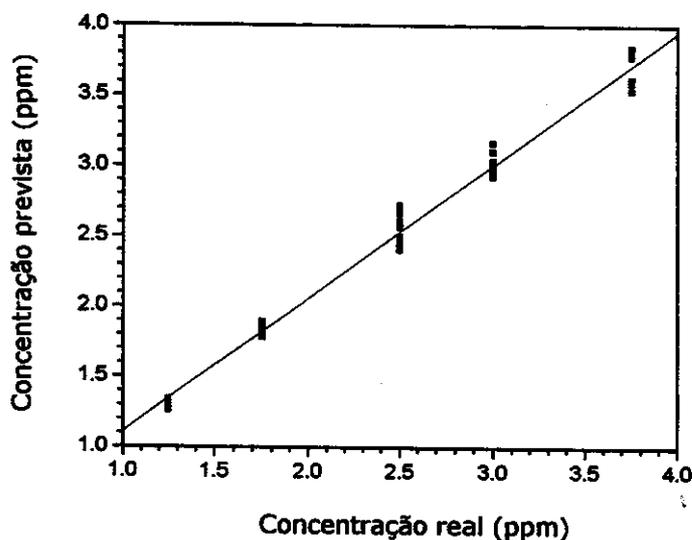


Figura 53: Concentração prevista pelo modelo de PLS tridimensional contra a concentração real para *p*-aminofenol

As tabelas 4, 5, e 6, mostram as concentrações previstas pelo modelo de N-PLS calculado com 4 variáveis latentes, para os isômeros *o*-, *m*- e *p*-aminofenol respectivamente, pertencentes a um conjunto de 6 misturas diferentes (preparadas em triplicata) não usadas na construção do modelo. A similaridade das estimativas do desvio padrão de cada isômero apresentados nestas tabelas, indicam a boa precisão da metodologia cinética utilizada em combinação com o método de N-PLS.

Tabela 4: Concentrações de *o*-aminofenol previstas em triplicata pelo modelo N-PLS construído.

Mistura	Concentrações previstas (ppm)			Estimativa do desvio padrão
30	7,25	7,12	7,20	0,06
31	12,01	12,16	12,04	0,08
32	7,01	7,07	7,13	0,06
33	12,14	12,18	12,31	0,09
34	7,19	7,09	7,11	0,05
35	12,24	12,46	12,62	0,19

Tabela 5: Concentrações de *m*-aminofenol previstas em triplicata pelo modelo N-PLS construído.

Mistura	Concentrações previstas (ppm)			Estimativa do desvio padrão
30	14,73	14,67	14,23	0,27
31	18,25	18,59	18,42	0,17
32	18,11	18,38	18,08	0,16
33	14,23	14,83	14,68	0,31
34	14,92	14,47	14,32	0,29
35	18,11	17,86	18,18	0,17

Tabela 6: Concentrações de *p*-aminofenol previstas em triplicata pelo modelo N-PLS construído.

Mistura	Concentrações previstas (ppm)			Estimativa do desvio padrão
30	1,86	1,88	1,81	0,04
31	2,93	2,97	3,04	0,05
32	2,99	2,94	3,01	0,04
33	1,89	1,78	1,85	0,06
34	3,11	3,17	2,96	0,11
35	1,84	1,82	1,87	0,03

Para avaliar o poder de previsão do modelo N-PLS construído, foram comparadas as concentrações previstas de cada isômero pelo modelo de N-PLS com suas respectivas

concentrações reais para o conjunto de misturas não usadas no modelo. O critério adotado para avaliar a previsão foi calcular o erro relativo e o erro padrão de previsão (RMSEP). As tabelas 7, 8 e 9 apresentam as concentrações reais e previstas pelo modelo de NPLS para cada isômero em 6 amostras diferentes. Em cada tabela está apresentado o valor do erro relativo para cada isômero em cada mistura e o valor de RMSEP de cada isômero para todo esse conjunto de misturas. Estes valores obtidos indicam que o modelo de NPLS tridimensional construído consegue prever de maneira bastante aceitável as concentrações de isômeros *o*-, *m*- e *p*-aminofenol em misturas ternárias de isômeros.

Tabela 7: Resultados da concentração prevista de *o*-aminofenol nas amostras teste usando PLS tridimensional

Mistura	Concentração real (ppm)	Concentração prevista (ppm) Média de 3 estimativas	Erro relativo %
30	7,00	7,19	3
31	12,00	12,07	1
32	7,00	7,07	1
33	12,00	12,21	2
34	7,00	7,13	2
35	12,00	12,44	4

RMSEP: 0.22

Tabela 8: Resultados da concentração prevista de *m*-aminofenol nas amostras teste usando PLS tridimensional

Mistura	Concentração real (ppm)	Concentração prevista (ppm) Média de 3 estimativas	Erro relativo %
30	14,00	14,54	4
31	18,00	18,42	2
32	18,00	18,19	1
33	14,00	14,57	4
34	14,00	14,57	4
35	18,00	18,05	1

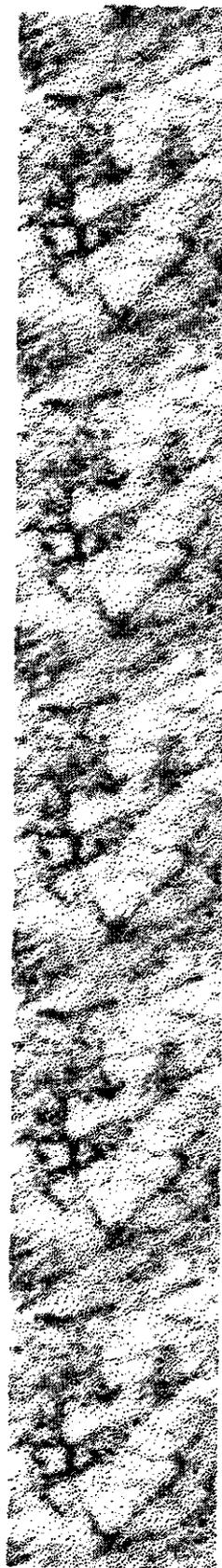
RMSEP: 0.44

Tabela 9: Resultados da concentração prevista de *p*-aminofenol nas amostras teste usando PLS tridimensional

