

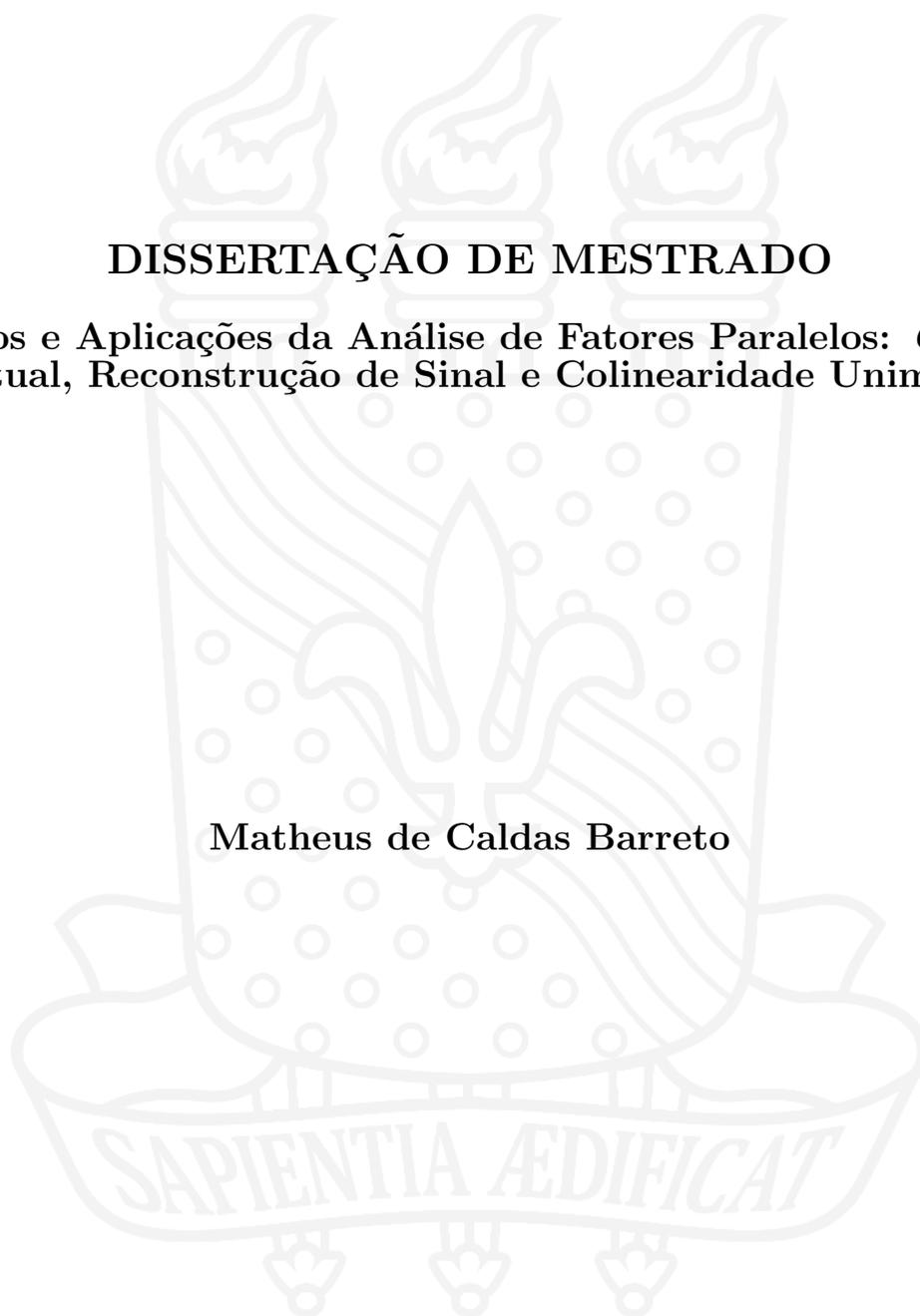


UNIVERSIDADE FEDERAL DA PARAÍBA  
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA  
DEPARTAMENTO DE QUÍMICA  
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA

## DISSERTAÇÃO DE MESTRADO

Aspectos e Aplicações da Análise de Fatores Paralelos: *Clean-up*  
Virtual, Reconstrução de Sinal e Colinearidade Unimodal

Matheus de Caldas Barreto



João Pessoa - PB - Brasil  
Julho/2021



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

## DISSERTAÇÃO DE MESTRADO

Aspectos e Aplicações da Análise de Fatores Paralelos: *Clean-up*  
Virtual, Reconstrução de Sinal e Colinearidade Unimodal

Matheus de Caldas Barreto\*

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

Orientador(a): Wallace Duarte Fragoso

\*Bolsista CNPq

João Pessoa - PB - Brasil  
Maio/2020

**Catálogo na publicação**  
**Seção de Catalogação e Classificação**

B273a Barreto, Matheus de Caldas.  
Aspectos e aplicações da análise de fatores paralelos :  
clean-up virtual, reconstrução de sinal e colinearidade unimodal /  
Matheus de Caldas Barreto. - João Pessoa, 2021.  
106 f. : il.

Orientação: Wallace Duarte Fragoso.  
Dissertação (Mestrado) - UFPB/CCEN.

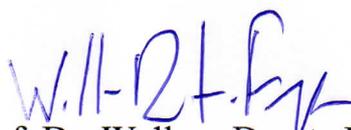
1. Química. 2. PARAFAC. 3. Fluorescência sincrônica.  
4. Métodos cinéticos. 5. Unicidade parcial. I. Fragoso, Wallace  
Duarte. II. Título.

UFPB/BC

CDU 54(043)

# **Aspectos e Aplicações de Análise de Fatores Paralelos: Clean-up Virtual, Reconstrução de Sinal e Colinearidade Unimodal.**

Dissertação de Mestrado apresentada pelo aluno Matheus de Caldas Barreto e aprovada pela banca examinadora em 30 de setembro de 2021.



Prof. Dr. Wallace Duarte Fragoso  
Departamento de Química - UFPB  
Orientador/Presidente



Profa. Dra. Rocio Pellegrino Vidal  
Universidad Nacional de La Plata - Argentina  
Examinador



Prof. Dr. Wagner de Mendonça Faustino  
Departamento de Química - UFPB  
Examinador

# Sumário

Lista de Figuras	i
Lista de Tabelas	iv
Lista de Abreviações	v
Notação Matemática	vi
Resumo	vii
Abstract	viii
<b>1 Introdução</b>	<b>1</b>
<b>2 Fundamentação Teórica</b>	<b>3</b>
2.1 Ordem Superior . . . . .	3
2.2 Análise de fatores paralelos (PARAFAC) . . . . .	4
2.3 Valores de mérito . . . . .	7
2.4 Espectroscopia de fluorescência e EEMs . . . . .	8
2.5 Métodos cinéticos de resolução de dados . . . . .	9
<b>3 Objetivos</b>	<b>11</b>
3.1 Objetivo geral . . . . .	11
3.2 Objetivos específicos . . . . .	11
<b>4 Fluorescência sincrônica e PARAFAC</b>	<b>12</b>
4.1 Conversão entre SFMs e EEMs . . . . .	15
4.2 Espaços vazios . . . . .	18
4.3 Aplicações simuladas . . . . .	19
4.4 Mistura de HPAs . . . . .	23
4.5 Conclusão . . . . .	27
<b>5 PARAFAC como <i>clean-up</i> virtual</b>	<b>28</b>
5.1 <i>Clean-up</i> virtual . . . . .	28
5.2 Aplicação - Determinação de melamina em leite . . . . .	28
5.2.1 Metodologia . . . . .	31
5.2.2 Resultados e discussão . . . . .	32
5.2.3 Conclusão . . . . .	37
<b>6 Sistemas espectrocinéticos</b>	<b>38</b>
6.1 Aplicação - Corantes alimentícios . . . . .	38
6.1.1 Reações de Fenton e Fenton-like . . . . .	38
6.1.2 Analitos utilizados . . . . .	39
6.1.3 Metodologia . . . . .	39
6.1.4 Resultados e discussão . . . . .	42

6.1.5	Aplicação em amostras reais . . . . .	45
6.2	Aplicação - Sistema dopamina/aminocroma . . . . .	46
6.2.1	Conversão de dopamina em aminocroma . . . . .	46
6.2.2	Metodologia . . . . .	47
6.2.3	Resultados e discussão . . . . .	49
6.3	Estudos simulados . . . . .	53
6.3.1	Reações do tipo $A \longrightarrow B$ . . . . .	53
6.3.2	Efeito de desvios da cinética de 1 <sup>a</sup> ordem . . . . .	54
6.4	Conclusão . . . . .	59
<b>7</b>	<b>Colinearidade unimodal no PARAFAC</b>	<b>61</b>
7.1	Causas e consequências da unicidade parcial causada por colinearidade unimodal . . . . .	61
7.2	Efeito da presença de interferentes colineares . . . . .	65
7.3	Estratégias para quebrar a colinearidade . . . . .	66
7.4	Conclusão . . . . .	73
<b>8</b>	<b>Conclusão</b>	<b>75</b>
<b>9</b>	<b>Perspectivas futuras</b>	<b>76</b>
	<b>Referências</b>	<b>83</b>
	<b>Apêndice A - Script para conversão de SFMs em EEMs</b>	<b>84</b>
	<b>Apêndice B - Simulador de dados espectrocinéticos</b>	<b>86</b>
B.1	Construção do simulador . . . . .	86
B.2	Simulador de dados espectrais . . . . .	88
B.3	Simulador de dados cinéticos . . . . .	89
B.4	Exemplo de geração de dados espectrocinéticos . . . . .	90
	<b>Apêndice C - Resultados da determinação de melamina</b>	<b>92</b>

## Lista de Figuras

1	Representação gráfica de amostras e conjuntos de diferentes ordens.	3
2	Representação gráfica da decomposição da matriz de dados segundo o PARAFAC. . . . .	5
3	(A) Exemplo de uma matriz de excitação-emissão; (B) Representação gráfica de uma série de EEMs organizadas em um tensor tridimensional com características trilineares. . . . .	9
4	Exemplo de amostra medida por espectrofluorimetria usando (A) SFMs, (B) EEMs. . . . .	12
5	(A) Região espectral de interesse, (B) Região que pode ser medida em uma EEM com remoção de espalhamentos Rayleigh, (C) Região que pode ser medida em uma EEM evitando os espalhamentos Rayleigh, (D) Região que pode ser medida com uma SFM evitando os espalhamentos Rayleigh e convertendo-a em uma EEM. . . . .	13
6	Exemplo de conversão de uma SFM para uma EEM usando uma transformação de cisalhamento. . . . .	16
7	Matriz EEM convertida de uma SFM de tamanho $10 \times 10$ com diferentes valores de $m_{min}$ e $n_{min}$ . . . . .	18
8	Componentes criados para testar o uso do PARAFAC em SFMs convertidas. . . . .	20
9	Componentes puros e componentes recuperados para os 5 casos. . . . .	22
10	SFM convertida para EEM reconstruída usando diferentes valores para preencher os espaços vazios. (A) Matriz original mostrando as regiões de espaços vazios, (B) Matriz reconstruída usando NaN em ambas as regiões, (C) Matriz reconstruída usando zeros em ambas as regiões, (D) Matriz reconstruída usando zeros na região próxima ao espalhamento Rayleigh e NaN na região próxima a difração de 2ª ordem do espalhamento Rayleigh. . . . .	24
11	Componentes puros e recuperados para o fluoranteno, antraceno, pireno, fenantreno e fluoreno. . . . .	25
12	Curvas de autopredição para o fluoranteno, antraceno, pireno, fenantreno e fluoreno feitas a partir dos valores no modo A do PARAFAC e as concentrações de referência. . . . .	26
13	Estrutura molecular da melamina. . . . .	29
14	EEM de uma amostra de leite pré-tratado não adulterada (a), padrão aquoso de melamina a 1,5 ppm (b) Melamina a 1,5 ppm em leite pré-tratado (c). . . . .	33
15	Curvas de calibração pseudo-univariada para melamina em quatro diluições diferentes. . . . .	34
16	Perfis de emissão e excitação de uma amostra de leite decomposta pelo PARAFAC. Linha azul sólida: melamina; linha tracejada vermelha: interferentes da amostra. . . . .	35
17	Valores de predição versus referência da concentração de melamina para diferentes tipos de leite obtidos utilizando PARAFAC. . . . .	36

18	Espectros obtidos da degradação de corantes alimentícios com o uso de absorção UV/Vis e reagentes de Fenton. . . . .	42
19	Resultado da deconvolução em 4 fatores para a degradação de AT e AC por reagentes de Fenton. . . . .	43
20	Resultado da deconvolução em 3 fatores para a degradação de AT e AC por reagentes de Fenton. . . . .	44
21	Resultado da calibração linear realizada a partir da deconvolução dos escores da matriz <b>A</b> em 4 fatores para o AT e AC. . . . .	44
22	Esquematisação das reações envolvidas na conversão de dopamina em aminocroma. . . . .	47
23	Esquematisação do processo de preparação de amostra para medição da matriz espectrocínética da conversão de DQ em AMC. . . . .	48
24	Espectros obtidos para a conversão de dopaminoquinonas em aminocroma, tendo levodopa adicionada como interferente, com o uso de absorção UV/Vis e hipoclorito de sódio. . . . .	50
25	Perfis espectrais e cinéticos deconvoluidos usando PARAFAC para a conversão de dopaminoquinonas em aminocroma. (A) Perfis espectrais usando dois componentes; (B) Perfis cinéticos usando dois componentes; (C) Perfis espectrais usando três componentes; (D) Perfis cinéticos usando três componentes. . . . .	51
26	Curva de calibração para determinação de dopaminoquinona a partir dos escores do modo A do PARAFAC usando: (A) 2 componentes; (B) 3 componentes. . . . .	52
27	Três simulações de reações do tipo $A \rightarrow B$ com diferentes níveis de sobreposição espectral. Cada coluna contem os três modos deconvoluidos pelo PARAFAC para cada simulação. . . . .	54
28	Resultado de duas simulações para dois componentes em um sistema no qual o ambos reagem com um terceiro de concentração inicial de 50 e 5 u.a. . . . .	56
29	Perfis recuperados para dois componentes afetados por um reagente limitante com tempos de simulação de 50 e 10 minutos. . . . .	56
30	Resultado de duas simulações para dois componentes em um sistema no qual o 1º reage com um 3º e o 2º é independente. As simulações foram feitas com concentrações iniciais do 3º de 8 e 1 u.a. . . . .	57
31	Perfis recuperados para o componente 1 afetado por um reagente limitante com tempos de simulação de 50 e 10 minutos. . . . .	58
32	Perfis recuperados para um analito seguindo uma cinética de 1ª ordem e um interferente seguindo uma de 2ª ordem em dois tempos de simulação distintos. . . . .	59
33	Perfis dos 3 modos para os dois componentes estudados. . . . .	62
34	Exemplo da representação dos pontos de um espectro de referência com o predito. . . . .	63
35	Comparação dos valores deconvoluidos com os valores de referência para os 3 modos em ambos os componentes tendo o modo A altamente correlacionado. . . . .	63

36	Comparação dos valores deconvoluidos com os valores de referência para os 3 modos em ambos os componentes tendo o modo B altamente correlacionado. . . . .	64
37	Comparação dos valores deconvoluidos com os valores de referência para os 3 modos em ambos os componentes tendo o modo C altamente correlacionado. . . . .	65
38	Comparação dos valores deconvoluidos para um sistema com dois componentes colineares no modo A e um terceiro linearmente independente nos três modos. . . . .	66
39	Representação da adição de uma matriz (em laranja) ao tensor de amostras original (em azul) contendo uma das amostras com adição de padrão para quebrar a colinearidade no modo A. . . . .	67
40	Comparação dos valores deconvoluidos com os valores de referência para os 3 modos em ambos os componentes tendo o modo A do tensor original altamente correlacionado, mas com uma amostra adicional que passou por adição de padrão. . . . .	68
41	Diferentes estratégias de implementação do deslocamento temporal ao longo do tensor. . . . .	69
42	Perfis espectrais recuperados para ambos os componentes da amostra.	70
43	Comparação dos valores deconvoluidos com os valores de referência para os 3 modos em ambos os componentes tendo o modo A do tensor original altamente correlacionado devido ao componente 2 ser produto do consumo do componente 1. . . . .	71
44	Comparação dos valores deconvoluidos com os valores de referência para os 3 modos em ambos os componentes tendo o modo A do tensor original altamente correlacionado, mas com uma amostra adicional sendo deslocada no tempo. . . . .	72
45	Resíduos do modelo PARAFAC para diferentes amostras do conjunto principal e para a amostra adicionada (deslocada no tempo). . . . .	72
46	Resultados da deconvolução com a adição de uma única amostra copiada com diferentes deslocamentos temporais. . . . .	73
47	Efeito dos parâmetros nas curvas gaussianas geradas. . . . .	86
48	Exemplo de espectro gerado pela soma de múltiplas gaussianas. . . . .	87
49	Fluxograma do processo de geração de dados simulados. . . . .	88
50	Exemplo de matriz de segunda ordem de absorção $\times$ tempo para uma reação do tipo $A \longrightarrow B$ . . . . .	88

## Lista de Tabelas

1	Níveis de concentração do conjunto de teste com os 7 HPAs. . . . .	23
2	Valores de mérito das curvas de autopredição do fluoranteno, antraceno, pireno, fenantreno e fluoreno. . . . .	26
3	Valores de mérito para a quantificação de melamina usando PARAFAC. . . . .	36
4	Concentração dos reagentes de Fenton utilizados para degradar os corantes alimentícios. . . . .	40
5	Conjunto de calibração para o AT e AC. . . . .	40
6	Conjunto de validação para o AT e AC. . . . .	41
7	Resultados da calibração pseudo-univariada usando deconvolução em 4 fatores para o AT e AC. . . . .	45
8	RMSEV e REP calculados para as concentrações preditas de AT e AC. . . . .	45
9	Conjunto de calibração para quantificação de dopamina. . . . .	48
10	Conjunto de validação do modelo para quantificação de dopamina com levodopa como interferente. . . . .	49
11	RMSEV e REP da quantificação de dopaminoquinonas e $s_{fit}$ e <i>CORCONDIA</i> calculados para o modelo PARAFAC usando dois e três componentes. . . . .	52

## Lista de Abreviações

**AC** Amarelo crepúsculo  
**ALS** Alternating least squares  
**AMC** Aminocroma  
**AT** Amarelo tartrazina  
**CAS** Chemical Abstracts Service  
**CORCONDIA** Core consistency diagnostic  
**DA** Dopamina  
**DQ** Dopaminoquinona  
**DTLD** Direct trilinear deconvolution  
**EEM** Excitation-emission matrix  
**HPLC-DAD** High performance liquid chromatography with diode array detection  
**LC** Leukoaminocroma  
**LoB** Limit of blank  
**LoD** Limit of detection  
**LoQ** Limit of quantification  
**MCR** Multivariate curve resolution  
**NaN** Not a number  
**PARAFAC** Parallel factor analysis  
**RAFA** Rank annihilation factor analysis  
**RMSE** Root mean square error  
**SFM** Synchronous fluorescence matrix  
**UHT** Ultra high temperature  
**U-PLS/RBL** Unfolded partial least squares with residual bilinearization

## Notação Matemática

Nesse trabalho serão usadas as seguintes notações:

- Elementos de objetos serão representados com a letra do objeto minúscula, em itálico e com a posição subscrita (Ex: Elemento  $a_{ij}$ , localizado na linha  $i$  e coluna  $j$  da matriz  $\mathbf{A}$ ).
- Matrizes serão representados por letras maiúsculas em negrito e itálico (Ex: Matriz  $\mathbf{A}$ ).
- Tensores de três ou mais dimensões serão representados por letras maiúsculas em negrito, itálico e sublinhadas (Ex: Tensor  $\underline{\mathbf{X}}$ ).

## Resumo

A análise de fatores paralelos (PARAFAC) é um algoritmo trilinear para a análise de dados de ordem superior que conta com a vantagem de segunda-ordem. Nesse trabalho, quatro estudos distintos foram feitos para explorar diferentes faces da aplicação do PARAFAC. O primeiro avalia o uso de matrizes de fluorescência sincrônica convertidas em matrizes de excitação-emissão, que são trilineares. O método proposto foi aplicado a uma mistura de HPAs e o PARAFAC pôde recuperar os perfis de componentes puros, e devido a trilinearidade, reconstruir o sinal na região de espaços vazios. O segundo estudo explora a vantagem de segunda-ordem para realizar um clean-up virtual de uma matriz complexa, ao utilizar um tensor contendo uma amostra por vez e um conjunto de calibração de padrões do analito puro. Essa técnica foi aplicada com sucesso para a determinação de melamina em leite adulterado em um método seguindo preceitos da química verde. O terceiro estudo avaliou a cinética de reação junto com a espectroscopia de absorção no UV-Vis como matriz de dados de segunda ordem. A degradação de corantes alimentícios por reação de fenton e a oxidação da dopamina em presença de levodopa como interferente foram os sistemas modelo estudados. Sistemas simulados também foram avaliados. Foi verificado que processos cinéticos que resultem em componentes correlacionados podem dificultar a deconvolução por PARAFAC, gerando problemas de convergência e perda da unicidade da solução. O quarto estudo usou de simulações para explorar o caso nos quais um dos modos é colinear para dois fatores. As simulações mostraram que o modo colinear dos componentes afetados pode ser recuperado consistentemente, mas os demais podem apresentar distorções consideráveis. Algumas estratégias foram desenvolvidas para quebrar a colinearidade no modo de concentração, comum em matrizes espectrocinéticas, e uma delas, o uso de matrizes adicionais deslocadas no tempo, foi testada com oxidação da dopamina e mostrou boa recuperação da consistência trilinear do modelo, o que sinaliza a correção das distorções nos modos espectrais e cinéticos. Essas aplicações demonstram que um entendimento mais aprofundado do processo de modelagem do PARAFAC pode ser usado para o desenvolvimento de estratégias analíticas inovadoras.

Palavras-chave: *PARAFAC, fluorescência sincrônica, métodos cinéticos, unicidade parcial.*

## Abstract

Parallel factor analysis (PARAFAC) is a trilinear algorithm for higher-order data analysis that relies on the second-order advantage. In this dissertation, four different studies were made to explore different aspects of PARAFAC application. The first evaluates the use of synchronic fluorescence matrices converted into excitation-emission matrices, which are trilinear. The proposed method was applied to a mixture of HPAs and PARAFAC was able to retrieve the profiles of pure components, and due to the trilinearity of the system, rebuild the signal in the regions containing empty spaces. The second study explores the second-order advantage by performing a virtual clean-up of a complex matrix by using a tensor containing one sample at a time and a calibration set consisting of analyte standards. This technique was successfully applied for the determination of melamine in adulterated milk in a method consistent with green chemistry precepts. The third study evaluated the use of reaction kinetics alongside UV-Vis absorption spectroscopy as source of second-order data. The degradation of food dyes by fenton reaction and dopamine oxidation in the presence of levodopa as an interferent were the model systems studied. Simulated systems were also evaluated. It was verified that kinetic processes that result in correlated components can hinder deconvolution by PARAFAC, generating problems of convergence and loss of the uniqueness of the solution. The fourth study used simulations to explore cases in which one of the modes is colinear for two factors. The simulations showed that the colinear mode of the affected components can be recovered consistently, but the other may present considerable distortions. Some strategies were developed to break the colinearity in the concentration mode, which common in spectrokinetic matrices, and one of them, the use of additional matrices displaced in time, was tested with oxidation of the dopamine and showed good recovery of the trilinear consistency of the model, which signals the correction of distortions in spectral and kinetic modes. These applications demonstrate that a deeper understanding of the PARAFAC modeling process can be used to develop innovative analytical strategies.

Keywords: *PARAFAC, synchronous fluorescence, kinetic methods, partial uniqueness.*

## 1 Introdução

Com a maior popularização de soluções comerciais para técnicas experimentais de segunda ordem a partir do final dos anos 90, matrizes de dados cada vez mais complexas foram se tornando mais comuns, popularizando técnicas de calibração e análise multivariada. Entre as motivações para o desenvolvimento dessas técnicas estão as maneiras que modelos usando dados de maior dimensionalidade são capazes de lidar com interferentes na matriz. Essas são chamadas de vantagens de 1ª e 2ª ordem. Modelos multivariados de 1ª ordem possuem a vantagem de 1ª ordem permitindo ao sistema identificar vieses de predição causados por interferentes. Métodos de análise de dados de ordem superior, por outro lado, se destacam devida a chamada vantagem de 2ª ordem, permitindo a modelagem desses interferentes sem necessidade de incluí-los no conjunto de calibração, garantindo maior robustez a presença de interferentes desconhecidos na matriz [1].

Um dos algoritmos que ganhou destaque para o análise de dados de ordem superior foi Análise de Fatores Paralelos (PARAFAC), podendo ser usado para recuperar informação tanto qualitativa quanto quantitativa do sistema [2].

Originalmente sendo utilizado para aplicações psicométricas, o PARAFAC começou a ser aplicado no contexto de química analítica a partir da segunda metade dos anos 90, com a tese de doutorado de Rasmus Bro e o tutorial publicado pelo mesmo em 1997 sendo considerados os marcos iniciais de maior importância para esse tipo de aplicação [2]. Desde então, o PARAFAC tem sido usado nas mais diversas aplicações, desde análise de sistemas ambientais [3, 4, 5, 6], alimentares [7, 8, 9, 10, 11], aplicações em biomedicina [12, 13, 14, 15], entre muitos outros.

Um dos motivos dessa ampla gama de aplicações se deve as propriedades intrínsecas ao método, como a unicidade de sua solução dado uma boa consistência trilinear dos dados e a capacidade de lidar com matrizes complexas, devido a vantagem de segunda ordem. Outra vantagem considerável é a recuperação de informações qualitativas sobre o sistema [2] que poderiam se perder com o uso de algoritmos como mínimos quadrados parciais desdobrados com bilinearização do resíduo (U-PLS/RBL).

Diversos tipos de técnicas conseguem gerar dados de ordem superior, um dos mais comuns sendo a espectrofluorimetria de excitação-emissão que é intrinsecamente trilinear e cromatografia com detecção espectral [16], que assim como outras técnicas hífenadas é capaz de gerar esses dados. A utilização de reações químicas

como fonte de variação em um sistema também pode ser aliada a um método de primeira ordem para gerar dados espectrocinéticos de segunda ordem.

Cada sistema capaz de gerar esses dados tem suas próprias peculiaridades, que devem ser levadas em consideração durante a análise, podendo ser exploradas para um entendimento mais aprofundado dos sistemas.

Por exemplo, certas implementações do PARAFAC podem aceitar tensores com valores vazios [17], permitindo a análise de sistemas que seriam intratáveis de outra forma. Sistemas cinéticos em certas condições podem gerar componentes com alta colinearidade (como um reagente e um produto), carregando suas próprias implicações.

Neste trabalho nos dedicamos a melhorar a compreensão de como o PARAFAC se comporta durante a modelagem em diversas situações limites, como forçar o número de componentes para dois, utilizar de implementações mais recentes para tratar dados com espaços vazios e modelar sistemas com alta colinearidade. Desenvolvendo essa compreensão com o auxílio de dados simulados, e aplicando-a a sistemas modelo experimentais, buscamos desenvolver novas metodologias analíticas e novas abordagens para problemas clássicos da modelagem com o PARAFAC.

## 2 Fundamentação Teórica

### 2.1 Ordem Superior

Atualmente, parte considerável dos métodos utilizados na quimiometria envolvem calibrações multivariadas. Os dados obtidos das técnicas experimentais podem ser organizados em objetos, sendo a ordem do objeto o número de dimensões que o arranjo de dados possui.

Dados univariados, como a absorbância em um comprimento de onda específico, são dados de ordem zero, sendo números escalares que representam a medida ou sinal, podendo ser organizados em um vetor de amostras. Um espectro de absorção com múltiplos comprimentos de onda seria um objeto de primeira-ordem, equivalente a um vetor por amostra que podem ser organizados em uma matriz. Uma matriz de fluorescência de excitação-emissão (EEM), com diversos comprimentos de onda de excitação e emissão em vias perpendiculares é considerado um objeto de segunda ordem, podendo ser organizado em um tensor tridimensional [1, 18]. Objetos com duas ou mais vias (*multiway*) são considerados dados de ordem superior. Uma esquematização desses diferentes tipos de dados pode ser visto na Fig. 1.

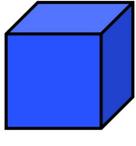
	Ordem 0	1ª ordem	2ª ordem
Amostra			
Conjunto de dados			

Figura 1: Representação gráfica de amostras e conjuntos de diferentes ordens.

Com o aumento da dimensionalidade do objeto contendo os dados da amostra, diferentes vantagens começam a se tornar aparentes. Modelos de ordem zero são sensíveis a presença de interferentes e, na prática, em química analítica moderna, só são úteis quando associados a métodos de separação, como a cromatografia. Modelos de primeira ordem, por exemplo, permitem detectar distorções causadas por interferentes, que causam um viés de predição. Essa é a chamada vantagem de

primeira-ordem. Modelos de segunda ordem, por outro lado, são robustos a presença de interferentes desconhecidos, permitindo que algoritmos desse tipo possam ser aplicados a matrizes complexas mesmo que apenas os analitos de interesse estejam no conjunto de calibração [18].

Essa vantagem permite a aplicação de diferentes estratégias para analisar as amostras e os analitos de interesse. Por exemplo, usando uma tensor de amostras complexas, desde que exista suficiente variabilidade entre as amostras, é possível fazer a caracterização de cada componente individualmente. No caso de analitos conhecidos, também é possível modelar uma amostra por vez junto com um conjunto de calibração puro para isolar os interferentes em um componente adicional ao reduzir a variabilidade dos interferentes.

