

**UNIVERSIDADE ESTADUAL DE CAMPINAS  
INSTITUTO DE QUÍMICA  
DEPARTAMENTO DE QUÍMICA ANALÍTICA**



**UNICAMP**

**APLICAÇÃO E VALIDAÇÃO DE MODELOS DE  
CALIBRAÇÃO DE SEGUNDA ORDEM EM QUÍMICA  
ANALÍTICA**

Tese de Doutorado

Autor: Jez Willian Batista Braga

Orientador: Prof. Dr. Ronei Jesus Poppi

Campinas, Fevereiro 2008

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

B73a	<p>Braga, Jez Willian Batista. Aplicação e validação de modelos de calibração de segunda ordem em química analítica / Jez Willian Batista Braga. -- Campinas, SP: [s.n], 2008.</p> <p style="text-align: center;">Orientador: Ronei Jesus Poppi.</p> <p style="text-align: center;">Tese - Universidade Estadual de Campinas, Instituto de Química.</p> <p style="text-align: center;">1. Validação. 2. Calibração multivariada. 3. Quimiometria. 4. Figuras de mérito. I. Poppi, Ronei Jesus. II. Universidade Estadual de Campinas. Instituto de Química. III. Título.</p>
------	--

**Título em inglês:** Application and validation of second order calibration models in analytical chemistry

**Palavras-chaves em inglês:** Validation, Multivariate calibration, Chemometrics, Figures of merit

**Área de concentração:** Química Analítica

**Titulação:** Doutor em Ciências

**Banca examinadora:** Ronei Jesus Poppi (orientador), Prof. Dr. Elias Ayres Guidetti Zagatto (CENA-USP-Piracicaba), Profa. Dra. Maria Fernanda D'Oliveira Pimentel (UFPE-DQ), Prof. Dr. Francisco Benedito Teixeira Pessine (IQ-UNICAMP), Prof. Dr. José Alberto Fracassi da Silva (IQ-UNICAMP)

**Data de defesa:** 22/02/2008

## **Banca Examinadora:**

Prof. Dr. Ronei Jesus Poppi (orientador)

Prof. Dr. Elias Ayres Guidetti Zagatto (CENA-USP)

Profa. Dra. Maria Fernanda D'Oliveira Pimentel (DQ-UFPE)

Prof. Dr. Francisco Benedito Teixeira Pessine (IQ-UNICAMP)

Prof. Dr. José Alberto Fracassi da Silva (IQ-UNICAMP)

## **Suplentes:**

Prof. Dr. Celio Pasquini (IQ-UNICAMP) (**1º Suplente**)

Profa. Dra. Solange Cadore (IQ-UNICAMP) (**2º Suplente**)

Prof. Dr. Roberto Kawakami Harrop Galvão (ITA) (**Suplente externo**)

*Dedico este trabalho à minha amada esposa Lilian Rodrigues Braga por seu amor e companheirismo. E aos meus amados pais Jez Batista Vieira e Laurinda Maria Batista. Obrigado pelo amor, exemplo, apoio e por sempre estarem juntos a mim durante toda a minha vida.*

## Agradecimentos

- Ao Prof. Dr. Ronei Jesus Poppi, pela oportunidade da realização desse trabalho, pela orientação, confiança, convivência e amizade.
- À UNICAMP, pelo programa de estágio docente (PED-A) (antigo programa de bolsa para instrutores graduados) que financiou o projeto e por fornecer toda a estrutura física, tecnológica e humana para a realização deste trabalho.
- Ao Instituto de Química da UNICAMP, por ter me concedido o privilégio e a honra de atuar como professor nas disciplinas dos cursos de graduação durante os quatro anos em que realizei o doutorado.
- Aos professores que foram meus coordenadores nas disciplinas que ministrei nos cursos de graduação da UNICAMP, pelo apoio e auxílio.
- Às Profas. Dras. Carla B. G. Botolli e Izabel C. S. F. Jardim pela ajuda e disponibilização do equipamento, padrões e materiais na realização da aplicação em HPLC-DAD na determinação de pesticidas e metabólitos em vinho.
- Aos Profs. Drs. Alejandro C. Olivieri, Héctor Goicoechea e Nicolaas M. Faber, pelas discussões sobre figuras de mérito e ajuda na fase inicial do desenvolvimento dos programas para BLLS e UPLS.
- Aos Profs. Drs. Lauro T. Kubota e Marcelo G. de Oliveira, pela disponibilização dos espectrofluorímetros utilizados para a realização das análises dos dois primeiros estudos de caso desta tese.
- À minha amada esposa Lilian Rodrigues Braga, por seu amor, companheirismo, incentivo e ajuda em todo esse período.

- A minha família, meus pais Jez Batista Vieira e Laurinda Maria Batista, meus irmãos Wanderley Batista Braga e Alfredo Ferreira Braga Neto e minha madrinha Guiomar Alves Braga, obrigado pelo amor, carinho, incentivo e força durante toda minha vida.
- Aos colegas do grupo LAQQA: Alessandra Borin, Gilmore A. da Silva, Marcelo M. de Sena, Marcelo G. Trevisan, Waldomiro B. Neto, Luiz C. M. Pataca, Patrícia Valderrama, Paulo H. Março, Sílvio L. T. Lima, Luciana Viviane, Diorginis B. M. Ribeiro, Werickson F. C. Rocha, Thiago P. Araújo, Joana G. Aguiar, Rosineide C. Simas, Ingrid K. de Oliveira, Júlio C. L. Alves e Danilo A. Maretto. Obrigado pelo apoio, companheirismo e amizade.
- Ao amigo e colega do grupo LAQQA Renato L. Carneiro, pela amizade, companheirismo e parceria na realização dos experimentos de três das quatro aplicações descritas nessa tese.
- Aos colegas do LEEDS obrigado pelo apoio e companheirismo.
- A todos os meus amigos de graduação e de minha cidade, que mesmo com a distância nunca deixaram de me apoiar e incentivar.
- A todos os técnicos dos laboratórios de ensino do Instituto de química e aos técnicos Ricardo e Cláudia, pela amizade e auxílio durante as aulas que ministrei nos laboratórios de ensino e nos experimentos de UV-VIS, fluorescência e cromatografia.
- Aos professores José Divino dos Santos, Olacir Alves de Araújo e Waldomiro Borges Neto, pelo apoio, ajuda e incentivo.
- Aos colegas Sérgio Bertazzo, Aloízio, Marcelo (caixa) e Jorge, pelo convívio, companheirismo e amizade.

- Aos funcionários, colegas e amigos do Instituto de Química da Unicamp e a todas as pessoas que, de uma forma ou outra, contribuíram para realização deste trabalho e para esta etapa de minha vida, os meus sinceros agradecimentos.

## SÚMULA CURRÍCULAR

### DADOS PESSOAIS:

Nome: Jez Willian Batista Braga

Nascimento: 23/08/1978

E-mail: [jezwillian@gmail.com](mailto:jezwillian@gmail.com)

Naturalidade: Anápolis/GO - Brasil

Estado Civil: casado

### FORMAÇÃO:

Graduação em Licenciatura em Química – 1998/2001

Universidade Estadual de Goiás, UEG, Goiás, Brasil

Mestre em Química, na área de Química Analítica – 2002/2004

Universidade Estadual de Campinas, UNICAMP, Campinas, Brasil

Título: Avaliação de figuras de mérito em Calibração Multivariada, aplicada na determinação de Carbamazepina por Espectroscopia no Infravermelho Próximo e Médio.

Orientador: Prof. Dr. Ronei Jesus Poppi

### EXPERIÊNCIA PROFISSIONAL:

Professor nas disciplinas de Física e Química – 1999/2001

Níveis: Ensino Médio Colegial e Técnico em Contabilidade.

Local: Colégio Estadual José Résio, Goianápolis-GO

Professor nas disciplinas de Química Geral e Química Analítica (experimentais e teóricas) – 2004/2007. Nível: Graduação.

Local: Universidade Estadual de Campinas, Campinas-SP.

### ATIVIDADES ACADÊMICAS:

Monitoria Acadêmica: Programa de Estágio Docente (PED-C) – 03/2003 a 06/2003.

Disciplina: QA313 – Análise Quantitativa

Curso: Engenharia Química

Instituto de Química – UNICAMP

Estágio em Sistemas Especialistas

Duração: 04/2001 a 10/2001

Linguagem Visual Prolog 4.1

Instituto de Química – UFG

## LISTA DE PUBLICAÇÕES:

- 1 – J. W. B. Braga e R. J. Poppi; *Figures of merit for determination of the polymorphic purity of carbamazepine by infrared spectroscopy and multivariate calibration*. J. Pharm. Sci. **2004**, 93, 2124.
- 2 – J. W. B. Braga e R. J. Poppi; *Validação de modelos de calibração multivariada: uma aplicação na determinação de pureza polimórfica de carbamazepina por espectroscopia no infravermelho próximo*. Quím. Nova, **2004**, 27, 1004.
- 3 – M. M. Sena, C. B. Freitas, L. C. Silva, J. W. B. Braga e R. J. Poppi; *Aplicação de calibração multivariada no controle de qualidade de fármacos*. Estudos (Goiânia), **2005**, 32, 1953.
- 4 – J. W. B. Braga e R. J. Poppi; *Comparison of variance sources and confidence limits in two PLSR models for determination of the polymorphic purity of carbamazepine*. Chemom. Intell. Lab. Syst. **2006**, 80, 50.
- 5 – J. W. B. Braga, C. B. G. Botolli, I. C. S. F. Jardim, H. C. Goicoechea, A. C. Olivieri e R. J. Poppi; *Determination of pesticides and metabolites in wine by high performance liquid chromatography and second-order calibration methods*. J. Chromatogr. A, **2007**, 1148, 200.
- 6 – R. L. Carneiro, J. W. B. Braga e R. J. Poppi; *Application of genetic algorithm for selection of variables for the BLLS method applied to determination of pesticides and metabolites in wine*. Anal. Chim. Acta, **2007**, 595, 51.
- 7 – P. Valderrama, J. W. B. Braga e R. J. Poppi; *Validation of multivariate calibration models in the determination of sugar cane quality parameters by near infrared spectroscopy*. J. Braz. Chem. Soc. **2007**, 18, 259.
- 8 – P. Valderrama, J. W. B. Braga e R. J. Poppi; *Variable selection, outlier detection and figures of merit estimation in a PLSR multivariate calibration model. A case study for the determination of quality parameters in alcohol industry by near infrared spectroscopy*. J. Agric. Food Chem. **2007**, 52, 8331.
- 9 – R. L. Carneiro, J. W. B. Braga, R. J. Poppi e R. Tauler; *Multivariate Curve Resolution of pH gradient flow injection mixture analysis with Schlieren effect*. Analyst, **2008**, artigo aceito para publicação.

## TRABALHOS EM EVENTOS INTERNACIONAIS:

- 1 – R. L. Carneiro, J. W. B. Braga e R. J. Poppi; *Aplicação de MCR-ALS na quantificação simultânea dos ácidos ascórbico (AA) e acetilsalicílico (AAS) em fármacos através de FIA e análise da influência do efeito Schlieren*. II Congresso Iberoamericano e IV Congresso Argentino de Química Analítica, **2007**, Buenos Aires, Argentina.
- 2 – R. L. Carneiro, J. W. B. Braga e R. J. Poppi; *Determinação simultânea de riboflavina e piridoxina em formulações infantis de leite em pó e suplementos alimentares por fluorescência de excitação e emissão e PARAFAC*. II Congresso Iberoamericano e IV Congresso Argentino de Química Analítica, **2007**, Buenos Aires, Argentina.
- 3 – R. L. Carneiro, J. W. B. Braga e R. J. Poppi; *Comparação entre modelos de calibração de segunda ordem BLLS-RBL e NPLS-AG na quantificação de ácido ascórbico e ácido acetilsalicílico em medicamentos por FIA-DAD*. II Congresso Iberoamericano e IV Congresso Argentino de Química Analítica, **2007**, Buenos Aires, Argentina.
- 4 – J. W. B. Braga, C. B. G. Botolli, H. C. Goicoechea, A. C. Olivieri e R. J. Poppi; *Comparison of BLLS-RBL, PLS-RBL and PARAFAC for pesticides and metabolites determinations in wine by SPE and HPLC-DAD*. 10th International Conference on Chemometrics in Analytical Chemistry, **2006**, Águas de Lindóia, Brasil.
- 5 – R. L. Carneiro, J. W. B. Braga, C. B. G. Botolli e R. J. Poppi; *Application of Genetic Algorithm for variables selection in BLLS method for pesticides and metabolites determinations in wine*. 10th International Conference on Chemometrics in Analytical Chemistry, **2006**, Águas de Lindóia, Brasil.
- 6 – P. Valderrama, J. W. B. Braga e R. J. Poppi; *Figures of merit for the determination of quality parameters in sugar cane industry by near infrared spectroscopy and multivariate calibration* In: 12th International Conference on Near-Infrared Spectroscopy, **2005**, Sky City, Nova Zelândia .
- 7 – J. W. B. Braga, R. J. Poppi e C. Mello; *Figures of merit in determination of polymorphic purity of Carbamazepine by Infrared Spectroscopy*. The Second International Conference on Advanced Vibrational Spectroscopy (ICAVS-2), **2003**, Nottingham, Inglaterra.
- 8 – J. W. B. Braga e R. J. Poppi. *Determination of polymorphic purity of carbamazepine by near infrared spectroscopy*. 4th Congress of Pharmaceutical Sciences–CIFARP, Ribeirão Preto, Brasil. Publicado no Braz. J. Pharm. Sci. **2003**, 39, 107.

9 – G. Magarelli, J. W. B. Braga, A. H. Fostier, e R. J. Poppi; *Study of the processes influencing the air/surface of mercury by using PCA analysis: measurements at the Negro River Basin, Amazon*. 4th International Symposium on Environmental Geochemistry in Tropical Countries, **2004**, Buzios, Brasil.

#### **APRESENTAÇÃO ORAL DE TRABALHOS EM EVENTOS NACIONAIS:**

1 – J. W. B. Braga, L. S. R. Morais, C. B. G. Botolli, I. C. S. F. Jardim e R. J. Poppi; *Aplicação de quadrados mínimos bilinear (BLLS) na determinação de pesticidas e metabólitos em amostras de vinho por HPLC-DAD*. 29ª Reunião Anual da Sociedade Brasileira de Química, **2006**, Águas de Lindóia-SP.

#### **TRABALHOS EM EVENTOS NACIONAIS:**

1 – J. W. B. Braga, R. L. Carneiro e R. J. Poppi; *Desenvolvimento e validação de um método analítico por FIA e BLLS, aplicado à determinação simultânea de AAS e AA em medicamentos*. 30ª Reunião Anual da Sociedade Brasileira de Química, **2007**, Águas de Lindóia.

2 – R. L. Carneiro, J. W. B. Braga e R. J. Poppi; *Emprego de algoritmos genéticos para seleção de variáveis na determinação das vitaminas B2 e B6 em preparações lácteas por fluorescência e NPLS*. 14º Encontro Nacional de Química Analítica, **2007**, João Pessoa.

3 – J. W. B. Braga, P. Valderrama e R. J. Poppi; *Identificação de amostras anômalas em calibração multivariada. Aplicação em determinação de BRIX em caldo de cana por NIR*. 28ª Reunião Anual da Sociedade Brasileira de Química, **2005**, Poços de Caldas.

4 – J. W. B. Braga e R. J. Poppi; *Determinação da variância dos erros de previsão e limites de confiança para modelos de calibração multivariada construídos por PLS*. XXVI Congresso Latinoamericano de Química e 27ª Reunião Anual da Sociedade Brasileira de Química, **2004**, Salvador.

5 – J. W. B. Braga e R. J. Poppi; *Construção e validação de um modelo de calibração multivariada com FTIR para determinação de formas polimórficas de carbamazepina*. 26ª Reunião Anual da Sociedade Brasileira de Química, **2003**, Poços de Caldas.

## Resumo

### APLICAÇÃO E VALIDAÇÃO DE MODELOS DE CALIBRAÇÃO DE SEGUNDA ORDEM EM QUÍMICA ANALÍTICA

Autor: Jez Willian Batista Braga

Orientador: Ronei Jesus Poppi

Modelos de calibração de segunda ordem vêm sendo cada vez mais aplicados, principalmente pela possibilidade de determinar analitos mesmo na presença de interferentes da amostra. Contudo, a validação desses modelos ainda não é possível por dificuldades na determinação de algumas figuras de mérito. Esta tese teve como objetivos estudar a determinação e avaliar a consistência de algumas figuras de mérito em modelos de calibração de segunda ordem, desenvolver programas que facilitem sua determinação. Também foram comparados a eficiência dos modelos de quadrados mínimos bilinear (BLLS) e quadrados mínimos parciais desdobrado (UPLS), ambos utilizando o processo de bilinearização residual (RBL), com os modelos de análise de fatores paralelos (PARAFAC) e resolução multivariada de curvas (MCR). Foram desenvolvidos quatro estudos, abordando as determinações de: (1) hidrocarbonetos policíclicos aromáticos, (2) riboflavina e piridoxina em matrizes lácteas (ambos por emissão e excitação de fluorescência), (3) ácido acetilsalicílico e ascórbico em fármacos por análise por injeção em fluxo com gradiente de pH e (4) pesticidas e metabólitos em vinho por cromatografia líquida de alta eficiência. Os resultados demonstraram que as figuras de mérito apresentam boa consistência, sendo observada influência de interferentes na sensibilidade, seletividade e nos erros obtidos. Contudo, observou-se a necessidade de desenvolvimentos nas estimativas de incerteza e seletividade, onde duas novas propostas para o modelo UPLS são descritas. Os modelos comparados foram equivalentes na maior parte dos casos, sendo que a escolha correta do número de fatores no RBL requer um conjunto de amostras externas.

## **Abstract**

### **APLICACION AND VALIDATION OF SECOND ORDER CALIBRATION MODELS IN ANALYTICAL CHEMISTRY**

Author: Jez Willian Batista Braga

Adviser: Ronei Jesus Poppi

Second order calibration models are becoming widely applied in analytical chemistry, mainly due to the possibility to determine analytes even in the presence of interferences. However, an appropriate validation of these models is not possible due to the difficult in the determination of some figures of merit. The aims of this thesis were to study and evaluate the determination of some figures of merit of the most applied second order calibration models, to develop softwares to perform its determination easier and to compare the performance of the bilinear least squares (BLLS) and the unfold partial least squares (UPLS), both applying the residual bilinearization process (RBL), with the models parallel factor analysis (PARAFAC) and multivariate curve resolution (MCR). Four applications were developed, which study the determination of: (1) polycyclic aromatic hydrocarbons, (2) riboflavin and pyridoxine in dairy supplements (both using excitation and emission fluorescence), (3) acetylsalicylic and ascorbic acids in pharmaceutical samples by FIA and pH gradient and (4) pesticides and metabolites in wine samples by HPLC. The results show that the estimated figures of merit present a good agreement with their expected values, also been observed the effects of the interferences of the sample in the sensitivity, selectivity and uncertainty values. It was also observed the necessity of new developments in the estimates of selectivity and uncertainty parameters, where two new approaches for the model UPLS are described. The compared models were equivalent in most of the situations and it was observed that the choice of the correct number of factors in the RBL requires the use of an independent data set.

## Sumário

<b>Introdução.....</b>	<b>1</b>
<b>Capítulo 1 - Modelos de calibração de segunda ordem e figuras de mérito.....</b>	<b>9</b>
<b>1.1 Modelos de calibração de segunda ordem. ....</b>	<b>14</b>
1.1.1 Análise de fatores paralelos (PARAFAC) .....	14
1.1.2 Resolução de curvas multivariada (MCR).....	19
1.1.3 Quadrados mínimos bilinear (BLLS).....	23
1.1.4 Quadrados mínimos parciais desdobrado (UPLS) .....	29
<b>1.2 Validação de modelos de calibração de segunda ordem: Figuras de Mérito.....</b>	<b>32</b>
1.2.1 Precisão.....	34
1.2.2 Erro médio de estimativa da propriedade de interesse .....	34
1.2.3 Exatidão .....	36
1.2.4 Sensibilidade (SEN) .....	36
1.2.4.1 Projeções ortogonais. ....	38
1.2.4.2 Sensibilidade para modelos PARAFAC e BLLS na ausência de interferentes. ....	38
1.2.4.3 Sensibilidade para modelos PARAFAC e BLLS com um analito calibrado e na presença de interferentes.....	40
1.2.4.4 Expressão geral para a determinação da sensibilidade em modelos PARAFAC e BLLS.....	40
1.2.4.5 Sensibilidade para o modelo UPLS.....	42
1.2.4.6 Sensibilidade em modelos MCR .....	44
1.2.5 Sensibilidade Analítica (SENA).....	44
1.2.6 Limite de detecção (LD).....	45
1.2.7 Seletividade (SEL) .....	46

1.2.8	Incerteza na concentração estimada da propriedade de interesse.....	49
<b>Capítulo 2 - Estudo da determinação de hidrocarbonetos policíclicos aromáticos em misturas .....</b>		
		57
<b>2.1</b>	<b>Procedimento experimental.....</b>	60
<b>2.2</b>	<b>Resultados e discussão. ....</b>	62
2.2.1	Escolha do número de fatores utilizado no RBL.....	64
2.2.2	Comparação dos modelos BLLS, UPLS e PARAFAC e figuras de mérito.....	75
2.2.3	Conclusões do capítulo.....	95
<b>Capítulo 3 - Determinação de riboflavina e piridoxina em leite em pó e suplementos alimentares .....</b>		
		97
<b>3.1</b>	<b>Procedimento experimental.....</b>	100
3.1.1	Estudo de estabilidade das soluções de trabalho .....	100
3.1.2	Procedimento das medidas de emissão e excitação de fluorescência (EEF).....	100
3.1.3	Preparo das soluções padrões .....	100
3.1.4	Preparo das amostras para as medidas de EEF .....	101
3.1.5	Procedimento das medidas cromatográficas utilizadas para a validação. ....	102
3.1.6	Preparo de soluções padrões e amostras para as medidas por HPLC.....	103
<b>3.2</b>	<b>Resultados e discussão .....</b>	104
<b>3.3</b>	<b>Conclusões do capítulo.....</b>	122

**Capítulo 4 - Determinação simultânea de ácido acetilsalicílico e ascórbico em medicamentos ..... 125**

<b>4.1</b>	<b>Procedimento experimental.....</b>	<b>128</b>
4.1.1	Construção do sistema FIA .....	128
4.1.2	Planejamento experimental para preparo dos padrões de calibração e misturas de validação.....	131
4.1.3	Preparo das amostras de medicamentos comerciais para a análise no sistema FIA. ....	133
4.1.4	Validação da determinação dos teores de AAS e AA em medicamentos.....	134
4.1.4.1	<i>Determinação de AAS por HPLC.....</i>	<i>135</i>
4.1.4.2	<i>Determinação de ácido ascórbico (AA) por Iodimetria <sup>112</sup>.....</i>	<i>136</i>
<b>4.2</b>	<b>Resultados e discussão .....</b>	<b>136</b>
<b>4.3</b>	<b>Conclusões do capítulo.....</b>	<b>159</b>

**Capítulo 5 - Determinação de pesticidas e metabólitos em amostras de vinho tinto ..... 161**

<b>5.1</b>	<b>Procedimento experimental.....</b>	<b>165</b>
5.1.1	Reagentes e soluções padrões.....	165
5.1.2	Instrumentação utilizada e condições cromatográficas .....	166
5.1.3	Procedimento para a extração em fase sólida.....	167
5.1.4	Amostras de vinho .....	168
<b>5.2</b>	<b>Pré-processamentos utilizados nos dados .....</b>	<b>169</b>
<b>5.3</b>	<b>Resultados e discussão .....</b>	<b>171</b>
5.3.1	Resultados obtidos pelo método proposto (MP) .....	175
5.3.2	Resultados obtidos pelo método de referência (MR) .....	190

5.3.3 Comparação entre os métodos propostos (MP) e de referência (MR).....	202
<b>5.4 Conclusões do capítulo.....</b>	<b>204</b>
<b>Capítulo 6 - Conclusões gerais.....</b>	<b>207</b>
<b>Capítulo 7 - Perspectivas futuras .....</b>	<b>213</b>
<b>Capítulo 8 - Referências.....</b>	<b>217</b>
<b>Apêndices.....</b>	<b>231</b>
<b>Apêndice A - Produtos matriciais de bloco.....</b>	<b>233</b>
1A - Produto de Kronecker.....	233
2A - Produto de Khatri-Rao .....	234
<b>Apêndices B - Programas desenvolvidos. ....</b>	<b>235</b>
B1 - Programa blls_rbl. ....	237
B2 - Sub-rotina func_mat_obj_rbl .....	251
B3 - Sub-rotina func_mat_obj_rbl_rd .....	253
B4 - Sub-rotina func_oct_obj_rbl. ....	255
B5 - Sub-rotina func_mat_obj_rbl_rd. ....	257
B6 - Programa upls1_rbl.....	259
B7 - Sub-rotina func_obj_mat_rbl_upls.....	269
B8 - Sub-rotina func_obj_oct_rbl_upls.m.....	271
B9 - Programa parafacpred.m. ....	273
B10 - Programa ortimef.m .....	277

## Lista de Abreviaturas

AA.....	Ácido ascórbico.
AAS.....	Ácido acetilsalicílico.
Abs.....	Intensidade de absorção.
ALS.....	Alternating least squares (quadrados mínimos alternantes).
AN.....	Antraceno.
ANVISA.....	Agência Nacional de Vigilância Sanitária.
ASTM.....	American Society for Testing Materials (Sociedade Americana para Testes em Materiais).
B2.....	Vitamina B2 ou riboflavina.
B6.....	Vitamina B6 ou piridoxina.
BAP.....	Benzo[a]pireno.
BLLS.....	Bilinear least squares (quadrados mínimos bilinear).
CAL.....	Calibração.
CV.....	Cross validation (validação cruzada).
DAD.....	Diode array detection (detecção por arranjo de diodos).
DTLD.....	Direct trilinear decomposition (decomposição trilinear direta).
EFA.....	Evolving factor analysis.
FAJ.....	Falta de ajuste.
EEF.....	Excitação e emissão de fluorescência.
FEN.....	Fenantreno.
FIA.....	Flow injection analysis (análise por injeção em fluxo).
GC.....	Gas chromatography (cromatografia gasosa) .

- HPLC..... High performance liquid chromatography (cromatografia líquida de alta eficiência).
- IC..... Intervalo de confiança.
- ICH..... International Conference on Harmonization (Conferência Internacional de Harmonização).
- INMETRO.. Instituto Nacional de Metrologia, Normalização e Qualidade Industrial.
- IUPAC..... International Union of Pure and Applied Chemistry (União Internacional de Química Pura e Aplicada).
- LD..... Limite de detecção.
- MCR..... Multivariate curve resolution (resolução de curvas multivariada).
- MLR..... Multiple linear regression (regressão linear múltipla).
- MP..... Método proposto
- MR..... Método de Referência
- MS..... Mass spectrometry (espectrometria de massas).
- MSEC..... Mean square error of calibration (erro médio quadrático de calibração).
- MSEP..... (Mean square error of prediction (erro médio quadrático de previsão).
- NAS..... Net analyte signal (sinal analítico líquido).
- NIPALS..... Nonlinear interative partial least squares.
- NIR..... Infravermelho próximo.
- NMR-2D.... Two dimensional nuclear magnetic resonance (ressonância nuclear magnética em 2 dimensões).
- NPLS..... N-way partial least squares (quadrados mínimos parciais N-

- dimensional).
- PC..... Principal components (componentes principais).
- PCA..... Principal components analysis (análise de componentes principais).
- PCR..... Principal components regression (regressão por componentes principais).
- PLS..... Partial least squares (regressão por quadrados mínimos parciais).
- PARAFAC.. Parallel factor analysis (análise de fatores paralelos).
- RD..... Resíduo de decomposição.
- RBL..... Residual bilinearization (bilinearização residual).
- REP..... Relative error of prediction (erro relativo de previsão).
- RMSEC..... Root mean square error of calibration (raiz quadrada do erro médio quadrático de calibração).
- RMSECV.... Root mean square error of cross validation (raiz quadrada do erro médio quadrático de validação cruzada).
- RMSEP..... Root mean square error of prediction (raiz quadrada do erro médio quadrático de previsão).
- SEL..... Seletividade.
- SEN..... Sensibilidade.
- SENA..... Sensibilidade analítica.
- SVD..... Singular value decomposition (decomposição em valores singulares).
- TCA..... Tri-chloro acetic acid (ácido tricloroacético).
- TLLS..... Trilinear least squares (quadrados mínimos trilinear).

USP..... United States Pharmacopoeia (Farmacopéia dos Estados Unidos).

UPLS..... Unfold partial least squares (quadrados mínimos parciais desdobrado).

VAL1..... Conjunto de validação.

VL..... Variável Latente.

## Lista de Símbolos

- Notação..... Letras maiúsculas em negrito e sublinhadas representaram cubos de dados, matrizes serão indicados por maiúsculas em negrito, vetores por letras minúsculas em negrito e escalares por letras normais maiúsculas ou minúsculas.
- A** ..... Matriz de pesos do PARAFAC proporcional à concentração das espécies presentes no meio.
- a**..... Vetor que contém uma das colunas da matriz **A** de pesos do PARAFAC.
- a**..... Elemento da matriz **A**.
- B**..... Matriz de pesos do PARAFAC que contém os perfis na dimensão J.
- b**..... Vetor que contém uma das colunas da matriz **B** de pesos do PARAFAC.
- b<sub>n</sub>**..... Vetor que contém o primeiro vetor singular de  $S_n$ , que corresponde ao perfil da dimensão J para a espécie n, estimado pelo modelo BLLS.
- b<sub>UPLS</sub>**..... Vetor de coeficientes de regressão do modelo UPLS.
- b<sub>RBL</sub>**..... Vetor de coeficientes de regressão do modelo UPLS levando em conta a utilização do RBL.
- b**..... Elemento da matriz **B**.
- C**..... Para o PARAFAC: matriz de pesos que contém os perfis na dimensão K.  
Para o MCR: matriz que contém os perfis de concentração das espécies presentes no meio.
- c**..... Elemento da matriz **C**.
- c<sub>n</sub>**..... Vetor que contém o primeiro vetor singular de  $S_n$ , que corresponde ao perfil da dimensão K para a espécie n, estimado pelo modelo BLLS.

- cal (subscrito)..... Indica uma amostra de calibração ou uma grandeza obtida por meio das amostras de calibração.
- E**..... Matriz de erros da decomposição da matriz **X**.
- e..... Elemento da matriz **E**.
- F**..... Número escolhido de fatores em um modelo PARAFAC, MCR e BLLS ou número de variáveis latentes em um modelo PLS.
- f**..... Vetor de resíduos de **y** deixados pela informação não modelada pelo modelo UPLS.
- $g_n$ ..... Primeiro valor singular de **S<sub>n</sub>**.
- G**..... Matriz que contém em sua diagonal principal todos os valores singulares de uma matriz **X**.
- $h_{un}$ ..... Medida da influência (do inglês “leverage”) da amostra **X<sub>un</sub>**.
- I**..... Matriz identidade de dimensões apropriada.
- I**..... Índice que representa o número de amostras.
- I<sub>c</sub>**..... Número de amostras de calibração.
- I<sub>v</sub>**..... Número de amostras utilizadas na etapa de validação ou teste.
- IC**..... Intervalo de confiança.
- int (subscrito)..... Indica que se trata de uma grandeza associada à um interferente.
- J**..... Índice que representa o número de variáveis da primeira dimensão de uma matriz (**J x K**).
- K**..... Índice que representa o número de variáveis da segunda dimensão de uma matriz (**J x K**).
- m**..... Número de replicatas.

$N_c$ .....	Número de espécies de interesse.
$\mathbf{P}$ .....	Pode representar a matriz de pesos (Loadings) de uma matriz $\mathbf{X}$ pelo modelo UPLS, ou uma matriz de projeção, dependendo do contexto.
$\mathbf{P}_{\text{eff}}$ .....	Matriz de pesos efetiva que é utilizada para estimar a sensibilidade no modelo UPLS quando há interferentes.
$\mathbf{q}$ .....	Vetor de pesos (Loadings) do vetor $\mathbf{y}$ .
$\mathbf{W}$ .....	Segunda matriz de pesos (Weight) de uma matriz $\mathbf{X}$ pelo modelo UPLS.
RD.....	Resíduo de decomposição de uma amostra.
SVD( ).....	Operador que calcula todos os valores e vetores singulares de uma matriz.
$\mathbf{S}$ .....	Matriz que contém os perfis espectrais do modelo MCR.
$\mathbf{S}_n$ .....	Matriz obtida por cada coluna de $\mathbf{V}_s$ , colocada na forma de uma matriz. Contém o sinal da espécie $n$ em concentração unitária.
$\mathbf{S}_{\text{cal}}$ .....	Matriz de dimensão $(JK, N_c)$ que contém as informações da regressão da etapa de calibração do modelo BLLS.
$\mathbf{S}_n^*$ .....	Matriz do sinal analítico líquido de 1 em concentração unitária.
$s_{\text{ruído}}$ .....	Estimativa do nível do ruído instrumental.
$s( )$ .....	Incerteza associada à estimativa de determinada grandeza.
$s_{\text{reg}}$ .....	Desvio padrão do erro de uma regressão univariada.
SEL.....	Seletividade.
SEN.....	Sensibilidade.
SENA.....	Sensibilidade Analítica.

- T** (sobrescrito)..... Representa a operação de transposição de uma matriz ou vetor.
- T** e **t**..... Matriz e vetor de escores (Scores) de uma matriz **X**.
- $t_{(1-\alpha)/2,v}$ ..... Valor da distribuição t-student com nível de confiança  $(1-\alpha)/2$  e  $v$  graus de liberdade.
- un** (subscrito)..... Indica uma amostra que esta sendo analisada.
- V**..... Matriz de loadings ortogonais de **X**.
- V(PE)**..... Variância dos erros de previsão.
- vec( )**..... Operador que realiza a vetorização de uma matriz.
- V<sub>X</sub>**..... Matriz de dimensão (JK,I) formada pela vetorização da matriz de dados de cada amostra I.
- V<sub>S</sub>**..... Matriz que contém as estimativas dos perfis puros dos compostos de interesse em concentração unitária.
- V( $\Delta X$ )** e **V( $\Delta x_{un}$ )**.. Variâncias das medidas instrumentais das amostras utilizadas na calibração e de uma amostra de previsão.
- V( $\Delta y$ )**..... Variância dos erros dos valores de referência das concentrações da espécie de interesse.
- V(e)** e **V(e<sub>un</sub>)**..... Variâncias do erro de previsão devido apenas ao ajuste do modelo para as amostras de calibração e para uma amostra de previsão.
- $x_{I,J,K}$ ..... Elemento do cubo de dados.
- X**..... Cubo de dados de dimensão (I,J,K) .
- X**..... Matriz que contém as respostas instrumentais de dimensões (J,K).
- Y** ou **y**..... Matriz ou vetor que contém as concentrações de referência de uma (**y**) ou mais de uma (**Y**) propriedade de interesse.
- $\alpha$ ..... Nível de recobrimento de um intervalo de confiança.

- $\varepsilon$ ..... Absortividade molar.
- $\lambda$ ..... Comprimento de onda.
- $\nu$ ..... Número de graus de liberdade perdidos na construção do modelo.
- $\omega$ ..... Inclinação da reta ajustada entre os pesos **a** do PARAFACe as concentrações **y** de referência.
- $\Theta$ ..... Matriz de projeção utilizada no calculo do sinal analítico líquido.
- $\| \ \|$ ..... Operador que calcula a norma Euclidiana de um vetor ou matriz.
- 1 (sobrescrito)..... Representa a operação de inversão de uma matriz ou vetor.
- $\wedge$  (sobrescrito)..... Símbolo que indica que a grandeza trata-se de um valor estimado.
- + (sobrescrito)..... Representa o cálculo da Moore-Penrose pseudo-inversa de uma matriz.
- $|\otimes|$ ..... Operador do produto de Katri-Rao.
- $\otimes$ ..... Operador do produto de Kronecker.

## Índice de Tabelas

<b>Tabela 1.</b> Algoritmo para a construção do modelo BLLS em situações onde não há presença de interferentes. ....	28
<b>Tabela 2.</b> Planejamento e níveis de concentração utilizados para construção das curvas de calibração e misturas de validação.....	61
<b>Tabela 3.</b> Valores de sensibilidade (SEN) e seletividade (SEL) obtidos para os modelos de calibração de segunda ordem utilizados quando a vantagem de segunda ordem não é utilizada. ....	77
<b>Tabela 4.</b> Resultados das figuras de mérito estimadas para antraceno (AN) nas quatro condições onde é determinado com BLLS, UPLS e PARAFAC.....	78
<b>Tabela 5.</b> Resultados das figuras de mérito estimadas para o benzo[a]pireno (BAP) nas quatro condições onde é determinado com BLLS, UPLS e PARAFAC.....	79
<b>Tabela 6.</b> Resultados das figuras de mérito estimadas para o fenantreno (FEN) nas quatro condições onde é determinado com BLLS, UPLS e PARAFAC.....	80
<b>Tabela 7.</b> Percentagens recuperadas para as probabilidades de recobrimento dos intervalos de confiança obtidos para PARAFAC e as duas estimativas com UPLS.....	88
<b>Tabela 8.</b> Concentrações esperadas, em $\mu\text{g/L}$ , para as vitaminas B2 e B6 nas análises por EEF para cada produto considerando o teor no rótulo. ....	102
<b>Tabela 9.</b> Concentrações esperadas (mg/L) nas análises por HPLC para as vitaminas B2 e B6 em cada produto levando em conta o valor atestado no rótulo por cada fabricante e o preparo de amostra.....	104

<b>Tabela 10.</b> Resultados das figuras de mérito obtidas para a determinação de B2, estimados comparando os resultados obtidos por HPLC e por EEF.....	110
<b>Tabela 11.</b> Recuperações obtidas para as determinações de B2.....	112
<b>Tabela 12.</b> Resultados das figuras de mérito obtidas para determinação de B6, estimados comparando os resultados obtidos por HPLC e por EEF.....	115
<b>Tabela 13.</b> Recuperações obtidas para as determinações de B6. ....	119
<b>Tabela 14.</b> Resultado dos teores das vitaminas B2 e B6 em mg/100g em cada produto, obtidos por EEF com o modelo BLLS, HPLC e os teores nominais atestados no rótulo.....	122
<b>Tabela 15.</b> Valores de t calculados e tabelados, com 99 % de confiança, para a comparação dos teores médios obtidos por EEF e HPLC. ....	122
<b>Tabela 16.</b> Concentrações em mg/L de AAS e AA nas amostras dos quatro medicamentos analisados. ....	135
<b>Tabela 17.</b> Resultados das figuras de mérito obtidas para determinação de AAS e AA, estimados comparando os resultados do método FIA com MCR e os valores de referência por HPLC e iodimetria para AAS e AA, respectivamente.....	144
<b>Tabela 18.</b> Número de fatores utilizados no RBL nos modelos BLLS e RBL que apresentam os menores erros em relação aos valores de referência. ....	149
<b>Tabela 19.</b> Resultados de RMSEP e REP obtidos para o modelo BLLS nos sistemas Octave 3.0.0 e Matlab 6.5 para a determinação de AAS e AA.....	151
<b>Tabela 20.</b> Resultados de SENA, SEL e LD obtidos como modelo BLLS com o sistema Octave 3.0.0.....	153

<b>Tabela 21.</b> Resultados obtidos para as figuras de mérito com o modelo UPLS em cada conjunto de dados analisado.....	155
<b>Tabela 22.</b> Recuperações percentuais obtidas para as determinações de AAS.....	157
<b>Tabela 23.</b> Recuperações percentuais obtidas para as determinações de AA.....	157
<b>Tabela 24.</b> Resultado dos teores de AAS e AA em mg/comprimido em cada medicamento, obtidos por FIA com o modelo BLS, HPLC ou iodimetria e os teores nominais atestados no rótulo. ....	159
<b>Tabela 25.</b> Intervalos de concentração utilizados nas soluções padrões de cada espécie de interesse. ....	166
<b>Tabela 26.</b> Pré-processamentos e número de fatores utilizados na construção dos modelos de calibração de segunda ordem foram empregados no MP. ....	177
<b>Tabela 27.</b> Resultados das figuras de mérito obtidas para a determinação dos pesticidas e metabólitos pelo método proposto (MP).....	189
<b>Tabela 28.</b> Resultados para as inclinações e interceptos das regressões lineares entre os valores de referência a estimados para as amostras EX no MP.....	190
<b>Tabela 29.</b> Pré-processamentos e número de fatores utilizados na construção dos modelos de calibração de segunda ordem que foram empregados no MR. ....	191
<b>Tabela 30.</b> Resultados das figuras de mérito obtidas para a determinação dos pesticidas e metabólitos pelo método de referência (MR).....	199
<b>Tabela 31.</b> Resultados para as inclinações e interceptos das regressões lineares entre os valores de referência a estimados para as amostras EX no MR. ....	200

**Tabela 32.** Valores obtidos para as recuperações percentuais nas amostras EX e VI comparando os valores obtidos no MP com as concentrações esperadas dos analitos que foram adicionados. ....203

**Tabela 33.** Valores obtidos para as recuperações percentuais nas amostras EX e VI comparando os valores obtidos no MR com as concentrações esperadas dos analitos que foram adicionados. ....203

## Índice de Figuras

- Figura 1-** Representação esquemática dos três tipos de dados em um sistema que adquire espectros em função do tempo. (a) de ordem-zero, representado por um escalar; (b) de primeira-ordem, vetor de dados; (c) segunda-ordem, matriz de dados..... 11
- Figura 2-** Representação de duas formas em que um cubo de dados pode ser arranjado para a análise em diferentes modelos de calibração de segunda ordem. (A) um cubo  $\underline{\mathbf{X}}(I,J,K)$  é colocado na forma de uma matriz  $\mathbf{X}(I,JK)$  pela vetorização da matriz uma das amostras I e (B)  $\underline{\mathbf{X}}(I, J, K)$  é re-arranjado em uma matriz  $\mathbf{X}(J,IK)$  pela colocação de cada uma das I amostras lado a lado. .... 14
- Figura 3-** Representação da decomposição trilinear dos dados nos perfis **a**, **b**, e **c** de cada dimensão, para um modelo PARAFAC de três fatores. .... 15
- Figura 4-** Esquema de decomposição do MCR quando se inicializa a otimização a partir dos perfis espectrais. Onde os sub-índices “cal” e “un” indicam amostras de calibração e de previsão, respectivamente; **X**, **C**, **S** e **E** são as matrizes de dados, perfis de concentração, espectrais e resíduos de decomposição, respectivamente. .20
- Figura 5-** Representação geométrica da propriedade do vetor NAS em ser ortogonal ao vetor do sinal dos interferentes. ....37
- Figura 6-** Fórmulas estruturais das espécies de interesse. ....60
- Figura 7-** Espectros de fluorescência de (A) emissão e (B) excitação normalizados, para as espécies de interesse no intervalo de comprimentos de onda ( $\lambda$ ) que foi selecionado para a construção das calibrações.....62

<b>Figura 8-</b> Superfície de fluorescência de uma solução de AN na concentração de 24,5 µg/L.....	63
<b>Figura 9-</b> Superfície de fluorescência de uma solução de BAP na concentração de 28,0 µg/L.....	63
<b>Figura 10-</b> Superfície de fluorescência de uma solução de FEN na concentração de 129,0 µg/L.....	64
<b>Figura 11-</b> Resíduos de decomposição para o BLLS nas situações onde três ou dois analitos estão calibrados.....	65
<b>Figura 12-</b> Concentrações de referência versus estimadas para as amostras de validação para o BAP na situação BAP FEN.....	67
<b>Figura 13-</b> (A) Resíduos de decomposição para o BLLS na calibração com apenas AN calibrado. (B) Ampliação de A. ....	68
<b>Figura 14-</b> (A) Resíduos de decomposição para o BLLS na calibração com apenas BAP calibrado. (B) Ampliação de A. Símbolos definidos como na Figura 13. ....	68
<b>Figura 15-</b> (A) Resíduos de decomposição para o BLLS na calibração com apenas FEN calibrado. (B) Ampliação de A. Símbolos definidos como na Figura 13. ....	69
<b>Figura 16-</b> Resíduos de decomposição para o UPLS nas situações onde três ou dois analitos estão calibrados.....	70
<b>Figura 17-</b> (A) Resíduos de decomposição para o UPLS na calibração com apenas AN calibrado. (B) Ampliação de A. Símbolos definidos como na Figura 16. ....	72
<b>Figura 18-</b> Espectros de emissão nos 9 diferentes comprimentos de onda de excitação utilizados.. ....	72

**Figura 19-** (A) Resíduos de decomposição para o UPLS na calibração com apenas BAP calibrado. (B) Ampliação de A.. .....73

**Figura 20-** (A) Resíduos de decomposição para o UPLS na calibração com apenas FEN calibrado. (B) Ampliação de A. Símbolos definidos como na Figura 13. ....74

**Figura 21-** Espectros de emissão nos 9 diferentes comprimentos de onda de excitação utilizados. ....74

**Figura 22-** Variação dos valores de SEL quando apenas AN e FEN estão calibrados. 82

**Figura 23-** Precisões obtidas para os analitos AN, BAP e FEN nas quatro situações em que foram determinados em cada amostra de validação (V1, V2, V3 e V4) nos três modelos de calibração utilizados. Número de graus de liberdade =20, exceto nas situações de divergência do RBL discutidas no texto. ....83

**Figura 24-** Incertezas estimadas para os analitos AN, BAP e FEN nas quatro situações em que foram determinados, estimadas para o PARAFAC pela equação 71 para o UPLS pelas equações 76 (UPLS) e 77 (UPLS-RBL), onde a equação 77 leva em conta o aumento da incerteza devido os interferentes. ....84

**Figura 25-** Histograma dos erros absolutos obtidos nas calibrações em que BAP|FEN (A e B) e FEN (C e D) estão calibrados. (A e C) PARAFAC, (B e D) UPLS.....87

**Figura 26-** Concentrações de referência versus estimadas obtidos nas calibrações em que BAP|FEN (A e B) e FEN (C e D) estão calibrados. (A e C) PARAFAC, (B e D) UPLS, intervalos de confiança de 99 % para inclinação e intercepto entre parênteses.....88

**Figura 27-** Intervalos de confiança de 95 % para as estimativas das concentrações de FEN pelo UPLS utilizando a equação 77 e a estimativa do ruído como 0,45 contagens. (\*) concentração de referência, (o) concentração estimada, (—) barras de erro.....90

**Figura 28-** Intervalos de confiança para a replicata 21 estimados com PARAFAC e com o UPLS utilizando a equação 76 e 77. (\*) concentração de referência, (o) concentração estimada, (—) barras de erro com 95 % de confiança. ....91

**Figura 29-** Valor absoluto dos perfis espectrais de emissão (esquerda) e excitação (direita) normalizados recuperados com o BLLS nas calibrações onde duas espécies de interesse são calibradas (— e —) e uma se apresenta como um interferente (—). Perfis experimentais são mostrados em preto (—).....92

**Figura 30-** Valor absoluto dos perfis espectrais normalizados de emissão (esquerda) e excitação (direita), recuperados com o BLLS nas curvas com apenas uma espécie calibrada (—) e duas se como interferentes (— e —). (Int 1) primeiro interferente e (Int 2) segundo interferente. ....93

**Figura 31-** Perfis normalizados estimados com BLLS para as replicatas da terceira amostra de validação na calibração onde apenas o AN é calibrado.....94

**Figura 32-** Estruturas moleculares das vitaminas B2 (esquerda) e B6 (direita). ....99

**Figura 33-** (A) Espectros da vitamina B2 adquiridos a cada 15 minutos na região UV-VIS em um intervalo de 21 horas. (B) Perfil cinético obtido pela subtração das absorbâncias em 445 nm e um comprimento de onda sem absorção (779 nm). ... 105

**Figura 34-** (A) Espectros da vitamina B6 adquiridos a cada 15 minutos na região UV-VIS em um intervalo de 21 horas. (B) Perfil cinético obtido pela subtração das absorvâncias em 292 nm e um comprimento de onda sem absorção (554 nm) .... 105

**Figura 35-** Superfícies de EEF para B2 (A) e B6 (B) obtidas para a solução padrão de concentração de 70,0 µg/L..... 106

**Figura 36-** Cromatogramas obtidos pelas análises por HPLC para as vitaminas B2 e B6 utilizando os comprimentos de onda de 450 e 290 nm, respectivamente. .... 107

**Figura 37-** Resíduos de decomposição para os modelos (A) BLLS e (B) UPLS obtidos na análise do produto Suplemento 1..... 108

**Figura 38-** Espectros normalizados de excitação (A) e emissão (B) obtidos pelo modelo BLLS na análise da vitamina B2 no produto Suplemento 1. .... 109

**Figura 39-** Espectros normalizados de excitação (A) e emissão (B) obtidos pelo modelo PARAFAC na análise da vitamina B2 no produto Suplemento 1..... 109

**Figura 40-** Gráficos das concentrações de referência (considerando as estimativas por HPLC) versus estimadas para a vitamina B2 nos produtos onde se utilizou calibração de ordem zero..... 111

**Figura 41-** Gráficos das concentrações de referência (considerando as estimativas por HPLC) versus estimadas para a vitamina B2 para o suplemento 1 utilizando calibração de segunda ordem..... 111

**Figura 42-** Resíduos de decomposição obtidos com o BLLS na análise de cada produto..... 113

<b>Figura 43-</b> Resíduos de decomposição obtidos com o UPLS na análise de cada produto.....	114
<b>Figura 44-</b> Espectros normalizados de excitação (A) e emissão (B) obtidos pelo modelo BLLS na análise da vitamina B6 no leite 1..	116
<b>Figura 45-</b> Espectros normalizados de excitação (A) e emissão (B) obtidos pelo modelo PARAFAC na análise da vitamina B6 no leite 1.....	117
<b>Figura 46-</b> Gráficos das concentrações de referência (considerando as estimativas por HPLC) versus estimadas para a vitamina B6 em cada produto utilizando o modelo BLLS.....	118
<b>Figura 47-</b> Incertezas médias estimadas para as determinações das vitaminas B2 (esquerda) e B6 (direita) nos cinco produtos analisados. ....	121
<b>Figura 48-</b> Estruturas moleculares dos ácidos acetilsalicílico (AAS) e ascórbico (AA).....	127
<b>Figura 49-</b> Esquema do sistema FIA construído. Onde L e D indicam saída do fluxo com as válvulas ligadas ou desligadas, respectivamente e C1, C2 e C2 são pontos de confluência.....	129
<b>Figura 50-</b> Gráfico do planejamento experimental utilizado. ....	131
<b>Figura 51-</b> Cromatogramas obtidos por HPLC para as análises de AAS e AA monitorados em 232 nm. ....	137
<b>Figura 52-</b> Curvas analíticas obtidas a partir das medidas em HPLC.....	138

**Figura 53-** Superfícies obtidas no sistema FIA para (A) AAS, (B) AA e (C) uma mistura contendo os dois analitos nas concentrações de 80 e 48 mg/L para AAS e AA, respectivamente. .... 139

**Figura 54-** Estimativas dos perfis espectrais (A) e de concentração (B) obtidas por MCR para uma amostra do conjunto VAL2. .... 141

**Figura 55-** Alturas dos perfis de concentração obtidos por MCR nas 33 replicatas das soluções de calibração..... 142

**Figura 56-** Valores de referência versus os estimados por FIA utilizando o MCR em todos os conjuntos de dados analisados. .... 145

**Figura 57-** Resíduos de decomposição obtidos com os modelos BLLS e UPLS na análise de cada conjunto de dados. .... 147

**Figura 58-** Concentrações de referência versus estimadas pelo método FIA com BLLS no medicamento 3..... 148

**Figura 59-** Gráfico das concentrações de referência versus estimadas, obtidos com UPLS para AAS nos conjuntos de dados: (A) VAL2, Medicamentos (B) 1 e (C) 2..... 149

**Figura 60-** Valores de referência versus os estimados pelo método FIA com o BLLS executado no sistema Octave para AAS (acima) e AA (abaixo)..... 152

**Figura 61-** Estimativas dos perfis espectrais (A) e de concentração (B) obtidas por BLLS para as amostras do conjunto VAL2. .... 154

**Figura 62-** Valores de referência versus os estimados pelo método FIA com o UPLS executado com o sistema Matlab. .... 156

<b>Figura 63-</b> Incertezas médias obtidas com os modelos MCR e UPLS. (MED) Medicamentos.....	158
<b>Figura 64-</b> Fórmulas estruturais dos quatro pesticidas (CBL, TIO, SIM e DMT) e dos três metabólitos de interesse que foram determinados nas amostras de vinho.....	164
<b>Figura 65-</b> Esquema que ilustra o procedimento empregado no preparo das amostras de vinho.....	168
<b>Figura 66-</b> (A) Matriz de dados formada por <b>N</b> e <b>M</b> colocadas lado a lado (adjuntas) apresentando desvio de tempo, (B) matriz de dados adjunta alinhada pelo deslocamento de <b>M</b> em relação a <b>N</b> até que o mínimo de RES seja atingido, (C) gráfico de RES em função do deslocamento de <b>M</b> em relação a <b>N</b> .....	171
<b>Figura 67-</b> Cromatogramas em 220 nm obtidos no MP. (A) Padrões de calibração (B) extratos.....	172
<b>Figura 68-</b> Cromatogramas em 220 nm no MP para os conjuntos de (—) calibração, (—) validação, (—) extratos, (—) vinho e (—) sinal de uma amostra de vinho sem adição dos analitos, destacando a região onde foi observada sobreposição entre os analitos ou pelos interferentes.....	172
<b>Figura 69-</b> Cromatogramas em 220 nm obtidos no MR. (A) Padrões de calibração (B) extratos.....	173
<b>Figura 70-</b> Cromatogramas em 220 nm no MR para os conjuntos de (—) calibração, (—) extratos e (—) vinho, destacando a região onde foi observada sobreposição entre os analitos ou pelos interferentes.....	174
<b>Figura 71-</b> Resíduos de decomposição obtidos com o modelo BLLS no MP.....	176

**Figura 72-** Cromatogramas em 220 nm para CBL após a correção dos desvios de tempo de retenção. .... 177

**Figura 73-** Variância residual (RES) obtida a partir da correção dos desvios de tempo para as amostras EX e VI na região de eluição dos composto CBL. .... 177

**Figura 74-** Cromatogramas em 220 nm para TIO e SIM após a correção dos desvios de tempo de retenção, onde a região selecionada foi aquela utilizada para a construção dos modelos..... 178

**Figura 75-** Variância residual (RES) obtida a partir da correção dos desvios de tempo para EX e VI na região de eluição dos compostos TIO e SIM. .... 178

**Figura 76-** Cromatogramas em 220 nm para DMT e FTA (A) antes e (B) após a correção dos desvios dos tempos de retenção em relação ao sinal do branco, onde a região selecionada foi aquela utilizada na próxima etapa do pré-processamento.. 179

**Figura 77-** Variância residual (RES), para a na região de eluição de DMT e FTA, obtida a partir da primeira correção dos desvios de tempo realizada para alinhar as amostras EX e VI com o sinal do branco. .... 180

**Figura 78-** Cromatogramas em 220 nm para DMT e FTA após a subtração do sinal do branco (A) antes e (B) depois da correção dos desvios de tempo de retenção em relação aos padrões. .... 181

**Figura 79-** Variância residual (RES), para DMT e FTA, obtida a partir da correção dos desvios dos tempos de retenção em relação a um padrão de calibração..... 182

**Figura 80-** Estimativas dos perfis (A) espectrais e (B) cromatográficos, ambos normalizados, obtidos por BLLS para a determinação de CBL nas amostras EX..... 183

**Figura 81-** Estimativas dos perfis (A) espectrais e (B) cromatográficos, ambos normalizados, obtidos por BLLS para a determinação de TIO e SIM nas amostras EX..... 184

**Figura 82-** Estimativas dos perfis (A) espectrais e (B) cromatográficos, ambos normalizados, obtidos por BLLS para a determinação de DMT e FTA nas amostras EX..... 185

**Figura 83-** Valores de referência versus os estimados pelo MP com o BLLS (TIO, SIM, DMT e FTA) e calibração de ordem zero (DCA e CBZ).. ..... 187

**Figura 84-** Valores de referência versus estimados para DCA pelo MP com BLLS..... 188

**Figura 85-** Cromatogramas em 220 nm para CBZ no MR (A) após a correção da linha de base, (B) após uma correção inicial dos desvios dos tempos das amostras e (C) após alinhamento das amostras de EX e VI em relação padrões..... 192

**Figura 86-** Estimativas dos perfis (A) espectrais e (B) cromatográficos, ambos normalizados, obtidos por PARAFAC para a determinação de CBZ nas amostras EX..... 192

**Figura 87-** Cromatogramas em 220 nm para SIM no MR (A) após correção da linha de base e um prévio alinhamento dos tempos de retenção de EX e VI, (B) após alinhamento do sinal das amostras EX e VI com o sinal do branco, (C) resultado da subtração do sinal do branco do sinal das amostras EX e VI e (D) alinhamento de EX e VI com os padrões..... 193

**Figura 88-** Estimativas dos perfis (A) espectrais e (B) cromatográficos, ambos normalizados, obtidos por PARAFAC para a determinação de SIM nas 10 replicatas das amostras EX..... 194

**Figura 89-** Resultado do alinhamento dos tempos de retenção entre as amostras (—) EX e (—) VI para a determinação de FTA pelo MR, onde a região selecionada foi aquela utilizada nos modelos UPLS e PARAFAC..... 195

**Figura 90-** Resultado do alinhamento dos tempos de retenção entre as amostras (—) EX, (—) VI para a determinação de DMT pelo MR, onde a região selecionada foi aquela utilizada nos modelos UPLS e PARAFAC..... 196

**Figura 91-** Estimativas dos perfis (A) espectrais e (B) cromatográficos, ambos normalizados, obtidos por PARAFAC para a determinação de FTA nas amostras EX e VI..... 197

**Figura 92-** Estimativas dos perfis (A) espectrais e (B) cromatográficos, ambos normalizados, obtidos por PARAFAC para a determinação de DMT nas amostras EX e VI.      197

**Figura 93-** Resíduos de decomposição obtidos com o modelo UPLS no MR.. ... 198

**Figura 94-** Valores de referência versus os estimados pelo MR com o UPLS..... 201

**Figura 95-** Valores de referência versus os estimados para SIM pelo MR com o UPLS..... 202

**Figura 96-** Incertezas médias para cada um dos compostos de interesse, estimadas para os (A) MP e (B) MR..... 205

# **Introdução**

Atualmente a utilização de métodos instrumentais em laboratórios químicos é cada vez mais freqüente, sendo esse crescimento motivado por fatores como a necessidade de métodos que gerem menos resíduos, que forneçam maior quantidade de informação do meio em estudo, que sejam mais rápidos, que necessitem de menor contato ou esforço do analista, etc.

Um fato marcante desse desenvolvimento é que vem sendo possível a aquisição de uma quantidade cada vez maior de informações sobre o meio e em diversos níveis de complexidade. A quantidade e a riqueza dos dados que podem ser adquiridos nesses instrumentos torna possível o entendimento de processos químicos e a análise de espécies de interesse em sistemas complexos. Contudo, a análise dessa quantidade de dados muitas vezes não constitui tarefa simples, sendo freqüentemente necessária a utilização de procedimentos mais elaborados para a obtenção das informações desejadas. A quimiometria é uma área que teve sua criação motivada devido à dificuldade de lidar com esse volume de dados, podendo ser definida como a pesquisa e a aplicação de métodos matemáticos e estatísticos para análise de dados químicos, de forma a extrair maior quantidade de informações e resultados de um conjunto de dados<sup>1</sup>. Suas linhas de pesquisa incluem temas como processamento de sinais, planejamento e otimização de experimentos, reconhecimento de padrões e classificação, construção de modelos de regressão, desenvolvimento e aplicação de sistemas de inteligência artificial voltados para química, etc.

Particularmente em química analítica, na maioria dos casos, a principal meta é a determinação quantitativa da concentração de uma ou mais espécies presentes no meio. Contudo, quando se utilizam técnicas instrumentais de análise, a concentração é uma propriedade que não pode ser observada diretamente, sendo determinada indiretamente por meio de uma relação com outra grandeza, como por exemplo medidas da intensidade de luz absorvida ou emitida, medidas de

condutividade ou da diferença de potencial elétrico, etc., em uma prática conhecida como calibração<sup>2,3</sup>. A calibração pode ser definida como uma série de operações que estabelecem, sob condições específicas, uma relação entre as medidas instrumentais e os valores da propriedade de interesse correspondentes. Os modelos de calibração existentes podem ser divididos, quanto à complexidade ou dimensionalidade dos dados que são utilizados, em calibração de ordem zero, primeira, segunda ordem e assim por diante<sup>4</sup>.

Modelos de calibração de ordem zero consistem de regressões univariadas, que são realizadas quando se utiliza apenas um número ou um escalar como resposta instrumental de cada amostra. Esses modelos são bem estabelecidos e amplamente utilizados rotineiramente em determinações químicas. Em modelos de primeira ordem é utilizado um vetor de dados por amostra resultando, portanto, em uma regressão multivariada. Atualmente esses modelos constituem uma importante linha de pesquisa em quimiometria, sendo que nos últimos anos a implementação desses modelos como ferramentas em programas (softwares) de equipamentos, como por exemplo espectrofotômetros na região do infravermelho próximo, vêm crescendo mediante principalmente quanto ao estabelecimento de métodos e critérios para determinação de suas figuras de mérito. Isso possibilita a validação, aceitação e o reconhecimento desses métodos por órgãos oficiais de fiscalização e normalização, tais como a Farmacopéia Americana (USP, do inglês “United States Pharmacopoeia”)<sup>5,6</sup> no setor farmacêutico, a Sociedade Americana de Testes e Materiais (ASTM, do inglês “American Society for Testing and Materials”)<sup>7</sup> e a União Internacional de Química Pura e Aplicada (IUPAC, do inglês “International Union of Pure and Applied Chemistry”)<sup>8,9</sup>.

Modelos de calibração de segunda ordem são construídos quando é possível obter uma matriz de respostas instrumentais por amostra. Vários desses modelos apresentam a possibilidade de permitir a determinação da concentração de espécies

de interesse em uma amostra mesmo na presença de interferentes desconhecidos e não levados em conta na etapa de construção do modelo, sem necessidade de separação física desses interferentes. Essa propriedade é conhecida como vantagem de segunda ordem<sup>4</sup> e torna esses modelos muito atrativos para o estudo e a análise de sistemas complexos.

Dentre os modelos de calibração de segunda ordem, pode-se citar como principais o modelo de análise de fatores paralelos (PARAFAC, do inglês “Parallel Factor Analysis”)<sup>10</sup> e o de resolução de curvas multivariada (MCR, do inglês “Multivariate Curve Resolution”)<sup>11</sup>. Os modelos de quadrados mínimos bilinear (BLLS, do inglês “Bilinear Least Squares”)<sup>12</sup> e quadrados mínimos parciais desdobrado (UPLS, do inglês “Unfold Partial Least Squares”)<sup>13</sup> vêm sendo recentemente propostos como modelos de calibração de segunda ordem mais simples e capazes de apresentar resultados equivalentes ao PARAFAC e MCR. Entretanto, ambos BLLS e UPLS utilizam a bilinearização de resíduos (RBL, do inglês “Residual Bilinearization”) para adquirir a vantagem de segunda ordem, enquanto que PARAFAC e MCR já possuem essa propriedade. Poucos trabalhos foram publicados até o momento abordando o BLLS e UPLS, de modo que ainda é necessário estudos que verifiquem a eficiência desses modelos e do processo de RBL em diferentes tipos de sistemas químicos e dados instrumentais.

Outro aspecto importante no contexto dos modelos de calibração de segunda ordem é a sua validação, a qual vem sendo tema de recentes pesquisas, na tentativa de propor formas de estimativas de figuras de mérito e solucionar inconsistências observadas para alguns parâmetros como sensibilidade e seletividade. Além disso, o número de trabalhos que determinam tais parâmetros em aplicações com diferentes sistemas químicos ainda é pequeno. Logo, para o desenvolvimento desses modelos, é necessário que um maior número de trabalhos que verifiquem a consistência das estimativas desses parâmetros seja realizado.

Tendo em vista os aspectos descritos, o presente trabalho de tese de doutorado teve como seus objetivos principais a aplicação dos modelos BLLS e UPLS em diferentes sistemas químicos, com dados obtidos em equipamentos diferentes e a comparação da eficiência desses modelos com o PARAFAC ou MCR, dependendo do caso. Uma atenção especial foi dada à determinação do número de fatores utilizados no processo RBL, verificando a concordância da forma como esse número é estimado e a atual proposta descrita na literatura. Para comparação entre os modelos, foram determinadas figuras de mérito como exatidão, precisão, sensibilidade, sensibilidade analítica, seletividade, limite de detecção e incerteza, onde se procurou abordar as propostas recentes descritas na literatura e avaliar sua consistência em cada estudo de caso abordado.

Uma vez que a determinação das figuras de mérito nesses métodos é de extrema importância para sua implementação desses modelos e a disponibilidade de programas que realizem esses cálculos ainda não é usual, foram desenvolvidos nesta tese programas para construção e estimativa de figuras de mérito dos modelos BLLS e UPLS. Para os modelos PARAFAC e MCR foram desenvolvidos programas apenas para estimar as figuras de mérito, uma vez que estes podem ser obtidos gratuitamente através dos professores Rasmus Bro<sup>14</sup> e Romá Tauler<sup>15</sup>, respectivamente. Os programas desenvolvidos são compatíveis com as linguagens de programação Matlab 6.5<sup>16</sup> e Octave 3.0.0<sup>17</sup>, tornando possível assim a utilização de um programa livre e disponível, sendo os códigos fonte apresentados no apêndice B da tese.

A presente tese está dividida em 5 capítulos mais conclusões, perspectivas futuras e as referências utilizadas. No primeiro capítulo, intitulado Modelos de Calibração de Segunda Ordem e Figuras de Mérito, são descritos os modelos PARAFAC, MCR, BLLS e UPLS, suas propriedades e a forma como são determinadas as figuras de mérito em cada modelo.

Nos capítulos de 2 a 5 são discutidos os quatro estudos de caso que diferem em complexidade do sistema químico ou no tipo de dados instrumentais, sendo que nos estudos de caso apresentados nos capítulos de 3 a 5 são analisadas amostras comerciais e os resultados validados através de uma técnica independente.

O capítulo 2 é intitulado Estudo da Determinação de Hidrocarbonetos Policíclicos Aromáticos em Misturas, que teve como objetivo verificar em uma aplicação simples, onde se tem um total conhecimento de todos os componentes presentes no meio, como as figuras de mérito dos modelos são influenciadas pela sobreposição entre dois ou três compostos. Também foi avaliada a maneira como a sobreposição entre as espécies pode estar atuando como interferente em uma hipotética amostra de composição desconhecida ou na forma de determinação simultânea de dois ou três compostos de interesse. Este estudo de caso foi desenvolvido utilizando a técnica de emissão e excitação de fluorescência (EEF), a qual fornece dados que apresentam estrutura adequada aos modelos de calibração de segunda ordem. No capítulo 3, intitulado Determinação de Riboflavina e Piridoxina em Leite em Pó e Suplementos Alimentares, novamente foi utilizada a técnica de emissão e excitação de fluorescência, mas consistindo de um sistema relativamente complexo onde é avaliada a presença de interferentes desconhecidos da amostra.

No capítulo 4, intitulado Determinação Simultânea de Ácido Acetilsalicílico e Ascórbico em Medicamentos, é avaliada a eficiência dos modelos BLLS, UPLS e MCR em um sistema de análise por injeção em fluxo (FIA) com gradiente de pH. Neste sistema são observadas duas espécies em equilíbrio para cada analito, situação que oferece certa dificuldade aos modelos de calibração de segunda ordem.

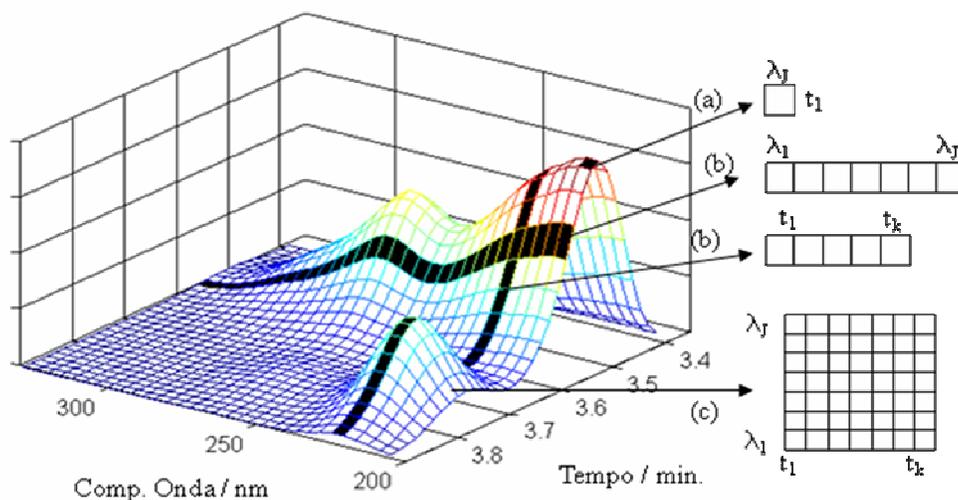
O capítulo 5, intitulado Determinação de Pesticidas e Metabólitos em Vinho Tinto, apresenta resultados do último estudo de caso que trata do desenvolvimento

de um método alternativo para determinação simultânea de quatro pesticidas e três metabólitos por cromatografia líquida de alta eficiência e detecção por arranjo de diodos. Neste sistema, são abordados procedimentos de pré-processamento como correção dos desvios de tempo de retenção e linha de base e comparados os resultados de dois métodos cromatográficos distintos.

Após o capítulo 5 são apresentadas as conclusões da tese, em termos gerais e quanto à eficiência de cada modelo de calibração de segunda ordem abordado, assim como as propriedades observadas para o processo RBL, os programas desenvolvidos, resultados das figuras de mérito estimadas e específicas obtidas em cada estudo de caso. Em seguida são apresentadas perspectivas futuras e as referências que foram utilizadas durante a tese.

# **Capítulo 1 - Modelos de calibração de segunda ordem e figuras de mérito**

Dependendo da instrumentação utilizada na aquisição dos dados instrumentais, estes podem ser constituídos por apenas um número ou escalar, um vetor, uma matriz ou mesmo um cubo de dados, arranjos que, de acordo com a nomenclatura comumente utilizada em álgebra linear, são denominados de tensores de ordem zero, primeira, segunda e terceira ordens, respectivamente<sup>4</sup>. A Figura 1 mostra uma representação esquemática dos tipos de dados que são utilizados em calibração de ordem zero, primeira e segunda ordem.



**Figura 1-** Representação esquemática dos três tipos de dados em um sistema que adquire espectros em função do tempo. (a) de ordem-zero, representado por um escalar; (b) de primeira-ordem, vetor de dados; (c) segunda-ordem, matriz de dados.

Em modelos de calibração de ordem zero um único valor escalar é obtido por amostra, sendo então uma calibração univariada. Segundo o exemplo mostrado na Figura 1 ele é representado por um ponto da superfície, que corresponde à leitura da absorvância em um comprimento de onda  $\lambda_1$  e em um tempo  $t_1$ . Exemplos típicos de instrumentos que geram esse tipo de dados incluem medidas potenciométricas, voltamétricas, fotométricas, espectrofotométricas, etc., com o monitoramento de um único ponto ou variável. As calibrações de ordem zero, por sua simplicidade, são as mais aplicadas em análises de rotina e dispõem de procedimentos de validação bem estabelecidos e descritos na literatura<sup>18,19</sup>.

Contudo, sua aplicação requer que a grandeza medida diretamente no sistema (ex.: absorvância) esteja livre de interferentes que possam provocar desvios em sua relação com a propriedade de interesse.

Métodos de calibração de primeira ordem fazem uso de um vetor de medidas instrumentais para cada amostra. Pelo exemplo da Figura 1, são possíveis dois tipos de dados: monitorando a absorvância em um comprimento de onda  $\lambda_i$  nos tempos de  $t_1$  a  $t_k$ , ou monitorar as absorvâncias dos comprimentos de onda  $\lambda_1$  a  $\lambda_j$  em um determinado tempo. Exemplos de dados multivariados incluem medidas espectrométricas e eletroquímicas em que se monitora a correspondente propriedade em um determinado intervalo espectral ou de variação de potencial, etc. Esses métodos possibilitam análises mesmo na presença de interferentes (desde que esses interferentes estejam presentes nas amostras de calibração), determinações simultâneas e determinações quando os sinais dos analitos estão sobrepostos entre si ou com os sinais dos interferentes calibrados. Diversos modelos de calibração multivariada vêm sendo utilizados, tais como: Regressão Linear Múltipla (MLR, do inglês “Multiple Linear Regression”), Regressão por Componentes Principais (PCR, do inglês “Principal Componentes Regression”) e Regressão por Quadrados Mínimos Parciais (PLS, do inglês “Partial Least Squares”). Esses modelos têm apresentado ótimos resultados com aplicações em diversas áreas, sendo um exemplo marcante de sua utilização a aplicação a dados de infravermelho próximo em análise de bebidas, madeira, polímeros, produtos farmacêuticos, agrícolas, etc.<sup>20</sup>. Porém, sua utilização possui a desvantagem de requerer um número relativamente grande de amostras de calibração dependendo da situação. O reconhecimento e a implementação desses modelos vêm crescendo nos últimos anos graças às pesquisas realizadas nessa área e o estabelecimento de critérios de validação<sup>8,9,21,22,23</sup>.

Calibrações de segunda ordem são construídas para métodos que geram uma

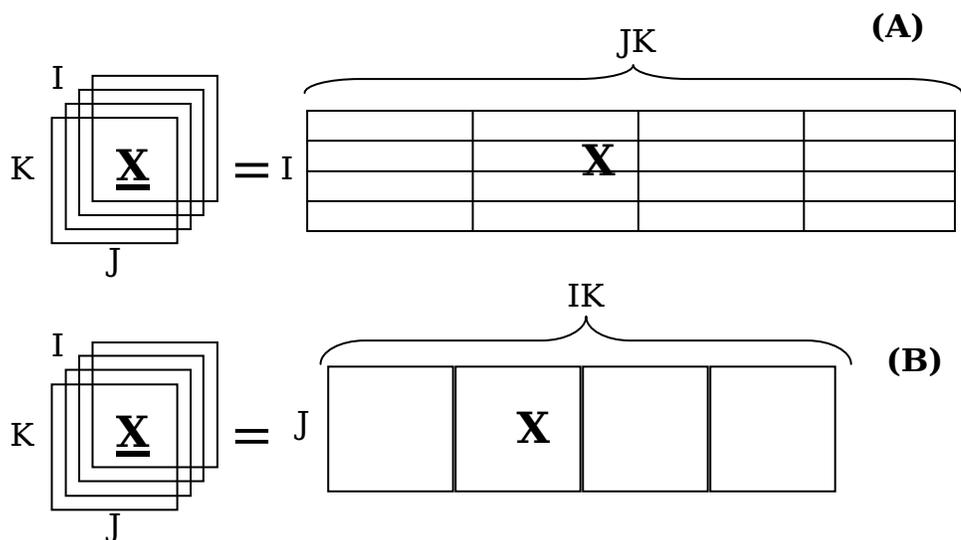
matriz de dados por amostra. Como é mostrado na Figura 1, é obtida uma matriz de dimensões (J, K), que define uma superfície para cada amostra. Caso sejam analisadas I amostras, tem-se um cubo de dados  $\underline{X}$  de dimensões (I, J, K). Esse tipo de dados pode ser gerado através de diversas técnicas, como cromatografia líquida de alta eficiência (HPLC, do inglês “High Performance Liquid Chromatography”) com detecção por arranjo de diodos (DAD, do inglês “Diode Array Detection”); cromatografia gasosa (GC, do inglês “Gas Chromatography”) com detecção por espectrometria de massas (MS, do inglês “Mass Spectrometry”); ressonância nuclear magnética em duas dimensões (NMR-2D, do inglês “Two Dimensional Nuclear Magnetic Resonance”); emissão e excitação de fluorescência molecular (EEF, do inglês “Excitation Emission Fluorescence”); análise por injeção em fluxo (FIA, do inglês “Flow Injection Analysis”) com gradiente de pH e DAD; dados de microscopias e imagens em geral, etc. Esses métodos têm a grande vantagem de permitir a determinação de espécies de interesse na presença de interferentes, mesmo que estes interferentes não tenham sido incluídos nas amostras de calibração, característica conhecida como “vantagem de segunda ordem”<sup>4</sup>. Além disso, o perfil característico (espectral ou cromatográfico, por exemplo) de cada composto independente presente na amostra pode ser estimado com dados de segunda ordem fornecendo, dessa forma, informações qualitativas sobre o sistema. Outra facilidade que esses modelos apresentam se refere ao número de amostras requerido para a construção das regressões, que é sensivelmente menor que aquele necessário para modelos de primeira ordem e aproximadamente igual ao utilizado em calibrações de ordem zero.

A seguir são descritos os modelos de calibração de segunda ordem abordados nessa tese e a forma como foram determinadas algumas de suas figuras de mérito. A notação utilizada na simbologia para a apresentação desses modelos segue o padrão geralmente utilizado em álgebra linear, onde letras maiúsculas em negrito e

sublinhadas representarão cubos de dados, matrizes serão indicados por maiúsculas em negrito, vetores por letras minúsculas em negrito e escalares por letras normais maiúsculas ou minúsculas.

## 1.1 Modelos de calibração de segunda ordem

Dependendo do modelo a ser utilizado na análise dos dados, estes devem ser organizados na forma de um cubo de dimensão (I, J, K) ou uma matriz, onde I é o número de amostras e J e K são os números de variáveis na primeira e na segunda dimensão, respectivamente. Na Figura 2 são exemplificados esses dois modos de organização ilustrando a conversão de um cubo de dados  $\underline{\mathbf{X}}$  em uma matriz  $\mathbf{X}$ .



**Figura 2-** Representação de duas formas em que um cubo de dados pode ser arranjado para a análise em diferentes modelos de calibração de segunda ordem. (A) um cubo  $\underline{\mathbf{X}}(I,J,K)$  é colocado na forma de uma matriz  $\mathbf{X}(I,JK)$  pela vetorização da matriz uma das amostras I e (B)  $\underline{\mathbf{X}}(I, J, K)$  é re-arranjado em uma matriz  $\mathbf{X}(J,IK)$  pela colocação de cada uma das I amostras lado a lado.

### 1.1.1 Análise de fatores paralelos (PARAFAC)

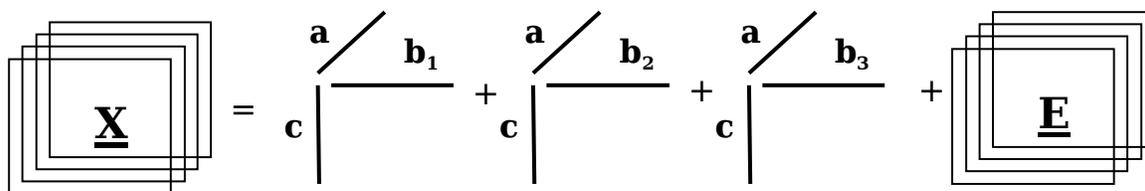
O modelo PARAFAC efetua uma decomposição do cubo de dados  $\underline{\mathbf{X}}(I, J, K)$  em três matrizes de pesos:  $\mathbf{A}(I, F)$ ,  $\mathbf{B}(J, F)$  e  $\mathbf{C}(K, F)$ , onde F é o número de fatores ou componentes do modelo, e as matrizes de peso  $\mathbf{A}$ ,  $\mathbf{B}$  e  $\mathbf{C}$  são os perfis de cada componente em cada dimensão, por exemplo, as estimativas dos espectros de

emissão e excitação de fluorescência. Esta decomposição pode ser representada pelas equações 1, 2 ou esquematicamente pela Figura 3<sup>10</sup>.

$$x_{I,J,K} = \sum_{i=1}^F a_{I,F} b_{J,F} c_{K,F} + e_{I,J,K} \quad (1)$$

$$\mathbf{X} = \mathbf{A}(\mathbf{C} | \otimes | \mathbf{B})^T + \mathbf{E} \quad (2)$$

onde,  $x_{I,J,K}$  é um elemento do cubo de dados,  $a$ ,  $b$  e  $c$  são os elementos das matrizes  $\mathbf{A}_{I,F}$ ,  $\mathbf{B}_{J,F}$  e  $\mathbf{C}_{K,F}$ ,  $e_{I,J,K}$  contém os erros de decomposição,  $\mathbf{X}_{I,JK}$  é a matriz de dados obtida depois da vetorização do cubo de dados (Figura 2A),  $| \otimes |$  é o operador do produto de Khatri-Rao (ver apêndice A2) e “T” sobrescrito a operação de transposição de um vetor ou matriz.



**Figura 3-** Representação da decomposição trilinear dos dados nos perfis  $\mathbf{a}$ ,  $\mathbf{b}$ , e  $\mathbf{c}$  de cada dimensão, para um modelo PARAFAC de três fatores.

Para a aplicação do modelo PARAFAC, os seguintes aspectos devem ser levados em conta: (1) a concordância da estrutura dos dados e a assumida pelo modelo, (2) o método de inicialização do algoritmo, as restrições impostas ao modelo e o critério de convergência, (3) o número de componentes ou fatores, (4) identificar os perfis fornecidos pelo modelo em cada fator com a espécie de interesse e os interferentes presentes e (5) construção de um modelo de regressão que permita estimar a concentração da espécie de interesse.