Mistura	Concentração real (ppm)	Concentração prevista (ppm) Média de 3 estimativas	Erro relativo %
30	1,75	1,85	6
31	3,00	2,98	1
32	3,00	2,98	1
33	1,75	1,84	5
34	3,00	3,08	3
35	1,75	1,84	5

RMSEP: 0.08

CONCLUSÕES



CONCLUSÕES

O trabalho desenvolvido demonstrou que os métodos quimiométricos de ordem superior podem fornecer um grande potencial em aplicações analíticas tanto em determinações quantitativas como qualitativas.

Foi demonstrado que os métodos de PARAFAC1 e PARAFAC2 conseguiram modelar as informações cinéticas espectrofotométricas de misturas de isômeros aminofenol, obtendo os espectros puros e as concentrações relativas dos produtos formados nas reações (*o*-, *m*- e *p*-azofenol). Destacando que o método de PARAFAC2 forneceu melhores resultados na previsão do espectro de *m*-azofenol que o método de PARAFAC1.

Além disso, o método de PARAFAC2 também levou a resultados satisfatórios na previsão das curvas cinéticas do *o*-azofenol (formado a partir do *o*-aminofenol), nas diferentes misturas de isômeros aminofenol.

Foram obtidos resultados aceitáveis na previsão das concentrações dos isômeros *o*-, *m*- e *p*-aminofenol na mistura (no início da reação), usando um método indireto, onde se fez uma regressão linear entre as concentrações reais dos isômeros e as concentrações relativas dos produtos formados (isômeros azofenol), na qual se assume que são iguais às concentrações relativas iniciais dos reagentes *o*-, *m*- e *p*-aminofenol.

Resultados do estudo preliminar da cinética de aminofenol com nitrito, mostrou que a cinética entre *o*-aminofenol e nitrito em pH=4, não é uma reação simples pois envolve várias etapas de acordo a ordem da reação calculada que foi 0,78. Também, foi demonstrado que

existe a possibilidade de sinergismo nas reações, já que existe um comportamento cinético diferente entre a reação de nitrito desenvolvida com o isômero *o*-aminofenol puro e com este isômero em uma mistura.

O método de PARAFAC1 pode ser aplicado para dados obtidos pelo monitoramento espectrofotométrico de uma ou mais reações de primeira ordem, onde se têm perfis cinéticos constantes em cada reação.

O método de PARAFAC2 pode ser aplicado para a modelagem de dados obtidos pelo monitoramento espectrofotométrico de reações que não têm perfis cinéticos constantes, como por exemplo reações afetadas por sinergismo.

O modelo de N-PLS construído para a determinação simultânea das concentrações dos isômeros *o*-, *m*- e *p*-aminofenol em misturas forneceu bons resultados na previsão das concentrações destes isômeros em misturas, tanto em amostras usadas para construir o modelo de calibração como em novas amostras usadas para sua validação.

Se o interesse for só a determinação da concentração de 1 ou mais analitos é recomendável usar N-PLS ou PLS pois são mais simples de utilizar que os modelos de PARAFAC1 e PARAFAC2.

APÊNDICES



APÊNDICE 1

ALGORITMO DO ALS

Os princípios de mínimos quadrados alternados⁵⁶ (ALS) foi proposto em 1933 por Yates.⁵⁷ Este algoritmo consiste em expressar cada parâmetro de uma equação, em uma função formada pelos parâmetros restantes. Assim, se há "n" parâmetros então têm-se "n" equações. Calcula-se cada parâmetro das equações, e após isso novamente é repetido este processo até observar-se uma mudança mínima nos parâmetros calculados.

Seja um modelo bilinear:

$$\mathbf{X} = \mathbf{A}\mathbf{B}^T \quad (1)$$

A estimativa de A e B simultaneamente pode ser feita usando a função de mínimos quadrados:

$$\min_{\mathbf{A}, \mathbf{B}} \|\mathbf{X} - \mathbf{A}\mathbf{B}^T\|_F^2 \quad (2)$$

Observa-se que a equação (2) é difícil de resolver, enquanto que estimando B em função de A e X, é mais simples:

$$\mathbf{B} = \mathbf{X}^T (\mathbf{A}^+)^T \quad (3)$$

Isto não resolve todo o problema, mais fornecerá apenas uma estimativa aproximada de B. A estimativa de A pode ser feita a partir de B:

$$\mathbf{A} = \mathbf{X}(\mathbf{B}^+)^T \quad (4)$$

Dessa maneira, uma nova estimativa de B pode ser feita pela equação 3, e assim sucessivamente até que os valores de A e B sejam tais que X é mínimo. Esta é a idéia básica

do algoritmo de mínimos quadrados alternados. Isto é, transformar a equação 2 em duas equações, e logo solucionar iterativamente essas equações até um critério de convergência previamente estabelecida. Por exemplo, para um conjunto de dados \underline{X} , expresso em função de A, B, C, \dots , descrito como:

$$\underline{X} = f(A, B, C, \dots) + E \quad (5)$$

O algoritmo ALS pode ser formulado para estimar os parâmetros A, B, C, \dots de acordo com o algoritmo mostrado abaixo.