## 2.2 Análise de fatores paralelos (PARAFAC)

A análise de fatores paralelos (PARAFAC) é um método numérico de decomposição de tensores de natureza trilinear em matrizes contendo os coeficientes de cada uma das vias separadas em fatores paralelos [2], como mostrado na Eq. (1):

$$x_{ijk} = \sum_{f=1}^F a_{if} b_{jf} c_{kf} + e_{ijk} \quad (1)$$

Em suma, o PARAFAC decompõe o tensor original  $\underline{\mathbf{X}}$  nas matrizes  $\mathbf{A}$ , contendo os coeficientes de concentração do analito  $f$  da amostra  $i$ ,  $\mathbf{B}$  e  $\mathbf{C}$ , contendo os coeficientes das duas vias para cada analito  $f$ . No caso de um tensor de dados trilinear formado pelo empilhamento de matrizes de excitação-emissão (EEMs) de diferentes amostras, temos que a matriz  $\mathbf{A}$  conteria o coeficiente de concentrações para cada analito, a matriz  $\mathbf{B}$  conteria os coeficientes de excitação para cada comprimento de onda de cada analito, e a matriz  $\mathbf{C}$  os coeficientes de emissão para cada comprimento de onda de cada analito. O tensor  $\underline{\mathbf{E}}$  terá os resíduos do modelo.

Uma representação gráfica do que o PARAFAC faz com o tensor tridimensional  $\underline{\mathbf{X}}$  pode ser vista na Fig. 2:

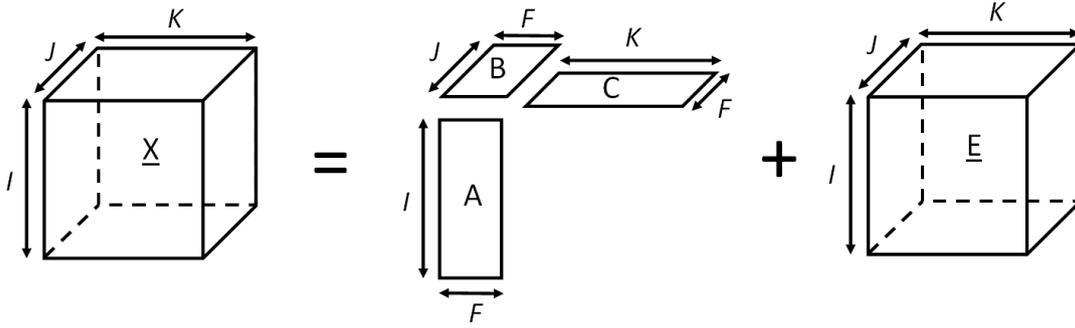


Figura 2: Representação gráfica da decomposição da matriz de dados segundo o PARAFAC.

Enquanto as matrizes  $\mathbf{B}$  e  $\mathbf{C}$  podem ser usados para a caracterização dos componentes presentes no tensor  $\underline{X}$ , a matriz  $\mathbf{A}$  pode ser usada para a calibração de uma curva pseudo-univariada. Para isso, é necessário que ambos os conjuntos de calibração e a amostra (ou conjunto de amostras) sejam incluídos no tensor e modelados juntos.

Um tensor 3D pode ser considerado trilinear se puder ser ajustado a partir da Eq. (1). A existência de trilinearidade requer que os dois perfis instrumentais sejam independentes uns dos outros e que cada componente tenha um único perfil comum a todas as amostras. Além disso, cada matriz de dados compondo o tensor deve ser bilinear.

Para resolver numericamente a decomposição, o algoritmo ALS (*Alternating Least Squares*) é utilizado [2], consistindo em calcular uma das matrizes ( $\mathbf{A}$ ,  $\mathbf{B}$  ou  $\mathbf{C}$ ) a partir de duas outras de maneira a alternar as matrizes a cada passo (estimando  $\mathbf{A}$  a partir de  $\mathbf{B}$  e  $\mathbf{C}$ ,  $\mathbf{B}$  a partir de  $\mathbf{A}$  e  $\mathbf{C}$  e assim em diante), como indicado na Eq. (2). As estimativas iniciais para as matrizes deconvoluídas podem ser calculadas por diversos métodos, como a decomposição em valores singulares (SVD), deconvolução trilinear direta (DTLD), entre outros [19].

$$\mathbf{A} = \mathbf{X}\mathbf{Z}^t(\mathbf{Z}\mathbf{Z}^t)^{-1}, \mathbf{Z} = (\mathbf{C}|\otimes|\mathbf{B})^t \quad (2)$$

com  $|\otimes|$  sendo o produto de Khatri-Rao [20].

Uma das vantagens do uso do PARAFAC é que ele possui resultado único, possibilitando que o algoritmo possa recuperar o perfil das variáveis dos componentes puros do sistema, desde que o sistema apresente boa consistência trilinear [1].

O PARAFAC também pode ser visto como uma versão restrita do *Tucker 3*, que pode ser entendido como uma análise de componentes principais de 3 vias [2]. Por exemplo, o PARAFAC pode modelar um tensor tridimensional através da Eq. (3):

$$\underline{\mathbf{X}} = \mathbf{A}(\mathbf{C}|\otimes|\mathbf{B})^t + \underline{\mathbf{E}} \quad (3)$$

O *Tucker 3* por sua vez pode representar o mesmo tensor através da Eq. (4):

$$\underline{\mathbf{X}} = \mathbf{AG}(\mathbf{C}|\otimes|\mathbf{B})^t + \underline{\mathbf{E}} \quad (4)$$

onde  $\underline{\mathbf{G}}$  é a matriz de núcleo do modelo *Tucker 3*. Caso o valor do tensor  $\underline{\mathbf{G}}$  seja uma superidentidade, ou seja, os elementos da sua superdiagonal sejam unitárias, enquanto os outros se aproximem de zero, o PARAFAC e o *Tucker 3* serão equivalentes.

O *CORCONDIA* (*CORe CONSistency DIAgnostic*) [21] se baseia nessa condição para fornecer informação sobre a consistência da trilinearidade do modelo, consistindo na aplicação do *Tucker 3* nos parâmetros do modelo PARAFAC estimado. Se espera em um arranjo de dados trilinear que o  $\underline{\mathbf{G}}$  seja uma superidentidade. Tendo isso em mente, o valor do *CORCONDIA* é definido pelo desvio quadrático do  $\underline{\mathbf{G}}$  comparado a uma superdiagonal, como mostrado na Eq. (5), com  $g_{def}$  sendo o elemento da matriz central calculado com o Tucker 3 a partir da saída do PARAFAC,  $h_{def}$  sendo os elementos da superidentidade e  $F$  o número de fatores.

$$CORCONDIA = 100 \times \left( 1 - \frac{\sum_{d=1}^F \sum_{e=1}^F \sum_{f=1}^F (g_{def} - h_{def})^2}{\sum_{d=1}^F \sum_{e=1}^F \sum_{f=1}^F h_{def}^2} \right) \quad (5)$$

Parte importante da modelagem por PARAFAC é a escolha do número de componentes, ou fatores. O *CORCONDIA* é um dos parâmetros mais comumente utilizados, junto com a análise dos valores do *fit* residual ( $s_{fit}$ , mostrado na seção 2.3). Um número excessivo de componentes pode resultar na quebra de trilinearidade do modelo, mas uma vez que esse não é o único fator que pode ser responsável pela quebra de consistência trilinear, é interessante que esse seja usado em conjunto com outros parâmetros, como a análise da mudança nos valores de  $s_{fit}$ , a variância explicada, entre outros [22, 23].

### 2.3 Valores de mérito

Ao decorrer desse trabalho, diversas métricas serão utilizados para avaliar quantitativamente os modelos criados pelo PARAFAC.

O ajuste do modelo aos dados pode ser avaliado através do parâmetro  $s_{fit}$ , que consiste da raiz quadrada da soma quadrática dos resíduos do modelo dividida pelo numero de elementos do tensor, fornecendo o resíduo médio através do tensor na mesma escala do sinal, como mostrado na Eq. (7). Tanto o  $s_{fit}$  quanto o *CORCONDIA* são ferramentas úteis para a determinação do número de componentes no modelo.

$$SSR = \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K (x_{ijk} - \hat{x}_{ijk})^2 \quad (6)$$

$$s_{fit} = \sqrt{\frac{SSR}{I \times J \times K}} \quad (7)$$

Os modelos quantitativos, obtidos através da calibração pseudo-univariada dos valores do modo A do PARAFAC, podem ser avaliados como qualquer outra calibração univariada. Os valores da raiz quadrada do erro médio quadrado (RMSE) dos valores obtidos e seu valor normalizados são especialmente úteis para avaliar os erros de calibração e previsão, como pode-se ver nas Eqs. (8) e (9).

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (8)$$

$$REP = \frac{RMSE}{y_{max} - y_{min}} \quad (9)$$

O limite de detecção (LoD) pode ser estimado de várias maneiras [24, 25], geralmente analisando o modelo ou suas previsões. O LoD é a concentração mínima de analito que o método pode detectar com confiança. Se for utilizado um modelo para cada amostra, o valor do LoD também pode ser calculado por amostra, o que em alguns casos pode não ser conveniente para uma avaliação global do método.

Em certas definições, como a citada por Olivieri e Escandar [25], o LoD de sistemas multiway pode ser definido em termos da sensibilidade de cada modelo.

A definição citada por Armbruster & Pry [24] utiliza apenas os valores previstos para um conjunto de espaços em branco e amostras de baixa concentração de analitos, levando assim a um valor de LoD generalizado para o método.

Com o LoD de um método sendo definido como o valor mínimo previsto que é estatisticamente diferente de brancos com 95% de confiança. Seguindo essa

definição, pode-se chegar nas equações (10) e (11). O limite de quantificação (LoQ) é comumente definido como três vezes o LoD, como visto em Eq. (12).

$$LoB = \mu_{brancos} + 1.645\sigma_{brancos} \quad (10)$$

$$LoD = LoB + t_{.95} s_{m.amostras} \quad (11)$$

$$LoQ = 3 \times LoD \quad (12)$$

O limite do branco (LoB) representa o maior valor das concentrações previstas para brancos, ou seja, o valor máximo esperado para um branco com 95% de confiança [24]. O LoD é calculado estimando a concentração mínima em que o valor previsto de uma amostra estaria acima do LoB com 95% de confiança, tendo em vista o desvio padrão para as amostras com menor concentração esperada nos conjuntos de validação. A distribuição *t* de *student* pode ser utilizada na Eq. (4) quando existe um menor número de réplicas na amostra de menor concentração.

## 2.4 Espectroscopia de fluorescência e EEMs

Na espectroscopia de fluorescência as moléculas absorvem fótons promovendo os elétrons a um estado de mais alta energia (fotoexcitação). Após uma série de processos não radiativos intrínsecos de cada molécula, o sistema relaxa radiativamente para o estado fundamental emitindo fótons que são detectados na forma de um padrão característico que chamamos espectro de emissão.

A variação da intensidade da luz emitida pelas amostras com a concentração é o que permite a quantificação por esse método. A intensidade de luz emitida pela amostra pode ser representada pela Eq.(13)[26].

$$I_{ij} = \sum_{f=1}^F \alpha_f \varepsilon_{fi} \delta_{fj} \quad (13)$$

onde *i* é o comprimento de onda de excitação, *j* é o comprimento de onda de emissão, *f* se refere a espécie química na solução,  $\alpha$  a concentração de cada espécie,  $\varepsilon$  ao coeficiente de absorção molar e  $\delta$  ao coeficiente de emissão relativa.

Como visto na Eq.(13), fixando o comprimento de onda de emissão e variando o de excitação, podemos gerar o que é chamado de um espectro de emissão. Caso o contrário seja feito, podemos obter um espectro de excitação. Ao empilhar esses espectros em uma matriz onde cada coluna corresponderá a um único comprimento de onda de emissão e cada linha um único comprimento de onda de excitação (ou

o contrário) é possível gerar uma matriz de dados chamada matriz de excitação-emissão (EEM), que devido a sua estrutura pode ser organizada em um tensor tridimensional intrinsecamente trilinear seguindo a Eq. (13), resultando em uma estrutura similar a mostrada na Fig. 3(B).

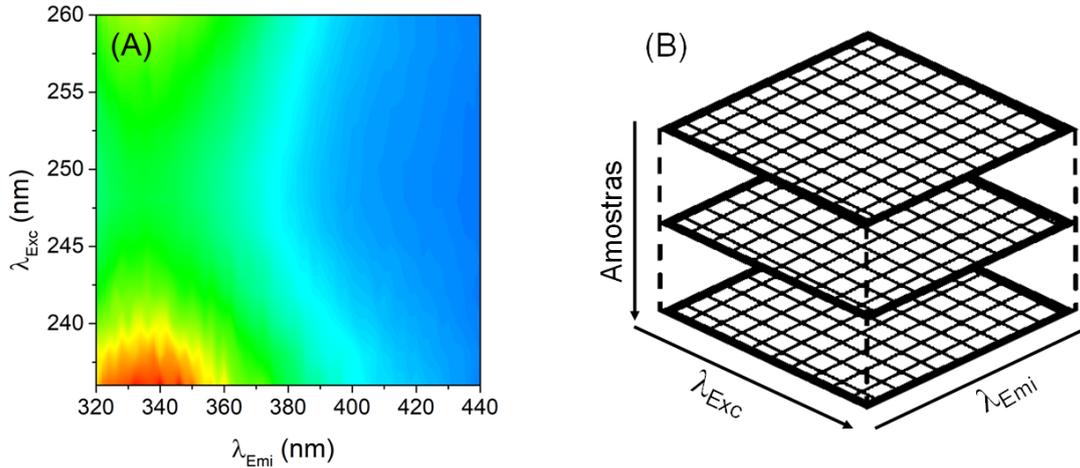


Figura 3: (A) Exemplo de uma matriz de excitação-emissão; (B) Representação gráfica de uma série de EEMs organizadas em um tensor tridimensional com características trilineares.

## 2.5 Métodos cinéticos de resolução de dados

Nesse trabalho, uma das maneiras propostas para a geração de dados de ordem superior é o uso de reações químicas como fonte de variação no tempo, gerando uma nova via em uma medição que antes era de uma ou duas vias, por exemplo, em uma mistura com diversos analitos pode-se medir a absorção em múltiplos comprimentos de onda, caso haja uma reação química pode-se medir esse espectro em diferentes pontos no tempo, resultando em uma matriz de dados para cada amostra. Essa é uma técnica conhecida, com estudos a usando até para gerar dados de 3<sup>a</sup> ordem [27, 28].

Essa aplicação no entanto tem seus próprios limites ao ser usada em conjunção com o PARAFAC, um deles sendo a necessidade da matriz resultante ser trilinear com a matrizes de coeficientes sendo perpendiculares umas as outras, ou seja, para que os dados sejam utilizáveis, as três variáveis devem ser independentes. No caso do exemplo dado anteriormente aplicado a Eq. 1 a matriz **A** vai representar as concentrações iniciais em cada amostra, a matriz **B** vai conter os perfis espectrais normalizados e a matriz **C** vai conter os perfis cinéticos normalizados. Nesse caso,

o perfil cinético tem que ser independente da concentração inicial, o que limita as reações que possam ser utilizadas, por exemplo, em reações de decomposição de 1ª e 2ª ordem, as concentrações ao longo do tempo assumem a forma a seguir (respectivamente):

$$C_A = C_{A0}e^{-kt} \quad (14)$$

$$C_A = \frac{C_{A0}}{1 + ktC_{A0}} \quad (15)$$

Como pode-se perceber ao comparar as Eq. (14) e (15) com a Eq. (1), o produto  $a_{if}c_{kf}$  da Eq. (1) pode facilmente representar o produto da Eq. (14), com  $a_{if}$  representando  $C_{A0}$  para a amostra  $i$  e  $c_{kf}$  representando  $e^{-kt}$  para cada tempo  $k$ . O mesmo não pode ser dito da Eq. (15), não sendo separável em um produto de dois perfis independentes, ou seja, reações de primeira ou pseudo-primeira ordem são compatíveis com a criação de conjuntos de dados trilineares, mas outras cinéticas podem não ser, como o caso das de 2ª ordem.

Existem ainda outras consequências desse tipo de matriz de dados que devem ser levadas em consideração, como diferentes componentes em solução possuindo o mesmo perfil cinético, ou casos com ambos produto e reagente tendo sinal na janela medida. Apesar de idealmente esses dois componentes serem fatores distintos nas matrizes decompostas, não seria incomum que a solução obtida pelo PARAFAC seja uma solução com dois ou mais fatores degenerados, uma vez que esses fatores possuem uma forte colinearidade, diminuindo a usabilidade de alguns fatores para análises qualitativas.

## 3 Objetivos

### 3.1 Objetivo geral

- Apresentar aspectos teóricos de modelos PARAFAC, explorando trilinearidade e vantagem de segunda ordem para propor novas metodologias analíticas e solucionar problemas de modelagem como quebras de unicidade e degenerescência de resultados.

### 3.2 Objetivos específicos

- Aplicar PARAFAC a espectroscopia de fluorescência sincrônica, um arranjo experimental não-bilinear, mas de natureza fundamental semelhante a espectroscopia de excitação-emissão.
- Deconvoluir as matrizes de fluorescência sincrônica convertidas usando o PARAFAC, recuperando os perfis espectrais puros, reconstruindo o sinal em espaços vazios.
- Avaliar o comportamento do PARAFAC quando o número de componentes é fixado em dois em um tensor, contendo apenas uma amostra complexa e uma curva de calibração construída com padrões, e aplicar essa metodologia como *clean-up* virtual na determinação de melamina em leite adulterado.
- Avaliar o funcionamento do PARAFAC em sistemas hifenados onde uma das vias corresponde a cinética de uma reação.
- Avaliar o efeito de colinearidade unimodal de um tensor em modelos PARAFAC e desenvolver maneiras de contorná-la durante a modelagem.

## 4 Fluorescência sincrônica e PARAFAC

Fluorescência sincrônica é uma técnica de espectroscopia de fluorescência na qual o comprimento de onda de emissão  $\lambda_{Emi}$  sendo medido tem sempre o valor do comprimento de onda de excitação  $\lambda_{Exc}$  somado a um deslocamento constante (ou *offset*)  $\Delta\lambda$ . Ou seja, caso a luz incidente em uma amostra tenha o comprimento de onda  $\lambda_{Exc}$ , a emissão será medida no comprimento de onda  $\lambda_{Exc} + \Delta\lambda$ .

Ao medir esses espectros em diferentes valores de  $\Delta\lambda$ , é possível criar uma matriz de fluorescência sincrônica (SFM), cujas linhas e colunas estarão em diferentes valores de  $\lambda_{Exc}$  e  $\Delta\lambda$ , como mostrado na Fig. 4(A).

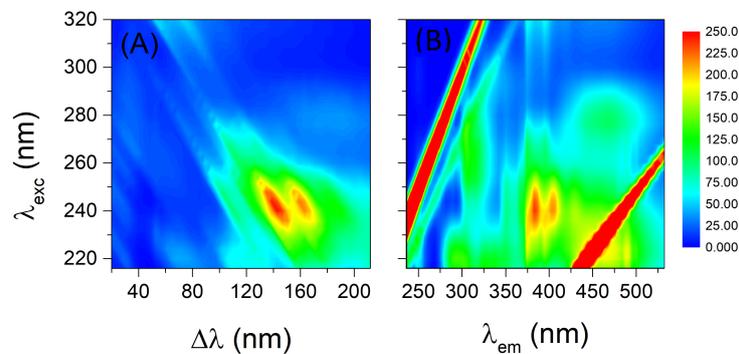


Figura 4: Exemplo de amostra medida por espectrofluorimetria usando (A) SFMs, (B) EEMs.

O uso da SFM vem com uma série de vantagens e desvantagens quando comparado a uma abordagem mais convencional como a matrizes de excitação-emissão (EEMs). Para entender essas diferenças, uma visualização do espaço espectral na qual medições fluorométricas se encontram é especialmente útil.

Como visto na Fig. 5(A), nos pontos ao redor de  $\lambda_{Emi} = \lambda_{Exc}$  e  $\lambda_{Emi} = 2\lambda_{Exc}$  o sinal é consideravelmente maior que o da luz emitida pelas amostras. Isso se deve a medição do espalhamento Rayleigh e sua difração de 2ª ordem<sup>1</sup> [29] nessas regiões, ou seja, a medição direta da luz emitida pela lâmpada do equipamento,

<sup>1</sup>Isso se deve a características intrínsecas das grades de difração usadas como monocromadores. Segundo a equação de grade (*grating equation*), em uma dada grade e ângulo de incidência, os fótons difratados a um certo ângulo tem o produto da ordem de difração e do comprimento de onda da luz difratada igual a uma constante. Portanto, a difração de 2ª ordem de um fóton de comprimento de onda  $\lambda$  ocorre no mesmo ângulo da difração de 1ª ordem de um fóton de comprimento de onda  $2\lambda$ .

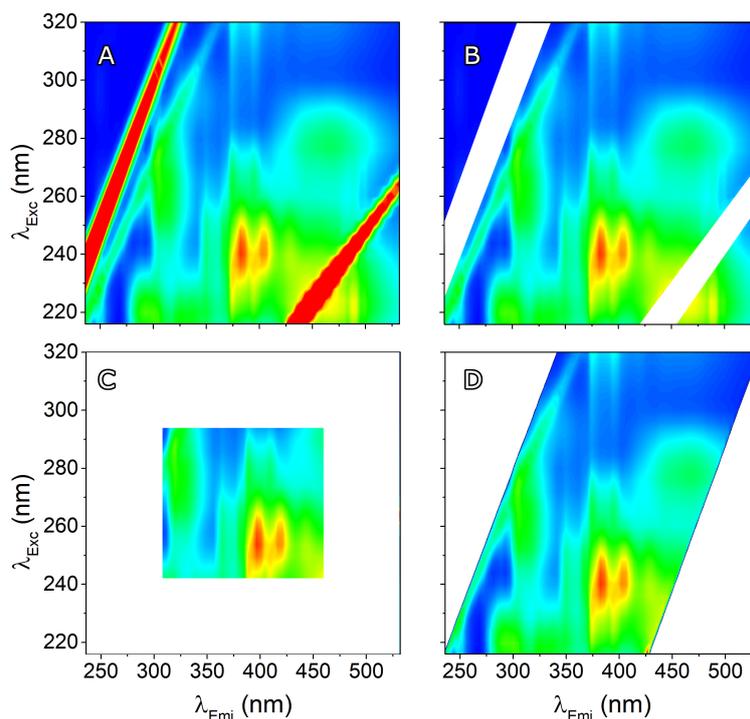


Figura 5: (A) Região espectral de interesse, (B) Região que pode ser medida em uma EEM com remoção de espalhamentos Rayleigh, (C) Região que pode ser medida em uma EEM evitando os espalhamentos Rayleigh, (D) Região que pode ser medida com uma SFM evitando os espalhamentos Rayleigh e convertendo-a em uma EEM.

cuja medição pode saturar e causar danos ao tubo fotomultiplicador (PMT) do fluorímetro.

Tanto pelas razões físicas quanto pelos efeitos desse sinal no tratamento de dados é interessante que essas regiões sejam evitadas. A Fig. 5 mostra três diferentes abordagens para evitar o sinal do espalhamento Rayleigh. Na primeira (Fig. 5(B)), uma EEM é medida contendo as regiões do espalhamento Rayleigh, mas as apagando da matriz. Essa abordagem permite medir toda a região espectral, gera uma matriz de dados trilinear com espaços vazios, mas satura o tubo fotomultiplicador (PMT), arriscando danos a longo prazo. A segunda abordagem (Fig. 5(C)) é a realização da medição de uma EEM entre os dois espalhamentos. Isso permite a geração de uma matriz trilinear completa, mas podendo deixar fora da matriz importantes frações do espaço espectral. Por fim, na terceira abordagem (Fig. 5(D)) uma SFM é medida na região a ser estudada.

Por consequência da medição em um  $\lambda_{Emi}$  a um offset constante de um  $\lambda_{Exc}$ , a SFM pode ser medida em uma região paralela ao espalhamento de Rayleigh e em seguida ser convertida em uma EEM, permitindo a medição de um espaço espectral consideravelmente maior que na segunda abordagem sem saturar o PMT. No entanto, a SFM gerada não possui características bilineares nem trilineares devido ao cisalhamento dos comprimentos de onda de emissão, o que inviabiliza seu uso em algoritmos como o PARAFAC, MCR-ALS e U-PLS/RBL. Na literatura existem diversos exemplos de estudos usando SFMs, mas as tratando como sinais não-trilineares e por isso recorrendo ao U-PLS/RBL [30] ou variações mais sofisticadas [31].

Dito isso, por compartilharem o mesmo espaço, SFMs e EEMs podem ser convertidas entre si com a formação de espaços vazios na matriz, devido a seus diferentes formatos. A conversão de uma SFM em uma EEM pode então permitir que a informação contida na SFM seja organizada em uma matriz com características trilineares, aproveitando as vantagens dos dois métodos. Implementações mais recentes do PARAFAC permitem explorar a trilinearidade do sistema para recuperar os perfis espectrais nos espaços vazios, permitindo a utilização dessa estratégia [17].

Essa abordagem foi sugerida anteriormente por Kumar [32], convertendo SFMs em matrizes com o layout de uma EEM. No entanto, a abordagem do script proposto por Kumar, mostrada na Eq. (16), apresenta certas limitações. Entre elas podemos citar que a matriz convertida inicia no menor comprimento de onda de excitação, levando a criação de várias colunas sem informação espectral alguma. A falta de um fator para aceitar variações no passo do comprimento de onda também pode causar distorções quando esses valores são diferentes de 1 nm.

$$\mathbf{X}_{EE}(i, i + ((a - \Delta d_\lambda) + \Delta d_\lambda j)) = \mathbf{X}_{Sync}(i, j); \quad (16)$$

As equações sugeridas por Kumar podem ser facilmente adaptadas para eliminar o bloco de espaços vazios antes da informação espectral e para permiti-la aceitar passos de excitação diferentes de 1 transformando a Eq. (16), fornecida por Kumar, na Eq. (17).

$$\mathbf{X}_{EE}(i, 1 + (i - 1)\Delta Ex + (\Delta d_\lambda(j - 1))) = \mathbf{X}_{Sync}(i, j); \quad (17)$$

No entanto, outra desvantagem é que SFMs medidas com passos mais altos também tem a tendência de aumentar consideravelmente de tamanho, devido ao aumento da quantidade de espaços vazios, o que carrega seus próprios problemas.

Nesse trabalho, um conversor foi desenvolvido focando em obter uma EEM compacta, que possa aceitar quaisquer valores para os passos e que seu formato dependa apenas da proporção entre os passos de excitação e *offset*.

#### 4.1 Conversão entre SFMs e EEMs

Em uma matriz obtida por fluorescência sincrônica, em cada linha  $i$  temos a seguinte série:

$$Em(i, j + 1) = Em(i, j) + \Delta d_\lambda \quad (18)$$

$$\therefore Em(i, j + n) = Em(i, j) + n\Delta d_\lambda \quad (19)$$

Com  $\Delta d_\lambda$  sendo o passo da separação da emissão.

Variando as linhas na mesma coluna, temos a seguinte série:

$$Em(i + 1, j) = Em(i, j) + \Delta Ex \quad (20)$$

$$\therefore Em(i + m, j) = Em(i, j) + m\Delta Ex \quad (21)$$

Com  $\Delta Ex$  sendo o passo do comprimento de onda da excitação.

O comprimento de onda de emissão para qualquer ponto da matriz pode ser dado pela relação a seguir:

$$Em(i, j) = Ex_0 + d_{\lambda_0} + (i - 1)\Delta Ex + (j - 1)\Delta d_\lambda \quad (22)$$

Nos pontos com os mesmos comprimentos de onda de emissão, temos que:

$$Em(i + m, j) = Em(i, j + n), \{m, n\} \in \mathbb{N} \quad (23)$$

$$Ex_0 + d_{\lambda_0} + (i + m - 1)\Delta Ex + j\Delta d_\lambda = Ex_0 + d_{\lambda_0} + (i - 1)\Delta Ex + (j + n - 1)\Delta d_\lambda \quad (24)$$

$$\therefore m\Delta Ex = n\Delta d_\lambda \quad (25)$$

Vemos portanto que  $m$  representa o afastamento vertical entre valores de mesmo comprimento de onda de emissão, enquanto  $n$  representa o afastamento horizontal. Ao reorganizar a Eq. (25) e aplicar valores genéricos de  $\Delta Ex$  e de  $\Delta d_\lambda$ , temos que:

$$\frac{\Delta d_\lambda}{\Delta Ex} = \frac{m_{min}}{n_{min}}, \{m_{min}, n_{min}\} \in \mathbb{N}^* \quad (26)$$

Com  $\frac{m_{min}}{n_{min}}$  sendo a razão mais simples entre os dois valores, temos que  $m_{min}$  será o intervalo de repetição das diferentes séries numéricas originárias da Eq. (20) e  $n_{min}$  será o deslocamento horizontal entre as diferentes linhas pertencentes a mesma série.

A conversão da matriz sincrônica para uma EEM é possível por uma transformação de cisalhamento quando todos os membros da série (21) também sejam membros da série (19), uma vez que os comprimentos de onda na mesma coluna após a operação devam coincidir em todas as linhas, logo:

$$Em(i + 1, j) = Em(i, j + n) \Leftrightarrow \Delta Ex = n\Delta d_\lambda, \{n\} \in \mathbb{N}^* \quad (27)$$

Logo, o uso da transformação de cisalhamento para conversão da matriz sincrônica para a EEM só será possível nos casos em que  $\Delta Ex$  for um múltiplo natural de  $\Delta d_\lambda$ . Com  $n_{min}$  sendo o deslocamento necessário entre linhas vizinhas em termos de posição na matriz para alinhar as séries, vemos que o mesmo assume o valor do fator de cisalhamento necessário para a transformação. Sendo assim, a transformação da matriz se dá pela Eq. (28).

$$\mathbf{X}_{EE}(i, j + (i - 1)n_{min}) = \mathbf{X}_{Sync}(i, j) \quad (28)$$

Um exemplo da conversão de uma SFM em uma EEM pode ser vista na Fig. 6.