Com relação à estrutura requerida pelo PARAFAC, considerando que o perfil relacionado à concentração de um componente  $F$  seja  $\mathbf{a}$ , é necessário que cada fator apresente um único perfil  $\mathbf{b}$  e  $\mathbf{c}$ . Esta restrição traz como consequência, por exemplo, que em dados obtidos por HPLC ou FIA, ambos com DAD, não

podem ocorrer variações nos tempos de retenção ou dos tempos de injeção, respectivamente. Estes efeitos, para o modelo PARAFAC, representariam um componente com mais de um perfil na referida dimensão temporal, o que é mencionado muitas vezes como desvios da trilinearidade dos dados. Como forma de contornar essa limitação, uma variação do modelo PARAFAC, chamada de PARAFAC2<sup>24</sup>, permite variações em um dos três perfis **A**, **B** ou **C**. Contudo, poucos trabalhos foram desenvolvidos aplicando esse modelo<sup>25,26</sup>, sendo a tentativa da correção desses desvios de trilinearidade mais comum nos trabalhos presentes na literatura<sup>27,28,29</sup>.

Na decomposição dos dados realizada pelo PARAFAC é utilizado o algoritmo de quadrados mínimos alternantes (ALS, do inglês “Alternating Least Squares”), o qual necessita de uma estimativa inicial dos perfis das dimensões (**B** e **C**) para sua inicialização. Diversos métodos podem ser utilizados para a obtenção desses perfis, sendo os mais utilizados os vetores singulares obtidos por decomposição em valores singulares (SVD, do inglês “Singular Value Decomposition”) ou perfis obtidos por decomposição trilinear direta (DTLD, do inglês “Direct Trilinear Decomposition”)<sup>30,31</sup>. Uma vez que uma estimativa inicial de **B** e **C** é obtida, uma estimativa de **A** pode ser obtida por quadrados mínimos como<sup>31,32</sup>:

$$\mathbf{A} = \mathbf{X} \left[ (\mathbf{C} | \otimes | \mathbf{B})^T \right]^+ \quad (3)$$

onde “+” o símbolo sobrescrito indica a Moore-Penrose pseudo-inversa da matriz **X**, que se encontra desdobrada como ilustrado na Figura 2A. Deste modo, o algoritmo ALS pode ser descrito de forma geral para um modelo PARAFAC com F componentes como<sup>31,32</sup>:

1) Inicialize **B** e **C**;

2) Obtenha **A**, por meio de **X**, **B** e **C** como:  $\mathbf{A} = \mathbf{X} \left[ (\mathbf{C} | \otimes | \mathbf{B})^T \right]^+ ;$

3) Obtenha  $\mathbf{B}$ , por meio de  $\mathbf{X}$ ,  $\mathbf{C}$  e a estimativa anterior de  $\mathbf{A}$  como:

$$\mathbf{B} = \mathbf{X}[(\mathbf{C} | \otimes | \mathbf{A})^T]^{\dagger};$$

4) Obtenha  $\mathbf{C}$ , por meio de  $\mathbf{X}$  e as estimativas de  $\mathbf{B}$  e  $\mathbf{A}$  dos passos anteriores como:

$$\mathbf{C} = \mathbf{X}[(\mathbf{B} | \otimes | \mathbf{A})^T]^{\dagger};$$

5) Retorne ao passo 1 utilizando  $\mathbf{B}$  e  $\mathbf{C}$  obtidos nos passos 3 e 4 até que a falta de ajuste dos dados entre duas iterações consecutivas seja menor que um valor crítico (geralmente  $10^{-6}$ ). A falta de ajuste (FAJ) é dada por<sup>32</sup>:

$$\text{FAJ} = \left( \frac{\sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^k (x_{i,j,k} - \hat{x}_{i,j,k})^2}{\sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^k (x_{i,j,k})^2} \right) \quad (4)$$

onde  $\hat{x}_{i,j,k}$  corresponde à estimativa do modelo para a resposta instrumental  $x_{i,j,k}$ .

As soluções do algoritmo ALS descrito podem ser muitas vezes melhoradas por meio da inclusão de restrições como: não-negatividade, o que faz com que os perfis não possam assumir valores negativos; ortogonalidade e unimodalidade, o que requer perfis com apenas um máximo<sup>10</sup>.

Dependendo da complexidade do sistema, a escolha do número de componentes pode ser feita com base no conhecimento prévio do número de espécies responsáveis pelo sinal instrumental medido. Quando isso não é possível, um parâmetro útil, conhecido como consistência do núcleo (do inglês “core consistency”), pode ser utilizado como guia. Na prática, o modelo PARAFAC é construído com um número crescente de componentes e este parâmetro é determinado em cada um, o número de componentes é então escolhido como um a menos daquele que o valor da consistência do núcleo cai de aproximadamente 100% para valores menores que 50%<sup>31</sup>. A utilização deste parâmetro é bastante útil quando os dados seguem perfeitamente o modelo trilinear, como por exemplo em

dados de EEF. Contudo, em sistemas que apresentam espécies em equilíbrio a utilização da consistência do núcleo não conduz ao número de componentes correto.

Uma vez obtidos os perfis **A**, **B** e **C**, a identificação do componente que corresponde à espécie de interesse é feita pela comparação entre os perfis obtidos e espectros conhecidos obtidos em padrões. Caso essa identificação não puder ser feita enquanto se procura o número ideal de componentes, a qualidade do ajuste da reta entre os pesos da coluna de **A** correspondente à espécie de interesse e as concentrações de referência para uma série de amostras de concentração conhecida pode ser utilizada como critério de escolha do número de componentes, pois espera-se que o número ideal de componentes forneça valores de **A** que se relacionem melhor com a concentração da espécie de interesse<sup>33</sup>.

O modelo de regressão para o PARAFAC é obtido através de uma regressão por quadrados mínimos entre a coluna de **A** relacionada com a concentração da espécie de interesse (**a**) e o vetor com as concentrações de referência (**y**) das amostras de calibração<sup>12</sup>.

$$\omega = \mathbf{y}^+ \mathbf{a} \quad (5)$$

onde,  $\omega$  corresponde à inclinação da reta ajustada entre os pesos **a** e as concentrações **y**. Na prática, para a análise de uma amostra de composição desconhecida, o cubo de dados é formado por  $I_c$  amostras de concentração conhecida da espécie de interesse (amostras de calibração) e uma amostra de composição desconhecida, que pode conter interferentes presentes ou não nas amostras de calibração, constituindo assim a vantagem de segunda ordem. Por fim, a concentração da amostra de composição desconhecida é obtida por<sup>12</sup>:

$$\hat{y}_{un} = \frac{\mathbf{a}}{\omega} \quad (6)$$

### 1.1.2 Resolução de curvas multivariada (MCR)

O MCR pode ser visto como um método de resolução de sinais que também pode ser utilizado para propósitos quantitativos como um modelo de calibração de segunda-ordem. Diversos trabalhos já demonstraram a sua capacidade em fornecer ambas as informações quantitativas como qualitativas para um sistema em estudo<sup>34,35,36,37,38</sup>. O MCR é baseado na decomposição bilinear dos dados realizada pelo algoritmo ALS, a qual pode ser matematicamente expressa por<sup>34,39</sup>:

$$\mathbf{X}_i = \mathbf{CS}^T + \mathbf{E} \quad (7)$$

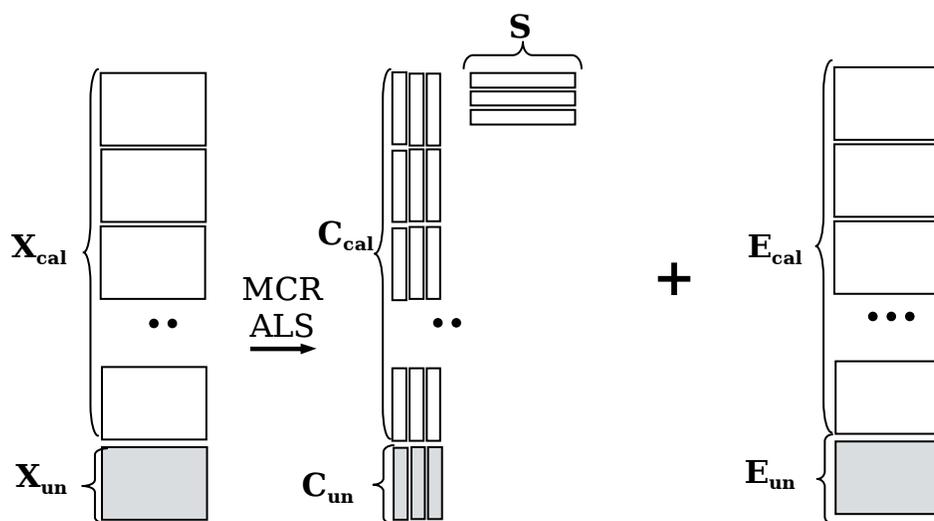
onde  $\mathbf{X}_i$  é uma matriz de dados para uma amostra “i”,  $\mathbf{C}$  e  $\mathbf{S}$  são as estimativas dos perfis relacionados com a concentração e os espectros das espécies químicas puras que estão presentes na mistura, respectivamente e  $\mathbf{E}$  é uma matriz de resíduos. Para a aplicação do MCR é necessária uma estimativa do número de espécies presentes na amostra e uma estimativa inicial dos perfis de concentração ou espectral para cada constituinte.

Ao contrário do PARAFAC, o MCR pode ser aplicado tanto na análise de várias amostras (mais de uma matriz de dados) simultaneamente quanto de uma única amostra (quando o objetivo seja a resolução ou deconvolução dos sinais), enquanto que o PARAFAC requer a presença de no mínimo 2 amostras diferentes. Essa diferença reside no fato de que o modelo MCR baseia-se na decomposição bilinear dos dados, enquanto que o PARAFAC utiliza uma decomposição trilinear, que requer mais de uma amostra para formar um cubo de dados. Uma das grandes conseqüências desse fato é que o MCR apresenta o problema de liberdade de rotação ou ambigüidade, isto é, mais de um conjunto de perfis que apresenta um mesmo ajuste aos dados pode ser obtido. Essa propriedade do MCR pode ser verificada por<sup>31</sup>:

$$\mathbf{CS}^T = \mathbf{CZZ}^{-1}\mathbf{S}^T \quad (8)$$

onde  $\mathbf{Z}$  é qualquer matriz não singular de dimensão  $(F,F)$ . As matrizes  $\mathbf{CZ}$  e  $\mathbf{Z}^{-1}\mathbf{S}^T$  representam outro conjunto de perfis nas dimensões  $J$  e  $K$ , mas apresentam exatamente o mesmo ajuste dos dados obtido com  $\mathbf{C}$  e  $\mathbf{S}^T$ . Esse problema é minimizado pela utilização de diversas restrições (descritas abaixo). O PARAFAC por basear-se em uma decomposição trilinear não apresenta esse tipo de problema de ambigüidade <sup>10</sup>.

Quando o MCR é aplicado na decomposição de mais de uma matriz, estas são organizadas de forma adjunta, uma abaixo da outra ou uma ao lado da outra dependendo das estimativas iniciais serem os perfis espectrais ou de concentração <sup>38</sup>. A Figura 4 mostra a forma como as matrizes são organizadas no caso de estimativas iniciais para os perfis espectrais. Deste modo, obtém-se um conjunto de espectros comum a todas as matrizes e um conjunto de perfis de concentração para cada amostra. Caso o algoritmo seja inicializado com os perfis de concentração o contrário é obtido.



**Figura 4-** Esquema de decomposição do MCR quando se inicializa a otimização a partir dos perfis espectrais. Onde os sub-índices “cal” e “un” indicam amostras de calibração e de previsão, respectivamente;  $\mathbf{X}$ ,  $\mathbf{C}$ ,  $\mathbf{S}$  e  $\mathbf{E}$  são as matrizes de dados, perfis de concentração, espectrais e resíduos de decomposição, respectivamente.

O número de componentes pode ser estimado utilizando o conhecimento prévio do sistema em estudo ou a partir dos resultados da decomposição em valores singulares (SVD) da matriz de dados constituída das amostras utilizadas na etapa de calibração. Na estimativa por SVD, o número de espécies é aproximado para o número de valores singulares acima do valor singular correspondente ao nível do ruído instrumental dos dados. Constituindo assim o número de componentes a princípio necessário para a descrição de toda a informação sistemática dos dados<sup>35</sup>.

As estimativas iniciais dos perfis que serão utilizados na inicialização do algoritmo podem ser obtidas por meio de medidas instrumentais com padrões, quando se dispuser de soluções das espécies de interesse separadamente. Quando isso não é possível, estimativas para perfis espectrais podem ser obtidas pelo método de pureza de variáveis<sup>40</sup>, enquanto que para perfis de concentração utiliza-se a análise evolucionária de fatores (EFA, do inglês “Evolving Factor Analysis”)<sup>41</sup>. A utilização da estimativa inicial dos perfis espectrais ou de concentração depende do conjunto de dados em questão e se estimativas iniciais melhores podem ser obtidas para os perfis espectrais ou de concentração. Dependendo das estimativas iniciais serem os perfis de concentração ou espectrais, o algoritmo ALS começa a otimização pelas equações 9 ou 10, respectivamente<sup>35,36</sup>:

$$\mathbf{S}^T = \mathbf{C}^+ \mathbf{X} \quad (9)$$

$$\mathbf{C} = \mathbf{X}(\mathbf{S}^T)^+ \quad (10)$$

Essas etapas são repetidas em um processo iterativo até que a mudança do ajuste entre duas iterações consecutivas seja menor que um valor crítico ou um número limite de iterações seja atingido<sup>35,39</sup>. Na Figura 4 é mostrada de forma esquemática a decomposição realizada pelo MCR em um conjunto de matrizes quando se inicializa o algoritmo com estimativas dos perfis espectrais.

Caso o MCR seja aplicado na decomposição de apenas uma matriz, restrições como: não negatividade, unimodalidade podem ser impostas ao modelo para a obtenção de soluções com melhor sentido físico. Caso várias amostras estiverem presentes, outras restrições podem ser impostas, tais como: posto local e trilinearidade<sup>35,36,38</sup>.

A restrição de posto local é utilizada quando se tem conhecimento previamente que determinado componente está ausente em uma ou mais amostras. Neste caso, esta informação é passada ao modelo fazendo com que o perfil de concentração deste componente seja igual a zero. Esta restrição é muito útil nos casos onde amostras contendo interferentes são decompostas conjuntamente como amostras de calibração, onde se requer que o perfil de concentração do interferente deva ser zero<sup>35,36,38</sup>.

A trilinearidade consiste em forçar que os perfis espectrais e de concentração não variem entre as diferentes amostras decompostas conjuntamente, o que é esperado no caso de dados de EEF. Se essa restrição não é imposta, o modelo permite que haja variações entre os perfis de uma das dimensões, como por exemplo o de concentração, na forma como os dados estão ilustrados na Figura 4. A imposição ou não da trilinearidade faz com que o MCR seja particularmente interessante em dados onde essas variações podem ocorrer, como por exemplo em dados de HPLC-DAD, onde são permitidas pequenas variações nos tempos de retenção.

Como exposto anteriormente, o modelo MCR é mais sujeito à imposição de restrições que o modelo PARAFAC, o que pode ser visto como algo positivo do ponto de vista de que o MCR pode ser otimizado especificamente para o tipo de dados em questão. Contudo, otimização dessas restrições requer um maior conhecimento do modelo e diversas vezes não consiste de uma tarefa rápida ou mesmo fácil. Logo, a maior flexibilidade do MCR pode ser considerada ao mesmo

tempo uma vantagem e uma desvantagem, dependendo do caso.

Uma vez realizada a decomposição das matrizes, o processo de calibração é feito com base nos perfis recuperados que contem a informação relacionada à concentração da espécie de interesse. Para o exemplo mostrado na Figura 4, são obtidos espectros normalizados e perfis de concentração que tem sua intensidade proporcional à concentração. Logo, uma regressão pseudo-univariada pode ser realizada com base na altura do máximo observado nos perfis de concentração, sua área ou a norma, e a concentração determinada a partir dessa regressão<sup>42</sup>.

### 1.1.3 Quadrados Mínimos Bilinear (BLLS)

Descrito por Linder e Sundberg em 1998<sup>43</sup>, o BLLS aparentemente não despertou grande interesse nos demais grupos de pesquisa em quimiometria, uma vez que dois outros trabalhos utilizando este modelo foram publicados apenas em 2002<sup>44,45</sup>, onde um deles tinha como autores seus próprios proponentes<sup>44</sup>. Contudo, a partir de 2004 diversos trabalhos vêm sendo publicados utilizando este modelo em diversos tipos de dados, os quais têm demonstrado que o BLLS é um modelo capaz de fornecer resultados equivalentes aos modelos PARAFAC e MCR em sistemas relativamente complexos<sup>12,46,47,48</sup> e com desvios de trilinearidade<sup>49</sup>. Ao mesmo tempo, outros trabalhos têm relatado avanços com este modelo envolvendo uma modificação do algoritmo para que este seja capaz de lidar com sistemas apresentando espécies em equilíbrio<sup>49,50</sup>, seleção de variáveis por meio de algoritmos genéticos<sup>51</sup>, o desenvolvimento de uma proposta para a determinação de sua sensibilidade em situações em que a vantagem de segunda ordem é aplicada<sup>52</sup> e a sua generalização para o modelo de quadrados mínimos trilinear (TLLS, do inglês “Trilinear Least Squares”)<sup>53</sup>, o qual é aplicado quando se tem um cubo de dados por amostra.

No BLLS, em contraste com o PARAFAC e o MCR, as concentrações dos

compostos de interesse são inseridas na análise dos dados instrumentais na etapa de calibração, onde obrigatoriamente apenas matrizes de composição conhecida devem ser utilizadas. Primeiramente, cada matriz de calibração  $\mathbf{X}$  é vetorizada e as  $I_c$  amostras de calibração são agrupadas em uma matriz  $\mathbf{V}_X$  de dimensão  $(JK, I_c)$  <sup>12,43,44,49,54</sup>:

$$\mathbf{V}_X = [\text{vec}(\mathbf{X}_1) | \text{vec}(\mathbf{X}_2) | \dots | \text{vec}(\mathbf{X}_{I_c})] \quad (11)$$

onde “vec” indica a operação de vetorização da matriz  $\mathbf{X}$  em questão na forma de um vetor coluna  $(JK,1)$ . Em seguida, através de um procedimento de quadrados mínimos clássico determina-se a matriz que contém as estimativas dos perfis “puros” dos compostos de interesse em concentração unitária ( $\mathbf{V}_s$ ) <sup>49,54</sup>:

$$\mathbf{V}_s = \mathbf{V}_X \mathbf{Y}^{T+} \quad (12)$$

onde  $\mathbf{Y}$  é a matriz que contém as concentrações de referência das  $N_c$  espécies de interesse. É interessante notar que a equação 12 é análoga a uma regressão univariada para uma espécie de interesse aplicando a lei de Beer-Lambert com quadrados mínimos clássicos expressa por:

$$\epsilon \mathbf{b} = \mathbf{Abs}(\mathbf{c}^+) \quad (13)$$

onde  $\mathbf{V}_x$  corresponderia aos dados de absorvância ( $\mathbf{Abs}$ ),  $\mathbf{y}$  as concentrações da espécie de interesse ( $\mathbf{c}$ ) e  $\mathbf{V}_s$  corresponde às absortividades molares ( $\epsilon$ ) multiplicadas pelo caminho óptico ( $\mathbf{b}$ ). Voltando à equação 12,  $\mathbf{V}_s$  contém as matrizes  $\mathbf{S}_n$  (as superfícies de “absortividades molares” de cada espécie presente na amostra) na forma de um vetor.

$$\mathbf{V}_s = [\text{vec}(\mathbf{S}_1) | \text{vec}(\mathbf{S}_2) | \dots | \text{vec}(\mathbf{S}_{N_c})] \quad (14)$$

De modo a obter os perfis nas dimensões J e K a partir das matrizes  $\mathbf{S}_n$ , um estimador baseado em SVD é empregado em cada matriz  $\mathbf{S}_n$  <sup>12,43,44,49,54</sup>:

$$(\mathbf{b}_n, \mathbf{g}_n, \mathbf{c}_n) = \text{SVD}(\mathbf{S}_n) \quad (15)$$

onde  $\mathbf{g}_n$  é o primeiro valor singular,  $\mathbf{b}_n$  e  $\mathbf{c}_n$  são os primeiros vetores singulares

direito e esquerdo de  $\mathbf{S}_n$ , respectivamente. Para dados que não possuem deficiência de posto, os perfis dos componentes são obtidos apenas pelos primeiros valores singulares  $g_n$ , e vetores singulares  $\mathbf{b}_n$  e  $\mathbf{c}_n$ . Quando existem espécies em equilíbrio, Goicoechea e Olivieri<sup>49</sup> propuseram que a presença de todas as espécies presentes é considerada a partir dos valores de  $g_n$ ,  $\mathbf{b}_n$  e  $\mathbf{c}_n$  seguintes. Por exemplo, caso existam três espécies de um mesmo componente F em  $\mathbf{S}_n$ , seus perfis são estimados pelos três primeiros valores e vetores singulares da decomposição  $\mathbf{S}_n$  por SVD<sup>49,50</sup>.

A obtenção dos perfis  $\mathbf{b}_n$ ,  $\mathbf{c}_n$  e  $g_n$  de cada espécie de interesse completa a etapa de calibração. As concentrações das espécies de interesse em uma amostra com composição desconhecida ( $\mathbf{X}_{un}$ ) são estimadas, contanto que não haja interferentes presentes, por outro procedimento de quadrados mínimos clássico<sup>49,54</sup>:

$$\hat{\mathbf{y}}_u = \mathbf{S}_{cal}^+ \text{vec}(\mathbf{X}_u) \quad (16)$$

onde  $\mathbf{y}_{un}$  é o vetor  $1 \times N_c$  que contém as concentrações estimadas para os  $N_c$  analitos em  $\mathbf{X}_{un}$ , e  $\mathbf{S}_{cal}(JK, N_c)$  é uma matriz que contém as informações da regressão obtidas na etapa de calibração, definida por<sup>49,54</sup>:

$$\mathbf{S}_{cal} = [g_1(\mathbf{c}_1 \otimes \mathbf{b}_1) | g_2(\mathbf{c}_2 \otimes \mathbf{b}_2) | \dots | g_{N_c}(\mathbf{c}_{N_c} \otimes \mathbf{b}_{N_c})] \quad (17)$$

onde  $\otimes$  é o operador do produto de Kronecker (ver apêndice A1). Em sistemas com espécies em equilíbrio, a matriz  $\mathbf{S}_{cal}$  contém as informações de todas as espécies, por exemplo, considerando o primeiro analito com três espécies em equilíbrio e o segundo com apenas uma a matriz  $\mathbf{S}_{cal}$  é calculada como<sup>49</sup>:

$$\mathbf{S}_{cal} = [g_{1,1}(\mathbf{c}_{1,1} \otimes \mathbf{b}_{1,1}) | g_{1,2}(\mathbf{c}_{1,2} \otimes \mathbf{b}_{1,2}) | g_{1,3}(\mathbf{c}_{1,3} \otimes \mathbf{b}_{1,3}) | g_2(\mathbf{c}_2 \otimes \mathbf{b}_2)] \quad (18)$$

Diferente do PARAFAC e do MCR, quando existe a presença de interferentes presentes em  $\mathbf{X}_{un}$ , a vantagem de segunda ordem é obtida por um processo posterior à etapa de calibração de bilinearização dos resíduos deixados pelo modelo (RBL, do inglês “Residual Bilinearization”). O RBL é empregado para encontrar os perfis dos interferentes, os quais são incorporados em uma versão

estendida da matriz  $\mathbf{S}_{\text{cal}}$ <sup>49,54</sup>:

$$\mathbf{S}_{\text{int}} = [\mathbf{S}_{\text{cal}} \mid \mathbf{g}_{\text{int}} (\mathbf{c}_{\text{int}} \otimes \mathbf{b}_{\text{int}})] \quad (19)$$

onde  $\mathbf{g}_{\text{int}}$ ,  $\mathbf{b}_{\text{int}}$  e  $\mathbf{c}_{\text{int}}$  são obtidos por SVD de uma matriz de resíduos ( $\mathbf{E}_{\text{un}}$ ), estimada pela diferença entre a dados instrumentais ( $\mathbf{X}_{\text{un}}$ ) e a soma da contribuição dos vários componentes  $N_c$  levados em conta na etapa de calibração<sup>12,49,54</sup>:

$$\mathbf{E}_{\text{un}} = \mathbf{X}_{\text{un}} - \sum_{n=1}^{N_c} \mathbf{g}_n \mathbf{b}_n (\mathbf{c}_n^T) \mathbf{y}_{\text{un},n} \quad (20)$$

$$(\mathbf{b}_{\text{int}}, \mathbf{g}_{\text{int}}, \mathbf{c}_{\text{int}}) = \text{SVD}(\mathbf{E}_{\text{un}}) \quad (21)$$

A concentração na amostra  $\mathbf{X}_{\text{un}}$  é então estimada a partir por:

$$\hat{\mathbf{y}}_{\text{un}} = \mathbf{S}_{\text{int}}^+ \text{vec}(\mathbf{X}_{\text{un}}) \quad (22)$$

O processo RBL pode ser realizado a partir de um procedimento iterativo<sup>44,54,55</sup>, onde as equações de 19 a 21 são repetidas até que a diferença entre a falta de ajuste ou os valores estimados das concentrações dos analitos entre duas iterações consecutivas seja menor que um valor crítico. Outra forma de realizar o processo RBL, proposta por Olivieri et al.<sup>49,54</sup>, é a utilização método de Gauss-Newton para a minimização dos resíduos  $\mathbf{E}_{\text{un}}$ , que segundo os trabalhos encontrados na literatura apresenta melhores resultados para dados que apresentem espécies de equilíbrio<sup>49,50</sup>.

Para que a estimativa da concentração das espécies de interesse na amostra seja feita de forma correta pelo modelo, é necessário que a presença de interferentes em  $\mathbf{X}_{\text{un}}$  seja identificada, assim como o número de componentes necessários para sua modelagem pelo processo RBL deve ser otimizado. Os resultados publicados na literatura até o momento sugerem que isso pode ser feito a partir da comparação dos resíduos de decomposição obtidos para a amostra  $\mathbf{X}_{\text{un}}$  com uma estimativa do desvio padrão do ruído instrumental, obtida por replicatas de amostras do branco, onde o resíduo de decomposição (RD) de  $\mathbf{X}_{\text{un}}$  é estimado

por<sup>46,47,49,50</sup>:

$$RD_{un} = \frac{\|\text{vec}(\mathbf{E}_{un})\|}{(JK - N_c)} \quad (23)$$

Onde  $\|\ \ \|$  representa a norma euclidiana de um vetor. Resíduos maiores que nível do ruído instrumental correspondem à presença de interferentes. Os trabalhos publicados propõem que o modelo deve ser construído adicionando fatores ao RBL até que o resíduo de decomposição se estabilize em um nível compatível com o nível do ruído instrumental<sup>46,47,49,50</sup>. Contudo, não é sugerido nenhum valor crítico com qual o resíduo de decomposição possa ser comparado e considerado compatível ou não com o nível do ruído instrumental ( $s_{\text{ruído}}$ ), tornando duvidosa a escolha do número de fatores do RBL em certas situações.

A maneira mais simples de se estabelecer um critério mais determinístico para a escolha do número de fatores é assumir que os resíduos de decomposição e o ruído instrumental seguem uma distribuição Normal, de forma que um valor crítico pode ser estabelecido requerendo que o resíduo  $RD_{un}$  deva ser menor que o nível do ruído instrumental com 99% de confiança, em um teste unilateral ( $RD_{un} < s_{\text{ruído}} \cdot 2,326$ ), sendo o número de fatores do RBL escolhido como o menor número necessário em que essa condição é satisfeita. Outro critério de valor crítico pode ser sugerido utilizando o resíduo médio de decomposição obtido nas amostras de calibração com 99% de confiança, no lugar do desvio padrão do ruído instrumental. Nesse caso, a utilização do resíduo médio é vantajosa por não requerer medidas em replicatas e amostras de branco.

O modelo BLLS, quando comparado com o PARAFAC ou o MCR, apresenta as vantagens de não requerer um processo de inicialização em seu algoritmo; não necessitar da imposição de restrições (como por exemplo não-negatividade ou ortogonalidade) durante o cálculo dos perfis em cada dimensão; poder ser implementado em um programa mais facilmente e ser mais rápido. A

facilidade da construção do modelo BLLS e análise de amostras onde o teor dos analitos é desconhecido pode ser melhor visualizada na Tabela 1, onde seu algoritmo é apresentado de forma compacta. Pode ser observado que o algoritmo não possui etapas iterativas, podendo ser construído e a concentração das espécies de interesse de uma amostra  $\mathbf{X}_{un}$  estimada através de apenas 6 operações de álgebra matricial, comumente disponíveis em programas especializados nesse tipo de cálculo (como exemplo Octave e Matlab). Contudo, o BLLS apresenta a grande desvantagem de requerer o conhecimento da composição de todas as amostras utilizadas na etapa de calibração, o que pode implicar uma séria restrição ao modelo em aplicações com amostras complexas, impedindo, por exemplo, a utilização do BLLS com métodos de adição de padrão.

**Tabela 1.** Algoritmo para a construção do modelo BLLS em situações onde não há presença de interferentes.

Operação	Comentário
$\mathbf{V}_x = [\text{vec}(\mathbf{X}_1)   \text{vec}(\mathbf{X}_2)   \dots   \text{vec}(\mathbf{X}_1)];$	Vetoriza as matrizes de dados.
$\mathbf{V}_s = \mathbf{V}_x \mathbf{Y}^{T+};$	Estima as matrizes de “absortividades molares”.
Para cada coluna de $\mathbf{V}_s$ : $\mathbf{S}_n = \text{reshape}(\mathbf{V}_s);$ $(\mathbf{b}_n, \mathbf{g}_n, \mathbf{c}_n) = \text{SVD}(\mathbf{S}_n);$	Estima os perfis das espécies presentes nas amostras de calibração.
$\mathbf{S}_{cal} = [g_1(\mathbf{c}_1 \otimes \mathbf{b}_1)   g_2(\mathbf{c}_2 \otimes \mathbf{b}_2)   \dots   g_{Nc}(\mathbf{c}_{Nc} \otimes \mathbf{b}_{Nc})];$	Determina a matriz $\mathbf{S}_{cal}$ .
Dada uma amostra $\mathbf{X}_{un}$ $\mathbf{y}_{un} = \mathbf{S}_{cal}^+ \text{vec}(\mathbf{X}_{un});$	Estima a concentração dos analitos na amostra.

No apêndice B1 é apresentado o código fonte de um programa, desenvolvido neste trabalho de tese, para a construção do modelo BLLS e determinação de algumas de suas figuras de mérito. O programa é compatível com os ambientes Octave e Matlab em suas versões 3.0.0 e 6.5, respectivamente.

### 1.1.4 Quadrados Mínimos Parciais Desdobrado (UPLS)\*<sup>1</sup>

O PLS atualmente é o modelo de calibração de primeira ordem mais utilizado e descrito na literatura<sup>56,57,58,59,60</sup>. A primeira tentativa de sua extensão para a utilização com dados de segunda ordem foi proposta por Wold et al.<sup>61</sup>, onde as matrizes de dados eram vetorizadas, como ilustrado na Figura 2A, e o modelo PLS de primeira ordem era aplicado. A principal crítica a este modelo era que, uma vez feita a vetorização, algumas das informações inerentes à estrutura dos dados de segunda ordem eram perdidas. Anos depois, Bro<sup>62</sup> apresentou outra proposta de generalização do PLS para segunda e ordens superiores, na qual a estrutura dos dados era preservada. Deste modo, os modelos propostos por Wold et al.<sup>61</sup> e Bro<sup>62</sup> passaram a ser designados como quadrados mínimos parciais desdobrado (UPLS, do inglês “Unfold Partial Least Squares”) e N-dimensional (NPLS, do inglês “N-way Partial Least Squares”). Contudo, ambos UPLS e NPLS apresentavam como desvantagem em relação aos demais modelos de calibração de segunda ordem o fato de não possuírem a vantagem de segunda ordem.

Öhman et al.<sup>55,63</sup> em 1990 desenvolveram o processo de RBL e uma proposta de algoritmo para sua implementação no UPLS, incorporando assim a vantagem de segunda ordem ao modelo UPLS. No entanto, a utilização do RBL com UPLS como proposto por Öhman, pelo nosso conhecimento, aparentemente não foi relatada a não ser por esses dois trabalhos até 2002, sendo o RBL utilizado de forma isolada como um modelo de calibração de segunda ordem em poucos trabalhos tais como o de Reis et al.<sup>64</sup>. Linder e Sundberg<sup>44</sup> em seu segundo trabalho descrevendo o BLS relataram um procedimento com o mesmo princípio do RBL, utilizando SVD como base do método, o qual foi utilizado por Olivieri<sup>65</sup> em um

---

\* Nessa tese os modelos UPLS foram construídos a partir de apenas uma propriedade de interesse, isto é, equivalentes a modelos PLS1 para calibração de primeira ordem. No entanto, como forma de simplificar a notação utilizada no texto a sigla utilizada neste caso será UPLS em vez de UPLS1 .

novo trabalho propondo o RBL junto com o UPLS. Recentes trabalhos publicados com UPLS empregando RBL comprovam que a presença de interferentes nas amostras analisadas é modelada satisfatoriamente, fornecendo resultados para as estimativas da concentração das espécies de interesse semelhantes às obtidas com BLLS, PARAFAC ou MCR, comprovando a incorporação da vantagem de segunda-ordem ao modelo UPLS<sup>66,67,68</sup>.

Como exposto acima, o UPLS consiste do mesmo modelo PLS aplicado à dados de segunda ordem, quando as  $I_c$  matrizes de calibração são vetorizadas dando origem a uma matriz  $\mathbf{X}$  de dimensão (I,JK), como ilustrado na Figura 2A, e as concentrações de referência em um vetor  $\mathbf{y}$ . Em seguida,  $\mathbf{X}$  e  $\mathbf{y}$  são decompostas por análise em componentes principais (PCA, do inglês “Principal Component Analysis”), a qual pode ser realizada com base em diversos algoritmos descritos na literatura<sup>58,59</sup>, e expressas por uma soma de matrizes de posto unitário por:

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E} = \sum_{i=1}^F \mathbf{t}_i \mathbf{p}_i^T + \mathbf{E} \quad (24)$$

$$\mathbf{y} = \mathbf{Tq}^T + \mathbf{f} = \sum_{i=1}^F \mathbf{T}_i \mathbf{q}_i^T + \mathbf{f} \quad (25)$$

onde, as matrizes  $\mathbf{E}$  e  $\mathbf{f}$  contém a informação de  $\mathbf{X}$  e  $\mathbf{y}$ , respectivamente, que não é explicada pelo modelo;  $\mathbf{T}$  é a matriz de escores;  $\mathbf{P}$  a matriz de pesos e  $F$  é o número de variáveis latentes utilizadas pelo modelo. Uma vez que o modelo foi construído, a estimativa do valor da propriedade de interesse ( $\hat{y}$ ) em uma amostra desconhecida “i” é obtida pela multiplicação do vetor resultado do desdobramento da matriz de dados instrumentais para essa amostra ( $\mathbf{x}_{un}$ ) por um vetor de coeficientes de regressão ( $\mathbf{b}_{UPLS}$ ) apropriado<sup>57</sup>:

$$\hat{y}_{un} = \mathbf{t}_{un} \mathbf{q}^T = \mathbf{x}_{un} \mathbf{b}_{UPLS} = \mathbf{x}_{un} \mathbf{W}(\mathbf{P}^T \mathbf{W})^{-1} \mathbf{q}^T \quad (26)$$

onde,  $\mathbf{W}$  é uma segunda matriz de pesos que obtida durante o algoritmo PLS. A

identificação de  $\mathbf{b}_{\text{UPLS}}$  como sendo igual a  $\mathbf{W}(\mathbf{P}^T\mathbf{W})^{-1}\mathbf{q}^T$  na equação 26 não é clara a primeira vista, mas lembrando das igualdades presentes no algoritmo do PLS:

$$\mathbf{X} = \mathbf{TP}^T \quad (27)$$

$$\mathbf{y} = \mathbf{Tq}^T = \mathbf{Xb}_{\text{UPLS}} \quad (28)$$

$$\mathbf{T} = \mathbf{XW} \quad (29)$$

Multiplicando por  $\mathbf{W}$  a equação 27, tem-se:

$$\mathbf{XW} = \mathbf{TP}^T\mathbf{W} \quad (30)$$

Multiplicando ambos os lados por  $(\mathbf{P}^T\mathbf{W})^{-1}$  de modo a isolar  $\mathbf{T}$  no lado direito da equação 30, tem-se:

$$\mathbf{XW}(\mathbf{P}^T\mathbf{W})^{-1} = \mathbf{TP}^T\mathbf{W}(\mathbf{P}^T\mathbf{W})^{-1} = \mathbf{TI} = \mathbf{T} \quad (31)$$

onde  $\mathbf{I}$  é uma matriz identidade de tamanho apropriado. Finalmente, ambos os lados da equação são multiplicados por  $\mathbf{q}$  para obter a relação expressa na equação 28:

$$\mathbf{XW}(\mathbf{P}^T\mathbf{W})^{-1}\mathbf{q}^T = \mathbf{Tq}^T = \mathbf{y} = \mathbf{Xb}_{\text{UPLS}} \quad (32)$$

obtendo assim a identificação de  $\mathbf{b}_{\text{UPLS}}$  como  $\mathbf{W}(\mathbf{P}^T\mathbf{W})^{-1}\mathbf{q}^T$ .

Uma vez construído o modelo UPLS, a presença de interferentes em amostras de composição desconhecida é considerada pela decomposição do sinal instrumental de uma amostra ( $\mathbf{X}_{\text{un}}$ ) em duas partes, uma descrita pelo modelo ( $\mathbf{X}_{\text{mod}}$ ) e outra não modelada que contém o interferente ( $\mathbf{X}_{\text{int}}$ ), expressa por<sup>65,67</sup>:

$$\mathbf{X}_{\text{un}} = \mathbf{X}_{\text{mod}} + \mathbf{X}_{\text{int}} \quad (33)$$

Uma vez que a parte que é descrita pelo modelo pode ser expressa em termos das variáveis latentes determinadas pelo modelo UPLS, a equação 33 pode ser representada como<sup>65,67</sup>:

$$\text{vec}(\mathbf{X}_{\text{un}}) = \mathbf{P}\mathbf{t}_{\text{un}} + \mathbf{e}_{\text{mod}} + \text{vec}(\mathbf{X}_{\text{int}}) \quad (34)$$

onde  $\mathbf{e}_{\text{mod}}$  vetor de erros deixados pelo modelo, que deve estar próximo do ruído instrumental. Se existir alguma estrutura bilinear em  $\mathbf{X}_{\text{int}}$ , ela pode ser modelada por SVD<sup>65,67</sup>:

$$\text{vec}(\mathbf{X}_{\text{un}}) = \mathbf{P}\mathbf{t}_{\text{un}} + \text{vec}(\mathbf{g}_{\text{int}}\mathbf{b}_{\text{int}}(\mathbf{c}_{\text{int}})^T) + \mathbf{e}_{\text{un}} \quad (35)$$

O que permite estimar os perfis dos interferentes presentes em  $\mathbf{X}_{\text{int}}$ , por<sup>65,67</sup>:

$$(\mathbf{b}_{\text{int}}, \mathbf{g}_{\text{int}}, \mathbf{c}_{\text{int}}) = \text{SVD}(\mathbf{X}_{\text{int}}) \quad (36)$$

onde, como no BLLS,  $\mathbf{b}_{\text{int}}$  e  $\mathbf{c}_{\text{int}}$  correspondem aos primeiros vetores singulares esquerdo e direito de  $\mathbf{X}_{\text{int}}$ , respectivamente, que correspondem aos perfis dos interferentes nas dimensões J e K,  $\mathbf{g}_{\text{int}}$  é o primeiro valor singular correspondente e  $\mathbf{e}_{\text{un}}$  é o erro de decomposição que engloba  $\mathbf{e}_{\text{mod}}$  mais a parte que não pode ser modelada como interferente. O processo RBL é feito de forma que  $\mathbf{P}$  é mantido constante e  $\mathbf{t}_{\text{un}}$  varia até que a norma do erro de decomposição ( $\|\mathbf{e}_{\text{un}}\|$ ) seja minimizada, o que pode ser feito por um algoritmo iterativo ou utilizando o método de minimização de Gauss-Newton. Após a convergência os valores obtidos para  $\mathbf{t}_{\text{u}}$  são utilizados para estimar a concentração da amostra pela primeira relação expressa na equação 26<sup>65,67</sup>.

## 1.2 Validação de modelos de calibração de segunda ordem: figuras de mérito.

Sempre que um método analítico é proposto este deve ser validado de modo a assegurar que os resultados obtidos por meio dele tenham a eficiência requerida nas condições em que ele será aplicado. Essa validação é feita através da determinação de diversos parâmetros que caracterizam a eficiência do método, os quais são denominados figuras de mérito. As especificações de como essas figuras

de mérito devem ser estimadas e os níveis requeridos em cada uma dependem da propriedade de interesse determinada ou da área a que o método se destina, sendo regidas por órgãos de fiscalização específicos como, por exemplo: a Farmacopéia Americana (USP, do inglês “United States Pharmacopoeia”)<sup>5,6</sup>, o protocolo da Conferencia Internacional de Harmonização de Requerimentos Técnicos para o Registro de fármacos para o uso humano (ICH, do inglês “International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use”)<sup>69</sup>, a Sociedade Americana de Testes e Materiais (ASTM, do inglês “American Society for Testing and Materials”)<sup>7</sup>, Instituto Nacional de Metrologia, Normalização e qualidade Industrial (INMETRO) e a Agência Nacional de Vigilância Sanitária (ANVISA)<sup>70</sup>. Dependendo da norma ou órgão de fiscalização a que o método está sujeito, a quantidade de figuras de mérito que devem ser determinadas pode ser diferente, sendo as principais: (1) exatidão, (2) precisão, (3) robustez, (4) extensão do intervalo em que a propriedade de interesse é determinada pelo método, (5) sensibilidade, (6) limite de detecção, (7) seletividade, (8) intervalos de confiança, (9) linearidade e (10) incerteza.

De acordo com um recente artigo de revisão<sup>71</sup>, a análise estatística envolvida em modelos de calibração de primeira ordem pode ser considerada relativamente complexa e alvo de pesquisas atualmente. Contudo, esses modelos já dispõem de algumas recomendações de órgãos oficiais como ASTM<sup>7</sup>, USP<sup>5,6</sup> e IUPAC<sup>8,9</sup>, para sua construção e validação, o que facilita a sua aceitação e implementação. Em particular para o setor farmacêutico, no Brasil trabalhos científicos vêm sendo desenvolvidos e publicados ressaltando a importância de uma norma de âmbito nacional<sup>23,72,73</sup>. Nesse sentido, uma proposta de norma já foi submetida à Farmacopéia Brasileira, que até o momento se encontra aguardando julgamento<sup>74</sup>.

Para modelos de calibração de segunda ordem a análise estatística envolvida também é complexa e ainda pouco desenvolvida<sup>71</sup>. Contudo, diversas pesquisas

atualmente tem tido como foco a determinação de figuras de mérito desses modelos, sendo possível encontrar trabalhos que discutem ou propõem formas para a determinação de parâmetros como: incerteza<sup>75,76,77</sup>, sensibilidade<sup>52,54,65,78</sup>, seletividade<sup>78</sup> e limites de detecção<sup>12,68,79</sup>. Esses trabalhos demonstram que para essas figuras de mérito devem-se levar em conta algumas particularidades de cada modelo, enquanto que outras figuras de mérito, como precisão e exatidão, podem ser determinadas de forma análoga ao que é feito em métodos univariados.

A seguir são descritas em maior detalhe algumas figuras de mérito abordadas nesse trabalho, as quais utilizam, dependendo do caso, propostas já consolidadas para sua determinação ou propostas recentes na literatura no caso de parâmetros que ainda apresentam discussões na forma de sua determinação.

### 1.2.1 Precisão

Expressa o grau de concordância entre os resultados de uma série de medidas feitas para uma mesma amostra homogênea em condições determinadas<sup>69</sup>. Em geral pode ser considerada em diversos níveis, tais como: repetitividade, precisão intermediária (interdias, interanalista, etc.) e reprodutibilidade<sup>69</sup>. Em todos os casos podendo ser determinada simplesmente por meio do cálculo do desvio padrão das  $n$  amostras realizadas em replicata ( $m$ )<sup>80</sup>:

$$\text{precisão} = \sqrt{\frac{\sum_{i=1}^n \sum_{j=1}^m (\hat{y}_{ij} - \hat{\bar{y}}_i)^2}{n(m-1)}} \quad (37)$$

A comparação entre as precisões de dois métodos pode ser feita através de um teste-F de significância.

### 1.2.2 Erro médio de estimativa da propriedade de interesse

Algumas vezes referido como uma estimativa da exatidão e utilizado para a comparação dos resultados obtidos em trabalhos com calibração

multivariada<sup>23,66,68,81</sup> a raiz quadrada do erro médio quadrático de previsão (RMSEP, do inglês “Root Mean Square Error of Prediction”) é definido por<sup>57</sup>:

$$\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^{I_v} (y_i - \hat{y}_i)^2}{I_v}} \quad (38)$$

onde,  $I_v$  é o número de amostras utilizadas na etapa de validação ou teste do modelo,  $y$  e  $\hat{y}$  são os valores de referência e estimativas da propriedade de interesse, respectivamente.

Como pode ser observado na equação 38, o RMSEP é uma medida de dispersão semelhante ao desvio padrão, mas que mede a dispersão entre os valores estimados pelo modelo e de referência. Outra propriedade que se assemelha à do desvio padrão, e que deve ser enfatizada, é que o RMSEP é uma medida que considera apenas erros aleatórios, que é uma decorrência da elevação dos erros ao quadrado na equação 38. Por exemplo, considerando os resultados de dois métodos distintos, supondo que um apresente erros sistemáticos negativos e o outro tenha erros com o mesmo valor em módulo mas que sejam distribuídos de forma aleatória, ambos fornecem os mesmos valores de RMSEP. Este exemplo ilustra um erro comumente cometido, que é a comparação de dois RMSEP por meio de um teste-F e a conclusão de que os resultados dos respectivos métodos são equivalentes. Portanto, a constatação de que dois RMSEP são estatisticamente equivalentes por meio de um teste-F apenas torna possível afirmar que os **erros médios** na estimativa da propriedade de interesse dos dois métodos são equivalentes não podendo ser utilizada para inferir sobre a exatidão do método.

Em termos relativos (REP, do inglês “Relative Error of Prediction”), o erro médio pode ser expresso por<sup>82,83</sup>:

$$\text{REP} = 100 \sqrt{\frac{\sum_{i=1}^{I_v} (y_i - \hat{y}_i)^2}{I_v y_i^2}} \quad (39)$$

### 1.2.3 Exatidão

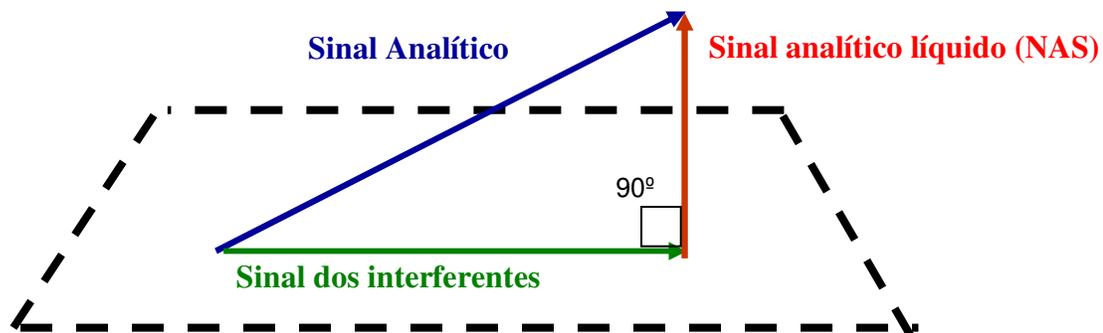
Expressa o grau de concordância entre um valor que é aceito por convenção como verdadeiro ou um valor de referência e o valor estimado pelo modelo de regressão<sup>69</sup>. Na comparação entre os resultados obtidos em dois métodos distintos, a exatidão pode ser acessada por meio da comparação dos valores obtidos para a inclinação e o intercepto de uma reta ajustada entre os valores de referência e os estimados pelo modelo. Caso os intervalos de confiança para esses parâmetros contiverem os seus valores esperados iguais a 1 e 0, para a inclinação e o intercepto respectivamente, os modelos podem ser considerados estatisticamente equivalentes no nível de confiança considerado. Valores estimados para a inclinação e intercepto fora de seus intervalos de confiança indicam erros sistemáticos proporcionais e constantes, respectivamente<sup>8,18</sup>.

### 1.2.4 Sensibilidade (SEN)

A estimativa da sensibilidade em modelos de calibração multivariada, particularmente em modelos de segunda ordem, depende do conceito de sinal analítico líquido (NAS, do inglês “Net Analyte Signal”), primeiramente desenvolvido por Lorber<sup>84,85</sup>, o qual o definiu como sendo a fração do sinal instrumental que é ortogonal às contribuições de outros possíveis constituintes presentes na amostra<sup>85</sup>. A Figura 5 mostra a representação geométrica da propriedade de ortogonalidade do NAS no caso de um vetor de dados.

Na extensão da teoria do NAS para calibração de segunda ou ordens superiores duas propostas foram desenvolvidas por Messick, Kalivas e Lang<sup>86</sup> e Ho, Christian e Davidson<sup>87</sup> nas quais duas formas independentes de estimar a sensibilidade foram deduzidas. Em um trabalho relativamente recente, Olivieri<sup>78</sup> utilizando dados simulados comparou as duas propostas em três situações distintas: (1) presença de dois ou mais analitos em uma amostra e sem interferentes. Neste

caso todas as espécies presentes são calibradas e o modelo não utiliza a vantagem de segunda ordem; (2) presença de apenas uma espécie de interesse e um ou mais interferentes, situação em que a vantagem de segunda ordem é utilizada; e (3) dois ou mais analitos calibrados e presença de interferentes. Foi observado que, por exemplo para o BLLS, a proposta de Messick, Kalivas e Lang<sup>86</sup> era adequada quando a vantagem de segunda ordem não era utilizada (situação 1), enquanto que a proposta de Ho, Christian e Davidson<sup>87</sup> era adequada quando a vantagem de segunda ordem era empregada em situações onde apenas um analito era calibrado (situação 2), sendo os resultados discordantes em ambas as propostas para a terceira situação. Uma expressão geral para a estimativa da sensibilidade em modelos bilineares de calibração de segunda-ordem, assim como PARAFAC e BLLS, que pudesse lidar com essas três situações foi deduzida em seguida por Olivieri e Faber<sup>52</sup>.



**Figura 5-** Representação geométrica da propriedade do vetor NAS em ser ortogonal ao vetor do sinal dos interferentes.

Uma observação importante com relação às estimativas de sensibilidade para modelos de calibração de segunda ordem, quando a vantagem de segunda ordem é utilizada, é que seus valores são dependentes dos interferentes presentes, isto é, a sensibilidade de um determinado analito é dependente do número de interferentes presentes e do grau de sobreposição destes interferentes em seu sinal, sendo os

valores de sensibilidade menores quanto maior for o número de interferentes e a sobreposição. Portanto em modelos que utilizam a vantagem de segunda ordem a sensibilidade se torna uma medida específica de cada amostra e não do método como um todo.

Para melhor compreender o raciocínio envolvido na dedução da equação proposta por Olivieri e Faber<sup>52</sup>, é necessário explicar como a projeção ortogonal expressa na Figura 5 é realizada e em seguida discutir as duas situações mais simples (descritas anteriormente) em que podem ser utilizadas as propostas de Messick, Kalivas e Lang<sup>86</sup> e Ho, Christian e Davidson<sup>87</sup>, respectivamente.

#### 1.2.4.1 Projeções ortogonais.

Dada uma matriz  $\mathbf{X}$  qualquer, a matriz  $\mathbf{XX}^+$  (onde o símbolo “+” sobrescrito indica a Moore–Penrose pseudoinversa da matriz em questão) é uma matriz de projeção que possui as seguintes propriedades<sup>84</sup>:

$$\mathbf{X} = (\mathbf{XX}^+) \mathbf{X} \quad (40)$$

$$\mathbf{X}^+ = \mathbf{X}^+ (\mathbf{XX}^+) \quad (41)$$

Pelas relações acima temos que, se qualquer vetor  $\mathbf{z}$  é uma combinação linear da matriz  $\mathbf{X}$ , a multiplicação de  $\mathbf{z}$  pela matriz  $\mathbf{XX}^+$  fornece como resultado o próprio vetor  $\mathbf{z}$ . Contudo, se multiplicarmos  $\mathbf{z}$  por  $(\mathbf{I}-\mathbf{XX}^+)$ , onde  $\mathbf{I}$  é uma matriz identidade de mesmo tamanho de  $\mathbf{XX}^+$ , resultará em um vetor de zeros. Portanto, a multiplicação de um vetor pela matriz,  $(\mathbf{I}-\mathbf{XX}^+)$ , fornece como resultado um vetor que é ortogonal à matriz  $\mathbf{X}$ .

#### 1.2.4.2 Sensibilidade para modelos PARAFAC e BLLS na ausência de interferentes.

Considerando um sistema formado por 3 espécies de interesse que são calibradas, a matriz do sinal instrumental de uma amostra  $\mathbf{X}_{un}$  pode ser

representada por<sup>52,54</sup>:

$$\mathbf{X}_{un} = y_1 \mathbf{S}_1 + y_2 \mathbf{S}_2 + y_3 \mathbf{S}_3 + \mathbf{E}_{un} \quad (42)$$

onde  $\mathbf{E}_{un}$  uma matriz de erro,  $y_n$  é a concentração do componente  $n$  e  $\mathbf{S}_n$  a sua matriz de sinal em concentração unitária. Uma vez que os modelos BLLS e PARAFAC são capazes de fornecer os perfis de cada componente, as matrizes  $\mathbf{S}_n$  podem ser determinadas. Para o BLLS, por exemplo, tem-se (ver equação 15)<sup>52</sup>:

$$\mathbf{S}_n = \mathbf{g}_n \mathbf{b}_n \mathbf{c}_n^T \quad (43)$$

Quando interferentes não estão presentes, as matrizes na equação 42 podem ser representadas igualmente na forma de vetores aplicando o operador “vec”<sup>52</sup>:

$$\text{vec}(\mathbf{X}_{un}) = y_1 \text{vec}(\mathbf{S}_1) + y_2 \text{vec}(\mathbf{S}_2) + y_3 \text{vec}(\mathbf{S}_3) \quad (44)$$

A partir da equação 44, o sinal de cada componente pode ser obtido através da multiplicação de ambos os lados por uma matriz de projeção apropriada. Considerando a espécie 1, essa matriz de projeção é expressa por<sup>52</sup>:

$$\mathbf{P}_1 = \mathbf{I} - [\text{vec}(\mathbf{S}_2) | \text{vec}(\mathbf{S}_3)][\text{vec}(\mathbf{S}_2) | \text{vec}(\mathbf{S}_3)]^+ \quad (45)$$

Após a multiplicação de  $\mathbf{P}_1$  em ambos os lados da equação 44, as contribuições das espécies 2 e 3 são removidas, obtendo-se<sup>52</sup>:

$$\mathbf{P}_1 \text{vec}(\mathbf{X}_{un}) = y_1 \mathbf{P}_1 \text{vec}(\mathbf{S}_1) \quad (46)$$

Portanto, o sinal analítico líquido em concentração unitária da espécie 1 é igual ao vetor  $\mathbf{s}_1^* = \mathbf{P}_1 \text{vec}(\mathbf{S}_1)$ , e a sensibilidade para esta condição em particular, que não apresenta interferentes, pode ser definida como<sup>52</sup>:

$$\text{SEN}_1 = \|\mathbf{s}_1^*\| = \|\mathbf{P}_1 \text{vec}(\mathbf{S}_1)\| \quad (47)$$

A sensibilidade para as espécies 2 e 3 é determinada de forma análoga à que foi feita para a espécie 1, substituindo  $\mathbf{P}_1$  pela matriz de projeção apropriada. A equação 47 é equivalente à expressão deduzida por Messick, Kalivas e Lang<sup>86</sup>.

### 1.2.4.3 Sensibilidade para modelos PARAFAC e BLLS com um analito calibrado e na presença de interferentes.

Neste caso, considerando a espécie 1 como de interesse e as espécies 2 e 3 como os interferentes, as projeções ortogonais que possibilitam a remover a contribuição de 2 e 3 são aplicadas na à  $\mathbf{X}_{un}$  na forma de uma matriz (equação 42) e não na forma de vetor como no caso anterior, obtendo como resultado<sup>52</sup>:

$$\mathbf{P}_J \mathbf{X}_{un} \mathbf{P}_K = y_1 \mathbf{P}_J \mathbf{S}_1 \mathbf{P}_K \quad (48)$$

onde  $\mathbf{P}_J$  e  $\mathbf{P}_K$  são as projeções que removem as contribuições das espécies 2 e 3 nas dimensões J e K, respectivamente, que são iguais a:

$$\mathbf{P}_J = \mathbf{I} - [\mathbf{b}_2 \mid \mathbf{b}_3][\mathbf{b}_2 \mid \mathbf{b}_3]^+ \quad (49)$$

$$\mathbf{P}_K = \mathbf{I} - [\mathbf{c}_2 \mid \mathbf{c}_3][\mathbf{c}_2 \mid \mathbf{c}_3]^+ \quad (50)$$

onde  $\mathbf{b}_2$  e  $\mathbf{b}_3$  são os perfis das espécies 2 e 3 na dimensão J e  $\mathbf{c}_2$  e  $\mathbf{c}_3$  os perfis de 2 e 3 na dimensão K. Pela equação 48, a matriz do sinal analítico líquido de 1 em concentração unitária é dada por  $\mathbf{S}_1^* = \mathbf{P}_J \mathbf{S}_1 \mathbf{P}_K$ , assim a sensibilidade neste caso é dada por<sup>52</sup>:

$$\text{SEN}_1 = \|\mathbf{S}_1^*\| = \|\mathbf{P}_J \mathbf{S}_1 \mathbf{P}_K\| \quad (51)$$

Sendo a equação 51 equivalente à proposta por Ho, Christian e Davidson<sup>87</sup>.

### 1.2.4.4 Expressão geral para a determinação da sensibilidade em modelos PARAFAC e BLLS.

A partir da combinação das duas situações descritas acima podem ser obtidas várias equações para as situações onde têm-se vários analitos calibrados e interferentes presentes. Contudo, a expressão deduzida por Olivieri e Faber<sup>52</sup> é geral e cobre todas as possibilidades.

Considerando agora a situação em que tem-se “N<sub>c</sub>” espécies de interesse que estão calibradas e a presença de “int” interferentes. Primeiramente a contribuição

dos interferentes nas dimensões J e K são removidas de  $\mathbf{X}_{un}$  pelas projeções  $\mathbf{P}_{J,int}$  e  $\mathbf{P}_{K,int}$ , respectivamente, dadas por<sup>52</sup>:

$$\mathbf{P}_{J,int} = \mathbf{I} - \mathbf{B}_{int} \mathbf{B}_{int}^+ \quad (52)$$

$$\mathbf{P}_{K,int} = \mathbf{I} - \mathbf{C}_{int} \mathbf{C}_{int}^+ \quad (53)$$

onde  $\mathbf{B}_{int}$  e  $\mathbf{C}_{int}$  contém em suas colunas os perfis dos interferentes nas dimensões J e K respectivamente. Multiplicando essas matrizes de projeções por  $\mathbf{X}_{un}$ , têm-se:

$$\mathbf{P}_{J,int} \mathbf{X}_{un} \mathbf{P}_{K,int} = \sum_{n=1}^{N_c} y_n \mathbf{P}_{J,int} \mathbf{S}_n \mathbf{P}_{K,int} \quad (54)$$

Uma vez removidas as contribuições dos interferentes, para a obtenção, por exemplo, do sinal analítico da espécie 1 é necessário remover a contribuição dos demais analitos calibrados. Isso é feito aplicando o operador “vec” à equação 54 e multiplicando ambos os lados por outra matriz de projeção  $\mathbf{P}_{cal}$ :

$$\mathbf{P}_{cal} \text{vec}(\mathbf{P}_{J,int} \mathbf{X}_u \mathbf{P}_{K,int}) = y_1 \mathbf{P}_{cal} \text{vec}(\mathbf{P}_{J,int} \mathbf{S}_1 \mathbf{P}_{K,int}) \quad (55)$$

onde  $\mathbf{P}_{cal}$  é determinado como:

$$\mathbf{P}_{cal} = \mathbf{I} - \mathbf{\Theta} \mathbf{\Theta}^+ \quad (56)$$

$$\mathbf{\Theta} = [\text{vec}(\mathbf{P}_{J,int} \mathbf{S}_2 \mathbf{P}_{K,int}) \mid \dots \mid \text{vec}(\mathbf{P}_{J,int} \mathbf{S}_{N_c} \mathbf{P}_{K,int})] \quad (57)$$

Por fim, a sensibilidade de uma espécie 1 é identificada independente do contexto em modelos PARAFAC e BLLS, como<sup>52</sup>:

$$\text{SEN}_1 = \left\| \mathbf{P}_{cal} \text{vec}(\mathbf{P}_{J,int} \mathbf{S}_1 \mathbf{P}_{K,int}) \right\| \quad (58)$$

onde  $\mathbf{P}_{J,int}$  e  $\mathbf{P}_{K,int}$  são matrizes identidade se não há interferentes presentes em  $\mathbf{X}_{un}$ . Segundo Olivieri e Faber, a equação 58 fornece valores coerentes para todas as possibilidades de espécies calibradas e interferentes, contudo, computacionalmente ela não é prática por requerer a determinação de várias matrizes de projeção que podem ter um tamanho excessivo dependendo dos dados. Rearranjando os termos e a partir das relações existentes em álgebra matricial, uma equação equivalente e mais adequada nesse sentido pode ser obtida como<sup>52</sup>:

$$SEN_n = g_n \left\{ \left( \mathbf{B}_{cal}^T \mathbf{P}_{J,int} \mathbf{B}_{cal} \right) \left( \mathbf{C}_{cal}^T \mathbf{P}_{K,int} \mathbf{C}_{cal} \right) \right\}_{n,n}^{-1/2} \quad (59)$$

onde na equação 59 todas as matrizes  $\mathbf{B}_{cal}$ ,  $\mathbf{C}_{cal}$ ,  $\mathbf{P}_{J,int}$  e  $\mathbf{P}_{K,int}$ , diferentes das equações anteriores que determinam SEN, contém apenas perfis normalizados,  $g_n$  é um fator de escala que corresponde ao sinal do componente “n” de interesse em concentração unitária, e o sub-índice “n,n” indica o elemento na enésima linha e coluna da matriz resultante entre chaves. Os valores de  $g_n$  para o BLLS correspondem aos valores de  $g_n$  obtidos por SVD<sub>1</sub> na equação 15 para cada espécie de interesse. No caso do PARAFAC, os valores de  $g_n$  equivalem às inclinações das retas ajustadas entre o vetor de escores  $\mathbf{a}$  das amostras de calibração e as concentrações de referência  $\mathbf{y}$  de cada espécie.

#### 1.2.4.5 Sensibilidade para o modelo UPLS.

Em modelos PLS de primeira ordem, a sensibilidade é uma extensão direta de modelos univariados. Sendo igual ao inverso da norma dos coeficientes de regressão ( $\mathbf{b}_{UPLS}$ ) do modelo:

$$SEN = \frac{1}{\|\mathbf{b}_{UPLS}\|} \quad (60)$$

o que em um modelo de calibração de ordem zero corresponde ao inverso do coeficiente angular da curva analítica. Sendo esta estimativa de sensibilidade idêntica à obtida pela teoria do NAS<sup>23,72</sup>.

No caso do UPLS em calibração de segunda ordem, quando não há interferentes presentes, a sensibilidade também é igual ao inverso da norma dos coeficientes de regressão. Sendo necessário portanto uma estimativa para situações onde se necessita da vantagem de segunda ordem utilizando o processo RBL.

Seguindo o mesmo raciocínio aplicado ao BLLS e ao PARAFAC, em um modelo UPLS onde a matriz  $\mathbf{X}_{un}$  é descrita pelos pesos  $\mathbf{P}$  e escores  $\mathbf{T}$ , a

contribuição dos interferentes estimados pelo processo RBL (equação 36) é removida por<sup>65</sup>:

$$\text{vec}(\mathbf{P}_{J,\text{int}} \mathbf{X}_{\text{un}} \mathbf{P}_{K,\text{int}}) = \text{vec}[\mathbf{P}_{J,\text{int}} \text{reshape}(\mathbf{P} \mathbf{t}_{\text{un}}) \mathbf{P}_{K,\text{int}}] \quad (61)$$

onde o operador “reshape” realiza a operação inversa ao operador “vec”, isto é, retorna um vetor de dimensão (JK,1) à forma de uma matriz de dimensão (J,K).

Segundo Olivieri, a equação 60 pode ser rearranjada e expressa como<sup>65</sup>:

$$(\mathbf{P}_{J,\text{int}} \otimes \mathbf{P}_{K,\text{int}})^T \text{vec}(\mathbf{X}_{\text{un}}) = (\mathbf{P}_{J,\text{int}} \otimes \mathbf{P}_{K,\text{int}})^T \mathbf{P} \mathbf{t}_{\text{un}} \quad (62)$$

Indicando que esta seria uma descrição do modelo UPLS depois de descontada a contribuição dos interferentes e indicando que os escores  $\mathbf{t}_{\text{u}}$  podem ser estimados por:

$$\mathbf{t}_{\text{un}} = [(\mathbf{P}_{J,\text{int}} \otimes \mathbf{P}_{K,\text{int}})^T \mathbf{P}]^+ \text{vec}(\mathbf{X}_{\text{un}}) = \mathbf{P}_{\text{eff}}^+ \text{vec}(\mathbf{X}_{\text{un}}) \quad (63)$$

onde  $\mathbf{P}_{\text{eff}}^+$  é a matriz que converte os dados instrumentais nos escores quando há interferentes presentes nas amostras de previsão  $\mathbf{X}_{\text{un}}$ . Portanto, pela definição de  $\mathbf{P}_{\text{eff}}^+$  e pela igualdade  $y_{\text{un}} = \mathbf{t}_{\text{un}} \mathbf{q}$ , Olivieri<sup>65</sup> propõe que o vetor dos coeficientes de regressão quando há interferentes presentes seja aproximado por:

$$\mathbf{b}_{\text{RBL}} = (\mathbf{P}_{\text{eff}}^+)^T \mathbf{q} \quad (64)$$

E por fim a sensibilidade é dada por<sup>65</sup>:

$$\text{SEN}_{\text{RBL}} = \frac{1}{\|\mathbf{b}_{\text{RBL}}\|} = \frac{1}{\|(\mathbf{P}_{\text{eff}}^+)^T \mathbf{q}\|} \quad (65)$$

Contudo, as operações envolvidas nas passagens da equação 60 para a 62 não são explicadas por Olivieri<sup>65</sup> em seu trabalho e cabe ressaltar que o vetor de coeficientes de regressão  $\mathbf{b}_{\text{RBL}}$  somente é utilizado para a estimativa da sensibilidade e não na estimativa da concentração da propriedade de interesse, sendo esta última obtida por meio dos escores  $\mathbf{t}_{\text{un}}$  obtidos pelo processo RBL e pelo vetor de pesos  $\mathbf{q}$ , através da primeira relação da equação 28.

#### 1.2.4.6 Sensibilidade em modelos MCR

Em contraste com a abordagem descrita para os modelos BLLS, PARAFAC e UPLS, os trabalhos descritos na literatura relatando a estimativa da sensibilidade abordando o MCR propõem considerar a sensibilidade igual ao inverso da inclinação da reta ajustada entre as áreas ou as alturas dos perfis de concentração estimados pelo MCR e as concentrações de referência das amostras de calibração, de forma análoga a modelos univariados<sup>42,82</sup>.

De fato, o valor obtido é uma estimativa da sensibilidade do modelo MCR, mas sua escala não corresponde à mesma dos dados instrumentais e portanto não torna possível uma comparação direta com métodos univariados ou as estimativas de sensibilidade obtidas nos outros modelos de calibração de segunda ordem.

#### 1.2.5 Sensibilidade Analítica (SENA)

Sendo  $s_{\text{ruído}}$  uma estimativa do ruído instrumental, que pode ser obtida por meio do desvio padrão de um conjunto de medidas do sinal instrumental quando nenhuma espécie está presente (amostras de branco) ou pelos resíduos de decomposição do modelo, define-se a sensibilidade analítica (SENA) de uma propriedade de interesse “n” como<sup>9,88,89</sup>:

$$\text{SENA}_n = \frac{\text{SEN}_n}{s_{\text{ruído}}} \quad (66)$$

onde o inverso de SENA tem unidades de concentração e expressa a menor diferença de concentração entre amostras que pode ser distinguida pelo método, desconsiderando o erro ou ajuste do modelo. Portanto, o inverso de SENA seria a menor diferença de concentração que a sensibilidade permite distinguir entre duas amostras quando limitada apenas pelo ruído instrumental.

Evidentemente, os valores do inverso de SENA são estimativas muito otimistas que não condizem com a diferença de concentração que se pode medir na

prática, pois essa também é limitada pelo erro do modelo e das concentrações de referência. Contudo, esse parâmetro é útil para comparação entre valores obtidos em diferentes instrumentos que podem ter níveis de ruídos diferentes, sendo uma medida de certa forma análoga à razão sinal ruído, mas expressa em termos de concentração.

### 1.2.6 Limite de detecção (LD)

De acordo com a definição adotada pela ICH<sup>69</sup> e USP<sup>90</sup>, o LD é a menor quantidade da espécie de interesse (concentração) que pode ser detectada, mas não necessariamente quantificada, sob condições experimentais estabelecidas. De acordo com as recomendações da IUPAC<sup>91</sup>, o LD deve levar em conta as probabilidades de ocorrência de erros falso positivo e falso negativo, usualmente nos níveis de 95% (0.05) de confiança.

Para modelos de calibração em modo geral, a maneira mais simples de estimar o LD é a partir da sensibilidade do método e da flutuação do ruído instrumental ( $s_{\text{ruído}}$ ):

$$LD_n = 3,3 \frac{s_{\text{ruído}}}{SEN_n} \quad (67)$$

onde a constante 3,3 corresponde ao nível de confiança de 95% para ambos os erros falso positivo e falso negativo. Contudo, para que a utilização da equação 67 leve à uma estimativa realista do LD, é necessário que diversas condições sejam satisfeitas, como: a distribuição dos erros da concentração da espécie de interesse no nível de concentração igual a zero e no nível do LD sigam uma distribuição normal, tenham uma mesma dispersão e que o desvio padrão populacional dessa distribuição seja igual a  $s_{\text{ruído}}$ <sup>91</sup>. Na prática as estimativas obtidas pela equação 67 são geralmente otimistas em decorrência da estimativa de  $s_{\text{ruído}}$  subestimar o seu valor verdadeiro ou pelas dispersões no nível igual a zero e em LD serem

diferentes. É importante notar que se  $s_{\text{ruído}}$  for a estimativa do ruído instrumental, o LD pela equação 67 será simplesmente o valor do inverso de SENA multiplicado por 3,3.

Os diversos trabalhos que descrevem a determinação do LD em calibração de segunda ordem diferem, em sua maioria, na forma como é calculado o valor de  $s_{\text{ruído}}$ . Entre estes trabalhos, pode-se citar o trabalho de Rodríguez-Cuesta et al.<sup>82</sup> que utiliza  $s_{\text{ruído}}$  como sendo igual ao desvio padrão da reta ajustada entre as áreas obtidas pelos perfis de concentração fornecidos pelo MCR e as concentrações de referência; e o trabalho de Boqué et al.<sup>79</sup> que utiliza como desvio padrão os valores estimados a partir de equações que estimam a incerteza do modelo por propagação de erro. Contudo, diversos trabalhos recentes têm reportado estimativas de LD através da equação 67 e  $s_{\text{ruído}}$  obtido por meio de medidas em replicata do sinal instrumental ou através do resíduo de decomposição obtido pelo modelo de calibração de segunda ordem<sup>12,68</sup>. Neste trabalho de tese e nos códigos dos programas que são apresentados nos apêndices B considera-se apenas o LD estimado a partir da equação 67, e  $s_{\text{ruído}}$  obtido pela variação instrumental dos dados ou, caso este último não for disponível, pelo resíduo de decomposição das amostras.

### 1.2.7 Seletividade (SEL)

De acordo com a recomendação da IUPAC, a SEL refere-se à extensão na qual um método pode ser utilizado para determinar analitos em misturas ou meios sem a interferência de outros componentes<sup>92</sup>. Apesar desta definição ser clara do ponto de vista de métodos de calibração de ordem zero (univariados), em calibração multivariada (primeira e ordens superiores) ela pode ser considerada ligeiramente deficiente, uma vez que métodos multivariados são capazes de estimar a concentração de espécies de interesse em um meio em que há presença de

interferentes (que podem estar presentes nas amostras utilizadas na calibração ou apenas nas amostras de previsão, sendo interferentes propriamente ditos).

Em um trabalho de revisão, Faber et al.<sup>93</sup> argumentam que uma medida apropriada de seletividade deve preencher diversos critérios, sendo os mais relevantes, do ponto de vista prático, que a seletividade deve estar relacionada com a incerteza das estimativas obtidas no método e que possa ser generalizado para calibrações multivariadas de ordens superiores. Nesse sentido sugerem que a proposta desenvolvida independentemente por Lorber<sup>84</sup> e Bergmann et al.<sup>94</sup> é a mais adequada por preencher estes critérios, a qual é definida como a razão entre o sinal analítico líquido ( $S_n^*$ ) do analito de interesse em concentração unitária (já é descontado da interferência de outras espécies presentes) e o seu sinal correspondente (também em concentração unitária) numa situação sem remover a fração que sofre influência dos interferentes ( $S_n$ ), o que pode ser matematicamente expresso para uma espécie n como<sup>93</sup>:

$$SEL_n = \frac{\|S_n^*\|}{\|S_n\|} \quad (68)$$

Contudo, essa definição restringe a medida da seletividade a modelos onde se dispõe do sinal de cada analito em concentração unitária, o que acontece apenas em modelos de calibração de primeira ordem baseados em quadrados mínimos clássico e modelos de segunda ordem que forneçam os perfis de cada espécie presente no meio, tais como PARAFAC, BLLS, e MCR. É interessante notar que a equação 68 é na verdade a razão da sensibilidade nas situações em que existem interferentes sobrepostos com o analito de interesse e quanto este está isolado dos demais. Portanto a definição de seletividade de acordo com Lorber<sup>84</sup> e Bergmann et al.<sup>94</sup> relata o quanto se perde em termos de sensibilidade devido à presença de outros componentes na amostra, sendo esta igual a unidade quando não há sobreposição e

diminuindo a medida que o número de interferentes e a sobreposição aumenta.

De acordo com essa definição de seletividade e com a equação geral de sensibilidade para modelos PARAFAC e BLLS (equação 59), que é transcrita abaixo, a determinação da seletividade para esses modelos pode ser expressa por<sup>12,78</sup>:

$$SEN_n = g_n \left\{ \left( \mathbf{B}_{cal}^T \mathbf{P}_{J,int} \mathbf{B}_{cal} \right) \left( \mathbf{C}_{cal}^T \mathbf{P}_{K,int} \mathbf{C}_{cal} \right) \right\}_{n,n}^{-1/2} \quad (59)$$

$$SEL_n = \left\{ \left( \mathbf{B}_{cal}^T \mathbf{P}_{J,int} \mathbf{B}_{cal} \right) \left( \mathbf{C}_{cal}^T \mathbf{P}_{K,int} \mathbf{C}_{cal} \right) \right\}_{n,n}^{-1/2} \quad (69)$$

onde ressalta-se novamente que nestas equações os perfis dos demais componentes calibrados e interferentes são normalizados. Comparando as equações 59 e 69 nota-se que ambas somente diferem pelo termo  $g_n$  que corresponde ao sinal total em concentração unitária do componente  $n$ .

Baseado neste conceito de seletividade, para o modelo UPLS (e conseqüente mente para o modelo PLS de primeira ordem) Olivieri<sup>65</sup> indica que a seletividade não pode ser determinada, uma vez que neste caso o modelo não é capaz de fornecer os sinais individuais das espécies de interesse em concentração unitária (o que corresponde ao denominador da equação 68). Com essa observação, nota-se que os valores de seletividade reportados por diversos trabalhos com PLS em calibração de primeira ordem, os quais baseiam-se principalmente nos trabalhos publicados por Lorber et al.<sup>85</sup> e Ferré et al.<sup>95</sup>, estão incorretos, uma vez que neles a seletividade nestes modelos é determinada como a razão do sinal analítico líquido e a norma do sinal instrumental originalmente adquirido (o qual contém contribuições de outros componentes além da espécie de interesse).

Voltando ao modelo UPLS, considerando a situação onde não há interferentes na amostra de previsão, isto é, a vantagem de segunda ordem não é utilizada, a limitação no cálculo da seletividade indicada por Olivieri<sup>65</sup> é clara. Contudo, quando a vantagem de segunda ordem é utilizada, o valor de

sensibilidade fornecido pela equação 65 corresponde a um valor de sensibilidade que leva em conta a perda de sensibilidade provocada pelo interferente da amostra de composição desconhecida (o que equivale ao numerador da equação 68). Logo interpretando o cálculo de seletividade como uma medida do quanto se perde em sensibilidade em um analito  $n$  pela ação dos interferentes, para o modelo UPLS utilizando a vantagem de segunda ordem uma medida da seletividade pode ser expressa por:

$$SEL_n = \frac{SEN_{n,RBL}}{SEN_n} \quad (70)$$

onde  $SEN_{n,RBL}$  é o valor de seletividade obtido pela equação 65 para a espécie  $n$  e  $SEN_n$  é o valor da sensibilidade quando não há interferentes na amostra de previsão. Sendo essa uma medida de seletividade que leva em consideração apenas o efeito dos interferentes não levados em conta na calibração. Pelo nosso conhecimento a relação expressa na equação 70 não foi reportada até o momento na literatura e será avaliada e comparada com os valores obtidos de seletividade nos modelos PARAFAC e BLLS como uma proposta de determinação desse parâmetro no modelo UPLS nas situações onde a vantagem de segunda ordem for utilizada.

Para o modelo MCR, propostas de medida de seletividade aparentemente não foram descritas literatura. Uma explicação para esse fato é que isso exigiria uma expressão para a sensibilidade desse modelo seguindo a mesma abordagem descrita para PARAFAC, BLLS, e UPLS, algo que pode ser deduzido de forma semelhante ao que foi feito para o PARAFAC e BLLS, uma vez que o MCR também fornece os perfis de cada componente da amostra. Contudo, isso ainda não foi aparentemente realizado.

### 1.2.8 Incerteza na concentração estimada da propriedade de interesse

A incerteza é uma das figuras de mérito mais importantes de um método

analítico, sendo também uma das que requer um maior rigor estatístico. Sua determinação geralmente é baseada em uma equação analítica que é deduzida pela teoria de propagação dos erros envolvidos em cada etapa, desde a preparação da amostra à estimativa dos parâmetros do modelo.

A proposição de equações para a determinação da incerteza no modelo PARAFAC foram feitas por Olivieri e Faber<sup>75,76</sup>. Resultados com dados simulados e adição de ruído sugerem que uma estimativa da incerteza com boa concordância com os valores esperados pode ser obtida tanto nas situações onde o PARAFAC utiliza ou não a vantagem de segunda ordem. Contudo, essa boa concordância é condicionada a boas estimativas da incerteza nas concentrações das amostras de calibração ou do método de referência e da incerteza proveniente dos dados instrumentais, as quais nem sempre são disponíveis. Por exemplo, em dados de EEF a incerteza dos dados experimentais pode ser estimada relativamente bem através do ruído presente nos dados, em situações onde o solvente ou o branco não apresenta fluorescência. No entanto, para dados de HPLC-DAD a incerteza dos dados estimada do mesmo modo pode conduzir a uma estimativa muito menor do que o seu valor verdadeiro, pois a incerteza que pequenos desvios no tempo de retenção causam ao modelo é difícil de ser estimada.