ALGORITMO DE ALS

1. Inicializar os parâmetros A, B, C, \dots , com valores aleatórios.
2. Calcular A da solução da equação:

$$\arg \min_A \|\underline{X} - f(A, B, C, \dots)\|_f^2 \quad (6)$$

3. Calcular B da equação:

$$\arg \min_B \|\underline{X} - f(A, B, C, \dots)\|_f^2 \quad (7)$$

4. Calcular C da equação:

$$\arg \min_C \|\underline{X} - f(A, B, C, \dots)\|_f^2 \quad (8)$$

5. Estimar os próximos parâmetros de modo similar.
6. Voltar à etapa 2 até a convergência.

A primeira etapa de inicialização dos parâmetros é realizada sob as condições de cada algoritmo específico. A última etapa avalia a convergência. Geralmente o algoritmo iterativo termina quando os valores dos parâmetros calculados não variam muito.

APÊNDICE 2

ALGORITMO DO PARAFAC 1

O modelo de PARAFAC 1 para um conjunto de dados tridimensionais, pode ser expresso como:

$$\mathbf{X}_{I \times JK} = \mathbf{A} (\mathbf{C} \otimes \mathbf{B})^T \quad (1)$$

Também pode ser escrito como:

$$\mathbf{X}_{J \times IK} = \mathbf{B} (\mathbf{C} \otimes \mathbf{A})^T \quad (2)$$

Ou:

$$\mathbf{X}_{K \times IJ} = \mathbf{C} (\mathbf{B} \otimes \mathbf{A})^T \quad (3)$$

O ajuste do modelo de PARAFAC1 pode ser expresso como:

$$\min_{\mathbf{A}, \mathbf{B}, \mathbf{C}} \|\mathbf{X}' - \mathbf{A}(\mathbf{C} \otimes \mathbf{B})^T\|^2 \quad (4)$$

Assim, o modelo de PARAFAC é obtido quando se minimiza a diferença quadrática entre os dados originais \mathbf{X}' e o modelo \mathbf{X} , sob um critério de convergência previamente estabelecido.

O algoritmo abaixo mostra o cálculo dos parâmetros \mathbf{A} , \mathbf{B} e \mathbf{C} do modelo de PARAFAC1 para um conjunto de dados tridimensionais arranjados numa matriz $I \times JK$.

ALGORITMO GERAL PARAFAC-ALS

1. Decidir o número de componentes.
2. Inicializar os parâmetros **B** e **C**.
3. Estimar **A** de **X**, **B**, **C** por regressão de mínimos quadrados.
4. Estimar **B** de **X**, **A**, **C** da mesma maneira anterior.
5. Estimar **C** da mesma maneira que em (3) e (4).
6. Verificar a convergência. Se não houver voltar para 3.

Cada vez que se calcula os valores de **A**, **B** e **C** a partir das etapas 3, 4 e 5 diz-se que ocorreu uma iteração.

1.- Determinação do número de componentes

A determinação do número de componentes pode ser realizada pela análise dos resíduos. Consiste em calcular modelos para diferentes números de componentes. Para cada modelo construído é calculado o resíduo, obtido pela diferença entre os dados originais X' e do modelo construído X . A partir disso, os elementos do resíduo **E** são elevados ao quadrado, e somados entre si, sendo este resultado conhecido como soma quadrática dos elementos do resíduo (SQR). Este pode ser expresso matematicamente como:

$$SQR = \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K (x'_{ijk} - x_{ijk})^2 \quad (5)$$

Onde:

- x'_{ijk} é um elemento do conjunto de dados tridimensionais originais.
- x_{ijk} é um elemento do conjunto de dados tridimensionais obtido pelo modelo de PARAFAC 1.

Calculado o valor de SQR para cada modelo, o passo seguinte é graficar os valores da somatória quadrática dos resíduos SQR contra o número de componentes.

O número de componentes adequado será aquele que apresente um valor pequeno de SQR e a diferença entre ele e o valor do SQR para o número de componentes anterior seja significativa.

2. - Inicialização dos parâmetros

Vários tipos de inicialização foram propostos por Harshman & Lundy em 1984.²² Foi proposta a inicialização dos parâmetros por o uso de valores aleatórios e começar a solução de vários pontos iniciais. Também foi proposta^{32,58-60} a inicialização dos parâmetros baseado na decomposição dos dados por SVD ou métodos similares como TLD.

Assim, pode-se inicializar os parâmetros A, B e C com TLD que decompõe os dados de X, de acordo à equação 1.

3. Cálculo de A

Fazendo:

$$\mathbf{Z} = \mathbf{C} \otimes \mathbf{B} \quad (6)$$

Substituindo na equação (1) tem-se:

$$\mathbf{X} = \mathbf{AZ}^T \quad (7)$$

Multiplicando por Z ambos lados e depois por $(\mathbf{Z}^T\mathbf{Z})^+$ tem-se:

$$\mathbf{XZ} (\mathbf{Z}^T\mathbf{Z})^+ = \mathbf{AZ}^T\mathbf{Z} (\mathbf{Z}^T\mathbf{Z})^+ \quad (8)$$

$$\mathbf{A} = \mathbf{XZ} (\mathbf{Z}^T\mathbf{Z})^+ \quad (9)$$

4. Calculo de B

Fazendo:

$$\mathbf{Z} = \mathbf{C} \otimes \mathbf{A} \quad (10)$$

Substituindo na equação (2) tem-se:

$$\mathbf{X} = \mathbf{BZ}^T \quad (11)$$

Multiplicando por \mathbf{Z} ambos lados e depois por $(\mathbf{Z}^T\mathbf{Z})^+$ tem-se:

$$\mathbf{XZ}(\mathbf{Z}^T\mathbf{Z})^+ = \mathbf{BZ}^T\mathbf{Z}(\mathbf{Z}^T\mathbf{Z})^+ \quad (12)$$