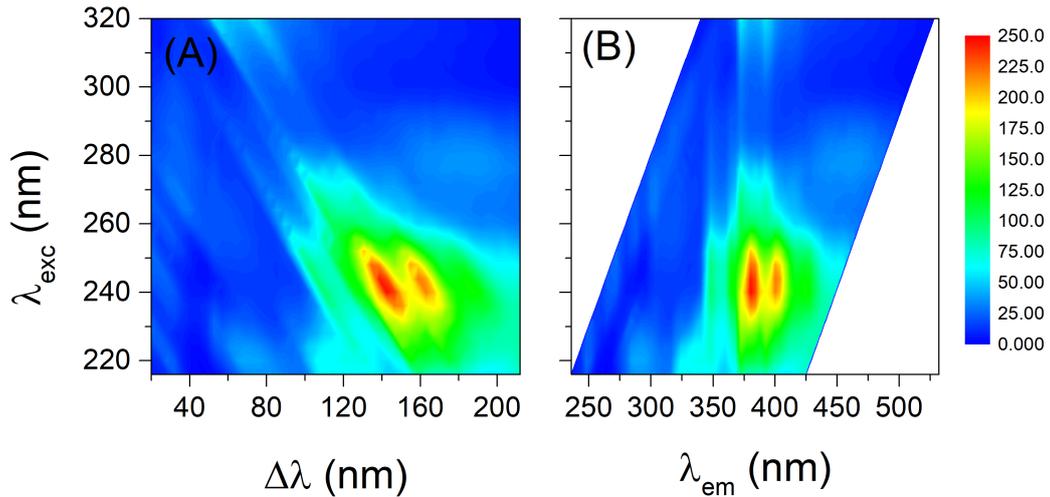


Figura 6: Exemplo de conversão de uma SFM para uma EEM usando uma transformação de cisalhamento. (A) SFM original, (B) EEM convertida.

Caso a relação mostrada na Eq. (27) não se aplique para uma determinada matriz de dados advinda de fluorescência sincrônica, com a coincidência de comprimentos de onda ocorrendo apenas em intervalos regulares, uma conversão para uma EEM com características trilineares ainda é possível, mas com formação de espaços vazios consideravelmente maior, já que não haverá continuidade na matriz no sentido dos comprimentos de onda de emissão.

Para desenvolver um algoritmo generalizado para conversão de SFMs para EEMs, as propriedades das séries (19) e (21) podem ser utilizadas. A dimensão que tanto o espaço da EEM quanto o espaço da SFM tem em comum é a do  $\lambda_{Exc}$ , sabendo disso e que será necessário um avanço de  $m_{min}$  posições nesse sentido para que haja um outro ponto com o mesmo  $\lambda_{Emi}$ , podemos dividir a matriz convertida verticalmente em grupos de  $m_{min}$  linhas, dentro do qual nenhuma das linhas terá valores de  $\lambda_{Emi}$  coincidentes. Por consequência cada  $\lambda_{Exc}$  só estará relacionado a um  $\lambda_{Emi}$  a cada  $m_{min}$  posições no sentido da emissão uma vez que é necessário que cada coluna tenha um único  $\lambda_{Emi}$ . Avançando no sentido dos valores de  $\lambda_{Exc}$  esse padrão aparecerá novamente, mas com um deslocamento horizontal de  $m_{min} \times n_{min}$  a partir da origem ao passar de um grupo para o outro. Isso resulta em uma diagonal de valores cercado por valores vazios em uma submatriz de dimensões  $m_{min} \times m_{min}$  dentro da matriz principal cujo formato se repete horizontalmente e cujo deslocamento a cada  $m_{min}$  linhas tem o valor de  $m_{min} \times n_{min}$ . Para a conversão, dividindo a matriz original em grupos verticais com  $m_{min}$  linhas, com um grupo adicional sendo considerado caso existam linhas sobrando, e usando um valor  $k$  para representar o grupo em que um ponto sendo convertido se encontra, pode-se chegar a Eq. (29) que funciona como uma equação generalizada para conversão de SFMs em EEMs.

$$\mathbf{X}_{EE}(i, i - m_{min}(k - 1) + m_{min}(j - 1) + m_{min}n_{min}(k - 1)) = \mathbf{X}_{Sync}(i, j) \quad (29)$$

A Eq. (29) tem algumas propriedades que valem a pena ser mencionadas. Com a mudança dos valores de  $m_{min}$  e  $n_{min}$  a matriz convertida pode ter formatos diferentes. Nos casos em que  $m_{min}$  tiver valor unitário, ou seja,  $\Delta Ex$  for um múltiplo natural de  $\Delta d_\lambda$ ,  $k$  sempre terá o valor de  $i$ , fazendo com que a Eq. (29) seja simplificada na Eq. (28), uma simples operação de cisalhamento com  $n_{min}$  sendo o fator de cisalhamento. Quando  $n_{min}$  tiver valores unitários, os fatores  $m_{min}(k - 1)$  se cancelam, resultando em uma matriz com linhas diagonais ininterruptas aparecendo

a cada  $m_{min}$  posições, como mostrado na Eq (30), similar a mostrada por Kumar & Mishra [32].

$$\mathbf{X}_{EE}(i, i + m_{min}(j - 1)) = \mathbf{X}_{Sync}(i, j) \quad (30)$$

A Fig 7 mostra os diferentes formatos que uma SFM de dimensões  $10 \times 10$  convertida pode assumir dado diferentes valores de  $m_{min}$  e  $n_{min}$ .

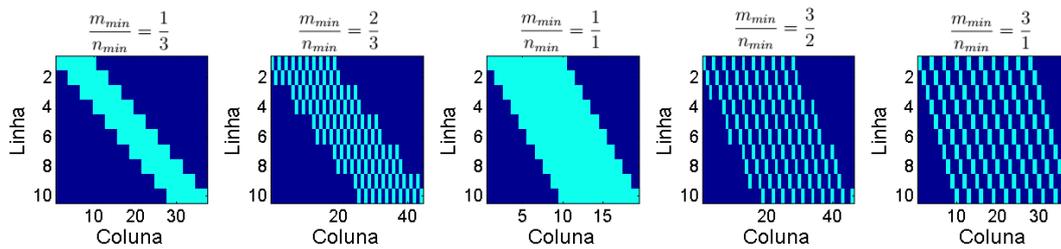


Figura 7: Matriz EEM convertida de uma SFM de tamanho  $10 \times 10$  com diferentes valores de  $m_{min}$  e  $n_{min}$ .

Vale a pena mencionar que quando  $m_{min}$  e  $n_{min}$  possuem valores diferentes de 1 simultaneamente, as colunas da EEM convertida não estarão ordenadas no sentido de  $\lambda_{Emi}$ . Apesar de isso não afetar a linearidade da matriz, é algo a se ter em mente ao fazer análises qualitativas de matrizes convertidas usando apenas a Eq. (29). Para evitar esse tipo de problema nessas situações, é necessário um passo adicional de organização horizontal da matriz para garantir a ordem crescente dos valores de  $\lambda_{Emi}$  nas colunas.

Um script para a conversão de SFMs para EEMs no *MATLAB* pode ser encontrado no Apêndice A.

## 4.2 Espaços vazios

Com as operações de conversão, as matrizes obtidas terão sempre um quantidade considerável de espaços vazios. Diferentes maneiras de lidar com esses espaços vazios e reconstruir o sinal já foram sugeridas na literatura, como o MDR-PCA e o MDR-PARAFAC, sugeridos por Tan *et al.* [33], usando o exemplo específico da aplicação dos algoritmos para restaurar espaços removidos devido a espalhamentos Rayleigh. Em sua implementação mais atual, o PARAFAC através da *N-way Toolbox* [34] é capaz de lidar com esses espaços através do uso de um algoritmo de imputação única, (*single imputation* no original), que maximiza uma função

probabilística para preencher esses espaços [17], que é a abordagem usada nesse trabalho. Dito isso, a presença e quantidade desses espaços pode aumentar consideravelmente a demanda computacional e o tempo de convergência devido aos cálculos adicionais necessários [17].

De maneira geral esses espaços permitem a projeção dos perfis espectrais de regiões medidas para regiões que não foram medidas, aproveitando das características trilineares do modelo, mas em certos casos também podem resultar na formação de picos fantasmas (*ghost peaks*), como descrito por Tomasi & Bro [17], onde sinais podem aparecer inteiramente em regiões onde não se esperava sinal algum podendo ser afetados por ruído ou outras estruturas não-multilineares.

Dependendo do sistema, deve-se considerar a substituição de frações desses espaços vazios por zeros, o que tende a reduzir os problemas citados acima, mas essa abordagem pode causar problemas se aplicada em uma região que interfira com o modelo dos analitos.

### 4.3 Aplicações simuladas

Para testar a aplicabilidade da Eq. (29) e usar as EEMs geradas no PARAFAC, um conjunto de dados simulado foi criado usando o gerador de espectros descrito no Apêndice B com 29 misturas de 4 componentes (25 misturas usando um planejamento de Brereton [35] em 5 níveis e 4 matrizes com cada componente puro) em um espaço espectral indo de 200 a 320 nm de excitação e de 200 a 528 nm na emissão, com passos de 1 nm em ambos os sentidos. As EEMs dos componentes criados podem ser vistas na Fig. 8.

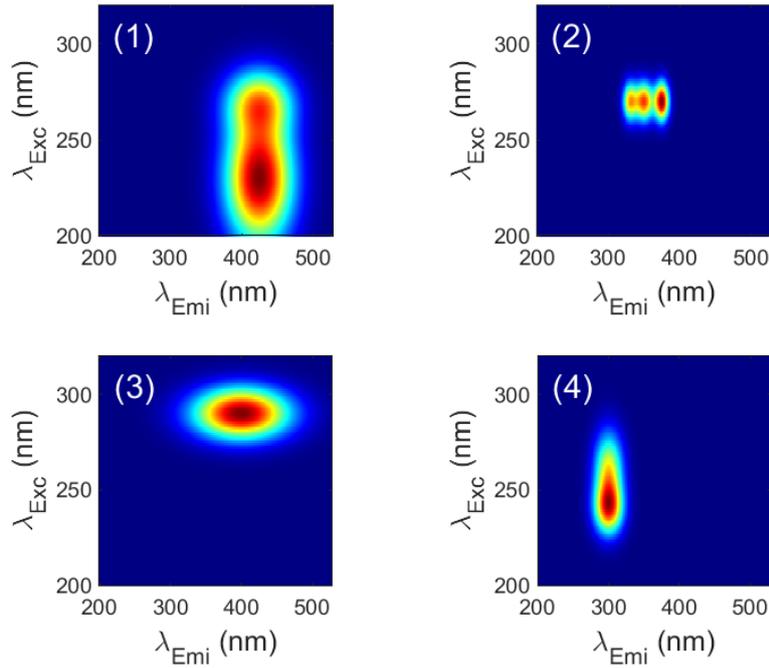


Figura 8: Componentes criados para testar o uso do PARAFAC em SFMs convertidas.

Cinco tensores compostos por SFMs simuladas, criadas a partir do conjunto de dados original, foram então montados usando  $\frac{\Delta d_\lambda}{\Delta Ex} = \frac{m_{min}}{n_{min}}$ , com valores de 1/3, 2/3, 1/1, 3/2, e 3/1 (chamados aqui de casos 1 a 5 respectivamente), os mesmos mostrados na Fig. 7. Todas as SFMs cobriram aproximadamente a mesma região espectral, mas com resoluções distintas. Todas as matrizes começaram com  $Ex_0$  em 200 nm e  $d_{\lambda_0}$  em 20 nm. Todos os tensores foram convertidos e deconvolvidos no PARAFAC em 4 componentes com restrições de não-negatividade aplicados nos 3 modos.

Os resultados dos componentes deconvolvidos e a comparação com os componentes originais puros para cada um dos tensores testados pode ser encontrado na Fig. 9.

Como visto na Fig.9, os SFMs convertidos puderam ser deconvolvidos usando o PARAFAC, recuperando os perfis individuais de cada componente com diferentes níveis de precisão nos perfis obtidos. A informação quantitativa também foi recuperada. Dito isso, ao analisar qualitativamente os perfis recuperados, é possível ver que os perfis obtidos quando  $m_{min}$  é unitário fornecem uma maior acurácia visual para identificação dos perfis espectrais quando comparado aos demais. As matrizes

convertidas também possuem uma proporção consideravelmente menor de espaços vazios, com 40,6 e 41,2% de espaços vazios nos casos 1 e 3 respectivamente, comparados a valores de 69,9, 80,0 e 80,3% para os casos 2, 4 e 5. Também foi possível ver a formação de picos fantasmas em regiões sem nenhuma informação espectral, ou seja, aparentes sinais em espaço sem informação espectral de nenhuma das vias, parecendo ser mais comuns nas extremidades da matriz, onde o modelo depende de um número consideravelmente menor de pontos medidos.

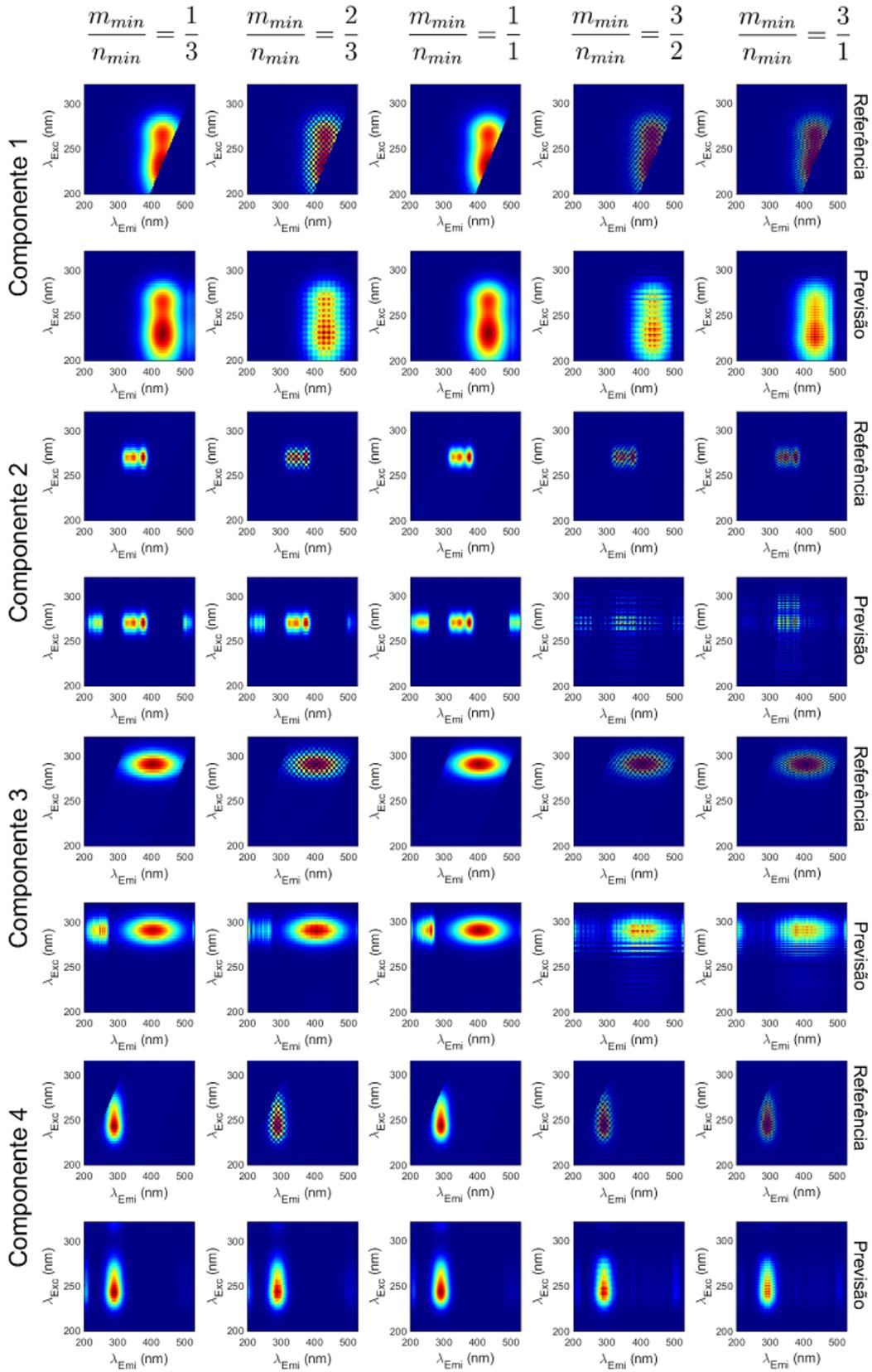


Figura 9: Componentes puros e componentes recuperados para os 5 casos.

#### 4.4 Mistura de HPAs

Para fornecer um exemplo de um sistema utilizando SFMs convertidas, um banco de dados foi construído com SFMs das misturas de 7 HPAs preparadas em água, fluoranteno, pireno, fenantreno, antraceno, fluoreno, acenafteno e naftaleno, com concentrações nos níveis indicados na Tab 1. Dos sete, fluoranteno, antraceno, pireno e fenantreno foram tratados como analitos e o fluoreno, acenafteno e naftaleno foram tratados como interferentes devido a seus estoques estarem vencidos. As SFMs foram medidas com excitação de 216 a 320 nm com passos de 4 nm e o offset de 20 a 208 nm com passos de 4 nm. A fenda de excitação foi ajustada para um *bandpass* de 20 nm e a de emissão para 5 nm.

Tabela 1: Níveis de concentração do conjunto de teste com os 7 HPAs.

Nível	FNT (ppb)	PYR (ppb)	PHE (ppb)	ANT (ppb)	FLU (ppb)	ACN (ppb)	NAF (ppb)
-2	15,0	1,5	0,8	0,3	0,8	0,3	1,5
-1	30,0	3,0	1,6	1,0	1,6	0,6	3,0
0	45,0	4,5	2,4	1,7	2,4	0,9	4,5
+1	60,0	6,0	3,2	2,3	3,2	1,2	6,0
+2	75,0	7,5	4,0	3,0	4,0	1,5	7,5

O tensor com as SFMs das amostras foi então convertido para EEMs e deconvoluído no PARAFAC em 6 componentes sob diferentes condições.

Um dos parâmetros a ser definido é o que fazer com os espaços vazios resultantes da conversão das matrizes. Como nesse caso  $\Delta E x = \Delta d_\lambda$ , temos que com a conversão existem duas regiões de espaços vazios nas matrizes convertidas, um a esquerda do espalhamento Rayleigh e um a direita da difração de 2ª ordem do espalhamento Rayleigh, como mostrado na Fig. 10. Aqui analisamos três diferentes possibilidades: Deixar ambas regiões vazias, com valores Not-a-Number (NaN), preencher ambas as regiões com zeros, e uma combinação das duas abordagens.

A Fig. 10 mostra a reconstrução de uma amostra do conjunto a partir dos perfis deconvoluidos com o PARAFAC. Na Fig. 10(B) temos a reconstrução da EEM usando NaN em ambos os espaços. Nesse caso, ocorreu a formação de um pico fantasma provavelmente devido a proximidade ao espalhamento Raman, que não é

trilinear<sup>2</sup>, mas que pode ser facilmente projetado no espaço vazio e permanecer com sinal naquela região devido a maneira que o algoritmo do PARAFAC funciona, tendo resíduo zero nas regiões com espaços vazios. Por outro lado, nas regiões sem esses espalhamentos o algoritmo conseguiu projetar os espectros dos HPAs, reconstruindo o sinal na região que teria a difração de 2<sup>a</sup> ordem do espalhamento Rayleigh, o que impossibilitaria a medição naquela região. Preencher ambos os espaços vazios com zeros evitou a formação de picos fantasmas, mas distorceu consideravelmente o espectro reconstruído na outra região, como mostrado na Fig. 10(C). Uma abordagem mista pode ser visto na Fig. 10(D), eliminando picos fantasmas na região onde não era esperado sinal (com  $\lambda_{Emi} < \lambda_{Exc}$ ) e permitindo que o PARAFAC projete o sinal das amostras na região onde é esperado que o sinal exista.

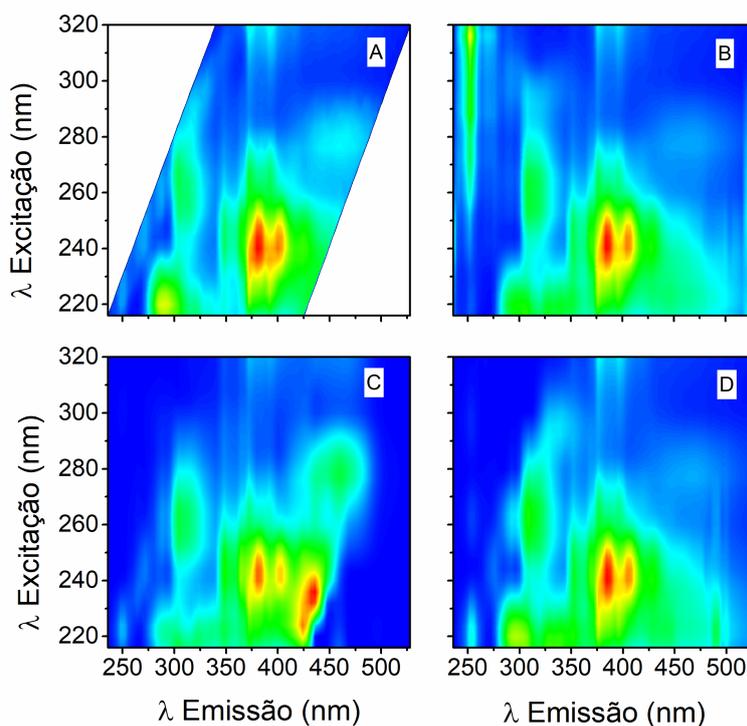


Figura 10: SFM convertida para EEM reconstruída usando diferentes valores para preencher os espaços vazios.

Usando a abordagem mista, o conjunto de dados foi processado usando o PARAFAC com 6 componentes, sendo capaz de recuperar e projetar os espectros do

<sup>2</sup>Bro reportou que a forma e aparição desses picos fantasmas parece estar intimamente ligada a presença de informação não-trilinear no tensor [17], como é o caso do espalhamento Raman.

fluoranteno, antraceno, pireno, fenantreno e fluoreno, exibindo-os sem interferentes como mostrado na Fig. 11.

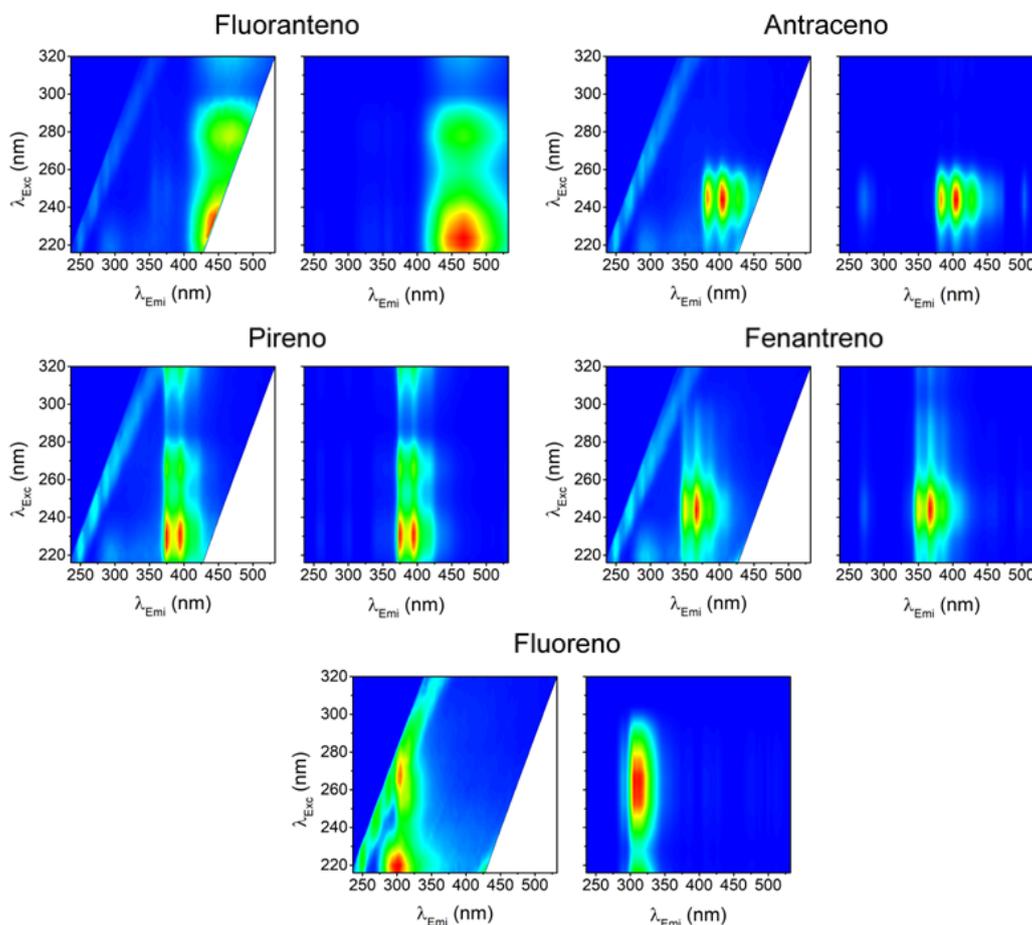


Figura 11: Componentes puros e recuperados para o fluoranteno, antraceno, pireno e fenantreno.

Vale a pena mencionar a utilidade do método especialmente em casos como o do fluoranteno, onde parte considerável do espectro se encontrava em uma região que não podia ser medida diretamente sem saturação da PMT, mas que pode fornecer informação o bastante no resto da matriz para recuperar seu perfil completo aproveitando da trilinearidade inerente a medidas espectrofluorimétricas.

Curvas de autopredição foram traçadas usando os valores no modo A do PARAFAC para os HPAs identificados, obtendo os RMSEs, REPs e  $r^2$  mostrados na Tab. 2.

Tabela 2: Valores de mérito das curvas de autopredição do fluoranteno, antraceno, pireno, fenantreno e fluoreno.

	FNT	ANT	PYR	PHE	FLU
<b>RMSE (ppb)</b>	3,69	0,10	0,64	0,29	0,30
<b>REP</b>	6,15%	3,73%	10,74%	8,98%	9,46%
<b>r<sup>2</sup></b>	0,97	0,99	0,92	0,94	0,93

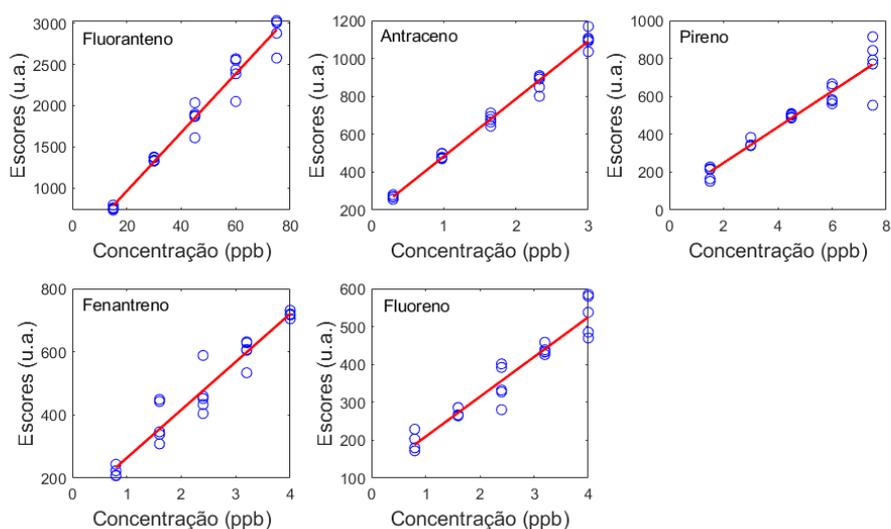


Figura 12: Curvas de autopredição para o fluoranteno, antraceno, pireno, fenantreno e fluoreno feitas a partir dos valores no modo A do PARAFAC e as concentrações de referência.

Essa abordagem que desenvolvemos foi utilizada por Araújo *et al.* para analisar água marinha contaminada com petróleo após o incidente de 2019, que resultou no derramamento de óleo cru em quantidades estimadas acima de 1.000 toneladas na costa nordestina. O método possibilitou a análise qualitativa de um conjunto com 55 amostras coletadas em diferentes pontos cobrindo 430 km da costa nordestina, permitindo a identificação de 4 compostos fluorescentes presentes nas amostras, dois dos quais foram ligados a compostos relacionados ao petróleo (similares ao naftaleno e dibenzotiofeno) [23].

## 4.5 Conclusão

A metodologia proposta foi capaz de converter SFMs em EEMs, permitindo a medição de uma região espectral maior, evitando os espalhamentos Rayleigh, mas mantendo as vantagens de um conjunto de dados trilinear. O algoritmo de conversão desenvolvido possui certas vantagens comparado ao proposto por Kumar [32], gerando matrizes mais compactas e com menos espaços vazios, também aceitando valores não-inteiros nos passos de excitação e offset.