Para o MCR, Saurina et al.<sup>42</sup> e Rodríguez-Cuesta<sup>82</sup> propõem que a estimativa da incerteza pode ser feita em analogia a calibração de ordem zero, utilizando as mesmas equações desses modelos já bem estabelecidos. Essa proposta baseia-se no fato de que na prática a estimativa da concentração no MCR é feita através de uma regressão “univariada” entre os valores de área ou alturas sobre os perfis de concentração recuperados pelo MCR, decompondo-se amostras de calibração e previsão conjuntamente, e as concentrações de referência. Dessa forma propõem que os valores de área ou altura carregam toda a incerteza presente nos dados instrumentais e da etapa de decomposição do MCR. Considerando um modelo de

calibração inversa, onde a concentração é função dos dados instrumentais (ou de uma medida provinda deles como é no MCR), a incerteza associada à estimativa da concentração de uma amostra ( $s(\hat{y}_u - y_u)$ ) pode se determinada como<sup>19</sup>:

$$s(\hat{y}_u - y_u) = s_{\text{reg}} \sqrt{\frac{1}{m} + \frac{1}{I_c} + \frac{(a_u - \bar{a})^2}{\sum (a_i - \bar{a})^2}} \quad (71)$$

onde  $m$  é o número de replicatas feitas na referida amostra,  $I_c$  é o número de amostras de calibração, “ $a$ ” são os valores de alturas ou áreas sobre perfis de concentração obtidos no MCR e  $s_{\text{reg}}$  é o desvio padrão do erro da regressão, expresso por<sup>19</sup>:

$$s_{\text{reg}} = \sqrt{\frac{\sum_{i=1}^{I_c} (\hat{y}_i - y)^2}{I_c - 2}} \quad (72)$$

onde no denominador são subtraídos dois graus de liberdade referentes à inclinação e intercepto da equação de regressão. Contudo, o número de trabalhos que expressão a incerteza em resultados experimentais com MCR ainda é pequeno para que essa proposta seja confirmada com segurança<sup>42,82</sup>.

Apesar de, pelo nosso conhecimento, nenhum trabalho ainda ter sido publicado propondo a mesma abordagem descrita acima no PARAFAC quando este efetua a decomposição das amostras de calibração mais uma de previsão, a utilização desse procedimento segue o mesmo raciocínio empregado no MCR, pois no PARAFAC após a decomposição dos dados a concentração também é estimada por uma regressão “univariada”.

A incerteza para o modelo UPLS, quando este não utiliza a vantagem de segunda ordem, segue a mesma abordagem empregada para o PLS de primeira ordem, a qual foi proposta por Faber e Kowalski<sup>96</sup> através de uma expressão que leva em consideração todas as fontes de erro, que pode ser expressa como<sup>96,97,98</sup>:

$$s(\hat{y}_{un} - y_{un}) = \sqrt{\underbrace{\left( h_{un} + \frac{1}{I_c} \right) \left( V(e) + V(\Delta y) + \|\mathbf{b}_{upls}\|^2 V(\Delta \mathbf{X}) \right)}_{(1)} + \underbrace{V(e_{un}) + \|\mathbf{b}_{upls}\|^2 V(\Delta \mathbf{x}_{un})}_{(2)}}_{(73)}$$

onde  $V(\Delta y)$  indica a variância dos erros dos valores de referência das concentrações da espécie de interesse;  $V(\Delta \mathbf{X})$ ,  $V(\Delta \mathbf{x}_{un})$  as variâncias nas medidas instrumentais das amostras utilizadas na calibração e de uma amostra de previsão, respectivamente;  $V(e)$  e  $V(e_{un})$  são as variâncias do erro de previsão devido apenas ao ajuste do modelo para as amostras de calibração e uma amostra de previsão, respectivamente; a fração  $1/I_c$  entre parênteses leva em conta um modelo centrado na média e deve ser suprimida se não for esse o caso e  $h_{un}$  uma medida da influência (do inglês “leverage”) da amostra  $\mathbf{X}_{un}$ , calculado como:

$$h_{un} = \sum_{i=1}^F \frac{t_{un,i}^2}{\mathbf{t}_i^T \mathbf{t}_i} \quad (74)$$

onde  $\mathbf{t}$  são os escores das amostras de calibração e  $\mathbf{t}_{un}$  o escore da amostra de previsão. Segundo Faber<sup>98</sup>, para melhor compreensão, a equação 73 pode ser dividida em duas partes: (1) o termo multiplicativo que corresponde à contribuição do modelo na etapa de calibração, (2) os termos seguintes que englobam as contribuições da variância do ajuste do modelo e a variação instrumental para a amostra de previsão. Sugerindo ainda que o erro médio quadrático das amostras de calibração (MSEC, do inglês “Mean Square Error of Calibration”) é uma medida que engloba  $\left( V(e) + V(\Delta y) + \|\mathbf{b}_{upls}\|^2 V(\Delta \mathbf{X}) \right)$ , onde o MSEC é definido como:

$$\text{MSEC} = \frac{\sum_{i=1}^{I_c} (\hat{y}_i - y)^2}{I_c - v} \quad (75)$$

onde  $v$  é o número de graus de liberdade perdidos na construção do modelo, que

pode ser estimado como  $I_c$ -F-1 para um modelo centrado na média e  $I_c$ -F para um modelo não centrado, ambas válidas em situações onde o modelo não apresenta sobreajuste. A partir da relação do MSEC com as variâncias no primeiro termo da equação 73 e das considerações que  $V(e_u) \approx V(e)$  e  $V(\Delta \mathbf{x}_{un}) \approx V(\Delta \mathbf{X})$ , Faber e Bro<sup>99</sup> sugeriram uma simplificação da equação 72, conduzindo à expressão:

$$s(\hat{y}_{un} - y_{un}) = \sqrt{\left(1 + h_{un} + \frac{1}{I_c}\right) \text{MSEC} - V(\Delta y)} \quad (76)$$

É importante ressaltar que caso a variância dos erros das concentrações de referência utilizadas na etapa de calibração ( $V(\Delta y)$ ) for insignificante, quando comparada com as demais fontes de erro, a equação 76 se reduz a uma equação proposta por Höskuldsson<sup>100</sup> que é adotada atualmente pela ASTM (norma E1655-00)<sup>7</sup>. Trabalhos utilizando a equação 76 para estimar os intervalos de confiança em modelos de calibração de primeira ordem com PLS já foram descritos na literatura mostrando que os resultados da equação 76 permitem estimar a incerteza de modelos PLS com uma ótima concordância com erros obtidos na prática<sup>23,72,97,98,101</sup>.

Até o momento, os trabalhos publicados com o UPLS em situações que utilizam a vantagem de segunda ordem, que é adquirida com o processo RBL, não apresentaram resultados ou sugestões para a determinação da incerteza das estimativas da concentração, as quais provavelmente devem apresentar uma maior incerteza do que em amostras onde não há presença de interferentes. Uma vez que devido à presença de interferentes tem-se uma queda na sensibilidade, intuitivamente pode-se presumir que as incertezas das estimativas do modelo devem também apresentar um aumento.

A partir da equação 76, não é claro como este aumento na incerteza pode ser levado em conta. Contudo, voltando à equação 73 têm-se as contribuições da etapa de calibração e de previsão em parcelas distintas. Logo, o aumento da incerteza devido à presença de interferentes deve estar incluída apenas na segunda parcela da

equação 73. Lembrando que a sensibilidade em modelos UPLS é dada pelo inverso da norma dos coeficientes de regressão do modelo (equações 60 e 65), o aumento da incerteza do modelo devido à presença dos interferentes pode ser medido através da utilização da norma dos coeficientes de regressão  $\mathbf{b}_{\text{RBL}}$  no lugar de  $\mathbf{b}_{\text{UPLS}}$  na segunda parcela da equação 73. Assumindo ainda que nessa situação as considerações  $V(\mathbf{e}_u) \approx V(\mathbf{e})$  e  $V(\Delta\mathbf{x}_{\text{un}}) \approx V(\Delta\mathbf{X})$  são válidas, chega-se à equação:

$$s(\hat{y}_{\text{un}} - y_{\text{un}}) = \sqrt{\left(h_{\text{un}} + \frac{1}{I_c}\right) \text{MSEC} + \left(V(\mathbf{e}) + \|\mathbf{b}_{\text{RBL}}\|^2 V(\Delta\mathbf{X})\right)} \quad (77)$$

onde  $V(\Delta\mathbf{X})$  deve ser estimado por medidas em replicatas e  $V(\mathbf{e})$ , segundo Faber<sup>97</sup>, estimado por:

$$V(\mathbf{e}) = \text{MSEC} - \left[V(\Delta y) + \|\mathbf{b}_{\text{UPLS}}\|^2 V(\Delta\mathbf{X})\right] \quad (78)$$

onde a equação 77 proposta, que é uma simplificação da equação geral de Faber e Kowalski<sup>96</sup> e deve ser válida quando o UPLS utiliza a vantagem de segunda ordem.

Para o modelo BLLS, Linder e Sundberg<sup>43</sup> deduziram e propuseram em seu trabalho inicial uma equação para a determinação de sua incerteza, mas apenas em situações onde a vantagem de segunda ordem não é aplicada<sup>12</sup>. Provavelmente devido a essa limitação, esta equação somente foi aplicada nos seus dois primeiros trabalhos publicados sobre o BLLS<sup>43,44</sup>.

Os resultados de estimativas de incerteza para os modelos que serão abordados nas aplicações descritas nos capítulos que se seguem serão restritas aos modelos PARAFAC, MCR e UPLS. Será considerada a equação 71 como uma estimativa nos modelos PARAFAC e MCR, enquanto que as equações 76 e 77 para o modelo UPLS nas situações onde este não emprega e emprega a vantagem de segunda ordem, respectivamente.

Uma vez estimada a incerteza na estimativa da concentração, o intervalo de confiança (IC) definido como o intervalo no qual se pode afirmar com certo grau de

confiança, ou probabilidade, que inclui o valor verdadeiro da propriedade de interesse podem ser obtidos para todos os modelos da mesma forma como é feita para calibração de ordem zero, como:

$$IC_{un} = \pm \left( t_{1-\alpha/2, v} \right) s(\hat{y}_{un} - y_{un}) \quad (79)$$

onde  $t_{(1-\alpha)/2, v}$  é o parâmetro t da distribuição t-Student com nível de confiança  $(1-\alpha)/2$  e  $v$  é o número de graus de liberdade. É importante ressaltar que os IC calculados com base na equação 79 admitem que a presença de erros sistemáticos não seja significativa, os erros obtidos em relação às concentrações estimadas e referência não são correlacionados e possuem uma distribuição que se aproxima da distribuição Normal.

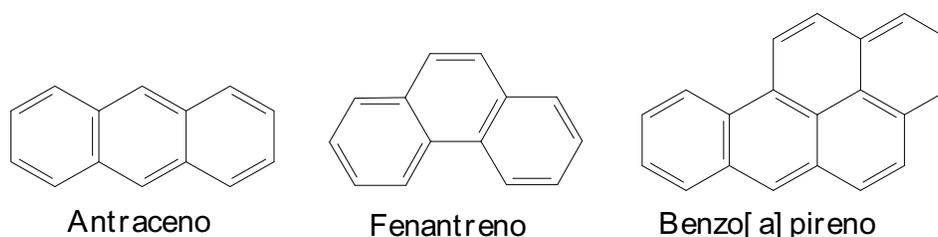
**Capítulo 2 - Estudo sobre a determinação de**  
**hidrocarbonetos policíclicos aromáticos em**  
**misturas**

O primeiro estudo de caso abordado neste trabalho consistiu no desenvolvimento de calibrações de segunda ordem utilizando PARAFAC, BLLS e UPLS para determinação de hidrocarbonetos policíclicos aromáticos em misturas ternárias por emissão e excitação de fluorescência (EEF). A determinação dessa classe de compostos por EEF já foi descrita em diversos trabalhos na literatura, inclusive com utilização de modelos de calibração de segunda ordem<sup>102,103,104,105</sup>. O grande interesse na determinação desses compostos justifica-se pela sua alta toxicidade e por serem carcinogênicos, sendo a EEF muito eficiente nessas determinações devido sua alta seletividade e alta intensidade de fluorescência que estes compostos apresentam.

Neste estudo em particular, os objetivos principais foram determinar e comparar algumas das figuras de mérito dos modelos BLLS e UPLS em relação ao modelo PARAFAC, que é geralmente empregado nesse tipo de dados. Foram comparados os parâmetros sensibilidade, seletividade e incerteza, sendo nesses dois últimos parâmetros apresentados e comparados os resultados duas sugestões propostas nesse trabalho para sua determinação no modelo UPLS. Ainda nesse estudo é verificado se o número de fatores necessários no processo RBL, utilizado para modelar espécies não calibradas, pode ser estimado com segurança pela comparação dos resíduos de decomposição das amostras de previsão com uma estimativa do ruído instrumental. Nesse sentido, foi sugerida e testada a utilização de dois limites críticos para o resíduo das amostras de previsão para guiar a escolha do número de fatores do RBL no BLLS e UPLS, baseados nos intervalos de confiança de uma estimativa do ruído instrumental ou do resíduo médio das amostras de calibração. Com base nesses objetivos, foram escolhidos como espécies de interesse os compostos antraceno (AN), benzo[a]pireno (BAP) e fenantreno (FEN), cujas fórmulas estruturais são apresentadas na Figura 6. Esses compostos foram escolhidos por apresentarem emissão de fluorescência em uma

região semelhante de comprimentos de onda de excitação e emissão, de modo que seus sinais estão parcialmente sobrepostos e, portanto, constituem um sistema interessante para o estudo.

Para possibilitar uma melhor avaliação dos modelos, foram estudadas sete situações, sendo três delas contendo como analito cada uma das espécies (AN, BAP e FEN), três contendo as misturas binárias AN|BAP, AN|FEN e BAP|FEN e a última consistindo da mistura ternária AN|BAP|FEN, além de misturas ternárias para a validação destas calibrações. Deste modo, têm-se sete sistemas que permitem o estudo dos modelos de calibração de segunda ordem quando as misturas de validação são analisadas nas situações em que: (a) três espécies de interesse calibradas e nenhum interferente está presente nas amostras de validação, (b) duas espécies calibradas e um interferente e (c) uma espécie calibrada e dois interferentes.



**Figura 6-** Fórmulas estruturais das espécies de interesse.

## 2.1 Procedimento experimental.

As soluções utilizadas para a construção das curvas analíticas e misturas de validação foram preparadas a partir dos reagentes antraceno (96,0 %, Fluka), benzo[a]pireno (97,0 %, Aldrich), fenantreno (97,0 %, Fluka), acetonitrila grau HPLC (99,99 %, Tedia), água destilada e material volumétrico devidamente calibrado.

Foram preparadas soluções estoque de cada um dos compostos de interesse em balões de 100,0 mL a partir da dissolução de 20,00 mg de cada um dos

compostos de interesse em acetonitrila. Em seguida foram realizadas duas diluições apropriadas dessas soluções utilizando como solventes misturas acetonitrila:água 50:50 (v/v) e 25:75 (v/v), respectivamente. Foram obtidas assim soluções de trabalho nas concentrações de 0,100, 0,150 e 0,500 mg/L para AN, BAP e FEN, respectivamente.

As soluções das curvas de calibração e misturas de validação foram constituídas por cinco e quatro níveis distintos de concentração, respectivamente, que foram preparados por diluições das soluções de trabalho em balões de 25,0 mL e volumes medidos com micropipetas de volume variável e capacidades máximas de 1,000 a 5,000 mL, cujo planejamento e concentrações são apresentadas na Tabela 2.

**Tabela 2.** Planejamento e níveis de concentração utilizados para construção das curvas de calibração e misturas de validação.

Solução	Espécies de interesse ( $\mu\text{g/L}$ )		
	AN	BAP	FEN
P1	2,50 ( $\pm 0,04$ )	28,0 ( $\pm 0,2$ )	65,5 ( $\pm 0,3$ )
P2	7,5 ( $\pm 0,1$ )	23,5 ( $\pm 0,2$ )	13,0 ( $\pm 0,3$ )
P3	17,5 ( $\pm 0,1$ )	18,5 ( $\pm 0,1$ )	38,5 ( $\pm 0,2$ )
P4	12,0 ( $\pm 0,1$ )	14,0 ( $\pm 0,1$ )	93,0 ( $\pm 0,4$ )
P5	24,5 ( $\pm 0,2$ )	9,5 ( $\pm 0,1$ )	129,0 ( $\pm 0,4$ )
V1	19,5 ( $\pm 0,2$ )	11,5 ( $\pm 0,1$ )	103,0 ( $\pm 0,4$ )
V2	4,87 ( $\pm 0,05$ )	16,2 ( $\pm 0,1$ )	51,7 ( $\pm 0,3$ )
V3	9,8 ( $\pm 0,1$ )	21,0 ( $\pm 0,2$ )	25,9 ( $\pm 0,2$ )
V4	14,5 ( $\pm 0,1$ )	25,5 ( $\pm 0,2$ )	77,8 ( $\pm 0,4$ )

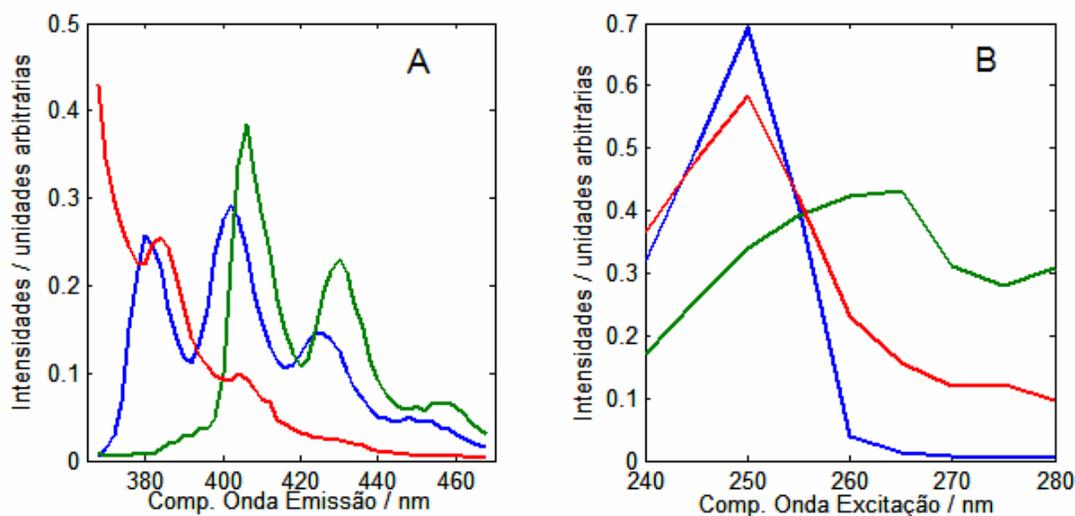
P: soluções padrões para a calibração, V: misturas para a validação. Concentrações e estimativas dos erros do preparo das soluções entre parênteses, obtidos por propagação de erro.

As superfícies de EEF foram adquiridas entre os comprimentos de onda de 230-400 nm e 240-600 nm com passos de 5 e 2 nm para excitação e emissão,

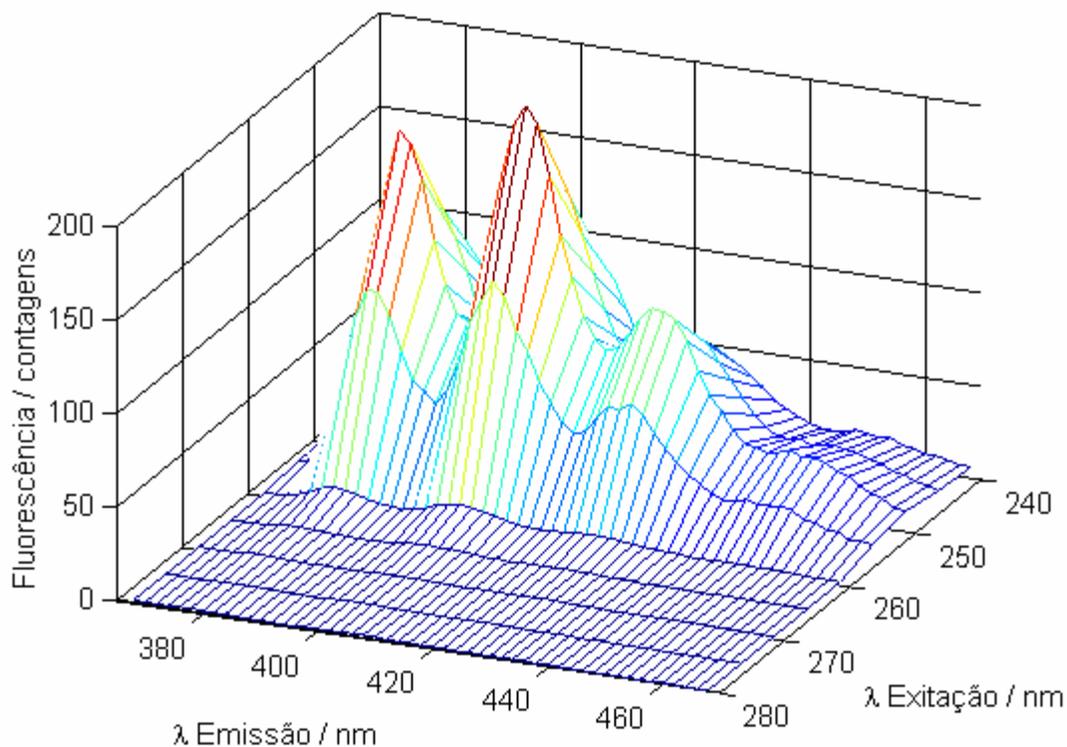
respectivamente. Foi utilizado um espectrofluorímetro Cary Eclipse da marca Varian, com as aberturas das fendas de ambos os monocromadores de excitação e emissão em 5 nm, varredura de modo mais rápido (“fastest”), ganho da fotomultiplicadora no nível médio e cubeta de 1 cm de caminho óptico. Foi realizada apenas uma leitura de cada solução, sendo preparadas 21 replicatas para cada mistura de validação e apenas 1 solução para cada padrão de calibração.

## 2.2 Resultados e discussão.

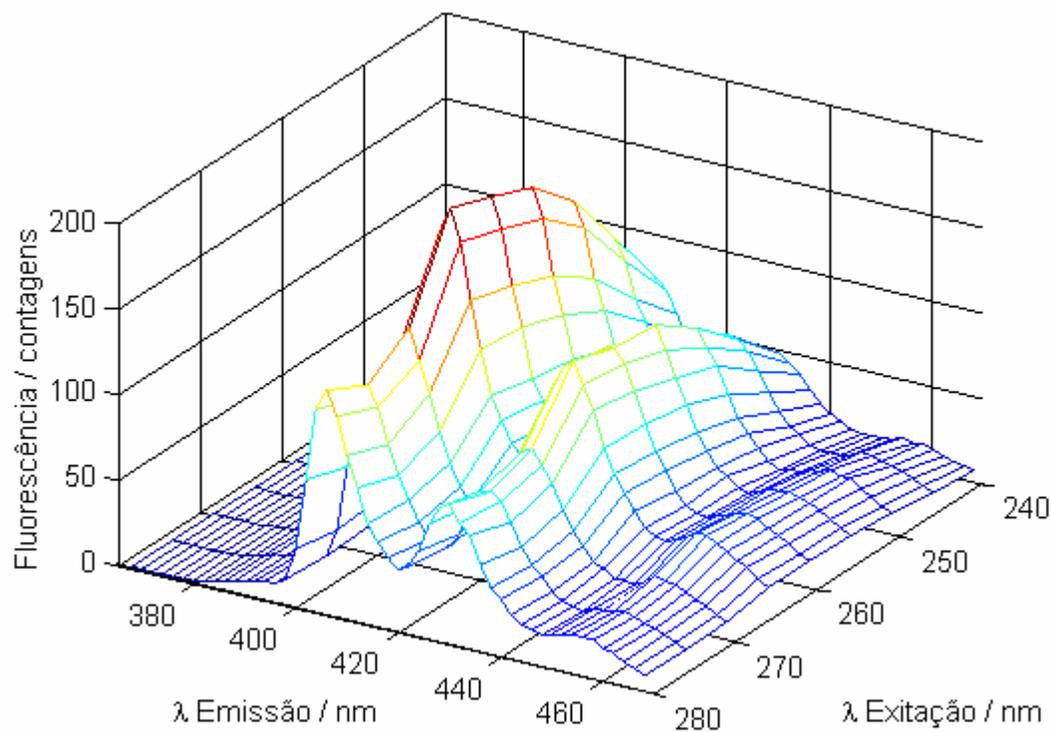
Primeiramente, para a construção dos modelos foi selecionada a região entre 240-280 nm para excitação e 368-468 nm para emissão, a qual apresenta os três compostos de interesse com espectros parcialmente sobrepostos e exclui espalhamentos Rayleigh e Raman ou outras ordens de difração da grade de difração. Cada matriz de dados foi constituída por 9 comprimentos de onda de excitação e 51 de emissão. As Figuras de 7 a 10 apresentam os espectros normalizados e superfícies obtidas para as soluções padrões de maior concentração que contém cada analito isoladamente, nas quais podem-se observar as regiões em que estes se sobrepõem.



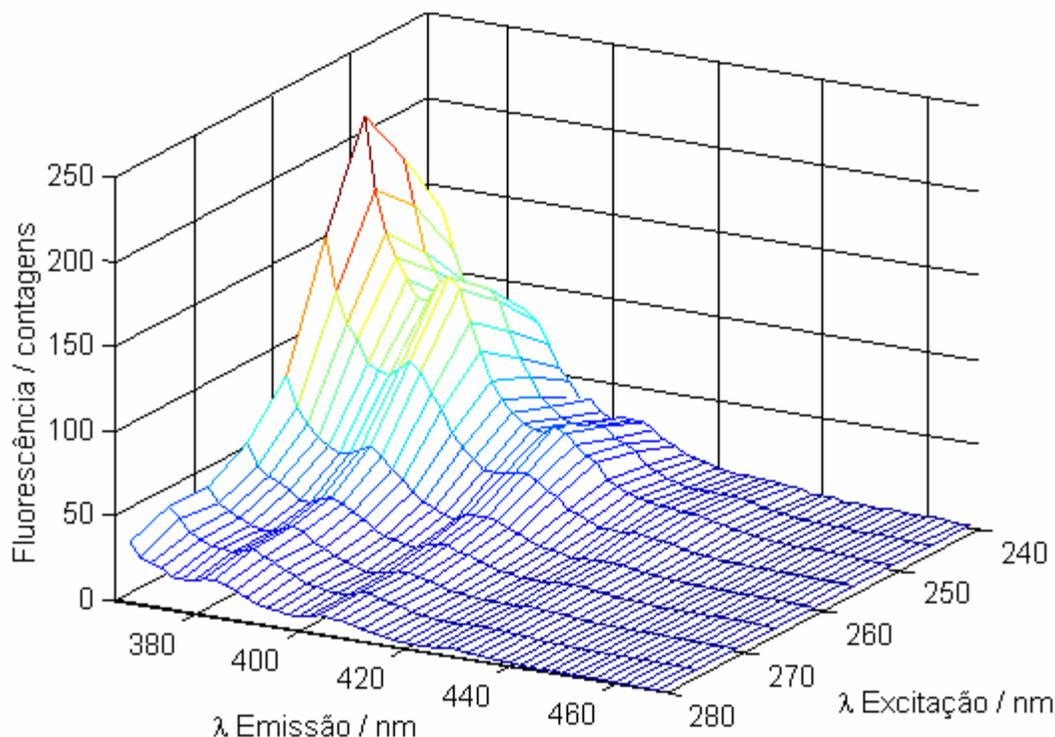
**Figura 7-** Espectros de fluorescência de (A) emissão e (B) excitação normalizados, para as espécies de interesse no intervalo de comprimentos de onda ( $\lambda$ ) que foi selecionado para a construção das calibrações. (—) AN, (—) BAP e (—) FEN.



**Figura 8-** Superfície de fluorescência de uma solução de AN na concentração de 24,5 µg/L.



**Figura 9-** Superfície de fluorescência de uma solução de BAP na concentração de 28,0 µg/L.



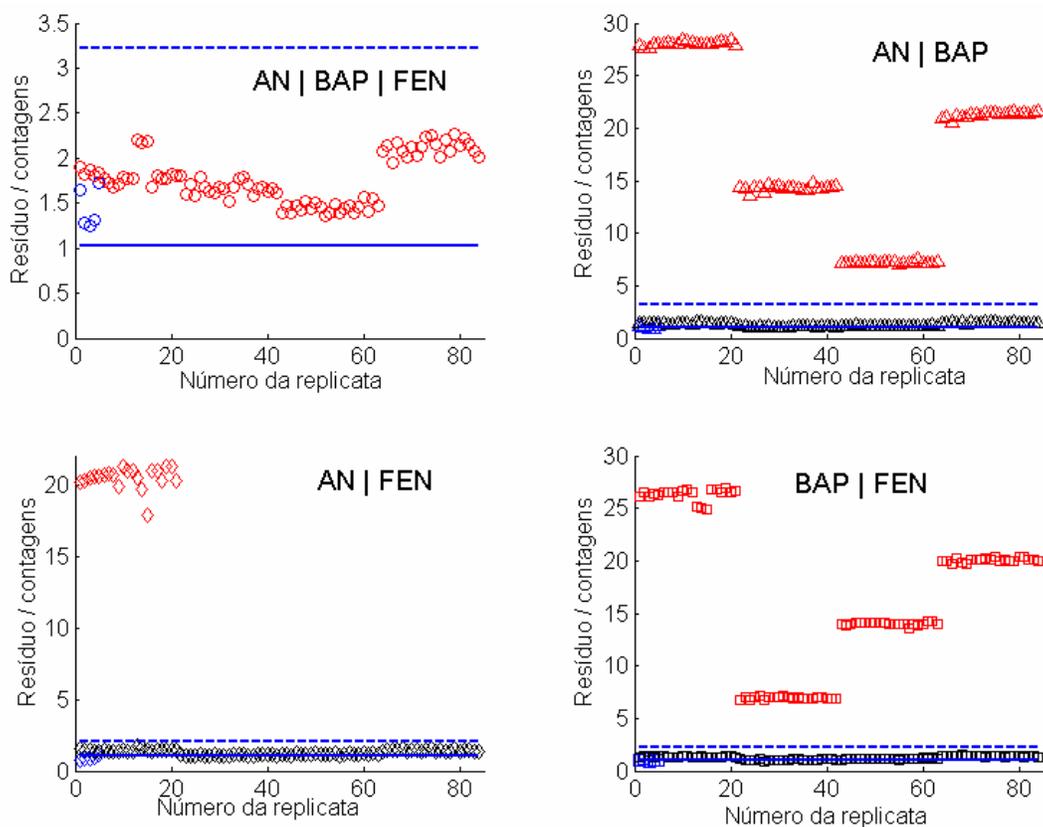
**Figura 10-** Superfície de fluorescência de uma solução de FEN na concentração de 129,0 µg/L.

### 2.2.1 Escolha do número de fatores utilizados no RBL.

Para construção das calibrações com o BLLS, a escolha do número de fatores restringe-se à que é utilizada no RBL, uma vez que na etapa de calibração o número de fatores já é automaticamente determinado pelo número de espécies presentes. No caso do UPLS, a escolha do número de fatores envolve duas etapas. A primeira consiste da escolha do número de fatores do modelo UPLS (as variáveis latentes), as quais modelam as espécies presentes nas amostras de calibração, geralmente estimado por procedimentos de validação cruzada<sup>57</sup>. Para este estudo de caso especificamente foi observado que o melhor número de variáveis latentes é igual ao número de espécies presentes nas amostras de calibração. Na segunda etapa, é escolhido o número de fatores que será utilizado no RBL, o qual a princípio pode ser diferente para cada amostra uma vez que a concentração do

analito nas amostras de previsão é estimada de forma seqüencial.

Como descrito nas seções 1.1.3 e 1.1.4, o número de fatores do RBL é estimado observando os resíduos de decomposição para as amostras de validação, obtidos pela equação 23. A Figura 11 apresenta o gráfico desses resíduos para o BLLS nas calibrações quando três ou dois analitos estão presentes nos padrões.



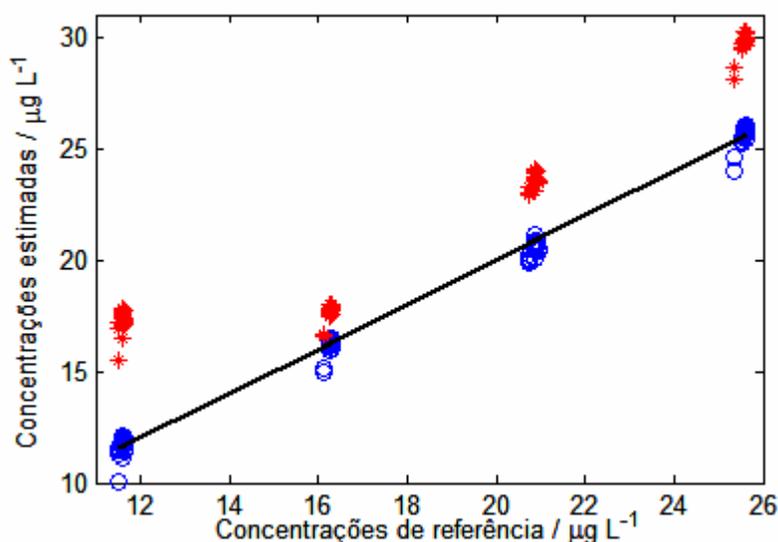
**Figura 11-** Resíduos de decomposição para o BLS nas situações onde três ou dois analitos estão calibrados. (azul) amostras de calibração, (vermelho) validação sem utilizar o RBL, (preto) adicionando um fator no RBL, (—) valor crítico pelo ruído instrumental e (---) valor crítico pelo resíduo médio da calibração, ambos com 99% de confiança. Cada 21 replicatas em seqüência correspondem a 1 amostra de validação.

Para a regressão AN|BAP|FEN, observa-se que os resíduos de decomposição das amostras de validação estão no mesmo patamar das amostras de calibração e abaixo do valor crítico, com 99% de confiança, estimado pelo resíduo médio das amostras de calibração. Esta observação está de acordo com o esperado, uma vez que as amostras de validação contém apenas as mesmas espécies presentes na

calibração. Contudo, ambos os resíduos das amostras de calibração e validação estão acima do valor crítico obtido pelo ruído instrumental (estimado em 0,46 unidades de fluorescência por 15 replicatas do branco), mostrando que este limite não foi coerente para este estudo.

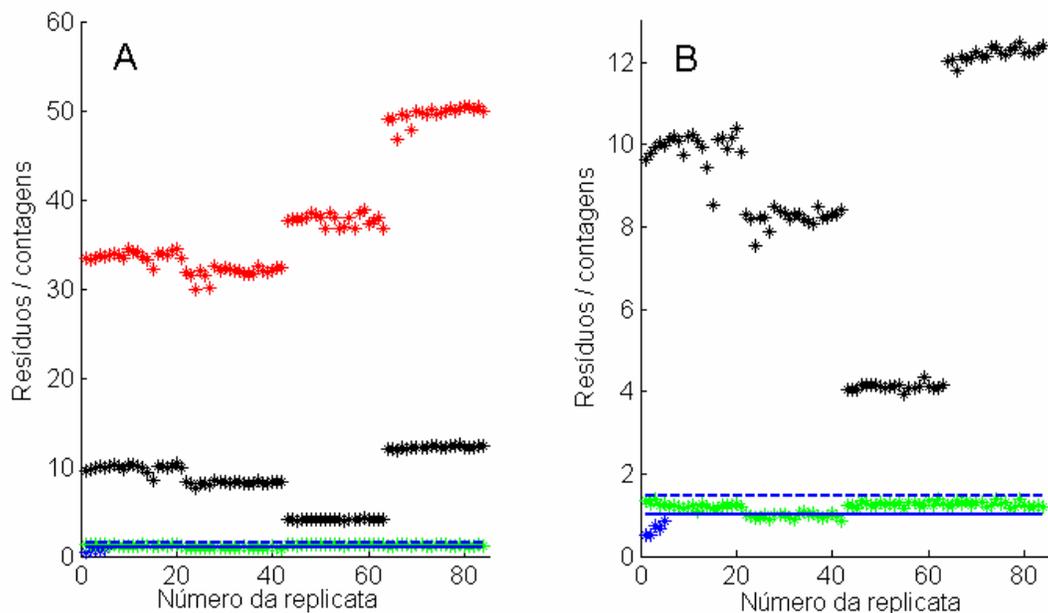
Para as curvas AN|BAP, AN|FEN e BAP|FEN observam-se resíduos significativamente mais altos que os observados nas amostras de calibração, quando não é adicionado nenhum fator ao RBL. Após a adição de um fator, os resíduos das amostras de validação reduzem-se praticamente para o mesmo patamar dos resíduos observados na calibração e abaixo do valor crítico obtido pelo resíduo médio da calibração, mas novamente acima do valor crítico obtido pelo ruído. Esses resultados levam a crer que o resíduo médio da calibração fornece um limite mais adequado para a escolha do número de fatores para o RBL.

A diminuição dos resíduos observada na Figura 11 demonstra a eficiência do RBL para modelar a espécie que não está presente nas amostras de calibração, o que é refletido nos erros da estimativa da concentração das espécies de interesse que são muito altos quando nenhum interferente é adicionado ao RBL nas curvas que contém apenas dois analitos calibrados. Na Figura 12 pode-se observar claramente a melhora nos resultados pela dispersão das amostras de validação com e sem a adição de um fator no RBL para a determinação de BAP, na situação em que BAP|FEN estão calibrados e AN está presente como interferente, onde os erros médios relativos são aproximadamente 20 e 3% com e sem a utilização do RBL, respectivamente. É importante ressaltar que em todas as situações onde se tinha espécies não calibradas os erros obtidos sem utilização do RBL foram de, no mínimo, 20 %, como exemplificado na situação da Figura 12, os quais são reduzidos significativamente com a utilização do RBL (ver Tabelas 4, 5 e 6 e discussão dos resultados na próxima seção).

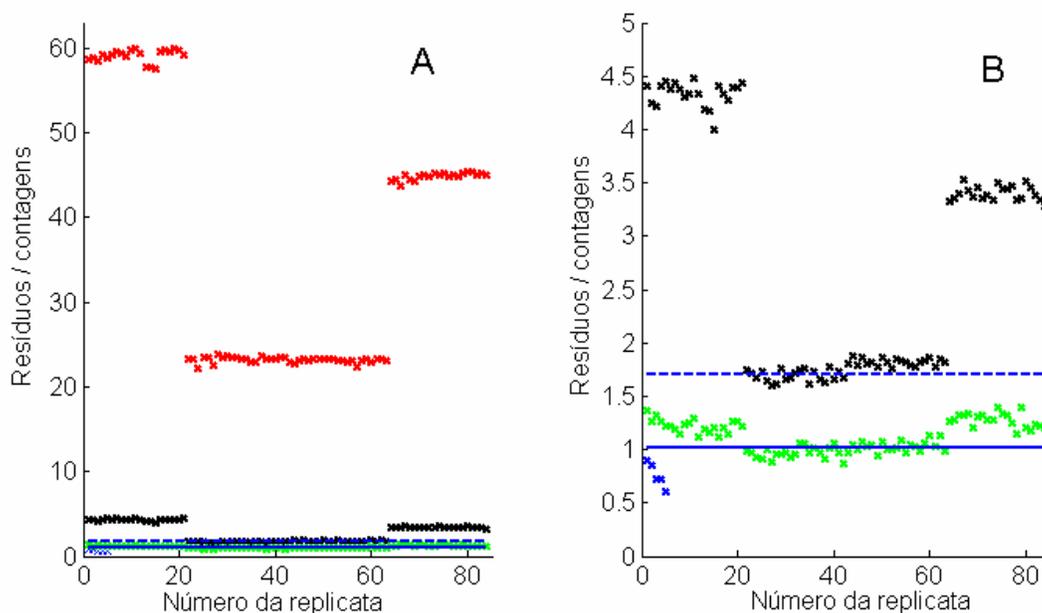


**Figura 12-** Concentrações de referência versus estimadas para as amostras de validação para o BAP na situação BAP|FEN. (\*) estimativas sem utilizar o RBL, (o) estimativas após a adição de um fator no RBL e (—) é a bissetriz do quadrante.

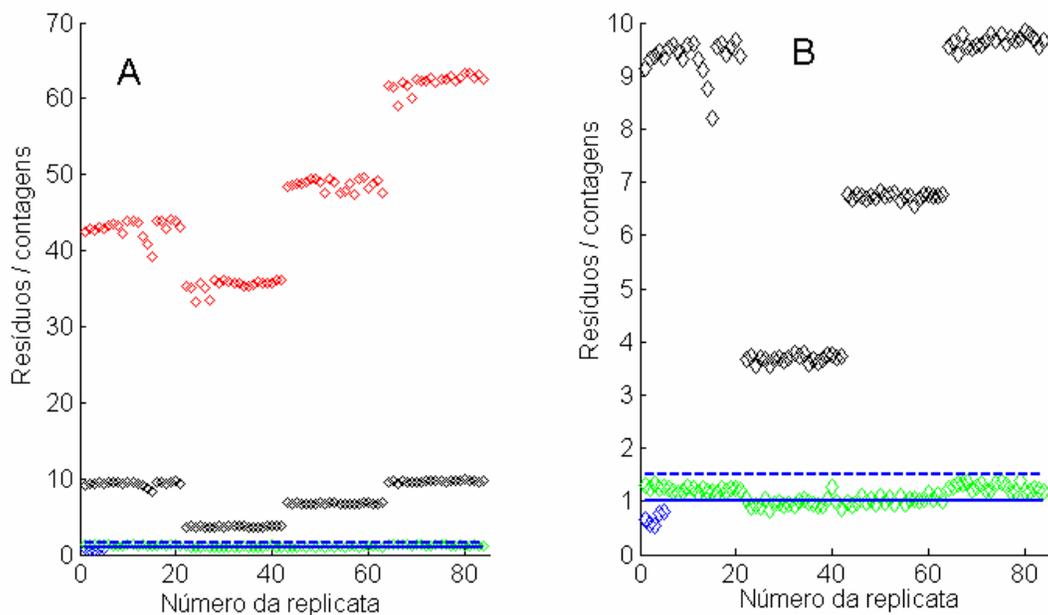
As Figuras 13 a 15 apresentam os gráficos de resíduos nas curvas onde apenas AN ou BAP ou FEN estão calibrados, respectivamente, tendo-se, portanto, dois interferentes nessas situações. Como esperado, os resíduos sem a utilização do RBL são maiores do que nas curvas com dois analitos calibrados. Quando apenas AN está calibrado, observa-se uma queda significativa nos resíduos pela adição de um fator no RBL, mas os resíduos ainda permanecem acima dos valores críticos, indicando a necessidade de outro fator. Quando outro fator é adicionado, obtêm-se uma redução dos resíduos de todas as amostras, mas bons resultados para as estimativas de concentração para apenas 3 das 4 amostras de validação (replicatas 1 a 42 e 64 a 84). Para a amostra 3 (replicatas 43 a 63), com apenas um fator são obtidos erros relativos em relação aos valores de referência em torno de 25 % (o que confirma a necessidade de outro fator), mas quando um novo fator é adicionado são obtidos valores absurdos nas estimativas das concentrações, indicando possível divergência do RBL.



**Figura 13-** (A) Resíduos de decomposição para o BLS na calibração com apenas AN calibrado. (B) Ampliação de A. (azul) amostras de calibração, (vermelho) validação sem utilizar o RBL, (preto) utilizando um fator no RBL, (verde) dois fatores no RBL, (—) valor crítico pelo ruído instrumental e (---) valor crítico pelo resíduo médio da calibração, com 99% de confiança. Cada 21 replicatas em seqüência correspondem a 1 amostra de validação.



**Figura 14-** (A) Resíduos de decomposição para o BLS na calibração com apenas BAP calibrado. (B) Ampliação de A. Símbolos definidos como na Figura 13.



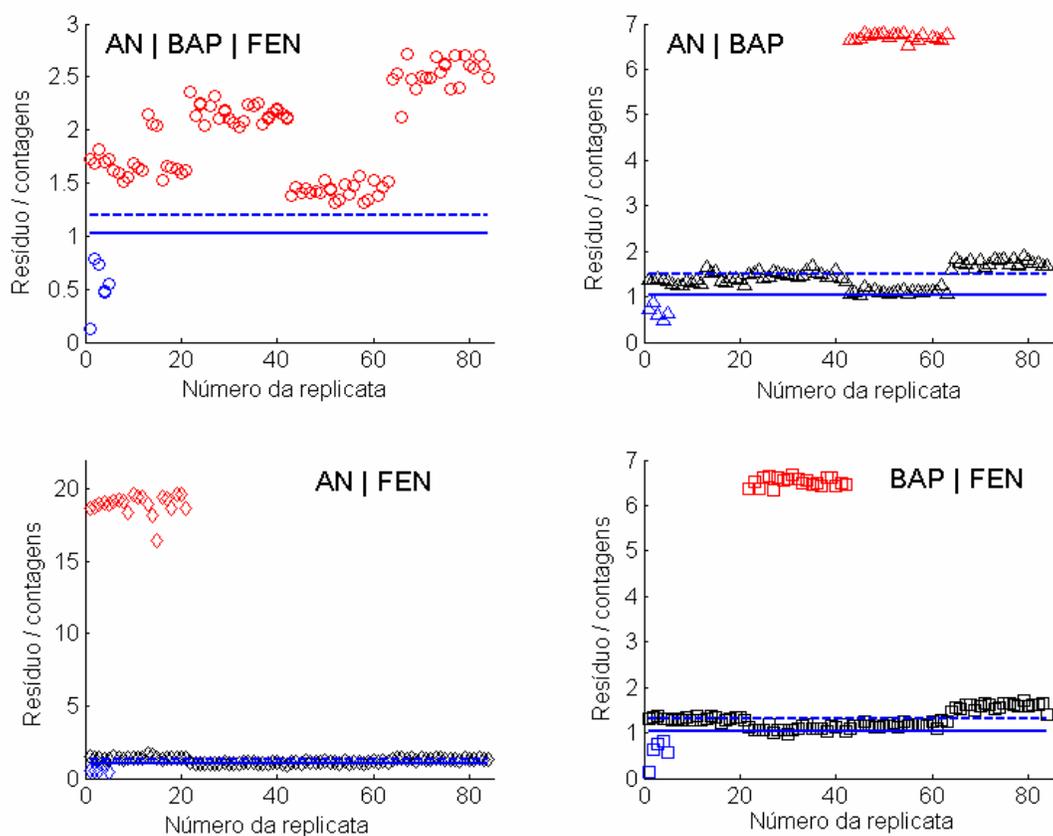
**Figura 15-** (A) Resíduos de decomposição para o BLS na calibração com apenas FEN calibrado. (B) Ampliação de A. Símbolos definidos como na Figura 13.

Para a os resíduos da curva onde apenas BAP está calibrado (Figura 14), observa-se que, com a adição de um fator no RBL, o resíduo das amostras de validação 2 e 3 (replicatas de 22 a 63) apresentam valores que estão abaixo do limite crítico calculado pela calibração para algumas replicatas, indicando que talvez para essas amostras um novo fator não seja necessário. Observando o erro médio relativo (REP) tem-se um REP de 25 % para a primeira amostra (replicatas de 1 a 21), enquanto que para as amostras 2, 3 e 4 (replicatas de 22 a 84) observa-se um erro médio de aproximadamente 4 %. Contudo, com apenas 1 fator foram observados erros sistemáticos negativos em todas as replicatas. Com a adição de um segundo fator no RBL o REP, considerando todas as 84 replicatas, é 4,1 % e obtém-se uma melhor distribuição dos erros, havendo também a redução dos resíduos para abaixo do valor crítico estabelecido pelo resíduo da calibração, mas ainda acima do limite pelo ruído instrumental.

Para a curva que contém apenas FEN (Figura 15), observa-se um comportamento semelhante ao descrito na Figura 13 para AN. Contudo, neste caso

o RBL apresenta bons resultados em 83 das 84 replicatas das amostras de validação utilizando dois fatores no RBL, sendo observada divergência para a replicata número 40. Ressalta-se que a divergência observada se refere a obtenção de um valor absurdo para a estimativa da concentração da replicata em questão, e não em relação ao valor do resíduo de decomposição dessa amostra, que está compatível com as demais amostras de validação.

A partir desses resultados observa-se que para o BLLS o valor crítico obtido a partir do resíduo médio das amostras de calibração permite estabelecer um critério mais coerente para a escolha do número de fatores do RBL do que o valor crítico obtido pelo ruído instrumental.



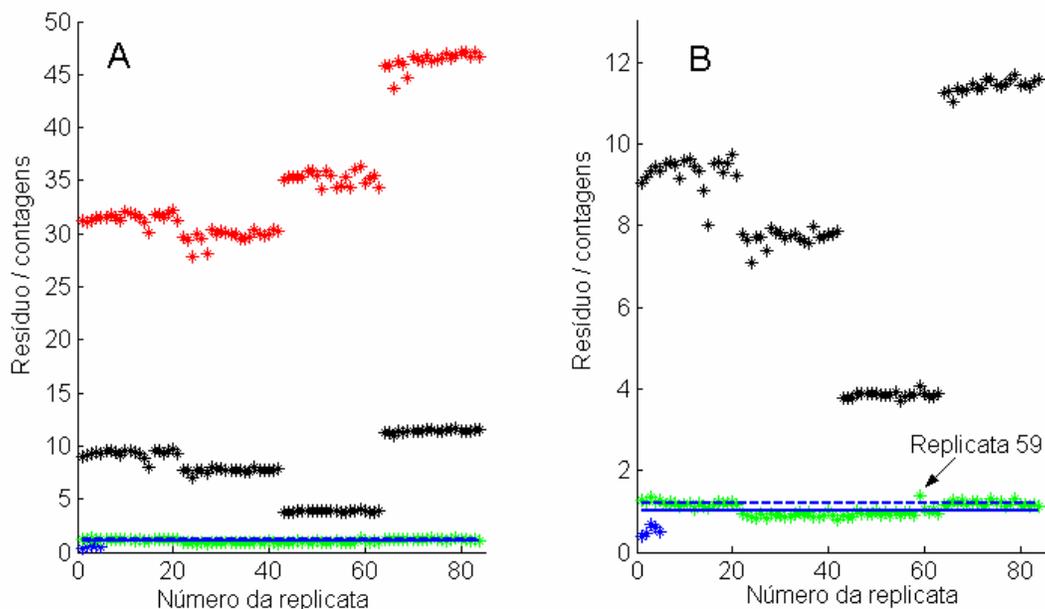
**Figura 16-** Resíduos de decomposição para o UPLS nas situações onde três ou dois analitos estão calibrados. (azul) amostras de calibração, (vermelho) validação sem utilizar o RBL, (preto) adicionando um fator no RBL, (—) valor crítico pelo ruído instrumental e (---) valor crítico pelo resíduo médio da calibração, ambos com 99% de confiança. Cada 21 replicatas em seqüência correspondem a 1 amostra de validação.

A Figura 16 mostra os gráficos de resíduos obtidos para o UPLS nas curvas calibrando três ou dois analitos de cada vez, onde os maiores resíduos obtidos para três das quatro amostras de validação quando o RBL não é utilizado foram omitidos para melhor clareza na parte inferior de gráfico. A principal diferença observada em relação aos gráficos do BLLS é que o valor crítico obtido pelo resíduo médio da calibração é menor para o UPLS do que para o BLLS, mas ainda é superior ao obtido pelo ruído instrumental. Observa-se que, na curva onde os três analitos estão calibrados, as amostras de validação apresentam resíduos superiores a ambos os limites críticos. A explicação para esse fato é que modelos PLS tendem a ajustar mais as amostras de calibração fornecendo, assim, resíduos menores.

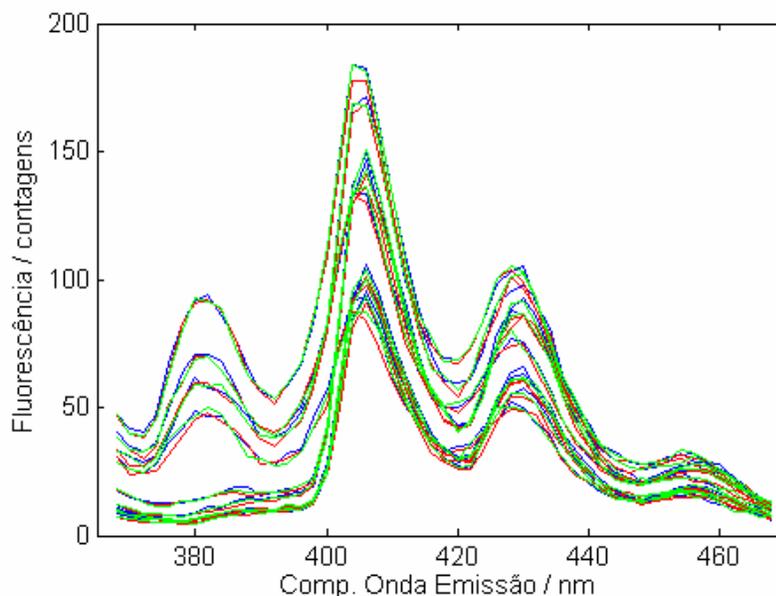
Nas curvas onde dois dos analitos estão calibrados observa-se que, após a adição de um interferente no RBL, a maior parte das replicatas apresenta resíduos abaixo do valor crítico obtido pela calibração, sendo obtidos bons resultados para as estimativas da concentração dos analitos. Caso outro interferente seja adicionado observa-se que para algumas replicatas valores semelhantes aos anteriores, enquanto que em outras o RBL apresenta estimativas das concentrações que fornecem REP da ordem de 50 %, indicando divergência do RBL. Portanto, nessas situações foi escolhido apenas um fator para as curvas com dois analitos calibrados, como esperado. Essas observações mostram que o limite estabelecido para o UPLS não fornece a mesma certeza na escolha do número de fatores observada no BLLS.

Na calibração do AN (Figura 17), observa-se claramente a necessidade de dois fatores no RBL, sendo a maior parte dos resíduos das amostras de validação se reduzem para abaixo do valor crítico estabelecido pelas amostras de calibração quando dois fatores são incluídos no RBL, mas ainda acima do limite pelo ruído instrumental. Com dois fatores, as estimativas das concentrações apresentam bons resultados, com exceção da replicata 59, onde o RBL diverge sem motivo aparente. A Figura 18 mostra os espectros de emissão das replicatas 58, 59 e 60, que

confirmam que a replicata 59 não apresenta diferença das demais replicatas e sugere que a divergência é decorrente de uma instabilidade do RBL. É interessante lembrar que no BLLS todas as replicatas dessa amostra apresentaram divergência.

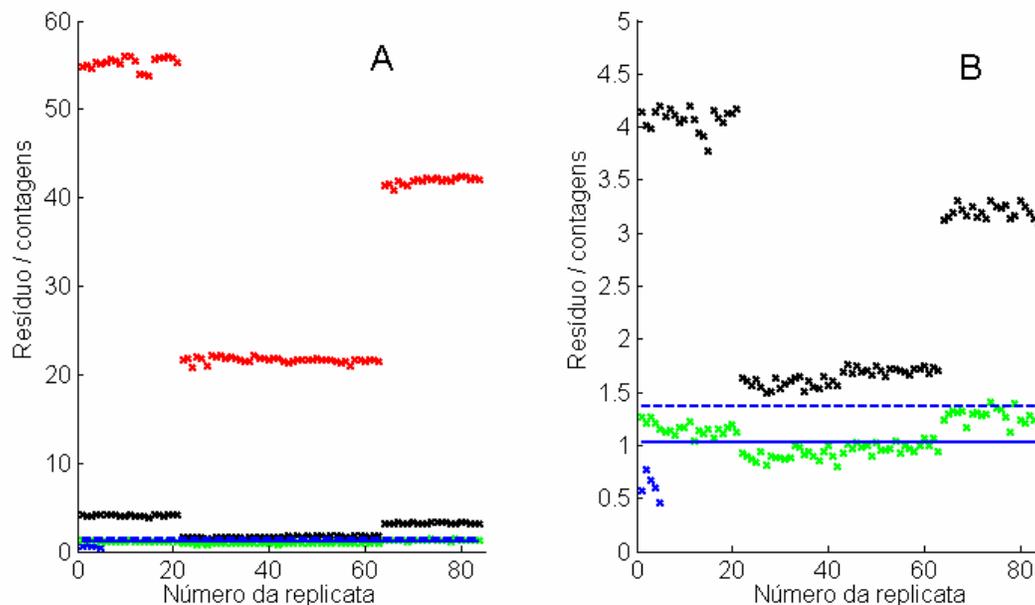


**Figura 17-** (A) Resíduos de decomposição para o UPLS na calibração com apenas AN calibrado. (B) Ampliação de A. Símbolos definidos como na Figura 16.



**Figura 18-** Espectros de emissão nos 9 diferentes comprimentos de onda de excitação utilizados. Replicatas (—) 58, (—) 59 e (—) 60, da terceira amostra de validação.

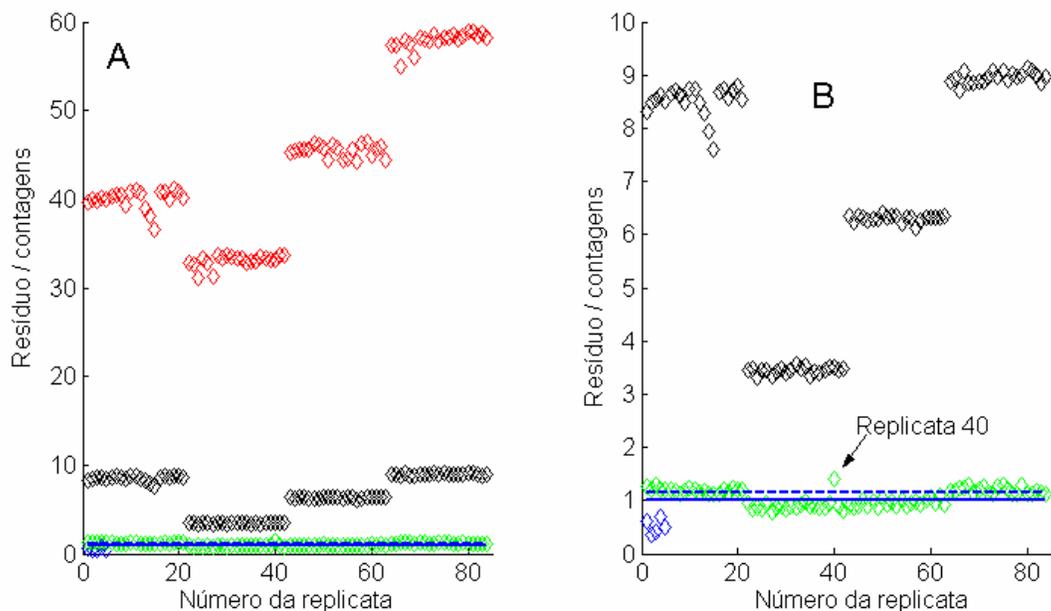
A Figura 19 mostra os resíduos para a curva com apenas BAP calibrado. Nesta situação, com dois interferentes obtêm-se bons resultados para as estimativas de concentração e valores coerentes dos resíduos para quase todas as replicatas, sem a observação de divergências no RBL.



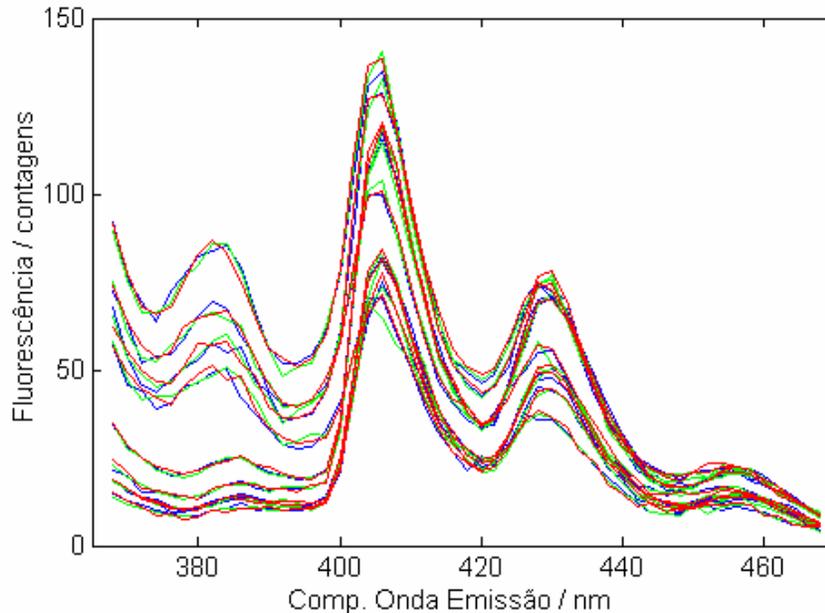
**Figura 19-** (A) Resíduos de decomposição para o UPLS na calibração com apenas BAP calibrado. (B) Ampliação de A. Símbolos definidos como na Figura 13.

A Figura 20 apresenta os resíduos para a curva com apenas FEN, os quais têm aproximadamente o mesmo perfil observado da Figura 17B, após a adição de dois fatores no RBL, sendo observada uma divergência na replicata 40. Assim como na replicata 59, a Figura 21 mostra que a replicata 40 não apresenta motivo aparente para a divergência do RBL. No BLLS, o RBL também apresenta divergência para essa situação nessa replicata. Verificando a Figura 15, observa-se que a replicata 40 apresenta um resíduo distinto das demais replicatas dessa mesma amostra. Esses resultados sugerem que o algoritmo do RBL que se baseia na minimização dos resíduos das amostras de validação, apesar de apresentar ótimos resultados na maior parte das situações, ainda requer melhoramentos quando mais

de um fator é necessário.



**Figura 20-** (A) Resíduos de decomposição para o UPLS na calibração com apenas FEN calibrado. (B) Ampliação de A. Símbolos definidos como na Figura 13.



**Figura 21-** Espectros de emissão nos 9 diferentes comprimentos de onda de excitação utilizados. Replicatas (—) 39, (—) 40 e (—) 41, da terceira amostra de validação.

Quanto aos valores críticos sugeridos para a escolha do número de fatores no RBL, pode-se concluir que os limites estabelecidos pelo resíduo médio das

amostras de calibração apresentam maior coerência com o número de fatores escolhidos. Para o BLS os limites críticos puderam identificar melhor o número de fatores necessários no RBL. A utilização do ruído instrumental nesse estudo mostrou-se ineficiente para estabelecer um limite crítico simplesmente baseando-se na hipótese de Normalidade e seu valor com 99 % de confiança. Esse resultado ilustra a dificuldade de utilizar este parâmetro, uma vez que estimativas menores ou maiores que os valores verdadeiros da amplitude do ruído podem ser feitas dependendo do número de medidas ou da qualidade das mesmas.

Desconsiderando as replicatas onde o RBL divergiu, o número de fatores foi igual ao esperado para cada situação tanto com o BLS como o UPLS. Assim, o número de fatores para a construção dos modelos foi igual ao número de espécies presentes nas amostras de calibração e o número de fatores no RBL igual ao número de espécies nas amostras de validação que não eram calibradas.

### **2.2.2 Comparação dos modelos BLS, UPLS e PARAFAC e figuras de mérito.**

Os resultados discutidos a seguir foram obtidos a partir do número de fatores escolhidos na seção anterior para o BLS e UPLS, utilizando, ainda como pré-processamento, os dados centrados na média, o qual também foi empregado para a escolha de fatores na seção anterior.

Para construção do modelo PARAFAC nas situações onde os três ou dois dos analitos estavam calibrados foram utilizadas como restrições não negatividade nas três dimensões, inicialização por SVD e uso de 3 fatores em todos os modelos. Nas curvas apresentando apenas um analito calibrado, o procedimento adotado envolveu uma decomposição inicial com as amostras de calibração e uma de validação, onde foram obtidas estimativas dos pesos **A**, **B** e **C** do PARAFAC. Em seguida, esses pesos foram utilizados como estimativas iniciais das decomposições

de cada amostra de validação, procedimento que é conhecido como “inicialização por pesos antigos”. Essa estratégia foi utilizada devido à dificuldade do PARAFAC em fornecer resultados satisfatórios quando apenas uma espécie está calibrada e têm-se dois interferentes.

Na Tabela 3 são apresentados os valores de SEN e SEL obtidos nas situações onde os modelos BLLS, UPLS e PARAFAC não utilizam a vantagem de segunda ordem (não há interferentes não calibrados), os quais foram obtidos através da estimativa da concentração das próprias amostras utilizadas na construção do modelo. Para as calibrações que contém apenas AN, BAP e FEN evidentemente não há necessidade da utilização de modelos de calibração de segunda ordem, pois tem-se apenas uma espécie presente no meio e, nesse caso, pode-se utilizar uma simples calibração de ordem zero (univariada). Os modelos foram construídos nessas situações apenas para obtenção dos valores de SEN e SEL. Para as condições presentes na Tabela 3, pode-se observar como a presença de mais de uma espécie nas amostras de calibração influencia nos valores de SEN e SEL, onde nota-se um decréscimo desses valores à medida que o número de espécies presentes (e que se sobrepõem) na calibração aumenta. Observa-se, ainda, que há uma boa concordância entre os valores obtidos para os três modelos, sendo as maiores diferenças nos valores obtidos pelo PARAFAC em relação aos outros dois modelos para o AN nas calibrações AN|FEN e AN|BAP|FEN.

É importante notar que, na prática, não é possível determinar valores de SEL para o UPLS nesta condição onde não há interferentes não calibrados, como sugeriu Olivieri, pois não se conhece os valores de SEN quando apenas a espécie de interesse está presente na calibração. Neste caso, os valores de SEL são apresentados apenas para ilustrar que, caso os valores de SEN fossem conhecidos (como é o caso nesse estudo), são obtidos valores equivalentes aos dos outros modelos simplesmente pela razão entre as sensibilidades da situação onde se tem

mais de uma espécie no meio e quando a espécie de interesse está isolada.

**Tabela 3.** Valores de sensibilidade (SEN) e seletividade (SEL) obtidos para os modelos de calibração de segunda ordem utilizados quando a vantagem de segunda ordem não é utilizada.

		PARAFAC		BLLS		UPLS	
		SEN <sup>a</sup>	SEL	SEN <sup>a</sup>	SEL	SEN <sup>a</sup>	SEL
AN		39,6	1,0	39,6	1,0	39,6	---
BAP		41,0	1,0	41,0	1,0	41,1	---
FEN		6,65	1,0	6,66	1,0	6,66	---
AN   BAP	AN	34,6	0,90	34,4	0,90	34,5	0,87
	BAP	36,5	0,90	36,3	0,90	36,6	0,89
AN   FEN	AN	27,7	0,78	31,0	0,82	31,2	0,79
	FEN	5,45	0,78	5,45	0,82	5,49	0,82
BAP   FEN	BAP	39,0	0,98	41,1	0,98	40,0	0,97
	FEN	6,59	0,98	6,83	0,98	6,65	1,00
AN   BAP   FEN	AN	26,9	0,74	29,3	0,71	28,9	0,73
	BAP	34,6	0,90	36,6	0,86	35,8	0,87
	FEN	5,23	0,80	5,32	0,81	5,36	0,80

(a) Expressa contagens/( $\mu\text{g/L}$ ).

Os resultados para a determinação dos AN, BAP e FEN com os três modelos nas três situações onde a vantagem de segunda ordem é empregada e quando os três analitos estão calibrados são apresentados nas Tabelas 4, 5 e 6, respectivamente. Estes modelos foram desenvolvidos desconsiderando as replicatas onde o RBL apresentou divergência.

Para o AN (Tabela 4), comparando os indicativos médios dos erros (RMSEP e REP), observa-se que a eficiência dos modelos é semelhante, com a exceção da regressão AN|BAP, onde o UPLS apresenta erro maior que os demais. Comparando as quatro situações (AN|BAP|FEN, AN|BAP, AN|FEN e AN), teoricamente esperava-se que os erros aumentassem à medida que o número de espécies não

calibradas aumenta, pois se tem um sistema mais complexo e que requer a utilização do RBL. Contudo, deve-se tomar cuidado nessa comparação, uma vez que em cada situação tem-se uma regressão diferente e com ajustes possivelmente diferentes (valores de RMSEC diferentes), que por conseqüência influenciam os valores de RMSEP. Um fator que contribuiu para a obtenção de diferentes ajustes na calibração foi a utilização de apenas uma replicata para cada padrão. Apesar dessa influência, pode-se perceber pelos valores da Tabela 4 que os erros obtidos na situação onde apenas AN é calibrado são maiores, como seria esperado. Essa tendência é também observada nos valores de precisão.

**Tabela 4.** Resultados das figuras de mérito estimadas para antraceno (AN) nas quatro condições onde é determinado com BLLS, UPLS e PARAFAC.

Situação	RMSEC <sup>a</sup>	RMSEP <sup>a</sup>	REP <sup>b</sup>	Precisão <sup>a,c</sup>	SEN <sup>d</sup>	SEL	LD <sup>a</sup>
<b>BLLS</b>							
[AN BAP FEN]	0,64	0,43	6,9	0,22	29,3	0,71	0,20
[AN BAP]	0,10	0,71	4,8	0,42	7,3	0,19	0,79
[AN FEN]	0,24	0,31	1,9	0,30	15,9	0,42	0,36
[AN]	0,11	1,3	9,4	0,93	3,1	0,08	1,9
<b>UPLS</b>							
[AN BAP FEN]	0,10	0,42	6,4	0,22	28,9	---	0,20
[AN BAP]	0,10	1,18	12	0,35	8,1	0,23	0,72
[AN FEN]	0,20	0,31	1,9	0,30	16,1	0,52	0,36
[AN]	0,11	1,36	9,4	0,81	3,1	0,08	1,9
<b>PARAFAC</b>							
[AN BAP FEN]	0,64	0,33	2,8	0,25	27,0	0,74	0,25
[AN BAP]	0,10	0,71	4,9	0,41	7,3	0,19	0,79
[AN FEN]	0,41	0,33	2,3	0,29	15,4	0,42	0,38
[AN]	0,19	1,28	9,6	0,81	3,1	0,08	1,9

Grandezas expressas em: (a)  $\mu\text{g/L}$ , (b) percentagem, (c) número de graus de liberdade igual a 80 e (d) contagens/ $(\mu\text{g/L})$ .

Quanto aos valores de SEN e SEL, observa-se que a diminuição nesses

valores devido à influência de uma mesma espécie é maior quando a referida espécie não esta presente na calibração (isto é, está atuando como um interferente), do que quando a mesma espécie é um analito calibrado. Este efeito pode ser visualizado comparando os valores de SEN e SEL da calibração AN|BAP|FEN com AN|BAP e AN|FEN, onde FEN e BAP estão calibrados na primeira e são interferentes nas outras duas calibrações, respectivamente.

**Tabela 5.** Resultados das figuras de mérito estimadas para o benzo[a]pireno (BAP) nas quatro condições onde é determinado com BLLS, UPLS e PARAFAC.

Situação	RMSEC <sup>a</sup>	RMSEP <sup>a</sup>	REP <sup>b</sup>	Precisão <sup>a,c</sup>	SEN <sup>d</sup>	SEL	LD <sup>a</sup>
<b>BLLS</b>							
[AN BAP FEN]	0,73	0,50	3,4	0,40	36,6	0,86	0,16
[AN BAP]	0,71	0,61	4,0	0,67	17,6	0,44	0,33
[BAP FEN]	0,24	0,42	2,6	0,42	20,9	0,50	0,28
[BAP]	0,30	0,62	4,1	0,51	7,32	0,18	0,82
<b>UPLS</b>							
[AN BAP FEN]	0,25	0,53	3,3	0,40	35,8	---	0,16
[AN BAP]	0,30	0,65	4,5	0,42	18,7	0,51	0,31
[BAP FEN]	0,15	0,39	2,4	0,40	22,1	0,55	0,26
[BAP]	0,30	0,61	4,4	0,50	7,7	0,19	0,78
<b>PARAFAC</b>							
[AN BAP FEN]	0,38	0,55	2,9	0,44	34,6	0,90	0,18
[AN BAP]	0,31	0,61	4,0	0,46	17,5	0,43	0,34
[BAP FEN]	0,34	0,46	2,8	0,43	20,2	0,51	0,29
[BAP]	0,37	0,55	3,8	0,65	7,8	0,19	0,75

Grandezas expressas em: (a)  $\mu\text{g/L}$ , (b) percentagem, (c) número de graus de liberdade igual a 80 e (d) contagens/ $(\mu\text{g/L})$ .

Na Tabela 4 são apresentados os resultados de SEL para o UPLS quando a vantagem de segunda ordem é utilizada com o RBL, obtidos pela equação 70 sugerida neste trabalho. Comparando esses valores com os obtidos para o BLLS e PARAFAC, observa-se que há uma boa concordância, sendo as diferenças

observadas pequenas e possivelmente devido às características do modelo.

Os valores de LD obtidos para o AN também são apresentados na Tabela 4. Para este parâmetro, na tentativa de se obter LD mais realistas, foi utilizado como estimativa do ruído instrumental ( $s_{\text{ruído}}$ ) que é utilizado no cálculo do LD (equação 67) o valor de 1,75, o qual foi obtido pela raiz quadrada da média das 4 variâncias das 21 replicatas das amostras de validação. Logo, o valor de 1,75 leva em consideração além do ruído instrumental a variação do preparo das amostras. Os valores obtidos sugerem alta sensibilidade para AN, com limites de detecção maiores que 1  $\mu\text{g/L}$ , apenas na condição onde ele é calibrado e têm-se dois interferentes, sendo a situação de menor sensibilidade.

**Tabela 6.** Resultados das figuras de mérito estimadas para o fenantreno (FEN) nas quatro condições onde é determinado com BLLS, UPLS e PARAFAC.

Situação	RMSEC <sup>a</sup>	RMSEP <sup>a</sup>	REP <sup>b</sup>	Precisão <sup>a,c</sup>	SEN <sup>d</sup>	SEL	LD <sup>a</sup>
<b>BLLS</b>							
[AN BAP FEN]	2,4	2,3	4,4	0,72	5,32	0,81	1,1
[AN FEN]	1,1	1,5	4,1	1,0	2,89	0,43	2,0
[BAP FEN]	1,0	1,2	2,1	0,86	1,65	0,24	3,5
[FEN]	1,4	3,0	4,7	2,6	0,56	0,08	10
<b>UPLS</b>							
[AN BAP FEN]	0,36	2,8	4,8	0,69	5,36	---	1,1
[AN FEN]	1,0	1,3	2,6	0,99	3,0	0,54	2,0
[BAP FEN]	0,80	2,0	3,9	0,83	1,7	0,26	3,4
[FEN]	1,38	2,6	5,2	2,5	0,59	0,09	10
<b>PARAFAC</b>							
[AN BAP FEN]	2,0	1,97	2,6	0,81	5,24	0,80	1,2
[AN FEN]	1,0	1,4	3,0	0,90	2,94	0,44	2,1
[BAP FEN]	0,90	0,80	1,3	0,82	1,59	0,24	3,7
[FEN]	1,7	1,7	3,7	1,6	0,53	0,08	11

Grandezas expressas em: (a)  $\mu\text{g/L}$ , (b) percentagem, (c) número de graus de liberdade igual a 80 e (d) contagens/ $(\mu\text{g/L})$ .

A interpretação dos resultados para BAP e FEN apresentados nas Tabelas 5 e 6 seguem aproximadamente o mesmo raciocínio do discutido para AN. Com diferença no fato do aumento dos erros médios (RMSE e REP) e da precisão para o BAP, onde não se observa um aumento significativo nesses valores na calibração contendo apenas BAP como é observado na Tabela 4 para AN, enquanto que para FEN observa-se um aumento apenas nos valores de precisão.

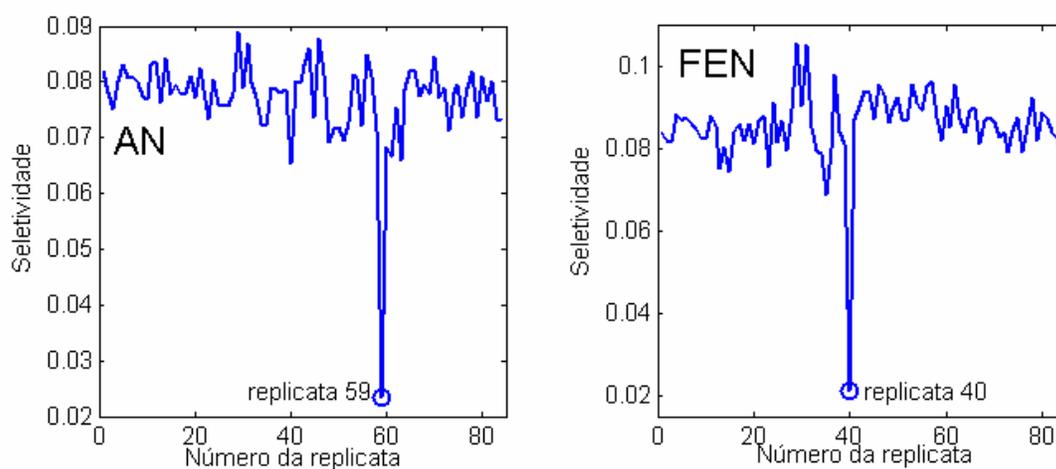
Comparando os valores de LD observa-se que BAP é o mais sensível dos três analitos, com LD inferiores a 1 µg/L em todas as situações, enquanto que para FEN têm-se valores que variam de 1 a 11 µg/L dependendo da situação.

Os valores de SEL obtidos para BAP e FEN com o UPLS também apresentam uma boa concordância com os valores obtidos para BLLS e PARAFAC, como foi observado para AN.

Outro resultado interessante com relação aos valores de SEN e SEL foi observado para as replicatas onde o RBL aparentemente não convergiu e apresentou erros altos nas estimativas da concentração em relação às concentrações de referência das amostras de validação. A Figura 22 mostra que os valores de SEL para essas replicatas sofrem uma queda fora do perfil aleatório das demais, o que equivale a uma queda incomum na sensibilidade dessa replicata, de modo que em uma aplicação em rotina, a ocorrência de possíveis divergências no algoritmo do RBL pode ser monitorada a partir desses valores, para os casos onde as concentrações estimadas em uma suposta divergência não sejam absurdas e deixem alguma dúvida quanto à decisão de descartar ou não a referida amostra.

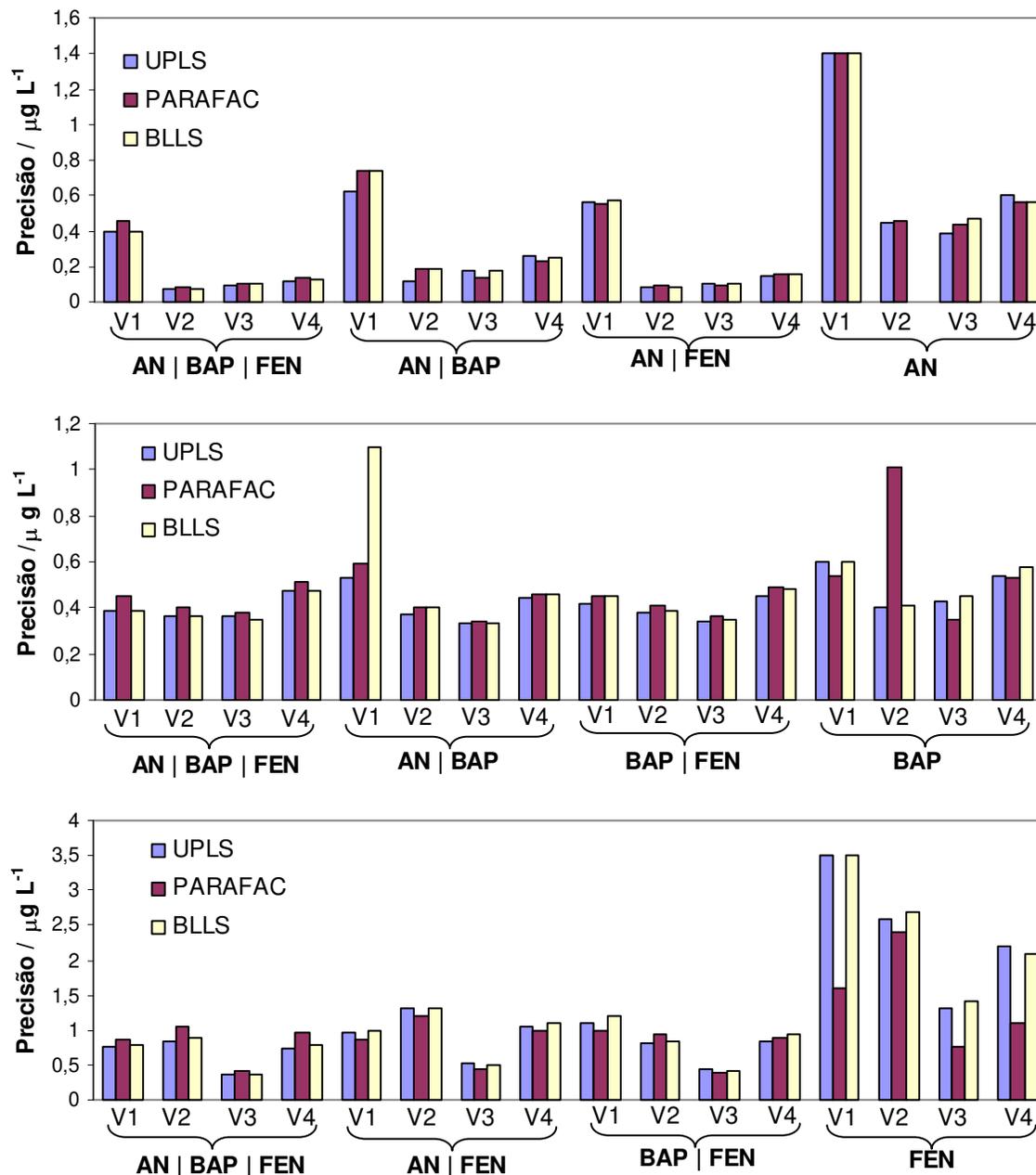
Outra observação com relação à Figura 22 é que a seletividade (da forma como é determinada para modelos de calibração de segunda ordem) depende se há ou não interferentes presentes e do número de interferentes presentes, mas não depende da intensidade do sinal do interferente na amostra, pois a cada 23 replicatas tem-se uma mudança das concentrações do analito e dos interferentes,

mas a seletividade estimada segue um perfil aleatório (com exceção das amostras anômalas) que não se relaciona à concentração dos interferentes. Essa observação mostra que a estimativa da seletividade pela atual proposta para modelos de calibração de segunda ordem<sup>78,94</sup> pode ser incompleta nesse aspecto, pois de acordo com a definição da IUPAC, esse parâmetro deveria ser capaz de indicar até quando um analito pode ser determinado sem que os interferentes prejudiquem as estimativas da sua concentração, o qual certamente depende da presença ou não dos interferentes e também da concentração desses interferentes.