$$\mathbf{B} = \mathbf{XZ}(\mathbf{Z}^T\mathbf{Z})^+ \quad (13)$$

5. Cálculo de C

Fazendo:

$$\mathbf{Z} = \mathbf{B} \otimes \mathbf{A} \quad (14)$$

Substituindo na equação (3) tem-se:

$$\mathbf{X} = \mathbf{CZ}^T \quad (15)$$

Multiplicando por \mathbf{Z} ambos lados e depois por $(\mathbf{Z}^T\mathbf{Z})^+$ tem-se:

$$\mathbf{XZ}(\mathbf{Z}^T\mathbf{Z})^+ = \mathbf{CZ}^T\mathbf{Z}(\mathbf{Z}^T\mathbf{Z})^+ \quad (16)$$

$$\mathbf{C} = \mathbf{XZ}(\mathbf{Z}^T\mathbf{Z})^+ \quad (17)$$

6. Verificação da convergência

Esta etapa do algoritmo verifica se o algoritmo conseguiu alcançar o critério de convergência previamente estabelecido. O critério de convergência por definição do programa utilizado (que calcula o modelo de PARAFAC 1) é 10^{-6} , isto quer dizer que se em uma interação a diferença entre o valor do SQR atual com o valor SQR da interação anterior está abaixo de 10^{-6} , o algoritmo convergiu. A medida que os dados contenham mais ruído é aconselhável que o critério de convergência aumente para não incluir o ruído na modelagem.

Usando o modelo de PARAFAC 1 sobre conjunto de dados de novas amostras

Os novos *scores* (*loading* que pertence à dimensão das amostras) para um novo conjunto de dados X^{nov} podem ser calculados mantendo fixas as matrizes *scores* dos dois modos restantes.

Se os *scores* pertencem ao primeiro modo então se estabelece a seguinte equação para calcular os novos *scores* A^{nov} :

$$X^{\text{nov}} = A^{\text{nov}} (C \otimes B)^T = A^{\text{nov}} Z^T \quad (18)$$

Resolvendo esta equação obtém-se:

$$A^{\text{nov}} = XZ (Z^T Z)^{-1} \quad (19)$$

Modelo de PARAFAC 1 para ordens maiores que 3

O modelo de PARAFAC1 também pode ser aplicado para dados multidimensionais de ordens maiores que 3. Por exemplo, para dados multidimensionais de quarta ordem o modelo de PARAFAC1 aplicado a esses dados pode ser estabelecido como:

$$X_{I \times JKM} = A (D \otimes C \otimes B)^T + E \quad (20)$$

O cálculo de A é realizado de maneira similar ao anteriormente visto. Neste caso Z é definido como:

$$Z = D \otimes C \otimes B \quad (21)$$

E A pode ser calculado como:

$$A = XZ (Z^T Z)^{-1} \quad (22)$$

Assim, observa-se que o cálculo dos parâmetros de um conjunto de dados de ordem mais alta que três é muito similar ao algoritmo proposto para um conjunto dados de terceira ordem.

APÊNDICE 3

ALGORITMO DO PARAFAC 2

H. A. L. Kiers⁶¹ em 1993 propôs um algoritmo para ajustar o modelo de PARAFAC2, que era bastante complexo e, em 1998, propôs outro algoritmo,⁶² mais simples.

Sabe-se que PARAFAC 2 calcula cada fatia do tensor de dados tridimensional do modelo X como:

$$X_k = AD_K (P_k B)^T \quad (1)$$

Esta equação também pode ser expressa como:

$$X'_k = AD_K (P_k B)^T + E_k \quad (2)$$

onde:

X'_k é a k -ésima fatia do tensor de dados originais.

E_k é o resíduo da k -ésima fatia.

O ajuste do modelo de PARAFAC2 é dado pela seguinte expressão:

$$\min_{A, C, B, P_k} k = \sum_{k=1}^{k=k} \| X'_k - AD_K (P_k B)^T \|^2 \quad (3)$$

O algoritmo apresentado abaixo mostra o PARAFAC2 para modelagem de um conjunto de dados tridimensionais.

ALGORITMO DE PARAFAC2 ALS

1. Determinar o número de componentes (F).
2. Inicializar os parâmetros **A**, **B**, e **C**.
3. Para cada valor de **k**, onde **k** varia de 1 até **K**, calcular:
 P_k de X_k , **A**, D_k , e **B**
4. Para cada valor de **k**, calcular:
 Y_k de X_k e P_k e formar $Y_{I \times FK}$ com as matrizes Y_k calculadas.
5. Calcular **A**, **B**, e **C** para uma iteração de PARAFAC / ALS sob $Y_{I \times FK}$.
6. Verificar a convergência. Se não houver voltar para 3.

Cada vez que se calculam os valores de **A**, **B**, **C** e P_k a partir das etapas 3, 4 e 5, diz-se que houve uma iteração.

1. Determinação do número de componentes

A escolha do número de componentes adequado para ser usado no modelo de PARAFAC2 é semelhante ao realizado para o PARAFAC 1. O valor de SQR é calculado para cada componente, segundo a equação:

$$SQR = \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K (x'_{ijk} - x_{ijk}) \quad (4)$$

onde:

- x'_{ijk} é um elemento do conjunto de dados tridimensionais originais.
- x_{ijk} é um elemento do conjunto de dados tridimensionais obtido pelo modelo de PARAFAC 2.

2. - Inicialização dos parâmetros

O uso de valores aleatórios é sugerido por Rasmus Bro, para inicializar os parâmetros A, B e C de PARAFAC 2, e este é o procedimento normalmente utilizado.