O PARAFAC se mostrou eficaz em recuperar os perfis espectrais nas regiões com espaços vazios, aproveitando das características trilineares do tensor e usando um método de imputação singular. O PARAFAC foi mais efetivo em matrizes que possam ser transformadas com uma operação de cisalhamento da SFM, ou seja, com o passo do comprimento de onda de excitação é um múltiplo do passo do offset. No entanto, a possibilidade da formação de picos fantasmas é uma possibilidade nesses sistemas que deve ser levada em conta ao interpretar os dados.

## 5 PARAFAC como *clean-up* virtual

### 5.1 *Clean-up* virtual

Uma consequência da vantagem de segunda ordem é a possibilidade de se construir curvas de calibração usando apenas soluções padrão preparadas no laboratório. Isso porque não é necessário que nenhuma informação do interferente esteja presente no conjunto de calibração. Se um único analito de interesse está presente na amostra, e a calibração segue a estratégia de modelar uma amostra por vez, podemos considerar fazer uma decomposição em apenas dois componentes. Nesse caso o analito será um dos componentes, definido como a espécie com o perfil espectral em comum com os padrões.

Os demais constituintes da amostra, que não estão nos padrões, constituirão o sinal de fundo capturado pela segunda componente do PARAFAC. Este pode até ser o sinal de uma mistura. No entanto não existe no sistema graus de liberdade suficiente para resolvê-las. Também não há necessidade em fazê-lo. Desde que o analito esteja modelado, e que não haja interesse no conjunto complexo de constituintes da amostra para além do analito, o problema analítico está resolvido.

Cromatografia virtual é uma denominação comumente usada para explicar os métodos de segunda ordem do ponto de vista da sua aplicação como métodos de separação de sinais. Aqui chamamos a atenção para casos onde a separação dos sinais não é de todo necessária, bastando na prática separar o analito.

Nesse contexto tratamos o PARAFAC como um *clean-up* virtual, ou seja, uma maneira de remover os interferentes da matriz, descartando-os em seguida, sem necessariamente resolvê-los. Uma aplicação dessa interpretação é dada a seguir:

### 5.2 Aplicação - Determinação de melamina em leite

Melamina (CAS 108-78-1) é um trímico de cianamida comumente usado na produção de plásticos e produtos anti-chama. Possuindo 67% de sua massa na forma de átomos de nitrogênio, melamina é um dos compostos mais comumente utilizados para adulterar leite e produtos lácticos, uma vez que o nitrogênio liberado durante a digestão química da melamina pode levar a uma medição incorreta da quantidade de proteínas pelo teste de Kjeldahl, permitindo que os produtos sejam diluídos. [36].

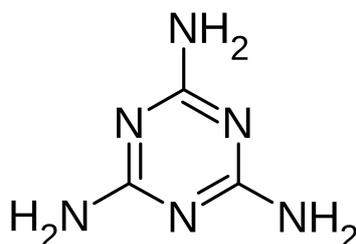


Figura 13: Estrutura molecular da melamina.

Além de ser um crime contra o consumidor, essa prática causa danos à saúde, pois, devido à sua baixa solubilidade, a melamina se acumula nos rins e pode levar à insuficiência renal [36]. Em 2008, a descoberta de um esquema de adulteração de leite de melamina na China envolvendo 22 empresas e atingindo mais de 290 mil crianças foi divulgada com seis mortes confirmadas [36]. Os primeiros casos foram relacionados à adulteração de leite em pó nas fábricas do Grupo *Sanlu*, que era um dos maiores produtores de leite do mundo, com a posterior investigação levando concluindo que a adulteração também foi realizada em ovos, cereais, biscoitos e outros alimentos [36]. Também houve casos anteriores de insuficiência renal causada em animais na República da Coreia (2004) e nos Estados Unidos (2007) devido a uma adulteração envolvendo melamina, mascarando a falta de proteína em rações animais [37].

Por ser uma matriz complexa, os métodos padrão de detecção de melamina no leite são baseados em técnicas cromatográficas, que normalmente requerem vários passos laboriosos de separação, geralmente sendo mais caros e demorados [38]. Há uma longa lista de técnicas não-cromatográficas para determinar melamina no leite usando infravermelho [39], Raman [40, 41] e espectrometria UV [42, 43], entre outros, como citado por Ritota e Manzi [44].

As técnicas padrão geralmente focam na determinação da melamina para avaliar a contaminação ou adulteração intencional de amostras de leite, mas cada uma dessas situações tem concentrações esperadas de melamina consideravelmente diferentes. Em relação à contaminação do leite, as concentrações máximas permitidas de melamina no leite variam entre 50 ppb e 2500 ppb dependendo da legislação local, sendo 1000 ppb o limite máximo recomendado pela Organização Mundial de Saúde [45]. Por outro lado, as quantidades esperadas são consideravelmente maiores nos casos de adição intencional de melamina para mascarar adulteração por diluição, atingindo centenas de partes por milhão. O uso de técnicas rápidas e de

baixo custo baseadas em espectroscopia para identificar essas ocorrências pode reduzir significativamente os custos e talvez até mesmo ser usado *in situ* dependendo da metodologia [41].

Bakain et al. [42] tentou determinar melamina em leite integral com o uso de dados espectrofotométricos UV-Vis obtidos em diferentes valores de pH usando PARAFAC, MCR-ALS e U-PLS/RBL, obtendo melhores resultados com U-PLS/RBL ao relatar resultados ruins para PARAFAC e MCR-ALS devido a problemas de deficiência de classificação no conjunto de dados. Liu et al. também usou espectrofotometria UV-Vis e variação de pH para gerar dados de segunda ordem, mas aplicando análise do fator de aniquilação de posto (RAFA) como algoritmo de calibração [43]. Em outros trabalhos, Ortiz et al. [46] usou PARAFAC<sup>3</sup> em dados HPLC-DAD para determinação melamina, mais especificamente melamina lixiviada de plástico, com o uso de PARAFAC2 sendo necessário devido a uma mudança no tempo de retenção da melamina.

Melamina tem um intenso sinal de fluorescência que pode potencialmente ser usado para sua quantificação. Usando EEMs em amostras de água/leite, esses sinais podem ser organizados como dados de segunda-ordem, fornecendo vantagem de segunda ordem, e permitindo o uso de algoritmos de calibração de ordem superior, como o PARAFAC. O PARAFAC tem sido usado com sucesso em matrizes alimentares desde sua proposição como ferramenta de química analítica por Bro [2] e em muitas outras aplicações [7, 8, 9, 10, 11]. No caso do leite, podemos destacar a detecção de resíduos antibióticos por espectroscopia de fluorescência [47].

Neste trabalho, o método baseado em calibração de segunda-ordem usando PARAFAC e espectroscopia de fluorescência EEM é proposto como uma alternativa para detectar adições fraudulentas de melamina ao leite para mascarar a adulteração do mesmo por diluição com água. O método proposto utiliza um único passo de *clean-up* com ácido acético, usando o PARAFAC para realizar um *clean-up* virtual dos interferentes restantes, resultando em um procedimento rápido, de baixo custo e consistente com os princípios de química verde.

---

<sup>3</sup>Modelo similar ao PARAFAC, desenvolvido com o objetivo de lidar com dados de ordem superior onde a trilinearidade quebra em uma das vias, ou seja, o modelo não requer trilinearidade estrita. Ao contrário do PARAFAC, o modelo é capaz de lidar com dados com diferentes tamanhos ou que possuam deslocamentos em uma das vias, o que o torna especialmente útil para sistemas cromatográficos.

### 5.2.1 Metodologia

#### Preparação das amostras adulteradas

Foram adquiridas 12 amostras de leite UHT no comércio varejista de João Pessoa-PB, compostas por 6 amostras de leite integral, 3 amostras de semi-desnatado e 3 amostras de leite desnatado entre cinco marcas diferentes. Estas amostras foram adquiridas durante um período de dois anos e foram utilizadas para produzir as amostras adulteradas.

A adulteração de leite por diluição com água foi realizada em laboratório em uma faixa de 2% a 12% de água adicionada. Assim, para compensar a redução no teor de proteína causada pela diluição, melamina foi adicionada para aumentar o teor de nitrogênio e, portanto, o teor proteico aparente ao nível esperado para o leite puro.

Considerando a razão de Kjeldahl para leite [48]:

$$m_{Proteinas} = 6,38m_N \quad (31)$$

e considerando que a melamina tem uma fração de massa de nitrogênio de 67%, a quantidade de melamina necessária para aumentar a quantidade de nitrogênio total para cada amostra adulterada foi calculada usando Eq. (32)

$$m_{melamina} = \frac{C_p V_T W_f}{4,27} \quad (32)$$

onde  $m_{melamina}$  é a quantidade de melamina necessária para um volume de leite final  $V_T$ , contendo uma fração de água adicionada  $W_f$  e uma concentração de proteína  $C_p$ . A constante 4,27 é o produto da razão de Kjeldahl para leite (6,38) e a fração de massa de nitrogênio em melamina (0,67). Usando esse procedimento, as amostras de leite adulterado variaram de 140 ppm a 870 ppm de melamina. No total, foram produzidas 40 amostras de leite adulterado e processadas juntamente com 18 amostras de leite puro. Entre as amostras adulteradas, 22 eram de leite integral, 7 de semi-desnatado e 11 de leite desnatado. 25 desses feitos em duplicatas autênticas. Para uma única amostra, foram feitas 8 réplicas para estimar o LoD. Das 18 amostras de leite puro, 8 eram de leite integral, 4 semi-desnatados e 6 desnatados. Vale ressaltar que, devido ao erro de pesagem, amostras no mesmo nível de diluição podem apresentar variações na concentração de melamina na ordem de 50-80 ppm. A concentração de melamina para cada amostra pode ser encontrada no Apêndice C.

### Procedimento analítico

As soluções padrão utilizadas para a obtenção dos modelos de calibração foram preparadas em água utilizando melamina (99%) na faixa de 60 ppm a 915 ppm. Amostras de validação foram preparadas seguindo uma metodologia baseada na primeira parte do procedimento descrito por Finete et al. [38], em que 7 mL de leite foram adicionados a uma mistura de 3 mL de água e 50  $\mu\text{L}$  de ácido acético, com a mistura sendo centrifugada por 20 minutos a 4000 rpm. O sobrenadante foi então transferido para um frasco volumétrico de 10 mL, com o precipitado sendo lavado duas vezes com 100  $\mu\text{L}$  de água, com a água usada na lavagem do precipitado também sendo adicionada ao frasco. O frasco volumétrico foi então preenchido com água e homogeneizado.

Tanto os padrões quanto as amostras de teste foram diluídas 200 vezes para a medição do EEM usando um espectrofluorômetro Cary Eclipse (Agilent). Os comprimentos de onda de excitação e emissão variaram de 240 nm a 260 nm e de 320 nm a 440 nm, respectivamente, ambos com um passo de 2 nm, 20 nm de bandpass, e tempo de integração de 0,2 segundos.

Para o processamento de dados e construção do modelo de calibração, o algoritmo PARAFAC foi executado pela *N-way Toolbox* para MATLAB [34].

### 5.2.2 Resultados e discussão

#### Avaliação do nível de diluição das amostras

O leite é uma matriz complexa, e mesmo após a processo de extração de caseína, ainda há muitas proteínas na amostra tratada. Fig. 14(a) mostra o EEM de uma amostra de leite após a limpeza conforme descrito na metodologia. Fig. 14(b) mostra o EEM de um 1. 5 ppm melamina padrão aquoso, e Fig. 14(c) o EEM de uma amostra de leite pré-tratada contendo 1,5 ppm melamina.

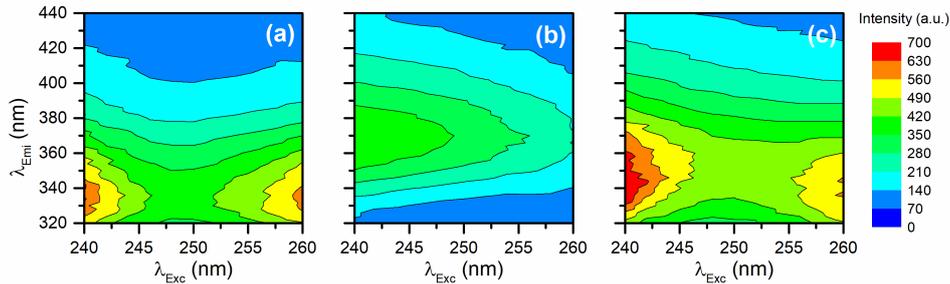


Figura 14: EEM de uma amostra de leite pré-tratado não adulterada (a), padrão aquoso de melamina a 1,5 ppm (b) Melamina a 1,5 ppm em leite pré-tratado (c).

Observa-se que a amostra de leite não adulterada tem um sinal de fluorescência intensa devido à presença de proteínas que não foram removidas no passo de remoção de caseína. Embora os métodos de calibração de segunda ordem possam lidar adequadamente com interferentes químicos separando seus sinais daqueles do analito, eles não são capazes de lidar com efeitos matriciais relacionados a propriedades físicas, como índice de refração e turbidez, por exemplo. Os processos de transferência de energia entre melamina e proteínas lácteas dependem das concentrações dessas espécies químicas e podem afetar o sinal de fluorescência da amostra, causando não linearidades na relação entre sinal analítico e concentração de analitos. A diluição da amostra poderia ser usada para aumentar a distância intermolecular média a um ponto onde essas transferências de energia sejam insignificantes.

O PARAFAC foi aplicado com dois componentes, utilizando restrições de não negatividade nos três modos, em quatro conjuntos diferentes de padrões contendo melamina e leite simulando diferentes diluições amostrais após a limpeza da amostra. A Fig. 15 mostra a curva de calibração pseudo-univariada de melamina para diferentes diluições. Cada nível de diluição tinha suas próprias configurações de fluorômetro para maximizar o sinal, portanto, as curvas em Fig. 15 não possuem intensidades comparáveis e são utilizados apenas para avaliar a relação linear entre os escores do PARAFAC e a concentração de melamina.

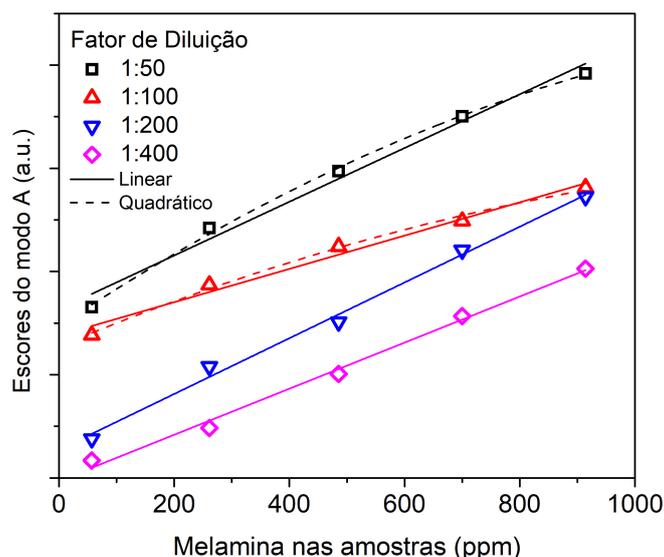


Figura 15: Curvas de calibração pseudo-univariada para melamina em quatro diluições diferentes.

Como visto na Fig. 15, para as diluições de 1:50 e 1:100, a curva segue um comportamento mais quadrático que linear. Uma curva preparada a partir de padrões aquosos de melamina sem leite é linear na faixa estudada. Assim, na presença de constituintes do leite, a relação entre concentração de melamina e sinal analítico não é linear, provavelmente devido às interações de melamina com componentes amostrais em diluições de 1:50 e 1:100. Tal efeito não é observado em diluições de 1:200 e 1:400, levando a curvas lineares.

Devido à sobreposição de espectros de excitação e emissão, a luz emitida no centro da cubeta pode ser reabsorvida pela própria amostra se ela for suficientemente concentrada. Isso é conhecido como o efeito filtro interno. É uma boa prática verificar se a absorção na faixa de emissão é inferior a 5%. As amostras diluídas mais de 200 vezes tiveram absorvância máxima abaixo de 0,05, indicando ausência do efeito do filtro interno.

### Resultados da deconvolução por PARAFAC

A Figura 16 mostra os perfis de excitação e emissão deconvoluídos (matrizes  $\mathbf{B}$  e  $\mathbf{C}$ , respectivamente) para um conjunto de dados contendo as EEMs dos padrões de melamina e uma amostra de leite contaminado. Dois componentes podem ser distinguidos, com a linha azul sólida correspondendo ao sinal de melamina, enquanto a linha vermelha tracejada corresponde ao sinal de interferentes na amostra, ori-

ginado das proteínas residuais não removidas no *clean-up* físico. A Fig. 16 mostra que aqui, o PARAFAC funciona como um passo virtual de limpeza-separando o sinal de melamina dos sinais relacionados ao leite.

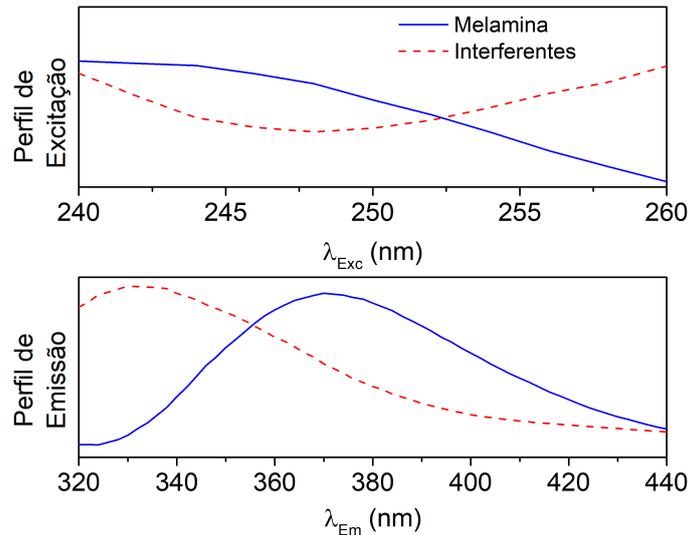


Figura 16: Perfis de emissão e excitação de uma amostra de leite decomposta pelo PARAFAC. Linha azul sólida: melamina; linha tracejada vermelha: interferentes da amostra.

Existem apenas padrões aquosos de melamina no conjunto de calibração e uma única amostra é modelada de cada vez. Assim, apenas dois componentes podem ser efetivamente separados: melamina, que varia em concentração ao longo dos padrões e amostra, e os interferentes relacionados ao leite, que estão presentes apenas na amostra. Diferentes amostras de leite podem resultar em diferentes perfis interferentes devido à variação natural na composição do leite, mas terão o mesmo perfil para melamina, que é determinado pelos perfis espectrais compartilhados do conjunto de calibração e da amostras.

A Figura 17 mostra a concentração prevista de melamina nos leites adulterados contra a referência de melamina adicionada. Pode-se ver uma boa relação linear entre os valores previstos e os de referência.

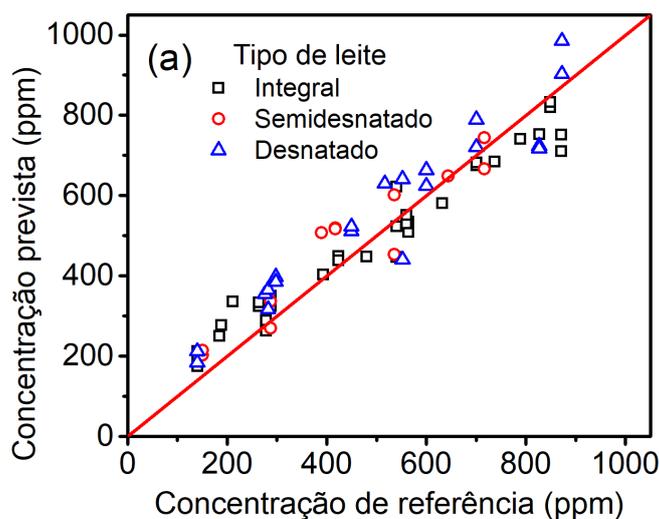


Figura 17: Valores de predição versus referência da concentração de melamina para diferentes tipos de leite obtidos utilizando PARAFAC.

Os valores de mérito para o modelo PARAFAC podem ser vistos na Tab. 3. O método proposto teve um LoD estimado de 120.6 ppm com nível 95% de confiança, que é o valor esperado para adulterações de pouco menos que 2% em volume. Nesse nível, o método pode ser considerado adequado para detectar diluição fraudulenta intencional do leite. Para amostras acima do LoQ, os valores de recuperação ficaram entre 80% e 134%. O Apêndice C contém esses valores para as amostras avaliadas.

Tabela 3: Valores de mérito para a quantificação de melamina usando PARAFAC.

Valores de mérito	
<b>RMSEP</b>	68,6 ppm
<b>REP</b>	8,02% ppm
<b><i>r</i></b>	0,97
<b>LoB</b>	98,0 ppm
<b>LoD</b>	120,6 ppm
<b>LoQ</b>	361,8 ppm

Os valores de mérito mostradas na Tab. 3 são comparáveis aos obtidos em outros métodos espectroscópicos utilizados para determinar melamina no leite [40, 41, 39]. Métodos cromatográficos estão disponíveis na literatura e são comumente usados para detectar contaminação e adulteração fraudulenta com melamina.

mina, geralmente sendo métodos com LoDs muito menores, mas também de maior complexidade experimental. Embora seja possível usá-los para determinar melamina no leite no contexto de adulterações, métodos experimentalmente mais simples e menos trabalhosos podem ser alternativas apropriadas para essas aplicações.

### 5.2.3 Conclusão

Através da restrição intencional do número de componentes do PARAFAC em 2, do conjunto de calibração construído apenas com o analíto puro e da modelagem de uma amostra por vez, foi possível isolar o analíto em uma componente e todos os demais interferentes em outra. Este procedimento foi denominado *clean-up* virtual. A estratégia para *clean-up* virtual aplicada nesse trabalho permitiu o desenvolvimento de um método analítico para determinação de melamina em leite adulterado que segue os preceitos de química analítica verde. O método proposto mostrou bons resultados analíticos, provando ser eficiente na detecção de adulterações a partir de 2% em volume, com LoD de 120.6 ppm, e RMSEP de 68,6 ppm. Esse trabalho foi publicado na *Food Chemistry* em 2021 [49].

## 6 Sistemas espectrocinéticos

A disseminação do uso de técnicas hifenadas para geração de dados de ordem superior a partir da combinação de duas técnicas [16] foi um dos responsáveis pela popularização de métodos de segunda ordem como o PARAFAC. Uma dessas técnicas consiste na utilização de reações químicas para mudar a composição de uma amostra com o tempo, o que por sua vez muda seu espectro de absorção. Esse tipo de técnica resulta na geração de uma matriz de dados de segunda ordem, com uma das vias no sentido dos comprimentos de onda e a outra no sentido do tempo de medição [16].

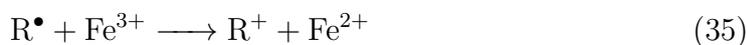
Para explorar os aspectos desse tipo de técnica e entender melhor suas peculiaridades, foram realizadas uma série de experimentos simulados e simulações apresentadas a seguir.

### 6.1 Aplicação - Corantes alimentícios

#### 6.1.1 Reações de Fenton e Fenton-like

Reações de Fenton são processos oxidativos avançados que consistem no uso de íons ferrosos, ou de outros metais com múltiplos estados de oxidação, para reagir com peróxido de hidrogênio produzindo radicais livres com alta capacidade oxidante [50].

As reações nesse sistema se dão da seguinte forma:



A baixa especificidade desse tipo de reação permite com que ela seja usada para degradar uma vasta gama de compostos orgânicos. Uma outra vantagem é que as cinéticas envolvidas com esse tipo de reação são em vários casos de pseudo-primeira ordem ou pseudo-primeira ordem combinada [51], o que teoricamente permitiria sua utilização na criação de dados de ordem superior para deconvolução com o PARAFAC.

Como visto pelas reações envolvidas nesse sistema, a constante de velocidade aparente da decomposição dos analitos dependerá fortemente das concentrações de

íons  $\text{Fe}^{2+}$ , de  $\text{H}_2\text{O}_2$  e do pH, portanto esses valores devem ser ajustados para que a velocidade das reações sejam compatíveis com as técnicas de medição utilizadas. Por exemplo, caso demore  $x$  segundos para medir um espectro de absorção, não é interessante que a concentração mude consideravelmente nesse intervalo de tempo, logo o ajuste desses parâmetros é essencial para a utilização dessa técnica. No caso de um tempo de medição em que a velocidade de reação seja considerável, a utilização ainda é possível com o PARAFAC e a trilinearidade é mantida, mas é preciso ter em mente que os perfis espectrais estarão multiplicados por constantes em cada comprimento de onda específico e de formato similar ao perfil cinético em um intervalo de  $x$  segundos no sentido da medição, o que pode afetar a análise qualitativa devido a mudança nos perfis espectrais deconvoluidos.

### 6.1.2 Analitos utilizados

Para testar o método, uma mistura de corantes alimentícios foi proposta como sistema modelo. Os corantes utilizados nesse estudo foram o amarelo tartrazina (CAS 1934-21-0) e o amarelo crepúsculo (CAS 2783-94-0), ambos amplamente utilizados em alimentos com tons amarelo/alaranjados, como isotônicos, gelatinas, refrigerantes, sucos, entre outros.

Nesse sistema, é possível acompanhar a degradação dos compostos através da medição do espectro de absorção na faixa visível da luz ao longo do tempo, possibilitando a geração de dados de segunda ordem. Um sistema fenton foi escolhido para realizar a degradação devido a sua baixa seletividade e forte capacidade oxidativa.

### 6.1.3 Metodologia

#### Reação de decomposição

São adicionados 25  $\mu\text{L}$  de uma solução de  $\text{H}_2\text{SO}_4$ , 1 mL da solução do analito, 1 mL de  $\text{Fe}^{2+}$  e por fim 1 mL de  $\text{H}_2\text{O}_2$  com a cubeta já inserida no suporte. O início da medição deve ser feito imediatamente após a adição do  $\text{H}_2\text{O}_2$ .

Para ajustar a velocidade de reação, foi necessário definir os valores desejados das concentrações de  $\text{Fe}^{2+}$ ,  $\text{H}_2\text{O}_2$  e do pH no meio reacional. Foi definido a duração total do experimento seria de 350 segundos. Para que a reação ocorresse dentro desse espaço de tempo, diferentes valores das concentrações dos reagentes e do pH foram testados, chegando nos valores mostrados na Tab. 4:

Tabela 4: Concentração dos reagentes de Fenton utilizados para degradar os corantes alimentícios.

Reagente	Estoque	Cubeta
	$C_0$ (mol/L)	$C_0$ (mol/L)
$H_2O_2$	2,25E-03	7,50E-04
$FeSO_4(H_2O)_7$	1,50E-04	5,00E-05

Os reagentes utilizados estão solubilizados em água em estoques com pH ajustado para 3. Uma solução de  $H_2SO_4$  1M foi utilizada para a acidificação tanto dos reagentes quanto do meio reacional, fazendo com que o meio reacional estivesse em um pH próximo a 3 no início da reação.

### Planejamento experimental

Um conjunto de amostras de calibração e um de validação contendo misturas entre os dois corantes estudados. O conjunto de calibração consistiu em 10 amostras, com 5 contendo apenas amarelo crepúsculo (AC) e 5 contendo apenas amarelo tartrazina (AT) indo de 4 a 20 mg/L. Os níveis de concentração para o grupo de calibração se encontram na Tab. 5:

Tabela 5: Conjunto de calibração para o AT e AC.

Nível	AC	AT
	$C_0$ (mg/L)	$C_0$ (mg/L)
-2	4,00	4,00
-1	8,00	8,00
0	12,00	12,00
+1	16,00	16,00
+2	20,00	20,00

Para o conjunto de validação, foram utilizadas diversas misturas com concentrações na faixa de calibração e brancos, contendo apenas os reagentes de fenton. As concentrações de AT e AC no conjunto de validação podem ser encontradas na Tab. 6:

Tabela 6: Conjunto de validação para o AT e AC.

Amostra	AC	AT
	$C_0$ (mg/L)	$C_0$ (mg/L)
1	4,00	4,00
2	20,00	4,00
3	8,00	8,00
4	16,00	8,00
5	10,00	10,00
6	12,00	12,00
7	8,00	16,00
8	16,00	16,00
9	4,00	20,00
10	20,00	20,00
11	0,00	0,00
12	0,00	0,00
13	0,00	0,00

Para gerar a matriz de dados de cada amostra, foi utilizado um espectrofotômetro Agilent Cary 60 UV/Vis configurado para medir a absorção de 310 a 600 nm em intervalos regulares, totalizando 30 espectros ao decorrer de 350 segundos.

### Tratamento dos dados

Para o tratamento dos dados foi utilizado a toolbox MVC2, uma interface gráfica do *MATLAB* criada para simplificar o acesso a diversos métodos de calibração multivariada de ordem superior [52]. Através do MVC2, é possível acessar a *n-way Toolbox*, que contém algoritmos para a modelagem de dados multilineares, um desses modelos sendo o PARAFAC [34].

A *n-way Toolbox* permite a aplicação de três tipos de restrições, não-negatividade, ortogonalidade e unimodalidade com não-negatividade. Nesse trabalho, apenas as restrições de não-negatividade foram aplicadas.

### 6.1.4 Resultados e discussão

#### Espectros obtidos

A Fig. 18 mostra os espectros obtidos para cada corante (18A e B), para uma mistura (18C) e para os reagentes de fenton (18D). É possível perceber a degradação dos corantes e a formação de produtos reacionais com o passar do tempo, o que fica claro na 18D, com um sinal crescente por volta de 310 nm. Então é esperado um número de fatores necessários para a deconvolução maior do que 2 (um para cada corante e mais fatores para os produtos).