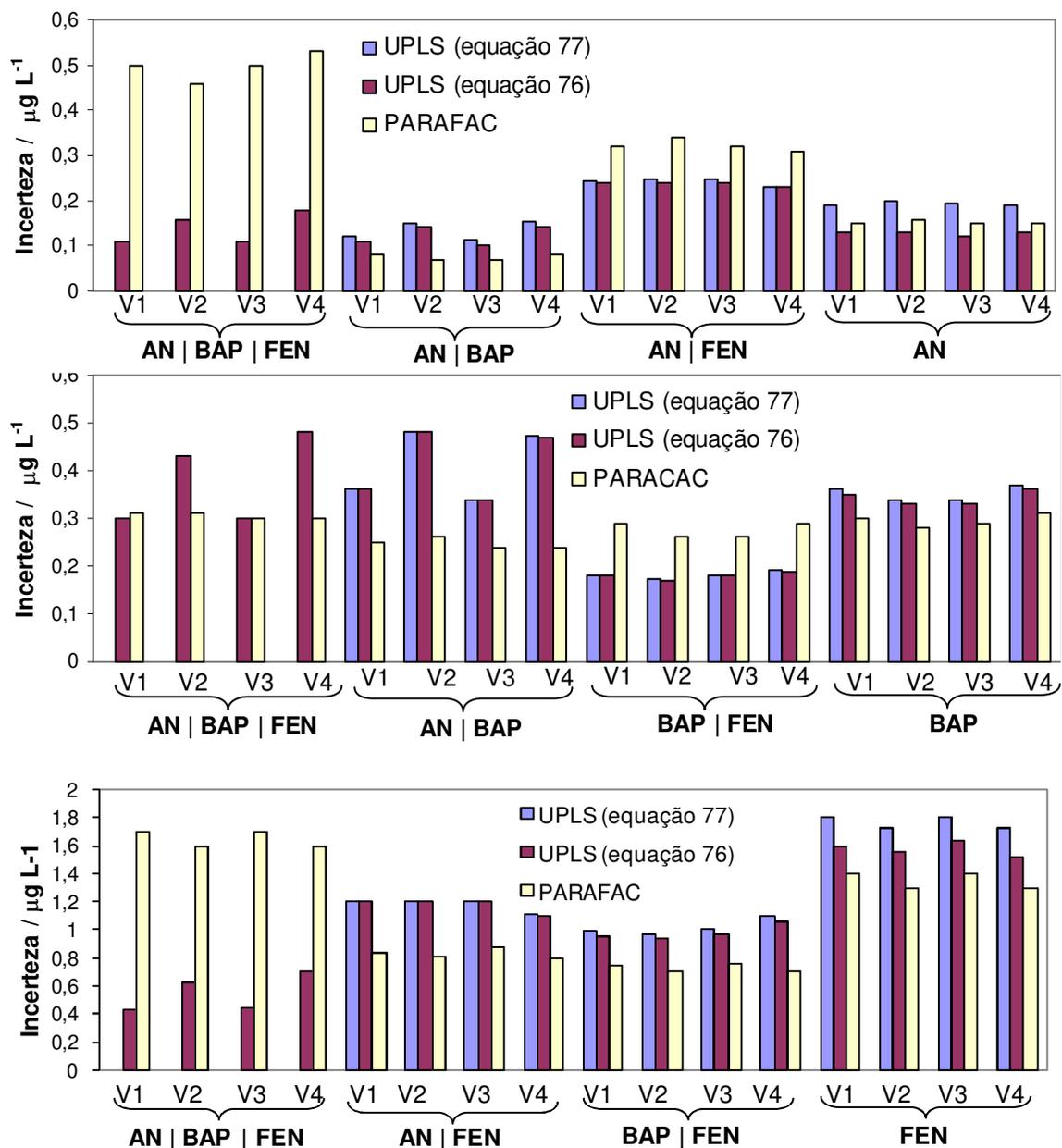


**Figura 22-** Variação dos valores de SEL quando apenas AN e FEN estão calibrados.

Os resultados de precisão apresentados nas Tabelas 4, 5 e 6 consistem das precisões médias observadas para as quatro amostras de validação. A Figura 23 mostra as precisões individuais para cada amostra, que confirma que os três modelos de calibração apresentam praticamente os mesmos resultados, com exceção de quando apenas FEN é calibrado, onde o PARAFAC apresenta uma melhor precisão em três das quatro amostras de validação.



**Figura 23-** Precisões obtidas para os analitos AN, BAP e FEN nas quatro situações em que foram determinados em cada amostra de validação (V1, V2, V3 e V4) nos três modelos de calibração utilizados. Número de graus de liberdade =20, exceto nas situações de divergência do RBL discutidas no texto.



**Figura 24-** Incertezas estimadas para os analitos AN, BAP e FEN nas quatro situações em que foram determinados, estimadas para o PARAFAC pela equação 71 para o UPLS pelas equações 76 (UPLS) e 77 (UPLS-RBL), onde a equação 77 leva em conta o aumento da incerteza devido os interferentes.

Os resultados para as estimativas de incerteza para os três analitos nas quatro situações em que foram determinados são apresentadas na Figura 24, as quais foram obtidas para o PARAFAC utilizando a equação 71 (que é análoga à utilizada em calibração de ordem zero) e para UPLS com as equações 76 (mesma equação

utilizada em calibração de primeira ordem) e 77 (equação que leva em conta a perda de sensibilidade devido à presença de interferentes). Neste último caso foi utilizado como estimativa da variação instrumental o valor de 0,45 obtida por meio de 15 medidas de fluorescência do solvente. Parte das diferenças que são observadas na Figura 24 entre as estimativas obtidas para o PARAFAC e UPLS (equação 76) se devem ao ajuste em cada modelo, que pode ser avaliado pelos valores de RMSEC. Por exemplo, para a determinação de AN na regressão onde os três analitos estão calibrados, os RMSEC são 0,64 e 0,10 para PARAFAC e UPLS, respectivamente, como conseqüência o PARAFAC tem estimativas de incerteza significativamente maiores que as obtidas para o UPLS.

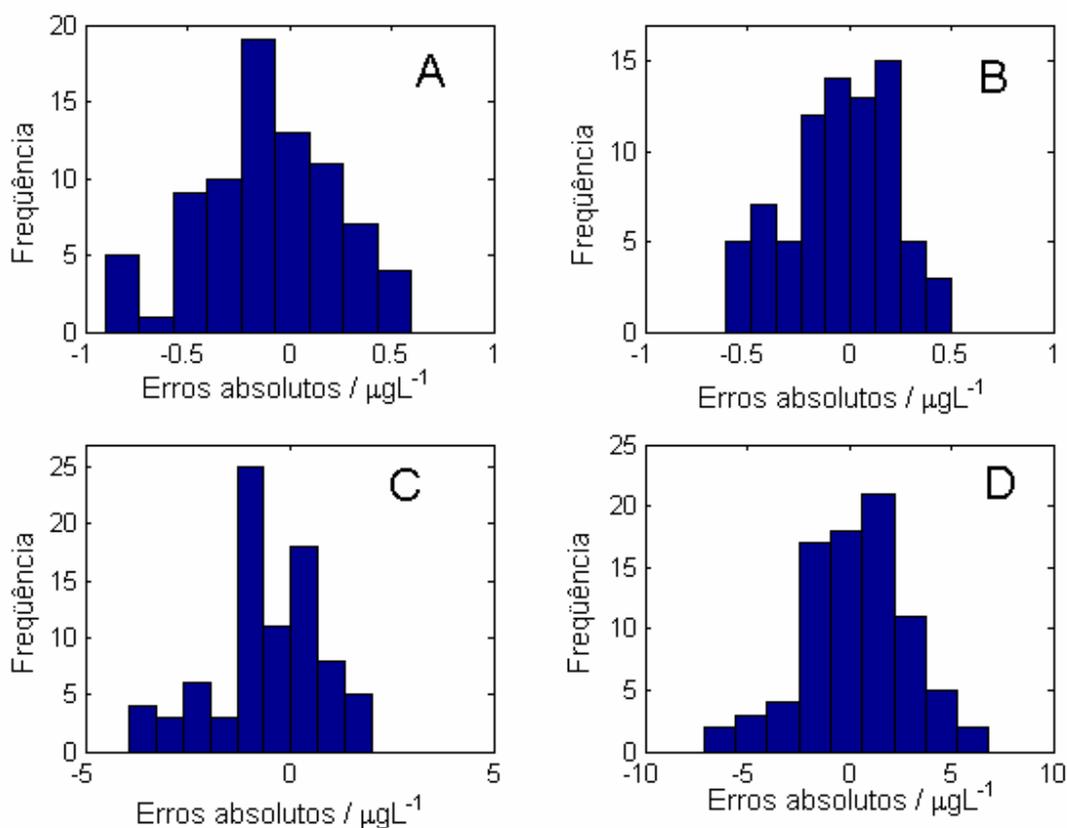
Comparando as incertezas obtidas apenas para o PARAFAC, observa-se que não há um aumento nos valores obtidos para modelos com praticamente o mesmo RMSEC e diferentes valores de sensibilidade. Esse efeito pode ser visualizado mais claramente comparando as incertezas nas calibrações AN|BAP|FEN e BAP, onde os RMSEC e SEN nesses casos são respectivamente iguais a (0,38 e 0,37) e (34,6 e 7,8), logo tem-se uma sensibilidade cerca de 4,5 vezes menor, mas praticamente as mesmas incertezas.

Para as estimativas de incerteza no UPLS, obtidas com as equações 76 e 77, observa-se um aumento na incerteza nas calibrações onde a sensibilidade é menor apenas nos resultados da equação 77, que é notado de forma mais significativa nas calibrações onde apenas um analito é calibrado. Os pequenos aumentos observados nas calibrações onde têm-se dois analitos calibrados, são referentes ao pequeno decréscimo nos valores de sensibilidade observados nesses casos, que é no máximo 4 vezes (obtido para AN). As diferenças mais marcantes entre as equações 76 e 77 são observados apenas nas calibrações de AN e FEN onde se observa uma diminuição de aproximadamente 9 vezes na sensibilidade em ambos os casos.

Uma maneira de confirmar se as estimativas das incertezas estão coerentes é

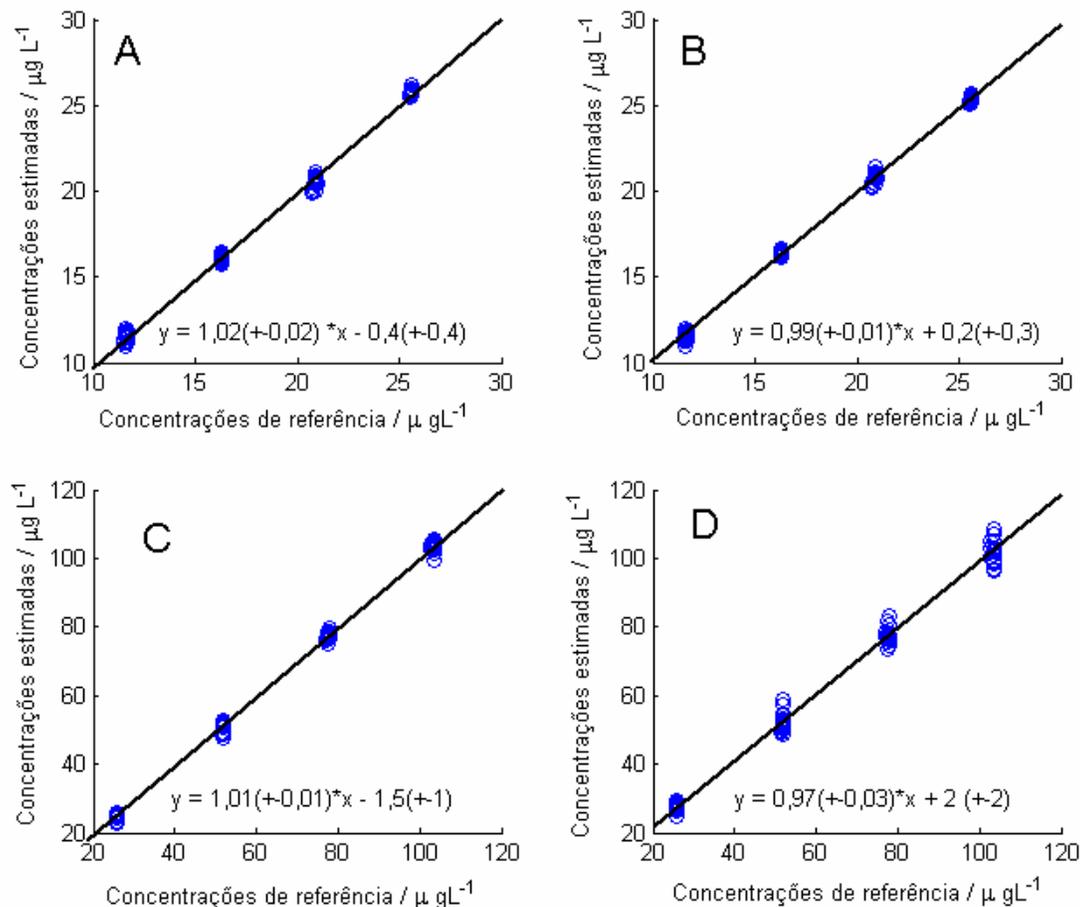
conferir se os intervalos de confiança conseguem abranger o valor de referência das amostras de validação em uma probabilidade de acerto próxima da esperada. Isto é, para um intervalo de confiança de 95 % espera-se que 95 em cada 100 amostras tenham o valor de referência dentro do intervalo de confiança. Contudo, para que esses intervalos sejam coerentes, é necessário que a distribuição dos erros das estimativas da concentração siga aproximadamente uma distribuição Normal e que a ocorrência de erros sistemáticos não seja significativa. Entre as situações estudadas, a calibração BAP|FEN para a determinação de BAP e a calibração que contém apenas FEN preenchem aproximadamente estes requisitos.

Os histogramas dos erros absolutos para o PARAFAC e UPLS nessas situações são apresentados na Figura 25. De acordo com o teste de significância de Jarque-Bera<sup>106</sup> para verificar a hipótese de Normalidade com 95 % de confiança ( $P=0,05$ ), a hipótese nula de que os quatro histogramas da Figura 25 seguem uma distribuição Normal não pode ser rejeitada, onde os valores de  $P$  obtidos para que essa hipótese seja rejeitada foram 0,668; 0,314; 0,108 e 0,514 para os histogramas da Figura 25 A, B, C e D, respectivamente. Quanto a erros sistemáticos, observando os resultados obtidos para a inclinação e o intercepto da reta ajustada entre os valores de referência e estimados e seus respectivos intervalos de confiança com 99 %, apresentados na Figura 26, verifica-se que para o UPLS a presença de erros sistemáticos não é significativa. Entretanto, para o PARAFAC os resultados mostram que, infelizmente, na calibração de apenas FEN um erro sistemático constante negativo não pode ser desconsiderado. Contudo, para propósitos de comparação entre os modelos UPLS e PARAFAC os intervalos de confiança com o PARAFAC nessa situação foram estimados e também são apresentados na Tabela 7.



**Figura 25-** Histograma dos erros absolutos obtidos nas calibrações em que BAP|FEN (A e B) e FEN (C e D) estão calibrados. (A e C) PARAFAC, (B e D) UPLS.

Pode-se observar na Tabela 7 que na situação em que BAP é determinado na calibração BAP|FEN o PARAFAC consegue fornecer intervalos de confiança que apresentam probabilidade praticamente igual à esperada, enquanto que o UPLS (independente da equação utilizada) apresenta erros um pouco maiores nas probabilidades recuperadas, mas que também podem ser considerados com boa concordância tendo em vista o número disponível de medidas (4 amostras com 21 replicatas cada). Para esse caso em particular, não foi observada diferença significativa nas percentagens recuperadas no UPLS com as equações 76 e 77, devido à queda de sensibilidade neste caso ser de apenas 1,7 vezes.



**Figura 26-** Concentrações de referência versus estimadas obtidos nas calibrações em que BAP|FEN (A e B) e FEN (C e D) estão calibrados. (A e C) PARAFAC, (B e D) UPLS, intervalos de confiança de 99 % para inclinação e intercepto entre parênteses.

**Tabela 7.** Percentagens recuperadas para as probabilidades de recobrimento dos intervalos de confiança obtidos para PARAFAC e as duas estimativas com UPLS.

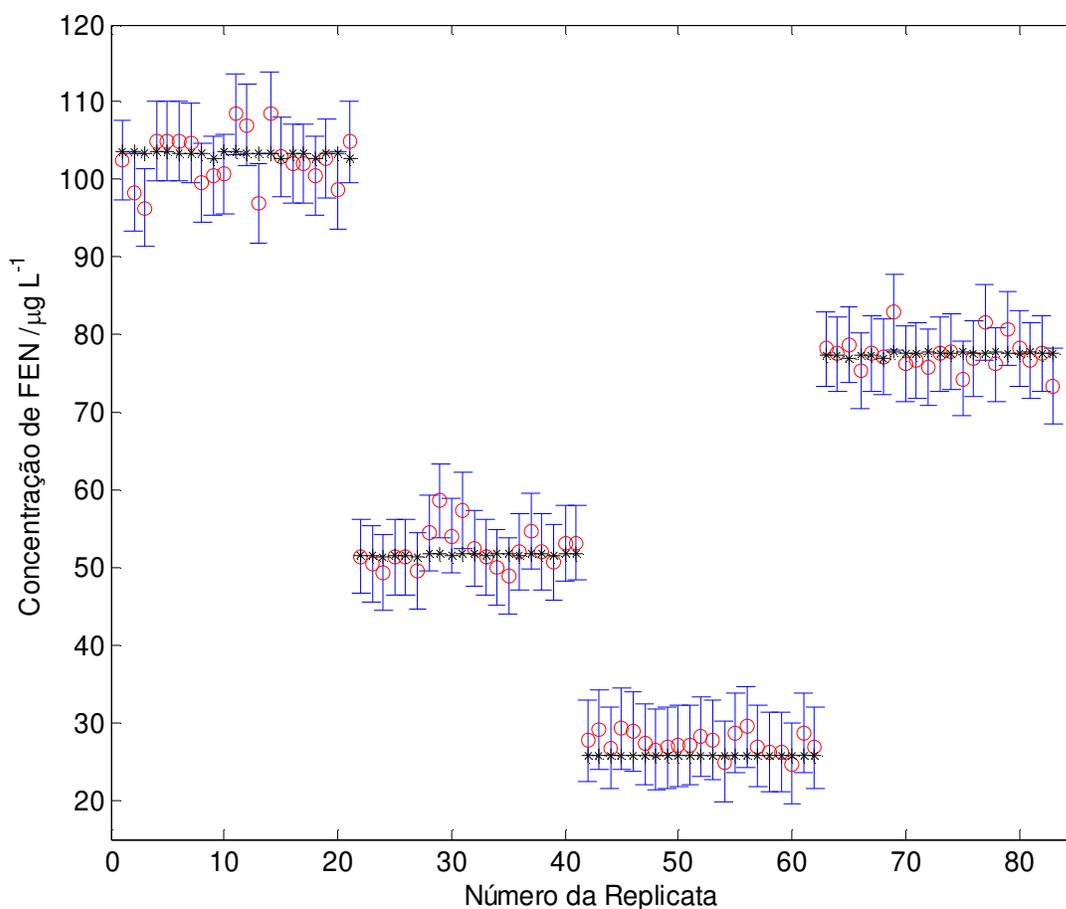
Probabilidade Esperada (%)	BAP (calibração BAP FEN)			FEN (calibração FEN)		
	PARAFAC	UPLS (eq. 76)	UPLS (eq 77)	PARAFAC	UPLS (eq. 76)	UPLS (eq 77)
99	100	100	100	100	100	100
95	96	99	99	100	93	95
90	94	90	91	95	86	87
80	81	75	77	84	70	75
50	49	38	42	60	36	46

Intervalos de confiança utilizando t de Student com 3 graus de liberdade.

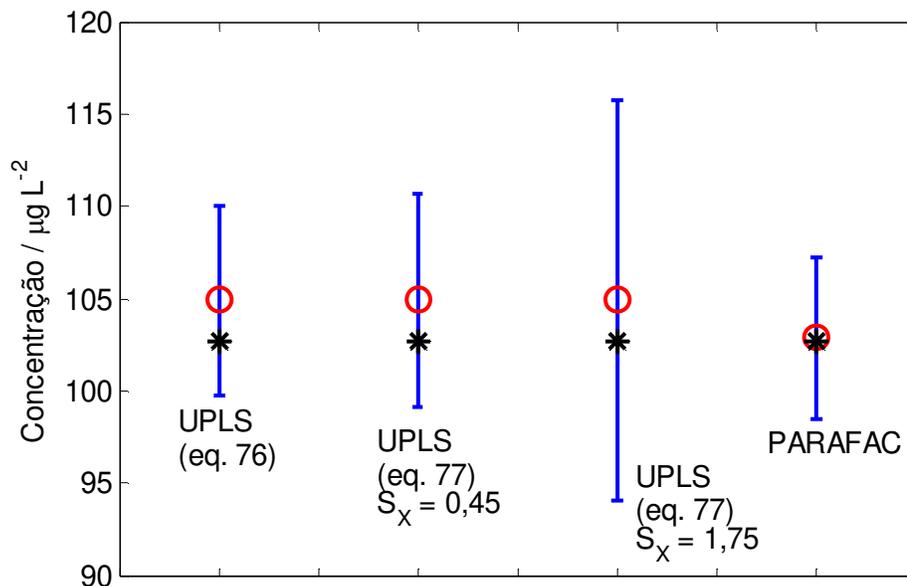
Para a situação onde FEN é determinado com AN e BAP como interferentes os resultados da Tabela 7 mostram que o PARAFAC desta vez fornece intervalos de confiança que são cerca de 5 % mais largos que o necessário, fato que pode ser devido à presença de erros sistemáticos significativos que foi constatada nesse caso. Para o UPLS com a equação 76 (que não leva em consideração a perda de sensibilidade) observam-se resultados um pouco mais coerentes que o PARAFAC nos níveis de 99 e 95 %, mas nos níveis de 90, 80 e 50 % observa-se erros variando de 5 a 15 %. Utilizando a equação 77 (que inclui o efeito de perda de sensibilidade) observam-se intervalos de confiança com uma boa concordância em todas as 5 probabilidades testadas, sendo obtidos erros de no máximo 5 %. Logo, apesar da pequena diferença observada nas incertezas estimadas entre as equações 76 e 77, a equação 77 do ponto de vista teórico é mais coerente, pois leva em conta o efeito dos interferentes nas amostras de previsão e pelos resultados obtidos parece fornecer incertezas que conduzem a intervalos de confiança com probabilidades de recobrimento mais próximas dos valores esperados. Contudo, uma desvantagem da utilização da equação 77 que é proposta reside na necessidade de uma boa estimativa da variância dos dados instrumentais. Nos resultados apresentados na Tabela 7 foi utilizado como estimativa dessa variância o quadrado do ruído instrumental (estimado em 0,45 contagens) obtido por 15 replicatas da fluorescência do solvente (solução acetonitrila:água 25:75 %). Para ilustrar a importância dessa variância, caso o valor do desvio padrão entre as replicatas (estimado em 1,75 contagens) for utilizado no lugar de 0,45, observa-se utilizando a equação 77 probabilidades de recobrimento de 100, 100, 99, 95 e 70 % para os valores esperados de 99, 95, 90, 80 e 50, respectivamente, que são mais largos que o esperado pois o valor de 1,75 contagens além da variação instrumental inclui o erro de preparo de amostra.

Na Figura 27 podem ser observados os intervalos de confiança para o UPLS

utilizando a equação 77, onde se percebe a sua aplicação de forma clara. Na Figura 28, tomando a replicata 21 como exemplo, observa-se os intervalos de confiança para PARAFAC e as duas estimativas com o UPLS, sendo os resultados da equação 77 mostrados utilizando ambos os valores de 0,45 e 1,75 como estimativas da variação instrumental. Na Figura 28 é nítida a importância de uma boa estimativa para a variação instrumental. Caso isso não seja possível, intervalos significativamente mais largos ou mais estreitos podem ser obtidos.

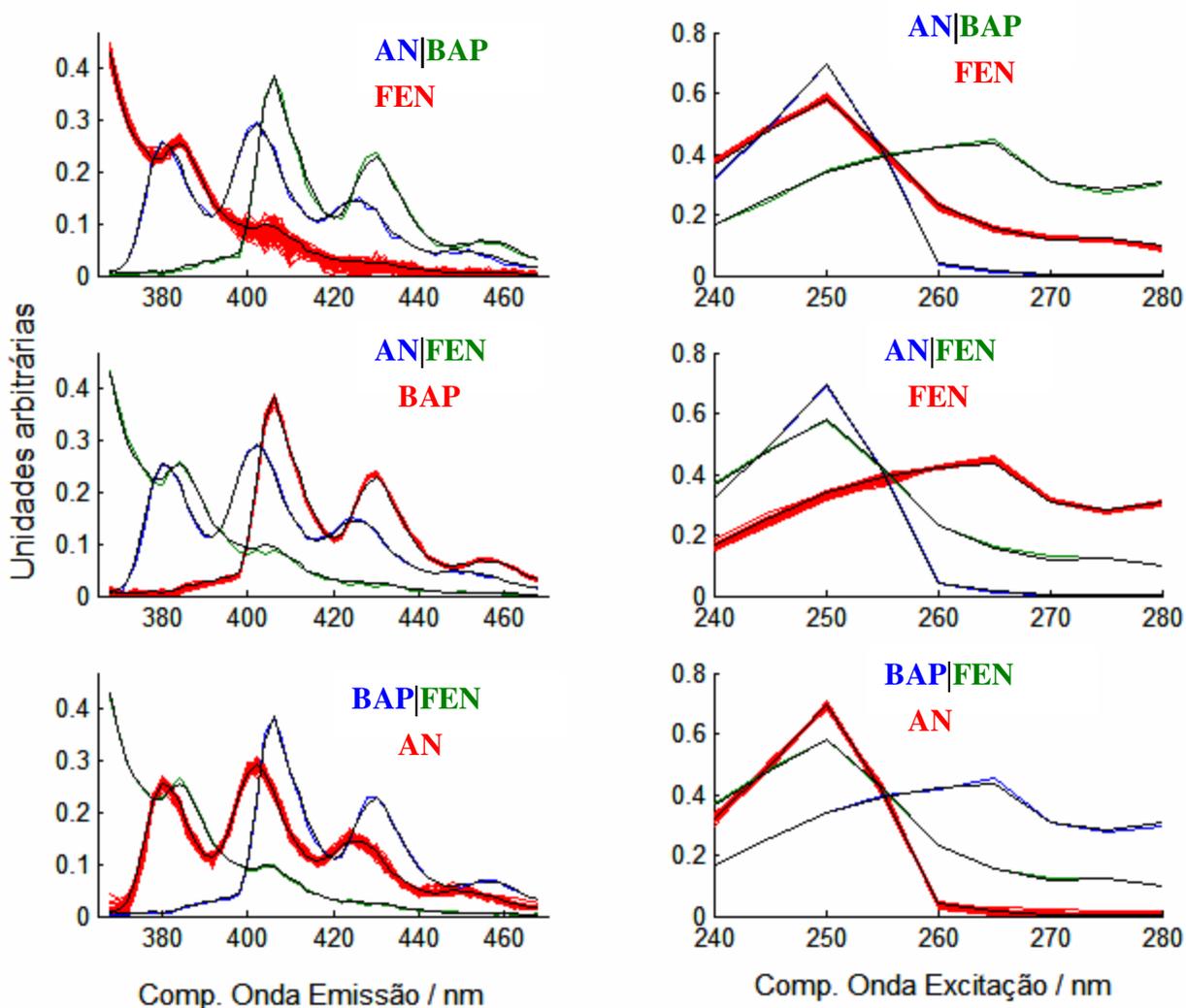


**Figura 27-** Intervalos de confiança de 95 % para as estimativas das concentrações de FEN pelo UPLS utilizando a equação 77 e a estimativa do ruído como 0,45 contagens. (\*) concentração de referência, (o) concentração estimada, (—) barras de erro.



**Figura 28-** Intervalos de confiança para a replicata 21 estimados com PARAFAC e com o UPLS utilizando a equação 76 e 77. (\*) concentração de referência, (o) concentração estimada, (—) barras de erro com 95 % de confiança.

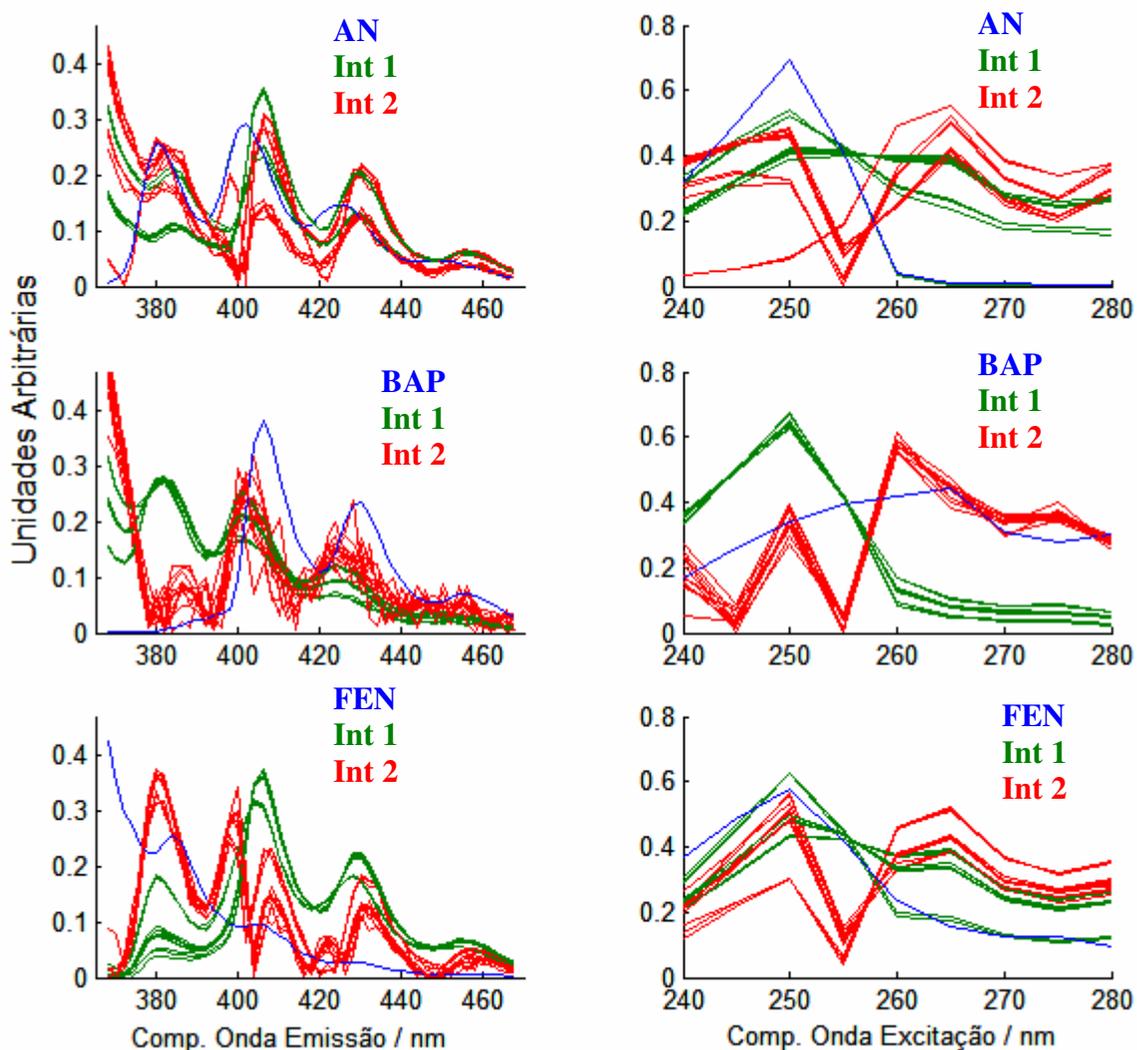
Por fim, quanto à obtenção de informações qualitativas através dos perfis que podem ser estimados pelos modelos PARAFAC e BLLS, têm-se um conjunto de perfis para cada replicata, em cada modelo e situação distinta uma vez que os modelos foram aplicados de modo que se analisava uma replicata de cada vez. A Figura 29 mostra os perfis obtidos para o BLLS quando dois analitos estão calibrados, onde são mostrados os perfis das 84 replicatas do conjunto de validação junto com os perfis obtidos experimentalmente em padrões isolados. Observa-se uma ótima concordância entre os perfis experimentais e estimados, sendo a maior variação observada no perfil da espécie que está presente como interferente (mostrada em vermelho), a qual na verdade é a única que é estimada uma vez em cada replicata através do RBL. Os perfis obtidos com o PARAFAC nessas situações apresentaram praticamente a mesma concordância e por isso não são apresentados.



**Figura 29-** Valor absoluto dos perfis espectrais de emissão (esquerda) e excitação (direita) normalizados recuperados com o BLS nas calibrações onde duas espécies de interesse são calibradas (— e —) e uma se apresenta como um interferente (—). Perfis experimentais são mostrados em preto (—).

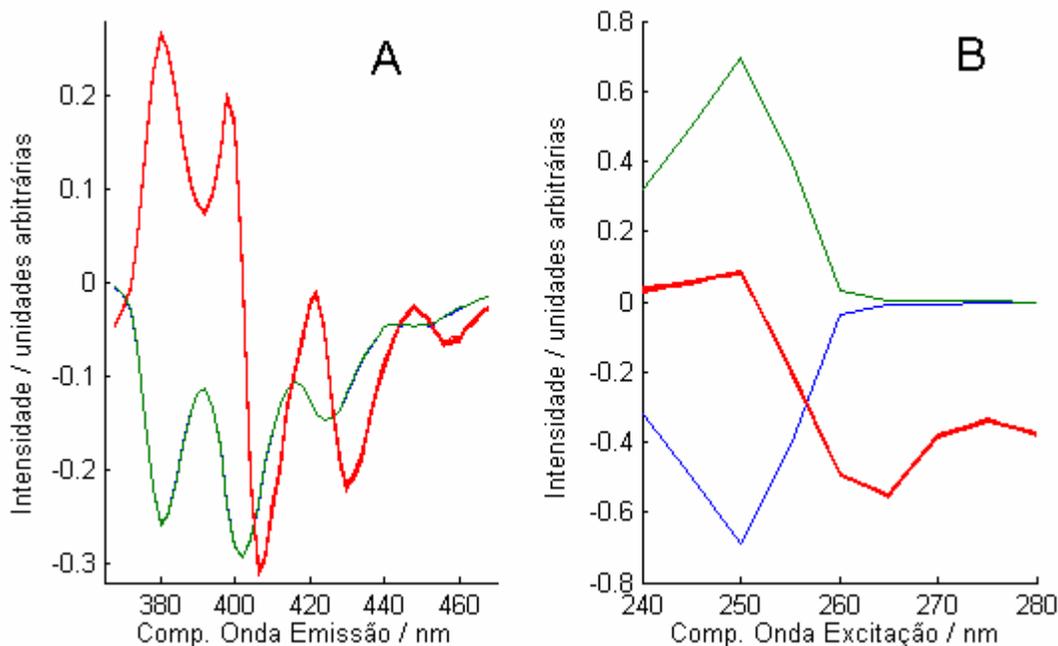
A Figura 30 mostra os perfis estimados com o BLS para as calibrações onde apenas uma espécie é calibrada e duas atuam como interferentes, onde ao contrário da Figura 29, são apresentadas apenas 3 replicatas de cada amostra de validação. Neste caso observa-se uma grande variação nos perfis estimados para os interferentes, principalmente para o segundo (mostrado em vermelho). A Figura 30 ilustra uma limitação do RBL em termos qualitativos, que é não ser capaz de

estimar o perfil do interferente quando mais de um interferente está presente na amostra. Nessas situações o perfil do primeiro interferente é uma combinação linear dos perfis dos vários interferentes presentes e o perfil seguinte uma combinação linear da variação sistemática que resta após a retirada do perfil do primeiro interferente. Ao contrário do BLLS, o PARAFAC consegue estimar os perfis das duas espécies presentes como interferentes apresentando uma boa concordância com os perfis experimentais.



**Figura 30-** Valor absoluto dos perfis espectrais normalizados de emissão (esquerda) e excitação (direita), recuperados com o BLLS nas curvas com apenas uma espécie calibrada (—) e duas se como interferentes (— e —). (Int 1) primeiro interferente e (Int 2) segundo interferente.

Outro fato que foi observado nas situações apresentadas na Figura 30, mas observando apenas os perfis estimados para as amostras que apresentaram divergência no RBL quando dois fatores foram utilizados (replicatas de 43 a 63 na calibração de AN e replicatas 40 para FEN), é que os perfis estimados para o primeiro interferente nessas amostras apresentam uma correlação muito alta (cerca de 0,9999) com o perfil do analito calibrado, sendo essa correlação positiva para um dos perfis de excitação ou emissão, e negativa para o outro perfil. A Figura 31 mostra esse comportamento para a terceira amostra de validação da situação onde apenas AN está calibrado, que teve divergência em todas as 21 replicatas. O primeiro interferente mostrado em verde inclui na verdade os 21 perfis das replicatas dessa amostra evidenciando uma variação extremamente pequena. Esse fato reforça a hipótese de problemas de convergência quando existe mais de um interferente em alguns casos.



**Figura 31-** Perfis normalizados estimados com BLS para as replicatas da terceira amostra de validação na calibração onde apenas o AN é calibrado. (—) AN, (—) primeiro interferente e (—) segundo interferente.

### **2.2.3 Conclusões do capítulo**

Através do estudo descrito neste capítulo foi possível observar algumas propriedades dos modelos PARAFAC, BLLS e UPLS sendo os dois últimos utilizando o RBL para estimar a concentração de uma espécie de interesse na presença de interferentes. Muitas dessas observações foram possíveis devido ao controle e conhecimento do sistema estudado que utilizou apenas misturas conhecidas de três hidrocarbonetos policíclicos aromáticos, o que reforça a importância do estudo de sistemas mais simples.

Foi possível verificar a coerência de uma proposta de limite ou valor crítico para guiar a escolha do número de fatores utilizados no RBL para modelar interferentes não calibrados. Os resultados dessa aplicação em particular sugerem que a utilização do valor crítico com 99 % de confiança baseado no resíduo médio de decomposição das amostras de calibração é uma maneira simples e eficaz para o modelo BLLS, o que apresenta ainda a vantagem de ser baseada nas informações das próprias amostras de calibração, não requerendo novas medidas. Por outro lado, esse valor crítico não apresentou a mesma eficiência no modelo UPLS, pois no número de fatores escolhido em algumas replicatas os resíduos estavam levemente acima e outras abaixo do limite, provavelmente devido ao fato do UPLS ajustar mais as amostras de calibração fornecendo assim resíduos menores.

Pôde ser observada a influência da presença de espécies sobrepostas como analitos calibrados ou interferentes nos valores de sensibilidade e seletividade, sendo a influência como interferente significativamente maior. Os resultados de seletividade nessa aplicação sugerem que a equação 70 proposta para o seu cálculo no modelo UPLS, quando este utiliza a vantagem de segunda ordem, é coerente com os valores obtidos nos modelos PARAFAC e BLLS, tornando possível a determinação desse parâmetro nessa situação para o UPLS. Contudo, a estimativa de seletividade para todos os modelos avaliados ainda deve ser reformulada de

forma a também considerar a concentração dos interferentes presentes na amostra e não apenas o fato do interferente estar ou não presente.

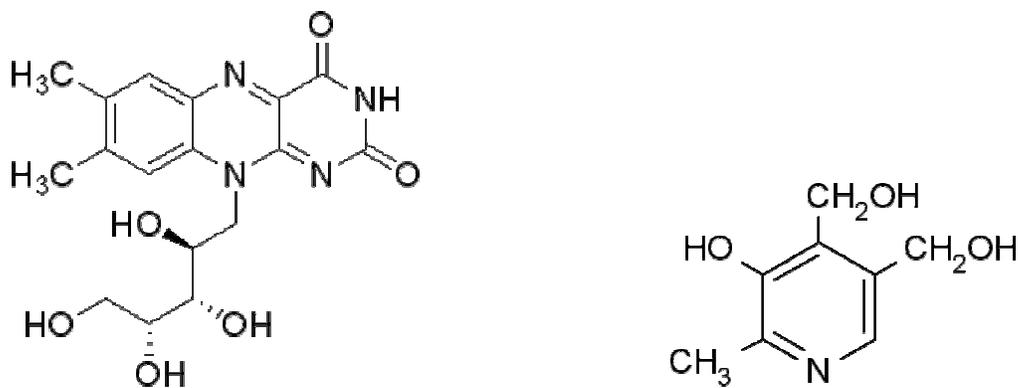
Puderam ser estimadas e comparadas incertezas nos modelos PARAFAC utilizando uma abordagem análoga à utilizada em calibração de ordem zero (equação 71) e no UPLS utilizando a equação empregada em calibração de primeira ordem e uma nova proposta que leva em conta a perda de sensibilidade devido à presença de interferentes (equações 76 e 77, respectivamente). Em duas situações foi possível a estimativa de intervalos de confiança, que apresentaram boa concordância com os valores esperados na calibração onde apenas um interferente estava presente tanto com PARAFAC como UPLS. Na situação onde dois interferentes estavam presentes as probabilidades recuperadas foram melhores com o UPLS utilizando a equação que foi proposta nesse trabalho.

Quanto à proposta de estimativa na incerteza no UPLS levando em conta a ação dos interferentes, foi observado que o aumento da incerteza para modelos com diminuição da sensibilidade parece ser coerente, mas depende de uma boa estimativa da variação instrumental dos dados. Os resultados dos erros obtidos não apresentaram um aumento significativo na incerteza, pois o ruído presente que foi estimado é baixo, tornando a contribuição deste termo pequena em relação à incerteza total. Contudo, os resultados apresentados apóiam a hipótese de que a incerteza nas estimativas da concentração obtidas pelo modelo UPLS, utilizando o RBL, aumentem quando esses interferentes provocarem uma perda na sensibilidade do modelo sugerindo a validade da equação proposta. No entanto, a confirmação da utilização da equação 77 deve ser avaliada em outros estudos tanto com dados simulados quanto experimentais.

Os resultados deste capítulo com os programas para a construção dos modelos BLLS e UPLS, que foram desenvolvidos neste trabalho, são os mesmos, independente da realização dos cálculos nos ambientes Matlab 6.5 ou Octave 3.0.0.

**Capítulo 3 - Determinação de riboflavina e  
piridoxina em leite em pó e suplementos  
alimentares**

O segundo estudo de caso apresentado neste trabalho utilizou como técnica instrumental novamente a EEF, mas ao contrário do capítulo anterior que tinha como foco o estudo de um sistema relativamente simples do ponto de vista analítico, neste caso trabalhou-se com amostras de extratos obtidos a partir de três marcas de leite em pó infantil e dois suplementos alimentares. Estes sistemas podem ser considerados complexos, uma vez que número de interferentes presentes deve ser maior que no estudo do capítulo anterior e não se tem o conhecimento deste número para a construção dos modelos de calibração de segunda ordem. As espécies de interesse determinadas nesses produtos foram a riboflavina e a piridoxina (vitaminas B2 e B6, respectivamente), cujas estruturas moleculares são apresentadas na Figura 32. Estes analitos apresentam grande importância em sua determinação devido às funções que apresentam no organismo e também pelo alto consumo por crianças com idade próxima de um ano ou o público em geral no caso dos suplementos alimentares. Portanto, os objetivos principais neste caso foram a determinação simultânea desses analitos em três marcas de leite em pó infantis, identificados por leite 1, 2 e 3 e em dois suplementos alimentares, identificados como suplemento 1 e 2 por EEF, evitando a necessidade de separação física dos componentes da amostra por HPLC, geralmente proposta na literatura <sup>107,108</sup>, e a comparação dos resultados obtidos com os modelos BLLS, UPLS e PARAFAC.



**Figura 32-** Estruturas moleculares das vitaminas B2 (esquerda) e B6 (direita).

## **3.1 Procedimento experimental**

### **3.1.1 Estudo de estabilidade das soluções de trabalho**

Inicialmente, para garantir a ausência de perdas por decomposição durante as medidas de EEF, foram feitos testes de estabilidade com as vitaminas B2 e B6 em um espectrofotômetro Agilent HP8453 utilizando uma solução de concentração de 5,0 mg/L de cada vitamina e um acompanhamento por um período de 21 horas e medidas a cada 15 minutos.

### **3.1.2 Procedimento das medidas de Emissão e Excitação de Fluorescência (EEF).**

As medidas de EEF foram realizadas em um espectrofluorímetro da marca PerkinElmer modelo LS 55, sendo os espectros de excitação cobrindo o intervalo de comprimentos de onda de 250 e 478 nm com incrementos de 12 nm e os espectros de emissão monitorados entre 300 a 600 nm com resolução de 0,5 nm e fendas dos monocromadores de emissão e de excitação mantidas em 10 nm. Todas as medidas foram realizadas em triplicata e a 24 °C mantida através de um banho termostaticado no qual as amostras foram conservadas desde o seu preparo até o momento da análise.

### **3.1.3 Preparo das soluções padrão**

Para a preparação das soluções empregadas na construção dos modelos de calibração foram utilizados padrões bioquímicos das vitaminas B2 e B6 com purezas de 98,38 % e 99,50 %, respectivamente. Inicialmente 25 mg de cada uma das vitaminas B2 e B6 foram pesadas em balança analítica, e em seguida dissolvidas em ácido clorídrico concentrado e água destilada, respectivamente. Após dissolução, as vitaminas foram transferidas para um balão volumétrico de 500 mL (que já continha determinado volume de água) e o volume completado com

água destilada, fornecendo uma solução estoque contendo as duas vitaminas na concentração de 50,00 mg/L. Em seguida, através das diluições apropriadas em tampão  $\text{H}_2\text{PO}_4^-/\text{HPO}_4^{2-}$  (0,12 mol/L) ajustado para pH=7,0 com NaOH, chegou-se a uma solução de trabalho de concentração 250  $\mu\text{g/L}$  e a partir dessa solução foram preparadas cinco soluções padrão contendo as vitaminas B2 e B6 nas concentrações de 5,00; 20,00; 37,50; 49,50 e 70,00  $\mu\text{g/L}$  de ambas as vitaminas. Após preparadas, todas as soluções foram mantidas em frascos âmbar ou cobertos por papel alumínio.

### **3.1.4 Preparo das amostras para as medidas de EEF**

Inicialmente foi preparada uma solução de cada produto através da dissolução em água de uma massa apropriada em um volume de 100,00 mL. Em seguida, em um balão volumétrico de 50,00 mL, foram adicionados 5,0 mL (para os produtos suplementos 1 e 2) ou 10,0 mL (para os leites 1, 2 e 3) de ácido tricloroacético (TCA) na concentração de 33 % (m/m) e completado o volume do balão com a solução preparada anteriormente de cada produto. Em seguida, a solução foi agitada e depois centrifugada por 5 minutos a 2000 rotações por minuto e o sobrenadante filtrado em funil de vidro com papel filtro comum. O extrato obtido foi diluído 10 vezes com  $\text{NaH}_2\text{PO}_4$  0,12 mol/L, corrigindo o pH para 7,0 sempre que necessário com NaOH na concentração de 2,0 mol/L. Em seguida, 25,0 mL da solução anterior de cada produto foram transferidos para um balão volumétrico de 50,0 mL e seu volume completado com tampão  $\text{HPO}_4^{2-}/\text{H}_2\text{PO}_4^-$  (pH=7,0). Logo, foi utilizado um fator de diluição de 20 vezes do extrato obtido após a extração com TCA. Este fator de diluição foi estabelecido através de testes com fatores de diluição de 10, 20, 30, 40 e 60 vezes, realizados para avaliar a possibilidade de supressão de fluorescência, onde se observou que a partir de uma diluição de 20 vezes não era observada supressão.

Para a realização de testes de recuperação, foram feitas adições das vitaminas B2 e B6 em três níveis distintos na última etapa do procedimento descrito no parágrafo anterior. Antes de completar o volume do balão na última etapa de diluição, foram adicionados 1,0; 2,0 e 3,0 mL da solução de trabalho de concentração 250,0 µg/L, que correspondem à incrementos na concentração da amostra de aproximadamente 5, 10 e 15 µg/L, respectivamente. A Tabela 10 apresenta os valores das concentrações esperadas de cada produto nas amostras sem adição, tendo em conta o procedimento experimental descrito e o valor nominal atestado no rótulo pelos fabricantes.

**Tabela 8.** Concentrações esperadas, em µg/L, para as vitaminas B2 e B6 nas análises por EEF para cada produto considerando o teor no rótulo.

Leite 1		Leite 2		Leite 3		Suplemento 1		Suplemento 2	
B2	B6	B2	B6	B2	B6	B2	B6	B2	B6
39,8	22,9	37,7	18,9	33,9	27,1	21,1	45,3	34,4	43,5

### **3.1.5 Procedimento das medidas cromatográficas utilizadas para a validação.**

Para validação dos resultados obtidos por EEF foram realizadas análises dos mesmos produtos utilizando um equipamento HPLC da marca Shimadzu Prominence equipado com amostrador automático, detector por arranjo de diodos, velocidade de aquisição de 1,5 pontos/segundo e uma coluna cromatográfica C18 da marca Microsorb MV com partículas de diâmetro de 5 µm. Foi utilizado como fase móvel acetonitrila:S1 20:80 (v/v), onde S1 é uma solução aquosa de dodecilsulfato de sódio na concentração 5 mg/L com pH ajustado para 3,0. Esta foi utilizada para facilitar a separação da vitamina B6 devido à formação de um par iônico fazendo com que esse analito tenha maior interação com a fase estacionária da coluna. Todas as soluções padrão foram preparados uma única vez e analisados em triplicata, enquanto que as amostras de cada produto foram tanto preparadas

como injetadas em triplicata, sendo os comprimentos de onda utilizados para a determinação das vitaminas B2 e B6 iguais a 450 e 290 nm, respectivamente.

### **3.1.6 Preparo de soluções padrão e amostras para as medidas por HPLC**

As soluções padrão foram preparadas de forma similar ao procedimento descrito para as medidas por EEF, sendo a única distinção o intervalo de concentrações utilizado para as calibrações de ordem zero. Foram utilizados seis padrões para construção das curvas analíticas, com concentrações iguais a de 0,25; 0,50; 1,00; 2,50; 5,00 e 7,50 mg/L para ambas as vitaminas.

Tendo em vista que o intervalo de concentrações nas calibrações por EEF está no nível de  $\mu\text{g/L}$ , enquanto que para as medidas de HPLC este intervalo de concentrações se encontra no nível de  $\text{mg/L}$ , o preparo das amostras de leite e suplementos alimentares teve de ser diferente em cada caso. Para as medidas por HPLC foi utilizada uma massa de 20 g para todos os produtos, a qual foi pesada, dissolvida em água destilada e transferida para um balão volumétrico de 100 mL sem completar o volume do balão. Em seguida esta solução foi deixada no ultrassom por 30 minutos, sendo três etapas de 10 minutos com uma agitação manual da solução entre cada etapa. Em seguida foram acrescentados 20 mL de solução de ácido tricloroacético (TCA) na concentração de 33 % (m/m) e completado o volume do balão volumétrico com água destilada. Após a adição da solução de TCA para a precipitação das proteínas, a solução foi novamente levada ao ultrassom por 30 minutos, com agitação manual a cada 10 minutos. Depois de decorridos os 30 minutos no ultrassom, a solução foi centrifugada por 5 minutos a 2000 rotações por minuto e o sobrenadante filtrado em funil de vidro com papel de filtro comum de laboratório. Os extratos obtidos foram armazenados em frascos âmbar, cobertos com papel alumínio e mantidos à 4 °C até o momento das medidas por HPLC, as quais foram realizadas no máximo no dia seguinte ao da preparação do extrato.

Imediatamente antes das injeções no HPLC, os extratos foram filtrados em membranas de 0,45µm (Millipore). A Tabela 9 apresenta as concentrações esperadas em cada produto nas análises por HPLC, tendo em vista o procedimento descrito.

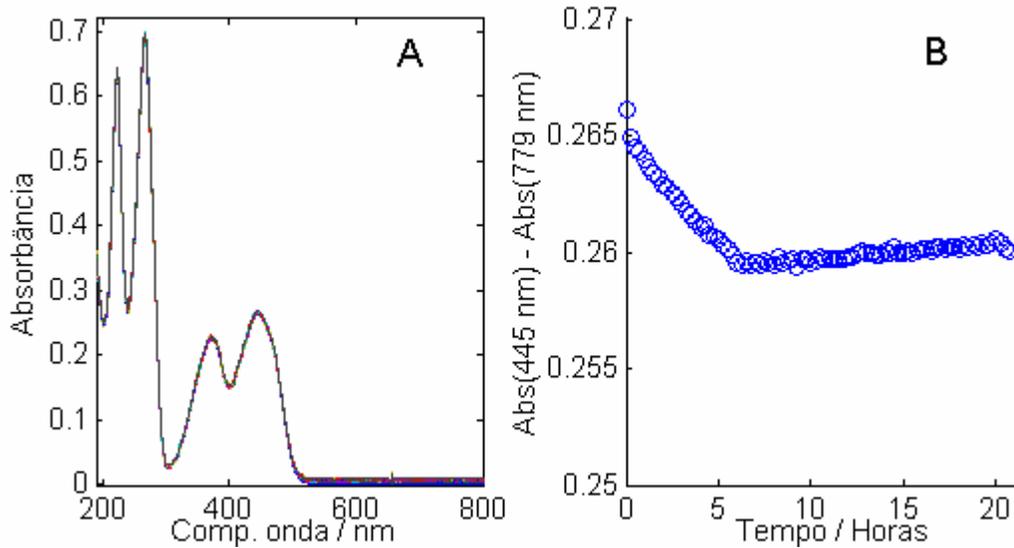
**Tabela 9.** Concentrações esperadas (mg/L) nas análises por HPLC para as vitaminas B2 e B6 em cada produto levando em conta o valor atestado no rótulo por cada fabricante e o preparo de amostra.

Leite 1		Leite 2		Leite 3		Suplemento 1		Suplemento 2	
B2	B6	B2	B6	B2	B6	B2	B6	B2	B6
1,79	1,04	1,56	0,78	2,40	1,92	2,67	5,73	3,81	4,81

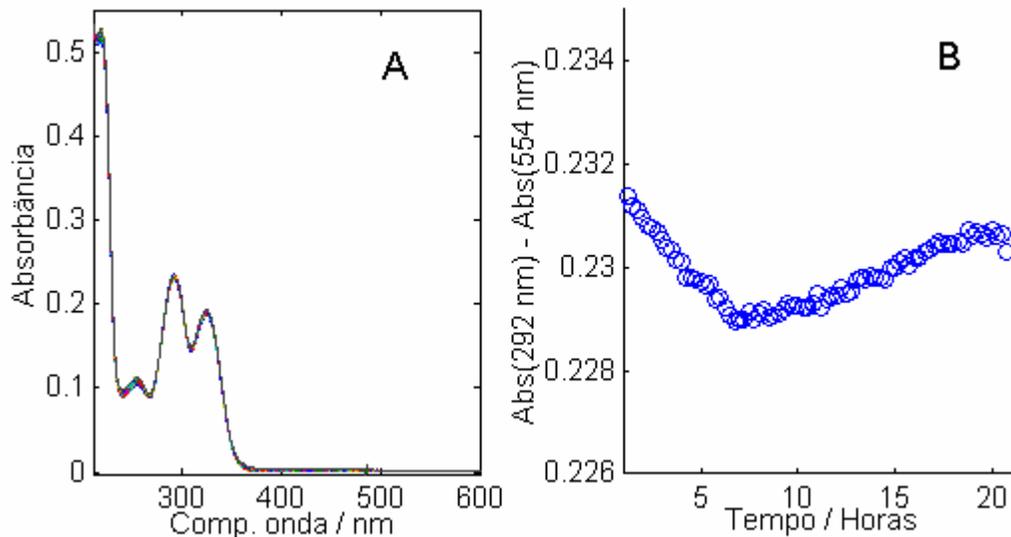
### 3.2 Resultados e discussão

As Figuras 33 e 34 apresentam os espectros adquiridos e os perfis cinéticos para cada vitamina no intervalo de 21 horas. Pelos espectros (Figuras 33A e 34A) observa-se que não há uma variação evidente dos valores de absorvância ao longo do tempo. Através do acompanhamento do perfil cinético nos comprimentos de onda de 445 e 292 nm para B2 e B6, respectivamente (não apresentados), notou-se uma variação não aleatória para ambas as vitaminas. Contudo, devido à flutuação da linha base não foi possível chegar a uma conclusão sobre a existência ou não de decomposição de ambos os compostos. Este problema aparentemente causado pela flutuação do sinal do equipamento foi corrigido pela subtração entre a absorvância dos comprimentos de onda escolhidos pela absorvância de um comprimento de onda que não apresentava absorção pelos analitos. As Figuras 33B e 34B mostram os resultados desse procedimento, onde se observou uma aparente queda correspondente à decomposição das espécies nas primeiras 6 horas. Contudo, essa queda corresponde à apenas aproximadamente  $5,0 \cdot 10^{-3}$  unidades de absorvância, o que corresponde a menos de 1% da absorvância, dependendo do comprimento de

onda considerado. Logo, foi considerado que as soluções estoque eram estáveis o suficiente para o manuseio no dia em que foram preparadas.



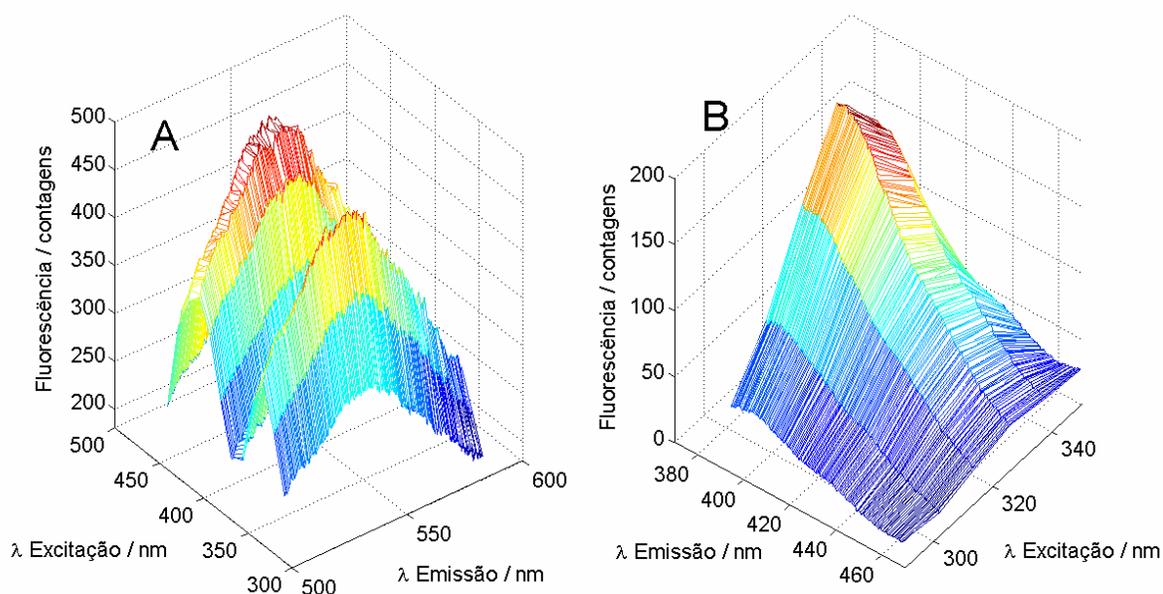
**Figura 33-** (A) Espectros da vitamina B2 adquiridos a cada 15 minutos na região UV-VIS em um intervalo de 21 horas. (B) Perfil cinético obtido pela subtração das absorvâncias em 445 nm e um comprimento de onda sem absorção (779 nm).



**Figura 34-** (A) Espectros da vitamina B6 adquiridos a cada 15 minutos na região UV-VIS em um intervalo de 21 horas. (B) Perfil cinético obtido pela subtração das absorvâncias em 292 nm e um comprimento de onda sem absorção (554 nm)

Na Figura 35, são apresentadas as superfícies obtidas para B2 e B6,

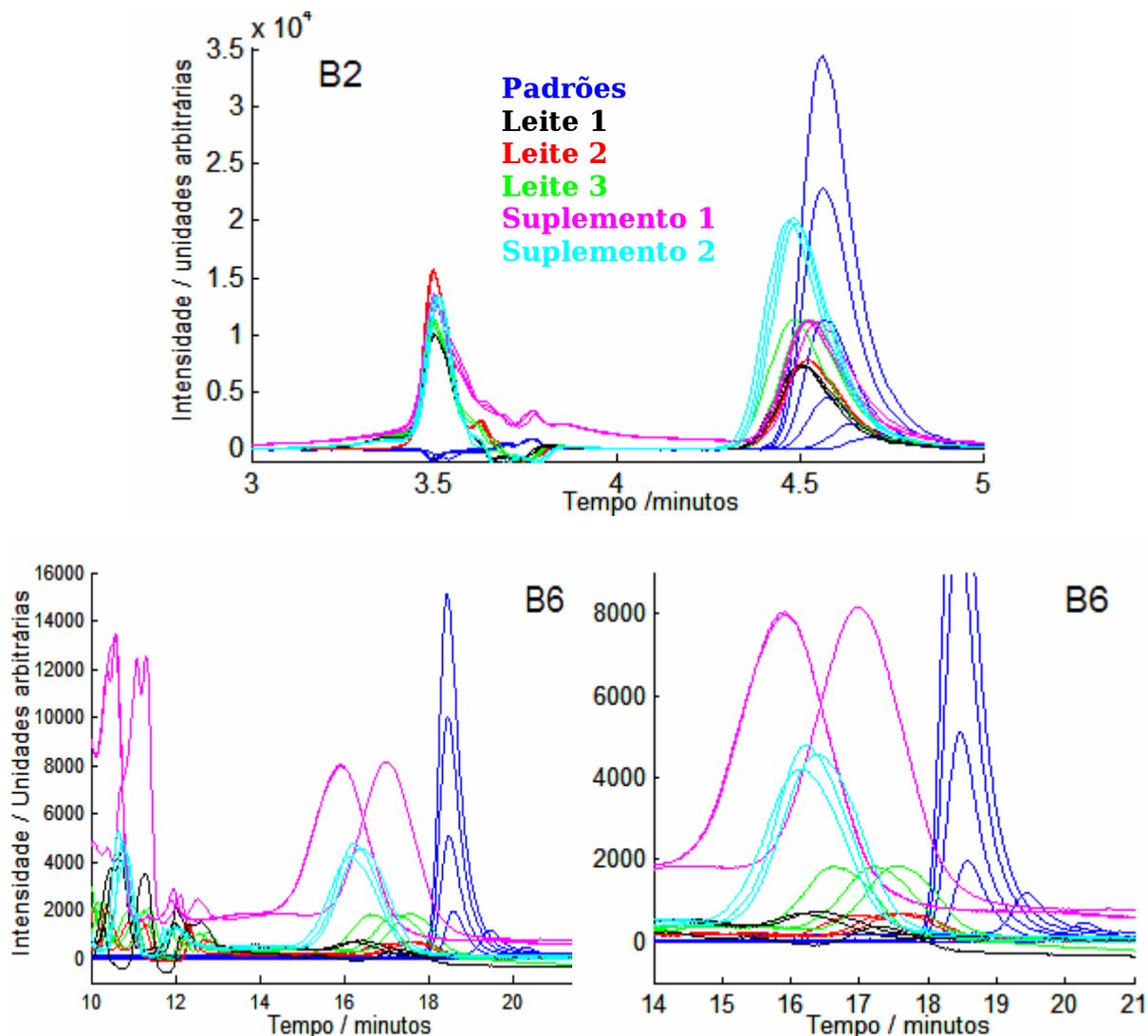
considerando apenas as regiões sem espalhamentos e que foram selecionadas para a construção dos modelos de calibração, em uma solução padrão de calibração de concentração 70 µg/L. Pelas intensidades de fluorescência observadas nota-se que, na ausência de interferentes, a vitamina B2 deve apresentar sensibilidade significativamente maior que a B6, além de apresentar um espectro de excitação mais característico e sua emissão se localizar na região de comprimentos de onda do visível, o que facilita a determinação desse analito por EEF.



**Figura 35-** Superfícies de EEF para B2 (A) e B6 (B) obtidas para a solução padrão de concentração de 70,0 µg/L.

A Figura 36 apresenta os cromatogramas obtidos para cada uma das vitaminas, onde se observa que para a vitamina B2 foram obtidos picos cromatográficos apresentando boa resolução em relação aos interferentes, perfis com aspectos aproximadamente gaussianos e tempos de retenção com pequena variação entorno de 4,5 a 4,6 minutos. Para a vitamina B6 observam-se que os cromatogramas dos produtos analisados apresentam maiores desvios tanto nos tempos de retenção como nos perfis dos picos cromatográficos, os quais apresentam maior alargamento em relação aos picos obtidos nas soluções padrões.

As regressões obtidas a partir dos cromatogramas dos padrões apresentaram coeficientes de correlação de ao menos 0,999 para as duas vitaminas e erros relativos de 3,0 e 2,0 % para B2 e B6, respectivamente.

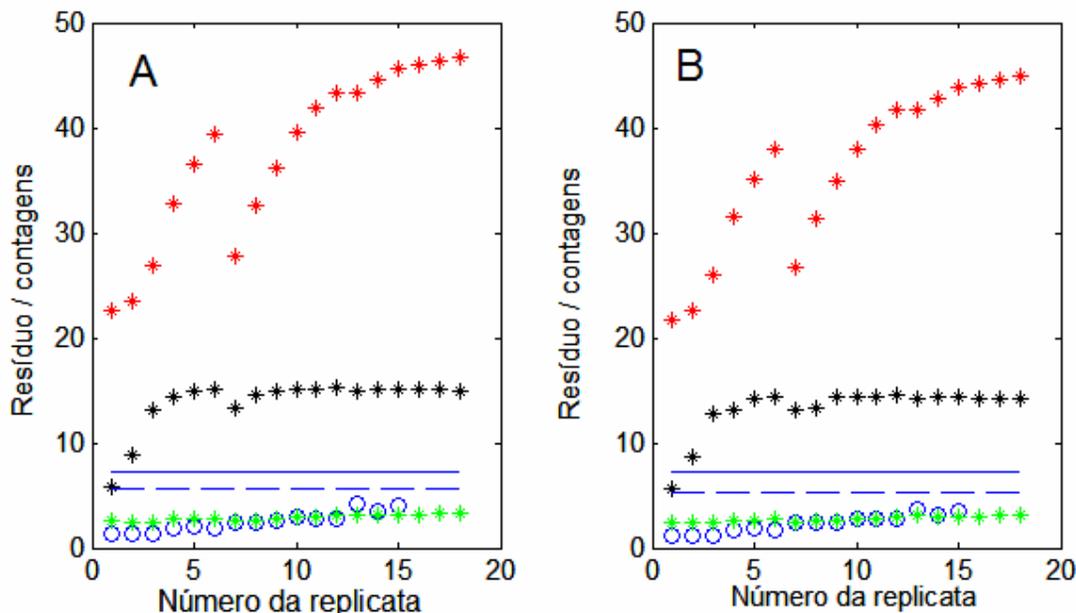


**Figura 36-** Cromatogramas obtidos pelas análises por HPLC para as vitaminas B2 e B6 utilizando os comprimentos de onda de 450 e 290 nm, respectivamente.

Para a determinação da vitamina B2 por EEF foi observado que apenas o Suplemento 1 apresentou interferência de outros compostos na mesma região de excitação e emissão. Logo, a utilização de calibração de segunda ordem somente foi necessária para esse produto, sendo a determinação nos demais produtos realizada com base em uma calibração de ordem zero a partir dos comprimentos de

onda 454 e 547 nm de excitação e emissão, respectivamente.

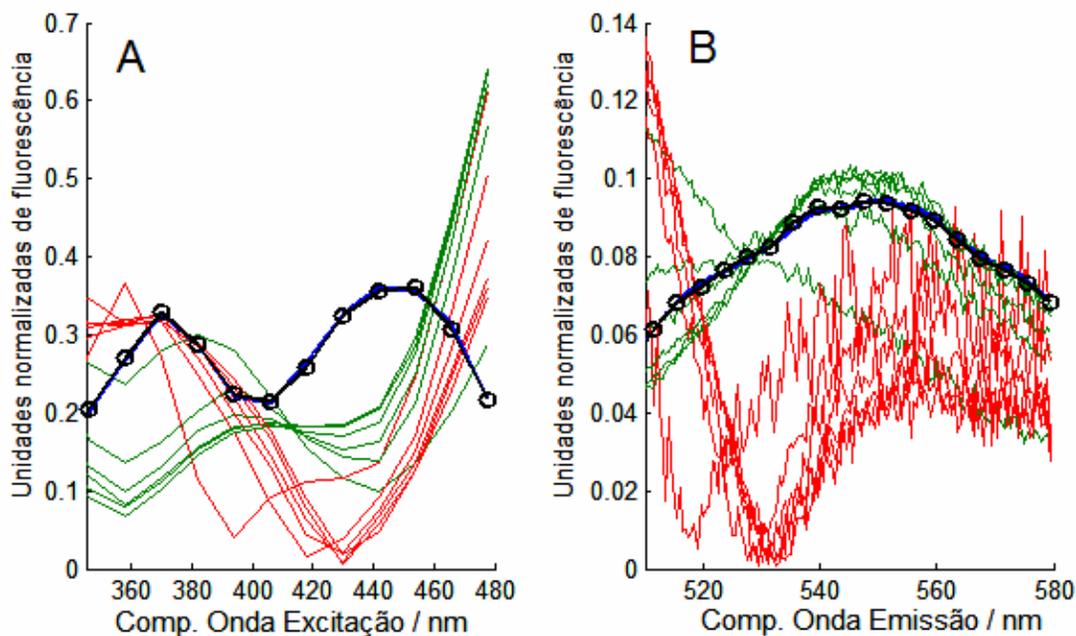
A Figura 37 mostra os resíduos de decomposição obtidos com os modelos BLLS e UPLS, os quais indicam a necessidade da utilização de dois fatores para modelar a interferência presente, número que de fato é o que proporciona os menores erros entre os resultados obtidos por EEF e HPLC. Para essa situação, observa-se que tanto os valores críticos baseados no nível do ruído instrumental como no resíduo médio das amostras de calibração são eficazes para estimar o número de fatores, com a exceção de uma amostra na qual o valor crítico pelo ruído instrumental indica a necessidade de apenas um fator. Para o PARAFAC o melhor resultado foi obtido por um modelo com dois fatores e a restrição de não-negatividade nas três dimensões.



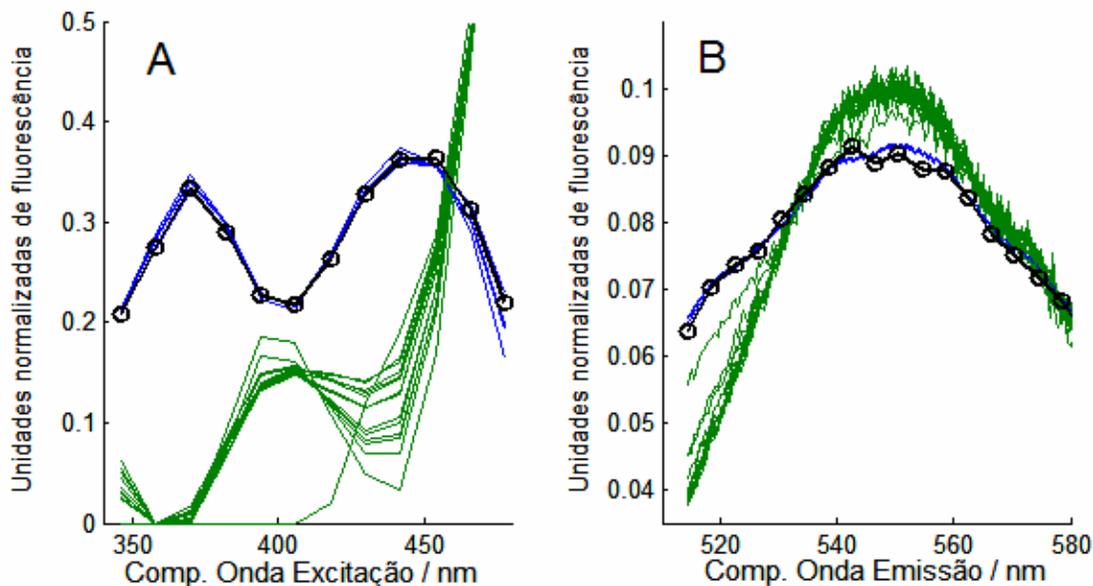
**Figura 37-** Resíduos de decomposição para os modelos (A) BLLS e (B) UPLS obtidos na análise do produto Suplemento 1. (\*) Nenhum (\*) um e (\*) dois fatores utilizados no RBL. (—) Valor crítico pelo ruído instrumental e (---) pelo resíduo médio das amostras de calibração, ambos com 99% de confiança.

Os espectros de excitação e emissão recuperados com os modelos BLLS e PARAFAC são apresentados nas Figuras 38 e 39, respectivamente, onde se observa que os perfis espectrais estimados para B2 apresentam ótima concordância com o

perfil experimental obtido nos padrões, enquanto que os perfis dos interferentes apresentam maior variação.



**Figura 38-** Espectros normalizados de excitação (A) e emissão (B) obtidos pelo modelo BLLS na análise da vitamina B2 no produto Suplemento 1. (—) Espectros estimados para B2, (—) para o primeiro e (—) segundo interferente e (-o-) perfil experimental da B2 obtido nos padrões.



**Figura 39-** Espectros normalizados de excitação (A) e emissão (B) obtidos pelo modelo PARAFAC na análise da vitamina B2 no produto Suplemento 1. (—) Espectros estimados para B2, (—) para o interferente e (-o-) perfil experimental da B2 obtido nos padrões.

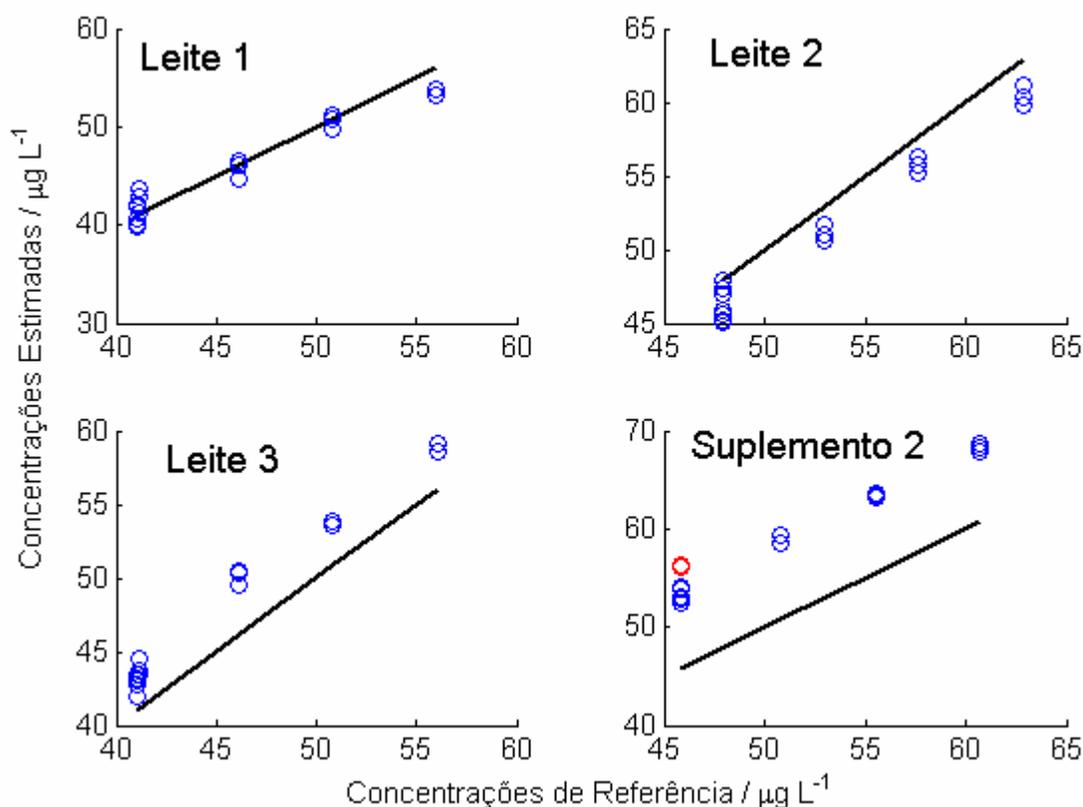
Os resultados das estimativas das figuras de mérito para determinação de vitamina B2 são apresentados na Tabela 10, onde se pode observar que os melhores resultados são obtidos para os leites 1, 2 e 3, nos quais se obtém erros relativos inferiores a 10%. No caso do suplemento 1, onde os modelos de calibração de segunda ordem foram empregados, observa-se que os modelos BLLS e UPLS apresentam melhor concordância com as concentrações obtidas por HPLC do que o PARAFAC, enquanto que este último apresenta uma melhor precisão e sensibilidade.

**Tabela 10.** Resultados das figuras de mérito obtidas para a determinação de B2, estimados comparando os resultados obtidos por HPLC e por EEF.

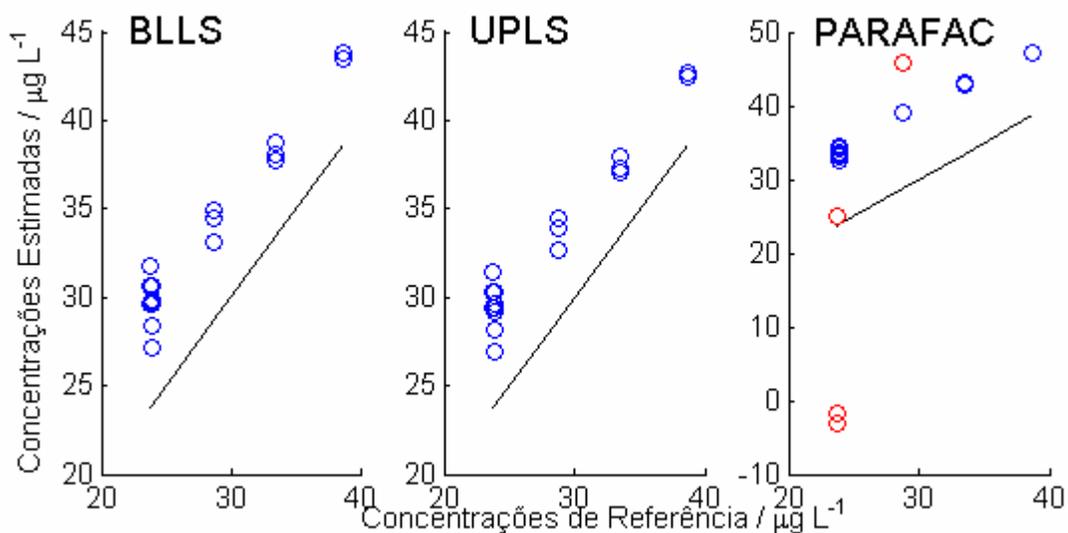
Modelo	RMSEC <sup>a</sup>	RMSEP <sup>a</sup>	REP <sup>b</sup>	Precisão <sup>a,c</sup>	(SENA) <sup>-1 d</sup>	SEL	LD <sup>a</sup>
Supl. 1-BLLS	0,2	5,6	21	±0,8	0,50(±0,01)	0,04(±0,01)	1,5(±0,3)
Supl. 1-UPLS	0,2	5,1	20	±0,8	0,50(±0,01)	0,04(±0,01)	1,6(±0,3)
Supl. 1-PARAFAC	0,2	9,5	34	±0,2	0,20(±0,04)	0,09(±0,01)	0,6(±0,1)
Leite 1 <sup>e</sup>	0,5	1,4	3	±0,8	0,42	1	1,4(±0,3)
Leite 2 <sup>e</sup>	0,5	2,1	4	±0,5	0,42	1	1,4(±0,3)
Leite 3 <sup>e</sup>	0,5	2,8	6	±0,5	0,42	1	1,4(±0,3)
Suplemento 2 <sup>e</sup>	0,5	7,8	15	±0,5	0,42	1	1,4(±0,3)

Grandezas expressas em: (a) µg/L, (b) percentagem, (c) número de graus de liberdade igual a 12, (d) (µg/L) e (e) obtido por regressão univariada. Desvios padrões apresentados entre parênteses. (Supl. 1) Suplemento 1

Observando as Figuras 40 e 41, as quais mostram os gráficos das concentrações de referência versus estimadas, nota-se que apenas para o leite 1 não se observa erros sistemáticos na determinação da vitamina B2, sendo estes erros sistemáticos negativos para o leite 2 e positivos nos demais produtos. Na Figura 41 também pode-se visualizar a melhor precisão do modelo PARAFAC, mas que apresenta altos erros na estimativa de 4 replicatas.