3. Cálculo de P_k

O valor de P_k pode ser calculado pela seguinte expressão:

$$\arg \min_{P_k} \| X_k - AD_k(P_k B)^T \|^2 \quad (5)$$

Fazendo:

$$Y_k = AD_k B^T \quad (6)$$

Substituindo a equação 6 em 5, tem-se:

$$\arg \min_{P_k} \| X_k - Y_k P_k^T \|^2 \quad (7)$$

A solução desta expressão é dada por⁴⁷ :

$$P_k = X_k^T Y_k (Y_k^T Y_k)^{-1} \quad (8)$$

Utilizando a condição que $P_k P_k^T = I$, onde I é a matriz identidade, tem-se:

$$P_k = X_k^T Y_k (Y_k^T X_k X_k^T Y_k)^{-1/2} \quad (9)$$

Fazendo:

$$Q_k = X_k^T Y_k \quad (10)$$

Substituindo a equação anterior em 9, pode-se simplificar a expressão de cálculo de P_k como:

$$P_k = Q_k (Q_k^T Q_k)^{-1/2} \quad (11)$$

4. Cálculo de Y_k e Y

O cálculo de Y_k pode ser feito pela equação (6). Sabe-se que X_k tem dimensões $I \times J$, e que P_k tem dimensões $J \times F$, portanto Y_k terá dimensões de $I \times f$. Agrupando K fatias Y_k , de acordo com a figura 1, se terá uma matriz Y de dimensões $I \times fK$.



Figura 1: Arranjo das matrizes Y_k em uma só matriz Y de dimensões $I \times fK$.

5. Cálculo de A , B e C

Da equação 6, pode-se observar que cada fatia Y_k tem a mesma forma de modelagem feito por PARAFAC 1. Assim, pode-se modelar Y pelo modelo de PARAFAC 1 e calcular os parâmetros A , B e C . Observar que B tem dimensões $f \times f$.

6. Verificação da convergência

A cada iteração calcula-se a matriz $X_{I \times JK}$ (matriz do modelo), a partir das K fatias X_k ($I \times J$) formado pelos parâmetros A , B , C e P_k . Depois, calcula-se a matriz resíduo a partir da diferença entre a matriz obtida pelo modelo X e a matriz de dados originais X' . Calcula-se o valor de SQR (somatória quadrática dos elementos do resíduo) e compara-se com o valor de SQR da iteração anterior. Se existir diferença deste dois valores de SQR (por exemplo 10^{-6}) menor ou igual ao critério estabelecido, então se diz que o algoritmo convergiu. O critério de convergência pode ser escolhido de maneira similar ao critério para o PARAFAC 1.

Usando o modelo de PARAFAC 2 em um conjunto com novas amostras

Calcular os *scores* (*loading* pertencente à dimensão das amostras) de novas amostras (concentrações relativas) é uma tarefa bastante complexa. Se os *scores* encontram-se na terceira dimensão, então para um conjunto de novas amostras, pode calcular-se os *scores* novos C^{nov}_k e a matrix P_k^{nov} pelo algoritmo de PARAFAC2 mantendo fixos os valores A e B obtidos pela modelagem anterior.

Modelo de PARAFAC 2 para ordem maior que 3

Os dados multidimensionais de ordem maior que três podem ser modelados de forma similar para dados de terceira ordem. Por exemplo, um conjunto de dados de quarta ordem de dimensões $I \times J \times K \times M$, desdobrado em uma matrix X de dimensões $IJ \times KM$ pode ser modelado por PARAFAC 2 sempre que os *loadings* da terceira dimensão (K) variem em função das variáveis da quarta dimensão (M).

O primeiro passo para a modelagem por PARAFAC 2 é estabelecer submatrizes de X em função das variáveis da quarta dimensão, como:

$$X_m = (A \otimes B) F_m C_m^T \quad (12)$$

onde:

- F é o número de componentes.
- A é uma matrix $I \times F$
- B é uma matrix $J \times F$
- F_m é matrix diagonal formada pelos elementos da m -ésima linha de D
- D é uma matrix $M \times F$
- m é a m -ésima variável da quarta dimensão ou modo.
- C_m é uma matrix $K \times F$
- X_m é uma matrix $I J \times K$

Fazendo:

$$C_m = P_m C \quad (13)$$

onde:

- P_m é uma matriz $K \times F$
- C é uma matriz $F \times F$

Obtém-se:

$$X_m = (A B) F_m (P_m C)^T \quad (14)$$

Da mesma forma que para um conjunto de dados tridimensional, o cálculo de P_m pode ser obtido pela solução da seguinte expressão:

$$\arg \min_{P_k} \| X_m - M_m P_m^T \|^2 \quad (15)$$

onde:

$$M_m = (A B) F_m C^T \quad (16)$$

Os cálculos dos parâmetros A , B , C , D podem ser realizados aplicando PARAFAC 1 a uma matriz de Y de dimensão $I \times J \times F \times M$, formada a partir das M matrizes Y_m cada uma calculada como $X_m P_m$.

APÊNDICE 4

ALGORITMO DE N-PLS

Nesta parte será discutido o algoritmo de Mínimos quadrados parciais multidimensionais (N-PLS) para dados de variáveis independentes de terceira ordem \underline{X} com dimensões $I \times J \times K$ (arranjados numa matrix $I \times JK$ para fazê-lo compatível com o algoritmo) e para os dados das variáveis dependentes de segunda ordem Y com dimensões $I \times N$. Neste caso, I indica o número de amostras. O algoritmo é mostrado abaixo.

ALGORITMO DO N-PLS

Fazer: $X_{res} = X$, $Y_{res} = Y$, e fazer $it = 0$. Onde, it é o número de iterações.