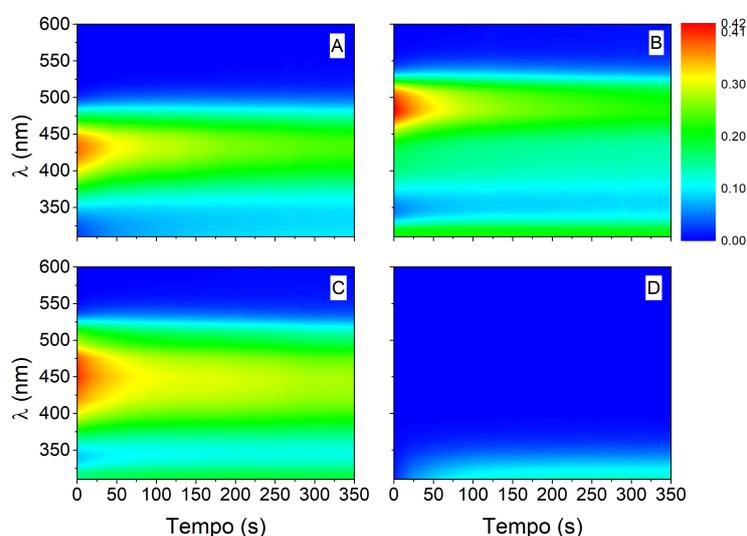


Figura 18: Espectros obtidos da degradação de corantes alimentícios com o uso de absorção UV/Vis e reagentes de Fenton. (A) AT 12 mg/L; (B) AC 12 mg/L; (C) AT e AC 8 mg/L; (D) Apenas os reagentes.

#### Resultados da deconvolução

Utilizando de restrições de não-negatividade nos três modos e uma deconvolução em 4 fatores, foi possível recuperar os perfis vistos na Fig. 19. Comparando os espectros puros dos corantes com os perfis cinéticos, podemos identificar o amarelo tartrazina como o fator 1 e o amarelo crepúsculo como o fator 3, enquanto o fator 4 aparenta estar relacionado aos produtos reacionais.

Os perfis espectrais encontrados na matriz **B** levam a algumas considerações. A primeira é que o perfil do amarelo crepúsculo parecer estar duplicado, aparecendo nos fatores 2 e 3, mas com perfis cinéticos distintos. Isso pode se dar devido a

uma série de razões diferentes, a mais simples sendo a existência de um produto reacional, evidenciado pela diminuição do fator 3 e aumento do fator 2 com o passar do tempo.

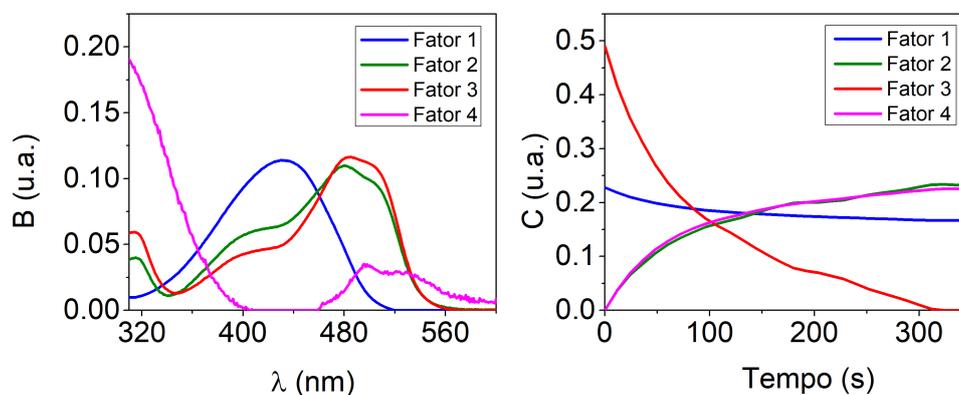


Figura 19: Resultado da deconvolução em 4 fatores para a degradação de AT e AC por reagentes de Fenton.

A coincidência dos perfis cinéticos do fator 2 e do fator 4 também fortalecem essa hipótese. Outro ponto é o perfil espectral do fator 4, visto na Fig. 19, que possui não apenas a banda vista na Fig. 18 (D) mas também uma banda secundária de 470 a 600 nm, o que pode ser interpretado como um produto de transferência de informação entre os fatores. Como o sistema estudado é um sistema cinético, é esperado que alguns fatores possuam certa colinearidade, o que pode resultar na perda parcial da unicidade da solução do PARAFAC, podendo até levar a formação de respostas degeneradas na deconvolução [53, 54], esse efeito também pode justificar a similaridade dos fatores 2 e 3.

É comum no uso do PARAFAC a aparição de fatores similares quando se utiliza um número de fatores maior do que o necessário, mas para uma comparação, temos na Fig 20 o resultado de uma deconvolução em 3 fatores:

Como pode-se ver na Fig. 20, o perfil mais similar ao amarelo crepúsculo mudou visivelmente de forma e o fator dos produtos ainda apresenta uma segunda banda, sugerindo que parte da informação que ele representa pode advir de outras fontes.

Esse fenômeno nesse tipo de sistema pode apresentar um obstáculo para a análise qualitativa dos dados deconvoluidos, no entanto, desde que os escores sejam internamente consistentes e ambos os analitos não estejam relacionados, a análise

quantitativa não deve ser fortemente afetada com o uso de um número adequado de fatores.

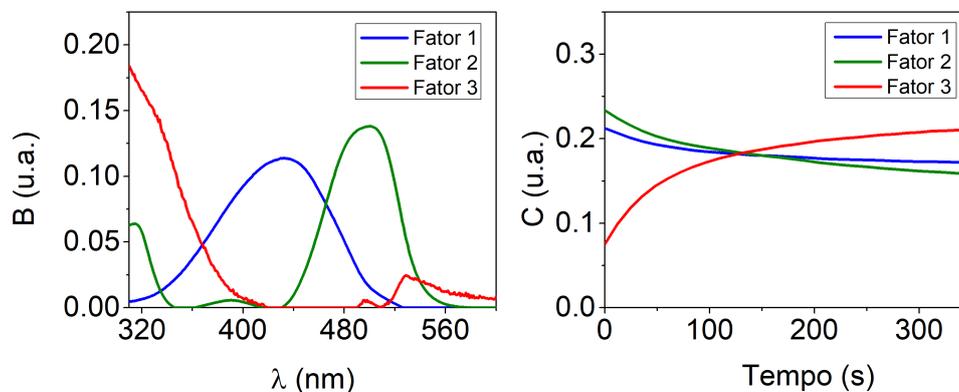


Figura 20: Resultado da deconvolução em 3 fatores para a degradação de AT e AC por reagentes de Fenton.

### Calibração pseudo-univariada

Utilizando os escores da matriz  $A$  obtidos na seção anterior, foi realizada uma calibração pseudo-univariada para recuperar as concentrações dos analitos no conjunto de validação. O resultado dessa calibração linear para ambos os pode ser encontrado na Fig. 21.

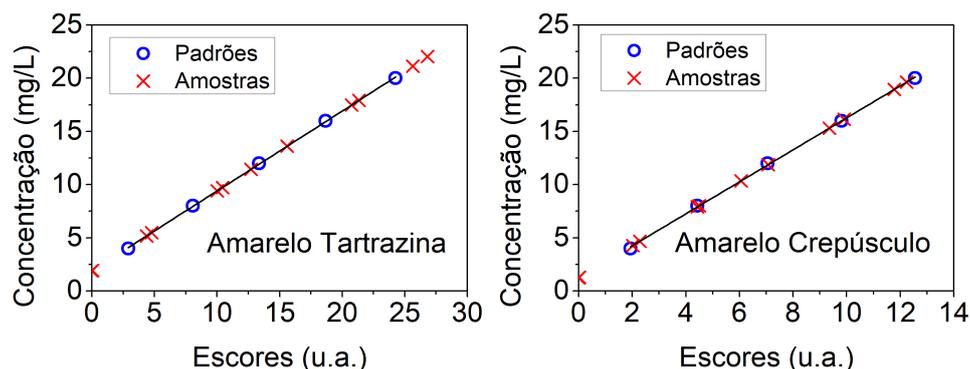


Figura 21: Resultado da calibração linear realizada a partir da deconvolução dos escores da matriz  $A$  em 4 fatores para o AT e AC.

A Tab. 7 contém os valores reais, preditos e os desvios das concentrações dos analitos no conjunto de validação. Usando os erros de previsão da Tab. 7 e a Eq.

(8), podemos calcular o RMSEV e o REP (RMSEV Proporcional) para a previsão de cada analito, obtendo os dados da Tab. 8.

Tabela 7: Resultados da calibração pseudo-univariada usando deconvolução em 4 fatores para o AT e AC.

<b>Amostra</b>	$C_{AT}$ mg/L	$C_{AT}^P$ mg/L	$E_{AT}$ mg/L	$C_{AC}$ mg/L	$C_{AC}^P$ mg/L	$E_{AC}$ mg/L
1	1,00	5,19	1,19	4,00	4,69	0,69
2	20,00	21,15	1,15	4,00	4,28	0,28
3	8,00	9,42	1,42	8,00	8,02	0,02
4	16,00	17,48	1,48	8,00	7,91	-0,09
5	10,00	11,45	1,45	10,00	10,35	0,35
6	12,00	13,62	1,62	12,00	11,88	-0,12
7	8,00	9,74	1,74	16,00	16,13	0,13
8	16,00	17,93	1,93	16,00	15,31	-0,69
9	4,00	5,48	1,48	20,00	19,64	-0,36
10	20,00	22,03	2,03	20,00	18,95	-1,05
11	0,00	1,91	1,91	0,00	1,25	1,25
12	0,00	1,9	1,9	0,00	1,27	1,27
13	0,00	1,93	1,93	0,00	1,29	1,29

Tabela 8: RMSEV e REP calculados para as concentrações preditas de AT e AC.

	$C_{AT}$	$C_{AC}$
<b>RMSEV (mg/L)</b>	1.656	0.748
<b>REP</b>	10.35%	4.68%

### 6.1.5 Aplicação em amostras reais

Uma vez que os analitos do sistema modelo são corantes alimentícios, o uso do método proposto para estimar as concentrações de amarelo tartrazina e amarelo crepúsculo em diferentes alimentos foi testado. Apesar dos resultados positivos com os corantes solubilizados em água, os testes em matrizes mais complexas como isotônicos e sucos se mostraram infrutíferas. Nas concentrações de reagentes

utilizadas no trabalho, a maior presença de compostos orgânicos, como açúcares ou adoçantes, na matriz inibiu a degradação dos corantes.

## 6.2 Aplicação - Sistema dopamina/aminocroma

Outra aplicação da modelagem de dados espectrocinéticos foi o processo de oxidativo da conversão de dopamina (CAS 51-61-6) em aminocroma.

Palladino *et al.* publicou um artigo discutindo como o ácido hipocloroso, formado em processos inflamatórios, pode afetar o estresse oxidativo neuronal ao produzir um excesso de dopamino-quinonas (DQ), também descrevendo a conversão dessas em aminocromas (AMC) [55]. Ao estudar esses sistemas, o autor criou condições que podem facilmente gerar dados espectrocinéticos usando espectrofotômetros UV-Vis, dado a velocidade e características cinéticas das reações e a presença de espectros em uma faixa utilizável.

### 6.2.1 Conversão de dopamina em aminocroma

Os processos envolvidos nessa série de reações foi descrita em detalhes por Palladino *et al.* [55], mostrando como o ácido hipocloroso pode formar dopamino-orto-quinona a partir de dopamina (DA) e a sua posterior conversão em leucoaminocroma (LC) e, por fim, aminocroma (AMC).

A oxidação de DA pode formar três isômeros estruturais de DQ em uma reação relativamente rápida. Dois desses isômeros passam então por um processo de ciclização gerando LC, que é rapidamente convertida em AMC.

As reações envolvidas nesse processo podem ser vistas na Fig. 22.

Segundo Palladino *et al.*, a conversão de DQ em AMC segue uma cinética de primeira-ordem, o que faz com que seja um bom candidato a utilização de métodos espectrocinéticos para modelagem com o PARAFAC.

Uma vez que as reações vão acontecer na cubeta, todos os reagentes devem ter seus valores ajustados para a própria adição garantir agitação o suficiente para homogeneizar a solução. Nesse caso, todos os reagentes foram adicionados em volumes acima de 500  $\mu\text{L}$ .

As reações mostradas na Fig. 22 podem ser iniciadas com adição de hipoclorito de sódio em uma solução contendo dopamina. No entanto, essas reações são consideravelmente sensíveis ao pH da solução, o que torna necessário a utilização de tampões fosfato que, pelo que a literatura indica, não inibem a reação de oxidação.

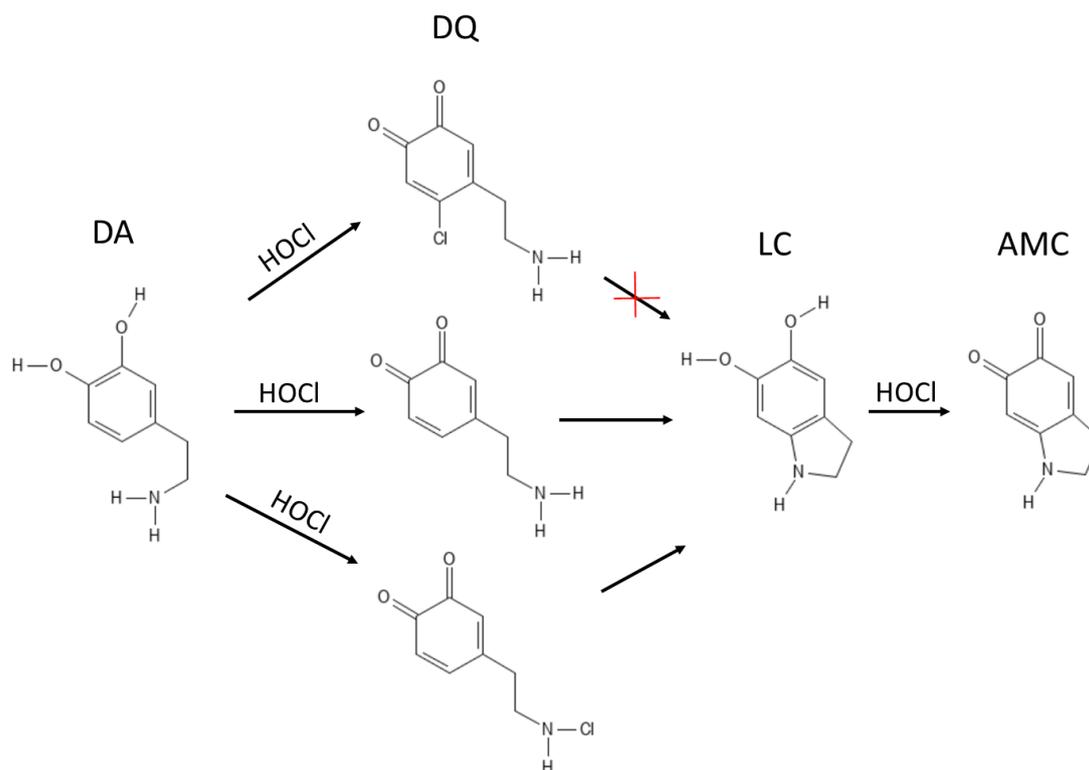


Figura 22: Esquemática das reações envolvidas na conversão de dopamina em aminocroma.

### Uso do pH para controlar a cinética

Em pHs ácidos, a conversão de DQ em AMC é relativamente lenta [55], mas pode ser consideravelmente acelerada em pHs neutros. Essa diferença de velocidades pode ser utilizada para acompanhar as reações do sistema separadamente, iniciando a reação em pH baixo, permitindo toda a conversão da DA em DQ sem ser afetada pela segunda reação e em seguida elevar o pH para a acompanhar a conversão de DQ em AMC.

### 6.2.2 Metodologia

Nesse trabalho dois tampões de fosfato foram utilizados para controlar o pH, com 1500  $\mu\text{L}$  de um tampão pH 2,71 a 0,5 mol/L adicionado antes dos reagentes e 550  $\mu\text{L}$  de um tampão de pH 7,2 a 1,0 mol/L sendo adicionado depois da conversão de DA em DQ.

Na faixa de concentração estudada (77 a 232 mg/L), a conversão de DA em DQ se deu como completa em pouco mais de 1 minuto após a adição da solução

de hipoclorito. O segundo tampão foi adicionado após dois minutos da adição de hipoclorito, permitindo acompanhar apenas a conversão de DQ em AMC. Medições da solução começaram 5 segundos após a adição do tampão neutro, medindo a absorvância de 300 a 500 nm, com passos de 1 nm, a cada 6 segundos até um tempo final de 240 segundos.

O procedimento experimental segue a esquematização mostrada na Fig. 23.

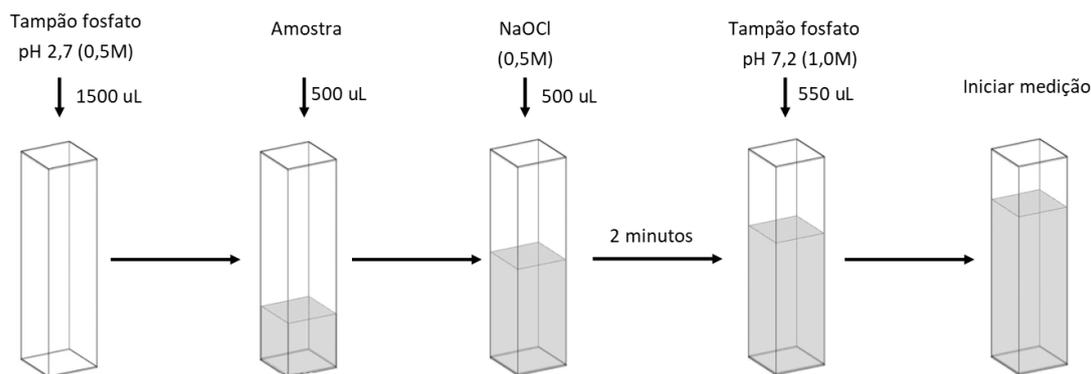


Figura 23: Esquematização do processo de preparação de amostra de dopamina para medição da matriz espectrocínética da conversão de DQ em AMC.

### Planejamento experimental

Para estudar o sistema, dois grupos de amostras foram construídos. Um conjunto de calibração contendo apenas dopamina pura e um grupo de validação, contendo misturas de dopamina e levodopa, que agirá como um interferente.

O conjunto de calibração para dopamina consistiu em uma curva em duplicata seguindo as concentrações da Tab. 9.

Tabela 9: Conjunto de calibração para quantificação de dopamina.

Nível	Dopamina $C_0$ (mg/L)
-2	77,2
-1	115,8
0	154,4
+1	193,0
+2	231,6

O conjunto de validação foi construído usando um planejamento fatorial  $4 \times 2$ , com quatro níveis de dopamina e dois de levodopa, usada para acrescentar um interferente no sistema. Levodopa possui uma estrutura similar a dopamina, reagindo nas mesmas condições, podendo até agir como precursor de dopamina e dopaminoquinonas. As concentrações dos níveis podem ser vistos na Tab. 10.

Tabela 10: Conjunto de validação do modelo para quantificação de dopamina com levodopa como interferente.

Nível	Dopamina	Levodopa
	$C_0$ (mg/L)	$C_0$ (mg/L)
-2	77,2	-
-1	115,8	77,6
+1	154,4	116,4
+2	193,0	-

### Tratamento dos dados

Assim como mostrado na seção 6.1.3, o MVC2 foi usado para modelar os dados. Foram criados modelos com 2 e 3 componentes, todos usando restrições de não-negatividade nos três modos e inicialização DTLTD. Apenas as medições de 0 a 120 segundos foram utilizadas nos modelos por apresentarem resultados mais consistentes e com menores erros médios.

### 6.2.3 Resultados e discussão

#### Espectros obtidos

A Fig. 24 contém exemplos das matrizes espectrocinéticas medidas após a adição do tampão neutro da solução, mostrando amostras contendo dopamina, levodopa e uma mistura das duas.

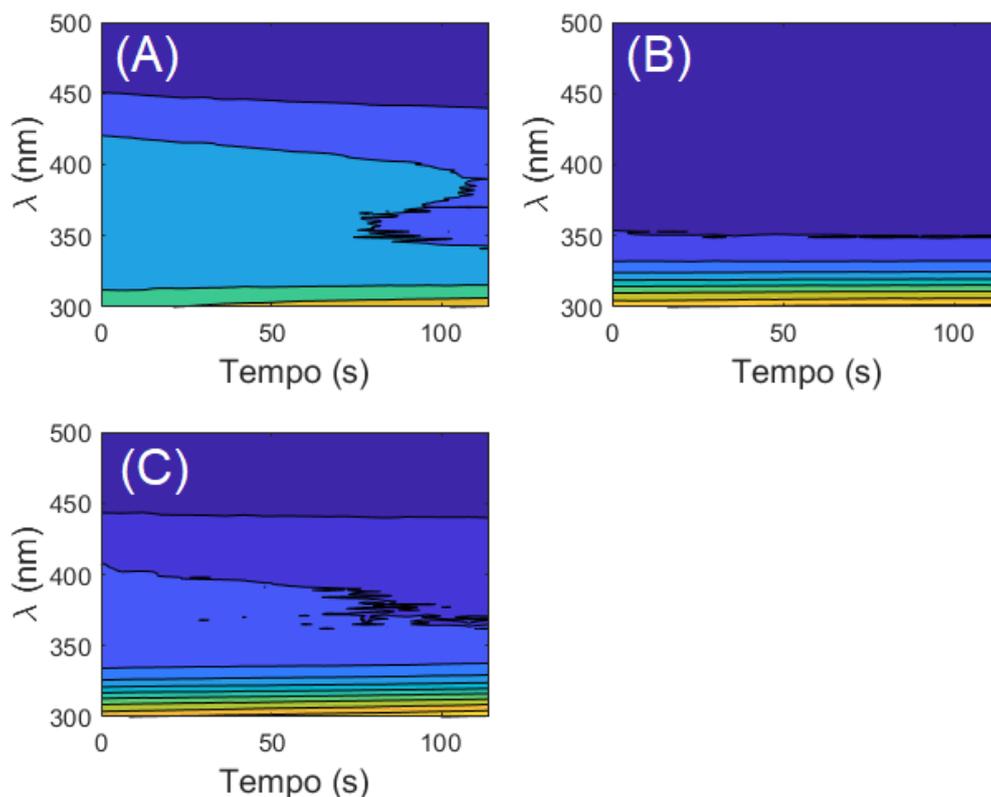


Figura 24: Espectros obtidos para a conversão de dopaminoquinonas em aminocroma, tendo levodopa adicionada como interferente, com o uso de absorção UV/Vis e hipoclorito de sódio. Amostras contendo: (A) dopamina 115,8 mg/L; (B) levodopa 116,4 mg/L; (C) mistura de dopamina e levodopa a 115,8 e 116,4 mg/L, respectivamente.

### Deconvolução

O modelo PARAFAC com dois componentes obteve um  $s_{fit}$  de 0,0041 e um  $CORCONDIA$  de 50,78%. O modelo com três componentes apresentou dificuldades em obter um resultado consistente, com o modelo ocasionalmente convergendo em pontos diferentes. Outra coisa a ser notada é a forte correlação entre os perfis cinéticos dos seus segundo e terceiro componentes, que pode ser a causa dos problemas de convergência. O resultado mostrado nesse trabalho para a deconvolução com três componentes apresentou  $s_{fit}$  de 0,0016 e um  $CORCONDIA$  de -3,59%.

### Perfis recuperados

Os perfis espectrais e cinéticos para ambas as deconvoluções podem ser encontrados na Fig. 25.

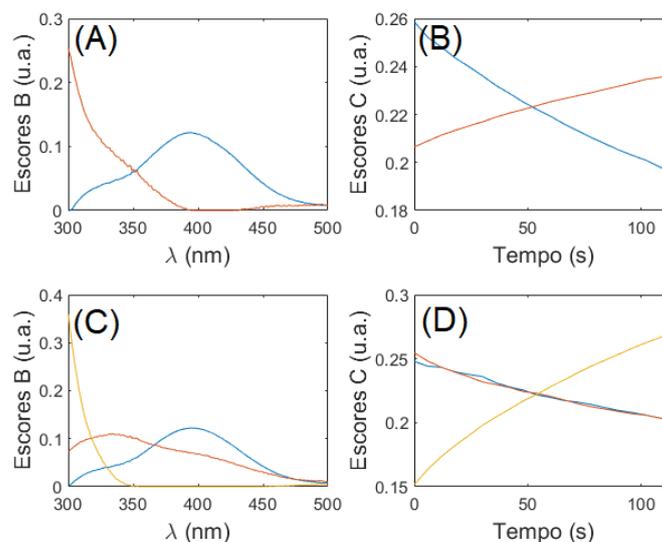


Figura 25: Perfis espectrais e cinéticos deconvoluidos usando PARAFAC para a conversão de dopaminoquinonas em aminocroma. (A) Perfis espectrais usando dois componentes; (B) Perfis cinéticos usando dois componentes; (C) Perfis espectrais usando três componentes; (D) Perfis cinéticos usando três componentes.

Ao observar os perfis cinéticos, podemos identificar o componente 1 (em azul) com as dopamino-o-quinonas (DQ). Com o aumento do número de componentes, o novo fator parece carregar informação presente nos dois anteriores, possuindo parte do perfil espectral do componente 2 e um perfil cinético praticamente coincidente com o componente 1, o que pode causar uma quebra de unicidade parcial (mantendo a solução única apenas para certos fatores e modos) da solução. Essa quebra pode justificar queda no valor do *CORCONDIA* e os problemas de convergência citados anteriormente.

### Calibração pseudo-univariada

As curvas de calibração pseudo-univariadas para o perfil da dopaminoquinona nos modelos PARAFAC com 2 e 3 componentes geradas podem ser vistas na Fig. 26.

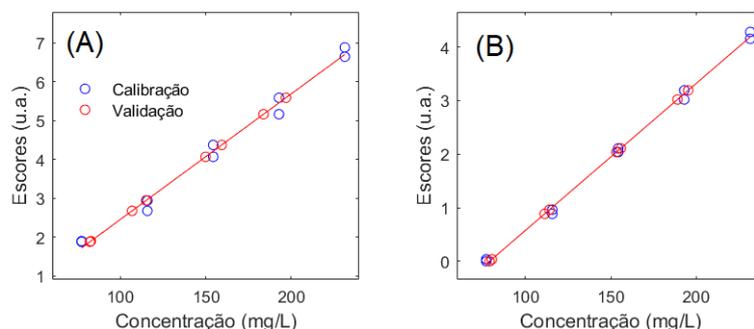


Figura 26: Curva de calibração para determinação de dopaminoquinona a partir dos escores do modo A do PARAFAC usando: (A) 2 componentes; (B) 3 componentes.

Ambas as curvas apresentaram resultados adequados, mas obtiveram menores erros médios no modelo com três componentes. Os valores de mérito para ambos os modelos pode ser visto na Tab. 11.

Tabela 11: RMSEV e REP da quantificação de dopaminoquinonas e  $s_{fit}$  e  $CORCONDIA$  calculados para o modelo PARAFAC usando dois e três componentes.

	2 Comp.	3 Comp.
<b>RMSEV</b>	5,9 ppm	2,8 ppm
<b>REP</b>	3,82%	1,79%
$s_{fit}$	0,0041	0,0016
<b><i>CORCONDIA</i></b>	50,78%	-3,59%

O modelo com 3 componentes é mais eficiente que o modelo com dois componentes e isso pode ser verificado em todos os valores de mérito com exceção do  $CORCONDIA$ . De fato, o  $CORCONDIA$  não mede a qualidade do modelo, mas sua consistência trilinear. O fato do analito e um outro fator serem praticamente indistintos em uma das vias, o perfil cinético, quebra essa consistência e isso deteriora o  $CORCONDIA$ , mesmo que a modelagem seja suficientemente bem sucedida para gerar resultados quantitativos.

### 6.3 Estudos simulados

Existem várias facetas da obtenção e uso de dados espectrocinéticos que não são necessariamente simples de se explorar experimentalmente. Em grande parte devido ao menor controle dos diversos parâmetros envolvidos em sistemas reais.

Para permitir maior liberdade na exploração desses sistemas, o simulador descrito no Apêndice B foi desenvolvido e utilizado para testar diferentes casos, permitindo explorar com total liberdade como o PARAFAC lida com cada um deles sem limitações experimentais.

#### 6.3.1 Reações do tipo $A \longrightarrow B$

Entre as reações mais comuns de 1ª ordem estão reações do tipo  $A \longrightarrow B$ , com um reagente sendo convertido em um produto.

O perfil cinético de ambos os componentes pode ser descrito a partir das Eqs.:

$$C_A(t) = C_{A0}e^{-kt} \quad (37)$$

$$C_B(t) = C_{A0}(1 - e^{-kt}) + C_{B0} \quad (38)$$

onde pode-se perceber que ambas podem ser divididas em um termo de  $C_{A0}$  e um outro dependente do tempo (considerando um valor baixo de  $C_{B0}$ ), sendo assim compatíveis com o modelo trilinear usado no PARAFAC.

No entanto, ao simular diferentes sistemas com esses componentes, pode-se perceber que os perfis espectrais e cinéticos apresentam claras distorções. O *CORCONDIA* também tendendo a valores baixos ou até mesmo negativos.

A Fig. 27 mostra diferentes resultados dessas simulações, geradas usando dois perfis unimodais com diferentes sobreposições espectrais.