**Figura 40-** Gráficos das concentrações de referência (considerando as estimativas por HPLC) versus estimadas para a vitamina B2 nos produtos onde se utilizou calibração de ordem zero. (—) Bissetriz do quadrante, (●) amostras, (●) replicatas consideradas anomalias.



**Figura 41-** Gráficos das concentrações de referência (considerando as estimativas por HPLC) versus estimadas para a vitamina B2 para o suplemento 1 utilizando calibração de segunda ordem. (—) Bissetriz do quadrante, (●) amostras, (●) replicatas consideradas anomalias.

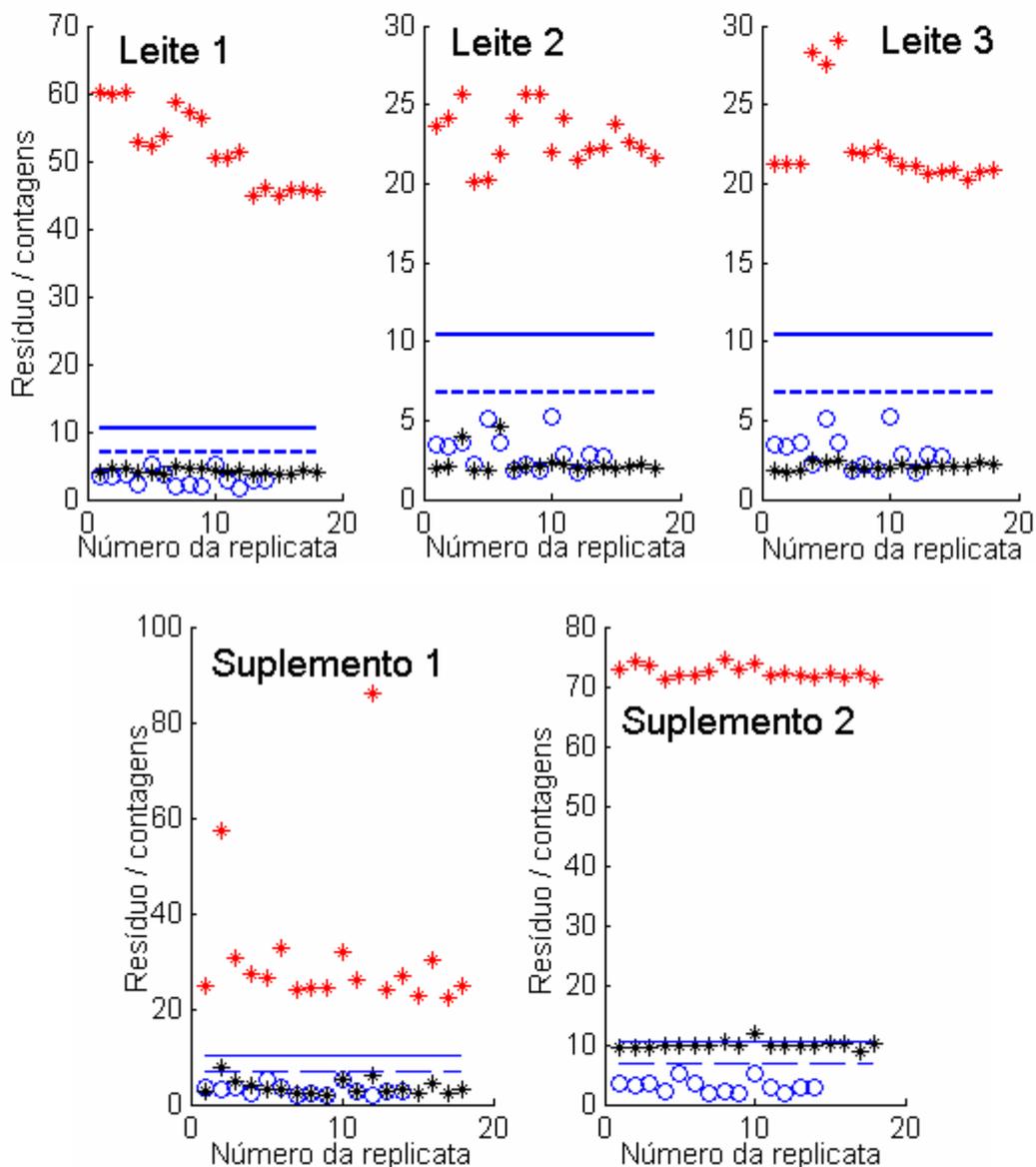
A Tabela 11 apresenta os resultados dos testes de recuperação que foram realizados. A alta variação observada pode ser explicada pela dispersão visualizada nas Figuras 40 e 41, nas quais o primeiro nível corresponde à concentração de B2 nos produtos analisados e os três níveis superiores às adições realizadas. As recuperações percentuais apresentadas na Tabela 11 variam de 82 a 137 %; no entanto, estas podem ser consideradas um bom resultado, uma vez que um erro de apenas 1 µg/L corresponde a um erro de 20 % no primeiro nível de concentração.

**Tabela 11.** Recuperações obtidas para as determinações de B2.

Adições (µg/L)	Recuperações (%)						
	Suplemento 1			Leite 1	Leite 2	Leite 3	Suplemento 2
	BLLS	UPLS	PARAFAC				
5	89(±19)	85 (±18)	109(±3)	88(±18)	101(±11)	137(±10)	115(±10)
10	88 (±5)	82 (±5)	97(±1)	95(±8)	100(±6)	108(±1)	104(±2)
15	93 (±1)	88 (±2)	92(±1)	82(±2)	96(±4)	105(±2)	100(±3)

Desvios padrões com 2 graus de liberdade apresentados entre parênteses.

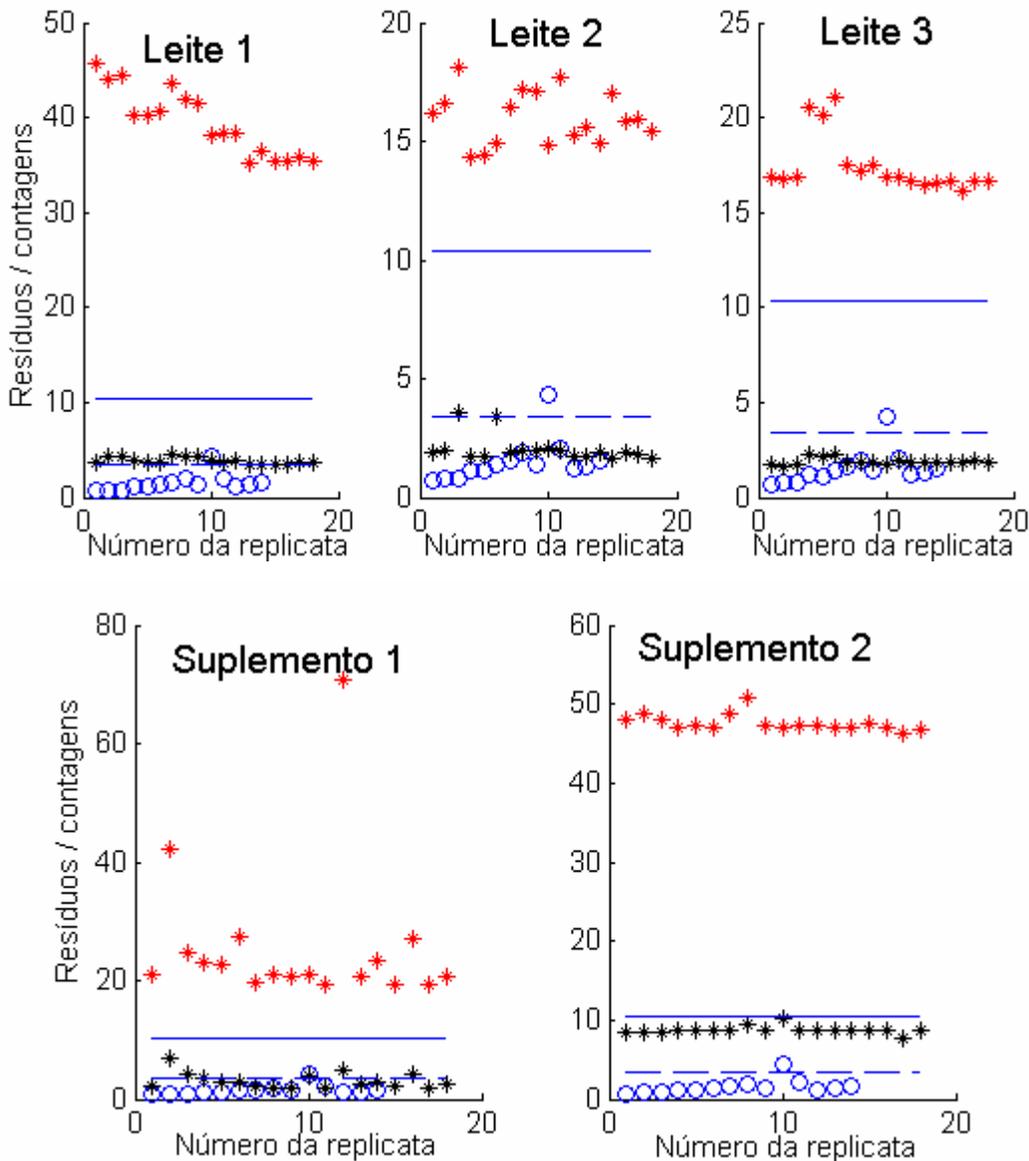
Para a determinação da vitamina B6 foi observada a ocorrência de interferentes em todos os produtos. As Figuras 42 e 43 mostram os gráficos de resíduos utilizados para a escolha do número de fatores para o RBL nos modelos BLLS e UPLS, respectivamente. Na Figura 42 observa-se que para os leites 1, 2 e 3 e para o suplemento 1 os limites críticos indicam a necessidade de apenas um fator no RBL, enquanto que para o suplemento 2 há uma discordância entre os limites, sendo o limite estabelecido pelo resíduo médio das amostras de calibração indicando a necessidade de dois fatores e pelo ruído apenas um. Pelos resultados das estimativas das concentrações, observou-se que com apenas um fator em todos os produtos tem-se estimativas mais concordantes com as obtidas pelo método HPLC.



**Figura 42-** Resíduos de decomposição obtidos com o BLS na análise de cada produto. (\*) Nenhum e (o) um fator no RBL. (—) Valor crítico pelo ruído instrumental e (---) pelo resíduo médio das amostras de calibração, ambos com 99% de confiança.

Observando os resíduos das amostras de calibração do modelo UPLS (Figura 43) nota-se que a replicata 10 apresenta um resíduo superior em relação aos observados nos demais padrões, no entanto, esta replicata não foi retirada para a construção do modelo devido sua estimativa da concentração apresentar um baixo erro. Observa-se que para quase todas as amostras dos leites 2 e 3 e do suplemento 1 os dois limites concordam com a utilização de apenas um fator no RBL, enquanto

que para leite 1 e suplemento 2 uma discordância semelhante aos resultados com o BLS foi observada. Assim como no BLS, os melhores resultados foram obtidos utilizando apenas um fator no RBL. Quanto às replicatas das amostras de previsão que apresentam um maior resíduo em relação às demais, observou-se que em alguns casos essas amostras possuem um alto erro na estimativa da concentração, sendo descartadas no cálculo das figuras de mérito.



**Figura 43-** Resíduos de decomposição obtidos com o UPLS na análise de cada produto. (\*) Nenhum e (\*) um fator no RBL. (—) Valor crítico pelo ruído instrumental e (---) pelo resíduo médio das amostras de calibração, ambos com 99% de confiança.

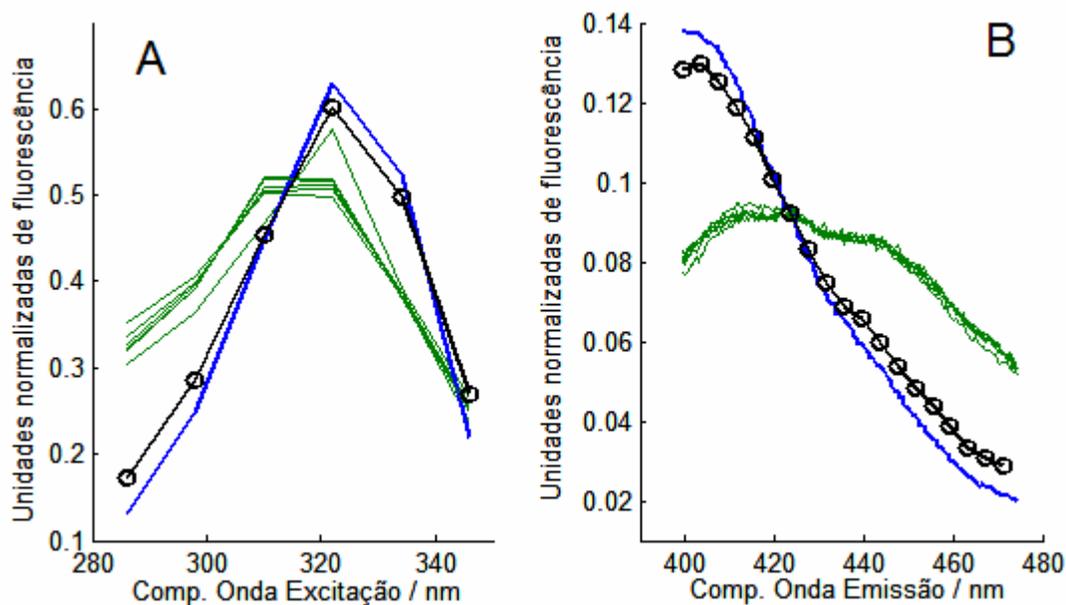
**Tabela 12.** Resultados das figuras de mérito obtidas para determinação de B6, estimados comparando os resultados obtidos por HPLC e por EEF.

Modelo/produto	RMSEC <sup>a</sup>	RMSEP <sup>a</sup>	REP <sup>b</sup>	Precisão <sup>a</sup>	(SENA) <sup>-1 c</sup>	SEL	LD <sup>a</sup>
<b>BLLS</b>							
Leite 1	2,2	3,8	24	±1,7	1,2(±0,1)	0,13(±0,01)	3,8(±0,3)
Leite 2	2,2	3,7	16	±0,8	0,9(±0,1)	0,17(±0,03)	3,0(±0,4)
Leite 3	2,2	7,0	27	±0,8	0,9(±0,1)	0,17(±0,01)	3,0(±0,2)
Suplemento 1	2,2	4,3	7	±2,9	1,1(±0,1)	0,13(±0,01)	3,7(±0,4)
Suplemento 2	2,2	4,8	12	±4,4	1,4(±0,3)	0,11(±0,03)	5(±1)
<b>UPLS</b>							
Leite 1	1,0	8,4	60	±1,9	1,0(±0,1)	0,19(±0,02)	3,3(±0,3)
Leite 2	1,0	5,6	27	±0,8	0,7(±0,1)	0,27(±0,04)	2,3(±0,3)
Leite 3	1,0	8,4	33	±0,5	0,8(±0,4)	0,23(±0,01)	2,7(±0,2)
Suplemento 1	1,0	3,0	4,3	±2,0	1,0(±0,1)	0,21(±0,03)	3,1(±0,4)
Suplemento 2	1,0	15	31	±4,5	1,3(±0,3)	0,16(±0,04)	4,2(±1)
<b>PARAFAC</b>							
Leite 1	1,2	9,8	65	±1,7	0,9(±0,3)	0,17(±0,02)	3,0(±0,1)
Leite 2	1,2	3,7	20	±0,3	0,8(±0,1)	0,19(±0,01)	2,7(±0,1)
Leite 3	1,2	9,2	36	±0,5	0,8(±0,1)	0,19(±0,01)	2,7(±0,3)
Suplemento 1	1,2	2,5	3,7	±0,9	0,9(±0,1)	0,18(±0,01)	3,1(±0,1)
Suplemento 2	1,2	8,6	22	±3,5	0,9(±0,2)	0,18(±0,02)	3,0(±0,1)

Desvios padrões entre parênteses e grandezas expressas em: (a)µg/L, (b) porcentagem, (c) (µg/L).

Pelos valores de RMSEP e REP obtidos para as determinações da vitamina B6 (Tabela 12) observa-se que para o suplemento 1 foram obtidos os melhores resultados, independente do modelo de calibração utilizado. Nos demais produtos observam-se erros maiores que, assim como na determinação da vitamina B2,

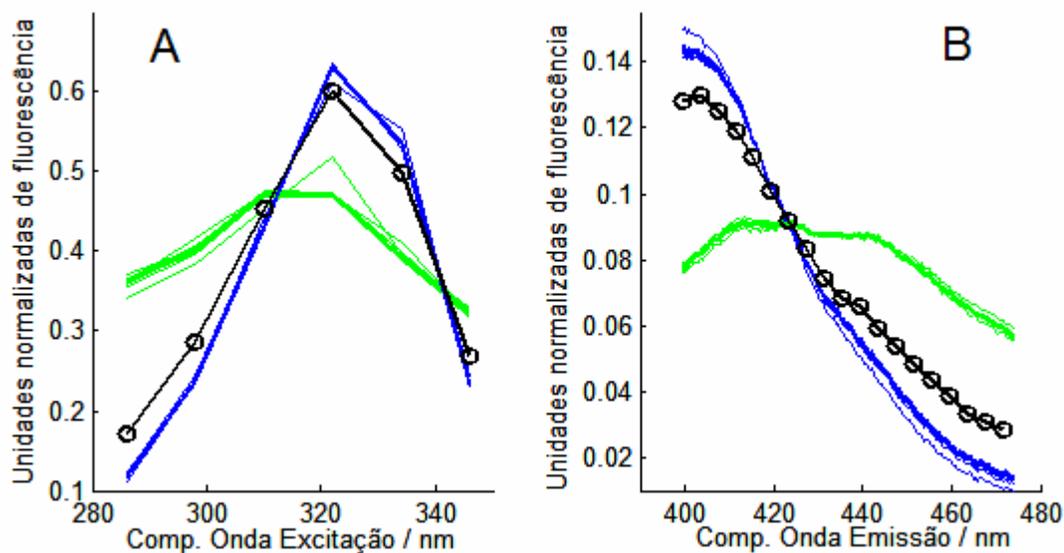
refletem a presença de erros sistemáticos que são provavelmente provenientes das diferenças entre os métodos de extração utilizados no preparo das amostras para as medidas em EEF e HPLC. Comparando os modelos pelos erros médios obtidos, as maiores diferenças são observadas para o leite 1 onde o BLS apresenta resultados significativamente melhores que o PARAFAC e UPLS. Pelos valores de RMSEP e REP, o BLS foi o modelo que apresentou melhor concordância com o método por HPLC, obtendo-se menores erros para quatro dos cinco produtos analisados (leites 1, 2 e 3 e suplemento 2).



**Figura 44-** Espectros normalizados de excitação (A) e emissão (B) obtidos pelo modelo BLS na análise da vitamina B6 no leite 1. (—) Espectros estimados para B6, (—) para o interferente e (-o-) perfil experimental da B2 obtido nos padrões.

Ainda observando os erros médios da Tabela 12, nota-se que em geral foram obtidos resultados inferiores para a determinação da vitamina B6 do que para a B2. Dois fatores que podem ter contribuído para esses resultados foram: (1) o intervalo de comprimentos de onda de excitação e emissão que pode ser selecionado para a construção dos modelos de calibração, que não engloba todo o sinal de fluorescência desse composto devido à presença de espalhamentos; (2) uma

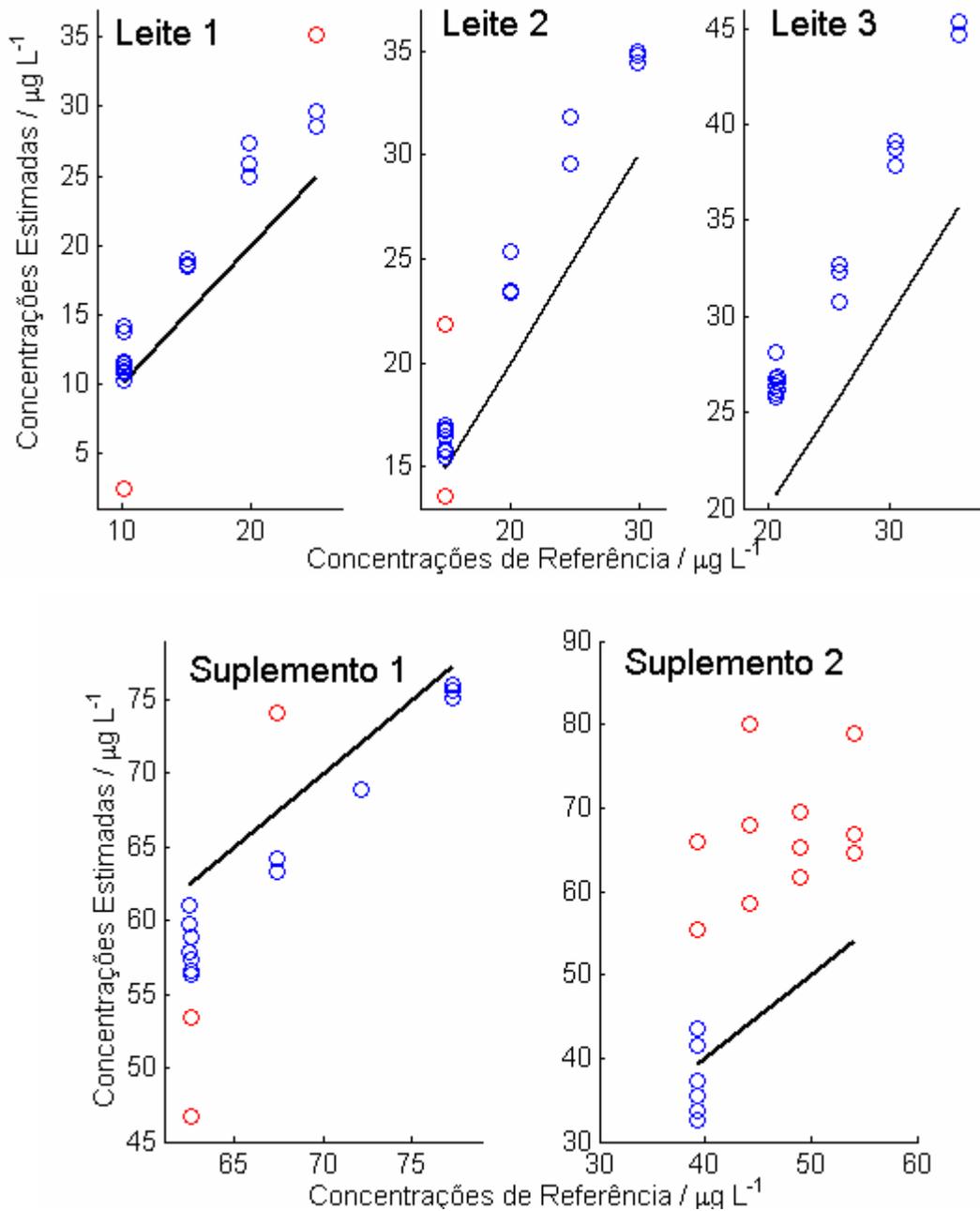
correlação entre o sinal do interferente e do analito, que pode ser observado nas Figuras 44 e 45 que mostram os perfis recuperados com o PARAFAC e BLS. Outra informação presente nessas figuras é que os espectros recuperados para a vitamina B6 apresentam maiores diferenças em relação aos espectros puros obtidos nos padrões do que para a vitamina B2, essa observação reflete uma maior dificuldade na resolução desse sistema.



**Figura 45-** Espectros normalizados de excitação (A) e emissão (B) obtidos pelo modelo PARAFAC na análise da vitamina B6 no leite 1. (—) Espectros estimados para B6, (—) para o interferente e (-o-) perfil experimental da B2 obtido nos padrões.

Pelos resultados de precisão observa-se que os menores desvios foram obtidos para os leites 2 e 3. Comparando os modelos observa-se que, assim como para a vitamina B2, o PARAFAC apresenta resultados mais precisos. Contudo, apenas no suplemento 1 o PARAFAC é significativamente mais preciso de acordo com um test-F com 95 % de confiança, onde os valores calculados e tabelados foram 4,9 e 3,18 (9 e 9 graus de liberdade), respectivamente. Deve-se notar que o número de graus de liberdade dos resultados de REP, RMSEP e precisão é diferente em cada produto, uma vez que foram observadas algumas replicatas com resultados significativamente diferentes em relação às demais replicatas, as quais

não foram utilizadas para o cálculo desses parâmetros. Na Figura 46 pode-se observar as replicatas que foram excluídas e também a dispersão em cada produto com o modelo BLLS, onde os números de graus de liberdade são 10, 10, 12, 9 e 4 para os leites 1, 2 e 3 e os suplementos 1 e 2, respectivamente.



**Figura 46-** Gráficos das concentrações de referência (considerando as estimativas por HPLC) versus estimadas para a vitamina B6 em cada produto utilizando o modelo BLLS. (—) Bissetriz do quadrante, (●) amostras, (○) replicatas consideradas anomalias.

Pelos resultados apresentados na Figura 46 para o suplemento 2 observa-se que as três adições feitas para a realização dos testes de recuperação apresentaram uma alta dispersão, que foi observada nos três modelos de calibração de segunda ordem utilizados. Este resultado proporcionou resultados insatisfatórios para as recuperações no suplemento 2, de modo que para este produto as recuperações obtidas não são apresentadas. A Tabela 13 apresenta as recuperações obtidas para os demais produtos, onde se observa resultados inferiores em relação aos obtidos com a vitamina B2. Deve ser ressaltado que os principais motivos para os resultados insatisfatórios obtidos para as recuperações foram os pequenos incrementos nos valores de concentração (de aproximadamente 5 µg/L) e os erros das regressões. Por exemplo, um RMSEP de 3 µg/L pode fazer com que um erro maior que 50 seja obtido para as recuperações, com é o caso do primeiro nível de recuperação do suplemento 1 com o UPLS.

**Tabela 13.** Recuperações obtidas para as determinações de B6.

Modelo/adições		Recuperações (%)		
<b>BLLS</b>	<b>Leite 1</b>	<b>Leite 2</b>	<b>Leite 3</b>	<b>Suplemento 1</b>
5 µg/L	136(±5)	143(±1)	107(±21)	121(±13)
10 µg/L	146(±12)	137(±1)	124(±7)	116(±1)
15 µg/L	116(±5)	124(±2)	122(±3)	119(±3)
<b>UPLS</b>	<b>Leite 1</b>	<b>Leite 2</b>	<b>Leite 3</b>	<b>Suplemento 1</b>
5 µg/L	128(±12)	155(±38)	103(±7)	172(±20)
10 µg/L	146 (±11)	135(±22)	122(±7)	147(±7)
15 µg/L	118(±8)	117(±1)	122(±2)	127(±15)
<b>PARAFAC</b>	<b>Leite 1</b>	<b>Leite 2</b>	<b>Leite 3</b>	<b>Suplemento 1</b>
5 µg/L	151(±8)	123(±10)	123(±13)	112(±9)
10 µg/L	151(±1)	106 (±4)	121(±3)	105(±3)
15 µg/L	137(±10)	105(±1)	118(±2)	106(±9)

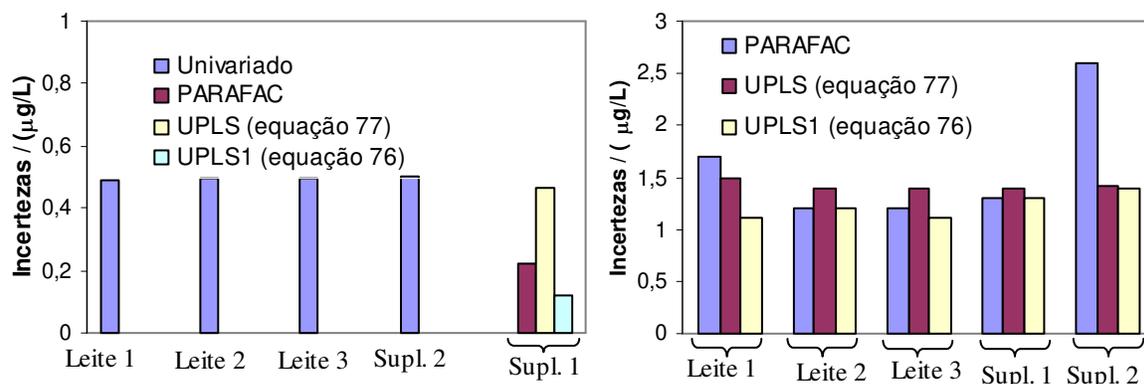
Desvios padrões entre parênteses.

Quanto aos resultados de SENA, SEL e LD (Tabela 12), observa-se que estes são resultados médios com seus respectivos desvios padrões, uma vez que quando há interferentes esses valores são específicos de cada amostra e não do método como um todo, diferente dos modelos de calibração de ordem zero e de primeira ordem.

Para os valores de LD (Tabela 12) observa-se que a concentração do primeiro ponto da curva analítica, que foi 5 µg/L, está muito próxima do LD estimado. Recordando que as estimativas de LD e SENA levam em conta apenas a variação devido ao ruído instrumental, o que sugere que esses valores provavelmente são estimativas otimistas dos seus valores verdadeiros. Logo, pode-se indicar o baixo intervalo de concentração como outro fator que pode ter contribuído para os erros obtidos na determinação e nas recuperações, pois, pelos valores de SENA e LD, se está trabalhando no limite da sensibilidade do método.

Os valores de seletividade tanto para B2 como para B6 mostram boa concordância entre os resultados obtidos com o UPLS e os modelos BLLS e PARAFAC, o que reforça a validade da proposta de sua determinação pelo UPLS.

Os valores das incertezas são apresentados na Figura 47, onde se observam valores significativamente maiores para a determinação da vitamina B6 em todos os modelos. Nesta aplicação, a maior diferença observada entre os valores de incerteza do UPLS com as equações 76 e 77 foi obtida na determinação da vitamina B2 no suplemento 1, sendo pequenas as diferenças observadas na determinação da vitamina B6. Infelizmente, devido aos erros sistemáticos observados e o pequeno número de amostras neste caso, a determinação e validação dos intervalos de confiança pelas probabilidades de recobrimento não foi possível.



**Figura 47-** Incertezas médias estimadas para as determinações das vitaminas B2 (esquerda) e B6 (direita) nos cinco produtos analisados.

Os teores em mg/100g obtidos para B2 e B6, considerando apenas as amostras sem adição no caso das determinações por EEF e com modelo BLLS que apresentou os melhores resultados na maior parte dos produtos, são apresentados na Tabela 14. Uma vez que na análise por HPLC não foram feitas adições, e sim a determinação de três amostras em triplicata em cada produto, a comparação entre os teores obtidos é feita através de um teste-t para comparação das médias obtidas em cada método. Pelos valores de t calculados e tabelados apresentados na Tabela 15, pode-se concluir que para a vitamina B2 apenas no produto leite 1 os resultados dos métodos EEF e HPLC são equivalentes, sendo os resultados de ambos os métodos significativamente maiores que os atestados no rótulo pelo fabricante, considerando uma variação de 10 % no valor nominal. Para a vitamina B6 os resultados entre os métodos EEF e HPLC são equivalentes em quatro dos cinco produtos, sendo eles leite 1 e leite 2 e os suplementos 1 e 2. Para a vitamina B6, nos leites 1 e 2 foram obtidos valores inferiores que os atestados no rótulo, enquanto que um valor superior foi obtido para o suplemento 1 e concordantes para leite 3 com EEF e suplemento 2 com tanto EEF e HPLC.

**Tabela 14.** Resultado dos teores das vitaminas B2 e B6 em mg/100g em cada produto, obtidos por EEF com o modelo BLLS, HPLC e os teores nominais atestados no rótulo.

Vit.	Leite 1	Leite 2	Leite 3	Suplemento 1	Suplemento 2	Método
B2	0,90(±0,09)	0,78(±0,08)	1,20(±0,12)	1,33(±0,13)	1,90(±0,19)	Nominal <sup>a</sup>
	0,93(±0,03)	0,95(±0,02)	1,53(±0,02)	1,9(±0,1)	2,94(±0,04)	EEF <sup>b</sup>
	0,93(±0,00)	0,99(±0,00)	1,46(±0,01)	1,50(±0,04)	2,53(±0,04)	HPLC <sup>b</sup>
B6	0,52(±0,05)	0,39(±0,04)	0,96(±0,1)	2,86(±0,3)	2,40(±0,24)	Nominal <sup>a</sup>
	0,26(±0,03)	0,33(±0,02)	0,94(±0,02)	3,6(±0,2)	2,0(±0,3)	EEF <sup>b</sup>
	0,23(±0,02)	0,31(±0,01)	0,74(±0,02)	3,9(±0,1)	2,2(±0,1)	HPLC <sup>b</sup>

Entre parênteses: (a) intervalo considerando uma variação de 10 % no teor nominal, (b) Desvios padrões com 6 graus de liberdade na maior parte dos produtos.

**Tabela 15.** Valores de t calculados e tabelados, com 99 % de confiança, para a comparação dos teores médios obtidos por EEF e HPLC.

Vitamina	Leite 1	Leite 2	Leite 3	Suplemento 1	Suplemento 2	Valor de t
B2	0,00	4,84	7,67	9,10	15,88	Calculado
	3,707 (6)	3,707(6)	3,055(12)	3,055 (12)	3,169 (10)	Tabelado
B6	1,98	2,09	17,32	3,10	1,55	Calculado
	3,106(11)	3,169(10)	3,055 (12)	3,169 (10)	3,169 (10)	Tabelado

Número de graus de liberdade apresentado entre parênteses.

### 3.3 Conclusões do capítulo

Os resultados apresentados neste capítulo demonstram a aplicação dos modelos de calibração de segunda ordem utilizados em dados de EEF em um sistema bastante complexo. Uma vez que em quatro das seis situações ( 1 para B2 e 5 para B6) em que os modelos de calibração de segunda ordem foram aplicados os resultados entre EEF e HPLC foram equivalentes, pode-se concluir que os modelos foram eficientes em estimar as concentrações dos analitos de interesse e os métodos de extração utilizados no preparo das amostras para EEF e HPLC não

influenciaram os resultados de forma significativa. Nas outras situações em que os resultados não foram concordantes devido à presença de erros sistemáticos, existe a possibilidade de erros tanto por parte dos modelos como as diferenças entre os métodos de preparo de amostra para EEF e HPLC.

Entre os modelos de calibração de segunda ordem abordados, o BLLS foi o que apresentou os melhores resultados para quatro dos cinco produtos analisados, podendo assim ser considerado o mais eficiente no estudo destes dados. Sendo também o modelo que é mais facilmente construído por não necessitar de inicialização, restrições ou otimização do número de fatores na etapa de calibração.

Quanto aos limites propostos para a escolha do número de fatores no RBL com o BLLS, apenas para o produto suplemento 2 houve uma discordância, sendo o limite obtido com a estimativa do ruído instrumental possibilitando a escolha correta do número de fatores em todos os produtos. Para os demais casos, foi observado que para a grande maioria das replicatas ambos os limites possibilitam a determinação correta do número de fatores.

Pode-se concluir que os resultados para os testes de recuperação foram aceitáveis, uma vez que, por exemplo para o BLLS, a ocorrência de erros aproximadamente iguais aos obtidos nas amostras de calibração (2,2 µg/L) era capaz de produzir altos erros nas recuperações obtidas devido aos pequenos incrementos das concentrações. Sendo também os erros de recuperação condizentes com o fato do intervalo de concentrações utilizadas estarem próximos aos limites de detecção dos métodos.

Os valores de seletividade estimados para o UPLS foram próximos aos obtidos nos outros modelos, reforçando a validade da proposta de determinação desse parâmetro para o UPLS.

Os valores das incertezas médias obtidas foram próximos nos três modelos utilizados. A maior diferença foi observada para o suplemento 1 com o UPLS

Capítulo 3 – Determinação de riboflavina e piridoxina em leite em pó e suplementos alimentares.

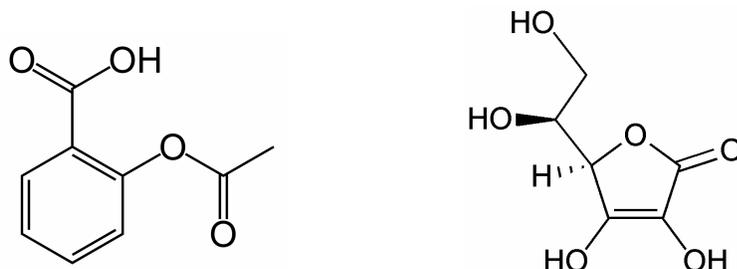
(equação 77) e o suplemento 2 com o PARAFAC para B2 e B6, respectivamente, sendo em ambos os casos os resultados maiores que as outras estimativas.

Como no capítulo anterior, a análise dos dados nos ambientes Matlab 6.5 e Octave 3.0.0 fornecem os mesmos resultados até o quarto algarismo significativo.

**Capítulo 4 - Determinação simultânea de ácido acetilsalicílico e ascórbico em medicamentos**

A terceira aplicação desenvolvida nesse trabalho consistiu na construção de métodos empregando calibração de segunda ordem por BLLS e UPLS em um sistema onde os analitos de interesse apresentassem mais de uma espécie em equilíbrio. Isto faz com que as matrizes de dados tenham deficiência de posto. Isto é, embora  $F$  espécies químicas estejam presentes no meio reacional a correlação ou dependência entre duas ou mais espécies faz com que o número de colunas ou linhas independentes na matriz de dados seja menor que  $F^{109}$ .

Como discutido no capítulo 2, o BLLS foi adaptado por Goicoechea e Olivieri<sup>49</sup> para lidar com sistemas apresentando espécies em equilíbrio, enquanto que segundo Olivieri<sup>65</sup> a aplicação do UPLS juntamente com o RBL nesse tipo de dados é possível sem nenhuma modificação no algoritmo, sendo um exemplo da aplicação desses modelos a determinação de ácido ascórbico<sup>49,65</sup> em sucos de laranja e de corantes em sucos em pó<sup>50</sup>. Contudo, o número de trabalhos estudando esse tipo de sistemas ainda é pequeno para confirmar a eficácia desses modelos. Logo, os objetivos principais da aplicação que é descrita nesse capítulo foram o desenvolvimento de um método para a determinação simultânea dos ácidos acetilsalicílico (AAS) e ascórbico (AA) (cujas estruturas moleculares são apresentadas na Figura 48) em medicamentos utilizando a técnica de FIA com gradiente de pH e detecção no ultravioleta e a comparação dos resultados obtidos por BLLS, UPLS e MCR, dessa forma tentando contribuir para a verificação da eficiência no estudo desses sistemas por meio desses modelos.



**Figura 48-** Estruturas moleculares dos ácidos acetilsalicílico (AAS) e ascórbico (AA).

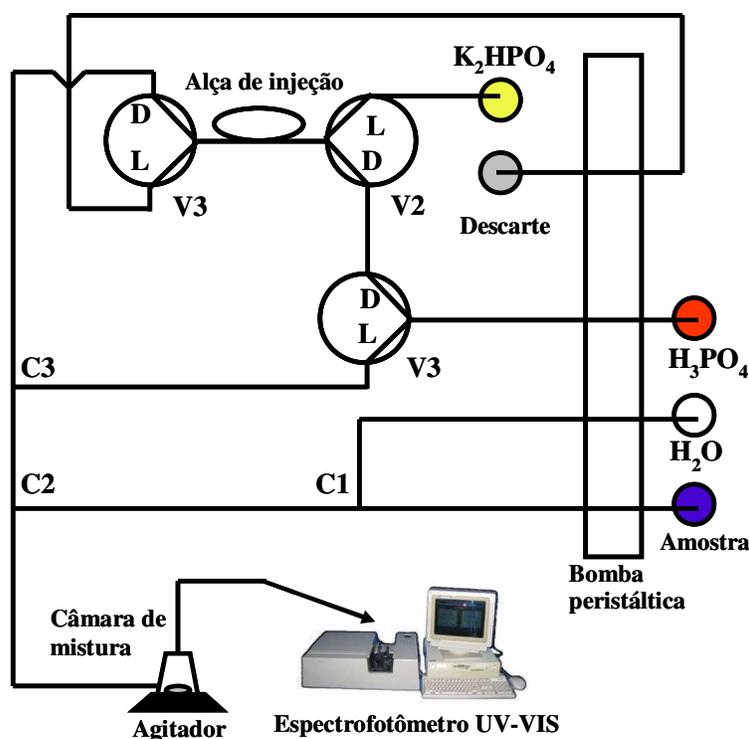
## **4.1 Procedimento experimental**

### **4.1.1 Construção do sistema FIA**

Como já descrito anteriormente, para a aplicação dos modelos de calibração de segunda ordem é necessária a aquisição de uma matriz de dados por amostra. Em um sistema FIA a aquisição dessas matrizes pode ser feita monitorando espectros ao longo do tempo. Contudo, a dimensão temporal deve apresentar informações que diferenciem os compostos de interesse entre si e de possíveis interferentes. Isso foi feito provocando um gradiente de pH no qual estivessem contidos os pKa's dos fármacos estudados. Para a construção desse sistema foram utilizadas 3 válvulas solenóide de três vias da marca N-Research que operam a 12 Volts e consomem 3 Watts; uma bomba peristáltica Ismatec IPC com tubos de Tygon; tubos de poli(tetrafluoroetileno) (PTFE) com 0,8 mm de diâmetro interno (1/32 polegadas); uma fonte de 12 Volts com potência de 24 Watts; um suporte; junções em “T” e uma câmara de mistura como volume de 0,90 mL, constituídas de acrílico; uma barra magnética e um agitador magnético comum. A saída do fluxo do sistema, representado na Figura 49, foi acoplada a uma célula de fluxo com 1 cm de caminho óptico e os espectros obtidos em função do tempo em um espectrofotômetro HP 8452 equipado com detector de arranjo de diodos, com resolução de 2 nm.

O sistema foi montado de modo que a amostra era introduzida continuamente através da cela de fluxo sofrendo uma diluição prévia constante de aproximadamente 11 vezes. O gradiente de pH foi gerado injetando uma solução aquosa de  $K_2HPO_4$  (0,05 mol/L) através de uma alça de injeção de 31,5 cm, equivalente a um volume de 160  $\mu$ L, e a utilização de uma solução aquosa de  $H_3PO_4$  de concentração 0,01 mol/L como carregador. No ponto C1 do sistema (Figura 49) há uma confluência do ácido com a amostra diluída, que em seguida se

conflui em C2 com o seguimento que contém a solução de  $K_2HPO_4$  após a injeção, provocando o gradiente de pH. Como existem dois fármacos constituídos de ácidos fracos monoproticos com  $pK_a$ 's dentro do gradiente de pH gerado pelo sistema FIA ( $pK_a$  de 4,2 para AA e 3,5 para AAS<sup>110</sup>), têm-se as espécies na forma desprotonada ( $pH \gg pK_a$ ) e na forma protonada ( $pK_a \gg pH$ ) dentro do intervalo do gradiente. Como os espectros UV da forma molecular protonada e desprotonada são diferentes, essa variação do espectro com o pH possibilita a utilização de modelos de calibração de segunda ordem.



**Figura 49-** Esquema do sistema FIA construído. Onde L e D indicam saída do fluxo com as válvulas ligadas ou desligadas, respectivamente e C1, C2 e C2 são pontos de confluência.

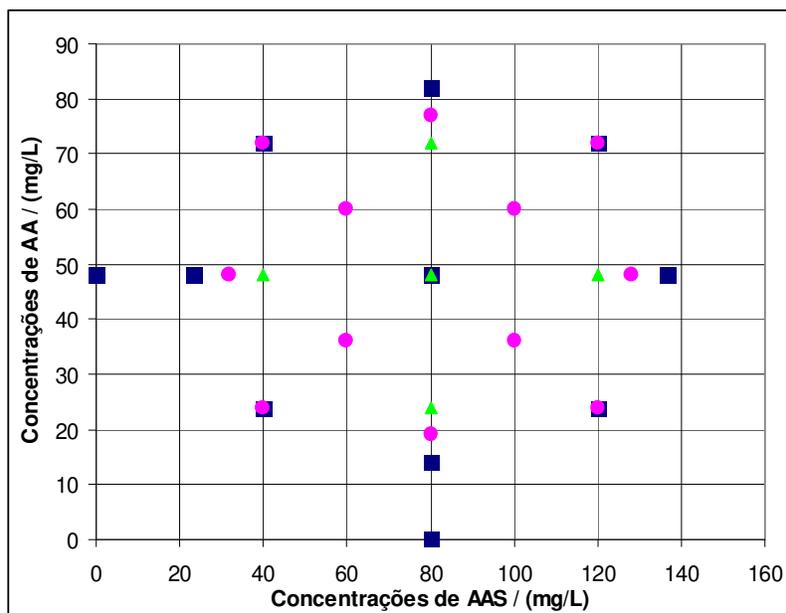
O controle das válvulas do sistema mostrado na Figura 48 é unificado, de forma que todas são ligadas ou desligadas ao mesmo tempo. Inicialmente preenche-se o sistema com os fluidos correspondentes a cada segmento do FIA. Quando as válvulas estão desligadas, a solução de  $H_3PO_4$  está passando pela alça de injeção. Assim que as válvulas são ligadas, o fluxo da solução de  $H_3PO_4$  é desviado pela

válvula solenóide 3 e por sucção as válvulas solenóides 1 e 2 permitem o preenchimento da alça de injeção com a solução de  $K_2HPO_4$ . Após as válvulas serem desligadas, o volume de solução da alça é então injetado no sistema. A dispersão do  $K_2HPO_4$  no  $H_3PO_4$  gera um gradiente de pH suave no fluxo com uma duração de aproximadamente de 135 segundos.

A fim de garantir que o intervalo de pH do gradiente fosse adequado para os compostos de interesse, o gradiente de pH do sistema foi verificado utilizando 11 soluções tampão em uma faixa de pH de 2,0 até 7,0. As soluções tampão foram alimentadas no caminho com maior vazão do sistema FIA (originalmente o que continha água para a diluição da amostra) e uma solução de indicadores no caminho de menor vazão (originalmente o da amostra) e água nos outros caminhos. As soluções tampão utilizadas consistiram de misturas de ácido cítrico e  $Na_2HPO_4$  enquanto que a solução de indicadores foi composta de uma solução aquosa de alaranjado de metila ( $3 \times 10^{-4}$  mol/L) e verde de bromocresol ( $1 \times 10^{-4}$  mol/L). Um modelo de calibração multivariada baseada no PLS foi construído entre os valores de pH das soluções tampão e os espectros na região do visível (400 a 650 nm) obtidos pela solução na saída do FIA. Após isso, o sistema foi preenchido com as soluções transportadoras utilizadas na análise ( $H_3PO_4$ ,  $K_2HPO_4$  e água) utilizando ainda a solução de indicadores no caminho de menor vazão e foi injetada a solução de  $K_2HPO_4$ . O espectrofotômetro monitorou o gradiente no fluxo através do seu programa de cinética, capturando 1 espectro por segundo durante 300 segundos. Através do modelo de calibração multivariada desenvolvido os espectros obtidos a cada segundo após a injeção do  $K_2HPO_4$  foram relacionados aos pH's<sup>111</sup>, de forma que foi verificado que o intervalo de pH alcançado no gradiente foi entre 2,0 e 7,0, permitindo o estudo das espécies ácidas (forma protonada) e básicas (forma desprotonada) dos analitos de interesse.

#### 4.1.2 Planejamento experimental para preparo das soluções padrão de calibração e misturas de validação.

Para construção e validação dos modelos de calibração quatro conjuntos de amostras foram preparadas, sendo: 11 padrões para a calibração (CAL); 12 misturas para validação que possuem apenas os compostos de interesse (VAL1); 5 misturas para validação nas quais foi adicionada cafeína a uma concentração constante (atuando como interferente não calibrado)(VAL2) e 24 amostras de medicamentos comerciais, sendo que em 12 amostras foram feitas adições dos analitos em três níveis distintos para testes de recuperação. A Figura 50 apresenta a disposição das amostras CAL, VAL1 e VAL2 no planejamento realizado, onde a concentração apresentada na Figura 50 refere-se à concentração da amostra antes de ser alimentada no sistema, ou seja, antes da etapa de diluição no sistema FIA. A escolha dos níveis de concentração foi feita por testes preliminares no sistema FIA proposto, utilizando soluções contendo apenas os analitos de interesse e observando a intensidade do espectro de absorção no UV.



**Figura 50-** Gráfico do planejamento experimental utilizado. (■) soluções padrão para a calibração (CAL), (●) misturas para validação sem interferente (VAL1) e (▲) misturas para validação com interferente (VAL2).

As soluções dos conjuntos CAL, VAL1 e VAL2 foram preparadas a partir de soluções estoque de AAS e AA com concentrações de 400 e 240 mg/L, respectivamente, que foram preparadas utilizando reagentes com grau de pureza analítica: ácido L(+) ascórbico (99,75 %, ECIBRA) e ácido acetilsalicílico (99,9 %, Synth) e material volumétrico calibrado. Para as soluções do conjunto VAL2, foi utilizada uma solução estoque de cafeína de 19,5 mg/L (99,9 %, Synth), sendo adicionada nas misturas desse conjunto de maneira que a concentração de cafeína em cada amostra fosse 4 mg/L. Essa concentração foi escolhida com base na proporção entre as concentrações de AAS e a cafeína presentes em um dos medicamentos analisados. De acordo com a propagação das incertezas nas massas e vidraria utilizada, o maior erro relativo teoricamente cometido na preparação das amostras é de 0,2 %, que corresponde ao nível de menor concentração para AA, o que implica que a incerteza no preparo das soluções deve ser insignificante comparada à incerteza final dos resultados.

É conhecido que AA oxida-se facilmente em solução<sup>112</sup>, de forma que devem ser tomados cuidados especiais desde a preparação dos padrões até sua análise. Logo, foi feito um estudo da estabilidade do AA durante 3 horas, variando a temperatura das soluções e a presença ou ausência de oxigênio dissolvido na água. Os resultados mostraram que a temperatura é o fator mais importante, sendo que após 2 horas do preparo da solução não foi observada decomposição significativa, dentro de uma incerteza de 3 %, na solução mantida à temperatura de 10 °C, enquanto que a 25 °C observou-se mais que 20 % de decomposição do AA presente em solução. A presença de oxigênio dissolvido mostrou-se significativa, mas em menor intensidade. Com base nesses testes, todos os padrões e amostras foram preparados utilizando água (Milli-Q, Millipore) à aproximadamente 10 °C (mantida por um banho de gelo) e saturada com nitrogênio.

### **4.1.3 Preparo das amostras de medicamentos comerciais para a análise no sistema FIA.**

Os medicamentos analisados foram:

- Medicamento 1: comprimido efervescente sabor limão, com teores nominais dos princípios ativos em cada comprimido de 400 mg/comp. de AAS e 240 mg/comp. de AA e o excipiente composto de carbonato ácido de sódio, ácido cítrico, ciclamato de sódio, sacarina sódica e um aromatizante não especificado pelo fabricante.
- Medicamento 2: comprimido efervescente sabor laranja, com teores nominais dos princípios ativos em cada comprimido de 400 mg/comp. de AAS e 240 mg/comp. de AA e excipientes não definidos na embalagem do medicamento.
- Medicamento 3: comprimido com teores nominais dos princípios ativos em cada comprimido de 500 mg/comp. de AAS, não contém AA e 30 mg/comp. de cafeína. O excipiente é composto de amido e corante CI nº 45.430 em quantidades não especificadas.
- Medicamento 4: comprimido efervescente, com teores nominais dos princípios ativos em cada comprimido de 1000 mg/comp. de AA, não contém AAS e 250 mg/comp. de cálcio ionizável na forma de lactobionato de cálcio e carbonato de cálcio e os excipientes bicarbonato de sódio, ácido cítrico, 1,8 g/comp. de sacarose, aromatizante, sacarina e polietilenoglicol.

No preparo das soluções de cada medicamento foi medida a massa de 10 comprimidos e determinada a massa média em cada caso. Em seguida, esses comprimidos foram macerados em cadinho cerâmico e homogeneizados. Foram preparados 500 mL de solução adicionando a massa necessária do medicamento para que fosse produzida uma solução com as concentrações dos analitos de interesse AAS e AA iguais a aproximadamente 160 e 94 mg/L, respectivamente (considerando os teores nominais presentes no rótulo), o que corresponde a uma

concentração duas vezes maior que a solução padrão com a concentração no centro de intervalo de concentrações definido nas soluções de calibração. Essa solução foi diluída em balões de 100 mL, de forma que a concentração das amostras em que não houveram adições ficasse no centro da curva de calibração.

Para cada medicamento foram preparadas 6 amostras, sendo que em três delas foram feitas adições dos analitos em 3 diferentes níveis a fim de verificar índices de recuperação. Essas adições foram feitas a partir da adição de volumes adequados das soluções estoque a 50 mL da solução dos medicamentos em que os analitos estavam com concentração 160 e 94 mg/L para AAS e AA, respectivamente, e completando o volume até 100 mL com água deionizada. As adições foram de 10, 25 e 60 % (percentagens relativas à concentração dos analitos em 80 e 47 mg/L para AAS e AA, respectivamente).

Para solubilização dos medicamentos, o medicamento 3 apresentou necessidade de solubilização em 10 mL de etanol (grau HPLC, Merck), pois este medicamento, por não ser efervescente, apresentou problemas na solubilização do AAS que tem baixa solubilidade em água. A quantidade de etanol utilizada não influenciou as medidas, tendo em vista que a quantidade de etanol adicionado é apenas 2 % do volume da solução de 500 mL inicialmente preparada e que a amostra era diluída várias vezes dentro do sistema FIA. Os outros medicamentos, todos efervescentes, não necessitaram de solubilização prévia em etanol. Na Tabela 16 é mostrado o valor das concentrações finais obtidas levando em conta o teor nominal de cada medicamento. Portanto, foram preparadas 6 amostras de cada medicamento e todas determinadas em triplicata no sistema FIA.

#### **4.1.4 Validação da determinação dos teores de AAS e AA em medicamentos.**

A validação dos teores estimados para os analitos em cada medicamento foi feita através de outro método que foi desenvolvido utilizando a técnica de HPLC

para AAS, enquanto que a determinação de AA foi feita de acordo com o método padrão por titulação com iodo<sup>113</sup>.

**Tabela 16.** Concentrações em mg/L de AAS e AA nas amostras dos quatro medicamentos analisados.

	Sem adição		Adição de 10 %		Adição de 25 %		Adição de 60 %	
	AAS	AA	AAS	AA	AAS	AA	AAS	AA
Medicamento 1	80,4	48,2	88,2	52,9	100,6	60,3	128,5	77,1
Medicamento 2	80,5	28,3	88,3	53,1	100,7	60,6	128,7	77,6
Medicamento 3	80,3	##	88,1	##	100,4	##	128,4	##
Medicamento 4	##	48,4	##	53,2	##	60,5	##	77,1

(#) Analitos ausentes.

#### 4.1.4.1 Determinação de AAS por HPLC.

Para a realização das medidas instrumentais foi utilizado HPLC-DAD da marca Shimadzu Prominence, equipado com bomba modelo LC20AT, detector por arranjo de diodos modelo APD-M20A e amostrador automático modelo SL20A. Os dados foram obtidos e exportados com o programa LCSolution (versão 1.21), os cromatogramas foram adquiridos a uma taxa de 0,64 por segundo, monitorando o intervalo de 190 a 400 nm. Para a separação foi utilizada uma coluna C18 Microsorb MV (VARIAN) (250 mm de comprimento por 4,6 mm de diâmetro interno) com partículas de 5 µm.

Os solventes para as preparações e análises cromatográficas foram acetonitrila (grau-HPLC, Tedia), água (Milli-Q, Millipore) e ácido fosfórico (Merck, 85 %), previamente filtrados utilizando membranas de fluoreto de poli(vinilideno) (Millipore). As separações foram realizadas com fase móvel composta por acetonitrila:água 15:85 (v/v), sendo a água acidificada para pH = 3,0 com ácido fosfórico. Foi utilizada vazão de 1,0 mL/min e eluição isocrática.

Foram preparadas três amostras de cada medicamento seguindo um

procedimento muito similar ao utilizado no FIA, onde a concentração final dos analitos na amostra foi de aproximadamente 32,0 e 19,0 mg/L para AAS e AA, respectivamente. A curva de analítica foi composta de cinco pontos igualmente espaçados compreendendo o intervalo de 10,00 a 50,00 mg/L e 6,00 a 30,00 mg/L para AAS e AA, respectivamente. Todas as injeções das amostras e padrões foram feitas em triplicata.

#### 4.1.4.2 *Determinação de ácido ascórbico (AA) por Iodimetria<sup>113</sup>.*

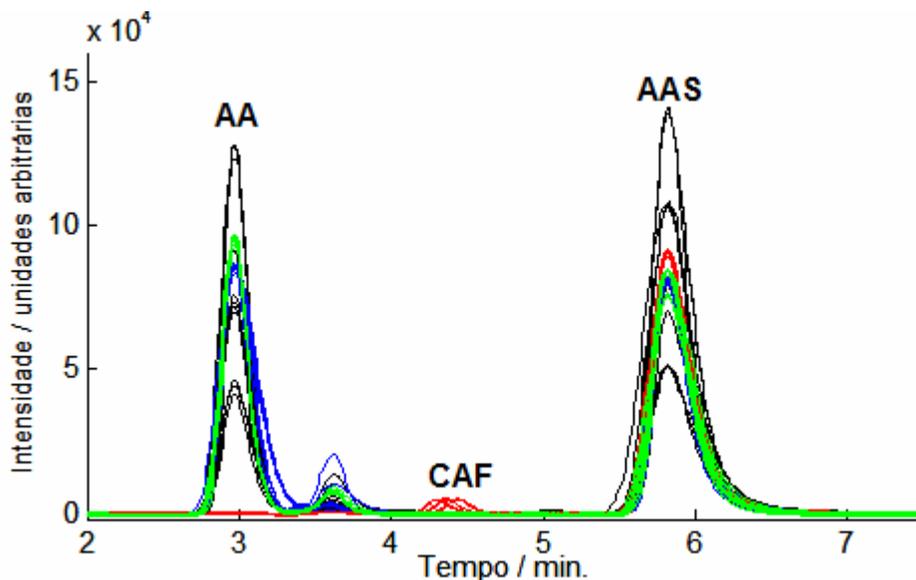
A padronização da solução de iodo utilizada nas determinações foi feita através da titulação de três alíquotas de 100,00 mg de AA padrão. Primeiramente, o AA foi dissolvido em 25,00 mL de água, adicionados 7,0 mL de H<sub>2</sub>SO<sub>4</sub> 4 mol/L e 1,0 mL de solução de amido. Em seguida, a solução de iodo foi adicionada gradativamente até o aparecimento da coloração azul (característica do complexo da goma de amido com o iodo) que persistisse por mais de 30 segundos.

O teor de AA dos medicamentos 1, 2 e 4 foi determinado em triplicata pela titulação de três alíquotas de massas de cada medicamento que deveriam conter aproximadamente 100,00 mg de AA. Foi utilizado o mesmo procedimento descrito para a padronização da solução de iodo, sendo a massa de AA padrão substituída pela massa correspondente de cada medicamento.

## **4.2 Resultados e discussão**

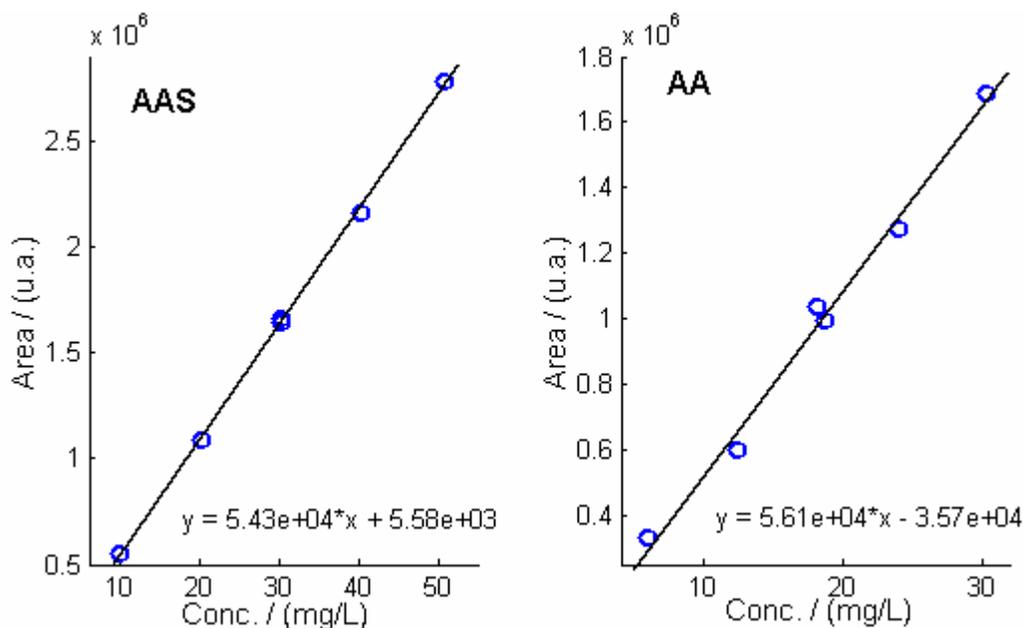
Os cromatogramas obtidos para validação das análises dos medicamentos são apresentados na Figura 51, onde se nota ótima separação dos picos, tempo de análise relativamente curto e pequenos picos que são atribuídos a interferentes dos medicamentos, entre os quais foi possível identificar a cafeína presente no medicamento 3. Para AA, observa-se a presença de pequenos picos de interferentes próximos ao seu pico. Segundo o recurso de medida de pureza de pico (disponível

no programa do equipamento), esses picos oferecem alguma interferência para algumas amostras.



**Figura 51-** Cromatogramas obtidos por HPLC para as análises de AAS e AA monitorados em 232 nm. (preto) padrões e medicamentos (azul) 1, (verde) 2 e (vermelho) 3.

Na Figura 52 pode-se observar as curvas analíticas obtidas para AAS e AA, que apresentam coeficientes de correlação iguais a 0,9996 e 0,9910, respectivamente. O ajuste inferior para AA em relação ao AAS é atribuído à alta taxa de decomposição que este possui quando em solução e pela possível ação de interferentes identificados pelo software do equipamento, os valores de RMSEC e REP para as regressões foram 0,25 e 0,75; 0,8 e 6,3% para AAS e AA, respectivamente. Quanto aos resultados obtidos para AA por iodimetria<sup>113</sup> comparados aos resultados obtidos pela regressão por HPLC, foi observada uma tendência sistemática nos resultados, sendo o resultado por HPLC sempre superior e com maior erro em relação ao valor nominal. Por essa razão, para AA os resultados obtidos por iodimetria foram considerados os valores de referência para a validação das medidas por FIA.

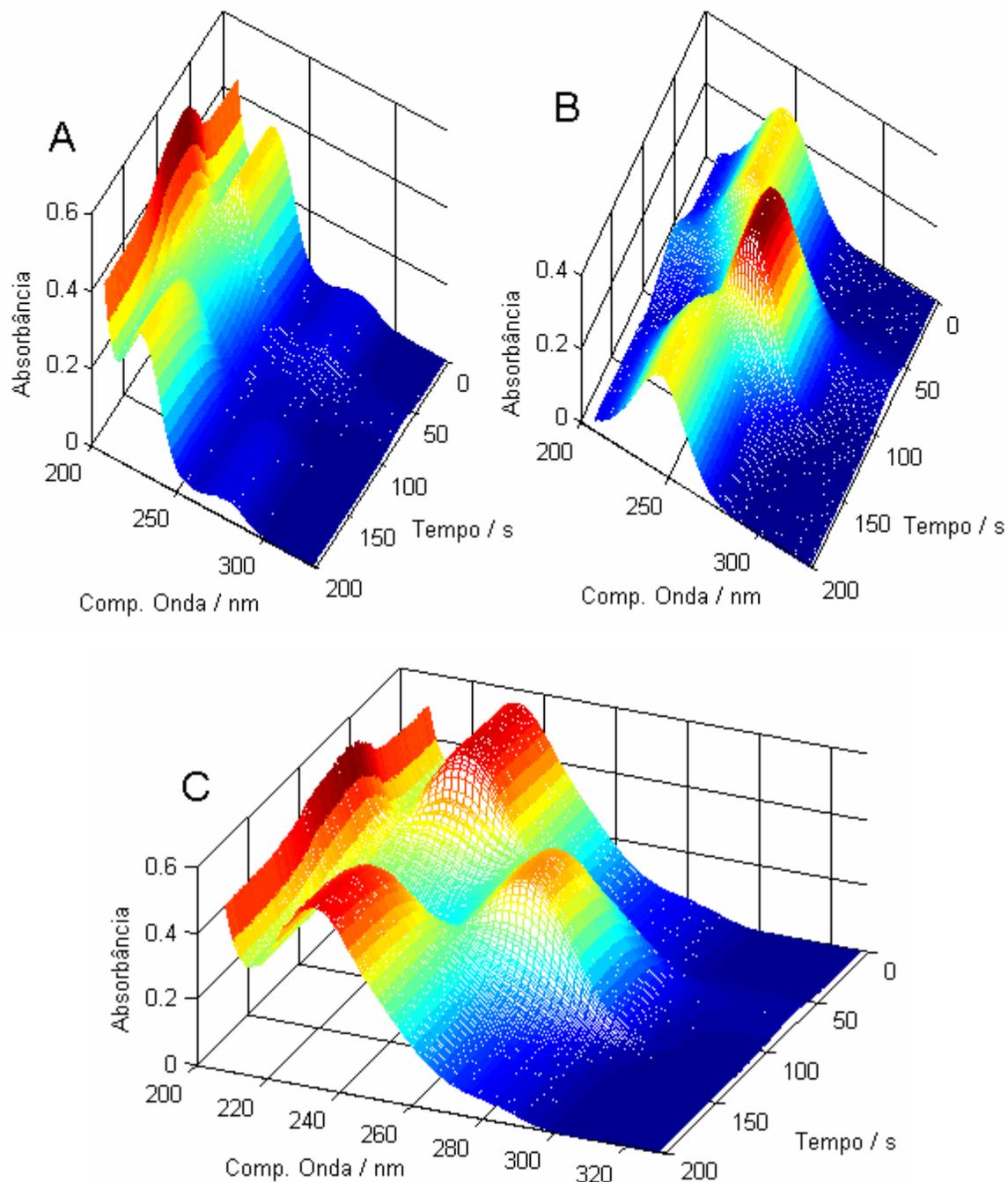


**Figura 52-** Curvas analíticas obtidas a partir das medidas em HPLC.

As análises realizadas no sistema FIA apresentaram tempo total de 10 minutos, incluindo as etapas de limpeza e preenchimento do sistema com uma nova amostra. A Figura 53 mostra as superfícies obtidas nos primeiros 200 segundos e a região espectral que foi utilizada para a construção dos modelos de calibração. Nestas Figuras pode-se perceber nitidamente a mudança dos espectros UV dos analitos devido ao gradiente de pH, que é caracterizada para AAS principalmente pelo desaparecimento de uma banda com máximo em 232 nm na transição da forma ácida (HA) para a forma básica ( $A^-$ ). Para AA, na transição de HA para  $A^-$  observa-se principalmente um deslocamento de uma banda centrada em 245 nm para 267 nm.

Para a construção dos modelos por BLLS, UPLS e MCR foram utilizados intervalos espectrais de 208 a 328 nm, enquanto que na dimensão temporal utilizaram-se os tempos de 94 a 150 segundos para BLLS e UPLS e de 1 a 291 para o MCR. Nos modelos BLLS e UPLS o único pré-processamento utilizado foi centrar os dados na média, enquanto que no MCR o melhores resultados foram obtidos pela imposição das restrições de não negatividade nas dimensões espectrais

e temporais, normalização dos espectros para comprimento unitário, trilinearidade e a inicialização do algoritmo pelas estimativas espectrais obtidas pelo método de pureza de variáveis<sup>40</sup>.



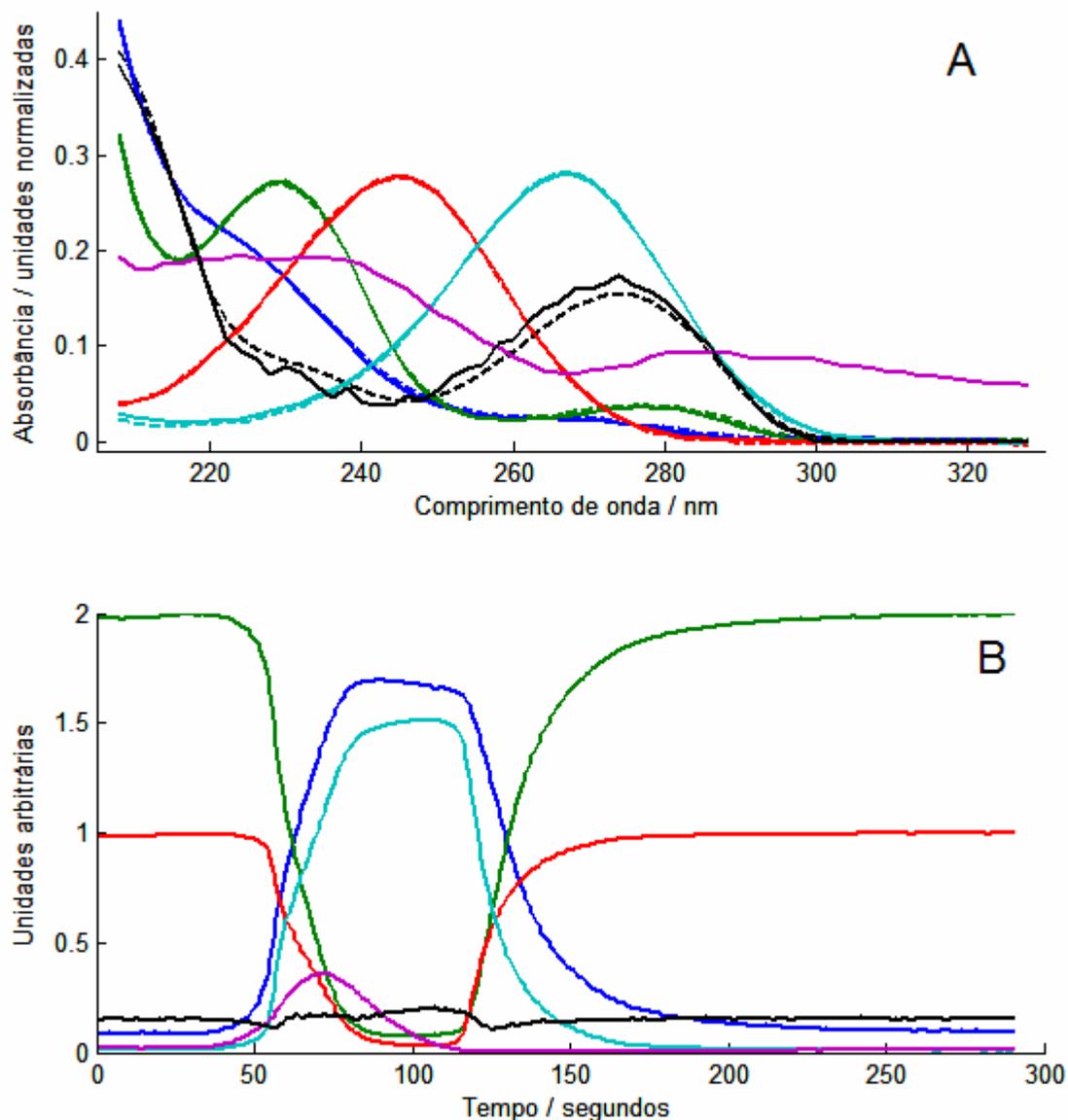
**Figura 53-** Superfícies obtidas no sistema FIA para (A) AAS, (B) AA e (C) uma mistura contendo os dois analitos nas concentrações de 80 e 48 mg/L para AAS e AA, respectivamente.

Primeiramente passando à discussão dos resultados obtidos por MCR, o qual

é o modelo de calibração de segunda ordem mais amplamente utilizado na análise desse tipo de dados. Apesar de neste caso os espectros experimentais de todas as espécies serem conhecidas e disponíveis, as estimativas iniciais foram obtidas pelo método de pureza de variáveis para executar o MCR como se estas não fossem disponíveis. Durante a estimativa, observou-se era possível que a obtenção de um perfil para um quinto fator contendo informação não aleatória. Portanto, o MCR foi executado com 4 e 5 fatores, para as amostras de calibração e um fator a mais para modelar interferentes no conjunto VAL2 e as amostras dos medicamentos. Para o fator referente ao interferente não foi utilizado o método de pureza de variáveis para a obtenção do perfil inicial, sendo inserido apenas um vetor contendo apenas o valor 0,1. Este procedimento foi adotado devido a falta de praticidade da necessidade de sempre estimar os perfil do interferente em cada nova amostra que é analisada no MCR em uma situação real de rotina. Logo o perfil do interferente é fruto apenas da otimização realizada pelo MCR.

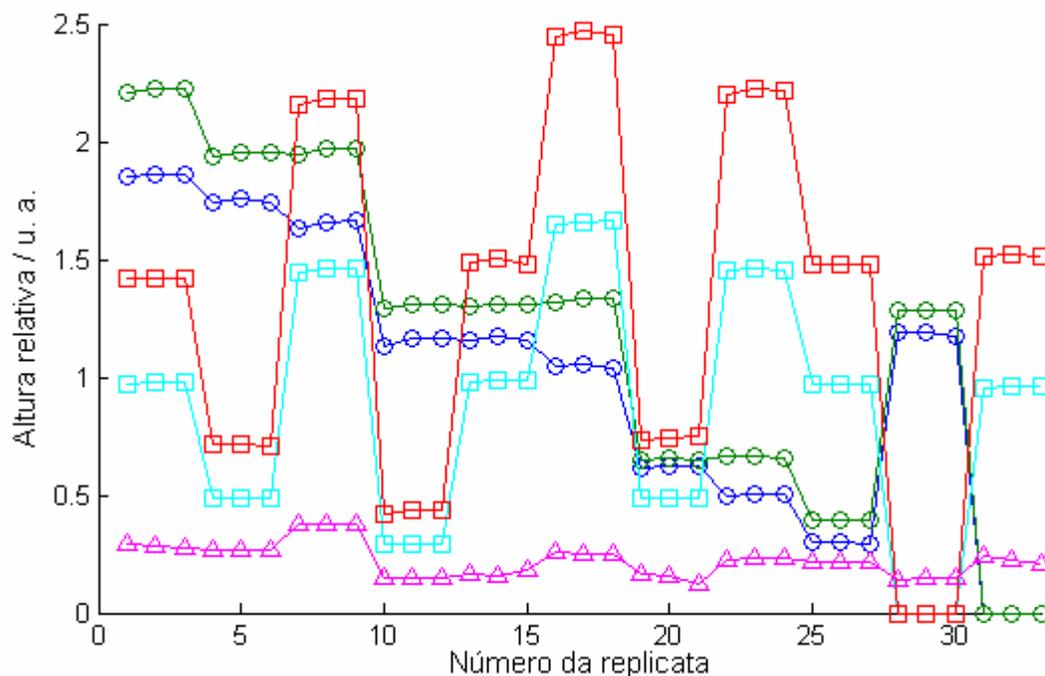
Foi observado que os resultados quantitativos obtidos com 4 e 5 fatores foram praticamente os mesmos; contudo, os perfis de concentração recuperados com 5 fatores apresentaram melhor concordância com os perfis esperados teoricamente. Na Figura 54 são apresentados os perfis recuperados pelo MCR com 5 fatores em uma amostra do conjunto VAL2, o qual possui como interferente a cafeína adicionada. Observa-se que o MCR consegue recuperar os espectros das espécies ácidas e básicas de AAS e AA e do perfil correspondente à cafeína com ótima concordância com os espectros obtidos experimentalmente em soluções isoladas de cada espécie. Quanto aos perfis de concentração observa-se também uma ótima concordância com os perfis esperados teoricamente, sendo obtidas curvas que representam a formação e o desaparecimento de cada uma das espécies dos dois analitos durante todo o tempo de análise. Para a cafeína, observa-se um perfil aproximadamente constante, que é coerente com o fato deste composto não

apresentar variações na sua estrutura devido à mudança do pH do meio.



**Figura 54-** Estimativas dos perfis espectrais (A) e de concentração (B) obtidas por MCR para uma amostra do conjunto VAL2. (contínuo) perfis recuperados com MCR, (tracejado) perfis experimentais, (—) AAS, forma ácida; (—) AAS, forma básica; (—) AA, forma ácida; (—) AA, forma básica; (—) efeito Schlieren, (—) interferente (cafeína).

Quanto ao quinto fator que foi adicionado na análise das amostras de calibração, que aparece em cor rosa na Figura 54, este é atribuído ao sinal devido à variação do índice de refração causado pela confluência da amostra no carregador ácido e da solução básica injetada, fenômeno físico denominado efeito Schlieren<sup>114,115</sup>. O fato dos resultados quantitativos não serem diferentes quando o MCR é executado com ou sem um fator para modelar esse efeito é explicado pela sua baixa intensidade que foi observada nessa aplicação, a qual se deve às boas condições de mistura, a alta dispersão (ambas as características adquiridas com a utilização da câmara de mistura) e pela pequena diferença de concentrações entre a solução carregadora e a base injetada para formar o gradiente. A Figura 55 mostra as alturas dos máximos de cada um dos perfis dos analitos e do efeito Schlieren obtidos nos padrões de calibração. Observa-se que as alturas para este efeito são pequenas em relação às dos demais analitos, o que reforça que o efeito Schlieren representa apenas uma pequena fração do sinal analítico nesta aplicação.



**Figura 55-** Alturas dos perfis de concentração obtidos por MCR nas 33 replicatas das soluções de calibração. (—●—) AAS, forma ácida; (—●—) AAS, forma básica; (—■—) AA, forma ácida; (—■—) AA, forma básica e (—▲—) efeito Schlieren.

É interessante notar que o MCR foi capaz de estimar o perfil de concentração desse fenômeno físico, mesmo este sendo pequeno em relação aos demais sinais presentes e estar localizado em uma região de transição de outros 4 perfis (Figura 54B). Uma possível explicação para essa ótima resolução é a existência de uma região seletiva no espectro do efeito Schlieren entre 305 e 330 nm (Figura 54A). A resolução observada pelo MCR neste caso sugere que em outras aplicações onde o efeito Schlieren seja significativo em termos quantitativos o MCR pode ser uma maneira eficiente em sua correção.

A Tabela 17 apresenta os resultados quantitativos obtidos com o MCR utilizando cinco fatores na análise do conjunto VAL1 e seis nos demais, realizando as regressões com base nas alturas dos perfis de concentração recuperados pelo MCR. Comparando os resultados obtidos com as formas ácidas (HA) e básicas ( $A^-$ ) observa-se que as formas ácidas dos dois analitos fornecem menores erros (RMSEP e REP) em relação aos valores de referência e valores mais precisos. Isso é justificado pelo formato dos perfis ácidos e básicos, as formas ácidas apresentam no início e no final de seu perfil patamares estáveis e sujeitos a menor variação, seja ela causada por alguma variação nas injeções ou do gradiente de pH. Para AA observa-se uma ótima concordância entre os valores obtidos pelo método por FIA e MCR e os obtidos por iodimetria, sendo 3,2 o maior erro relativo obtido para esse analito considerando a forma HA. Para AAS também foi observada boa concordância entre os métodos, com exceção do medicamento 2, no qual o método FIA apresenta um erro de quase 10 % em relação aos valores obtidos por HPLC. A Figura 56 mostra os gráficos dos valores de referência versus os estimados pelo método FIA com MCR, onde se pode observar que no caso do AAS no medicamento 2 os valores estimados por MCR apresentam um erro sistemático positivo, que é responsável pelo maior erro observado nesse medicamento.

**Tabela 17.** Resultados das figuras de mérito obtidas para determinação de AAS e AA, estimados comparando os resultados do método FIA com MCR e os valores de referência por HPLC e iodimetria para AAS e AA, respectivamente.