Inicializar $f = 1$

1. Inicializar os vetores u_f e t_f .
2. Fazer: $tgl = t_f$ e $it = it + 1$.
3. Calcular w_f de X_{res} e u_f .
4. Calcular w_f^J e w_f^K de w_f .
5. Calcular t_f de X_{res} e os vetores calculados na etapa 4.
6. Calcular q_f de Y_{res} e t_f .
7. Calcular u_f de Y_{res} e q_f .
8. Verificar a convergência. Se houver continuar, senão voltar para 2.
9. Calcular b_f de u_f e t_f .
10. Calcular os resíduos X_{res} e Y_{res}

11. Fazer $f = f + 1$, voltar à etapa 1. Repetir até que f seja igual a F , onde F é o número de variáveis latentes usadas na construção do modelo, para o qual se obtém uma apropriada descrição de Y .

1. Inicializar os vetores u_f e t_f

Os vetores u_f e t_f são inicializados por valores aleatórios, com elementos compreendidos entre 0 e 1.

3. Calcular w_f

O vetor de w_f , de dimensões $JK \times 1$ é calculado a partir de X_{res} e o vetor u_f , de acordo a seguinte equação:

$$w_f = X_{res}^T u_f \quad (1)$$

O vetor calculado é arranjado numa matriz W_f de dimensões $J \times K$.

4. Calcular w_f^J e w_f^K

Estes vetores unitários são calculados pela decomposição por valores singulares SVD da matriz W_f , de acordo a seguinte expressão matemática:

$$[w_f^J, s, w_f^K] = \text{svd}(W_f, 1) \quad (2)$$

5. Calcular t_f

O vetor t_f pode ser calculado pela seguinte equação:

$$t_f = X_{res} (w_f^K \otimes w_f^J) \quad (3)$$

6. Calcular q_f

O cálculo do vetor unitário q_f pode ser feito a partir de Y_{res} e t_f como:

$$\mathbf{q}_f = \frac{\mathbf{Y}_{res}^T \mathbf{t}_f}{|\mathbf{Y}_{res}^T \mathbf{t}_f|} \quad (4)$$

7. Calcular \mathbf{u}_f

O vetor \mathbf{u}_f é calculado pela seguinte equação:

$$\mathbf{u}_f = \mathbf{Y}_{res} \mathbf{q}_f \quad (5)$$

8. Verificar a convergência

Se $\frac{|\mathbf{t}_f - \mathbf{t}_{gl}|}{|\mathbf{t}_f|} > 10^{-8}$ e it (iteração) é menor que 20, voltar a etapa 2, senão continuar com as seguintes etapas do algoritmo:

9. Calcular \mathbf{b}_f

Sabe-se que:

$$\mathbf{u}_f = \mathbf{t}_f \mathbf{b}_f \quad (6)$$

Multiplicando ambos lados por \mathbf{t}^T e depois por $(\mathbf{t}^T \mathbf{t})^{-1}$ tem-se:

$$\mathbf{b}_f = (\mathbf{t}^T \mathbf{t})^{-1} \mathbf{t}^T \mathbf{u}_f \quad (7)$$

10. Calcular \mathbf{X}_{res} e \mathbf{Y}_{res}

Estas matrizes residuais são calculadas pelas seguintes equações:

$$\mathbf{X}_{res} = \mathbf{X}_{res} - \mathbf{t}_f (\mathbf{w}_f^K \otimes \mathbf{w}_f^J)^T \quad (8)$$

$$\mathbf{Y}_{res} = \mathbf{Y}_{res} - \mathbf{u}_f \mathbf{q}_f^T \quad (9)$$

Fazendo previsões para novas amostras

As variáveis dependentes $Y^{\text{nov}}o$ de um conjunto de dados de novas amostras $X^{\text{nov}}o$ podem ser calculadas a partir dos *loadings* conhecidos w_f e q_f , de acordo com o seguinte algoritmo:

1. fazer: $X_{\text{res}} = X^{\text{nov}}o$
2. Inicializar $f = 1$
3. Calcular $t_f = X_{\text{res}} (w_f^K \otimes w_f^J)$
4. Calcular $X_{\text{res}} = X_{\text{res}} - t_f (w_f^K \otimes w_f^J)$
5. Fazer $f=f + 1$, e voltar para a etapa 1, e continuar até $f= F$, onde F é o número de variáveis latentes usadas na construção do modelo de calibração.

Os F vetores t_f podem ser arranjados numa matriz $T^{\text{nov}}o$ de dimensões $I \times F$. Conhecido os *scores* de $X^{\text{nov}}o$ pode-se calcular os *scores* de $Y^{\text{nov}}o$ pela seguinte equação:

$$U^{\text{nov}}o = T^{\text{nov}}o B \quad (10)$$

A matriz B contém os F vetores b_f obtidos no modelo de calibração.

Substituindo esta equação em:

$$Y^{\text{nov}}o = U^{\text{nov}}o Q^T \quad (11)$$

onde a matriz Q contém os F vetores q_f obtidos no modelo de calibração.

Pode-se calcular a matriz $Y^{\text{nov}}o$, de acordo a seguinte equação:

$$Y^{\text{nov}}o = T B Q^T \quad (12)$$

APÊNDICE 5

PARÂMETROS ESTATÍSTICOS

Os parâmetros estatísticos utilizados para avaliação dos resultados obtidos pelos modelos Quimiométricos foram os seguintes:

1. - Erro médio quadrático de previsão²⁷ (MSEP)

É um método de validação externa, calculado a partir da diferença quadrática média entre valores de concentração real e prevista para um conjunto de amostras não presentes na calibração.

$$MSEP = \Sigma(yr - yp)^2 / N_p \quad (1)$$

- yr é o valor real

- yp é o valor previsto pelo modelo.

- N_p é o número de amostras de previsão.