Como visto na Fig. 27, apesar de seus perfis espectrais e cinéticos sofrerem distorções, os coeficientes do modo A do PARAFAC se manterem consistentes, com ambos podendo ser usados para a quantificação de  $C_{A0}$ . Esse resultado é uma consequência da alta colinearidade entre os dois componentes no tensor no sentido das amostras, resultando em uma perda parcial de unicidade da resposta do PARAFAC. Esse fenômeno é explorado mais a fundo na seção 7.

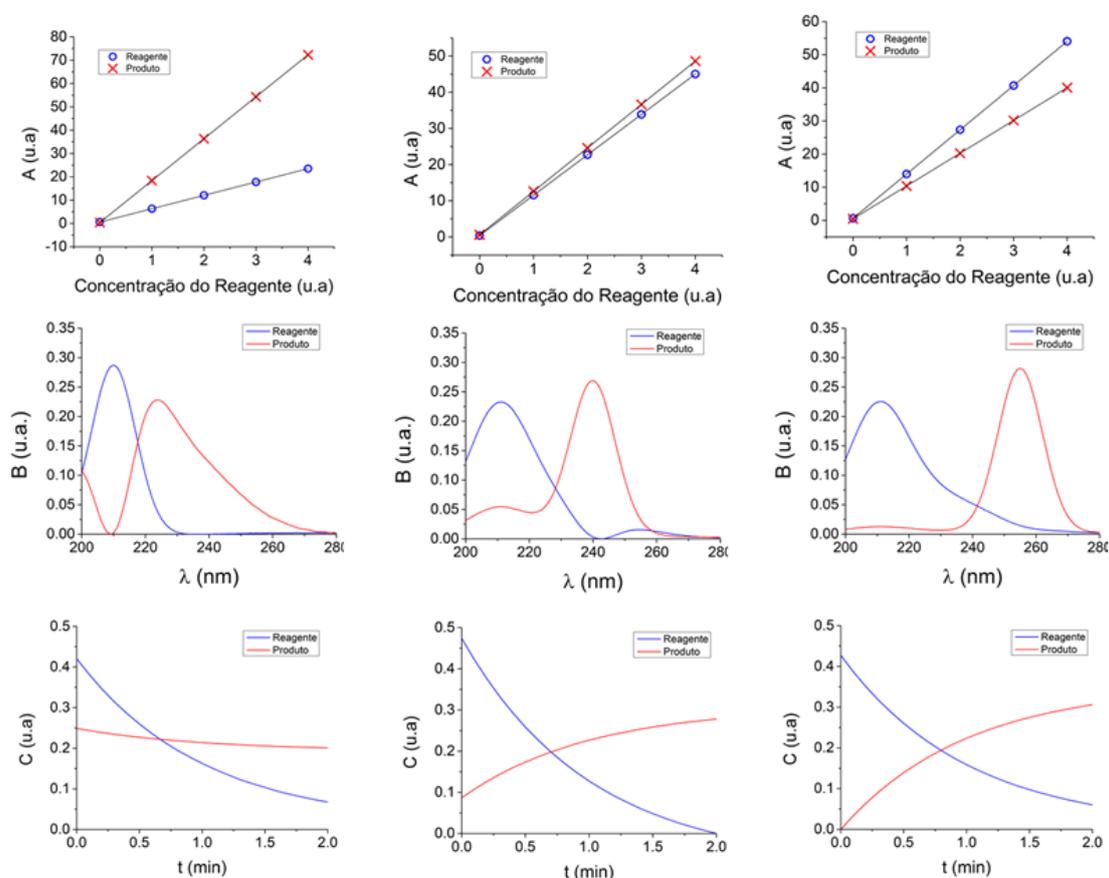


Figura 27: Três simulações de reações do tipo  $A \rightarrow B$  com diferentes níveis de sobreposição espectral. Cada coluna contém os três modos deconvoluidos pelo PARAFAC para cada simulação.

### 6.3.2 Efeito de desvios da cinética de 1ª ordem

#### Reações paralelas de 2ª ordem

Cinéticas de primeira e pseudo-primeira ordem são capazes de gerar sistemas trilineares, no entanto, em sistemas reais diversos fatores podem gerar desvios nas curvas cinéticas e afetar a modelagem do sistema. O objetivo dessa seção é estudar essas distorções e as consequências na deconvolução dos dados usando PARAFAC.

Um caso comum a ser avaliado é o de uma cinética do tipo  $A + B \rightarrow C$ , que quando  $C_B$  é consideravelmente maior que  $C_A$ , pode ser aproximada como uma cinética de primeira ordem dependente apenas de  $C_A$ . Quando o valor de  $C_A$  se aproxima de  $C_B$ ,  $C_B$  irá mudar de valor consideravelmente com o decorrer da reação, fazendo com que desvie de um perfil de primeira ordem.

Para estudar esses casos, um sistema com quatro componentes foi simulado com as seguintes reações paralelas:



com constantes de velocidade de 0,04 e 0,02, respectivamente e apenas A e B possuindo sinal.

Tensores com 25 amostras de dados espectrocinéticos foram gerados usando um planejamento de Brereton [35] em 5 níveis para os componentes A e B (concentrações com valores de 0, 1, 2, 3, e 4 u.a.), mantendo uma única concentração de C para todas as amostras e usando um tempo de reação de 50 minutos.

Simulações com diferentes concentrações de C, indo de  $C_{C0}$  igual a 50 u.a. até 5 u.a., foram feitas e seus resultados foram analisados. Valores abaixo de 5 u.a. reagindo com dois componentes, ou seja, com concentração na mesma escala de A e B], não serão discutidos nessa seção devido resultarem em um aumento da correlação dos valores deconvolvidos de concentração dos componentes A e B. Os efeitos causados por dois componentes terem um de seus modos correlacionados serão discutidos de maneira mais aprofundada na seção 7.

Nessa situação, o modelo do PARAFAC se mostrou relativamente robusto, no entanto foi possível perceber um aumento nos resíduos com a diminuição dos valores de  $C_{C0}$ . Leves desvios nos perfis espectrais e cinéticos foram percebidos, como mostrado na Fig. 28, com o PARAFAC modelando o perfil cinético como um compatível com um modelo trilinear, o distanciando do real. O perfil de concentrações nas amostras (modo A) foi o mais afetado, apresentando desvios negativos consideráveis em amostras com maior concentração, quando o sistema se distancia de uma cinética de 1ª ordem.

As simulações também mostraram que mesmos em casos mais extremos, com  $C_{C0}$  igual a 5 u.a., as distorções nos valores do modo A podem ser reduzidas ao reduzir o tempo de reação. A Fig. 29 mostra os resultados da deconvolução desse sistema com tempos de simulação de 50 (Fig. 29A) e 10 minutos (Fig. 29B).

Vale a pena mencionar que nesses casos, o valor do *CORCONDIA* permaneceu sempre acima de 95%, com os únicos indicativos do desvio sendo o aumento dos resíduos do modelo.

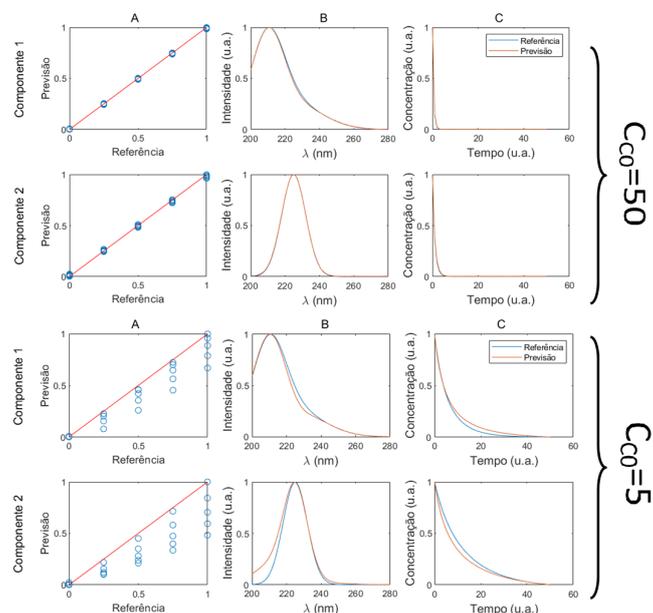


Figura 28: Resultado de duas simulações para dois componentes em um sistema no qual o ambos reagem com um terceiro de concentração inicial de 50 e 5 u.a.

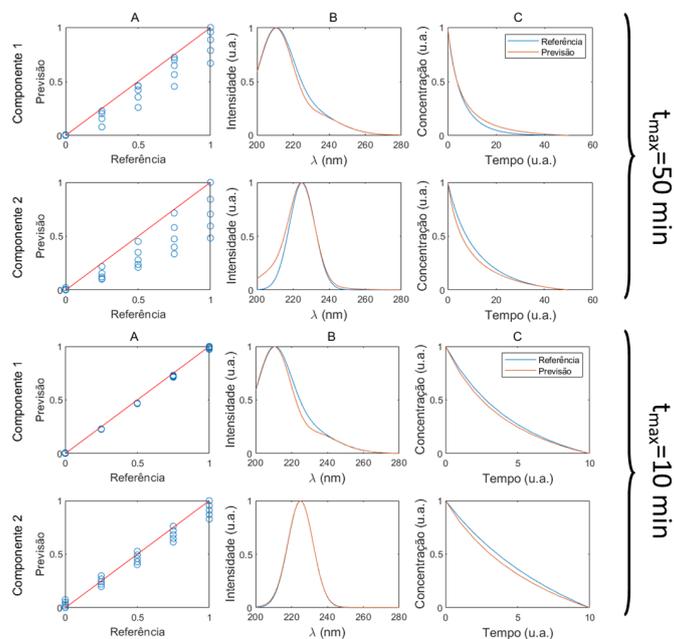
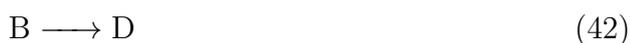


Figura 29: Perfis recuperados para dois componentes afetados por um reagente limitante com tempos de simulação de 50 e 10 minutos.

### Caso com reagentes limitantes

Um segundo sistema foi simulado em condições similares, mas com apenas a primeira reação sendo dependente da concentração de C. Portanto, seguindo as seguintes reações:



Nesse caso, o modelo se mostrou muito menos sensível a valores mais baixos de  $C_{C0}$ , com distorções significantes nos perfis do componente A, sendo mais próximos de uma relação quadrática que linear, apenas aparecendo com valores abaixo de  $C_{C0}$  abaixo de 5 u.a., como mostrado na Fig. 30. O componente B não teve sua deconvolução afetada e seus perfis foram recuperados em todos os casos.

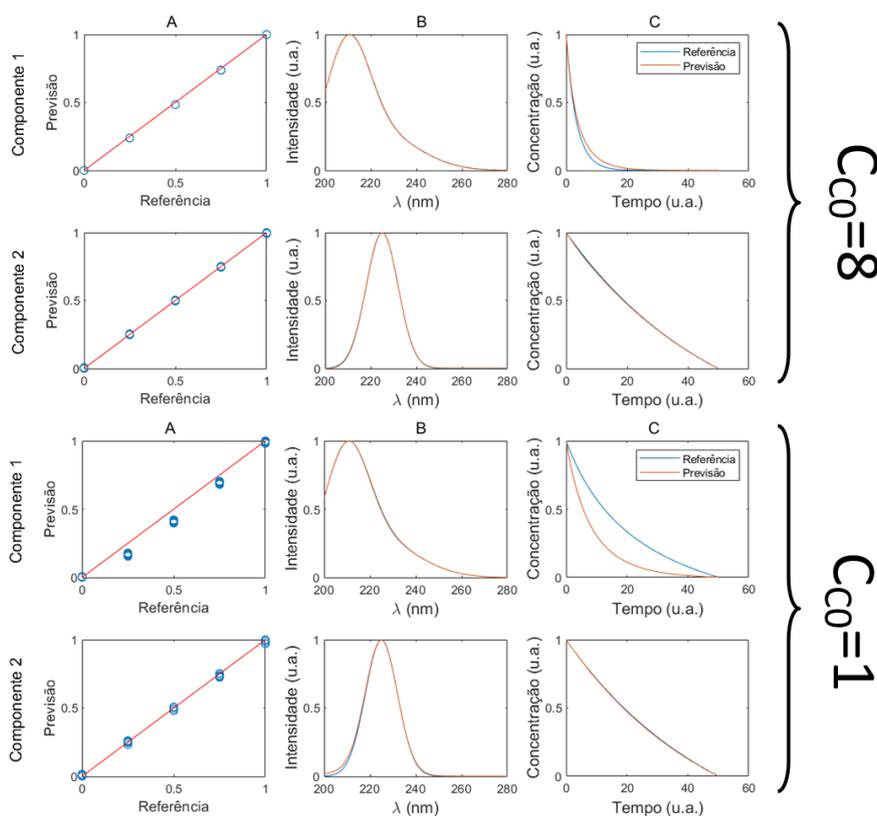


Figura 30: Resultado de duas simulações para dois componentes em um sistema no qual o 1º reage com um 3º e o 2º é independente. As simulações foram feitas com concentrações iniciais do 3º de 8 e 1 u.a.

No entanto, a perda de linearidade entre os valores do modo A e a concentração de referência quando  $C_{C0}$  possui um valor abaixo de 5 u.a. também pode ser mitigada ao limitar o tempo de reação. Limitar o tempo de reação permitiu reduzir consideravelmente as distorções no componente A, permitindo uma melhor recuperação das concentrações relativas das amostras e os perfis espectrais. Por exemplo, duas simulações com  $C_{C0}$  de 1 u.a. foram feitas com diferentes tempos de reação e seus resultados estão mostrados na Fig. 31.

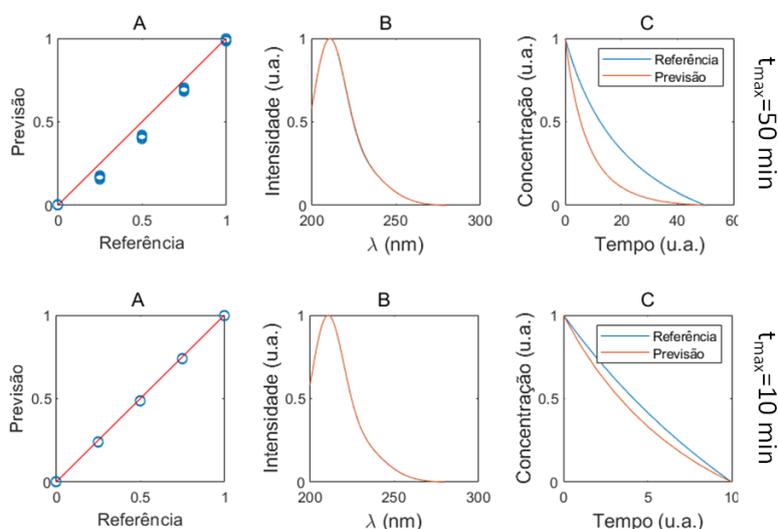


Figura 31: Perfis recuperados para o componente 1, afetado por um reagente limitante, com tempos de simulação de 50 e 10 minutos.

### Modelagem de interferentes com cinéticas incompatíveis

Em certos casos, é possível que apesar de os analitos de interesse seguirem uma cinética de 1<sup>a</sup> ordem, interferentes não o façam. A presença desses interferentes pode afetar a modelagem do sistema, logo é importante entender quais os efeitos desses desvios no analito de interesse.

Sistemas simulados contendo um componente seguindo uma cinética de 1<sup>a</sup> ordem e interferentes seguindo uma de 2<sup>a</sup> foram gerados para analisar esses casos, com dois exemplos podendo ser vistos na Fig. 32.

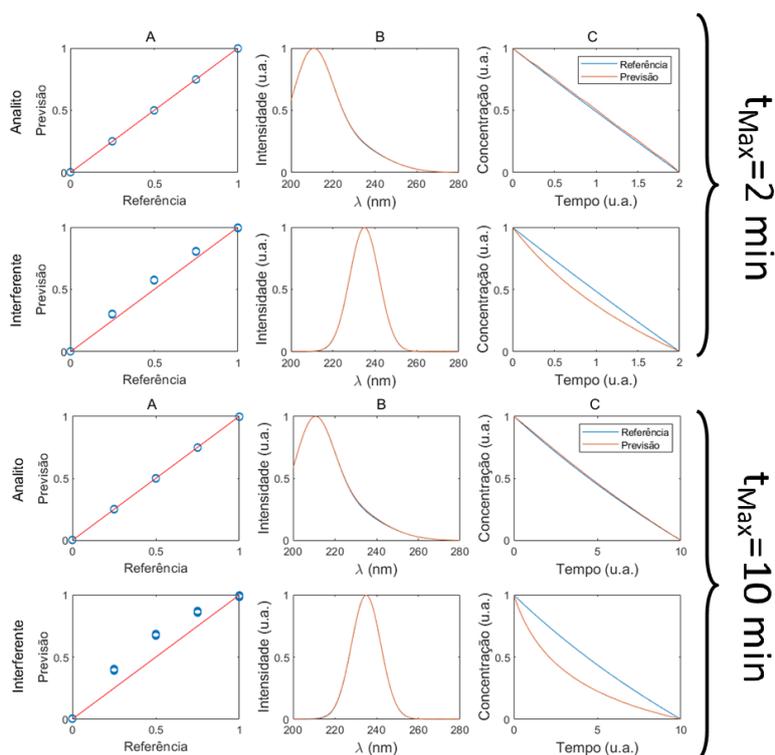


Figura 32: Perfis recuperados para um analito seguindo uma cinética de 1<sup>a</sup> ordem e um interferente seguindo uma de 2<sup>a</sup> ordem em dois tempos de simulação distintos.

Como visto na Fig. 32, o analito não parece ser afetado pelo interferente. O perfil espectral dos interferentes é recuperado sem distorções consideráveis, mas seus perfis de concentração e cinéticos são distorcidos. Dito isso, a relação entre perfil de concentrações e a concentração de referência depende fortemente do tempo de reação, podendo facilmente ser aproximada como linear em reações mais lentas em tempos curtos de medição, podendo assim ser usada para a criação de curvas de calibração.

## 6.4 Conclusão

Técnicas hífenadas como a combinação de medições espectroscópicas e reações químicas são capazes de gerar dados de segunda ordem, podendo até gerar tensores com características trilineares dado que as cinéticas envolvidas tenham perfis compatíveis, como no caso de cinéticas de primeira e pseudoprimeira-ordem.

O PARAFAC se mostrou relativamente robusto em casos que os perfis cinéticos se desviavam de um de primeira ordem, sendo capaz de gerar modelos aceitáveis

mesmo em situações limitantes, com ajustes no tempo de reação podendo facilitar a modelagem.

Componentes correlacionados em um dos modos, como reagentes e produtos ou reagentes com uma taxa de reação similar, podem causar problemas de convergência ou resultar em resultados degenerados com a perda de unicidade da solução do PARAFAC.

## 7 Colinearidade unimodal no PARAFAC

Em situações ideais, um modelo PARAFAC aplicado em um tensor multilinear deveria resultar em soluções essencialmente únicas (únicas desconsiderando permutações nos vetores decompostos e efeitos de escalamento), no entanto, dado certas condições alguns sistemas podem apresentar resultados que apesar de serem matematicamente válidos, apresentam deconvoluções com perfis distorcidos de dois ou mais componentes, resultantes de uma condição de unicidade parcial da solução [53].

Esse é um fenômeno conhecido desde as origens do método nos anos 70 [56], sendo normalmente associados a casos de *overfactoring* (aplicação de um número de fatores maior que o que a variância do sistema suportaria) e outras deficiências de variabilidade no tensor [53, 56]. Resultados parcialmente únicos podem levar a modelos de difícil análise, seja por distorções nos perfis qualitativos e quantitativos, por dificuldades na determinação do número de componentes ou pelo sistema simplesmente não convergir consistentemente em um único resultado. Harshman reportou que fatores com correlações de 0,75 ou maior já podiam afetar consideravelmente a convergência gerando platôs ao redor do mínimo do sistema e desacelerando a convergência do algoritmo em 4 a 10 vezes [56].

As razões para esse fenômeno, os efeitos nos modelos obtidos e algumas estratégias para aproximar um tensor obtendo unicidade parcial de um com solução essencialmente única serão exploradas nas seções a seguir.

### 7.1 Causas e consequências da unicidade parcial causada por colinearidade unimodal

Em sua forma mais simples, a perda parcial de unicidade dos resultados do PARAFAC pode ser interpretada como consequência da deficiência de posto parcial do tensor. De acordo com a prova fornecida pelo Dr. Robert Jennrich, publicada por Harshman [56], para um tensor de  $N$  fatores, cada uma das matrizes deconvoluídas deve possuir um posto igual ou maior que o número de fatores para o sistema possuir solução única, portanto, deficiências de posto nos modos podem levar a perda parcial unicidade.

Essas deficiências de posto são comumente associadas a falta de variação entre diferentes componentes no tocante as vias relacionadas aos sensores, no entanto

certos sistemas podem possuir características que fazem essa deficiência acontecer na via das amostras, as vezes de forma intrínseca ao método.

No caso específico de sistemas cinéticos esse fenômeno é facilmente expressado em sistemas com reações do tipo  $A \rightarrow B$ , que possuirão os vetores no modo A relativos aos componente A e B colineares, uma vez que ambos serão correlacionados a  $C_{A0}$  (concentração de A antes do início da reação). Outros casos podem ser vistos na análise exploratória de amostras com dois ou mais componentes com concentrações correlacionadas. Nas vias relativas aos sensores, diversos outros fatores podem gerar vetores correlacionados, como reações paralelas com leis de velocidade similares resultando em perfis correlacionados, por exemplo.

Apesar de o sistema possuir múltiplos resultados válidos, a maneira com a qual a decomposição se comporta segue certos padrões, dependendo fortemente de qual modo está correlacionado. Para avaliar as consequências da presença de componentes com fatores correlacionados nos diferentes modos, o simulador desenvolvido na seção B.1 foi utilizado para gerar dados de segunda ordem com resultados unicidade parcial advinda da correlação entre os modos de diferentes componentes.

Modelos absorção  $\times$  tempo foram simulados usando dois componentes distintos com um perfil de concentrações iniciais (modo A) sendo originado de um planejamento de Brereton [35] em 5 níveis, espectros de absorção distintos mas com certa sobreposição (modo B) e perfis cinéticos gerados com constantes de velocidade com valores de 1.0 e 0.6 (modo C). Para criar um sistema com resultados parcialmente únicos, um dos modos por vez tem o vetor do componente 2 igualado ao do componente 1. A Fig. 33 mostra os três modos para ambos os componentes.

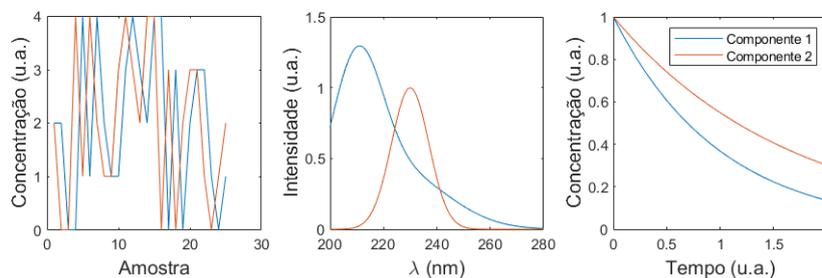


Figura 33: Perfis dos 3 modos para os dois componentes estudados.

Rodando o PARAFAC com os dois componentes com todos os modos linearmente independentes, é possível recuperar ambos os componentes sem desvios

consideráveis, como mostrado na Fig. 34. Esse modelo obteve *CORCONDIA* de 99,96% e  $s_{fit}$  de 0,0176.

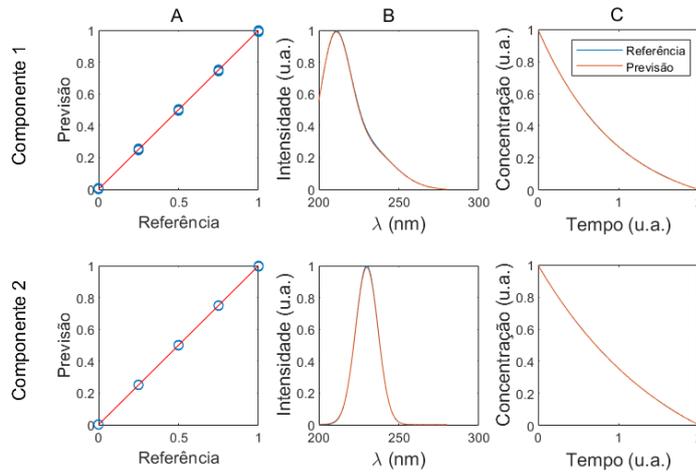


Figura 34: Exemplo da representação dos pontos de um espectro de referência com os preditos.

Ao usar o modelo em um tensor com o modo A altamente correlacionado, ao colocar as concentrações iniciais do segundo componente iguais as do primeiro, pode-se obter os resultados vistos na Fig. 35, com *CORCONDIA* de -14,72% e  $s_{fit}$  de 0,0174.

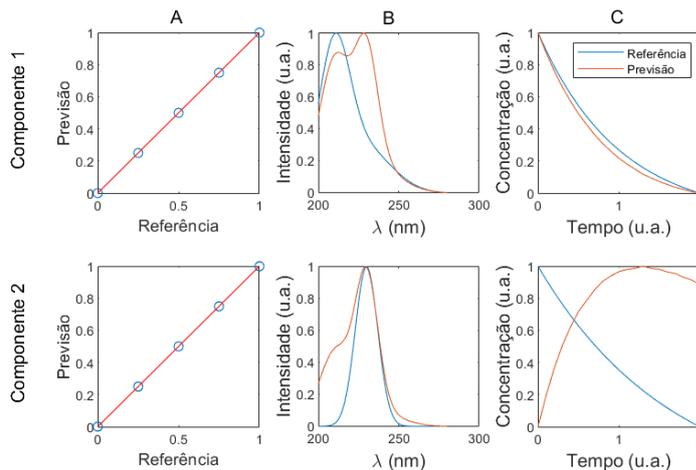


Figura 35: Comparação dos valores deconvoluidos com os valores de referência para os 3 modos em ambos os componentes tendo o modo A altamente correlacionado.

Como visto nas Figs. 35, a recuperação dos perfis no modo correlacionado não parece ser afetada, enquanto os outros podem sofrer distorções consideráveis. O mesmo se verifica em casos que a colineariedade se encontra nos modos B e C.

Ao usar o modelo em um tensor com o modo B altamente correlacionado, igualando o perfil espectral do segundo componente ao do primeiro, pode-se obter os resultados vistos na Fig. 36, com *CORCONDIA* de -6,25% e  $s_{fit}$  de 0,0165.

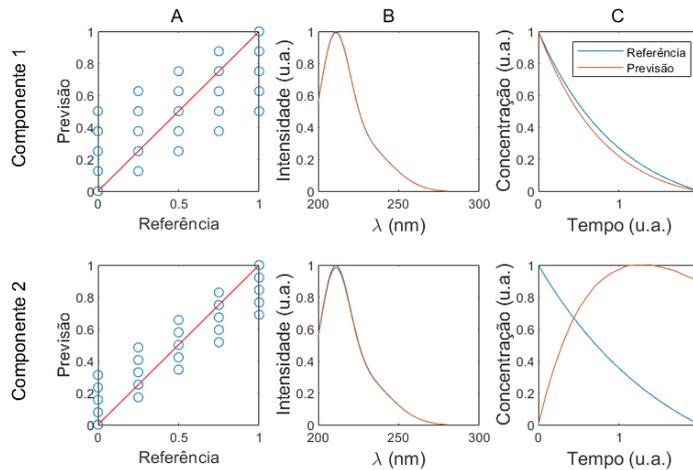


Figura 36: Comparação dos valores deconvoluidos com os valores de referência para os 3 modos em ambos os componentes tendo o modo B altamente correlacionado.

Ao usar o modelo em um tensor com o modo C altamente correlacionado, igualando a constante de velocidade do segundo componente iguais as do primeiro, pode-se obter os resultados vistos na Fig. 35, com *CORCONDIA* de 11,43% e  $s_{fit}$  de 0,0182.

É possível notar que o valor do *CORCONDIA* é fortemente afetado, podendo alcançar valores negativos. O valor de parâmetros como o  $s_{fit}$  não são muito úteis para identificar esses casos, uma vez que apesar das distorções, o modelo pode recuperar o mesmo tensor, com os resultados válidos possuindo um *fit* similar.

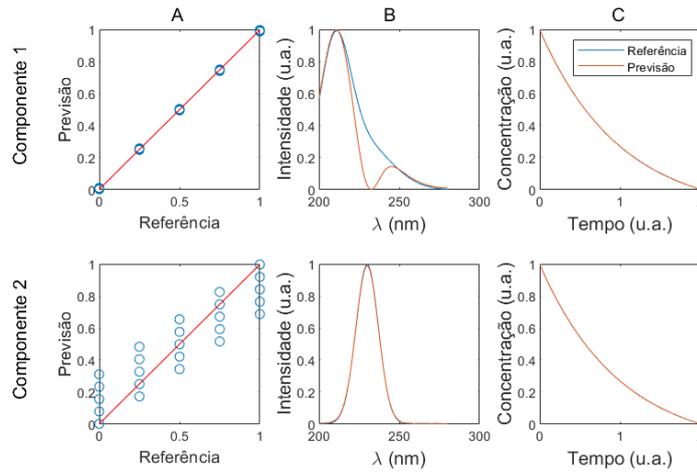


Figura 37: Comparação dos valores deconvolvidos com os valores de referência para os 3 modos em ambos os componentes tendo o modo C altamente correlacionado.

## 7.2 Efeito da presença de interferentes colineares

A fim de avaliar o efeito da colinearidade nos perfis de um terceiro composto não relacionado e linearmente independente, o tensor com o modo A dos componentes 1 e 2 colinear foi adicionado de um terceiro componente com um planejamento independente e uma constante de velocidade com valor 1,2. A deconvolução obteve *CORCONDIA* de 3,41% e  $s_{fit}$  de 0,0166, com seus resultados podendo ser vistos na Fig. 38.