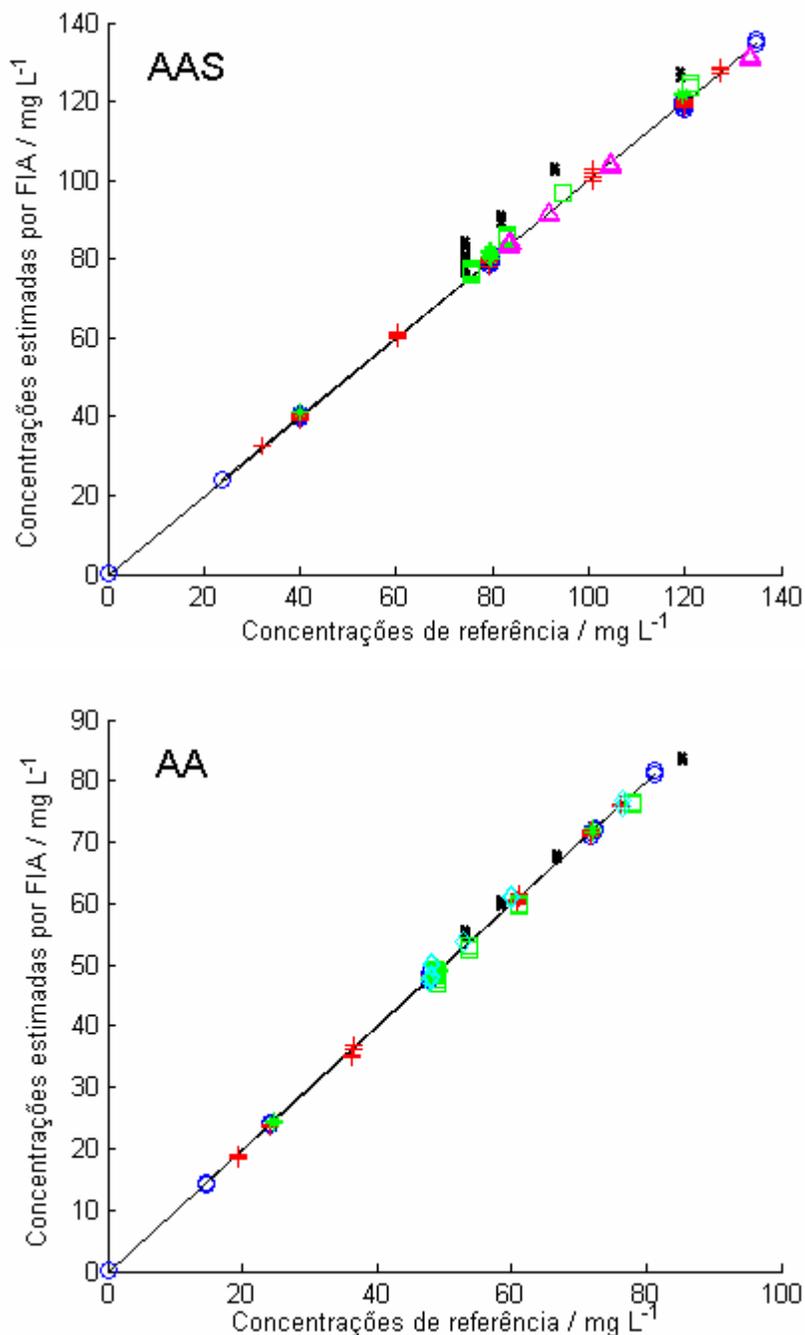
AAS								
Dados	HA				A <sup>-</sup>			
	RMSEP <sup>a</sup>	REP <sup>b</sup>	Precisão <sup>a</sup>	LD <sup>a</sup>	RMSEP <sup>a</sup>	REP <sup>b</sup>	Precisão <sup>a</sup>	LD <sup>a</sup>
VAL1	0,9	1,3	0,55	0,85	3,0	4,2	0,85	0,97
VAL2	1,8	2,2	0,48	0,87	2,9	3,6	1,17	1,00
Medicamento 1	2,0	2,2	0,57	0,89	17	19	1,84	1,21
Medicamento 2	7,9	9,3	0,83	0,93	23	26	0,76	1,03
Medicamento 3	1,0	0,8	0,34	0,90	13	15	0,53	1,00

AA								
Dados	HA				A <sup>-</sup>			
	RMSEP <sup>a</sup>	REP <sup>b</sup>	Precisão <sup>a</sup>	LD <sup>a</sup>	RMSEP <sup>a</sup>	REP <sup>b</sup>	Precisão <sup>a</sup>	LD <sup>a</sup>
VAL1	0,51	1,5	0,29	0,70	0,88	1,7	0,44	0,46
VAL2	0,34	1,0	0,31	0,70	0,95	2,2	0,52	0,47
Medicamento 1	1,12	1,9	0,27	0,69	1,3	2,7	0,43	0,48
Medicamento 2	1,84	3,2	0,42	0,69	3,0	5,4	0,31	0,48
Medicamento 4	1,03	2,0	0,28	0,67	3,8	7,3	0,57	0,47

Grandezas expressas em: (a)mg/L e (b) percentagem.

Quanto à sensibilidade do método desenvolvido por MCR observa-se que não há uma grande diferença entre os resultados dos LD entre as formas ácidas e básicas, para ambos analitos. Pela Tabela 17, os LD para as espécies ácidas estão em aproximadamente 0,9 e 0,7 mg/L para AAS e AA, respectivamente. Contudo, deve-se lembrar que para a construção das regressões não foi considerada a diluição em linha de aproximadamente 11 vezes existente no sistema, o que implica que os valores de LD devem ser na verdade 11 vezes menores que os apresentados na Tabela 17, iguais a 0,08 e 0,06, respectivamente para AAS e AA.

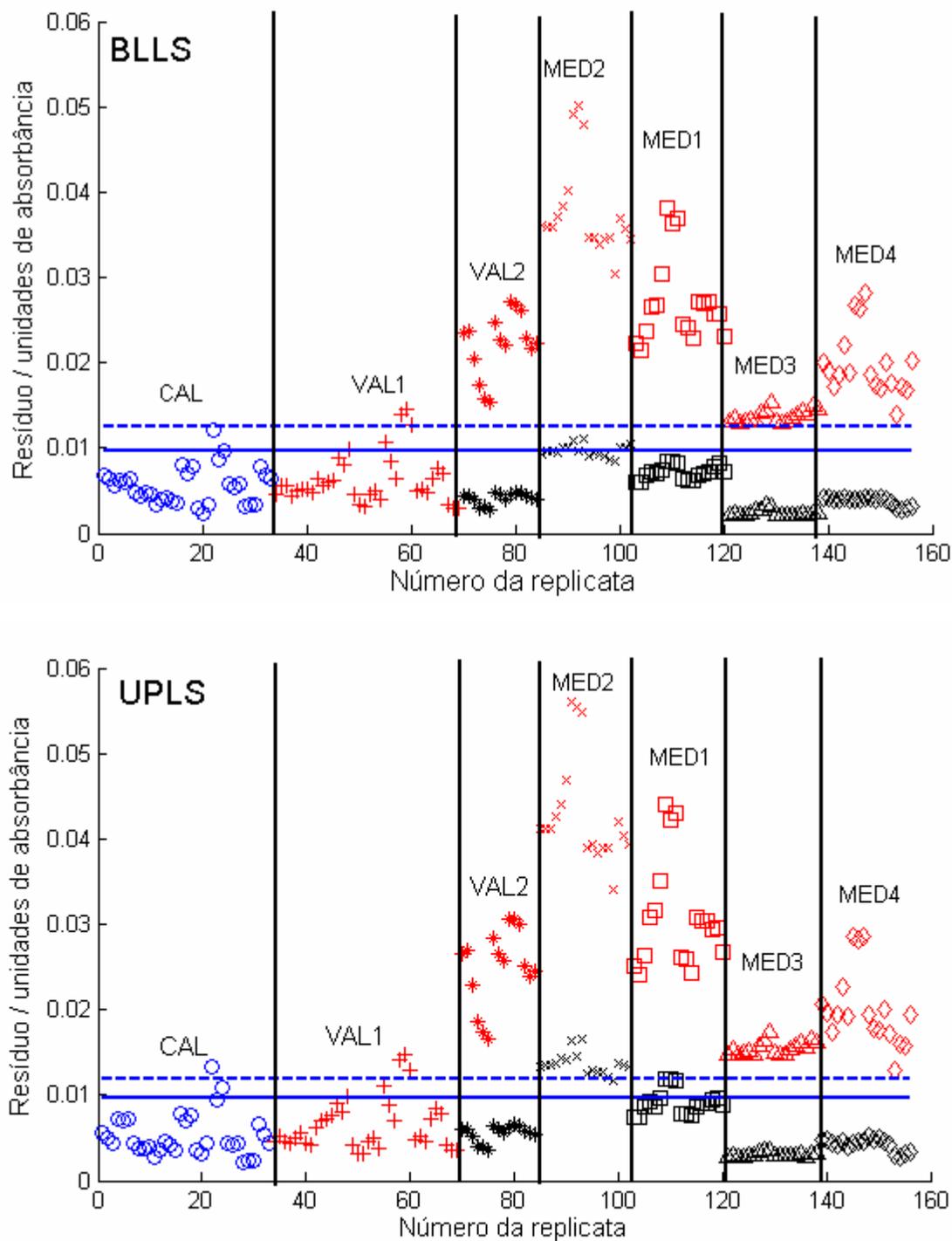


**Figura 56-** Valores de referência versus os estimados por FIA utilizando o MCR em todos os conjuntos de dados analisados. (—) Bissetriz do quadrante, (○) CAL, (+) VAL1, (\*) VAL2 e Medicamentos (□) 1, (x) 2, (△) 3 e (◇) 4.

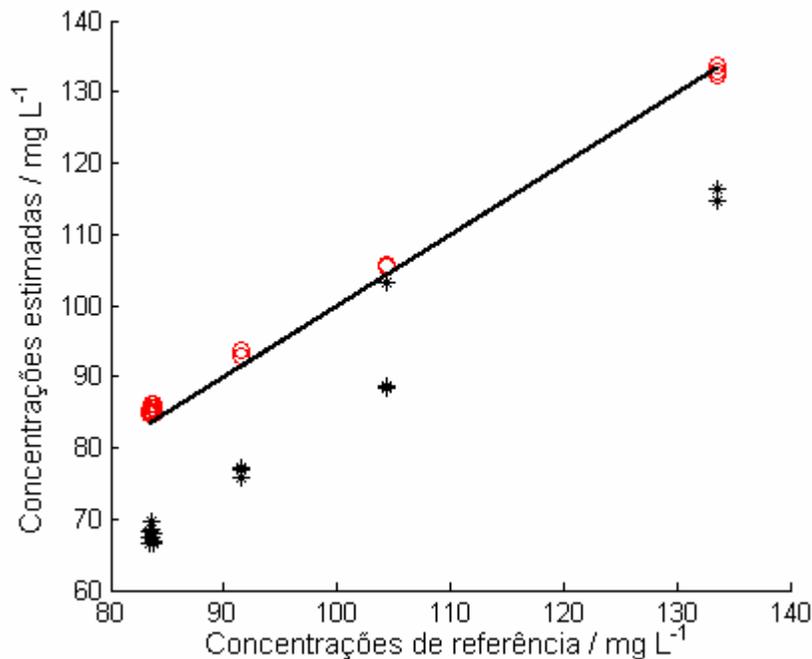
Passando à discussão dos resultados dos outros dois modelos, a Figura 57 mostra os gráficos de resíduos para ambos BLLS e UPLS em todos os conjuntos de dados analisados. Para o BLLS observa-se que, com exceção de algumas replicatas

do conjunto VAL1, o limite baseado no resíduo médio das amostras de calibração indica que não é necessária a utilização do RBL, enquanto que para o conjunto VAL2 e em todos os medicamentos sugere a necessidade de um fator no RBL, o que é coerente com a presença da cafeína adicionada em VAL2 e com a interferência devido aos excipientes de cada medicamento. Para a determinação de AA os melhores resultados são obtidos com um fator no RBL nos medicamentos e em VAL2, como sugerem os resíduos. Contudo, pela comparação entre os valores estimados por BLLS e os obtidos por HPLC para AAS no medicamento 3 (mostrada na Figura 57), observou-se que com um fator no RBL tem-se um alto erro sistemático negativo em 17 das 18 replicatas. Logo, neste caso os limites propostos não permitem a escolha correta do número de fatores no RBL. O comportamento dos resíduos para esse medicamento ressalta que a escolha deste parâmetro no modelo BLLS é muito importante e não é simples como sugerida nos artigos até o momento publicados na literatura<sup>49,50,52,53</sup>, pois, os medicamentos 3 e 4 apresentam resíduos de decomposição muito próximos, mas apenas o medicamento 4 requer a adição de um fator no RBL. Outro fato que reforça essa importância é que os valores de concentração estimados para AAS no medicamento 3 com um fator no RBL não chegam a ser absurdos, logo, em uma aplicação real de rotina onde não se tem conhecimento do valor de referência, o analista poderia ser levado a crer que as concentrações estimadas com um fator são as mais corretas, o que não seria verdade.

Para o modelo UPLS observam-se resíduos com aspecto semelhante aos obtidos com o BLLS. Contudo, neste caso, os limites sugerem a necessidade de dois fatores no RBL para o medicamento 2. Para a determinação de AA os melhores resultados foram obtidos com apenas um fator no RBL, como foi observado no BLLS. Logo, a única inconsistência entre os limites e o número de fatores do RBL neste analito ocorre para o medicamento 2.

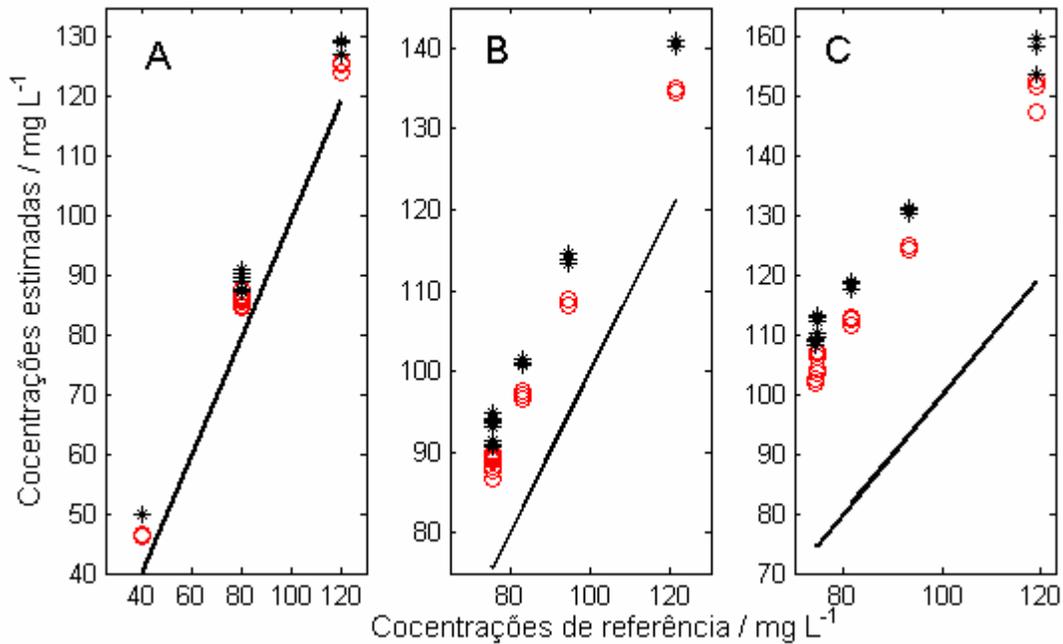


**Figura 57-** Resíduos de decomposição obtidos com os modelos BLS e UPLS na análise de cada conjunto de dados. Onde: (o) CAL, (+) VAL1, (\*) VAL2, Medicamentos (MED) (□) 1, (x) 2, (Δ) 3 e (◇) 4, (vermelho) Nenhum e (preto) um fator no RBL. (—) Valor crítico pelo ruído instrumental e (---) pelo resíduo médio das amostras de calibração, ambos com 99% de confiança.



**Figura 58-** Concentrações de referência versus estimadas pelo método FIA com BLLS no medicamento 3. (—) Bissetriz do quadrante, resultados obtidos com (o) nenhum e (\*) um fator no RBL.

Para a determinação de AAS com o UPLS, ao contrário do que sugerem os resíduos, os melhores resultados são obtidos com nenhum fator no RBL para o conjunto de amostras VAL2 e os medicamentos 1 e 2. A Figura 59 mostra esse comportamento, evidenciando que a diferença entre as estimativas para a concentração é pequena nesses conjuntos de dados, mas é melhor sem utilizar o RBL. Observa-se também que para essas amostras são obtidos erros sistemáticos positivos significativos, demonstrando que o UPLS não é eficiente em estimar a concentração de AAS nessas amostras. A Tabela 18 apresenta o número de fatores escolhido para o RBL na construção dos modelos BLLS e UPLS que conduziram à obtenção de menores erros em relação aos valores e referência e foram considerados para a determinação das figuras de mérito de cada modelo.



**Figura 59-** Gráfico das concentrações de referência versus estimadas, obtidos com UPLS para AAS nos conjuntos de dados: (A) VAL2, Medicamentos (B) 1 e (C) 2. Onde: (—) Bissotrizolol do quadrante, resultados obtidos com (o) nenhum e (\*) um fator no RBL.

**Tabela 18.** Número de fatores utilizados no RBL nos modelos BLLS e RBL que apresentam os menores erros em relação aos valores de referência.

Dados	BLLS		UPLS	
	AAS	AA	AAS	AA
VAL1	0	0	0	0
VAL2	1	1	0	1
Medicamento 1	1	1	0	1
Medicamento 2	1	1	0	1
Medicamento 3	0	0	1	1
Medicamento 4	1	1	1	1

Como já comentado anteriormente, os programas desenvolvidos nesta tese para a construção dos modelos BLLS e UPLS foram executados nos ambientes Matlab 6.5 e Octave 3.0.0. Para os dados obtidos por FIA, ao contrário do que foi observado nos resultados obtidos para os dados de EEF discutidos nos dois

capítulos anteriores, o modelo BLLS não apresentou o mesmo resultados nos dois ambientes em todos os casos, sendo observada uma diferença significativa na determinação de AAS no medicamento 2, o que pode ser observado na Tabela 19 que mostra os resultados de RMSEP e REP obtidos nos dois ambientes. A principal razão para as diferenças observadas na Tabela 19 é a utilização de um programa de otimização para a minimização da matriz de resíduos no algoritmo do RBL, que é distinto para os dois sistemas. Observa-se que os resultados médios são equivalentes entre os sistemas, com exceção da determinação de AAS para o medicamento 2 onde as concentrações estimadas pelo Octave apresentam melhor resultado. Isto demonstra que o sistema livre Octave é capaz de fornecer resultados numéricos equivalentes, ou até superiores como neste caso, aos obtidos com sistema comercial Matlab, que é o sistema mais utilizado nos grupos de pesquisa em quimiometria.

Pela Tabela 19, observa-se que os melhores resultados para AAS e AA são obtidos com a forma ácida e básica, respectivamente, diferente do MCR que utiliza as duas formas ácidas. Considerando essas espécies e os resultados obtidos com o Octave como os melhores resultados com o modelo BLLS, observam-se que os maiores erros médios obtidos foram de aproximadamente 4 mg/L e 6 % para RMSEP e REP, respectivamente, nos conjuntos VAL2 e o medicamento 2. A dispersão dos resultados para BLLS pode ser observada na Figura 60, sendo melhor para AAS comparando com os resultados obtidos por MCR (Figura 56). Contudo, também foi observado na Figura 60 que foram obtidos erros sistemáticos negativos para AA com o BLLS para o medicamento 1 e 2, onde o MCR tem melhor resultado para esse analito. Quanto aos valores de precisão, observou-se uma variação de 0,5 a 0,9 mg/L para AAS e 0,3 a 0,8 mg/L para AA, dependendo do conjunto de dados.

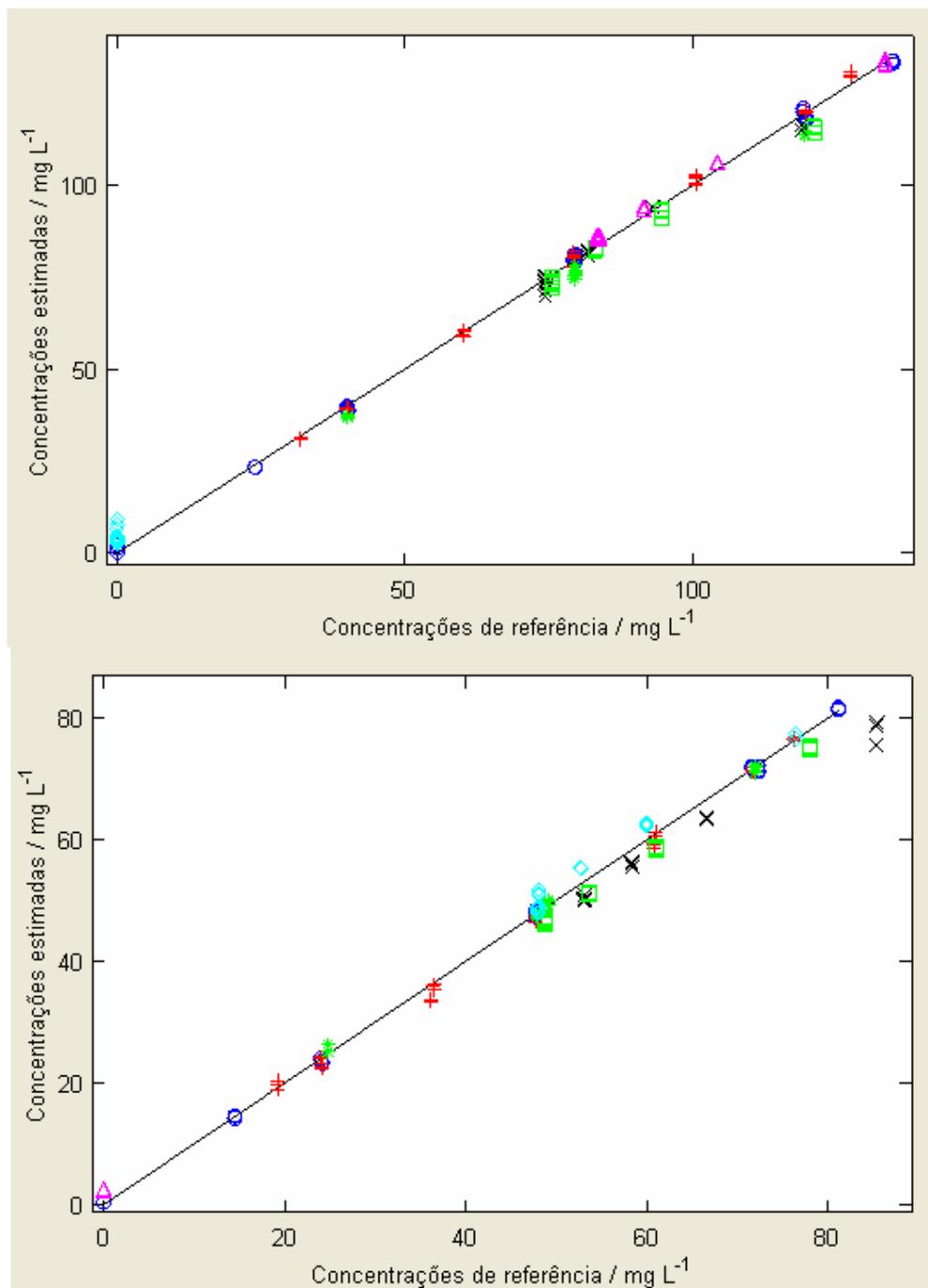
**Tabela 19.** Resultados de RMSEP e REP obtidos para o modelo BLLS nos sistemas Octave 3.0.0 e Matlab 6.5 para a determinação de AAS e AA.

MATLAB								
Dados	AAS				AA			
	HÁ		A <sup>-</sup>		HA		A <sup>-</sup>	
	RMSEP <sup>a</sup>	REP <sup>b</sup>						
VAL1	1,3	2,0	1,3	2,0	1,5	3,8	1,1	3,0
VAL2	3,0	4,2	11	17	2,8	5,5	0,7	2,1
Medicamento 1	3,5	4,0	18	22	7,6	13	2,4	4,3
Medicamento 2	8,5	9,4	38	45	10	17	4,0	5,7
Medicamento 3	1,6	1,8	7,0	7,8	---	---	---	--
Medicamento 4	---	---	---	---	1,6	2,8	2,1	4,0

OCTAVE								
Dados	AAS				AA			
	HA		A <sup>-</sup>		HA		A <sup>-</sup>	
	RMSEP <sup>a</sup>	REP <sup>b</sup>						
VAL1	1,3	2,0	1,3	2,0	1,5	3,8	1,1	3,0
VAL2	4,1	5,5	11	17	2,2	4,2	0,7	2,1
Medicamento 1	3,2	3,2	19	22	6,9	12	2,4	4,4
Medicamento 2	2,4	2,8	37	45	9,1	15	4,1	5,9
Medicamento 3	1,6	1,8	7,0	7,8	---	---	---	--
Medicamento 4	---	---	---	--	1,5	2,7	2,1	4,0

Grandezas expressas em: (a) mg/L e (b) percentagem.



**Figura 60-** Valores de referência versus os estimados pelo método FIA com o BLS executado no sistema Octave para AAS (acima) e AA (abaixo). (—) Bissetriz do quadrante, (○) CAL, (+) VAL1, (\*) VAL2, Medicamentos (□) 1, (x) 2, (△) 3 e (◇) 4.

A Tabela 20 que mostra os resultados de SENA, SEL e LD obtidos com BLS e Octave, onde pode-se observar que a forma ácida para AAS apresenta valores de seletividade significativamente maiores que a forma básica, o que está

de acordo com os melhores resultados obtidos para essa forma. Quanto aos valores de SENA e LD, observa-se que a forma ácida do AAS é menos sensível que a forma básica. Contudo, esse resultado foi contrário ao que seria esperado intuitivamente, uma vez que o espectro da forma ácida é mais característico e intenso. Uma hipótese para esse resultado é que o efeito da sobreposição das quatro espécies e do efeito Schlieren presente nos dados não pode ser interpretado apenas pela visualização dos perfis espectrais e de concentração, os quais são apresentados na Figura 61. Em contraste com os perfis do MCR, o BLLS não consegue recuperar os perfis espectrais e de concentração dos analitos, assim como para os perfis da cafeína e do efeito Schlieren que estão misturados no perfil do interferente.

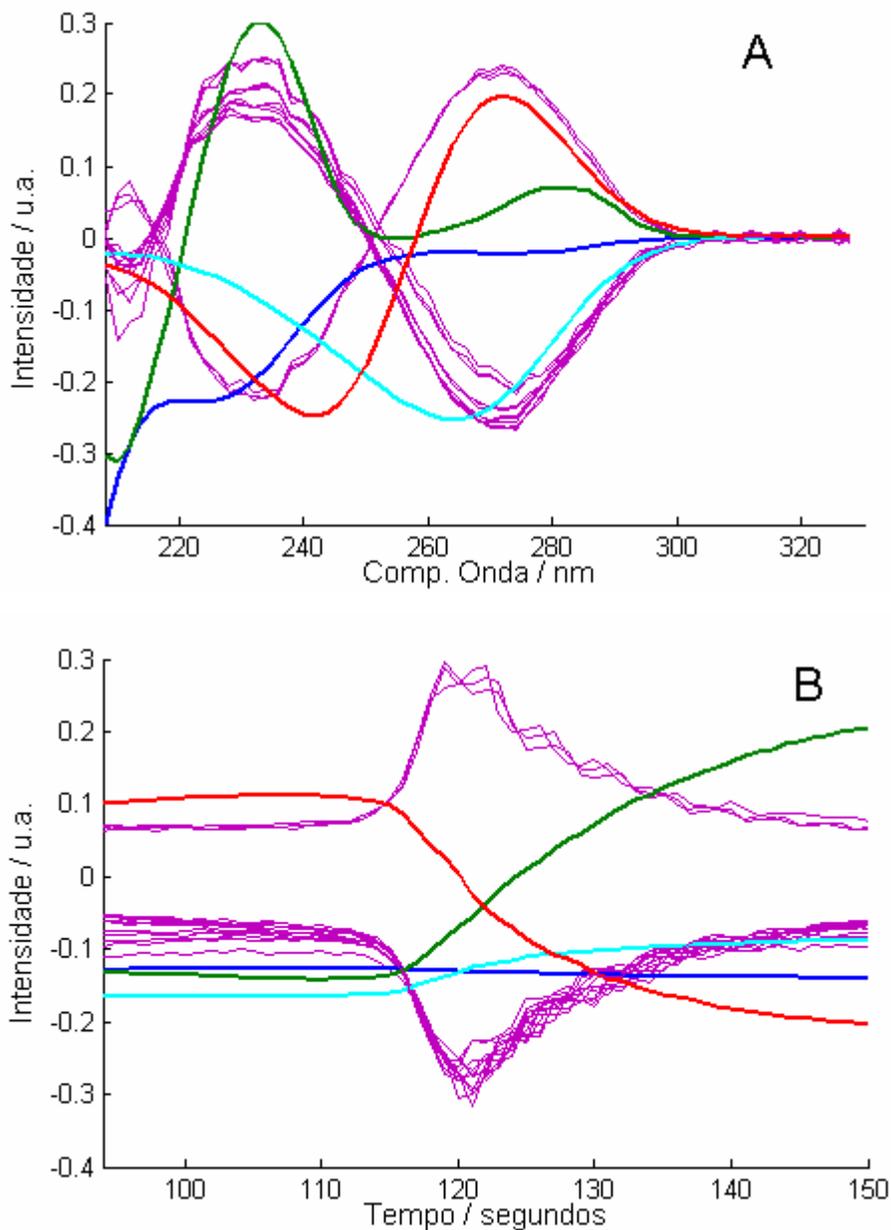
**Tabela 20.** Resultados de SENA, SEL e LD obtidos como modelo BLLS com o sistema Octave 3.0.0.

AAS						
Dados	HA			A <sup>-</sup>		
	(SENA) <sup>-1 a</sup>	SEL <sup>b</sup>	LD <sup>a</sup>	(SENA) <sup>-1 a</sup>	SEL <sup>b</sup>	LD <sup>a</sup>
VAL1	0,32	0,95	1,1	0,04	0,94	0,13
VAL2	0,35(0,03)	0,87(0,06)	1,2(0,1)	0,10(0,01)	0,37(0,03)	0,33(0,03)
Medicamento 1	0,35(0,01)	0,86(0,02)	1,20(0,03)	0,09(0,01)	0,43(0,03)	0,29(0,02)
Medicamento 2	0,35(0,01)	0,88(0,01)	1,20(0,01)	0,09(0,01)	0,40(0,02)	0,33(0,03)
Medicamento 3	0,32	0,95	1,1	0,04	0,94	0,13

AA						
Dados	HA			A <sup>-</sup>		
	(SENA) <sup>-1 a</sup>	SEL <sup>b</sup>	LD <sup>a</sup>	(SENA) <sup>-1 a</sup>	SEL <sup>b</sup>	LD <sup>a</sup>
VAL1	0,08	0,94	0,25	0,024	0,94	0,08
VAL2	0,16(0,01)	0,45(0,04)	0,53(0,04)	0,062(0,004)	0,38(0,02)	0,20(0,01)
Medicamento 1	0,12(0,01)	0,58(0,03)	0,41(0,02)	0,057(0,003)	0,41(0,02)	0,19(0,01)
Medicamento 2	0,12(0,01)	0,59(0,03)	0,53(0,04)	0,064(0,005)	0,37(0,02)	0,20(0,01)
Medicamento 4	0,14(0,01)	0,52(0,02)	0,46(0,02)	0,046(0,002)	0,51(0,02)	0,15(0,01)

Grandezas expressas em: (a) mg/L e (b) sem unidade.



**Figura 61-** Estimativas dos perfis espectrais (A) e de concentração (B) obtidas por BLS para as amostras do conjunto VAL2. (—) AAS, forma ácida; (—) AAS, forma básica; (—) AA, forma ácida; (—) AA, forma básica; (—) perfil do interferente estimado pelo RBL.

As figuras de mérito para o modelo UPLS são apresentadas na Tabela 21, para este modelo os sistemas Matlab e Octave levaram aos mesmos resultados considerando os quatro primeiros algarismos significativos. Observa-se mais claramente que o UPLS não foi eficiente na determinação de AAS, sendo obtidos erros de 10 % ou superiores para o conjunto VAL2 e em todos os medicamentos

analisados. Contudo, para AA foram obtidos resultados semelhantes aos dos modelos MCR e BLLS, sendo a dispersão dos resultados em todos os conjuntos de dados melhor visualizada na Figura 62. Os valores de precisão observados foram semelhantes aos observados com os outros modelos, enquanto que a sensibilidade estimada para o UPLS foi aparentemente melhor. Os valores de seletividade entre os modelos BLLS e UPLS não podem ser comparados neste caso, pois, enquanto o UPLS não leva em conta as espécies em equilíbrio no meio, o BLLS considera 2 espécies para cada analito.

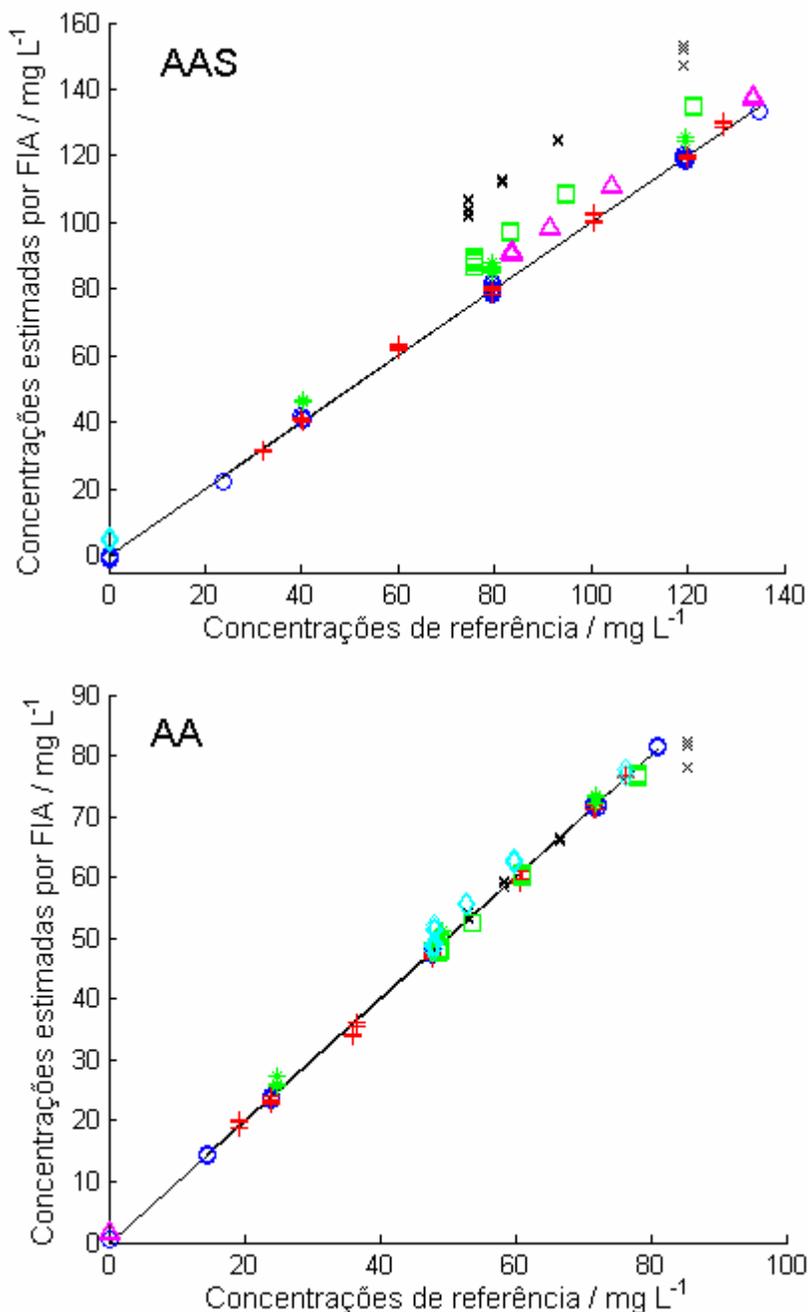
**Tabela 21.** Resultados obtidos para as figuras de mérito com o modelo UPLS em cada conjunto de dados analisado.

AAS						
Conjunto	RMSEP <sup>a</sup>	REP <sup>b</sup>	Precisão <sup>a</sup>	(SENA) <sup>-1 a</sup>	SEL <sup>c</sup>	LD <sup>a</sup>
VAL1	1,38	2,1	0,40	0,04	---	0,13
VAL2	6,20	9,5	0,90	0,04	---	0,13
Medicamento 1	13	15	0,45	0,04	---	0,13
Medicamento 2	31	37	1,26	0,04	---	0,13
Medicamento 3	9,7	11	0,48	0,095(0,002)	0,41(0,1)	0,31(0,01)

AA						
Conjunto	RMSEP <sup>a</sup>	REP <sup>b</sup>	Precisão <sup>a</sup>	(SENA) <sup>-1 a</sup>	SEL <sup>c</sup>	LD <sup>a</sup>
VAL1	0,94	2,6	0,33	0,02	---	0,08
VAL2	1,52	4,0	0,57	0,043(0,002)	0,55(0,04)	0,14(0,01)
Medicamento 1	0,95	1,7	0,27	0,039(0,002)	0,60(0,03)	0,13(0,01)
Medicamento 2	1,29	1,7	0,37	0,039(0,002)	0,61(0,03)	0,13(0,01)
Medicamento 4	2,26	4,3	0,48	0,044(0,002)	0,54(0,03)	0,15(0,01)

Grandezas expressas em: (a) mg/L, (b) percentagem e (c) sem unidade.



**Figura 62-** Valores de referência versus os estimados pelo método FIA com o UPLS executado com o sistema Matlab. (—) Bissetriz do quadrante, (○) CAL, (+) VAL1, (\*) VAL2, medicamentos (□) 1, (x) 2, (△) 3 e (◇) 4.

As Tabelas 22 e 23 apresentam os valores de recuperação para as adições realizadas, onde observa-se que bons resultados foram obtidos com todos os modelos.

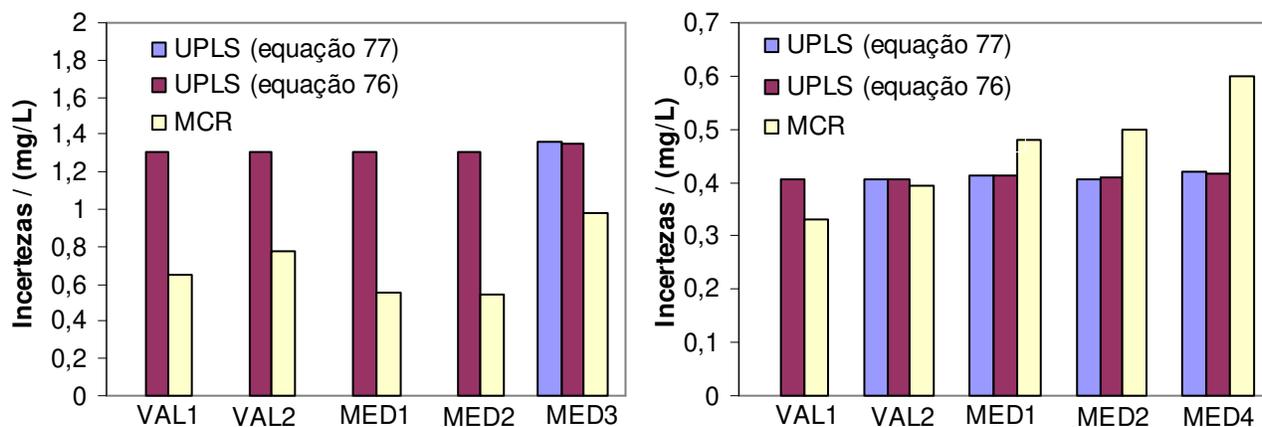
**Tabela 22.** Recuperações percentuais obtidas para as determinações de AAS.

Modelo/adições	Recuperações (%)		
	Medicamento 1	Medicamento 2	Medicamento 3
<b>BLLS</b>			
7,9 mg/L	100(1)	100(1)	102(1)
20,2 mg/L	98(1)	101(1)	101(1)
48,2 mg/L	95(1)	97(2)	100(2)
<b>UPLS</b>			
7,9 mg/L	120(6)	99(8)	107(1)
20,2 mg/L	104(2)	99(1)	106(2)
48,2 mg/L	98(1)	95(6)	103(1)
<b>MCR</b>			
7,9 mg/L	103(2)	111(2)	100,0(0,2)
20,2 mg/L	102(1)	111(1)	99(1)
48,2 mg/L	103(1)	107(1)	98(1)

**Tabela 23.** Recuperações percentuais obtidas para as determinações de AA.

Modelo/adições	Recuperações (%)		
	Medicamento 1	Medicamento 2	Medicamento 4
<b>BLLS</b>			
4,8 mg/L	96(1)	96(1)	105(1)
12,3 mg/L	96(1)	95(1)	104(2)
29,3 mg/L	96(2)	91(2)	100(2)
<b>UPLS</b>			
4,8 mg/L	98(1)	101(1)	105(1)
12,3 mg/L	99(1)	99(1)	105(1)
29,3 mg/L	98(1)	96(1)	101(1)
<b>MCR</b>			
4,8 mg/L	99(1)	103(1)	102(1)
12,3 mg/L	98(1)	101(1)	102(2)
29,3 mg/L	98(1)	98(1)	100(1)

A Figura 63 mostra as incertezas estimadas com UPLS e MCR, onde se observa que em concordância com os erros obtidos, uma vez que o UPLS apresenta maior incerteza para AAS. Nas situações onde houve a ocorrência de interferentes, observa-se que as duas propostas para UPLS são praticamente idênticas, resultado que é reflexo do baixo ruído instrumental estimado como 0,004 unidades de absorbância e pelo efeito que os interferentes provocam na sensibilidade dos modelos não ser muito pronunciado.



**Figura 63-** Incertezas médias obtidas com os modelos MCR e UPLS. (MED) Medicamentos.

Considerando que os melhores resultados quantitativos com o método FIA foram obtidos com o modelo BLLS (executado no sistema Octave), os teores médios em mg/comprimido para a determinação dos analitos em cada medicamento podem ser observados na Tabela 24. Através de um teste-t para comparação de médias conclui-se com 99% de confiança que para AAS o método por FIA apresenta resultados equivalentes aos obtidos por HPLC para os medicamentos 2 e 3 e diferentes para o medicamento 1, sendo os valores de t calculados respectivamente iguais a 1,41; 2,94 e 4,02. Para AA concluí-se que os resultados são equivalentes apenas para o medicamento 4, sendo os valores de t calculados iguais a 7,46; 4,16 e 1,84 para os medicamentos 1, 2 e 4, respectivamente, onde os valores críticos para AAS e AA com 12 e 8 graus de liberdade é 3,055 e 3,355 respectivamente. Contudo, em todos os casos ambos os métodos estão dentro da

variação permitida de 10 % para o teor do ativo desses medicamentos segundo a farmacopéia americana<sup>113,116</sup>.

**Tabela 24.** Resultado dos teores de AAS e AA em mg/comprimido em cada medicamento, obtidos por FIA com o modelo BLLS, HPLC ou iodimetria e os teores nominais atestados no rótulo.

Analitos	Medicamento 1	Medicamento 2	Medicamento 3	Medicamento 4	Método
AAS	400(± 40)	400(± 40)	500(± 50)	###	Nominal <sup>a</sup>
	366(±6)	362(±10)	526(±4)	###	FIA <sup>b</sup>
	377(±3)	368(±3)	520(±3)	###	HPLC <sup>b</sup>
AA	240(±24)	240(±24)	###	1000(±100)	Nominal <sup>a</sup>
	232(±5)	250(±2)	###	1020(±20)	FIA <sup>b</sup>
	243(±3)	264(±3)	###	990(±20)	Iodimetria <sup>b</sup>

(a) Erro de 10 % admitido pela USP. Números de graus de liberdade: (b) 6 e (c) 2.

### 4.3 Conclusões do capítulo

Os resultados apresentados para essa aplicação representam mais um exemplo que comprova a eficiência de modelos de calibração de segunda ordem em dados que apresentam analitos com espécies em equilíbrio.

Os resultados demonstram que o MCR foi capaz de fornecer informações tanto quantitativas quanto qualitativas do sistema em estudo com uma ótima concordância com os valores esperados. Além dos perfis dos analitos e dos interferentes químicos o MCR foi capaz de estimar o perfil do efeito Schlieren, mesmo este sendo relativamente pequeno nesta aplicação. Resultado que sugere que a sua aplicação para a correção deste efeito pode ser possível.

Os gráficos de resíduos para a escolha do número de fatores do RBL possibilitaram a escolha do número de fatores na maior parte das situações para o BLLS, sendo o valor crítico baseado no resíduo médio da calibração o mais coerente nessa aplicação. Contudo, para a determinação de AAS no medicamento 3

os limites induzem à utilização de um fator, enquanto que na realidade sem modelar o interferente tem-se melhores resultados. Isso reforça a importância da escolha correta desse fator e que em determinadas situações os limites propostos não permitem a escolha correta. Para o UPLS observaram-se três situações em que os limites falham na escolha do número correto de fatores, todas com o AAS, demonstrando que nesse modelo provavelmente seja necessário um conjunto de amostras onde se conheça o valor de referência para que a escolha do número de fatores seja feita com segurança.

O sistema Octave 3.0.0 mostrou-se capaz de apresentar resultados para as determinações de AAS e AA com os modelos BLLS e UPLS com a mesma qualidade, ou até superior no caso de AAS no medicamento 2, que as obtidas com o sistema Matlab 6.5. Estes resultados demonstram que a utilização desse sistema livre é possível em grupos de pesquisa situados em instituições que não dispõem do ambiente Matlab, sendo o sistema de programação em linha de comando essencialmente o mesmo nos dois sistemas.

Observou-se nesta aplicação que o modelo UPLS foi menos eficiente para a determinação de AAS que os outros modelos, sendo que o BLLS apresentou os melhores resultados para o AAS, enquanto que o MCR foi melhor para AA. Comparando os resultados obtidos por FIA com os valores de referência obtidos por HPLC e iodimetria para AAS e AA, respectivamente, conclui-se que os resultados são estatisticamente equivalentes na metade das situações, independente dos modelos BLLS ou MCR. Entretanto, mesmos nos casos onde a diferença entre os métodos foi significativa os resultados foram muito próximos, o que justifica concluir que a determinação desses ácidos pelo método FIA utilizando calibração de segunda ordem é possível e viável para fins de análise de rotina quando esses modelos forem reconhecidos por órgãos de fiscalização do setor farmacêutico, como as Farmacopéias.

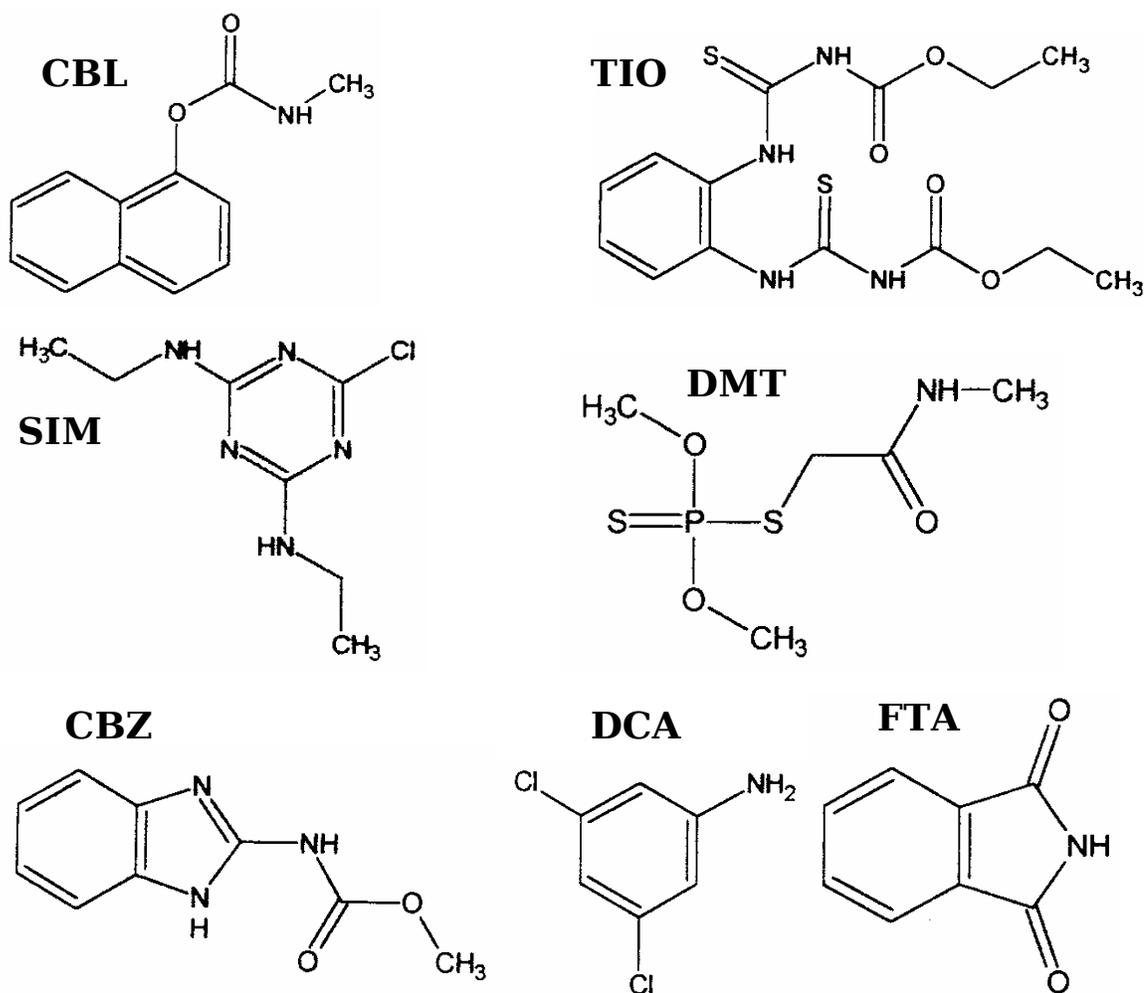
**Capítulo 5 - Determinação de pesticidas e  
metabólitos em amostras de vinho tinto**

A cromatografia líquida de alta eficiência (HPLC) é uma técnica amplamente aplicada em análises de rotina em diversas áreas, como farmacêutica, biológica, ambiental e alimentícia. Na maioria dos casos essa técnica é capaz de proporcionar a separação de todas as espécies de interesse em um sistema. Contudo, quando uma amostra complexa é analisada a sobreposição do sinal das espécies de interesse com interferências da amostra é frequentemente observada. Assim, é necessária a modificação e a otimização das condições cromatográficas de modo a adquirir novamente uma resolução dos compostos, tarefa que demanda considerável trabalho experimental, gasto de reagentes, aumento do tempo de análise, etc.

A utilização de sistemas de detecção multicanal como espectrometria de massas (MS) ou DAD pode reduzir o problema da sobreposição entre o sinal dos analitos e os interferentes da amostra através da aquisição de espectros em função do tempo de retenção, aumentando a probabilidade de encontrar um ponto do espectro que seja seletivo e a possibilidade da utilização de calibração de ordem zero. Contudo, quando HPLC é aplicado com DAD a existência de um comprimento de onda seletivo para o analito pode não ser observada devido às bandas largas de absorção geralmente observadas na região UV. Nessas situações a utilização de modelos de calibração de segunda ordem podem ser empregados devido à capacidade desses modelos em realizar uma separação matemática dos sinais e pela vantagem de segunda ordem.

A última aplicação descrita nesse trabalho teve como objetivo o desenvolvimento de um método alternativo por HPLC-DAD para determinação dos pesticidas e metabólitos em amostras de vinho tinto, onde foram comparadas as eficiências dos modelos PARAFAC, BLLS e UPLS. Foram escolhidos como espécies de interesse os pesticidas carbaril (CBL), tiofanato metílico (TIO), simazina (SIM) e dimetoato (DMT) e os metabólitos carbendazim (CBZ), 3,5-dicloroanilina (DCA) e ftalimida (FTA). A Figura 64 apresenta as estruturas

moleculares dos sete compostos de interesse, os quais são utilizados em defensivos agrícolas geralmente empregados em plantações de uvas e cujo teor de seus resíduos em vinhos é limitado e fiscalizado por órgãos como ANVISA e o *Codex Alimentarius*<sup>117</sup>. Esta aplicação foi desenvolvida em colaboração com as Profas. Dras. Carla B. G. Bottoli e Isabel C. S. F. Jardim, as quais dispunham de padrões de cada um dos compostos de interesse, possibilitaram a utilização de um equipamento HPLC-DAD e já haviam desenvolvido um método cromatográfico que proporcionava boa separação dos compostos em solução, que foi tomado como método de referência para validação do método alternativo desenvolvido.



**Figura 64-** Fórmulas estruturais dos quatro pesticidas (CBL, TIO, SIM e DMT) e dos três metabólitos de interesse que foram determinados nas amostras de vinho.

## 5.1 Procedimento experimental.

### 5.1.1 Reagentes e soluções padrão

Todo o material volumétrico utilizado no preparo das soluções foi calibrado. Os solventes utilizados para a análise cromatográfica e preparação de amostras foram acetonitrila (grau HPLC, Tedia) e água Milli-Q (Millipore), os quais foram filtrados utilizando membranas de fluoreto de poli(vinilideno) com poros de 0,45  $\mu\text{m}$ .

Os padrões de pesticidas foram: CBL (99,8 %, Supelco), SIM (98,3 %, Novartis), TIO (98,5 %, Riedel-de-Häen) e DMT (99,0 %, Riedel-de-Häen); e os metabólitos CBZ (99,1 %, Chem Service), DCA (99,1 %, Riedel-de-Häen) e FTA (99,9 %, Riedel-de-Häen).

Foram preparadas soluções estoque em acetonitrila de cada um dos analitos nas seguintes concentrações: 1046 mg/L para FTA, 1077 mg/L CBL; 1028 mg/L TIO; 1084 mg/L DCA, 1011 mg/L DMT; 402,8 mg/L SIM e 404,8 CBZ. A partir dessas soluções estoque foram preparadas soluções intermediárias através de diluições em uma solução acetonitrila:água 50:50(v/v) levando às concentrações de 20,15 mg/L para FTA; 105,7 mg/L CBL; 39,63 mg/L TIO; 106,5 mg/L DCA; 99,21 mg/L DMT; 19,40 mg/L SIM e 40,51 mg/L para CBZ, sendo para os compostos FTA, TIO, DCA e SIM realizadas duas etapas de diluição. Essas soluções foram estocadas em frascos âmbar e mantidas a de 4 °C. Para a construção dos modelos de calibração foram preparadas seis soluções padrão, as quais foram utilizadas em ambos os métodos de referência (MR) e proposto (MP), abrangendo os intervalos de concentração apresentados na Tabela 25. Os intervalos de concentração foram estabelecidos com base nas recomendações do *Codex Alimentarius*<sup>117</sup> para os limites máximos permitidos de cada espécie de interesse, ou com base em análises preliminares utilizando o MP e observando as áreas dos picos

obtidas acompanhando os cromatogramas em 220 nm.

**Tabela 25.** Intervalos de concentração utilizados nas soluções padrões de cada espécie de interesse.

Analito	Intervalo de concentrações (mg/L)
DMT	1,00 – 7,50
FTA	0,10 – 1,40
TIO	0,50 – 5,37
SIM	0,10 – 1,24
CBL	1,00 – 6,00
CBZ	0,50 – 6,75
DCA	1,50 – 11,30

Para o MP foram observadas sobreposições entre DMT e FTA e entre TIO e SIM. Com o intuito de verificar a exatidão e a precisão dos modelos de calibração de segunda-ordem no MP nessas situações (que não apresentavam interferentes da amostra), foram preparadas nove misturas teste (conjunto VAL) contendo todos os compostos de interesse, mas seguindo um planejamento fatorial  $3^2$  para os quatro compostos que apresentaram sobreposição e com concentrações crescentes para os demais.

### 5.1.2 Instrumentação utilizada e condições cromatográficas

As análises foram realizadas em um equipamento HPLC-DAD da marca Shimadzu VP Series equipado com um amostrador automático modelo SIL-10AXL e um detector por arranjo de diodos modelo SPD-10AVP. Os dados foram adquiridos e exportados utilizando o programa Class VP versão 6.1, onde os espectros UV foram monitorados entre 190-370 nm, com velocidade de aquisição de 4 pontos/segundo no MP e 1,5 pontos/segundo no MR. Para as separações foi utilizada coluna Novapack C18 da marca Waters de dimensões (150 mm x 4,6 mm)

e preenchida com partículas de 4 µm de diâmetro e uma coluna de guarda similar.

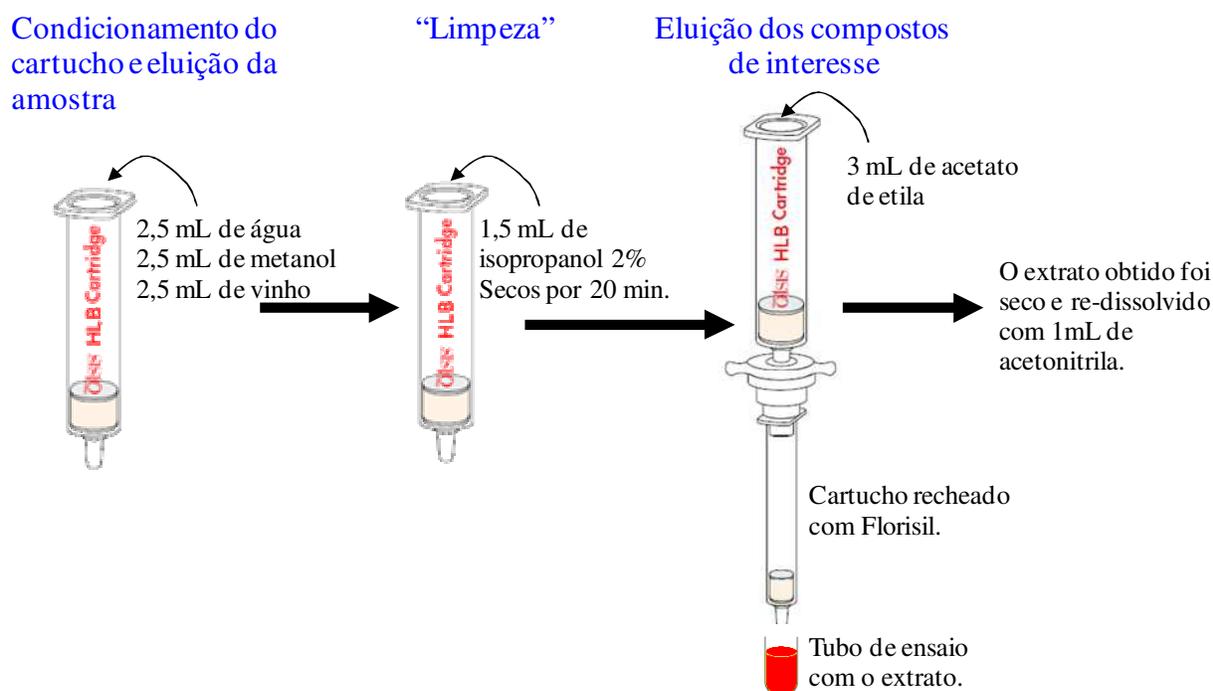
Nas separações pelo MP utilizou-se eluição isocrática com fase móvel composta por acetonitrila:água 50:50 (v/v), sendo o pH da água ajustado para 3,0 com ácido fosfórico e uma vazão de 0,60 mL/min. O MR empregou eluição por um gradiente binário composto por água acidificada a pH=3,0 com ácido fosfórico (solvente A) e acetonitrila (solvente B), de acordo com o programa: 15-25 % de B em 20 min (0,60 mL/min), 25-30 % de B em 10 min (0,60 mL/min), 30-40 % de B em 10 min (0,60 mL/min), 40-50 % de B em 10 min (1,00 mL/min), 50-55 % de B por 5 min (1,00 mL/min), manter 55 % de B por 5 min (1,00 mL/min), 55-15 % de B em 10 min (1,00 mL/min), manter 15 % de B por 10 min para retornar às condições iniciais.

### **5.1.3 Procedimento para extração em fase sólida**

Antes das análises cromatográficas, as amostras de vinho foram submetidas a uma etapa de extração em fase sólida com o objetivo de reduzir o número de interferentes presentes na amostra. Foram empregados cartuchos de extração da marca Oasis HLB com capacidade de 1,0 mL e um aparato para realização das extrações pela sucção dos reagentes e da amostra que seriam percoladas pelo cartucho (“*manifold*” a vácuo), ambos da marca Waters.

A Figura 65 ilustra o procedimento empregado na etapa de extração em fase sólida. Inicialmente os cartuchos foram condicionados através da passagem de 2,50 mL de metanol e 2,50 mL de água. Em seguida 2,50 mL de uma amostra de vinho tinto foi adicionada e percolada lentamente através do cartucho. Os cartuchos foram então lavados com 1,50 mL de uma solução de isopropanol 2 % (v/v) e secos por 20 minutos, deixando passar ar através dos cartuchos durante este tempo. Em seguida, os compostos de interesse foram eluídos com 3,00 mL de acetato de etila diretamente em um cartucho recheado com Florisil preparado no laboratório. O

acetato de etila foi percolado através dos dois cartuchos e o extrato obtido coletado em tubo de ensaio, o qual foi em seguida seco à temperatura ambiente sob fluxo de nitrogênio. A amostra seca foi então re-dissolvida com 1,00 mL de acetonitrila, obtendo um fator de concentração de 2,5 vezes, filtrados em uma membrana de poli(vinilideno) com poros de 0,45  $\mu\text{m}$  e finalmente transferidos para um frasco de amostra (*vial*) que foi colocado no amostrador automático do equipamento.



**Figura 65-** Esquema que ilustra o procedimento empregado no preparo das amostras de vinho.

#### 5.1.4 Amostras de vinho

As amostras de vinho tinto analisadas foram da marca Juan Carrau (Santana do Livramento-RS, Brasil), o qual de acordo com o fabricante foi produzido sem a utilização de pesticidas no cultivo das uvas. Logo, foi considerado que o vinho tinto utilizado neste trabalho não continha inicialmente os compostos de interesse. Foram analisadas dez amostras, as quais foram retiradas de uma única garrafa no momento de sua abertura.

A etapa de extração em fase sólida, descrita na seção anterior, representa

uma fonte significativa de erro sistemático, tendo em vista que uma recuperação de 100 % de todos os compostos de interesse dificilmente é obtida. Para melhor avaliar os resultados dos métodos construídos, as dez amostras de vinho analisadas foram divididas em dois grupos: (EX) seis amostras nas quais os analitos foram adicionados apenas após a etapa de extração em fase sólida, de forma que nessas amostras deve-se ter uma recuperação de 100 %; (VI) outras 4 amostras onde os analitos foram adicionados diretamente ao vinho, antes da etapa de extração em fase sólida, de forma que as perdas do método de extração pudessem ser verificadas.

## **5.2 Pré-processamentos utilizados nos dados**

No Capítulo 2 foi comentado que a aplicação da maior parte dos modelos de calibração de segunda ordem requer que as matrizes de dados sejam bilineares, isto é, possam ser representadas pelo produto de dois vetores que representam os perfis dos compostos presentes em cada dimensão e que os perfis normalizados sejam os mesmos para todas as amostras analisadas (o que corresponde a dizer que os dados formados pelo conjunto de várias amostras devem ser trilineares). Contudo, em dados cromatográficos a ocorrência de desvios de tempo de retenção é comumente observada. Isso representa desvios da trilinearidade dos dados, fazendo com que altos erros para as estimativas de concentração dos analitos sejam obtidos pelos modelos de calibração de segunda ordem.

Vários algoritmos já foram descritos na literatura para a correção desses desvios<sup>27,28,29,118,119,120</sup>. Contudo, apenas os métodos propostos por Prazen et al.<sup>27,28</sup> e Comas et al.<sup>29</sup> levam em consideração a utilização de matrizes de dados. Neste trabalho foi aplicado o método proposto por Prazen et al.<sup>27</sup>, a qual se baseia na decomposição SVD de duas matrizes adjuntas (N|M), uma tomada como referência (N) (que na prática consiste de uma das soluções padrão) e a outra a amostra a ser

corrigida (M) em relação a N, o que é expresso por:

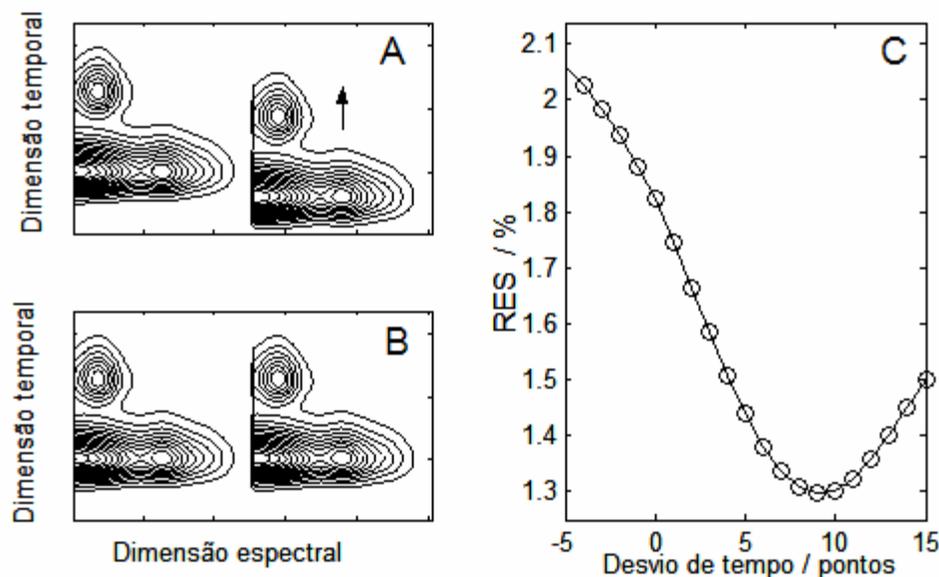
$$(\mathbf{B}, \mathbf{G}, \mathbf{C}) = \text{SVD}(\mathbf{N} | \mathbf{M}) \quad (80)$$

onde  $\mathbf{G}$  é uma matriz diagonal que contém os valores singulares e  $\mathbf{B}$  e  $\mathbf{C}$  são as matrizes dos vetores singulares esquerdo e direito, respectivamente. A correção é feita pelo cálculo de uma variância residual (RES), determinada para diversos deslocamentos da matriz M em relação à matriz N que é mantida constante, sendo que o número de deslocamentos é geralmente estabelecido pela inspeção visual dos cromatogramas. Os valores de RES são determinados a partir do vetor  $\mathbf{g}$  que contém os valores singulares de  $\mathbf{N} | \mathbf{M}$ , contidos na diagonal principal de  $\mathbf{G}$ , conforme a seguinte equação:

$$\text{RES} = 100 \frac{\sum_{a=A+1}^{\min(J,K)} g_a^2}{\sum_{a=1}^{\min(J,K)} g_a^2} \frac{JK}{(J - A)(K - A)} \quad (81)$$

onde A é o número de valores singulares significativos, idealmente igual ao número de espécies presentes em  $\mathbf{N} | \mathbf{M}$  acrescido de 1 e J e K o número de variáveis em cada dimensão. Idealmente, quando as duas matrizes estão alinhadas os valores de RES, determinados com A valores singulares, devem atingir um valor mínimo e dessa forma obtém-se a correção. Na Figura 66 este processo é ilustrado pela correção de dois padrões obtidos no MP na região onde os compostos TIO e SIM se sobrepõem.

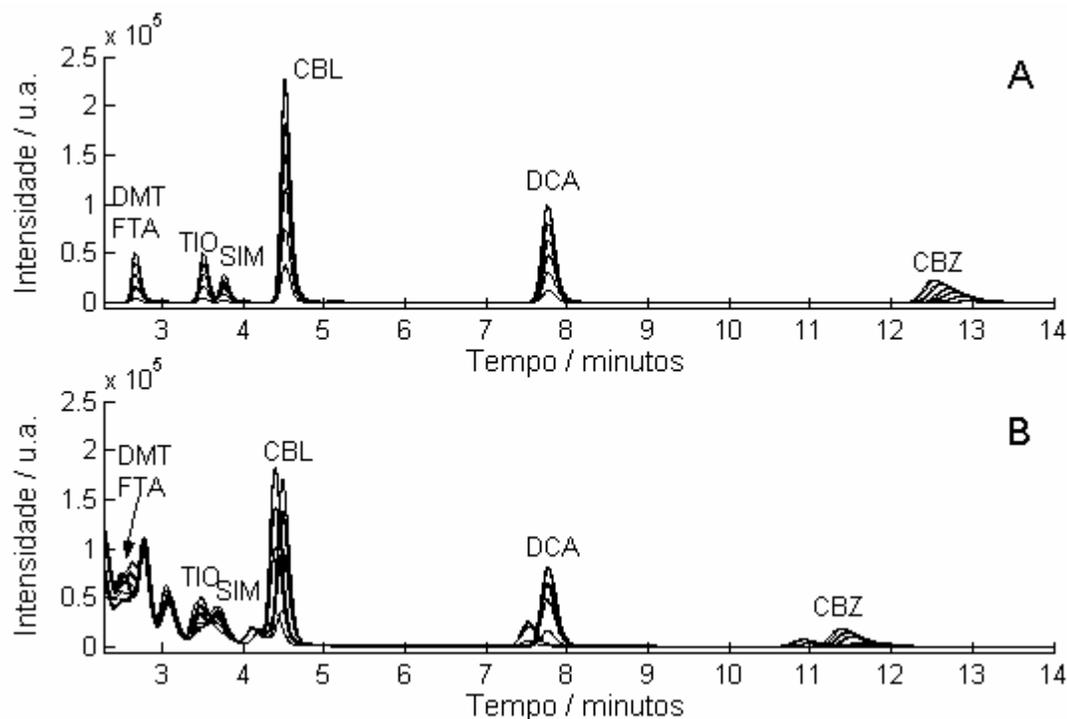
O programa utilizado para as correções dos tempos de retenção pelo método proposto por Prazen et al.<sup>27</sup> foi construído baseado na descrição dos trabalhos presentes na literatura<sup>27,28</sup> e também é compatível com ambos os sistemas Octave 3.0.0 e Matlab 6.5, cujo código fonte é apresentado no apêndice B10.



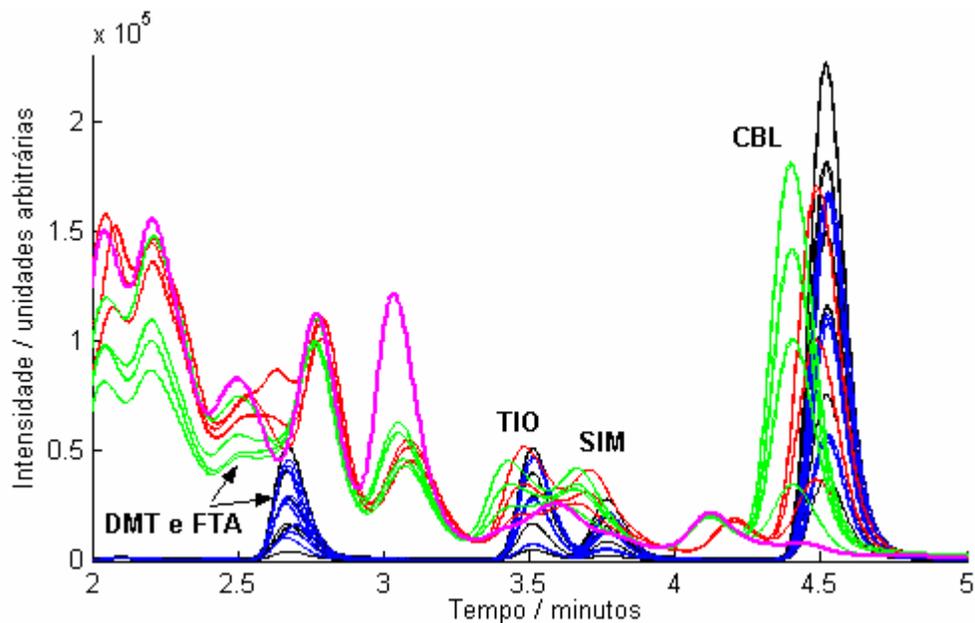
**Figura 66-** (A) Matriz de dados formada por  $N$  e  $M$  colocadas lado a lado (adjuntas) apresentando desvio de tempo, (B) matriz de dados adjunta alinhada pelo deslocamento de  $M$  em relação a  $N$  até que o mínimo de RES seja atingido, (C) gráfico de RES em função do deslocamento de  $M$  em relação a  $N$ .

### 5.3 Resultados e discussão

Nas Figuras 67 e 68 podem ser visualizados os cromatogramas em 220 nm obtidos pelo MP, onde se observa três regiões distintas em que os modelos de calibração de segunda-ordem são necessários: (1) para CBL, onde se tem apenas uma pequena sobreposição com os interferentes da amostra; (2) para TIO e SIM, onde é observada uma pequena sobreposição entre os picos dos analitos e a presença de interferentes da amostra; (3) para DMT e FTA, onde se observa uma grande sobreposição entre os picos dos analitos nos padrões e também a presença de interferentes da amostra que chegam a quase encobrir o pico dos analitos. A presença dos desvios de tempo de retenção também é observada, a qual pode ser visualizada mais claramente na Figura 68, sendo esses desvios diferentes entre as amostras EX e VI ou mesmo entre os compostos.



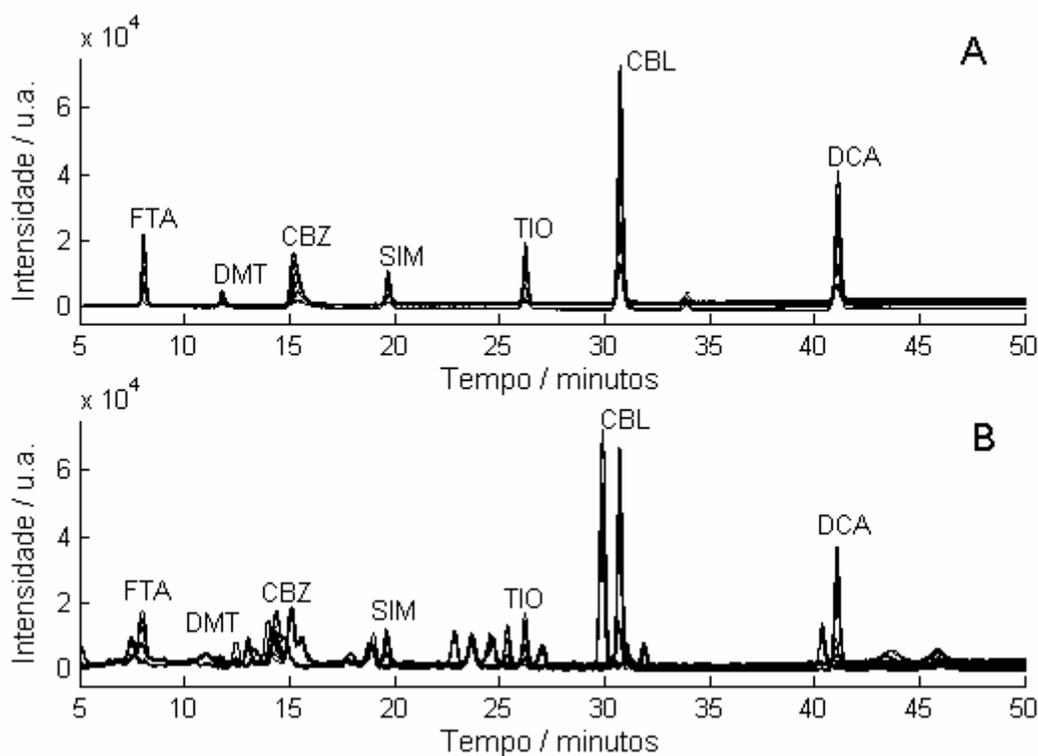
**Figura 67-** Cromatogramas em 220 nm obtidos no MP. (A) Padrões de calibração (B) extratos



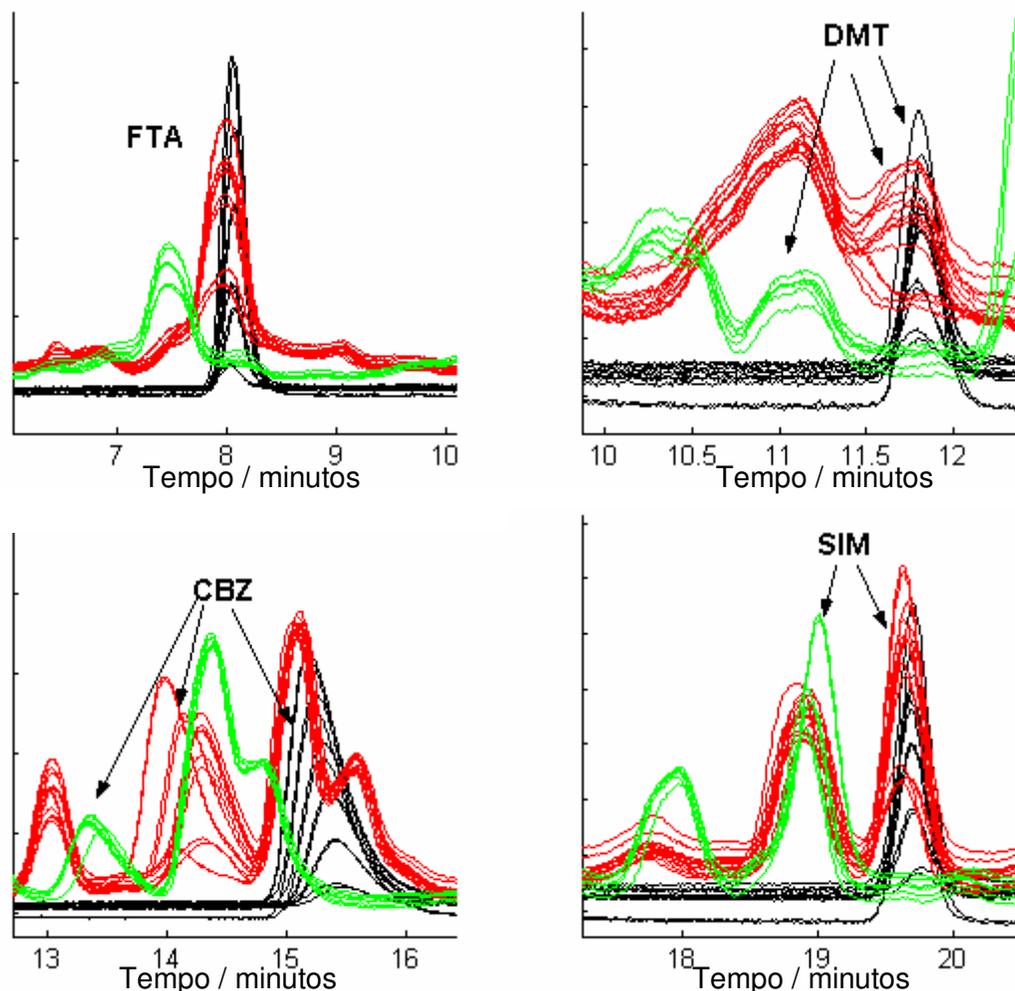
**Figura 68-** Cromatogramas em 220 nm no MP para os conjuntos de (—) calibração, (—) validação, (—) extratos, (—) vinho e (—) sinal de uma amostra de vinho sem adição dos analitos, destacando a região onde foi observada sobreposição entre os analitos ou pelos interferentes.

Para o MR, observa-se que todos os compostos apresentam boa resolução nos padrões (Figura 69A). Contudo, quando uma amostra (EX ou VI) é analisada

observam-se interferentes significativos nos compostos: FTA, DMT, CBZ e SIM (Figura 69B). A Figura 70 mostra em destaque as quatro regiões em que esses compostos são eluídos, onde se observa que além dos desvios nos tempos de retenção têm-se desvios de linha de base significativos. Para a correção desses desvios de linha base foi utilizado o método proposto por Boelens et al.<sup>121</sup>, cujo programa utilizado foi cedido gentilmente por P.H.C. Eilers<sup>121</sup>. Logo, apesar do MR ter sido desenvolvido com o intuito de obter valores independentes por calibração de ordem zero para comparar e validar os resultados obtidos pelos modelos de calibração de segunda ordem no MP, estes modelos também tiveram que ser aplicados a quatro situações no MR de modo a possibilitar a obtenção de resultados para a comparação entre os métodos.



**Figura 69-** Cromatogramas em 220 nm obtidos no MR. (A) Padrões de calibração (B) extratos.



**Figura 70-** Cromatogramas em 220 nm no MR para os conjuntos de (—) calibração, (—) extratos e (—) vinho, destacando a região onde foi observada sobreposição entre os analitos ou pelos interferentes.

Comparando os cromatogramas dos dois métodos observa-se que a ordem de eluição dos analitos é diferente, sendo as principais mudanças os aumentos dos tempos de retenção dos analitos TIO, SIM, CBL e DCA quando se passa do MP para o MR (que passam a ser eluídos depois de CBZ na mudança do MP para o MR) e pela inversão da ordem de eluição entre TIO e SIM. A ocorrência dessas mudanças é devido às diferenças na fase móvel, sendo um dos principais fatores responsáveis o pH, que é constante no MP que utiliza eluição isocrática e variável no MR, uma vez que a proporção entre acetonitrila e água em pH 3,0 é modificada ao longo da corrida no MR.