2. - Erro quadrático médio da estimativa²⁷ (MSEE)

É um método de validação interna, calculado pela diferença quadrática média entre valores de concentração real e prevista das próprias amostras da calibração.

$$MSEE = \Sigma(yr - yp)^2 / N_c \quad (2)$$

- yr é o valor real

- yp é o valor previsto pelo modelo.

- N_c é o número de amostras de calibração.

3. - Erro padrão de previsão²⁷ (RMSEP)

O cálculo do erro padrão de previsão é usado para avaliar o poder de previsão do modelo, sendo dado pela equação:

$$RMSEP = \sqrt{\frac{\sum (y_r - y_p)^2}{N}} \quad (3)$$

- y_r é o valor real.

- y_p é o valor previsto.

- N pode ser, dependendo o caso, número de amostras ou número de variáveis, como por exemplo, número de comprimentos de onda.

4. - Estimativa do desvio padrão⁶³ (S)

É calculado a partir do quadrado da diferença entre o valor médio de concentrações previstas e a concentração estimada.

$$S = \sqrt{\frac{\sum_{i=1}^{i=N} (y_i - y_m)^2}{N-1}} \quad (4)$$

- y_i é o valor previsto.

- Y_m é o valor médio calculado a partir de N estimativas de y .

- N é o número de estimativas.

5. - Erro relativo⁶³ (Er)

É um critério empregado para avaliar o erro de previsão, sendo definido pela seguinte equação:

$$Er = \frac{(y_r - y_p)}{y_r} \times 100 \quad (5)$$

- y_r é o valor real.

- y_p é o valor previsto.

REFERÊNCIAS BIBLIOGRÁFICAS

1. B. Wise, N. Gallagher, "PLS_Toolbox 2.0", Eigenvector Research Inc., U.S.A., 1998.
2. P. Geladi, "Analysis of multi-way (multimode) data", *Chem. Intell. Lab. Syst.*, **7**, 11, 1989.
3. L. Kiers, "Three-way methods for the analysis qualitative and quantitative two way data", DSWO Press, Leiden, 1989.
4. K. Esbensen, S. Wold, P. Geladi, "Relationships between higher-order data array configurations and problem formulations in multivariate data analysis", *J. Chemom.*, **3**, 33, 1988.
5. N. Ho, G. Christian, E. Davidson, "Application of the method of rank annihilation to quantitative analysis of multicomponent fluorescence from the video fluorometer", *Anal. Chem.*, **50**, 1108, 1978.
6. J. Johnson, R. Yost, "Tandem mass spectrometry for trace analysis", *Anal. Chem.*, **57**, 758A, 1985.
7. P. Hindmarch, K. Kavianpour, RG Brereton., " Evaluation of parallel factor analysis for the resolution of kinetic data by diode array high performance liquid chromatography", *Analyst*, **122**, 871, 1997.
8. R. Tauler, I. Marqués, E. Casassas, "Multivariate curve resolution applied to three-way trilinear data: study of a spectrofluorimetric acid-base titration of salicylic acid at three excitation wavelengths", *J. Chemom.*, **12**, 55, 1998.
9. R. Tauler, K. Smilde, M. Henshaw, W. Burgess, B. Kowalski, "Multicomponent determination of chlorinated hydrocarbons using a reaction based chemical sensor. 2. Chemical speciation using multivariate curve resolution", *Anal. Chem.*, **66**, 3337, 1994.
10. S. Swinbourne, "Analysis of kinetic data", Ed. Nelson, London, 1971.
11. B. Yatsimirskii, "Kinetic methods of analysis", Pergamom Press, London, 1966.

12. M. Blanco, J. Gene, H. Ituriaga, S. Maspach, J. Riba, "Diode array detectors in flow injection analysis mixture resolution by multi wavelength analysis", *Talanta*, 34, 987, 1987.
13. A. Carey, J. Sundberg, "Advanced organic chemistry", Plenum Press, New York, 1990.
14. C. Atkins, A. Carey, "Organic Chemistry", Mc Graw - Hill, New York, 1990.
15. R. Bro, "Multi-way analysis in the food industry", Ph. D. Thesis, University of Amsterdam, Amsterdam, 1998.
16. T. Hirschfeld, "The hy-phen-ated methods", *Anal. Chem.*, 52, 297A, 1980.
17. T. Hirschfeld, "Instrumentation in the next decade", *Science*, 230, 286, 1985.
18. C. Demir, P. Hindmarch, RG Brereton, "Deconvolution of a three component co-eluting peak cluster in gas chromatography-mass spectrometry", *Analyst*, 125, 287, 2000.
19. C. Wilkins, "Linked gas chromatography-infrared-mass spectrometry", *Anal. Chem.*, 59, 571A, 1987.
20. M. Kronenberg, "Three-mode principal component analysis: theory and applications", DSWO Press, Leiden, 1983.
21. B. Kruskal, "Multilinear methods", *Proc. Symp. Appl. Math.*, 28, 75, 1983.
22. A. Harshman, E. Lundy, "Data pre-processing and the extended PARAFAC model, Inc: Research methods for multimode data analysis", Praeger, New York, 1984.
23. M. Otto, "Chemometrics: Statistics and computer application in analytical chemistry", John Wiley & Son, New York, 1999.
24. R. Bro, "PARAFAC. Tutorial and applications", *Chem. Intell. Lab. Syst.*, 38, 149, 1997.
25. S. Wold, K. Esbensen, P. Geladi, "Principal components analysis", *Chem. Intell. Lab. Syst.*, 2, 37, 1987.
26. S. Burdick, "An introduction to tensor products with applications to multiway data analysis", *Chem. Intell. Lab. Syst.*, 28, 229, 1995.
27. M. Adams, "Chemometrics in analytical spectroscopy", The Royal Society of Chemistry, Cambridge, 1995.
28. P. Geladi, B. Kowalski, "PLS tutorial", *Anal. Chim. Acta.*, 185, 1, 1986.