Como visto na Fig. 38, apesar dos componentes 1 e 2 estarem com consideráveis distorções espectrais e cinéticas, essas não afetaram a deconvolução de um terceiro componente independente. Em um caso real em que esse terceiro componente seja o analito de interesse, a presença dos compostos com colinearidade pode não representar um problema no tocante aos resultados da deconvolução, mas pode afetar a determinação do número de componentes devido ao seu efeito em valores de mérito como o *CORCONDIA*.

Esse resultado é consistente com o descrito por Berge [53], apontando que o resultado do componente não-correlacionado nos três fatores é parcialmente único enquanto os outros são afetados independentemente.

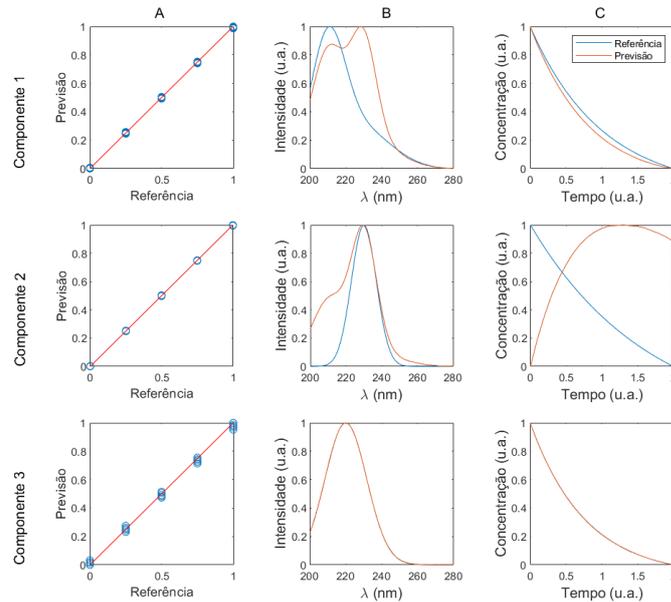


Figura 38: Comparação dos valores deconvolvidos para um sistema com dois componentes colineares no modo A e um terceiro linearmente independente nos três modos.

### 7.3 Estratégias para quebrar a colinearidade

Uma vez que a unicidade parcial das soluções do PARAFAC é originada na correlação entre diferentes componentes em um modo, estratégias para contornar o tipo de problema podem focar em pelo menos duas abordagens distintas: Utilizar de restrições matemáticas no modelo, como por exemplo *hard-modeling*<sup>4</sup> quando o sistema permite, ou tentar eliminar a correlação entre os vetores. Nesse trabalho, focaremos na segunda abordagem.

A principal vantagem de se tentar eliminar a correlação entre os vetores é que o algoritmo do PARAFAC não será forçado a um resultado específico, permitindo que ele se ajuste ao modelo independente de parâmetros fornecidos, que nem sempre podem se aplicar ao sistema sendo estudado. Claro, a maneira com a qual estratégias para diferenciar os vetores vão ser definidas podem depender fortemente de um entendimento mais aprofundado do sistema, as vezes com informação que pode não estar disponível.

<sup>4</sup>Restrições que se baseiam em restringir ou direcionar o formato de um dos modos para que esse seja compatível com um modelo químico predefinido.

Em seu formato mais tradicional, um tensor de segunda ordem usado no PARAFAC estará organizado de forma que a matriz do modo A contenha informação sobre a concentração dos componentes enquanto as matrizes dos modos B e C contém informação sobre os sensores. Dado a diferente natureza desses modos, abordagens para diferenciar os vetores das matrizes serão naturalmente distintos.

Os modos B e C normalmente dependem fortemente das condições na qual as amostras estão sendo medidas, com as possibilidades de possíveis estratégias sendo fortemente dependentes do método. Por exemplo, métodos espectroscópicos podem ter perfis espectrais dependentes do pH, em casos com alta correlação espectral, uma mudança de pH pode diferenciar os espectros. Em sistemas cinéticos o mesmo se aplica, incluindo mudanças de temperatura para diferenciar as constantes de velocidade, presença de diferentes catalisadores, entre outros.

O modo A por outro lado pode ser o mais facilmente manipulado experimentalmente em certos sistemas. Em sistemas com componentes conhecidos, a adição de uma matriz adicional no tensor com adição de um padrão ou com o padrão puro de um dos compostos correlacionados, como mostrado na Fig. 39 já pode ser o bastante para quebrar a colinearidade entre os vetores.

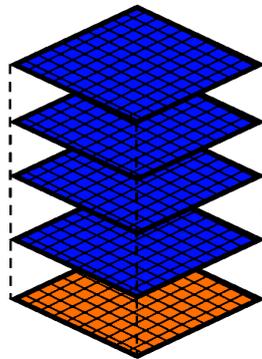


Figura 39: Representação da adição de uma matriz (em laranja) ao tensor de amostras original contendo uma das amostras com adição de padrão para quebrar a colinearidade no modo A.

Como exemplo desse tipo de estratégia, usaremos o mesmo tensor do exemplo mostrado na Fig. 35, mas uma matriz adicional será colocada contendo uma cópia da primeira amostra somada de 20% da sua quantidade original de componente 2. Os resultados dessa deconvolução podem ser vistos na Fig. 40.

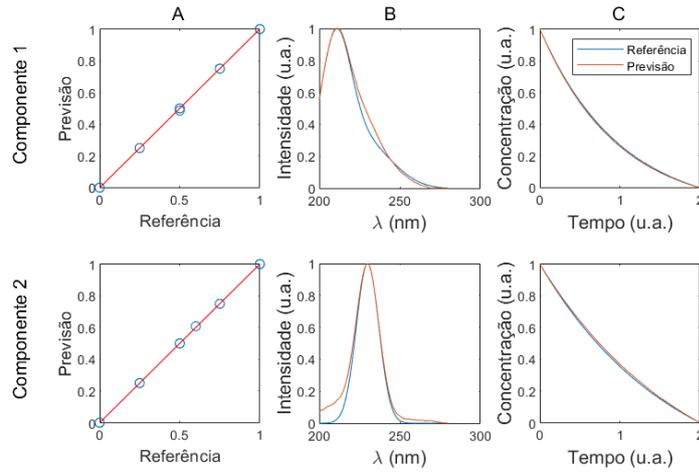


Figura 40: Comparação dos valores deconvoluidos com os valores de referência para os 3 modos em ambos os componentes tendo o modo A do tensor original altamente correlacionado, mas com uma amostra adicional que passou por adição de padrão.

É possível ver que apesar de certa distorção tanto espectral quanto cinética ainda existir, os resultados são bem mais próximos dos esperados. O valor do *CORCONDIA* mudou de -14,62% para 99,17%, indicando maior consistência trilinear. O  $s_{fit}$ , como esperado, não mudou consideravelmente, indo de 0,0174 para 0,0179.

Vale a pena mencionar que essa estratégia não precisa ser necessariamente aditiva. Por exemplo, o uso de adsorventes em quantidades limitadas que sejam capazes de adsorver diferentemente os componentes de uma amostra em um pré-tratamento podem diferenciá-los o bastante em uma única amostra para quebrar a colinearidade do vetor, permitindo que esse método seja usado em amostras com componentes desconhecidos.

Em sistemas cinéticos de primeira ordem, como os estudados na seção 6, algumas propriedades da lei de velocidade podem ser exploradas para realizar essa diferenciação sem a necessidade da medição de amostras adicionais.

A lei de velocidade de primeira ordem para um reagente tem a forma da Eq. (43):

$$C_A(t) = C_{A0}e^{-kt} \quad (43)$$

Considerando que a parte da equação relevante para o modo C é apenas a exponencial, sabemos que o perfil cinético normalizado será consistente independente da concentração inicial. Uma das consequências dessa propriedade é que o perfil normalizado também será consistente independente do tempo inicial, uma vez que passado certo tempo e realizando a medida com as mesmas condições, a única diferença será a concentração inicial dos componentes, ou seja, um efeito similar ao obtido com adição de padrão.

Portanto, uma matriz contendo dados de segunda ordem envolvendo cinéticas de primeira ordem podem ser reorganizadas de diferentes formas para quebrar a colinearidade no modo A. Uma das possíveis maneiras de fazer isso é selecionar um certo número de colunas no sentido do tempo para formar o corpo principal do tensor e adicionar a cópia de uma amostra representativa com o mesmo número de colunas, mas com a posição selecionada sendo deslocada, como mostrado na Fig. 41(A). Uma outra abordagem, que pode ser aplicada em uma única amostra é a mostrada na Fig. 41(B), onde uma única amostra é copiada ao longo do tensor com diferentes deslocamentos temporais.

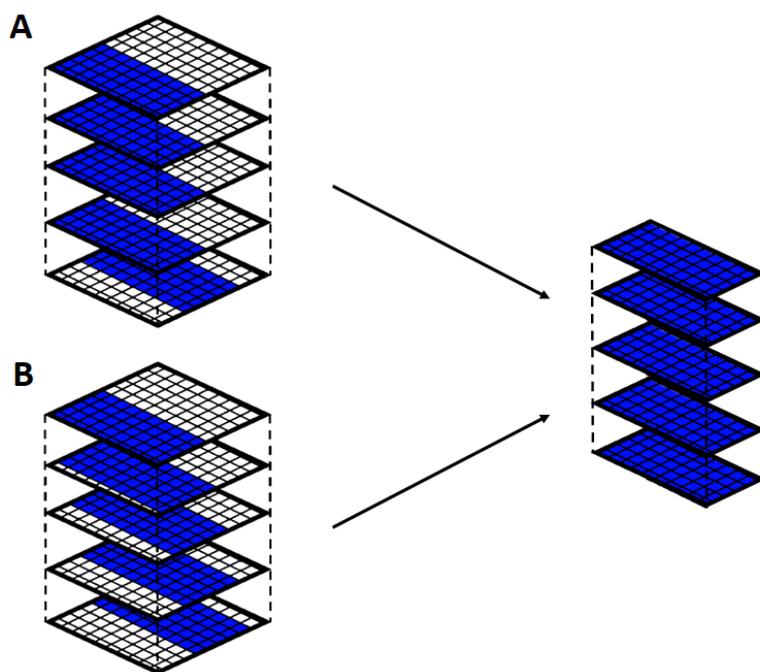


Figura 41: Diferentes estratégias de implementação do deslocamento temporal ao longo do tensor. As células destacadas (em azul) no tensor original sendo selecionadas para o tensor final.

Para mostrar a utilidade desse tipo de estratégia em amostras desconhecidas, uma única amostra foi simulada usando os componentes 1 e 2 descritos acima, com ambos se degradando com constantes de velocidade distintas. A estratégia mostrada na Fig. 41(B) foi usada nessa única matriz, com ela sendo copiada sete vezes com diferentes seleções de colunas. A matriz gerada teve 21 colunas no sentido do tempo, portanto, a primeira cópia selecionou as colunas de 1 a 15, a segunda de 2 a 16 e assim em diante. A deconvolução por PARAFAC foi realizada obtendo *CORCONDIA* de 99,96% e  $s_{fit}$  0,0140. Os espectros originais e recuperados podem ser vistos na Fig. 42.

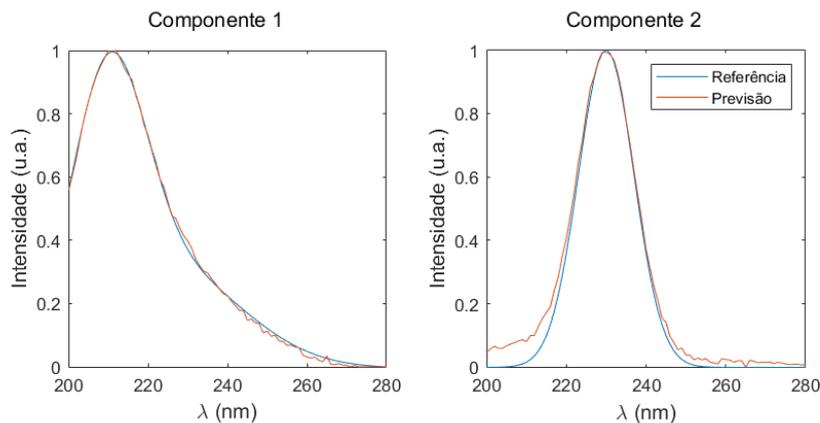


Figura 42: Perfis espectrais recuperados para ambos os componentes da amostra.

Como pode ser visto na Fig. 42, apesar de uma única amostra ter sido utilizada, a estratégia de deslocamentos sucessivos permitiu que o modelo do PARAFAC fosse capaz de deconvoluir o tensor, mantendo sua consistência trilinear, e recuperar os perfis espectrais, o que em tese permitiria a análise qualitativa dos componentes de uma amostra completamente desconhecida.

Claro, a aplicação desse tipo de estratégia depende fortemente da consistência do perfil cinético das amostras, com desvios nesses perfis podendo não se ajustar ao modelo.

Se tratando de reações de primeira ordem do tipo  $A \longrightarrow B$  com ambos tendo sinal, apesar desse método poder ser utilizado, ele pode ser mais limitado devido ao formato do perfil cinético do produto. Esse perfil segue o formato visto na Eq. (44)

$$C_B(t) = C_{A0}(1 - e^{-kt}) + C_{B0} \quad (44)$$

O valor de  $C_{B0}$  na Eq. (44) pode representar uma quebra na trilinearidade do tensor em diferentes níveis dependendo do seu valor. Dito isso, em um tensor de dados com esse sistema, o modelo do PARAFAC pode ser capaz de aproximar o perfil do produto B a um perfil compatível, com o seu modo A sendo colinear ao modo A do reagente. Ao usar a estratégia de deslocamento temporal descrita acima pode-se manipular o valor de  $C_{B0}$  em amostras adicionais para quebrar essa colinearidade.

Usando o componente 1, mostrado na Fig. 33, como reagente e colocando o componente 2 como produto dele (e tendo concentração inicial igual a zero), pode-se ver uma significativa distorção nos espectros de ambos os componentes, como mostrado na Fig. 43 obtendo um *CORCONDIA* de 11,58% e  $s_{fit}$  de 0,0175.

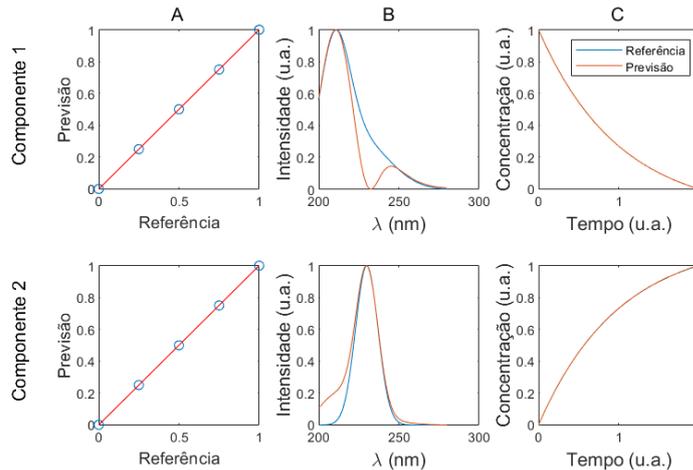


Figura 43: Comparação dos valores deconvolvidos com os valores de referência para os 3 modos em ambos os componentes tendo o modo A do tensor original altamente correlacionado devido ao componente 2 ser produto do consumo do componente 1.

Aplicando a estratégia mostrada na Fig. 41(A), com uma única amostra sendo copiada e deslocada no tempo, o modelo foi capaz de reduzir consideravelmente a distorção no componente 2, mas afetando pouco o componente 1, como visto na Fig. 44. O valor do *CORCONDIA* subiu para 86,90%, mas o  $s_{fit}$  também se elevou, tendo um valor de 0,0262, o que pode se dar devido ao leve desvio do modelo trilinear causado pela constante na cinética do produto, com indicações disso no plot dos resíduos do modelo, que apresentarão valores maiores e estruturados na matriz adicional como mostrado na Fig. 45.

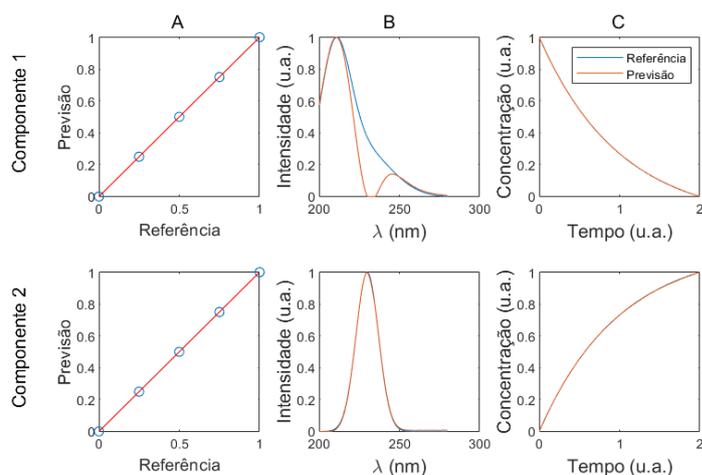


Figura 44: Comparação dos valores deconvolvidos com os valores de referência para os 3 modos em ambos os componentes tendo o modo A do tensor original altamente correlacionado, mas com uma amostra adicional sendo deslocada no tempo.

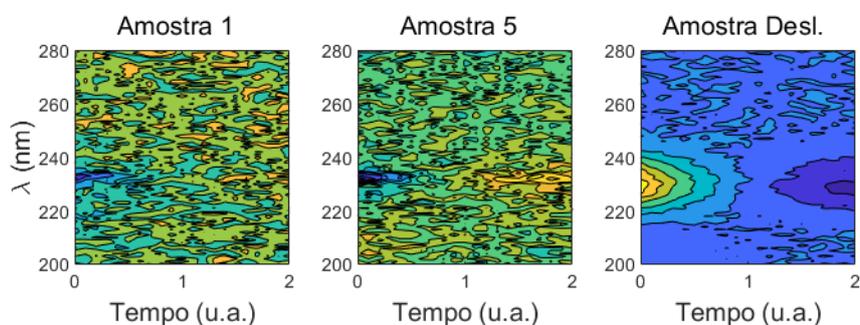


Figura 45: Resíduos do modelo PARAFAC para diferentes amostras do conjunto principal e para a amostra adicionada (deslocada no tempo).

Para um exemplo real da aplicação dessa estratégia, podemos utilizar a curva de calibração usada na seção 6.2, na qual *dopamina-o-quinona* se converte em *aminocroma* seguindo uma cinética de pseudo-primeira ordem[55].

Aplicando o PARAFAC com dois componentes e não-negatividade nos três modos na curva de calibração, se obtém dois perfis que aparentam estar levemente distorcidos, obtendo um *CORCONDIA* de 48,74% e  $s_{fit}$  0,0042. Copiando uma das amostras no tensor, selecionando 20 das 40 colunas e realizando a deconvolução do modelo com diferentes deslocamentos para a amostra copiada, podemos obter os

resultados vistos na Fig. 46, obtendo *CORCONDIA*s de 50,34% sem deslocamento até 93,78% deslocando 20 colunas. O valor do  $s_{fit}$  aumenta levemente com o deslocamento, mas ainda permanece consideravelmente baixo comparado com o sinal das amostras (ordem de  $10^{-3}$  comparado a  $10^{-1}$ ).

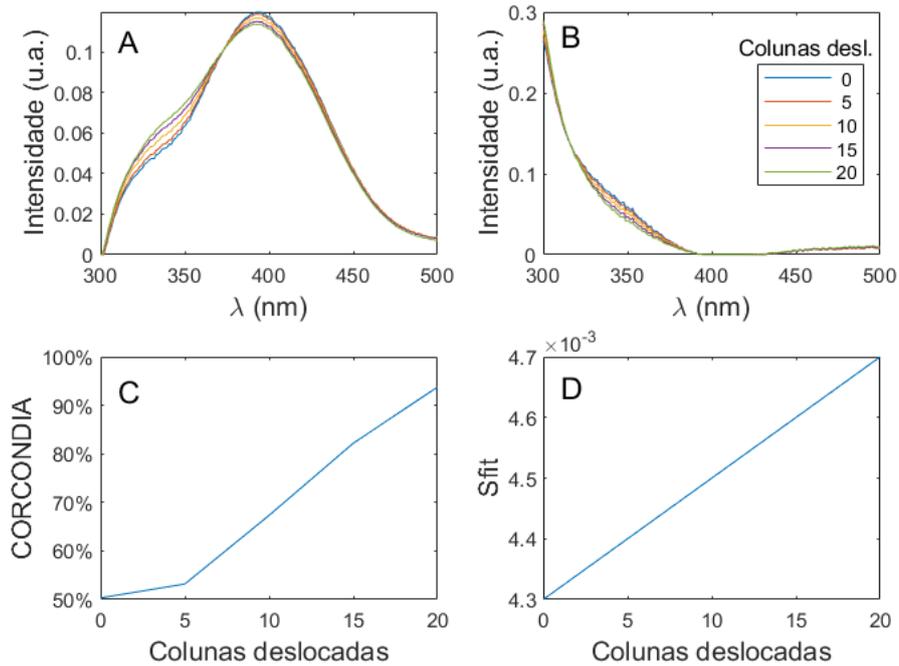


Figura 46: Resultados da deconvolução com a adição de uma única amostra copiada com diferentes deslocamentos temporais. (A) Perfil recuperado do reagente, (B) Perfil recuperado do produto, (C) Valores do *CORCONDIA* para os diferentes deslocamentos da amostra copiada, (D) Valores de  $s_{fit}$  para os diferentes deslocamentos.

Apesar de esse exemplo não possuir uma distorção considerável usando o tensor original, a melhora no valor do *CORCONDIA* é considerável com a aplicação da amostra deslocada, mantendo valores similares de  $s_{fit}$ , o que pode indicar um modelo que melhor reflita os perfis originais dos componentes.

## 7.4 Conclusão

A presença de componentes em um tensor trilinear com um de seus modos altamente correlacionado pode dificultar a convergência do sistema, levando a resultados parcialmente únicos. O modo colinear pode ser recuperado consistente-

mente, mas os demais podem apresentar distorções consideráveis. Componentes não-relacionados não parecem ter sua deconvolução afetada.

As simulações realizadas nessa seção mostraram que diferentes estratégias podem ser usadas para quebrar essa colinearidade dependendo do tipo de amostra e de técnica. Componentes correlacionados no modo de concentrações podem quebrar a colinearidade com a adição de padrão em uma amostra adicional. No caso de dados espectrocinéticos, assumindo reações de primeira-ordem, o uso de matrizes adicionais deslocadas no tempo pode ajudar a quebrar a colinearidade do modo de concentrações e corrigir parte dos problemas causados pela mesma.

## 8 Conclusão

Os estudos realizados nesse trabalho permitiram explorar variadas facetas da aplicação de modelos PARAFAC, usando suas características intrínsecas e implementações para resolver problemas analíticos e explorar diversas propriedades de diferentes sistemas.

O uso de fluorescência sincrônica ajudou a contornar limitações do sistema e o desenvolvimento de um algoritmo de conversão para EEMs, usado graças a implementação de algoritmos de imputação singular, permitiu a deconvolução dos dados e o estudo de uma região espectral ampla.

A vantagem de segunda ordem inerente ao modelo permitiu seu uso como clean-up virtual, permitindo a simplificação de um processo de tratamento de amostra.

O estudo de sistemas espectrocinéticos mostrou a relativa robustez de modelos PARAFAC em sistemas espectrais com resolução temporal, mesmo em casos que os perfis cinéticos se desviavam de um de primeira ordem. O PARAFAC foi capaz de gerar modelos aceitáveis em situações limitantes. A alta colinearidade dos componentes em sistemas do tipo  $A \longrightarrow B$  também direcionou o estudo para o desenvolvimento de métodos para contorná-la e recuperar a unicidade da resposta do PARAFAC, entre as quais métodos de adição de padrão e deslocamento temporal se mostram promissoras.

## 9 **Perspectivas futuras**

Para estudos futuros, uma análise mais aprofundada do uso de reações Fenton e Fenton-like pode possibilitar a criação de um método generalizado para a análise de amostras complexas a partir da degradação oxidativa de seus componentes. Tendo assim o potencial de poder ser aplicado a toda uma miríade de sistemas, seja para analisar qualitativa quanto quantitativa de seus componentes. Outro sistema de interesse é a análise de sistemas espectroeletróquímicos com sinal fluorescente. Essa técnica possui características similares a sistemas cinéticos, incluindo uma tendência a gerar sistemas com alta colinearidade em casos que ambas as formas oxidada e reduzida de um componente possuam sinal, o que pode permitir o aprofundamento dos estudos realizados durante esse trabalho.

## Referências

- [1] M. M. Sena, M. G. Trevisan, and R. J. Poppi, “PARAFAC: uma ferramenta quimiométrica para tratamento de dados multidimensionais. aplicações na determinação direta de fármacos em plasma humano por espectrofluorimetria,” *Química Nova*, vol. 28, pp. 910–920, Oct. 2005.
- [2] R. Bro, “Parafac. tutorial and applications,” *Chemometrics and Intelligent Laboratory Systems*, vol. 38, no. 2, pp. 149 – 171, 1997.
- [3] J. V. dos Santos, L. G. Fregolente, A. B. Moreira, O. P. Ferreira, S. Mounier, B. Viguier, H. Hajjoul, and M. C. Bisinoti, “Humic-like acids from hydrochars: Study of the metal complexation properties compared with humic acids from anthropogenic soils using PARAFAC and time-resolved fluorescence,” *Science of The Total Environment*, vol. 722, p. 137815, June 2020.
- [4] C. Qian, L.-F. Wang, W. Chen, Y.-S. Wang, X.-Y. Liu, H. Jiang, and H.-Q. Yu, “Fluorescence approach for the determination of fluorescent dissolved organic matter,” *Analytical Chemistry*, vol. 89, pp. 4264–4271, Mar. 2017.
- [5] W. G. Mendoza, D. D. Riemer, and R. G. Zika, “Application of fluorescence and PARAFAC to assess vertical distribution of subsurface hydrocarbons and dispersant during the deepwater horizon oil spill,” *Environmental Science: Processes & Impacts*, vol. 15, no. 5, p. 1017, 2013.
- [6] J. H. Christensen, G. Tomasi, J. Strand, and O. Andersen, “PARAFAC modeling of fluorescence excitation-emission spectra of fish bile for rapid en route screening of PAC exposure,” *Environmental Science & Technology*, vol. 43, pp. 4439–4445, May 2009.
- [7] M. Cabrera-Bañegil, D. Martín-Vertedor, E. Boselli, and I. Durán-Merás, “Control of olive cultivar irrigation by front-face fluorescence excitation-emission matrices in combination with PARAFAC,” *Journal of Food Composition and Analysis*, vol. 69, pp. 189–196, 2018.
- [8] T.-Q. Peng, X.-L. Yin, W. Sun, B. Ding, L.-A. Ma, and H.-W. Gu, “Developing an Excitation-Emission Matrix Fluorescence Spectroscopy Method Coupled with Multi-way Classification Algorithms for the Identification of

- the Adulteration of Shanxi Aged Vinegars,” *Food Analytical Methods*, vol. 12, pp. 2306–2313, jul 2019.
- [9] R. Ríos-Reina, S. M. Azcarate, J. M. Camiña, and R. M. Callejón, “Sensory and spectroscopic characterization of Argentinean wine and balsamic vinegars: A comparative study with European vinegars,” *Food Chemistry*, vol. 323, p. 126791, sep 2020.
- [10] L. Rubio, S. Sanlloriente, L. A. Sarabia, and M. C. Ortiz, “Determination of cochineal and erythrosine in cherries in syrup in the presence of quenching effect by means of excitation-emission fluorescence data and three-way PARAFAC decomposition,” *Talanta*, vol. 196, pp. 153–162, 2019.
- [11] L. Rubio, S. Sanlloriente, L. A. Sarabia, and M. C. Ortiz, “Fluorescence determination of cochineal in strawberry jam in the presence of carmoisine as a quencher by means of four-way PARAFAC decomposition,” *Food Chemistry*, vol. 290, pp. 178–186, 2019.
- [12] M. G. Trevisan and R. J. Poppi, “Determination of doxorubicin in human plasma by excitation–emission matrix fluorescence and multi-way analysis,” *Analytica Chimica Acta*, vol. 493, pp. 69–81, Sept. 2003.
- [13] L.-Q. Hu, C.-L. Yin, Y.-H. Du, and Z.-P. Zeng, “Simultaneous and direct determination of vancomycin and cephalixin in human plasma by using HPLC-DAD coupled with second-order calibration algorithms,” *Journal of Analytical Methods in Chemistry*, vol. 2012, pp. 1–8, 2012.
- [14] F. Samari, B. Hemmateenejad, and M. Shamsipur, “Spectrophotometric determination of carminic acid in human plasma and fruit juices by second order calibration of the absorbance spectra–pH data matrices coupled with standard addition method,” *Analytica Chimica Acta*, vol. 667, pp. 49–56, May 2010.
- [15] A. Gardlo, A. K. Smilde, K. Hron, M. Hrdá, R. Karlíková, D. Friedecký, and T. Adam, “Normalization techniques for PARAFAC modeling of urine metabolomic data,” *Metabolomics*, vol. 12, June 2016.
- [16] A. C. Olivieri and G. M. Escandar, “Chapter 3 - experimental three-way/second-order data,” in *Practical Three-Way Calibration* (A. C. Olivieri and G. M. Escandar, eds.), pp. 27–45, Boston: Elsevier, 2014.