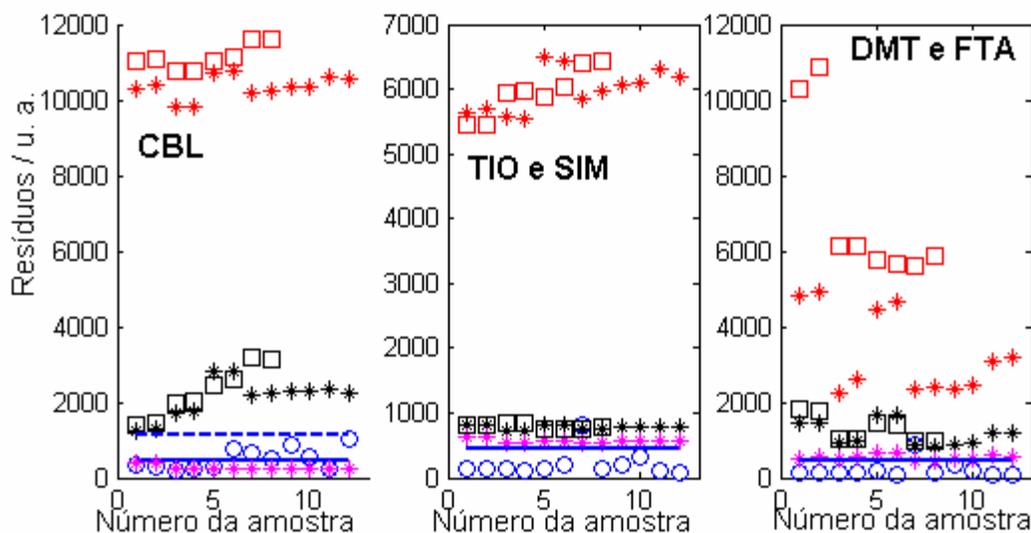
Para melhor clareza na visualização dos picos dos analitos nas Figuras 67 e 68, as corridas completas nos MP e MR não são apresentadas, sendo os tempos de análise aproximadamente iguais a 23 e 80 minutos para MP e MR, respectivamente. O prolongamento das corridas até este tempo foi necessário devido à eluição de um interferente das amostras de vinho em aproximadamente 20 minutos no MP e para restabelecer as condições iniciais de análise do gradiente utilizado no MR. Com esses tempos de corrida tem-se que o MP é aproximadamente 3,5 vezes mais rápido que o MR.

A construção de apenas um modelo de calibração de segunda ordem para a determinação de todos os analitos que apresentaram sobreposição em cada método é possível, mas conduz a resultados inferiores na determinação dos analitos, quando comparada à construção de modelos locais para cada região de sobreposição. Esse maior erro obtido com um modelo global deve-se principalmente por se tratar de uma situação mais complexa, que envolve um número maior de fatores que dificulta a resolução do sistema por parte dos modelos, além da correção dos desvios de tempos de retenção de vários picos simultaneamente constituir uma tarefa muito mais difícil. Logo, foram construídos modelos de calibração para cada região em que foi observada sobreposição separadamente, constituindo assim sete situações distintas: três no MP e quatro no MR.

### **5.3.1 Resultados obtidos pelo método proposto (MP)**

Na Figura 71 observa-se os resíduos para a escolha do número de fatores do RBL nas amostras EX e VI no BLLS, onde os limites sugerem que são necessários ao menos dois fatores para reduzir o resíduo de decomposição dessas amostras para valores inferiores ou aproximadamente iguais aos limites. Contudo, observando a concordância das estimativas das concentrações dos analitos com os valores nominais, obtêm-se melhores resultados utilizando apenas um fator no RBL. Para

esses dados, uma hipótese para essa discordância é a imperfeição da correção dos tempos de retenção, que devido à resolução dos pontos da matriz de dados não permite um alinhamento exato entre as diferentes amostras, o que possivelmente contribui para a obtenção de maiores resíduos nas amostras. Para o modelo UPLS, os resíduos obtidos apresentaram praticamente o mesmo comportamento e por isso não são apresentados, sendo as observações e conclusões as mesmas obtidas para o modelo BLLS pela Figura 71. A Tabela 26 mostra o número de fatores e os pré-processamentos que foram aplicados em cada situação do MP com os modelos BLLS, UPLS e PARAFAC.



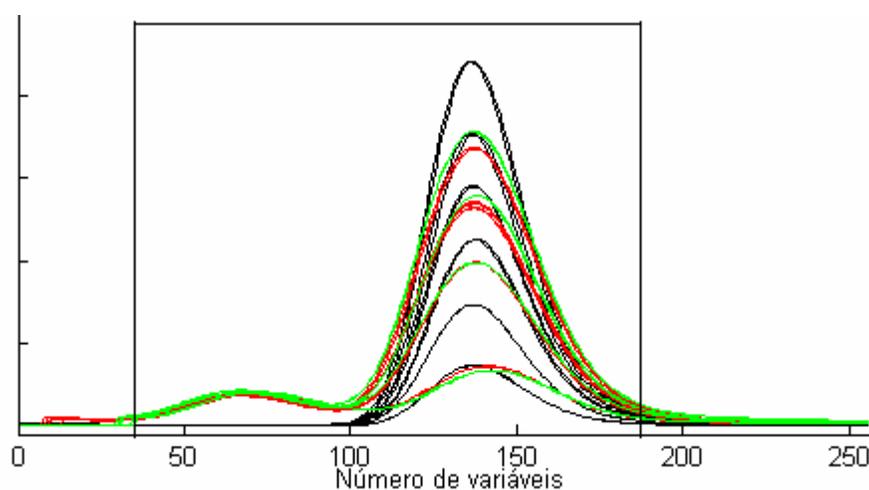
**Figura 71-** Resíduos de decomposição obtidos com o modelo BLLS no MP. Onde: (o) padrões de calibração, (\*) EX, (□) VI, (vermelho) Nenhum, (preto) um e (rosa) dois fatores no RBL. (—) Valor crítico pelo ruído instrumental e (---) pelo resíduo médio das amostras de calibração, ambos com 99% de confiança.

Na Figura 72 pode-se observar que após a correção dos desvios de tempo para o CBL os cromatogramas apresentam uma ótima concordância. A Figura 73 apresenta os gráficos de RES em função dos deslocamentos, que são obtidos para a correção desses desvios, onde se observa a ocorrência de um mínimo das curvas, o que caracteriza o alinhamento das matrizes de dados.

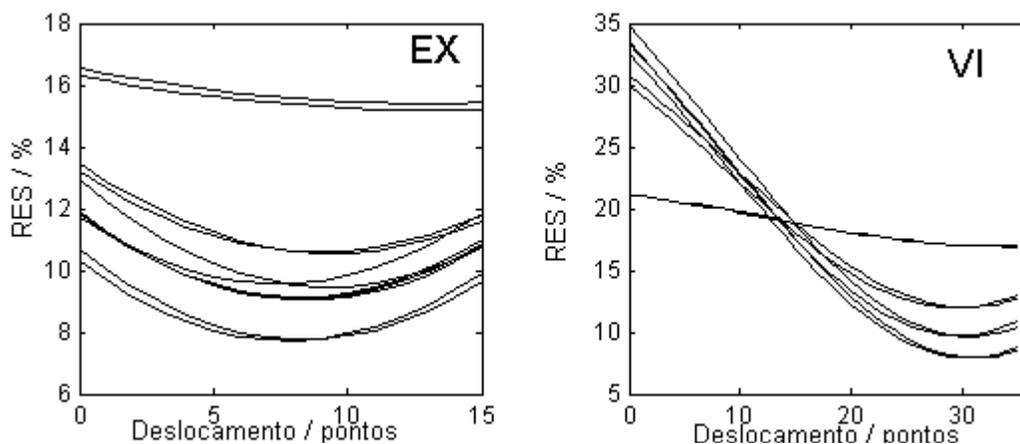
**Tabela 26.** Pré-processamentos e número de fatores utilizados na construção dos modelos de calibração de segunda ordem foram empregados no MP.

Situação	Pré-processamentos	Número de fatores		
		BLLS <sup>a</sup>	UPLS <sup>a</sup>	PARAFAC
CBL	Alinhamento dos tempos de retenção.	1+1	1+1	2
TIO e SIM	Alinhamento dos tempos de retenção.	2+1	2+1	3
FTA e DMT	Subtração do sinal do branco e alinhamento dos tempos de retenção.	2+1	2+1	3

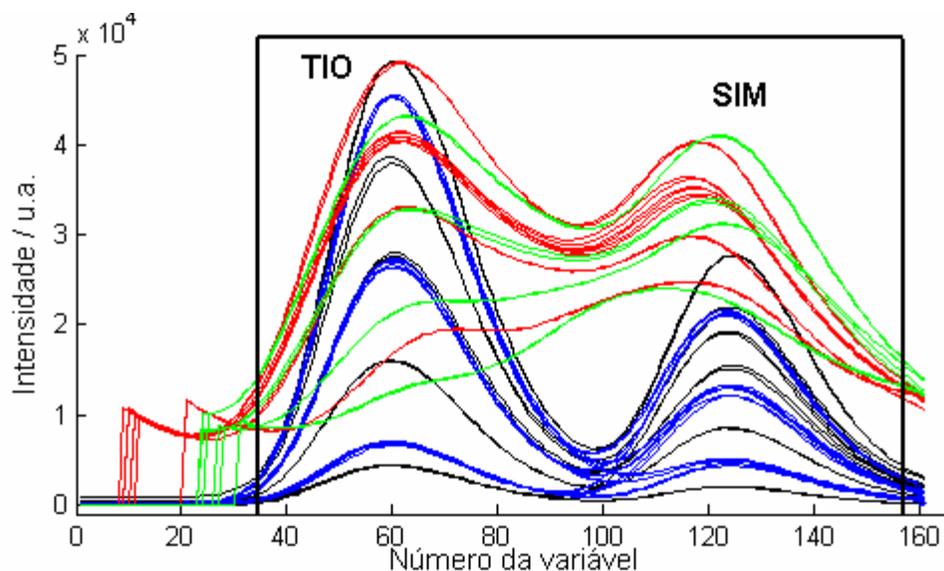
(a) o primeiro e o segundo número referem-se ao número de fatores utilizados na etapa de calibração e no RBL, respectivamente e modelos construídos com dados centrados na média.



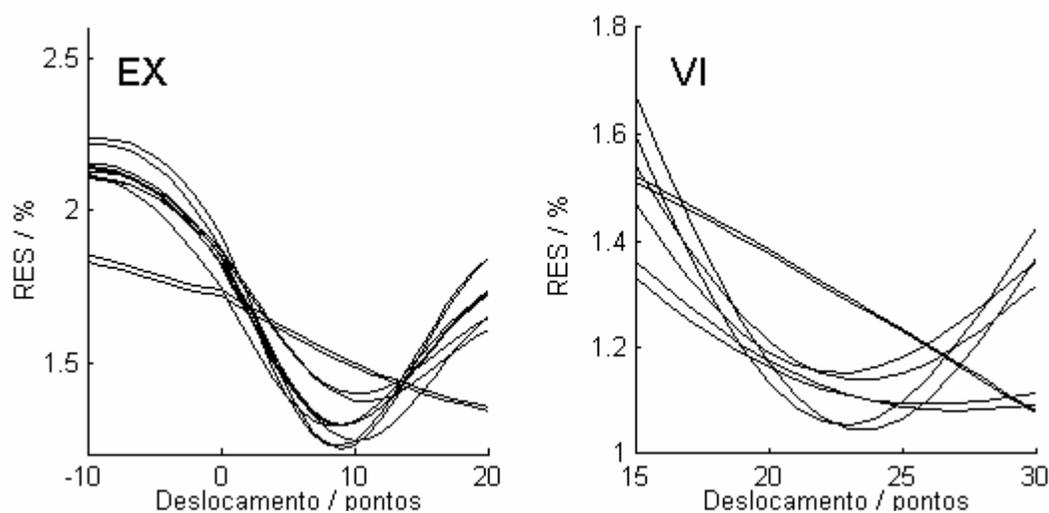
**Figura 72-** Cromatogramas em 220 nm para CBL após a correção dos desvios de tempo de retenção, onde a região selecionada foi a utilizada para a construção dos modelos, (—) padrões de calibração, (—) EX e (—) VI.



**Figura 73-** Variância residual (RES) obtida a partir da correção dos desvios de tempo para as amostras EX e VI na região de eluição dos composto CBL.



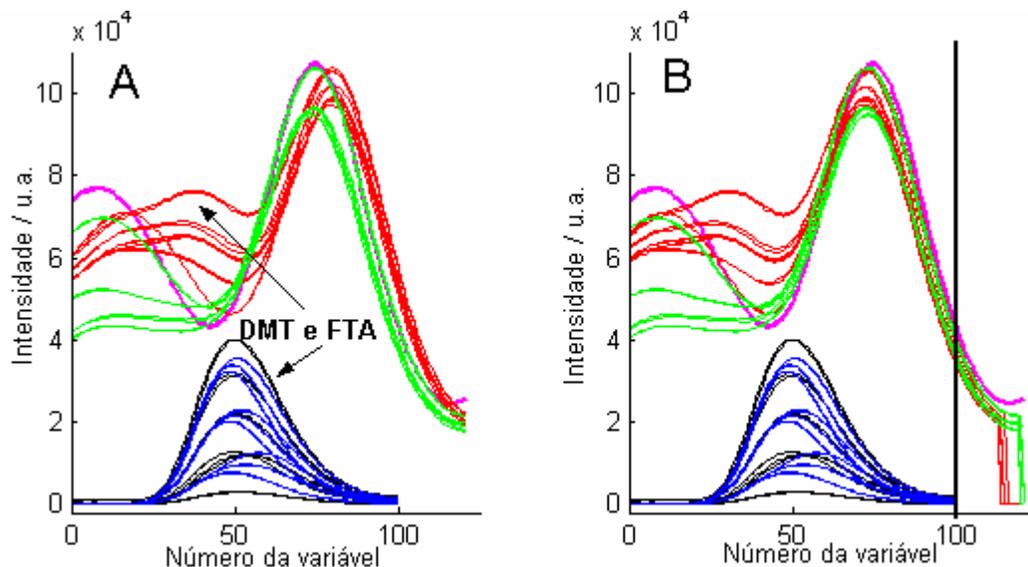
**Figura 74-** Cromatogramas em 220 nm para TIO e SIM após a correção dos desvios de tempo de retenção, onde a região selecionada foi aquela utilizada para a construção dos modelos, (—) padrões de calibração, (—) misturas de validação, (—) amostras EX e (—) VI.



**Figura 75-** Variância residual (RES) obtida a partir da correção dos desvios de tempo para EX e VI na região de eluição dos compostos TIO e SIM.

Para a correção dos desvios de tempo da região de eluição dos compostos TIO e SIM, também foi observado que aparentemente todas as amostras EX e VI tiveram os tempos de retenção alinhados em relação aos padrões com boa concordância (Figura 74). Contudo, na Figura 75 observa-se que para as duas replicatas das amostras EX e VI de menor concentração não atingem um valor mínimo, para deslocamentos de até 20 e 30 pontos para EX e VI, respectivamente.

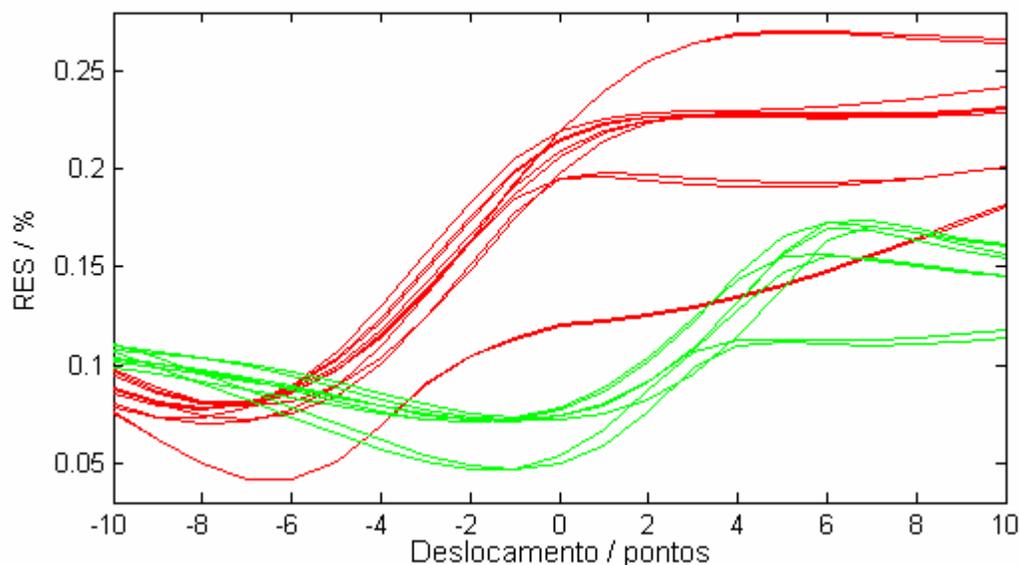
Caso se faça deslocamentos de aproximadamente 50 pontos, chega-se a atingir um valor mínimo. Entretanto, pela concentração estimada para a amostra EX de menor concentração, este não corresponde ao alinhamento correto, uma vez que são obtidos altos erros para essa amostra. Esse resultado indica uma limitação do método de alinhamento para baixas concentrações, que provavelmente deve-se ao total encobrimento do sinal dos analitos pelos interferentes da amostra.



**Figura 76-** Cromatogramas em 220 nm para DMT e FTA (A) antes e (B) após a correção dos desvios dos tempos de retenção em relação ao sinal do branco, onde a região selecionada foi aquela utilizada na próxima etapa do pré-processamento, (—) padrões de calibração, (—) misturas de validação, (—) EX, (—) VI e (—) sinal do branco.

Para a determinação de DMT e FTA, cujos cromatogramas sem correção são apresentados na Figura 76A, a correção dos desvios dos tempos de retenção das amostras EX e VI em relação aos padrões não apresentaram bons resultados devido à alta intensidade do sinal dos interferentes que são eluídos juntamente com os analitos. Mesmo com uma tentativa de alinhamento manual dos cromatogramas os modelos de calibração não são capazes de obter bons resultados para as estimativas das concentrações DMT e FTA, pois a alta sobreposição entre os analitos juntamente com os interferentes impossibilita a determinação dessa forma. Tendo em vista esses problemas, na tentativa de reduzir o sinal dos interferentes e

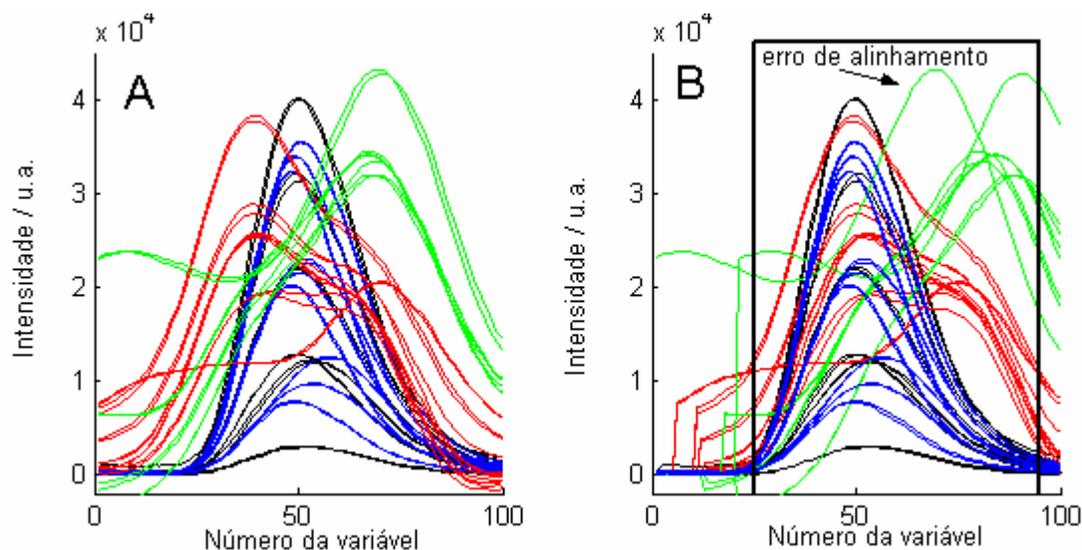
possibilitar a determinação dos analitos nessa situação, além da correção dos tempos de retenção foi utilizada uma subtração do sinal de uma amostra de vinho em que os analitos não tinham sido adicionados (amostra de branco). De modo que o pré-processamento neste caso consistiu de: (1) alinhamento das amostras EX e VI com uma amostra do branco, (2) subtração do sinal do branco previamente multiplicado por uma constante escolhida de modo a evitar a obtenção de sinais negativos nos cromatogramas, (3) alinhamento do sinal das amostras EX e VI após a subtração do sinal do branco com os padrões de calibração e (4) seleção da região para a construção do modelo. É importante ressaltar que o procedimento de subtração do sinal do branco que foi empregado somente foi possível porque os interferentes são sempre os mesmos em todas as amostras, uma vez que todas as amostras de vinho foram retiradas de uma mesma garrafa de vinho.



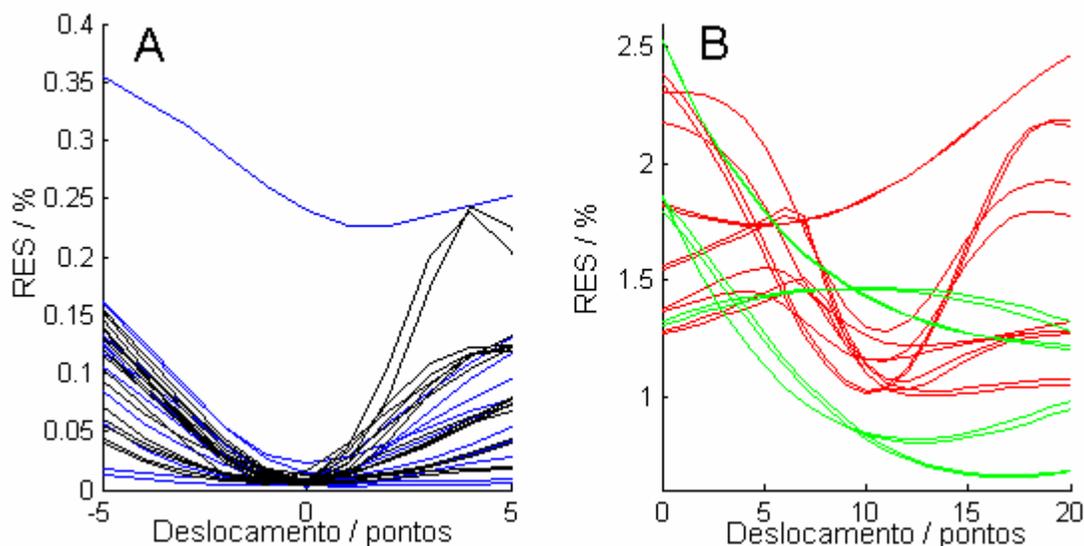
**Figura 77-** Variância residual (RES), para a na região de eluição de DMT e FTA, obtida a partir da primeira correção dos desvios de tempo realizada para alinhar as amostras EX e VI com o sinal do branco. Onde: (—) EX e (—) VI

A Figura 76B mostra os resultados dos alinhamentos das amostras EX e VI com o sinal do branco, sendo o gráfico dos resíduos desses alinhamentos apresentado na Figura 77, onde se observa a ocorrência de um mínimo e todas

replicatas de todas as amostras. Na Figura 78A e 78B podem ser observados os cromatogramas após a subtração do sinal do branco antes e após a segunda correção dos desvios de tempo de retenção. Observa-se que nas amostras EX o pico cromatográfico dos analitos, após os pré-processamentos, apresenta uma maior intensidade em relação aos interferentes e visualmente com um bom resultado para o alinhamento. Contudo, para as amostras VI observa-se visualmente que ao menos uma amostra apresenta um erro no alinhamento dos tempos de retenção. Na Figura 78B também pode ser observado que provavelmente são obtidas perdas significativas pelo método de extração para esses analitos, uma vez que a intensidade observada para o pico dos analitos é menor que nas amostras EX. Os gráficos de resíduos obtidos para o segundo alinhamento (Figura 79) reforçam que para a os padrões, misturas de validação e amostras EX são obtidos bons resultados para o alinhamento. Entretanto, para as amostras VI é observado que duas das oito replicatas não apresentam gráfico com a obtenção de um mínimo, indicando um possível erro no alinhamento dessas amostras.

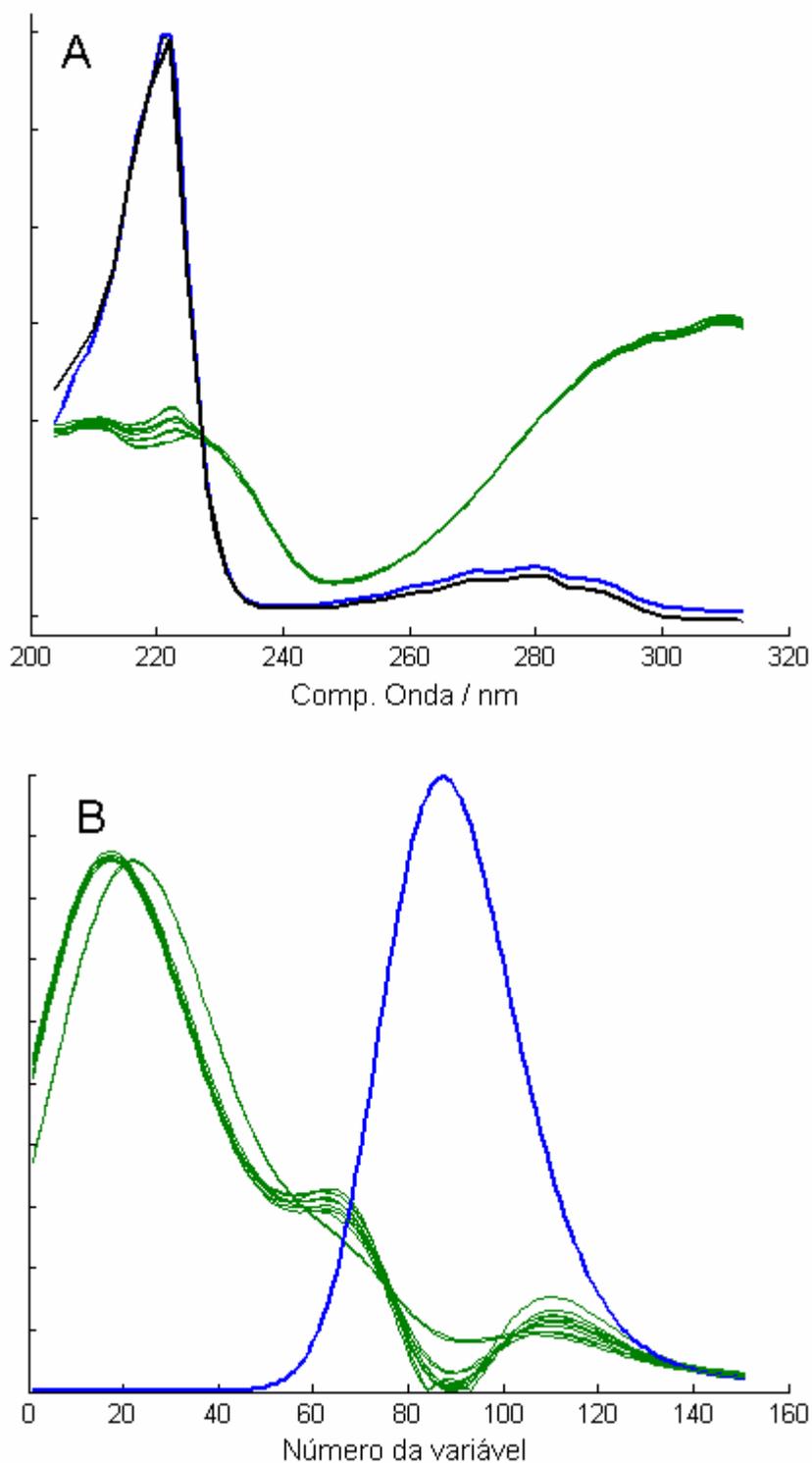


**Figura 78-** Cromatogramas em 220 nm para DMT e FTA após a subtração do sinal do branco (A) antes e (B) depois da correção dos desvios de tempo de retenção em relação aos padrões, onde a região selecionada foi aquela utilizada na construção dos modelos de calibração, (—) padrões de calibração, (—) misturas de validação, (—) EX e (—) VI.

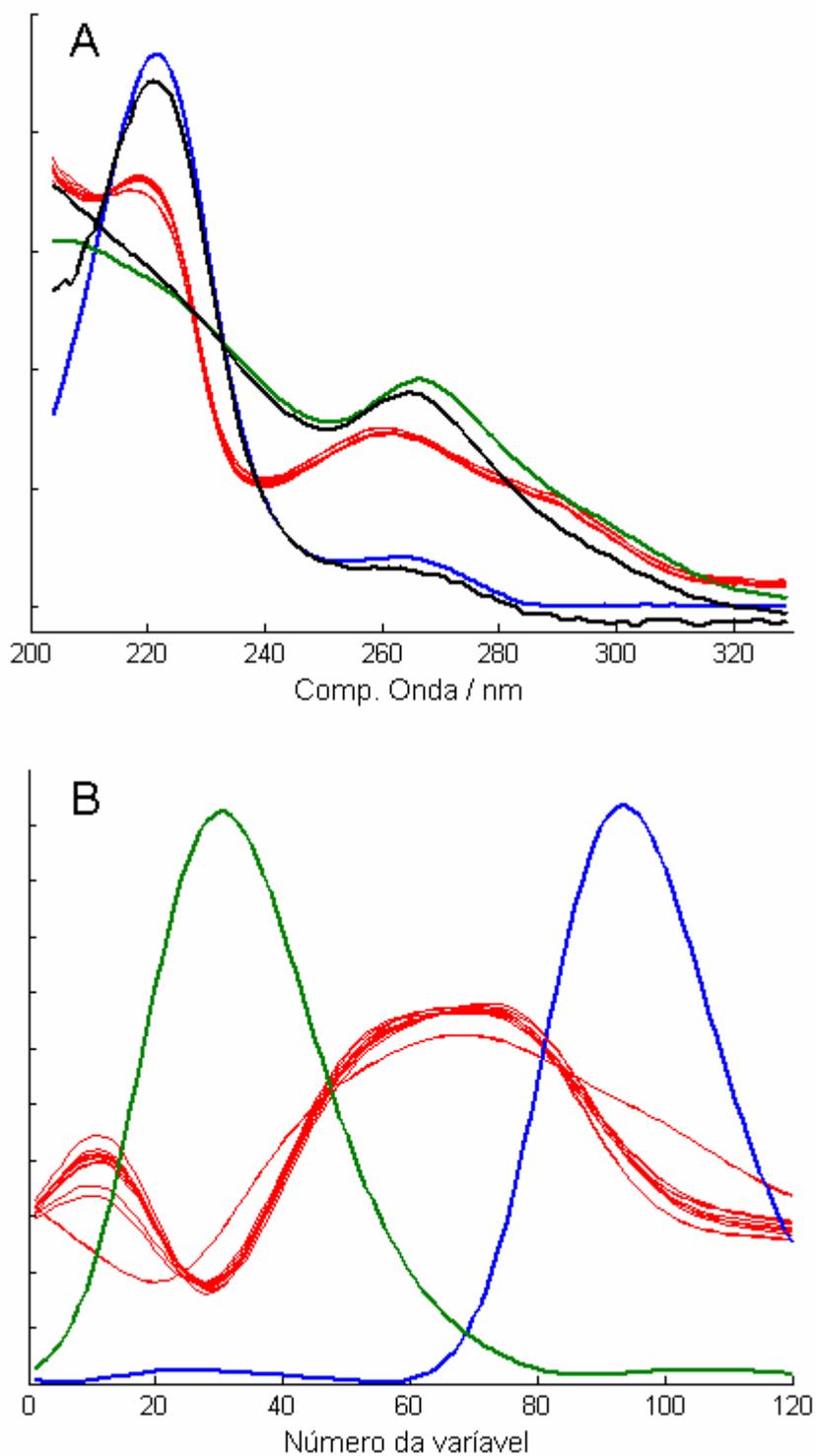


**Figura 79-** Variância residual (RES), para DMT e FTA, obtida a partir da correção dos desvios dos tempos de retenção em relação a um padrão de calibração. Onde: (A) (—) padrões de calibração e (—) misturas de validação; e (B) para (—) EX e (—) VI.

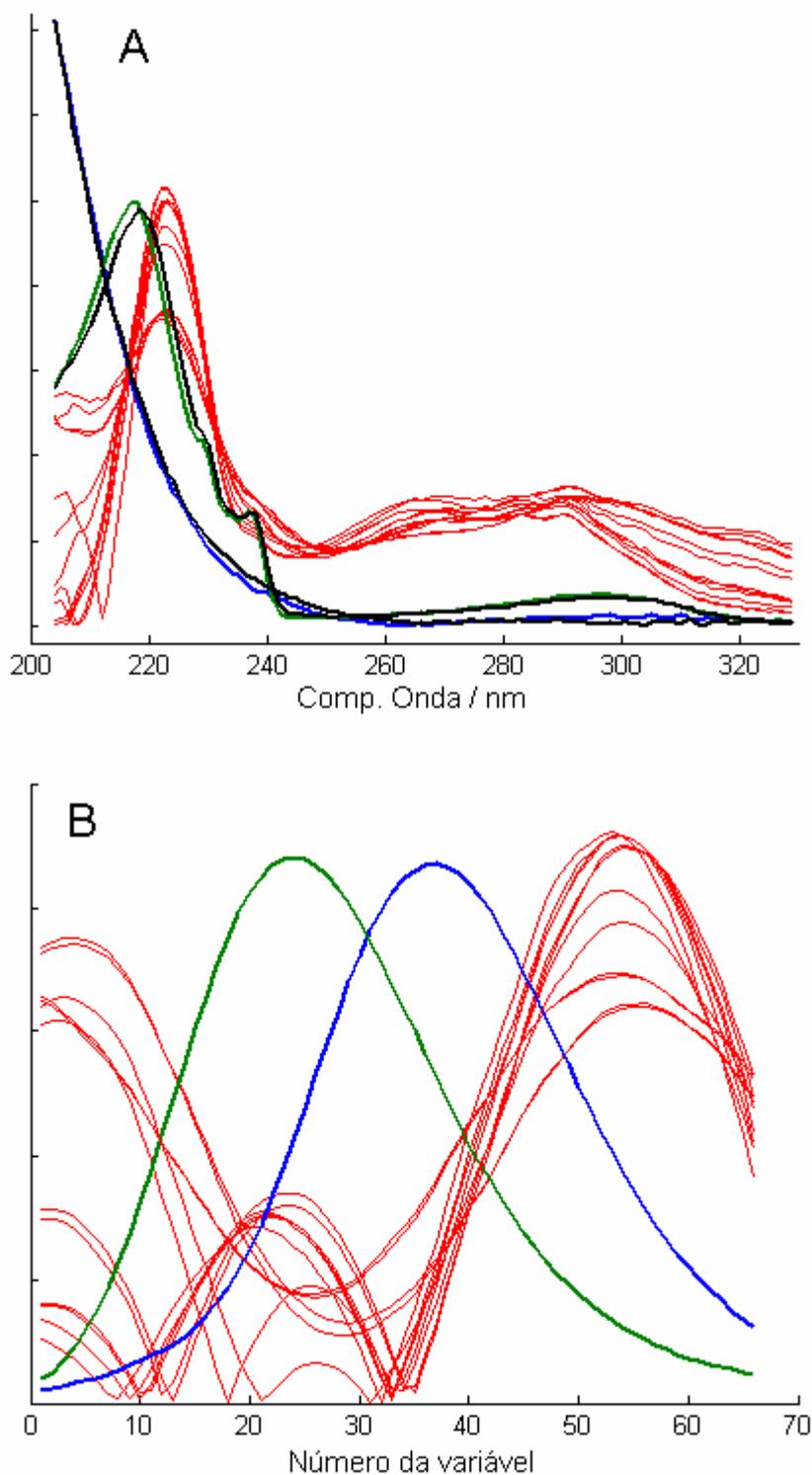
As Figuras 80, 81 e 82 apresentam os perfis espectrais e cromatográficos estimados por BLLS para CBL; TIO e SIM; e DMT e FTA, respectivamente. Observa-se que há uma boa concordância entre os perfis espectrais experimentais obtidos nos padrões do MR e os estimados pelo BLLS, sendo obtidos coeficientes de correlação entre esses perfis de ao menos 0,98. Quanto aos perfis cromatográficos estimados, observa-se que os perfis do interferente para a determinação de CBL e de TIO e SIM apresentam uma pequena variação, com a exceção do perfil estimado para as duas replicatas de menor concentração em cada caso, o que provavelmente reflete uma possível imperfeição da correção dos desvios de tempos de retenção dessas amostras. Para a determinação de DMT e FTA, é observado que há uma grande variação nos perfis dos interferentes, a qual também é atribuída a eventuais imperfeições na correção dos desvios nos tempos de retenção, que neste caso envolveu duas etapas. Os perfis obtidos com o modelo PARAFAC apresentaram aproximadamente os mesmos resultados, inclusive para a variação observada no perfil dos interferentes e por isso não são apresentados.



**Figura 80-** Estimativas dos perfis (A) espectrais e (B) cromatográficos, ambos normalizados, obtidos por BLLS para a determinação de CBL nas amostras EX. Onde: (—) perfil experimental obtido pelo MR, (—) estimado para CBL e (—) para o interferente estimado pelo RBL.

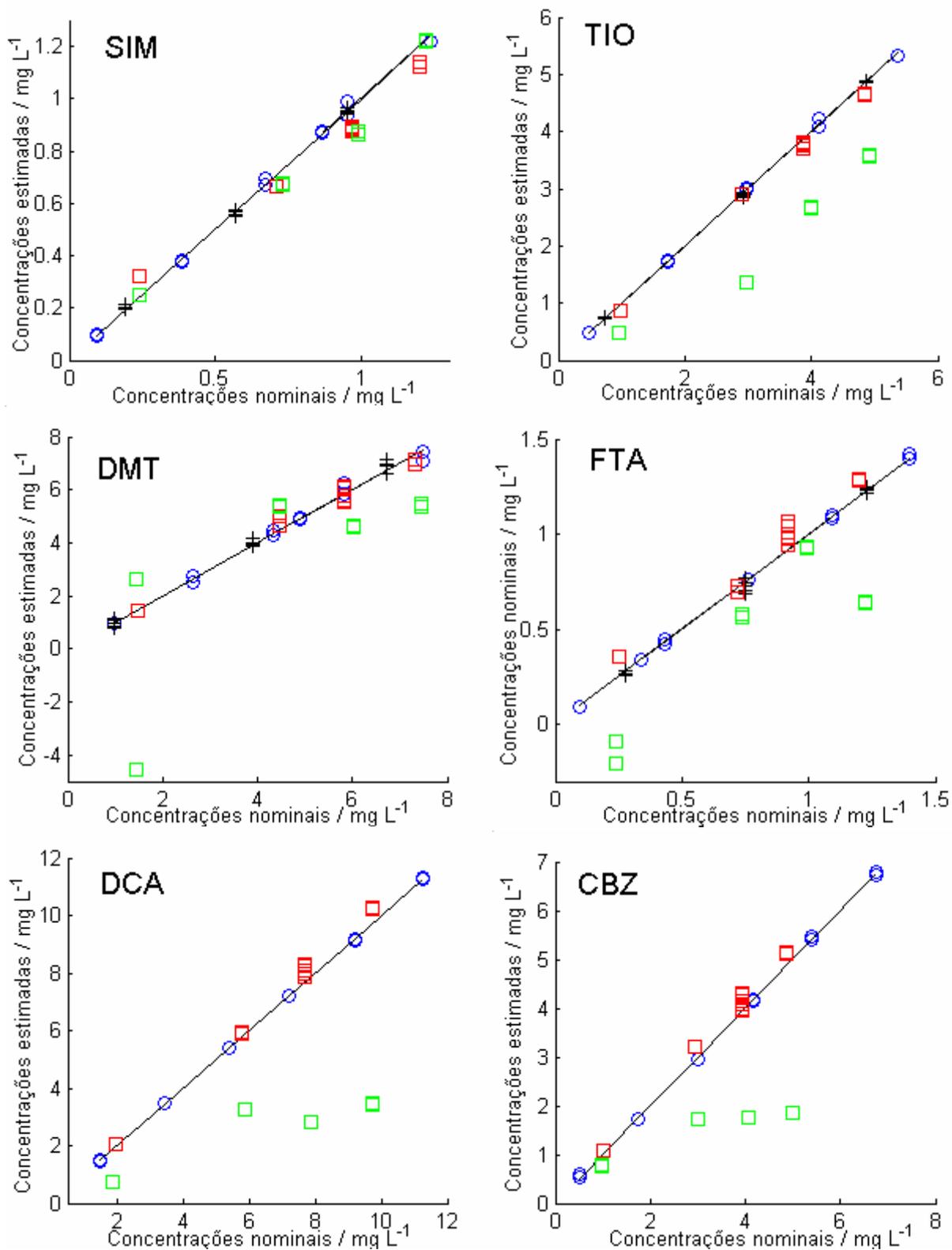


**Figura 81-** Estimativas dos perfis (A) espectrais e (B) cromatográficos, ambos normalizados, obtidos por BLLS para a determinação de TIO e SIM nas amostras EX. Onde: (—) perfis experimentais obtidos pelo MR, (---) estimado para SIM, (—) TIO e (---) para o interferente estimado pelo RBL.

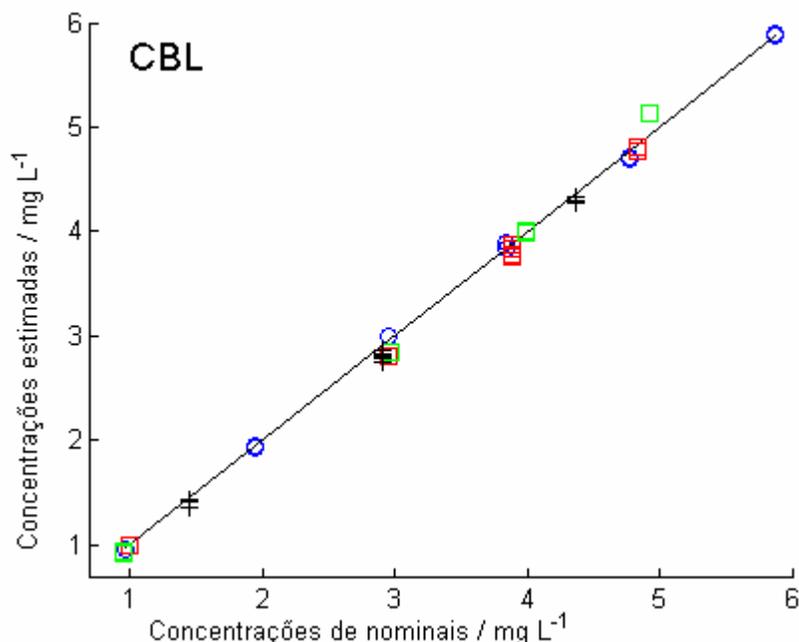


**Figura 82-** Estimativas dos perfis (A) espectrais e (B) cromatográficos, ambos normalizados, obtidos por BLLS para a determinação de DMT e FTA nas amostras EX. Onde: (—) perfis experimentais obtidos pelo MR, (—) estimado para DMT, (—) FTA e (—) para o interferente estimado pelo RBL.

As Figuras 83 e 84 mostram os gráficos dos valores de referência versus estimados pelo modelo BLLS para CBL, TIO, SIM, DNT e FTA e por calibração de ordem zero para CBZ e DCA. Em todos os casos foram observados bons resultados para as amostras EX, que são visualizados pela pequena dispersão ou pelos erros médios (RMSEP e REP) apresentados na Tabela 27. Para as amostras VI foram observadas maiores dispersões e erros, com exceção de CBL e SIM, que também apresentaram bons resultados nessas amostras. Com relação às amostras de menor concentração nos compostos TIO e SIM, as quais não tiveram uma boa correção dos desvios de tempo de retenção de acordo com os gráficos de RES, observou-se que apenas SIM apresentou um alto erro nessa amostra, o qual foi de aproximadamente de 30 %, enquanto que para TIO essa amostra apresentou um erro de apenas 11 %. Logo, para SIM essa amostra não foi considerada para o cálculo das figuras de mérito apresentadas na Tabela 27. Do mesmo modo, pela Figura 83 para DMT e FTA, observa-se que é obtido um alto erro para as replicatas da amostra VI de menor concentração, as quais provavelmente não foram alinhadas corretamente. Quanto à ocorrência de outras amostras com altos erros em relação às demais nos outros compostos, considerando o BLLS e o UPLS, apenas FTA apresentou um erro de aproximadamente 40 % para a amostra EX de menor concentração, a qual também não foi considerada para os cálculos das figuras de mérito e índices de recuperação. Para o PARAFAC, além das amostras já comentadas, foram observadas outras duas replicatas com altos erros, uma no composto DMT e outra no FTA. Com exceção dessas últimas amostras para DMT e FTA, os gráficos dos valores de referência versus estimados para os modelos UPLS e PARAFAC (que não são apresentados) foram similares aos obtidos com BLLS.



**Figura 83-** Valores de referência versus os estimados pelo MP com o BLLS (TIO, SIM, DMT e FTA) e calibração de ordem zero (DCA e CBZ). (—) Bissetriz do quadrante, (●) padrões de calibração, (+) amostras de validação, (□) EX e (□) VI.



**Figura 84-** Valores de referência versus estimados para DCA pelo MP com BLLS. (—) Bissetriz do quadrante, (●) padrões de calibração, (+) amostras de validação, (□) EX e (□) VI.

De forma geral, observando a Tabela 27, nota-se que os resultados obtidos pelos modelos BLLS e UPLS são muito semelhantes. As maiores diferenças são observadas para DMT e TIO, onde o BLLS apresenta erros menores, mas que não chegam a ser estatisticamente significativos. Quanto ao modelo PARAFAC, este apresenta resultados similares aos obtidos por BLLS e UPLS, com exceção dos compostos DMT e FTA, sendo observado para DMT um valor de RMSEP cerca de 2,5 vezes maior que o observado com o modelo BLLS além da ocorrência de altos erros em algumas replicatas como comentado anteriormente.

Os valores de seletividade calculados foram coerentes entre os três modelos, havendo uma maior diferença apenas nos compostos DMT e FTA para o UPLS. Já os valores estimados para os limites de detecção são praticamente os mesmos para os três modelos, em todos os compostos, e indicam que os compostos CBL, SIM e FTA são os que apresentam uma maior sensibilidade, a nível de unidades de  $\mu\text{g/L}$ , também apresentando os menores erros médios e a melhor precisão.

**Tabela 27.** Resultados das figuras de mérito obtidas para a determinação dos pesticidas e metabólitos pelo método proposto (MP).

	RMSEP <sup>a</sup>		REP <sup>b</sup>		Precisão <sup>a</sup>		SEL		LD <sup>c</sup>	
	VAL	EX	VAL	EX	VAL	EX	VAL	EX	VAL	EX
<b>BLLS</b>										
DMT	0,20	0,24	7,5	4,8	0,12	0,13	0,75	0,51(±0,03)	16	23(±3)
FTA	0,02	0,08	4,1	8,0	0,01	0,01	0,75	0,39(±0,03)	1,4	2,6(±0,2)
TIO	0,03	0,13	1,6	5,6	0,01	0,02	1	0,21(±0,01)	2,0	9,6(±0,3)
SIM	0,01	0,08	5,0	7,9	0,006	0,004	1	0,37(±0,00)	1,4	3,9(±0,1)
CBL	0,07	0,08	3,0	2,5	0,01	0,01	1	0,84(±0,01)	2,0	1,2(±0,0)
<b>UPLS</b>										
DMT	0,20	0,32	8,7	6,8	0,11	0,19	---	0,68(±0,04)	15	23(±1)
FTA	0,02	0,08	3,8	8,7	0,01	0,01	---	0,54(±0,04)	1,4	2,5(±0,2)
TIO	0,03	0,13	1,5	9,0	0,01	0,02	---	0,21(±0,01)	2,0	9,6(±0,4)
SIM	0,01	0,08	5,6	7,8	0,006	0,004	---	0,37(±0,00)	1,4	3,9(±0,0)
CBL	0,07	0,08	3,0	2,5	0,01	0,01	1	0,84(±0,01)	2	1,2(±0,0)
<b>PARAFAC</b>										
DMT	0,19	0,61	5,7	12	0,11	0,33	0,79	0,50(0,03)	15	24(2)
FTA	0,02	0,12	2,3	13	0,01	0,12	0,79	0,45(0,03)	1,3	2,3(±0,2)
TIO	0,03	0,12	1,5	7,2	0,01	0,02	1	0,21(±0,01)	2,0	9,7(±0,3)
SIM	0,01	0,07	3,3	7,5	0,005	0,004	1	0,37(±0,00)	1,5	4,0(±0,0)
CBL	0,07	0,08	3,0	2,4	0,01	0,01	1	0,86(±0,01)	2	1,2(±0,0)
<b>Univariado</b>										
DCA	0,08	0,37	1,3	5,2	0,03	0,12	1	1	6,2	6,2
CBZ	0,12	0,24	3,2	7,2	0,04	0,04	1	1	7,3	7,3

Grandezas expressas por: (a) mg/L, (b) percentagem e (c) 10<sup>-3</sup> mg/L.

Uma vez que os resultados obtidos no MP com os três modelos de calibração de segunda ordem são muito próximos e que o modelo BLLS é o mais simples e que também é capaz de fornecer informações tanto quantitativas quanto qualitativas

neste caso, este foi considerado o mais eficiente neste método. A exatidão para as determinações dos analitos no MP no BLLS foi confirmada verificando se os intervalos de confiança para a inclinação e o intercepto da regressão linear entre os valores nominais e estimados para as amostras EX contém seus valores esperados 1 e 0, respectivamente. Estes valores são apresentados na Tabela 28, e confirmam que as determinações podem ser considerados exatas com 99 % de confiança para todos os analitos, não sendo observados erros sistemáticos significativos tanto constantes quanto proporcionais para nenhum dos compostos de interesse.

**Tabela 28.** Resultados para as inclinações e interceptos das regressões lineares entre os valores de referência a estimados para as amostras EX no MP.

<b>Analito</b>	<b>Inclinação</b>	<b>Intercepto</b>
DMT	0,97(±0,13)	-0,18(±0,7)
FTA <sup>a</sup>	0,99(±0,17)	-0,07(±0,15)
TIO	0,98(±0,05)	-0,04(±0,19)
SIM <sup>a</sup>	0,94(±0,12)	-0,02(±0,11)
CBL	0,99(±0,04)	-0,02(±0,16)
DCA	1,05(±0,05)	-0,03(±0,35)
CBZ	1,04(±0,09)	0,07(±0,22)

Intervalos de confiança com 99 % entre parênteses, com 8 graus de liberdade para FTA e SIM e 10 nos demais compostos. (a) Não consideram a amostra EX de menor concentração.

### 5.3.2 Resultados obtidos pelo método de referência (MR)

A Tabela 29 mostra os pré-processamentos que foram utilizados nas quatro situações em que foi necessária a aplicação dos modelos de calibração de segunda ordem ao MR, sendo que neste método também foi necessária a correção dos desvios de linha de base para os analitos CBZ, SIM e DMT.

Na Figura 85 podem ser visualizados os cromatogramas na região onde o composto CBZ é eluído após cada etapa do pré-processamento. Primeiramente foi feita uma correção dos desvios de linha de base, cujo resultado pode ser visualizado

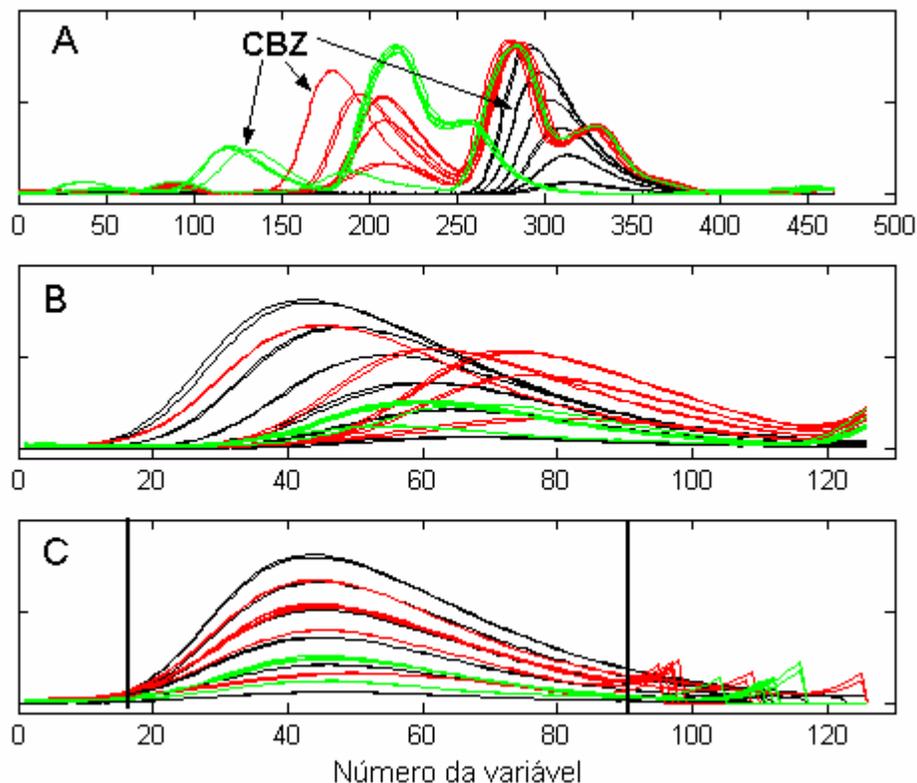
comparando as Figuras 70 para CBZ e Figura 85A. Em seguida, devido ao grande desvio dos tempos de retenção nas amostras VI, EX e os padrões de calibração, foi feita uma correção prévia dos desvios de tempo das amostras VI em relação a uma das amostras EX e selecionada manualmente uma região nos cromatogramas dos padrões para facilitar a correção pelo programa (Figura 85B). Por fim as matrizes de dados das amostras EX e VI foram alinhadas às matrizes dos padrões e selecionada uma região para a construção do modelo (Figura 85C). Como apresentado na Tabela 29, após a realização dos pré-processamentos, a construção dos modelos para este analito foi realizada utilizando 2 fatores para o PARAFAC e 1 para a etapa de calibração e 1 fator a mais para a análise das amostras EX e VI com o RBL nos modelos BLLS e UPLS.

**Tabela 29.** Pré-processamentos e número de fatores utilizados na construção dos modelos de calibração de segunda ordem que foram empregados no MR.

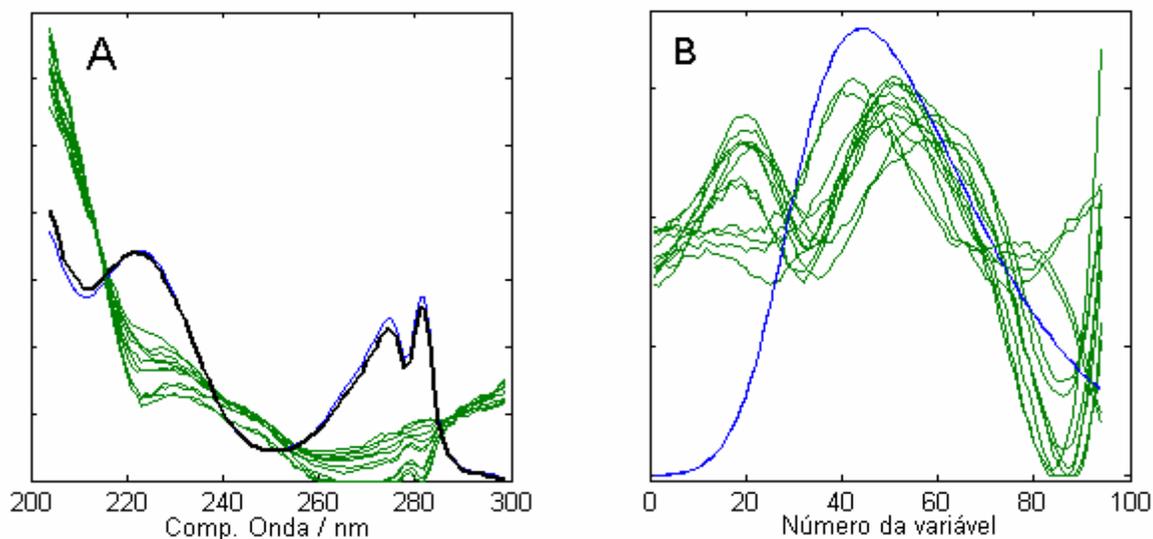
Situação	Pré-processamentos	Nº. de fatores		
		BLLS <sup>a</sup>	UPLS <sup>a</sup>	PARAFAC
CBZ	Correção de linha de base e alinhamento dos tempos de retenção.	1+1	1+1	2
SIM <sup>b</sup>	Correção de linha de base, subtração do sinal do branco e alinhamento dos tempos de retenção.	1+0	1+0	1
DMT	Correção de linha de base, alinhamento dos tempos de retenção.	---	2+0	2
FTA	Correção dos tempos de retenção.	---	2+1	2

(a) O primeiro e o segundo número referem-se aos fatores utilizados na calibração e no RBL, respectivamente e modelos construídos com dados centrados na média. (b) A amostra EX de menor concentração utilizou 1 fator no RBL e 2 fatores no PARAFAC.

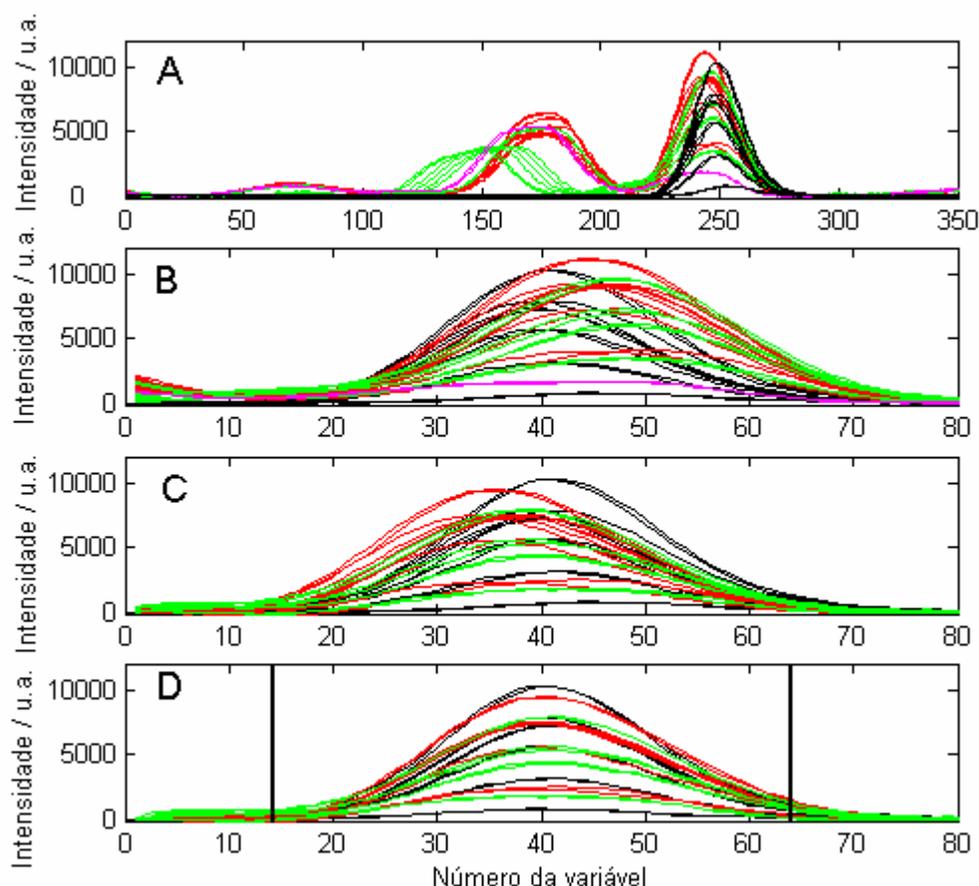
A comparação entre os perfis recuperados pelo PARAFAC e os obtidos experimentalmente nos padrões do MR é apresentada na Figura 86A, onde se observa uma boa concordância. Já os perfis do interferente apresentam grande variação, que em parte provavelmente deve-se à sua baixa intensidade.



**Figura 85-** Cromatogramas em 220 nm para CBZ no MR (A) após a correção da linha de base, (B) após uma correção inicial dos desvios dos tempos das amostras e (C) após alinhamento das amostras de EX e VI em relação padrões. Onde a região selecionada foi aquela utilizada na construção dos modelos, (—) padrões de calibração, (—) EX e (—) VI.



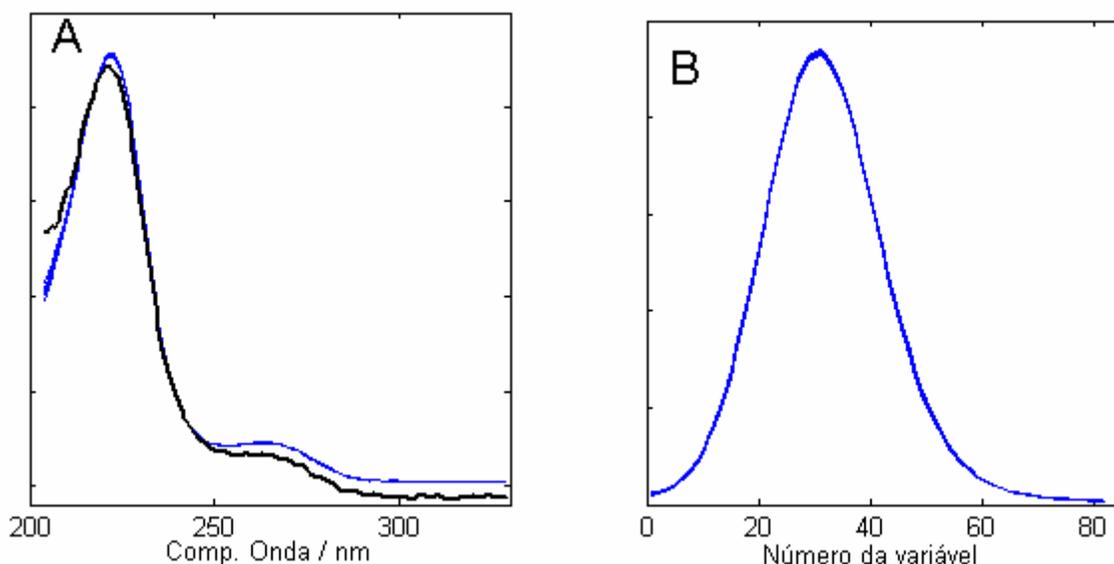
**Figura 86-** Estimativas dos perfis (A) espectrais e (B) cromatográficos, ambos normalizados, obtidos por PARAFAC para a determinação de CBZ nas amostras EX. (—) Perfis experimentais obtidos nos padrões do MR, (—) estimado para FTA e (—) para o interferente.



**Figura 87-** Cromatogramas em 220 nm para SIM no MR (A) após correção da linha de base e um prévio alinhamento dos tempos de retenção de EX e VI, (B) após alinhamento do sinal das amostras EX e VI com o sinal do branco, (C) resultado da subtração do sinal do branco do sinal das amostras EX e VI e (D) alinhamento de EX e VI com os padrões. Onde a região selecionada foi aquela utilizada nas calibrações, (—) padrões, (—) EX, (—) VI e (—) sinal do branco.

A seqüência dos pré-processamentos utilizados para a determinação de SIM é apresentada na Figura 87, onde se observa que inicialmente foi feita a correção da linha de base e um alinhamento inicial das amostras VI, as quais apresentaram um maior desvio (Figura 87A). Em seguida foi selecionada uma menor região e feito o alinhamento das amostras EX, VI em relação ao sinal de uma amostra de branco (Figura 87B) e o sinal do branco foi subtraído do sinal das amostras EX e VI (semelhante ao que foi feito para a determinação de DMT e FTA no MP) (Figura 87C). Este último resultado foi alinhado ao sinal dos padrões (Figura 87D) e

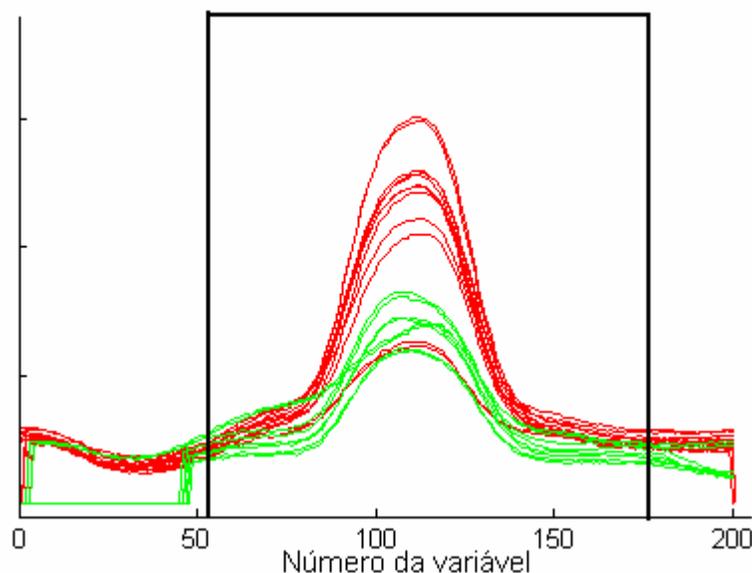
selecionada a região para a construção dos modelos de calibração. Este procedimento de subtração do sinal do branco foi utilizado porque os modelos PARAFAC, BLLS e UPLS foram incapazes de separar esse sinal do sinal do analito, sendo obtidas concentrações sempre superiores em cerca de 30 % as esperadas para as amostras EX. Após este procedimento, com exceção da amostra EX de menor concentração que necessitou de 2 fatores no PARAFAC e 1 fator no RBL com BLLS e UPLS, utilizou-se apenas um fator nos três modelos, como se não houvesse interferentes, sendo obtidos erros médios relativos para as amostras EX de 3,5 % (Tabela 30). A Figura 88 apresenta os perfis recuperados para este analito com apenas um fator no modelo PARAFAC, onde se observa uma boa concordância com o perfil espectral do analito.



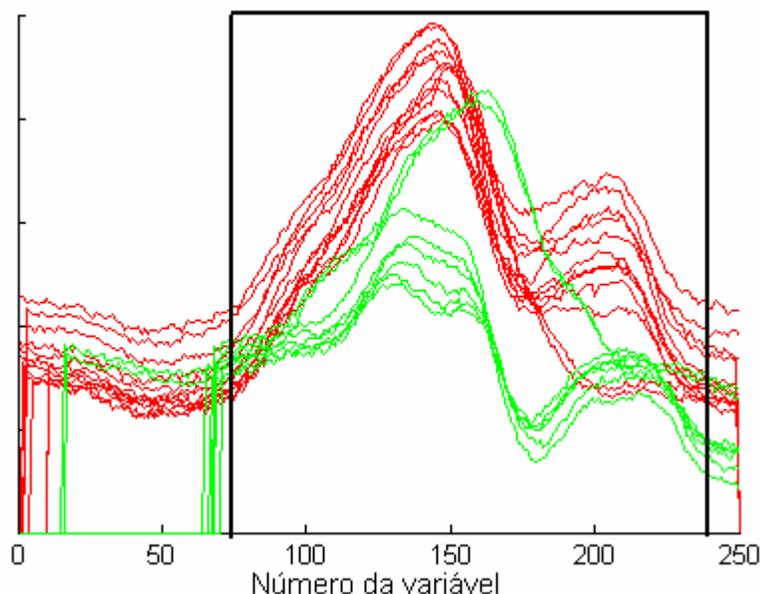
**Figura 88-** Estimativas dos perfis (A) espectrais e (B) cromatográficos, ambos normalizados, obtidos por PARAFAC para a determinação de SIM nas 10 replicatas das amostras EX. (—) perfis experimentais obtidos nos padrões do MR e (—) estimado para SIM.

Os cromatogramas para os analitos FTA e DMT após os pré-processamentos são mostradas nas Figuras 89 e 90. Nessas Figuras observa-se que apenas foi realizado o alinhamento das amostras EX e VI em relação a uma própria amostra EX que foi utilizada como referência, sem a utilização dos padrões. Para esses dois

analitos não foi possível a utilização dos padrões porque foi observado que aparentemente o perfil do pico cromatográfico nas amostras EX e VI foi diferente do perfil observado nos padrões, resultado que impossibilitou a construção dos modelos de calibração com os padrões e a partir deste modelo estimar a concentração dos analitos nas amostras. Isso implica que nessas situações os modelos não apresentaram a vantagem de segunda ordem. Portanto, para possibilitar a determinação desses analitos no MR, foi necessário construir os modelos de calibração com as amostras EX e a partir desse modelo estimar as concentrações na amostras VI, procedimento que se assemelha ao método de adição de padrão, pois a calibração é construída na amostra. Contudo, neste caso espera-se que a concentração dos analitos corresponda apenas à quantidade que foi adicionada de cada um nas amostras EX e VI. Como a calibração para esses analitos foi feita utilizando as amostras EX, as quais não se têm conhecimento de todos os componentes, o modelo BLLS não pôde ser utilizado.



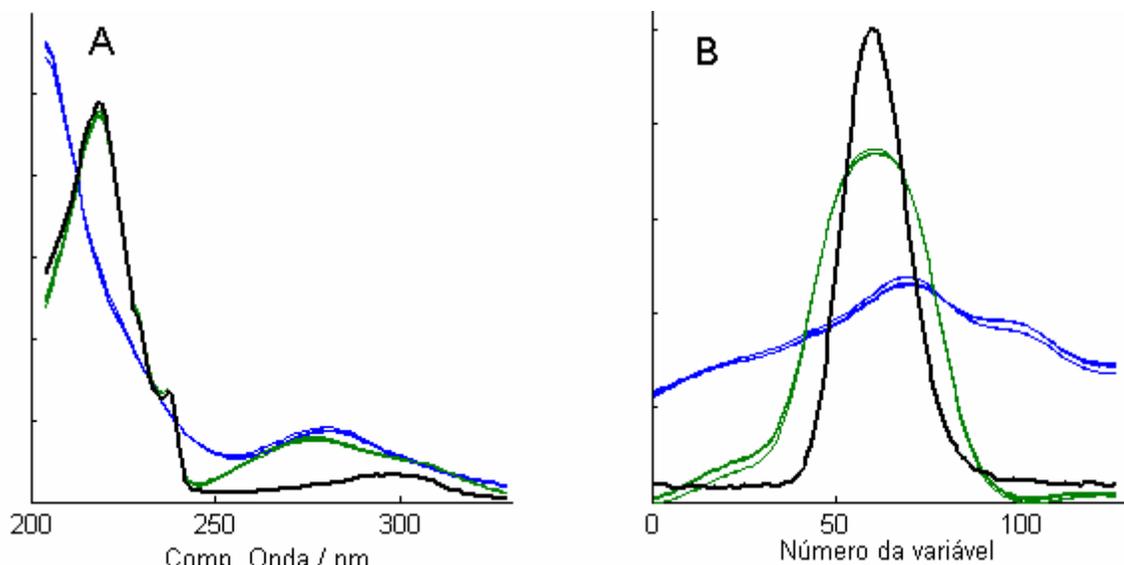
**Figura 89-** Resultado do alinhamento dos tempos de retenção entre as amostras (—) EX e (—) VI para a determinação de FTA pelo MR, onde a região selecionada foi aquela utilizada nos modelos UPLS e PARAFAC.



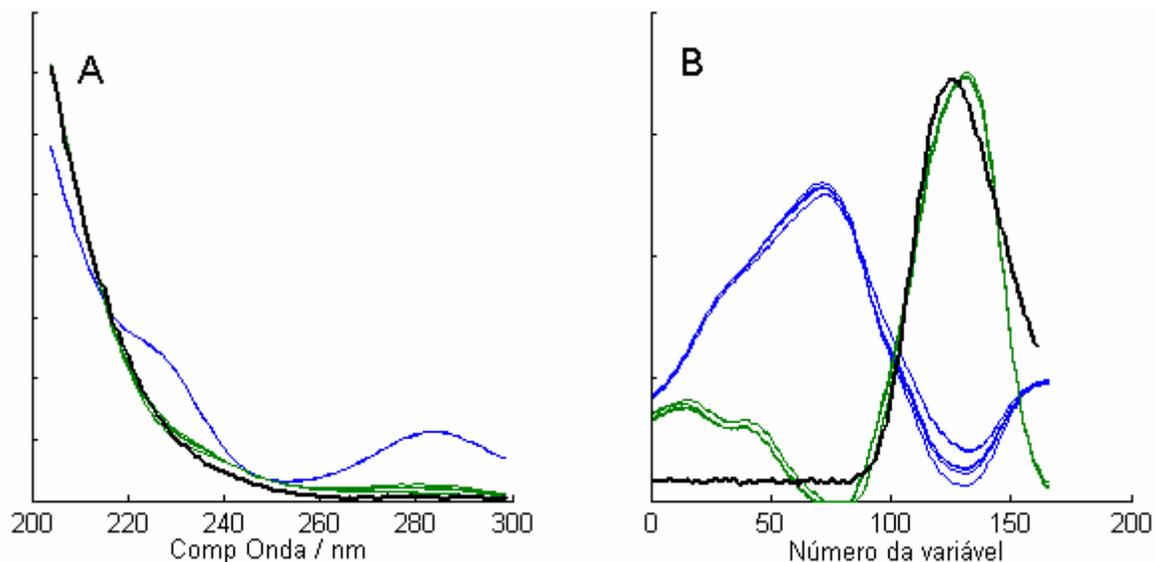
**Figura 90-** Resultado do alinhamento dos tempos de retenção entre as amostras (—) EX, (—) VI para a determinação de DMT pelo MR, onde a região selecionada foi aquela utilizada nos modelos UPLS e PARAFAC.

Utilizando o pré-processamento descrito para os analitos FTA e DMT, para a obtenção dos melhores resultados foram necessários dois fatores no PARAFAC, dois fatores para a construção do modelo UPLS utilizando um fator no RBL para FTA e nenhum para DMT (Tabela 29).

As Figuras 91 e 92 mostram os perfis recuperados pelo PARAFAC para FTA e DMT, respectivamente, onde foi observada uma boa concordância entre os perfis espectrais em ambos os analitos com os perfis obtidos experimentalmente. Quanto aos perfis cromatográficos, para FTA observa-se que os perfis recuperados nas amostras são diferentes dos obtidos nos padrões, o que justifica o fato das calibrações baseadas nos padrões não apresentarem bons resultados. Para DMT, os perfis cromatográficos não se mostraram coerentes, pois apresentam o analito em duas regiões distintas no cromatograma.



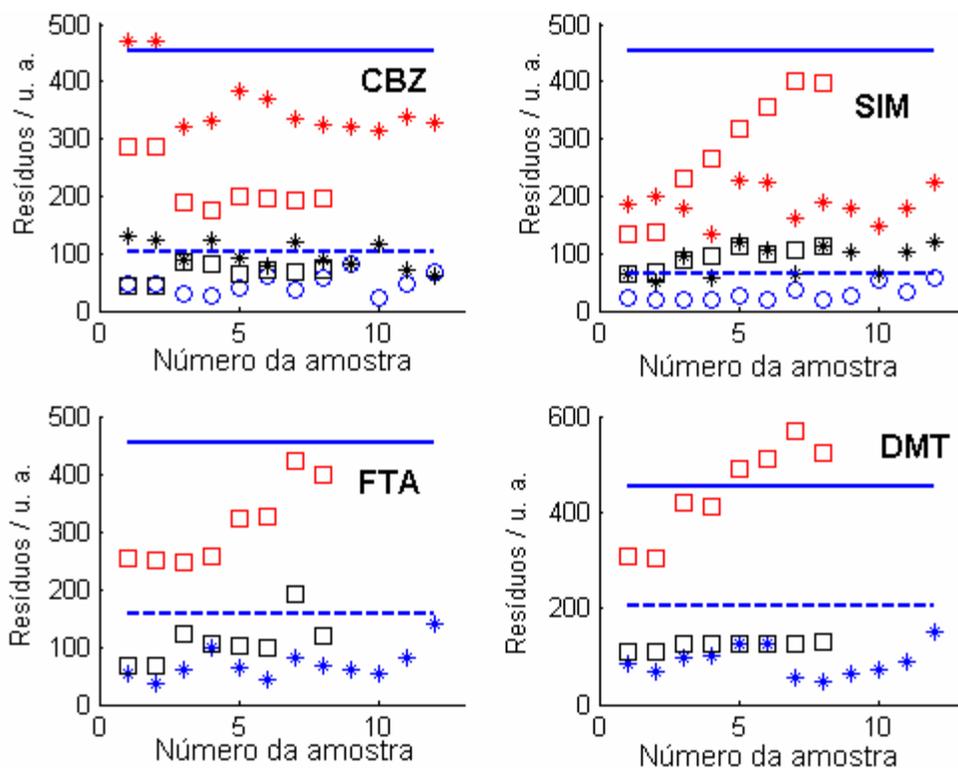
**Figura 91-** Estimativas dos perfis (A) espectrais e (B) cromatográficos, ambos normalizados, obtidos por PARAFAC para a determinação de FTA nas amostras EX e VI. (—) Perfis experimentais obtidos nos padrões do MR, (—) estimado para FTA e (—) para o interferente estimado pelo RBL.



**Figura 92-** Estimativas dos perfis (A) espectrais e (B) cromatográficos, ambos normalizados, obtidos por PARAFAC para a determinação de DMT nas amostras EX e VI. (—) Perfis experimentais obtidos nos padrões do MR, (—) estimado para FTA e (—) para o interferente estimado pelo RBL.

Os gráficos de variância residual versus os deslocamentos para os analitos CBZ, SIM, DMT e FTA sempre atingiram uma região de mínimo em todos os alinhamentos que foram realizados no MR, sugerindo que o alinhamento foi bem sucedido em todos os casos.

Na Figura 93 são apresentados os gráficos de resíduos de decomposição obtidos pelo modelo UPLS para as quatro situações do MR. Neste caso observa-se que os limites baseados no ruído instrumental foram sempre superiores aos baseados no resíduo médio das amostras de calibração. Baseando-se os cálculos dos limites no ruído, não seria necessário adição de nenhum fator ao RBL em nenhum caso, enquanto que pelos limites obtidos pelas amostras de calibração seria necessário um fator em todas as quatro situações. Como pela Tabela 29 os melhores resultados foram obtidos sem utilizar o RBL nos analitos SIM e DMT e utilizando 1 fator no RBL nos analitos CBZ e FTA, conclui-se que nessas quatro situações os limites propostos não são eficazes na escolha do número de fatores do RBL. O mesmo comportamento foi observado nos resíduos para o BLLS nos analitos SIM e CBZ.



**Figura 93-** Resíduos de decomposição obtidos com o modelo UPLS no MR. Onde: (o) padrões, (\*) EX, (□) VI, (vermelho) Nenhum e (preto) um fator no RBL. (—) Valor crítico pelo ruído instrumental e (---) pelo resíduo médio das amostras de calibração, ambos com 99% de confiança.

A Tabela 30 mostra os resultados obtidos para o MR. Como no MP, os resultados são praticamente os mesmos nos três modelos, sendo a grande distinção do UPLS e do PARAFAC em relação ao BLLS a possibilidade de sua construção quando não se tem o conhecimento de todos os componentes nas amostras utilizadas na etapa de calibração, como no caso dos compostos DMT e FTA.

**Tabela 30.** Resultados das figuras de mérito obtidas para a determinação dos pesticidas e metabólitos pelo método de referência (MR).