29. T. Næs, H. Martens, "Principal component regression in NIR analysis", *J. Chemom.*, 2, 155, 1988.
30. E. Sanchez, B. Kowalski, "Generalized Rank Annihilation factor analysis", *Anal. Chem.*, 58, 496, 1986.
31. E. Sanchez, B. Kowalski, "Tensorial calibration: II. Second Order calibration", *J. Chemom.*, 2, 265, 1988.
32. E. Sanchez, B. Kowalski, "Tensorial Resolution: a direct trilinear decomposition", *J. Chemom.*, 4, 29, 1996.
33. D. Carroll, J. Chang, "Analysis of individual differences in multidimensional scaling via an N-way generalization of Eckart - Young decomposition", *Psychometrika*, 35, 283, 1970.
34. A. Harshman, Foundation of the PARAFAC procedure: Model and conditions for an explanatory multi-mode factor analysis", UCLA, *Working papers in phonetics*, 16, 1 1970.
35. L. Kiers, T. Berge, R. Bro, "PARAFAC2 - Part I. A direct fitting algorithm for the PARAFAC2 model", *J. Chemom.*, 13, 275, 1999.
36. R. Bro, A. Andersson, L. Kiers, "PARAFAC2 - Part II. Modeling chromatographic data with retention time shifts", *J. Chemom.*, 13, 295, 1999.
37. T. Berge, L. Kiers, "Some uniqueness results for PARAFAC2", *Psychometrika*, 61, 123, 1996.
38. S. Wold, P. Geladi, K. Esbensen, J. Ohman, "Multiway-principal components and PLS - analysis", *J. Chemom.*, 1, 41, 1987.
39. V. der Kloot, M. Kronenberg, "External analysis with three - mode principal component models", *Psychometrika*, 50, 479, 1985.
40. R. Bro, "Multiway calibration multilinear PLS", *J. Chemom.*, 10, 47, 1996.
41. Yu-Long Xie, J. Baeza-Baeza, G. Ramis-Ramos, "Kinetic spectrophotometric resolution of binary mixtures using three - way partial least squares", *Chem. Intell. Lab. Syst.*, 27, 211, 1995.

42. R. Tauler, A. Smilde, "Selectivity, local rank, three - way data analysis and ambiguity in multivariate curve resolution", *J. Chemom.*, 9, 31, 1995.
43. M. Martens, "Determining sensory quality of vegetables: a multivariate study", Ph. D. thesis, Agriculture University of Norway, Norway, 1986.
44. A. Skoog, J. Leary, "Principles of instrumental analysis", Harcourt Brau College Publishers, New York, 1992.
45. R. Bro, S. de Jong, "A fast non negativity constrained linear least squares algorithm for use in multi - way algorithms", *J. Chemom.*, 11, 393, 1997.
46. R. Bro, N. Sidiropoulos, "Least squares algorithms under unimodality and non negativity constraints", *J. Chemom.*, 12, 223, 1998.
47. N. Cliff, "Orthogonal rotation to congruence", *Psychometrika*, 31, 33, 1966.
48. S. Leurgans, R. T. Ross, R. B. Abel, "A decomposition for three-way arrays", *SIAM J. Matrix and Appl.*, 14, 1064, 1993.
49. B. Cattell, "Parallel proportional profiles and other principles for determining the choice of factors by rotation", *Psychometrika*, 9, 267, 1944.
50. H. Martens, T. Næs, "Multivariate calibration", John Wiley & Sons, Chichester, 1989.
51. S. de Jong, A. Phatak, "Partial least squares regression in: Recent advances in total least squares techniques and errors in variables modeling", Ed. Van Huffel, Philadelphia, 1997.
52. L. Ståhle, "Aspects of analysis of three- way data", *Chem. Intell. Lab. Syst.*, 7, 95, 1989.
53. "HP 8452A diode-array spectrophotometer Handbook, Hewlett Packard", Germany, 1990.
54. A. V. Rossi, E. Ferreira, 20^a Reunião Anual da sociedade Brasileira de Química, Livro de Resumos, Poços de Caldas, p. ED047.
55. D. Perriss, B. Dempsey, "Buffers for pH and metal ion control", Chapman and Hall, London, 1984.
56. J. de Leeuw, W. Young, Y. Takane, "Additive structure in qualitative data: an alternating least squares method with optimal scaling features", *Psychometrika*, 41, 471, 1976.
57. F. Yates, "The analysis of replicated experiments when the field results are incomplete", *The empire Journal of experimental agriculture*, 1, 129, 1933.

58. R. Sands, W. Young, "Component models for three-way data: an alternating least squares algorithm with optimal scaling features", *Psychometrika*, 45, 39, 1980.
59. D. Burdick, X. Tu, L. Mc Grown, D. Millican, "Resolution of multicomponent fluorescent mixtures by analysis of the excitation-emission-frequency array", *J. Chemom.*, 4, 15, 1990.
60. S. Li, J. Hamilton, P. Gemperline, "Generalized rank annihilation using similarity transformations", *Anal. Chem.*, 64, 599, 1992.
61. L. Kiers, "An alternating, least squares algorithm for PARAFAC2 and Three-way", *DEDICOM Comp. Stat. Data Anal.*, 16, 103, 1993.
62. L. Kiers, "A three- step algorithm for CANDECOMP/PARAFAC analysis of large data sets with multicollinearity", *J. Chemom.*, 12, 155, 1998.
63. R. Caulcut, R. Boddy, "Statistics for analytical chemists", Chapman and Hall, London, 1989.