- [17] G. Tomasi and R. Bro, "PARAFAC and missing values," *Chemometrics and Intelligent Laboratory Systems*, vol. 75, pp. 163–180, Feb. 2005.
- [18] A. C. Olivieri and G. M. Escandar, "Chapter 1 - calibration scenarios," in *Practical Three-Way Calibration* (A. C. Olivieri and G. M. Escandar, eds.), pp. 1–9, Boston: Elsevier, 2014.
- [19] A. C. Olivieri and G. M. Escandar, "Chapter 5 - parallel factor analysis: Trilinear data," in *Practical Three-Way Calibration* (A. C. Olivieri and G. M. Escandar, eds.), pp. 65–92, Boston: Elsevier, 2014.
- [20] A. Smilde, R. Bro, and P. Geladi, "Array definitions and properties," in *Multi-Way Analysis with Applications in the Chemical Sciences*, ch. 2, pp. 13–34, John Wiley & Sons, Ltd, 2004.
- [21] R. Bro and H. A. L. Kiers, "A new efficient method for determining the number of components in PARAFAC models," *Journal of Chemometrics*, vol. 17, no. 5, pp. 274–286, 2003.
- [22] P. Valderrama, P. Março, N. Locquet, F. Ammari, and D. Rutledge, "A procedure to facilitate the choice of the number of factors in multi-way data analysis applied to the natural samples: Application to monitoring the thermal degradation of oils using front-face fluorescence spectroscopy," *Chemometrics and Intelligent Laboratory Systems*, vol. 106, pp. 166–172, Apr. 2011.
- [23] K. C. Araújo, M. C. Barreto, A. S. Siqueira, A. C. P. Freitas, L. G. Oliveira, M. E. P. Bastos, M. E. P. Rocha, L. A. Silva, and W. D. Fragoso, "Oil spill in northeastern brazil: Application of fluorescence spectroscopy and PARAFAC in the analysis of oil-related compounds," *Chemosphere*, vol. 267, p. 129154, Mar. 2021.
- [24] D. A. Armbruster and T. Pry, "Limit of blank, limit of detection and limit of quantitation.," *The Clinical biochemist. Reviews*, vol. 29 Suppl 1, no. August, pp. S49–52, 2008.
- [25] A. C. Olivieri and G. M. Escandar, "Chapter 6 - analytical figures of merit," in *Practical Three-Way Calibration* (A. C. Olivieri and G. M. Escandar, eds.), pp. 93–107, Boston: Elsevier, 2014.

- [26] M. G. Trevisan, *Aplicação de metodos quimiometricos de ordem superior e fluorescencia molecular na analise em matrizes biologicas*. 81f. Dissertação (mestrado), Universidade Estadual de Campinas, Instituto de Quimica, 2003.
- [27] W. D. Fragoso and A. C. Olivieri, “Chemometric modeling of kinetic-fluorescent third-order data for thiamine determination in multivitamin complexes,” *Microchemical Journal*, vol. 128, pp. 42–46, Sept. 2016.
- [28] M. D. Carabajal, J. A. Arancibia, and G. M. Escandar, “Excitation-emission fluorescence-kinetic data obtained by fenton degradation. determination of heavy-polycyclic aromatic hydrocarbons by four-way parallel factor analysis,” *Talanta*, vol. 165, pp. 52–63, Apr. 2017.
- [29] D. Skoog, *Principles of instrumental analysis*. Philadelphia: Saunders College, 1980.
- [30] A. V. Schenone, M. J. Culzoni, A. D. Campiglia, and H. C. Goicoechea, “Total synchronous fluorescence spectroscopic data modeled with first- and second-order algorithms for the determination of doxorubicin in human plasma,” *Analytical and Bioanalytical Chemistry*, vol. 405, pp. 8515–8523, Aug. 2013.
- [31] A. V. Schenone, A. de Araújo Gomes, M. J. Culzoni, A. D. Campiglia, M. C. U. de Araújo, and H. C. Goicoechea, “Modeling nonbilinear total synchronous fluorescence data matrices with a novel adapted partial least squares method,” *Analytica Chimica Acta*, vol. 859, pp. 20–28, Feb. 2015.
- [32] K. Kumar and A. K. Mishra, “Parallel factor (PARAFAC) analysis on total synchronous fluorescence spectroscopy (TSFS) data sets in excitation–emission matrix fluorescence (EEMF) layout: Certain practical aspects,” *Chemometrics and Intelligent Laboratory Systems*, vol. 147, pp. 121–130, Oct. 2015.
- [33] L. Tan, W. Du, Y. Zhang, L.-J. Tang, J.-H. Jiang, and R.-Q. Yu, “Rayleigh scattering correction for fluorescence spectroscopy analysis,” *Chemometrics and Intelligent Laboratory Systems*, vol. 203, p. 104028, Aug. 2020.
- [34] C. A. Andersson and R. Bro, “The n-way toolbox for MATLAB,” *Chemometrics and Intelligent Laboratory Systems*, vol. 52, pp. 1–4, Aug. 2000.

- [35] R. Brereton, *Chemometrics : data analysis for the laboratory and chemical plant*. Chichester, West Sussex, England Hoboken, NJ: Wiley, 2003.
- [36] C. M.-E. Gossner, J. Schlundt, P. B. Embarek, S. Hird, D. Lo-Fo-Wong, J. J. O. Beltran, K. N. Teoh, and A. Tritscher, “The Melamine Incident: Implications for International Food and Feed Safety,” *Environmental Health Perspectives*, vol. 117, pp. 1803–1808, dec 2009.
- [37] J.-Y. Yhee, C. A. Brown, C.-H. Yu, J.-H. Kim, R. Poppenga, and J.-H. Sur, “Retrospective Study of Melamine/Cyanuric Acid-Induced Renal Failure in Dogs in Korea between 2003 and 2004,” *Veterinary Pathology*, vol. 46, pp. 348–354, mar 2009.
- [38] V. d. L. M. Finete, M. M. Gouvêa, F. F. de Carvalho Marques, and A. D. P. Netto, “Validation of a method of high performance liquid chromatography with fluorescence detection for melamine determination in UHT whole bovine milk,” *Food Control*, vol. 51, pp. 402–407, may 2015.
- [39] R. Yang, R. Liu, and K. Xu, “Detection of adulterated milk using two-dimensional correlation spectroscopy combined with multi-way partial least squares,” *Food Bioscience*, vol. 2, pp. 61–67, 2013.
- [40] M. K. Nieuwoudt, S. E. Holroyd, C. M. McGoverin, M. C. Simpson, and D. E. Williams, “Raman spectroscopy as an effective screening method for detecting adulteration of milk with small nitrogen-rich molecules and sucrose,” *Journal of Dairy Science*, vol. 99, no. 4, pp. 2520–2536, 2016.
- [41] M. K. Nieuwoudt, S. E. Holroyd, C. M. McGoverin, M. C. Simpson, and D. E. Williams, “Rapid, sensitive, and reproducible screening of liquid milk for adulterants using a portable Raman spectrometer and a simple, optimized sample well,” *Journal of Dairy Science*, vol. 99, no. 10, pp. 7821–7831, 2016.
- [42] R. Z. A. Bakain, Y. S. A. Degs, A. H. E. Sheikh, and S. H. Arar, “Spectrophotometric Determination of Melamine in Liquid Milk by Multivariate Second Order Calibration,” *Current Analytical Chemistry*, vol. 12, no. 1, pp. 74–84, 2015.
- [43] Y. Liu, J. Deng, L. An, J. Liang, F. Chen, and H. Wang, “Spectrophotometric determination of melamine in milk by rank annihilation factor analysis based

- on pH gradual change-UV spectral data,” *Food Chemistry*, vol. 126, no. 2, pp. 745–750, 2011.
- [44] M. Ritota and P. Manzi, “Melamine Detection in Milk and Dairy Products: Traditional Analytical Methods and Recent Developments,” *Food Analytical Methods*, vol. 11, no. 1, pp. 128–147, 2018.
- [45] F. World Health Organization and A. O. of the United Nations, “Proposed Draft Maximum Levels for Melamine in Food and Feed (N13-2009),” *Uniform Procedure for the Elaboration of Codex Standards and Related Texts*, no. March, p. 26, 2010.
- [46] M. C. Ortiz, S. Sanllorente, A. Herrero, C. Reguera, L. Rubio, M. L. Oca, L. Valverde-Som, M. M. Arce, M. S. Sánchez, and L. A. Sarabia, “Three-way PARAFAC decomposition of chromatographic data for the unequivocal identification and quantification of compounds in a regulatory framework,” *Chemometrics and Intelligent Laboratory Systems*, vol. 200, no. December 2019, 2020.
- [47] H. Liu, W. Liu, D.-H. Han, and S.-P. Wang, “Three-Dimensional Fluorescence Fingerprint Technique for Milk Quality Evaluation: Antibiotic Residual Detection and Heat-Treated Evaluation,” *Guang Pu Xue Yu Guang Pu Fen Xi/Spectroscopy and Spectral Analysis*, vol. 38, pp. 1633–1639, 2018.
- [48] J. L. Maubois and D. Lorient, “Dairy proteins and soy proteins in infant foods nitrogen-to-protein conversion factors,” *Dairy Science and Technology*, vol. 96, no. 1, pp. 15–25, 2016.
- [49] M. C. Barreto, R. G. Braga, S. G. Lemos, and W. D. Fragoso, “Determination of melamine in milk by fluorescence spectroscopy and second-order calibration,” *Food Chemistry*, vol. 364, p. 130407, Dec. 2021.
- [50] M. Fioreze, E. P. dos Santos, and N. Schmachtenberg, “Processos oxidativos avançados: Fundamentos e aplicação ambiental,” *Revista Eletrônica em Gestão, Educação e Tecnologia Ambiental*, vol. 18, Apr. 2014.
- [51] S. Wang, “A comparative study of fenton and fenton-like reaction kinetics in decolourisation of wastewater,” *Dyes and Pigments*, vol. 76, pp. 714–720, Jan. 2008.

- [52] A. C. Olivieri, H.-L. Wu, and R.-Q. Yu, “MVC2: A MATLAB graphical interface toolbox for second-order multivariate calibration,” *Chemometrics and Intelligent Laboratory Systems*, vol. 96, pp. 246–251, Apr. 2009.
- [53] J. M. F. ten Berge, “Partial uniqueness in CANDECOMP/PARAFAC,” *Journal of Chemometrics*, vol. 18, pp. 12–16, Jan. 2004.
- [54] J. B. Kruskal, R. A. Harshman, and M. E. Lundy, *How 3-MFA Data Can Cause Degenerate Parafac Solutions, among Other Relationships*, p. 115–122. NLD: North-Holland Publishing Co., 1989.
- [55] P. Palladino, F. Torrini, S. Scarano, and M. Minunni, “Colorimetric analysis of the early oxidation of dopamine by hypochlorous acid as preliminary screening tool for chemical determinants of neuronal oxidative stress,” *Journal of Pharmaceutical and Biomedical Analysis*, vol. 179, p. 113016, Feb. 2020.
- [56] R. Harshman, “Foundations of the PARAFAC procedure: Models and conditions for an ”explanatory” multi-model factor analysis,” UCLA Working Papers Phonetics, 1970.
- [57] A. Stegeman and N. D. Sidiropoulos, “On kruskal’s uniqueness condition for the candecomp/parafac decomposition,” *Linear Algebra and its Applications*, vol. 420, pp. 540–552, Jan. 2007.
- [58] G. Stocco, M. Pazzola, M. L. Dettori, P. Paschino, G. Bittante, and G. M. Vacca, “Effect of composition on coagulation, curd firming, and syneresis of goat milk,” *Journal of Dairy Science*, vol. 101, no. 11, pp. 9693–9702, 2018.

## Apêndice A - Script para conversão de SFMs em EEMs

```
1 function X=universalsync(X_s,d_exc,d_l)
2 % Universal synchronous fluorescence converter
3 %
4 % Converts a synchronous fluorescence matrix into an
   excitation-emission matrix.
5 %
6 %Inputs:
7 % X_s: Synchronous fluorescence matrix with the
   excitation wavelength changing rowise.
8 % d_exc: Excitation step.
9 % d_l: Offset step.
10 %
11 %Outputs:
12 % X: Converted excitation-emission matrix.
13
14 if nargin==1
15     d_exc=1;
16     d_l=1;
17 end
18
19 %% EEM format and conversion parameters
20 [l,c]=size(X_s);
21
22 % Determining the values of m_min and n_min
23 [m_min,n_min]=rat(d_l/d_exc);
24
25 %Determining the number of vertical groups in the
   converted matrix
26 groups=floor(l/m_min);
27
28 %% Empty spaces
29 if l~=(groups*m_min)
```

```

30     X=NaN(1,(1-m_min*((groups+1)-1))+m_min*(c-1)+m_min*
        n_min*((groups+1)-1));
31 else
32     X=NaN(1,(1-m_min*(groups-1))+m_min*(c-1)+m_min*n_min*(
        groups-1));
33 end
34
35 %% EEM conversion
36 i=0;
37 for k=1:groups+1 % The +1 accounts for the rest of the
        lines not filling an entire group.
38     if k<=groups
39         counter=m_min*k; % The last line of the current
                vertical group
40     else
41         counter=1; % The final line
42     end
43     for i=(i+1):counter
44         for j=1:c
45             X(i,(i-m_min*(k-1))+m_min*(j-1)+m_min*n_min*(k-1))
                =X_s(i,j);
46         end
47     end
48 end
49
50 %% Reordering
51 if m_min~=1&& n_min~=1
52     [~,ce]=size(X);
53     ng=ceil(ce/m_min);
54     s_subm=[0:n_min:n_min*(m_min-1)];
55     s_total= repmat(s_subm,1,ng)+repelem((0:m_min:(ng-1)*
        m_min),m_min);
56     X_seq=sortrows([s_total(1,1:ce); X] ',1)';
57     X=X_seq(2:l+1,:);
58 end

```

## Apêndice B - Simulador de dados espectrocinéticos

### B.1 Construção do simulador

Para estudar os diferentes fatores que afetam sistemas cinéticos, foi necessário criar um simulador que pudesse gerar dados de ordem superior seguindo uma série de parâmetros.

Para esse estudo, matrizes com um componente espectral e um temporal foram escolhidos como sistemas padrão (similar a sistemas analisando o espectro de absorção em função do tempo), com os dados de cada um desses componentes sendo gerados separadamente.

Os espectros simulados foram gerados a partir da soma de curvas gaussianas adaptadas para melhor manipular seu formato, como mostrado na Eq. (45):

$$S = \sum_{i=1}^n \left[ I(i) \times \exp \left( \frac{(R - C(i))^2 \ln(F)}{W(i)^2} \right) \right] \quad (45)$$

Na qual  $I$  representa a intensidade máxima de cada curva componente,  $R$  é um vetor com todos os valores de  $\lambda$ ,  $C$  é o comprimento de onda central de cada curva, e  $W$  é a distância desse centro na qual a intensidade de cada curva alcançará uma fração  $F$  da intensidade máxima. Um exemplo do efeito de cada parâmetro pode ser visto na Fig. 47 e um exemplo de espectro gerado pela soma de múltiplas gaussianas pode ser visto na Fig. 48.

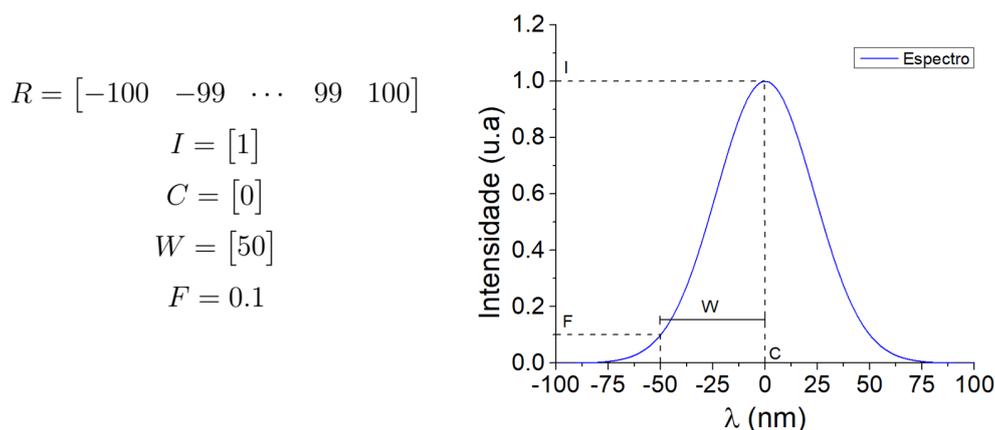


Figura 47: Efeito dos parâmetros nas curvas gaussianas geradas.

$$R = [400 \quad 401 \quad \dots \quad 649 \quad 650]$$

$$I = [1 \quad 0.6]$$

$$C = [520 \quad 560]$$

$$W = [30 \quad 60]$$

$$F = 0.1$$

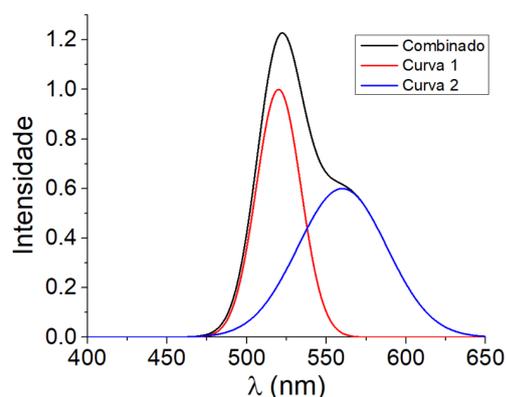


Figura 48: Exemplo de espectro gerado pela soma de múltiplas gaussianas.

O componente cinético da matriz por sua vez pode ser gerado pela integração numérica das leis de velocidade do sistema a ser simulado. Nesse trabalho um integrador numérico usando Runge-Kutta de 4ª ordem com correção de 5ª e passo adaptativo foi escolhida para a simulação (função *ode45* nativa ao MATLAB). Um passo adicional de interpolação é então necessário para que todas as matrizes tenham as mesmas dimensões e que cada elemento seja sempre referente a um tempo específico. Isso é necessário devido a natureza da integração usando passo adaptativo, que tende a ter passos menores em regiões com derivadas de valor elevado.

O produto do vetor com as informações espectrais pelo vetor de concentrações fornecerá então uma matriz de dados simulada, que por fim é somada a uma matriz adicional contendo ruído aleatório com valores por volta de 5% do sinal. Um fluxograma mostrando o processo de geração de dados pode ser visto na Fig. 49.

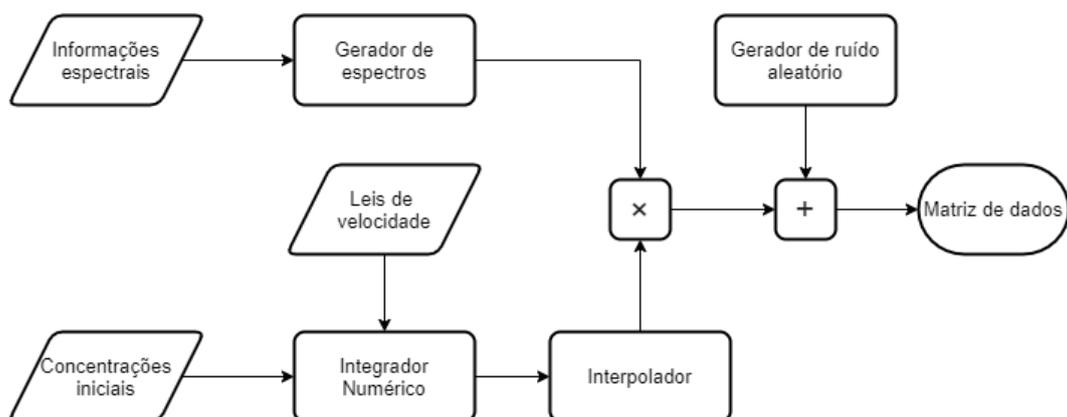


Figura 49: Fluxograma do processo de geração de dados simulados.

Um exemplo contendo uma matriz simulada com uma reação do tipo  $A \longrightarrow B$  pode ser vista na Fig. 50.

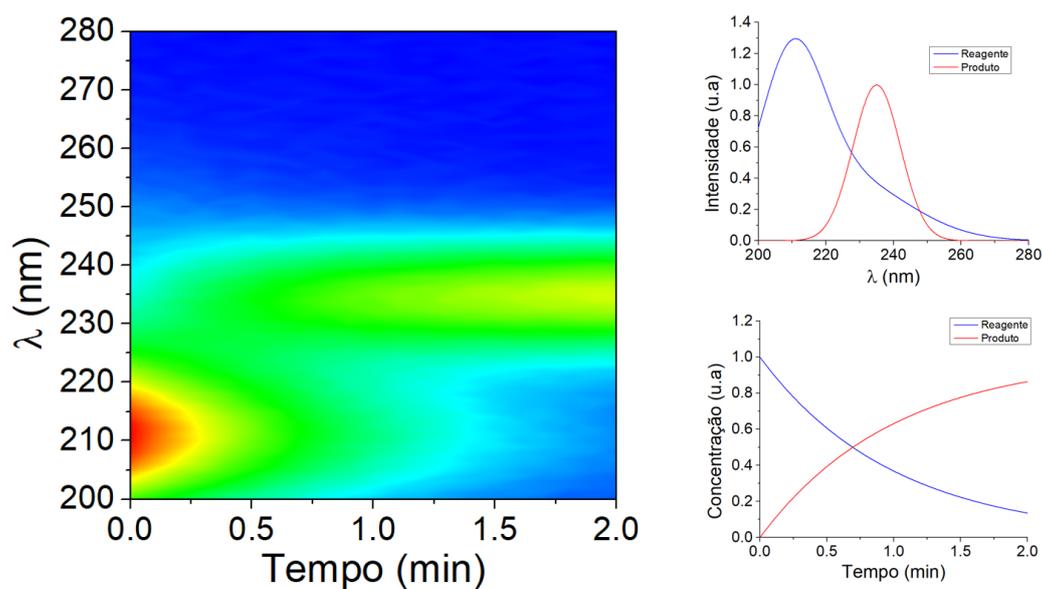


Figura 50: Exemplo de matriz de segunda ordem de absorção  $\times$  tempo para uma reação do tipo  $A \longrightarrow B$ .

## B.2 Simulador de dados espectrais

```

1 function [profiles,sensor]=sprofile_generator(centres,
        width,intensity,range)
2 %Inputs:
  
```

```
3 % Centres: Matrix with the centres for the different
   bands in the spectral profile. Different bands in the
   same spectra must use different columns in the same
   line, different spectras use different lines.
4 % Width: Distance from the center in which the intensity
   of a band reaches 10% of the value of the peak. This
   percentage can be changed by changing the value of
   the Int_fraction variable.
5 % Intensity: Relative intensity of each peak.
6 % Range: Spectral range for the spectra. Ex: 200 to 800
   nm with 10 nm spaces, range=[200:10:800];
7 %
8 %Outputs:
9 % Profiles: Generated profiles, each profile in a
   different column.
10
11 Int_fraction=0.1;
12 [l,c]=size(centres);
13 [~,pf_size]=size(range);
14 profiles=zeros(pf_size,l);
15 for i=1:l
16     for j=1:c
17         profiles(:,i)=profiles(:,i)+intensity(i,j)*exp(-(
           range-centres(i,j)).^2*1/(width(i,j)^2/(-log(
           Int_fraction))))';
18     end
19 end
20
21 %Sensor data
22 sensor=[range(1) range(pf_size) range(2)-range(1)];
```

### B.3 Simulador de dados cinéticos

```
1 function [t,C]=kinetic_curves(dC,t,C0)
2 % Kinetic curves generator
```

```
3 %
4 % Uses an 4th order Runge-Kutta numerical integrator to
   simulate
5 % kinetic curves.
6 %
7 %Inputs:
8 % dC: Matlab function containing the reaction rates for
   each
9 % component. (Ex: dC=@(t,Cx) [-Cx(1);-1.5*Cx(2);Cx(1)];)
10 % t: Time vector for the simulation (start at t=0).
11 % C0: Vector with the initial concentrations for each
   component.
12 %
13 %Outputs:
14 % t: Vector with the time values for the simulation
15 % C: Vector with the concentrations for each component
   in each
16 %time value for the simulation
17
18 [ti,Ci]=ode45(dC,[min(t), max(t)], C0);
19 C=interp1(ti,Ci,t);
20 end
```

#### B.4 Exemplo de geração de dados espectrocinéticos

```
1 %% Concentracoes iniciais das amostras
2 A(:,1)=[1;2;3;4;5;6;7;8];
3 A(:,2)=[0;0;0;0;0;0;0;0];
4 A(:,3)=[1;0;8;3;5;7;2;5];
5
6 %% Espectros
7 centres=[210 220; 0 230; 240 0];
8 width= [20 40; 0 15; 25 0];
9 intensity=[1 0.4; 0 1.3; 1 0];
10 range=[200:1:280];
```

```
11 [B,sensor]=sprofile_generator(centres,width,intensity,
    range);
12
13 %% Cinetica
14 dC= @(t_int,Cx) [-1.0*Cx(1); 1.0*Cx(1);-1.3*Cx(3)];
15 t=[0:0.1:2];
16
17 %% Geracao do tensor
18 for i=1:size(A)
19     C0=A(i,:);
20     [t,C]=kinetic_curves(dC,t,C0);
21     X(i,:,:)=B*C';
22 end
23
24 %Ruido
25 X=X+0.05*rand(size(X));
```

## Apêndice C - Resultados da determinação de melamina

Table C.1: Amostras de teste and concentrações / ppm

Adulteração (%)	Samples				PARAFAC	
	Marca	Tipo	Batealda*	Concentração de melamina	Concentração prevista	Recuperação
1**	A	Desnatado	1	140.0	184.0	< LOQ
					212.1	< LOQ
2**	B	Integral	3	140.0	210.0	< LOQ
					197.2	< LOQ
					195.8	< LOQ
					192.8	< LOQ
					181.7	< LOQ
					182.2	< LOQ
					174.8	< LOQ
					187.5	< LOQ
3	C	Integral	1	188.0	277.4	< LOQ
4	A	Semidesnatado	1	150.4	202.6	< LOQ
					214.3	< LOQ
5	A	Integral	1	140.0	191.0	< LOQ
					212.5	< LOQ
6	C	Integral	1	184.0	249.7	< LOQ
					335.8	< LOQ
7	B	Integral	1	211.0	324.1	< LOQ
					333.9	< LOQ
8	A	Integral	2	264.0	354.5	< LOQ
					289.1	< LOQ
9	D	Desnatado	1	275.8	262.9	< LOQ
					366.1	130%
10	B	Integral	2	278.2	317.3	< LOQ
					269.5	< LOQ
11	E	Desnatado	1	282.0	336.8	< LOQ
					397.4	134%
12	A	Semidesnatado	1	287.6	385.6	130%
					349.7	< LOQ
13	A	Desnatado	1	297.6	324.8	< LOQ
					403.3	103%
14	A	Integral	1	287.2	507.7	130%
					519.5	125%
15	C	Integral	1	393.0	516.5	124%
					510.9	114%
16	B	Semidesnatado	1	389.2	522.9	116%
					449.5	106%
17	A	Semidesnatado	1	417.0	438.4	103%
					447.4	93%
18	A	Desnatado	1	449.6	449.5	106%
					438.4	103%
19	A	Integral	1	423.8	449.5	106%
					438.4	103%
20	C	Integral	1	480.0	447.4	93%
					447.4	93%

APÊNDICE C - RESULTADOS DA DETERMINAÇÃO DE MELAMINA

21		C	Integral	1	540.0	622.5	115%
22		A	Integral	2	540.4	447.0	83%
						523.1	97%
23		D	Desnatado	1	600.0	663.4	111%
						624.2	104%
24		B	Integral	2	564.2	508.8	90%
	8%					534.7	95%
25		E	Desnatado	1	552.4	440.5	80%
						641.4	116%
26		A	Semidesnatado	1	535.4	453.4	85%
						601.6	112%
27		A	Desnatado	1	516.4	629.9	122%
28		A	Integral	1	560.0	530.0	95%
						551.5	98%
29		C	Integral	1	632.0	580.3	92%
30	9%	C	Semidesnatado	1	643.6	648.0	101%
31		C	Integral	1	737.0	683.9	93%
32		A	Semidesnatado	1	716.8	666.2	93%
						743.4	104%
33	10%	A	Desnatado	1	700.6	720.7	103%
						789.5	113%
34		A	Integral	1	700.4	675.5	96%
						682.6	97%
35	11%	C	Integral	1	789.0	740.6	94%
36		C	Integral	1	827.0	752.3	91%
37		A	Integral	2	849.2	820.1	97%
						833.1	98%
38		D	Desnatado	1	826.6	723.8	88%
	12%					716.9	87%
39		B	Integral	2	871.4	710.7	82%
						751.5	86%
40		E	Desnatado	1	872.0	986.1	113%
						903.5	104%

\* Números de batelada não podem ser comparados entre leites de diferentes marcas e tipos.

\*\* Amostras usadas para determinação de LoD.

Tabela C.2: Leite puro (Predição dos brancos para calculo do LoB)

	Marca	Batelada	Tipo	Predição (ppm)
1	C	1	Integral	48.4
2	C	1	Integral	98.7
3	C	1	Semidesnatado	31.4
4	B	1	Integral	63.0
5	B	1	Semidesnatado	46.4
6	A	2	Integral	16.2
7	E	1	Desnatado	51.8
8	E	1	Desnatado	4.2
9	B	2	Integral	-5.9
10	B	2	Integral	10.7
11	F	1	Desnatado	-44.8
12	F	1	Desnatado	-20.8
13	A	2	Semidesnatado	-44.3
14	A	2	Semidesnatado	-47.3
15	A	2	Desnatado	88.5
16	A	2	Desnatado	6.0
17	A	1	Integral	55.9
18	A	1	Integral	69.3