	RMSEP <sup>a</sup>		REP <sup>b</sup>		Precisão <sup>a</sup>		SEL		LD <sup>c</sup>	
	CAL	EX	CAL	EX	CAL	EX	CAL	EX	CAL	EX
<b>BLLS</b>										
DMT	---	---	---	---	---	---	---	---	---	---
FTA	---	---	---	---	---	---	---	---	---	---
SIM	0,01	0,02	1,9	3,4	0,004	0,01	1	0,3(±0,1)	4,5	14(±5)
CBZ	0,06	0,31	1,3	9,0	0,03	0,03	1	0,33(±0,03)	8,3	25(±2)
<b>UPLS</b>										
DMT	---	0,24	---	5,4	---	0,11	--	---	---	43
FTA	---	0,02	---	2,6	---	0,01	---	---	---	3,3
SIM	0,01	0,02	2,0	3,5	0,004	0,01	--	0,3(±0,1)	4,5	14(±4)
CBZ	0,06	0,31	1,3	8,9	0,03	0,03	--	0,33(±0,03)	8,3	25(±2)
<b>PARAFAC</b>										
DMT	---	0,17	---	6,3	---	0,20	---	0,32	---	115
FTA	---	0,03	---	9,6	---	0,01	---	0,28	---	9,7
SIM	0,01	0,02	2,4	1,6	0,006	0,01	1	1	4,5	4,5
CBZ	0,06	0,27	1,6	7,4	0,03	0,03	1	0,29(±0,02)	8,3	29(±2)
<b>Univariado</b>										
DCA	0,06	0,24	1,0	3,6	0,05	0,08	1	1	11	11
TIO	0,10	0,23	14	5,8	0,02	0,02	1	1	13	13
CBL	0,08	0,41	3,4	10	0,02	0,02	1	1	3,5	3,5

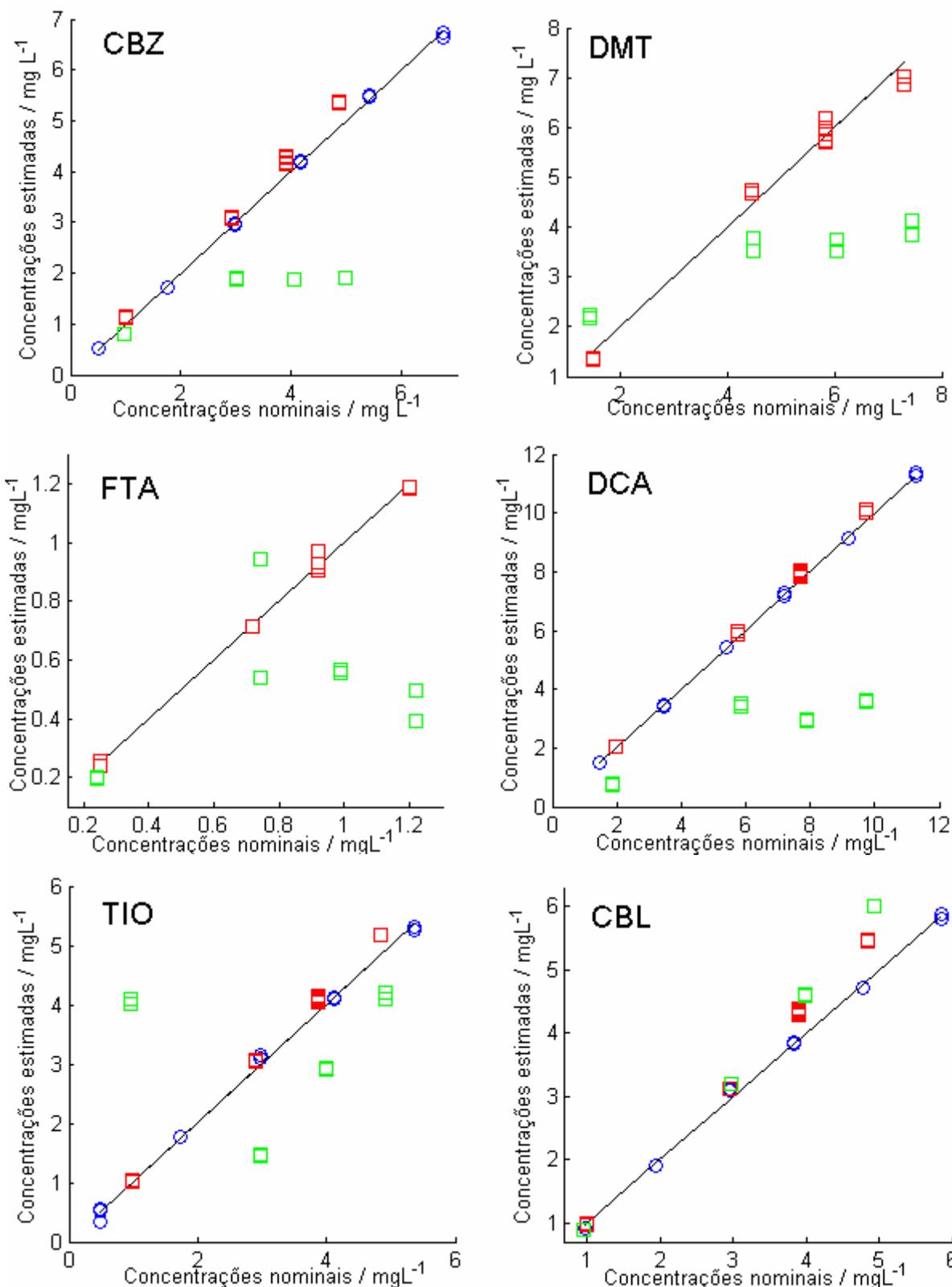
Grandezas expressas por: (a) mg/L, (b) percentagem e (c) 10<sup>-3</sup> mg/L.

Nas Figuras 94 e 95 pode-se observar que o ajuste dos valores obtidos pelo modelo UPLS e por calibração de ordem zero nas amostras EX, foi adequado para quase todos os compostos de interesse, sendo as concentrações estimadas para o analito CBL as que mais se distanciam dos valores esperados. Pelos valores da inclinação e do intercepto da regressão linear entre os valores nominais e estimados para as amostras EX e seus respectivos intervalos de confiança, apresentados na Tabela 31, pode-se concluir que, além do analito CBL, os compostos TIO e CBZ também apresentam um erro sistemático proporcional positivo significativo no nível de 99 % de confiança, uma vez que as inclinações estimadas são maiores que o valor esperado 1,00 teoricamente.

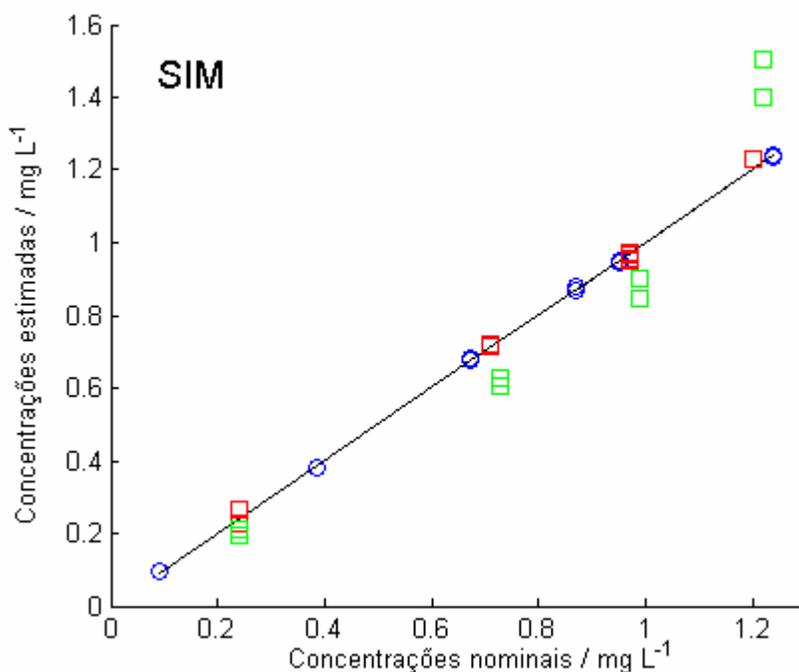
**Tabela 31.** Resultados para as inclinações e interceptos das regressões lineares entre os valores de referência a estimados para as amostras EX no MR.

Analito	Inclinação	Intercepto
DMT <sup>a</sup>	0,98(±0,13)	0,09(±0,70)
FTA <sup>a</sup>	1,00(±0,06)	0,00(±0,05)
CBZ <sup>a</sup>	1,09(±0,06)	-0,02(±0,21)
SIM <sup>a</sup>	1,00(±0,05)	0,00(±0,05)
TIO <sup>b</sup>	1,07(±0,02)	-0,05(±0,08)
DCA <sup>b</sup>	1,03(±0,03)	0,03(±0,19)
CBL <sup>b</sup>	1,17(±0,05)	-0,23(±0,18)

Valores estimados por (a) UPLS e (b) calibração de ordem zero. Intervalos de confiança com 99 % entre parênteses, com 10 graus de liberdade.



**Figura 94-** Valores de referência versus os estimados pelo MR com o UPLS. (—) Bissetriz do quadrante, (●) padrões de calibração, (□) amostras EX e (□) VI.



**Figura 95-** Valores de referência versus os estimados para SIM pelo MR com o UPLS. (—) Bissetriz do quadrante, (●) padrões de calibração, (◻) amostras EX e (◻) VI.

### 5.3.3 Comparação entre os métodos propostos (MP) e de referência (MR).

Considerando que os melhores resultados foram obtidos no MP e no MR pelos modelos BLLS e UPLS, respectivamente, observa-se pelas Tabelas 30 e 27 que em ambos os métodos obtiveram-se erros médios menores que 10 % na análise dos extratos de todos os compostos de interesse, o que pode ser considerado um bom resultado considerando a complexidade da amostra. Contudo, pelos valores estimados das inclinações e interceptos das regressões entre os valores nominais e estimados, e seus intervalos de confiança, foram obtidos melhores resultados no MP, uma vez que não foram observados erros sistemáticos significativos no nível de 99% de confiança. Ainda pelos valores de limite de detecção estimados, o MP também é mais sensível para os analitos DMT, SIM e CBZ.

**Tabela 32.** Valores obtidos para as recuperações percentuais nas amostras EX e VI comparando os valores obtidos no MP com as concentrações esperadas dos analitos que foram adicionados.

Extratos (EX) <sup>a</sup>				Vinho (VI) <sup>b</sup>		
Analito	BLLS	UPLS	PARAFAC	BLLS	UPLS	PARAFAC
DMT	100(±5)	100(±7)	103(±11)	90(±26)	91(±24)	93(±20)
FTA	106(±6)	107(±5)	106 (±6)	74(±21)	77(±21)	38(±35)
TIO	96(±4)	94(±8)	95(±6)	59(±13)	57(±15)	57(±15)
SIM	92(±2)	92(±2)	93(±2)	96(±7)	94(±6)	96(±6)
CBL	98(±2)	98(±2)	99(±2)	100(±3)	100(±3)	100(±3)
DCA <sup>c</sup>	105(±2)			41(±9)		
CBZ <sup>c</sup>	106(±3)			54(±16)		

Desvios padrões entre parênteses com (a) 6 e (b) 4 graus de liberdade, com exceção de SIM e FTA que são (a) 5 e (b) 3, respectivamente.(c) Determinados por calibração de ordem zero.

**Tabela 33.** Valores obtidos para as recuperações percentuais nas amostras EX e VI comparando os valores obtidos no MR com as concentrações esperadas dos analitos que foram adicionados.

Extratos (EX) <sup>a</sup>				Vinho (VI) <sup>b</sup>		
Analito	BLLS	UPLS	PARAFAC	BLLS	UPLS	PARAFAC
DMT	---	99(±6)	98(±6)	---	65(±15)	58(±13)
FTA	---	100(±2)	104(±10)	---	69(±28)	49(±19)
CBZ	109(±3)	109(±3)	107(±23)	57(±20)	57(±20)	50(±21)
SIM	101(±11)	101(±11)	102(±12)	85(±3)	85(±2)	80(±19)
TIO <sup>c</sup>	106(±1)			69(±16)		
DCA <sup>c</sup>	103(±1)			44(±10)		
CBL <sup>c</sup>	108(±5)			106(±10)		

Desvios padrões entre parênteses com (a) 6 e (b) 4 graus de liberdade, com exceção de TIO, CBL e SIM que são (a) 6 e (b) 3, respectivamente.(c) Determinados por calibração de ordem zero.

As Tabelas 32 e 33 mostram as recuperações obtidas para nas amostras EX e VI nos MP e MR, respectivamente. Como esperado, nas amostras EX foram

obtidas recuperação próximas de 100% e com desvios padrões relativamente baixos em ambos os métodos. Pelas recuperações médias, nas amostras VI também foi observada uma boa concordância entre os métodos, sendo as maiores discordâncias observadas para DMT e FTA, que também foram os compostos que apresentaram os maiores problemas em sua determinação tanto no MP quanto no MR.

Avaliando o método de extração empregado pelos resultados das amostras VI, observam-se menores recuperações e maiores desvios padrões, que refletem as perdas do método de extração e pela sua baixa reprodutibilidade em alguns analitos. Os melhores resultados foram obtidos para os pesticidas CBL e SIM, no qual o método pode ser considerado eficiente. Para DCA este método de extração pode ser considerado de baixa eficiência, sendo mediana para os demais analitos.

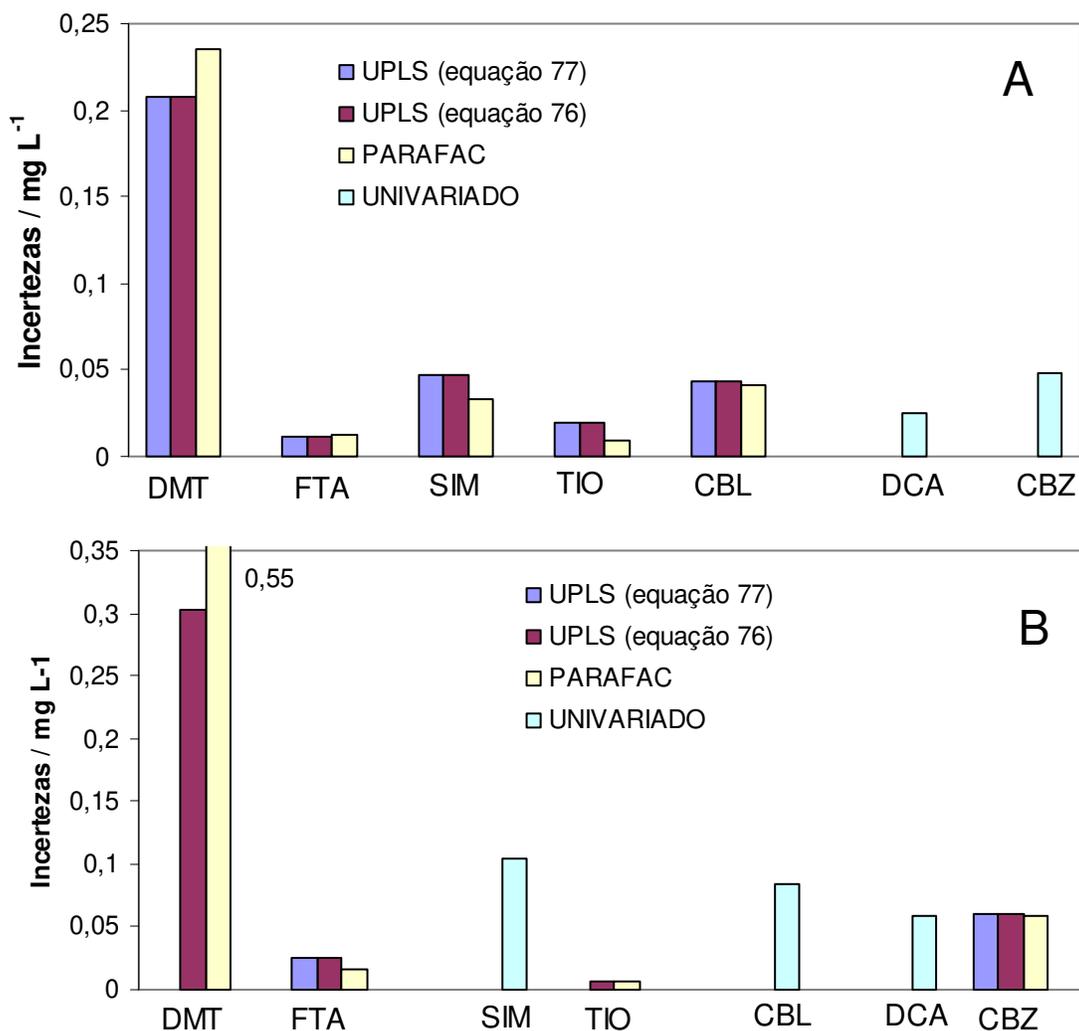
As incertezas médias estimadas são mostradas na Figura 96, onde se observa que as maiores diferenças entre os métodos ocorrem para os compostos DMT e TIO, sendo o MP com menores incertezas para DMT e o MR para TIO. Para os analitos onde se obteve estimativas com os modelos UPLS e PARAFAC, não foi observada diferença significativa. Para o UPLS foram obtidos praticamente os mesmos resultados com as equações 76 e 77 em todas as situações.

## **5.4 Conclusões do capítulo**

A partir dos resultados apresentados neste capítulo pôde-se avaliar a aplicação dos modelos BLLS, UPLS e PARAFAC em dados obtidos pela técnica HPLC-DAD. Como nas demais aplicações, foi observado que os resultados obtidos pelos três modelos foram equivalentes na grande maioria dos casos, sendo uma das diferenças às amostras anômalas que foram observadas no modelo PARAFAC e não nos outros dois modelos. Outra diferença relevante foi a impossibilidade da determinação dos compostos DMT e FTA no MR pelo BLLS, fato que evidenciou sua maior limitação em relação aos outros modelos, que é a necessidade do

conhecimento de todos os componentes das amostras na etapa de calibração.

Neste estudo também foi observada a aplicação de um procedimento para a correção dos desvios dos tempos de retenção proposto por Prazen et al.<sup>27</sup> e que foi implementado em um programa desenvolvido nesse trabalho. As correções ou alinhamentos dos tempos de retenção mostraram-se eficientes na maior parte dos casos, sendo observados erros apenas nas amostras em que os analitos encontravam-se em baixa concentração e na presença de interferentes.



**Figura 96-** Incertezas médias para cada um dos compostos de interesse, estimadas para os (A) MP e (B) MR.

Os limites propostos para a estimativa do número de fatores no RBL não

foram eficientes nessa aplicação na maior parte dos resultados, sendo a principal hipótese para isso a imperfeição das correções dos desvios de tempo de retenção, que embora foram suficientes para possibilitar bons resultados nas estimativas de concentração dos analitos, talvez não tenham sido adequadas o suficiente para fazer com que os resíduos de decomposição das amostras obedecessem os limites quando o número de fatores correto era utilizado no RBL. Este resultado indica que, assim como nos dados obtidos por FIA, deve-se ter um conjunto de amostras externas em que se saiba o valor das concentrações dos analitos para se ter certeza da escolha do número correto de fatores no RBL.

Quanto aos resultados de seletividade estimados para o modelo UPLS, estes mostraram-se novamente concordantes com os obtidos para os outros modelos, enquanto que as estimativas de incertezas para um mesmo analito em diferentes modelos não apresentaram diferenças significativas.

Ambos os métodos avaliados (MP e MR) mostraram-se capazes de determinar todos os compostos de interesse com erros médios abaixo de 10 %. Contudo o MP pode ser considerado mais eficiente em diversos aspectos, tais como: (1) apresentar um tempo de análise de aproximadamente 3,5 vezes menor; (2) utilizar um método cromatográfico por eluição isocrática, que requer uma instrumentação mais simples; (3) possibilita construção de modelos de calibração com padrões em solução e a estimativa da concentração dos analitos na amostra, isto é a vantagem de segunda ordem foi adquirida pelos modelos em todas as situações necessárias; e (4) as etapas de pré-processamento utilizadas no MP podem ser consideradas mais simples que as utilizadas no MR.

O método de extração em fase sólida que foi utilizado mostrou-se eficiente apenas para os analitos CBL e SIM. De forma que esta etapa do preparo da amostra deve ser otimizada para a aquisição de melhores valores de recuperação para os demais compostos.

## **Capítulo 6 - Conclusões gerais**

Pelos resultados apresentados nesta tese de doutorado foi possível verificar em quatro estudos de caso, com diferentes conjuntos de dados, que os modelos de calibração de segunda ordem BLLS e UPLS são capazes de fornecer estimativas para as concentrações de espécies de interesse equivalentes às obtidas pelos modelos PARAFAC ou MCR, modelos que já tem sua aplicação mais difundida em grupos de pesquisa de quimiometria. Esta capacidade foi verificada principalmente em situações em que foi observada a presença de interferentes nas amostras analisadas pelo modelo de calibração, onde o método RBL mostrou-se eficiente na implementação da vantagem de segunda ordem aos modelos BLLS e UPLS. Contudo, apesar dos bons resultados na grande maioria das situações, foram observados casos onde o RBL apresentou divergência, indicando que melhoramentos no seu algoritmo ainda são necessários.

Conclui-se que, apesar da semelhança nos resultados observados nos estudos de caso apresentados, que o modelo BLLS foi o mais eficiente do ponto de vista quantitativo, sendo um modelo mais simples de ser construído que os modelos PARAFAC, MCR e UPLS. Contudo, este apresenta como desvantagem a impossibilidade de realizar determinações quando não se tem conhecimento de todos os componentes presentes nas amostras de calibração, limitação não apresentada pelos outros três modelos citados. Do ponto de vista qualitativo o BLLS foi capaz de fornecer boas estimativas dos perfis dos analitos e interferente nos dados de EEF e HPLC. Entretanto, apresentou-se muito inferior em relação ao MCR nos dados de FIA com gradiente de pH, sendo este modelo o mais eficiente quando uma resolução qualitativa também for necessária nesse tipo de dados.

O limite proposto para a escolha do número de fatores para o RBL baseado no resíduo médio das amostras de calibração, com 99% de confiança, mostrou-se eficiente no modelo BLLS em quase todos os modelos construídos com dados de EEF. Logo, concluí-se que para a aplicação do BLLS utilizando o RBL nesse tipo

de dados este limite pode ser utilizado como um guia na escolha do número de fatores. Contudo, os resultados obtidos na análise de dados obtidos por FIA com gradiente de pH e por HPLC (dados estes que estão sujeitos a variações no gradiente de pH utilizado ou nos tempos de retenção entre diferentes amostras) sugerem que tanto o limite baseado nas amostras de calibração como no ruído instrumental podem levar à escolha de um número não adequado de fatores no RBL. Logo, é prudente sugerir que sempre que este modelo for aplicado seja utilizado um conjunto de amostras externas (validação externa), nas quais se conheça o valor das concentrações dos analitos, para a confirmação do número de fatores utilizado.

Para a aplicação do RBL ao modelo UPLS a utilização dos limites tanto baseados no ruído instrumental como do resíduo médio das amostras de calibração não foram eficientes para guiar a escolha do número de fatores. Logo, a utilização de um conjunto de amostras externas também é sugerida na utilização do RBL com o UPLS em qualquer tipo de dados.

Foi possível avaliar a consistência das figuras de mérito: sensibilidade analítica, limite de detecção, seletividade, exatidão, precisão e incerteza obtidas por calibração de segunda ordem e os valores obtidos por métodos independentes. Para os modelos PARAFAC, BLLS e UPLS pode-se concluir que a proposta mais recente para a determinação de sensibilidade<sup>52</sup> forneceu valores coerentes com os intervalos de concentração estudados nos quatro estudos de caso abordados. Ela é capaz de levar em conta a influência dos interferentes presentes em amostras de composição desconhecida.

Os resultados obtidos para a proposta de determinação de seletividade no modelo UPLS, quando este utiliza a vantagem de segunda ordem, sugerem que esta é coerente com a sugestão mais atual de medida de seletividade para os modelos BLLS e PARAFAC, o que torna possível a determinação desse parâmetro nesse

modelo. Entretanto ainda são necessárias pesquisas sobre este parâmetro para que os resultados atendam a definição da IUPAC.

Quanto às estimativas de incerteza para o UPLS, observou-se que equação proposta para situações em que o UPLS utiliza a vantagem de segunda ordem (expressa pela equação 77) é equivalente na grande maioria dos casos à equação 76 proposta por Faber et al<sup>96</sup> e é utilizada em modelos de calibração de primeira ordem. Contudo, para uma situação em que a presença de interferentes influenciou muito na sensibilidade do modelo, foi observado que a equação 77 fornece uma melhor medida da incerteza do modelo. Logo concluí-se que quando a influência dos interferentes na sensibilidade não for alta ( $SEL > 0,1$ ) e o ruído dos dados for baixo a equação 76 pode ser utilizada com o UPLS utilizando a vantagem de segunda ordem. Entretanto, em um caso específico onde  $SEL < 0,1$  e o ruído dos dados for alto, a equação 77 proposta deve fornecer melhores estimativas de incerteza.

Quanto à estimativa de incertezas com o modelo PARAFAC, concluí-se que estas podem ser estimadas de forma análoga a modelos de calibração de ordem zero quando o modelo PARAFAC for construído decompondo simultaneamente os padrões e as amostras a serem analisadas.

Na determinação de pesticidas e metabólitos por amostras de vinho tinto foi aplicado e verificada a eficiência do método proposto por Prazen et al. para a correção dos desvios de tempo de retenção em dados obtidos por HPLC-DAD. Portanto, concluí-se que este método deve ser sugerido para esse tipo de correção em novos estudos com desvios similares.

Os programas desenvolvidos nesta tese para: correção de desvios de tempos de retenção, construção dos modelos BLLS e UPLS (ambos utilizando o RBL) e de previsão e determinação de figuras de mérito para o modelo PARAFAC, mostraram-se eficientes em todas as aplicações. Pode-se ainda destacar que o

programa desenvolvido para o modelo BLLS já foi utilizado na dissertação de mestrado de Renato L. Carneiro, também orientada pelo Prof. Dr. Ronei J. Poppi<sup>122</sup>, apresentando bons resultados. Os programas desenvolvidos são apresentados nos apêndices B da tese e são compatíveis com os ambientes Octave 3.0.0 e Matlab 6.5, de forma que possibilitam a utilização de um sistema livre para o desenvolvimento desses modelos. Os resultados obtidos em ambos os ambientes foram equivalentes em quase todas as situações descritas nessa tese.

Por fim concluí-se que os resultados dos estudos de caso descritos ressaltam as vantagens da aplicação de métodos de calibração de segunda ordem na determinação da concentração de espécies de interesse em um meio complexo. Espera-se que os resultados e discussões apresentadas, assim como os programas desenvolvidos, possam contribuir para o desenvolvimento, a difusão e a aplicação desses modelos, de modo que no futuro estes possam ser utilizados com segurança em análises de rotina trazendo benefícios como a redução dos tempos de análise, minimizar o uso de reagentes e geração de resíduos, além de facilitarem ou possibilitarem a determinação de espécies de interesse em sistemas complexos, onde não sejam conhecidos os interferentes presentes na amostra.

## **Capítulo 7 - Perspectivas futuras**

Nos últimos 5 anos a redescoberta do método RBL possibilitou a implementação da vantagem de segunda ordem a modelos que originalmente não apresentavam essa propriedade, fato que tem incentivado a pesquisa na utilização do BLLS e UPLS e a proposta de novos modelos de calibração para ordens superiores, como por exemplo o TLLS e o NPLS com RTL<sup>53</sup>. Contudo, devido a sua proposta e implementação ser relativamente recente e pelos aparentes problemas de convergência observados no primeiro estudo de caso dessa tese, ainda é necessário que se pesquise e compare outros algoritmos para a sua implementação nesses modelos procurando dessa forma minimizar a ocorrência de divergências do método.

O método RBL, em seu algoritmo, estima os perfis dos interferentes com dados na forma matricial. Entretanto, curiosamente esse método foi aplicado até o momento apenas a modelos que trabalham com as matrizes de dados na forma desdobrada (BLLS e UPLS) e não na forma matricial, como por exemplo, a sua implementação ao N-PLS, o que seria aparentemente o raciocínio mais direto. Em um artigo recente, Olivieri et al.<sup>123</sup> apenas comentam terem conseguido a união desses dois modelos, indicando nesse trabalho que o programa encontra-se em fase de testes e que resultados serão apresentados em breve. Ao propor o NPLS, Bro compara e sugere que este é superior ao UPLS por utilizar os dados em sua forma original (sem realizar o desdobramento), logo a utilização do RBL como o NPLS provavelmente deve trazer vantagens em relação à sua utilização com o UPLS. A implementação do RBL ao NPLS para incorporar a vantagem de segunda ordem e o estudo de suas propriedades são com certeza temas que devem ser verificados em trabalhos futuros.

Como discutido brevemente nesse trabalho, a estimativa de seletividade atualmente empregada para modelos de calibração de segunda ordem não preenche um requisito importante e que está presente na definição da IUPAC para esse

parâmetro. Requisito este que é não apresentar variações em seus valores em função da concentração de um interferente presente na amostra, variando apenas devido a sua presença ou ausência. Essa é uma propriedade que aparentemente tem desafiado os pesquisadores em quimiometria com trabalhos desenvolvidos nesse tema. Portanto, a busca de um método para a determinação desse parâmetro que leve em conta tanto a presença quanto a intensidade dos interferentes em uma amostra ainda é necessária, de modo que este parâmetro possa informar esses efeitos em um método e preencha a sua definição pela IUPAC.

Quanto à proposta de determinação do erro padrão das estimativas de concentração para o modelo UPLS utilizando o RBL, devem ser realizados novos estudos para verificar a sua eficácia quando à ação dos interferentes na amostra interferirem significativamente na sensibilidade ou seletividade da amostra que é analisada. Para isso, devem ser utilizados conjuntos de dados, tanto simulados como experimentais, com um número relativamente grande de amostras ou replicatas e a obtenção de uma medida independente do erro padrão para comparação, o que pode ser conseguido por métodos numéricos ou de reamostragem, tais como métodos de adição de ruído ou *Bootstrap*.

Do ponto de vista analítico, os modelos de calibração de segunda ordem oferecem uma possibilidade imensa de aplicações com amostras complexas. Devem-se desenvolver novas aplicações com dados de origem cromatográfica, principalmente de forma conjunta com grupos de pesquisa em cromatografia, difundindo desta forma esses modelos a esses grupos e introduzindo a possibilidade da sua utilização em sistemas de difícil separação.

Outro resultado que foi observado deve ser comprovado em novos estudos é a eficácia do MCR em corrigir o efeito Schlieren em aplicações com FIA que gerem matrizes de dados.

## **Referências**

- [1] M. Otto, *Chemometrics*, Wiley, Weinheim, **1999**, cap. 1.
- [2] K. Danzer e L. A. Currie; *Guidelines for calibration in analytical chemistry – Part 1. Fundamentals and single component calibration*. Pure & Appl. Chem. **1998**, 70, 993.
- [3] B. Barros Neto, M.F. Pimentel e M.C.U. Araújo; *Recomendações para Calibração em Química Analítica – Parte I. Fundamentos de Calibração com um Componente (Calibração Univariada)*. Quím. Nova, **2002**, 25, 856.
- [4] K. S. Booksh e B. R. Kowalski; *Theory in Analytical Chemistry*. Anal. Chem. **1994**, 66, 782A.
- [5] Near-infrared spectrometry, Chapter 1119, United States Pharmacopoeia USP28 NF23, **2006**, 2691.
- [6] Raman Spectrometry, Chapter 1120, United States Pharmacopoeia USP28 NF23, **2006**, 2983.
- [7] Annual Book of ASTM Standards, E1655-05: *Standards practices for infrared multivariate quantitative analysis*, American Society for testing and Materials International, West Conshohocken, PA, 2005.
- [8] K. Danzer, M. Otto e L. A. Currie; *Guidelines for calibration in analytical chemistry part 2. Multispecies calibration*. Pure Appl. Chem. **2004**, 76, 1215.
- [9] A. C. Olivieri, N. M. Faber, J. Ferré, R. Boqué, J. H. Kalivas e H. Mark; *Uncertainty estimation and figures of merit for multivariate calibration*. Pure Appl. Chem. **2006**, 78, 633.
- [10] R. Bro; *PARAFAC. Tutorial and applications*. Chemom. Intell. Lab. Syst. **1997**, 38, 149.
- [11] R. Tauler, A. Smilde e B.R Kowalski; *Selectivity, local rank, 3-way data analysis and ambiguity in multivariate curve resolution*. J. Chemom. **1995**, 9, 31.
- [12] P. C. Damiani, A. J. Nepole, M. Bearzotti e A. C. Olivieri; *A test field for the second-order advantage in bilinear least-squares and parallel factor analyses: fluorescence determination of ciprofloxacin in human urine*. Anal. Chem. **2004**, 76, 2798.

- [13] M. J. Culzoni, H. C. Goicoechea, A. P. Pagani, M. A. Cabezón e A. C. Olivieri. *Evaluation, of partial least-squares with second-order advantage for the mult-way spectroscopy analysis of complex biological samples in the presence of analyte-background interactions*. *Analyst*, **2006**, 131, 718.
- [14] <http://www.models.kvl.dk/source/> – acessada em 30 de novembro de 2007.
- [15] <http://www.ub.edu/mcr/ndownload.htm> – acessada em 30 de novembro de 2007.
- [16] MATLAB 6.5, The Mathworks Inc. Natick, MA, 2000.
- [17] Gnu Octave 3.0.0, John W. Eaton and others, University of Wisconsin, Department of Chemical Engineering, Madison, 2007.
- [18] J. N. Miller e J. C. Miller; *Statistics and chemometrics for analytical chemistry*. 4<sup>a</sup> ed. Prentice Hall, New York, **2000**, cap.5.
- [19] B. de Barros Neto, I.S. Scarmínio e R. E. Bruns; *Como Fazer Experimentos: Pesquisa e desenvolvimento na ciência e na industria*, 2<sup>a</sup> ed., Editora da UNICAMP: Campinas, **2001**, cap. 5.
- [20] D. A. Burns e E. W. Ciurczak; *Handbook of Near-Infrared Analysis*, 2<sup>a</sup> ed., Marcel Dekker, New York, **2001**, cap. 7.
- [21] A. Lorber, K. Faber e B. R Kowalski; *Net analyte signal calculation in multivariate calibration*. *Anal. Chem.* **1997**, 69, 1620
- [22] M. Laasonen, T. H. Pulkkinen, C. Simard, M. Rasanen e H. Vuorela; *Development and validation of a near-infrared method for the quantitation of caffeine in intact single tablets*. *Anal. Chem.* **2003** 75, 754
- [23] J. W. B. Braga e R. J. Poppi; *Figures of merit for the determination of polymorphic purity of carbamazepine by infrared spectroscopy and multivariate calibration*. *J. Pharm. Sci.* **2004**, 93, 2124.
- [24] L. Kiers, J. M. F. Ten Berge e R. Bro; *PARAFAC2 – Part I. A direct fitting algorithm for PARAFAC2 model*. *J. Chemom.* **1999**, 13, 275.
- [25] R. Bro, C. Andersen e H. A. L. Kiers; *PARAFAC2 – Part II. Modeling chromatographic data with retention time shifts*. *J. Chemom.* **1999**, 13, 295.

- [26] J. M. Dueñas Cueva, A. V. Rossi e R. J. Poppi; *Modeling kinetic spectrophotometric data of aminophenol isomers by PARAFAC2*. Chemom. Intell Lab. Syst. **2001**, 55, 125.
- [27] B. J. Prazen, R. E. Synovec e B. R. Kowalski; *Standardization of second-order chromatographic/spectroscopic data for optimum chemical analysis*. Anal. Chem. **1998**, 70, 218.
- [28] C. G. Fraga, B. J. Prazen e R. E. Synovec; *Objective data alignment and chemometric analysis of comprehensive two-dimensional separations with run-to-run peak shifting on both dimensions*. Anal. Chem. **2001**, 73, 5833.
- [29] E. Comas, J. Ferré e F. X. Rius; *Graphical criterion for assessing trilinearity and selecting the optimal number of factors in the generalized rank annihilation method using liquid chromatography-diode array detection data*. Anal. Chim. Acta, **2004**, 515, 23.
- [30] E. Sanches e B. R. Kowalski; *Tensorial resolution: a direct trilinear decomposition*. J Chemom. **1990**, 4, 29.
- [31] R. Bro, *Multi-way analysis in the food industry – Models, algorithms and applications*. Tese de doutorado, Universidade de Amsterdam, Holanda, **1998**.
- [32] N. M. Faber, R. Bro e P. K. Hopke; *Recent developments in CANDECOMP/PARAFAC algorithms: a critical review*. Chemom. Intell Lab. Syst. **2003**, 65, 119.
- [33] A. M. De La Pena, A. E. Mansilla, D. G. Gómez, A. C. Olivieri e H. C. Goicoechea; *Interference-free analysis using three-way fluorescence data and the parallel factor model. Determination of fluoroquinone antibiotics in human serum*. Anal. Chem. **2003**, 75, 2640.
- [34] R. Tauler, A. Smilde e B. Kowalski; *Selectivity, local rank, 3-way data-analysis and ambiguity in multivariate curve resolution*. J. Chemom. **1995**, 9, 31.
- [35] J. Diewok, A. Juan, M. Maeder, R. Tauler e B. Lendl; *Quantitation of mixtures of diptotic organic acids by FT-IR flow titrations and multivariate curve resolution*. Appl. Spectrosc. **2002**, 56, 40.
- [36] J. Diewok, A. Juan, M. Maeder, R. Tauler e B. Lendl; *Application of a combination of hard and soft modeling for equilibrium systems to the quantitative analysis of pH-modulated mixture samples*. Anal. Chem. **2003**, 75, 641-647.

- [37] J. Saurina, S. Hernández-Cassou, R. Tauler e A. Izquierdo-Ridorsa; *Continuous-flow and flow injection pH gradients for spectrophotometric determinations of mixtures of nucleic acid components*. Anal. Chem. **1999**, 71, 2215.
- [38] A. Izquierdo-Ridorsa, J. Saurina, S. Hernández-Cassou e R. Tauler; *Second-order multivariate curve resolution applied to rank-deficient data obtained from acid base spectrophotometric titrations of mixtures of nucleic acids*. Chem. Intell. Lab. Syst. **1997**, 38, 183.
- [39] R. Tauler; *Multivariate curve resolution applied to second order data*. Chemom. Intell. Lab. Syst. **1995**, 30, 133.
- [40] W. Windig e J. Guilment; *Interactive self-modeling mixture analysis*. Anal. Chem. **1991**, 63, 1425.
- [41] M. Maeder, A. D. Zuberbuehler; *The resolution of overlapping chromatographic peaks by evolving factor-analysis*. Anal. Chim. Acta, **1986**, 181, 287.
- [42] J. Saurina, C. Leal, R. Compañó, M. Granados, M. D. Prat e R. Tauler; *Estimation of figures of merit using univariate statistics for quantitative second-order multivariate curve resolution*. Anal. Chim. Acta, **2001**, 432, 241.
- [43] M. Linder e R. Sundberg; *Second-order calibration: bilinear least squares regression and a simple alternative*. Chemom. Intell. Lab. Syst. **1998**, 42, 159.
- [44] M. Linder e R. Sundberg; *Precision of prediction in second-order calibration, with focus in bilinear regression method*. J. Chemom. **2002**, 16, 12.
- [45] N. M. Faber, J. Ferré, R. Boque e J. H. Kalivas; *Second-order bilinear calibration: the effects of vectorising the data matrices of the calibration set*. Chemom. Intell. Lab. Syst. **2002**, 63, 107.
- [46] R. S. Valverde, M. D. G. García, M. M. Galera e H. C. Goicoechea; *Highly collinear three-way photoinduced spectrofluorimetric data arrays modeled with bilinear least squares: Determination of tetracyclines in surface water samples*. Talanta, **2006**, 70, 774.

- [47] A. M De La Pena, A. E. Mansilla, N. Mora, Díez, D. B. Gil, A. C. Olivieri e G. M. Escandar; *Second-order calibration of excitation-emission matrix fluorescence spectra for the determination of N-phenylanthranilic acid derivatives*. Appl. Spectrosc. **2006**, 60, 330.
- [48] J. W. B. Braga, C. B. G. Botolli, I. C. S. F. Jardim, H. C. Goicoechea, A. C. Olivieri e R. J. Poppi; *Determination of pesticides and metabolites in wine by high performance liquid chromatography and second-order calibration methods*. J. Chromatogr. A, **2007**, 1148, 200.
- [49] H. C. Goicoechea e A. C. Olivieri; *New robust bilinear least squares method for the analysis of spectral-pH matrix data*. Appl. Spectrosc. **2005**, 59, 926.
- [50] N. R. Marsili, A. Lista, B. S. F. Band, H. C. Goicoechea e A. C. Olivieri, *Evaluation of complex spectral-pH three-way arrays by modified bilinear least-squares: determination of four different dyes in interfering systems*. Analyst, **2005**, 130, 1291.
- [51] R. L. Carneiro, J. W. B. Braga e R. J. Poppi; *Application of genetic algorithm for selection of variables for the BLLS method applied to determination of pesticides end metabolites in wine*. Anal. Chim. Acta, **2007**, 595, 51.
- [52] A. C. Olivieri e N. M. Faber; *A closed-form expression for computing the sensitivity in second-order bilinear calibration*. J. Chemom. **2005**, 19, 583.
- [53] J. A. Arancibia, A. C. Olivieri, D. B. Gil, A. E. Mansilla, I. Durán-Merás e A. M De La Peña; *Trilinear least-squares and unfolded-PLS coupled to residual trilinearization: New chemometric tools for the analysis of four-way instrumental data*. Chemom. Intell. Lab. Syst. **2006**, 80, 77.
- [54] A. Haimovich, R. Orselli, G. M. Escandar e A. C. Olivieri; *Spectroscopic bilinear least-squares methods exploiting the second-order advantage. Theoretical and experimental study concerning accuracy, sensitivity and prediction error*. Chemom. Intell. Lab. Syst. **2006**, 80, 9.
- [55] J. Öhman, P. Geladi e S. Wold; *Residual bilinearization. Part 1: Theory and algorithms*. J. Chemom. **1990**, 4, 79.
- [56] R. G. Brereton; *Chemometrics – Data analysis for the laboratory and chemical plant*. 1ª ed., John Wiley & Sons, Chichester, **2003**, Cap. 5.

- [57] H. Martens e T. Næs; *Multivariate Calibration*. 1<sup>a</sup> ed., John Wiley & Sons, Chichester, **1989**, Cap 3.
- [58] D. M. Haaland e E. V. Thomas; *Partial least-squares for spectral analyses. 1. Relation to other quantitative calibration methods and the extraction of qualitative information*. Anal. Chem. **1988**, 60, 1193.
- [59] B. S. Dayal e J. F. Macgregor; *Improved PLS algorithms*. J. Chemom. **1997**, 11, 73.
- [60] S. de Jong; *SIMPLS - An alternative approach to partial least-squares regression*. J. Chemom. **1993**, 18, 251.
- [61] S. Wold, P. Geladi, K. Esbensen e J. Öhman; *Mult-way principal components and PLS-analysis*. J. Chemom. **1987**, 1, 41.
- [62] R. Bro; *Multway calibration. Multilinear PLS*. J. Chemom. **1996**, 10, 47.
- [63] J. Öhman, P. Geladi e S. Wold; *Residual Bilinearization. Part 2: Application to HPLC-diode array data and comparison with rank annihilation factor analysis*. J. Chemom. **1990**, 4, 135.
- [64] M. M. Reis, S. P. Gurden, A. K. Smilde e M. M. C. Ferreira; *Calibration and detailed analysis of second-order flow injection analysis data with rank overlap*. Anal. Chim. Acta, **2000**, 422, 21.
- [65] A. C. Olivieri; *On a versatile second-order multivariate calibration method based on partial least-squares and residual bilinearization: Second-order advantage and precision properties*. J. Chemom. **2005**, 19, 253.
- [66] M. J. Culzoni, H. C. Goicoechea, A. P. Pagani, M. A. Cabezón e A. C. Olivieri; *Evaluation of partial least squares with second-order advantage for the mult-way spectroscopic analysis of complex biological samples in the presence of analyte-background interactions*. Analyst, **2006**, 131, 718.
- [67] D. B. Gil, A. M. De La Pena, J. A. Arancibia, G. M. Escandar e A. C. Olivieri; *Second-order advantage achieved by unfolded-partial least-squares/residual bilinearization modeling of excitation – emission data presenting inner filter effects*. Anal. Chem. **2006**, 78, 8051.
- [68] A. García-Reiriz, P. C. Damiani e A. C. Olivieri; *Different strategies for the direct determination of amoxicillin in human urine by second-order multivariate analysis of kinetic-spectrophotometric data*. Talanta, **2007**, 71, 806.

- [69] International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use, *ICH Harmonized Tripartite Guideline – Validation of Analytical Procedures: Text and Methodology*, Federal Register. **1995**, 60, 11260.
- [70] Agência Nacional de Vigilância Sanitária; *Guia para a validação de métodos analíticos e bioanalíticos*. Resol. 899, 29 março de 2003.
- [71] G. M. Escandar, N. M. Faber, H. C. Goicoechea, A. M. De La Peña, A. C. Olivieri e R. J. Poppi; *Second- and third-order multivariate calibration: data, algorithms and applications*. Trends Anal. Chem. **2007**, 26, 752.
- [72] J. W. B. Braga e R. J. Poppi; *Validação de modelos de calibração multivariada: uma aplicação na determinação de pureza polimórfica de carbamazepina por espectroscopia no infravermelho próximo*. Quím. Nova, **2004**, 27, 1004.
- [73] M. M. Sena, C. B. Freitas, L. C. Silva, J. W. B. Braga e R. J. Poppi; *Aplicação de calibração multivariada no controle de qualidade de fármacos*. Estudos, **2006**, 32, 1953.
- [74] J. W. B. Braga, R. J. Poppi, F. A. L. Ribeiro, F. D. Barbosa, M. C. Breitzkreitz e J. A. Martins; *Calibração multivariada em análises farmacêuticas*. Norma submetida à Farmacopéia Brasileira em **2006**.
- [75] A. C. Olivieri e N. M. Faber; *Standard error of prediction in parallel factor analysis of three-way data*. Chemom. Intell Lab. Syst. **2004**, 70, 75.
- [76] A. C. Olivieri; *Sample-specific standard prediction errors in three-way parallel factor analysis (PARAFAC) exploiting the second-order advantage*. J. Chemom. **2004**, 18, 363.
- [77] N. M. Faber e R. Bro; *Standard error of prediction for multway PLS 1. Background and a simulation study*. Chemom. Intell Lab. Syst. **2002**, 61, 133.
- [78] A. C. Olivieri; *Computing sensitivity and selectivity in parallel factor analysis and related multway techniques: the need for further development in net analyte signal theory*. Anal. Chem. **2005**, 77, 4936.
- [79] R. Boque, J. Ferré, N. M. Faber e F. X. Rius; *Limit of detection estimator for second-order bilinear calibration*. Anal. Chim. Acta, **2002**, 451, 313.

- [80] A. C. Moffat, A. D. Trafford, R. D. Jee e P. Graham; *Meeting the International Conference on Harmonization's Guidelines on Validation Procedures: Quantification as exemplified by a near-infrared reflectance assay of paracetamol in intact tablets*. *Analyst*, **2000**, 125, 1341
- [81] G. Ragno, G. Ioele e A. Risoli; *Multivariate calibration techniques applied to the spectrophotometric analysis of on-to-four component systems*. *Anal. Chim. Acta*, **2004**, 512, 173.
- [82] M. J. Rodríguez-Cuesta, R. Boqué, F. X. Rius, J. L. M. Vidal e A. G. Frenich; *Development and validation of a method for determining pesticides in groundwater from complex overlapped HPLC signals and multivariate curve resolution*. *Chemom. Intell. Lab. Syst.* **2005**, 77, 251.
- [83] N. J. Miller-Ihli, T. C. O'Haver e J. M. Harnly; *Calibration and curve fitting for extended range AAS*. *Spectrochim Acta*, **1984**, 39, 1603.
- [84] A. Lorber, *Error propagation and figures of merit for quantification by solving matrix equations*. *Anal. Chem.* **1986**, 58, 1167.
- [85] A. Lorber, K. Faber e B. R. Kowalski, *Net analyte signal calculation in multivariate calibration*. *Anal. Chem.* **1997**, 69, 1620.
- [86] N. J. Messik, J. H. Kalivas e P. M. Lang; *Selectivity and related measures for nth-order data*. *Anal. Chem.* **1996**, 68, 1572.
- [87] C. N. Ho, G. D. Christian e E. R. Davidson; *Application of the method of rank annihilation to fluorescent multi-component mixtures of polynuclear aromatic hydrocarbons*. *Anal. Chem.* **1980**, 52, 1071.
- [88] J. Mandel e R. D. Stiehler; *Sensitivity – A criterion for the comparison of methods of test*. *J. Res. Natl. Bur. Stan.* **1954**, 53, 633.
- [89] L. C. Rodríguez, A. M. G. Campaña, C. G. Lindares e M. R. Ceba; *Estimation of performance characteristics of an analytical method using the data set of the calibration experiment*. *Anal. Lett.* **1993**, 26, 1243.
- [90] *Validation of Compendial Methods*, Chapter 1225, United States Pharmacopoeia USP28 NF23, **2005**, 2748.

- [91] L. A. Currie; *Nomenclature in evaluation of analytical methods including detection and quantification capabilities*. Pure & Appl. Chem. **1995**, 67, 1699.
- [92] J. Vessman, R.I. Stefan, J.F. Van Staden, K. Danzer, W. Lindner, D.T. Burns, A. Fajgelj e H. Muller; *Selectivity in Analytical Chemistry (IUPAC Recommendations 2001)*. Pure & Appl. Chem. **2001**, 73, 1381.
- [93] N. M. Faber, J. Ferre, R. Boqué e J. H. Kalivas; *Quantifying selectivity in spectrophotometric multicomponent analysis*. Trends Anal. Chem. **2003**, 22, 352.
- [94] G. Bergmann, V. von Oepen e P. Zinn; *Improvement in the definitions of sensitivity and selectivity*. Anal. Chem. **1987**, 59, 2522.
- [95] J. Ferré, S.D. Brown e F.X. Rius; *Improved calculation of net analyte signal in inverse multivariate calibration*. J. Chemom. **2000**, 15, 537
- [96] K. Faber e B. R. Kowalski; *Propagation of errors for the validation of predictions obtained by principal components regression and partial least squares*. J. Chemom. **1997**, 11, 181.
- [97] N. M. Faber; *Comparison of two recently proposed expressions for partial least squares regression prediction error*. Chemom. Intell. Lab. Syst. **2000**, 52, 123.
- [98] J. A. F. Pierna, L. Jin, F. Wahl, N. M. Faber e D. L. Massart; *Estimation of partial least squares regression prediction uncertainty when values carry a sizeable measurement error*. Chemom. Intell. Lab. Syst. **2003**, 65, 281.
- [99] N. M. Faber e R. Bro; *Standard error of prediction for multway PLS 1. Background and a simulation study*. Chemom. Intell. Lab. Syst. **2002**, 61, 133.
- [100] A. Höskuldsson; *Estimation of prediction error for multivariate calibration*. J. Chemom. **1998**, 2, 93.
- [101] M. L. Griffiths, R. P. Barbagallo e J. T. Keer; *Multiple and simultaneous fluorophore detection using fluorescence spectrometry and partial least-squares regression with sample-specific confidence intervals*. Anal. Chem. **2006**, 78, 513.
- [102] R. D. Jiji, G. A. Cooper e K. L. Booksh; *Excitation-emission fluorescence based determination of carbamate pesticides and polycyclic aromatic hydrocarbons*. Anal. Chim. Acta, **1999**, 397, 61.

- [103] J. Amador-Hernández, J. M. Fernández-Romero e M. D. L. Castro; *Flow injection screening and semiquantitative determination of polycyclic aromatic hydrocarbons in water by laser induced spectrofluorimetry*. Anal. Chim. Acta, **2001**, 448, 61.
- [104] E. Selli, C. Zaccaria, F. Sena, G. Tomasi e G. Bidoglio; *Application of multi-way models to the time-resolved fluorescence of polycyclic aromatic hydrocarbons mixtures in water*. Water Res. **2004**, 38, 2269.
- [105] M. L. Nahorniak e K. S. Booksh; *Excitation-emission matrix fluorescence spectroscopy in conjunction with multway analysis for PAH detection in complex matrices*. Analyst, **2006**, 131, 1308.
- [106] C. M. Jarque e A. K. Bera; *A test for normality of observations and regressions residuals*. Int. Stat. Rev. **1987**, 55, 163.
- [107] R. Bianchini e M. D. V. C. Penteado; *Teores de tiamina riboflavina e piridoxina em leites bovinos comercializados na cidade de São Paulo*. Ciênc. Tecnol. Aliment. **2000**, 20, 291.
- [108] D. C. Woolard e H. E. Indyk; *Rapid determination of thiamine, riboflavin, pyridoxine and niacinamide in infant formulas by liquid chromatography*. J. AOAC Int. **2002**, 85, 945.
- [109] R. Bro, R. A. Harshman e N. D. Sidiropoulos. *Rank-deficient models for multway data. Part I: Theory*. J. Chemom., artigo submetido.
- [110] M. M. Sena, J. C. B. Fernandes, L. Rover Jr., R. J. Poppi e L. T. Kubota; *Application of two- and three-way chemometric methods in the study of acetylsalicylic acid and ascorbic acid mixtures using ultraviolet spectrophotometry*. Anal. Chim. Acta, **2000**, 409, 159.
- [111] R. J. Poppi; *Construção de um espectrofotômetro com transformada de Hadamard e sua aplicação na análise por injeção em fluxo*, Tese de Doutorado, Universidade Estadual de Campinas, Brasil, **1993**.
- [112] J. C. B. Fernandes, G. Oliveira Neto e L. T. Kubota; *Use of column with modified silica for interfering retention in a FIA spectrophotometric method for direct determination of vitamin C in medicine*. Anal. Chim. Acta, **1998**, 366, 11.

- [113] The United States Pharmacopoeia- USP 28, The National Formulary – NF23, Official monographs, Rockville, 2005, pg 179 e 183.
- [114] J. Krug, H. Bergamin Fº, E. A. G. Zagatto, S. S. Joergensen; *Rapid determination of sulphate in natural waters and plant digests by continuous flow injection turbidimetry*. Analyst, **1977**, 102, 503.
- [115] A. C. B. Dias, E. P. Borges, E. A. G. Zagatto e P. J. Worsfold; *A critical examination of the components of the Schlieren effect in flow analysis*. Talanta, **2006**, 68, 1076.
- [116] The United States Pharmacopoeia- USP 28, The National Formulary – NF23, Official monographs, Rockville, 2005, pg. 180.
- [117] [http://www.codexalimentarius.net/mrls/pestdes/jsp/pest\\_q-e.jsp](http://www.codexalimentarius.net/mrls/pestdes/jsp/pest_q-e.jsp). Acessada em julho de 2005.
- [118] C. P. Wang e T. L. Isenhour; *Time-warping algorithm applied to chromatographic peak matching gas chromatography/Fourier transform infrared/mass spectrometry*. Anal. Chem. **1987**, 59, 649.
- [119] N. P. V. Neilsen, J. M. Carstensen e J. Smedsgaard; *Aligning of single and multiple wavelength chromatographic profiles for chemometric data analysis using correlation optimized warping*. J. Cromatogr. A, **1998**, 805, 17.
- [120] V. Pravdova, B. Walczak e D. L. Massart; *A comparison of two algorithms for warping of analytical signals*. Anal. Chim. Acta, **2002**, 456, 77.
- [121] H. F. M. Boelens, R. J. Dijkstra, P. H. C. Eilers, F. Fitzpatrick e J. A. Westerhuis; *New background correction method for liquid chromatography with diode array detection, infrared spectroscopy detection and Raman spectroscopy detection*. J. Cromatogr. A, **2004**, 1057, 21.
- [122] R. L. Carneiro. *Algoritmos genéticos para a seleção de variáveis em métodos de calibração de segunda ordem*. Dissertação de mestrado, Universidade Estadual de Campinas, Brasil, **2007**.
- [123] P. C. Damiani, I. Duran-Merás, A. García-Reiriz, A. Jiménez-Girón, A. M. de la Peña e A. C. Olivieri; *Multway partial least-squares coupled to residual trilinearization: a genuine multidimensional tool for the study of third-order data. Simultaneous analysis of procaine and its metabolite p-aminobenzoic acid in equine serum*. Anal. Chem. **2007**, 79, 6949.

## **Apêndices**

## Apêndice A: Produtos matriciais de bloco

### A1 Produto de Kronecker<sup>2</sup>

O produto de Kronecker, também conhecido como produto de blocos ou tensorial e geralmente expresso pelo símbolo “ $\otimes$ ”. Dadas duas matrizes de tamanho arbitrário  $\mathbf{A}(m,n)$  e  $\mathbf{B}(p,q)$ , o produto de Kronecker de  $\mathbf{A}$  por  $\mathbf{B}$  é expresso por  $\mathbf{A} \otimes \mathbf{B}$  e resulta em uma matriz de dimensões  $(mp,nq)$ , definida como:

$$\begin{bmatrix} a_{1,1}\mathbf{B} & \cdots & a_{1,n}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{m,1}\mathbf{B} & \cdots & a_{m,n}\mathbf{B} \end{bmatrix}$$

Por exemplo, dadas as matrizes:  $\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 3 & 5 \\ 4 & 0 \end{bmatrix}$  e  $\mathbf{B} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ , o produto  $\mathbf{A} \otimes \mathbf{B} = \mathbf{C}$

fornece como resultado:

$$\mathbf{C} = \begin{bmatrix} 2 & 4 & 1 & 2 \\ 6 & 8 & 3 & 4 \\ 3 & 6 & 5 & 10 \\ 9 & 12 & 15 & 20 \\ 4 & 8 & 0 & 0 \\ 12 & 16 & 0 & 0 \end{bmatrix}$$

---

<sup>2</sup> Referência: J. R. Magnus e Heinz Neudecker; Matrix differential calculus with applications in statistics and econometrics. John Wiley & Sons, Chichester, 1988, Cap. 2, pg. 27.

## A2 Produto de Khatri-Rao<sup>3,4</sup>

Assim como o produto de Kronecker, o produto de Khatri-Rao também pode ser considerado um produto matricial em blocos, sendo geralmente expresso por “ $|\otimes|$ ”. Para que este produto seja aplicado a duas matrizes, é necessário que estas tenham o mesmo número de colunas  $F$ . Logo, tomando como exemplo as matrizes  $\mathbf{A}(2,2)$  e  $\mathbf{B}(3,2)$  abaixo:

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \text{ e } \mathbf{B} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}$$

onde  $F=2$ . O produto de Khatri-Rao  $\mathbf{A}|\otimes|\mathbf{B}=\mathbf{C}$  fornece como resultado:

$$\mathbf{C} = \begin{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 3 \\ 5 \end{bmatrix} & \begin{bmatrix} 2 \\ 4 \\ 6 \\ 2 \end{bmatrix} \\ \begin{bmatrix} 1 \\ 3 \\ 4 \\ 5 \end{bmatrix} & \begin{bmatrix} 2 \\ 4 \\ 4 \\ 6 \end{bmatrix} \end{bmatrix} = \begin{bmatrix} 1 & 4 \\ 3 & 8 \\ 5 & 12 \\ 3 & 8 \\ 9 & 16 \\ 15 & 24 \end{bmatrix} = [\mathbf{a}_1 \otimes \mathbf{b}_1, \mathbf{a}_2 \otimes \mathbf{b}_2]$$

$\mathbf{a}_1$  e  $\mathbf{b}_1$  correspondem a vetores contendo apenas as primeiras colunas de  $\mathbf{A}$  e  $\mathbf{B}$ , respectivamente e  $\mathbf{a}_2$  e  $\mathbf{b}_2$  vetores contendo as segundas colunas. Deste modo, se percebe que o produto de Khatri-Rao pode ser visualizado como uma aplicação sequencial do produto de Kronecker.

---

<sup>3</sup> R. Bro, *Multi-way analysis in the food industry – Models, algorithms and applications*. Tese de doutorado, Universidade de Amsterdam, Holanda, **1998**, Cap. 3. pg. 21.

<sup>4</sup> S. Liu; *Matrix results on the Khatri-Rao and tracy-sing products*. Lin. Alg. Appl., 1999, 289, 267.

## **Apêndices B: Programas desenvolvidos.**

Os códigos fonte dos programas e sub-rotinas desenvolvidos nesta tese são apresentados a seguir, sendo estes compatíveis com os ambientes de programação Matlab e Octave em suas versões 6.5 e 3.0.0.

Para a utilização de cada programa, deve-se transcrever (ou selecionar e colar) todo o código fonte do programa em qualquer editor de texto simples (ex.: bloco de notas) e salvar o programa com seu respectivo nome e a extensão “.m”. Em seguida este programa e suas sub-rotinas devem ser colocados no diretório de trabalho do ambiente em questão e executado segundo as instruções presentes no texto de ajuda de cada programa, o qual pode ser acessado na janela de comando de cada ambiente pelo comando help seguido do nome do respectivo programa sem a extensão do arquivo.

As linhas ou sentenças precedidas do símbolo “%” no código correspondem a comentários inseridos para facilitar o entendimento do código fonte. Para minimizar a ocorrência de erros no código fonte, não são utilizados acentos ou caracteres como “ç” dentro do código fonte do programa, inclusive nos comentários.

## B1 Programa `blls_rbl`: construção do modelo BLLS utilizando o RBL quando há interferentes.

```
% BLLS-RBL : Este programa constroi o modelos de quadrados minimos bilinear (BLLS) com
% bilinearizacao dos residuos (RBL)[1-4,10,11].
% A utilizacao e distribuicao deste programa e livre e permitida desde que devidamente
% citada a sua fonte [12]. Contudo, os autores nao oferecem nenhuma garantia da sua funcionalidade
% ou ajuste na analise de qualquer conjunto de dados.
%
% Sintaxe: modelo=blls_rbl(Xc,yc,Xun,ni,center,eq_species,ruido,sistema);
%
% Entradas:
%   Xc : cubo de dados contendo as amostras de calibracao, onde a
%   primeira dimensao deve ser a das amostras;
%
%   yc : vetor (ou matriz) contendo os valores de referencias para as
%   concentracoes nas amostras de calibracao;
%
%   Xun : cubo de dados contendo as amostras para analise;
%
%   ni : fatores que seram utiliados no rbl para modelar os interferentes em Xun;
%
%   center: variavel logica que informa se os dados devem ser centrados na media (center=qualquer
%   numero diferente de 0) ou nao (center =0);
%
%   eq_species: vetor que indica o numero de especies em equilibrio de cada composto.
%   Ex: para dois analitos e duas especies para o primeiro e 1 para o
%   segundo eq_species=[2 1]. Caso na haja especies em equilibrio,
%   deve-se utilizar como entrada no programa [].
%
%   ruido: estimativa do ruido instrumental. Caso nao seja dispoivel, deve-se inserir utilizar [] como
%   entrada do programa, que utilizara o residuo obtido pela decomposicao da amostra como uma estimativa.
%
%   sistema: variavel de texto informa qual o utilizado. Deve-se inserir 'matlab' ou 'octave', entre apostofos.
%
% Saidas: variavel estruturada modelo
%
%   yun: estimativas da propriedade de interesse para as amostras analisadas;
%
%   yc_est: estimativas para a propriedade de interesse para as amostras de calibracao;
%
%   rmsec: raiz quadrada do erro medio quadratico de calibracao;
%
%   rep: estimativa do erro relativo das estimativas das amostras de
%   calibracao (%) [10];
%
%   perfil_B e perfil_C: perfis normalizados estimados nas dimensoes
%   segunda e terceira dimensao do cubo de dados;
%
%   k_n: valores singulares estimados na determinacao dos analitos;
%
%   residuo_c: residuo de decomposicao para as amostras de calibracao;
```

```

%
% residuo_un: residuo de decomposicao para as amostras de
% previsao;
%
% SEN: valores de sensibilidade [5]
% SEL: valores de seletividade [11]
% SENA: valores de sensibilidade e analitica [7,8]
% LD: limite de deteccao [9]
% LQ: limite de quantificacao [9]
%
%
% OBS: Se ni>0:
% 1)os perfis serao Be e Ce (que possuem o perfil dos analitos e dos interferentes).
% 2)residuos_un=residuo da deixado pelo modelo para as amostras de de previsao,
% sem considerar os interferentes.
% 3)residuos_uni=residuo da deixado pelo modelo para as amostras de de previsao,
% considerando os interferentes.
%
% Referencias (indicadas entre colchetes):
% 1- M. Linder, R. Sundberg, Chemom. Intell. Lab. Syst., 1998, 42, 159.
% 2- M. Linder, R. Sundberg, J. Chemom., 2002, 16, 12.
% 3- A.C. Olivieri, Anal. Chem., 2005, 77(15), 4936.
% 4- N. R. Marsili, A. Lista, B. S. F. Band, H. C. Goicoechea, A. C. Olivieri. Analyst, 2005,130,1291.
% 5- A. C. Olivieri, N. M. Faber, J. Chemom. 2005, 19, 583.
% 6- A. Haimovich, R. Orselli, G. M. Escandar, A. C. Olivieri, Chemom. Intell. Lab. Syst., 2006, 80, 99.
% 7- J. Mandel e R. D. Stiehler; J. Res. Natl. Bur. Stan. 1954, 53, 633.
% 8- A. Munoz de la Pena, A. E. Mansilla, D. G. Gomes, A. C. Olivieri, H. C. Goicoechea, Anal. Chem. 2003,
% 75, 2640.
% 8- L. A. Currie, Anal. Chim. Acta, 1999, 391, 105-126.
% 9- N. J. Miller-Ihli, T. C. OHaver e J. M. Harnly; Spectrochim Acta, 1984, 39(12),1603.
% 10- H. C. Goicoechea e A.C. Olivieri; Appl. Spectrosc., 2005, 59, 926.
% 11- P. C. Damiani, A. J. Nepole, M. Bearzotti e A. C. Olivieri; Anal. Chem., 2004, 76, 2798.
% 12- J. W. B. Braga; Aplicacao e validacao de modelos de calibracao de segunda ordem em quimica analitica.
% Tese de Doutorado, Universidade estadual de Campinas, Brasil, 2008.
%
% Em caso de duvidas, sugestoes ou qualquer problema na execucao deste programa,
% por favor, entre em contato com os autores:
% Jez Willian Batista Braga (email: jezwillian@gmail.com)
% Ronei Jesus Poppi (e-mail: ronei@iqm.unicamp.br)
%
% LAQQA - Laboratorio de Quimiometria em Quimica Analitica
% Instituto de Quimica - UNICAMP
% 21/01/2008
%
function modelo=blls_rbl(Xc,yc,Xun,ni,center,eq_species,ruido,sistema);
[L1,C1,H1]=size(Xc);
[L2,C2]=size(yc);
[L3]=size(Xun,1);
[L4]=size(ruido);
[C3]=size(eq_species,2);

Yc=yc;
% Centrar os dados na media
if center==1
my=mean(yc);
yc=yc-ones(L1,1)*my;

```

```

for i=1:C1
MX(i,:)=mean(squeeze(Xc(:,i,:)));
end

for i=1:L1
Xc(i,:,:)=squeeze(Xc(i,:,:))-MX;
end

for i=1:L3
Xun(i,:,:)=squeeze(Xun(i,:,:))-MX;
end

end

% Vectoriza o tensor X(i,j,k) em Vx(i,j*k)
for i=1:L1;
Vx(:,i)=reshape(Xc(i,:,:),1,C1*H1)';
end

% quadrados minimos classico entre Vx e y
Vs=Vx*pinv(yc');

if C3==0 % condição para decidir se os dados sao deficientes de rank ou nao.

% Obtencao dos perfis das dimenssoes j e k
for i=1:C2
[U,S,V]=svd(reshape(Vs(:,i),C1,H1));
S=diag(S);
k_n(i)=S(1,1);
B(:,i)=U(:,1)/norm(U(:,1));
C(:,i)=V(:,1)*k_n(i);
Sc(:,i)=S(1,1)*kron(V(:,1),U(:,1));
end

% Previsao das amostras Xc
for i=1:L2
yc_est(:,i)=pinv(Sc)*reshape(Xc(i,:,:),1,C1*H1)';
Yest=yc_est(:,i);
Res1=squeeze(Xc(i,:,:))-B*diag(Yest)*C';
residuo_c(i)=norm(Res1,'fro')/sqrt((C1-1)*(H1-1));
end

yc_est=yc_est';

if center==1
DF=C2+1;
else
DF=C2;
end

if ni==0 % Condiçao para decidir se ha interferencia ou nao.

% Normalisa os perfis B e C estimados
for i=1:C2
Bcal(:,i)=B(:,i)/norm(B(:,i));

```

```

Ccal(:,i)=C(:,i)/norm(C(:,i));
end

ii=inv((Bcal*Bcal).*(Ccal*Ccal));

% Previsao das amostras Xun
for i=1:L3
yun(:,i)=pinv(Sc)*reshape(Xun(i,,:),1,C1*H1)';
Yest=yun(:,i);
Res1=squeeze(Xun(i,,:))-B*diag(Yest)*C';
residuo_un(i)=norm(Res1,'fro')/sqrt((C1-1)*(H1-1));

% Estimativa do ruído
if L4==0
ruído=residuo_un(i);
end

% Calculo de figuras de merito
for j=1:C2
sen(j)=k_n(j)/sqrt(ii(j,j));
sel(j)=1/sqrt(ii(j,j));
sen_anal(i,j)=(k_n(j)/sqrt(ii(j,j)))/ruído;
SD_conc(i,j)=1./sen_anal(i,j);
LD(i,j)=3.3/sen_anal(i,j);
LQ(i,j)=10/sen_anal(i,j);
end

end
yun=yun';

% volta a media
if center==1
yc_est=yc_est+ones(L1,1)*my;
yc=yc+ones(L1,1)*my;
yun=yun+ones(L3,1)*my;
end

for i=1:C2
res=Yc(:,i)-yc_est(:,i);
rmsec(i)=sqrt((res*res)/(L1-DF));
rep(i)=sqrt(sum((res.^2)/((Yc(:,i).^2)*(L1-DF))))*100;
end

%FIGURAS
figure;plot(residuo_c,'ko');
hold on;plot(residuo_un,'m*');
sxm(1:L3)=mean(residuo_c(:));
hold on;plot(sxm*2.326,'-');
if L4~=0
RUIDO(1:L3)=ruído;
hold on;plot(RUIDO*2.326,'-');
end

if ni>0;hold on;plot(residuo_uni,'^');end
title('Residuos de decomposicao');xlabel('Numero da amostra');ylabel('Residuos');

```

```

figure
for i=1:C2
hold on;subplot(1,C2,i);plot(yc(:,i),yc_est(:,i),'o');
xlabel('Valores de referencia');ylabel('Valores estimados');
end

% Salva saidas
modelo.yun=yun;
modelo.yc_est=yc_est;
modelo.rmsec=rmsec;
modelo.rep=rep;
modelo.perfil_B=Bcal;
modelo.perfil_C=Ccal;
modelo.k_n=k_n;
modelo.residuo_c=residuo_c;
modelo.residuo_un=residuo_un;
modelo.SEN=sen;
modelo.SEL=sel;
modelo.SENA=sen_anal;
modelo.LD=LD;
modelo.LQ=LQ;

else

% BILINEARIZACAO RESUDIAL
% Previsao das amostras Xun
for i=1:L3
XT=squeeze(Xun(i,:,:));           % seleciona a amostra
Yest0=pinv(Sc)*XT(:);           %primeira estimativa da concentraçã
Res1=XT-B*diag(Yest0)*C';       %estimativa dos residuos
residuo_un(i)=norm(Res1,'fro')/sqrt((C1-1)*(H1-1)); %norma dos residuos

%Minimizacao para o RBL
g=Yest0;                         % inicialização de Y para a otimização
In=Res1;                         % inicialização dos residuos para a otimização
[P,S,Q]=svd(In,0);              % estimativa dos perfis dos interferentes
Q=S*Q'; Q=Q(1:ni,:);P=P(:,1:ni); % seleciona apenas os ni interferentes
In_r=P*Q;                       % matris com o sinal dos interferentes recuperada
g1=pinv(Sc)*reshape(XT-In_r,1,C1*H1)'; % estimativa de Y sem o sinal dos interferentes

while abs(g1-g)> abs(0.000001*g)   % Loop iterativo para melhorar a estimatica inicial
g=g1;
[P,S,Q]=svd(In_r,0);
Q=S*Q'; Q=Q(1:ni,:);P=P(:,1:ni);
In_r=P*Q;
g1=pinv(Sc)*reshape(XT-In_r,1,C1*H1)';
end
beta0=g1';

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
if sistema=='matlab'
Ynl=XT(:);
save tempj Ynl C1 H1 ni B C Sc % salva as variaveis necessarias
Xnl=zeros(size(Ynl)); % matriz apropriada para nlinfit
[Yest_gn,Resf]=nlinfit(Xnl,Ynl,'func_mat_obj_rbl',beta0);
Resf=reshape(Resf,C1,H1);

```



```

Ccal=[Ccal,Ce(:,j)];
end

Bint=[];
for j=C2+1:ni+C2
Bint=[Bint,Be(:,j)];
end

Cint=[];
for j=C2+1:ni+C2
Cint=[Cint,Ce(:,j)];
end

Pb_int=eye(C1)-Bint*pinv(Bint);
Pc_int=eye(H1)-Cint*pinv(Cint);

ii=inv((Bcal'*Pb_int*Bcal).*(Ccal'*Pc_int*Ccal));

for j=1:C2
sen(i,j)=k_n(j)/sqrt(ii(j,j));
sel(i,j)=1/sqrt(ii(j,j));
sen_anal(i,j)=(k_n(j)/sqrt(ii(j,j)))/ruído;
SD_conc(i,j)=1./sen_anal(i,j);
LD(i,j)=3.3/sen_anal(i,j);
LQ(i,j)=10/sen_anal(i,j);
end
end

%volta a media
if center==1
yc_est=yc_est+ones(L1,1)*my;
yc=yc+ones(L1,1)*my;
yun=yun+ones(L3,1)*my;
end

for i=1:C2
res=Yc(:,i)-yc_est(:,i);
rmsec(i)=sqrt((res'*res)/(L1-DF));
rep(i)=sqrt(sum((res.^2)/((Yc(:,i).^2)*(L1-DF))))*100;
end

for i=1:C2+ni
Be(:,i)=Be(:,i)/norm(Be(:,i));
Ce(:,i)=Ce(:,i)/norm(Ce(:,i));
end

%FIGURAS
figure;plot(residuo_c,'ko');
hold on;plot(residuo_un,'m*');
sxm(1:L3)=mean(residuo_c(:));
hold on;plot(sxm*2.326,'-');
if L4~=0
RUIDO(1:L3)=ruído;
hold on;plot(RUIDO*2.326,'-');
end
if ni>0;hold on;plot(residuo_uni,'^');end

```

```
title('Residuos de decomposicao');xlabel('Numero da amostra');ylabel('Residuos');
```

```
figure
for i=1:C2
hold on;subplot(1,C2,i);plot(yc(:,i),yc_est(:,i),'o');
xlabel('Valores de referencia');ylabel('Valores estimados');
end
```

```
% Salva saidas
modelo.yun=yun;
modelo.yc_est=yc_est;
modelo.rmsec=rmsec;
modelo.rep=rep;
modelo.perfil_Be=Be;
modelo.perfil_Ce=Ce;
modelo.k_n=k_n;
modelo.residuo_c=residuo_c;
modelo.residuo_un=residuo_un;
modelo.residuo_uni=residuo_uni;
modelo.SEN=sen;
modelo.SEL=sel;
modelo.SENA=sen_anal;
modelo.LD=LD;
modelo.LQ=LQ;
end
```

```
elseif C3>0
```

```
% Obtencao dos perfis das dimenssoes j e k
w=0;
```

```
for i=1:C2
[U,S,V]=svd(reshape(Vs(:,i),C1,H1));
S=diag(S);
```

```
for j=1:eq_species(i)
k_n(w+j)=S(j);
B(:,w+j)=U(:,j)/norm(U(:,j));
C(:,w+j)=V(:,j)*k_n(j+w);
Sc(:,w+j)=S(j)*kron(V(:,j),U(:,j));
end
```

```
w=i*j;
end
```

```
% Previsao das amostras Xc
for i=1:L2
yc_est(:,i)=pinv(Sc)*reshape(Xc(i, :, :),1,C1*H1);
Yest=yc_est(:,i);
Res1=squeeze(Xc(i, :, :))-B*diag(Yest)*C';
residuo_c(i)=norm(Res1,'fro')/sqrt((C1-1)*(H1-1));
end
```

```
yc_est=yc_est';
```

```
if center==1
```

```

DF=C2+1;
else
DF=C2;
end

if ni==0 % Condição para decidir se há interferência ou não.

% Normaliza os perfis B e C estimados
for i=1:sum(eq_species)
Bcal(:,i)=B(:,i)/norm(B(:,i));
Ccal(:,i)=C(:,i)/norm(C(:,i));
end

ii=inv((Bcal'*Bcal).*(Ccal'*Ccal));

% Previsão das amostras Xun
for i=1:L3
yun(:,i)=pinv(Sc)*reshape(Xun(i, :, :), 1, C1*H1);
Yest=yun(:,i);
Res1=squeeze(Xun(i, :, :))-B*diag(Yest)*C';
residuo_un(i)=norm(Res1, 'fro')/sqrt((C1-1)*(H1-1));

% Estimativa do ruído
if L4==0
ruído=residuo_un(i);
end

% Cálculo de figuras de mérito
for j=1:sum(eq_species)
sen(j)=k_n(j)/sqrt(ii(j,j));
sel(j)=1/sqrt(ii(j,j));
sen_anal(i,j)=(k_n(j)/sqrt(ii(j,j)))/ruído;
SD_conc(i,j)=1./sen_anal(i,j);
LD(i,j)=3.3/sen_anal(i,j);
LQ(i,j)=10/sen_anal(i,j);
end

end
yun=yun';

% volta a média
if center==1
% Média para as espécies em equilíbrio
w=0;
for i=1:C2
for j=1:eq_species(i)
myy(w+j)=my(i);
end
w=i*j;
end
yc_est=yc_est+ones(L1,1)*myy;
yun=yun+ones(L3,1)*myy;
end

% cálculo de rmsep e rep
w=0;

```

```

for i=1:C2
for j=1:eq_species(i)
res=Yc(:,i)-yc_est(:,j+w);
rmsec(j+w)=sqrt((res'*res)/(L1-DF));
rep(i+w)=sqrt(sum((res.^2)./((Yc(:,i).^2)*(L1-DF))))*100;
end
w=i*j;
end

%FIGURAS
figure;plot(residuo_c,'ko');
hold on;plot(residuo_un,'m*');
sxm(1:L3)=mean(residuo_c(:));
hold on;plot(sxm*2.326,'-');
if L4~=0
RUIDO(1:L3)=ruido;
hold on;plot(RUIDO*2.326,'-');
end
if ni>0;hold on;plot(residuo_uni,'^');end
title('Residuos de decomposicao');xlabel('Numero da amostra');ylabel('Residuos');

% Salva saidas
modelo.yun=yun;
modelo.yc_est=yc_est;
modelo.rmsec=rmsec;
modelo.rep=rep;
modelo.perfil_B=Bcal;
modelo.perfil_C=Ccal;
modelo.k_n=k_n;
modelo.residuo_c=residuo_c;
modelo.residuo_un=residuo_un;
modelo.SEN=sen;
modelo.SEL=sel;
modelo.SENA=sen_anal;
modelo.LD=LD;
modelo.LQ=LQ;

else

% BILINEARIZACAO RESIDIAL
% Previsao das amostras Xun
for i=1:L3
XT=squeeze(Xun(i,:,:));
Yest=pinv(Sc)*reshape(XT,1,C1*H1)';
Res1=XT-B*diag(Yest)*C';
residuo_un(i)=norm(Res1,'fro')/sqrt((C1-1)*(H1-1));

beta0=[];
indi=0;
for k=1:C2
for j=1:eq_species(k)
beta0=[beta0,mean(Yest(indi+1:indi+eq_species(k)))];
end
indi=indi+eq_species(k);
end
end

```

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
if sistema=='matlab'
Ynl=XT(:);save temp2 Ynl B C C1 H1 C2 Sc ni eq_species
Xnl=zeros(size(Ynl));clear Yest_gn

[Yest_gn,residue]=nlinfit(Xnl,Ynl,'func_mat_obj_rbl_rd',beta0);

Res=(XT-B*diag(Yest_gn)*C');
[u0,s0,v0]=svd(Res,0);

Be=[B,(u0(:,1:ni))];
Ce=[C,(v0(:,1:ni))*s0(1:ni,1:ni)];

Sexp=Sc;
for z=1:ni
Sexp=[Sexp,s0(z,z)*kron(v0(:,z),u0(:,z))]; % Expande Scal com os perfis dos interferentes
end
residuo_uni(i,1)=norm(residue,'fro')/sqrt((C1-1)*(H1-1));
yun(i,:)=Yest_gn(1:sum(eq_species));

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
else sistema=='octave'
Ynl=XT(:);
save -binary temp2gnu Ynl B C C1 H1 C2 Sc ni eq_species
Xnl=zeros(size(Ynl));

[Yest_gn, obj, info, iter, nf, lambda] = sqp (beta0', @func_oct_obj_rbl_rd, [], []);

Res=(XT-B*diag(Yest_gn)*C');
[u0,s0,v0]=svd(Res,0);
Be=[B,(u0(:,1:ni))];
Ce=[C,(v0(:,1:ni))*s0(1:ni,1:ni)];

Sexp=Sc;
for z=1:ni
Sexp=[Sexp,s0(z,z)*kron(v0(:,z),u0(:,z))]; % Expande Scal com os perfis dos interferentes
end
Yest_gn2=pinv(Sexp)*Ynl; %estima a concentração otimizada
Res=(XT-Be*diag(Yest_gn2)*Ce');
residuo_uni(i,1)=norm(Res,'fro')/sqrt((C1-1)*(H1-1));
yun(i,:)=Yest_gn(1:sum(eq_species));
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% Estimativa do ruido
if L4==0
ruido=residuo_uni(i);
end

for j=1:sum(eq_species)+ni
Be(:,j)=Be(:,j)/norm(Be(:,j));
Ce(:,j)=Ce(:,j)/norm(Ce(:,j));
end

% Calculo de figuras de merito

```

```

Bcal=[];
for j=1:sum(eq_species)
Bcal=[Bcal,Be(:,j)];
end

Ccal=[];
for j=1:sum(eq_species)
Ccal=[Ccal,Ce(:,j)];
end

Bint=[];
for j=sum(eq_species)+1:ni+sum(eq_species)
Bint=[Bint,Be(:,j)];
end

Cint=[];
for j=sum(eq_species)+1:ni+sum(eq_species)
Cint=[Cint,Ce(:,j)];
end

Pb_int=eye(C1)-Bint*pinv(Bint);
Pc_int=eye(H1)-Cint*pinv(Cint);

ii=inv((Bcal'*Pb_int*Bcal).*(Ccal'*Pc_int*Ccal));

% Calculo de figuras de merito
for j=1:sum(eq_species)
sen(i,j)=k_n(j)/sqrt(ii(j,j));
sel(i,j)=1/sqrt(ii(j,j));
sen_anal(i,j)=(k_n(j)/sqrt(ii(j,j)))/ruído;
SD_conc(i,j)=1./sen_anal(i,j);
LD(i,j)=3.3/sen_anal(i,j);
LQ(i,j)=10/sen_anal(i,j);
end
end

% volta a media
if center==1
% Media para as especies em equilibrio
w=0;
for i=1:C2
for j=1:eq_species(i)
myy(w+j)=my(i);
end
w=i*j;
end
yc_est=yc_est+ones(L1,1)*myy;
yun=yun+ones(L3,1)*myy;
end

% calculo de rmsep e rep
w=0;
for i=1:C2
for j=1:eq_species(i)
res=Yc(:,i)-yc_est(:,j+w);
rmsec(j+w)=sqrt((res'*res)/(L1-DF));

```

```
rep(i+w)=sqrt(sum((res.^2)/((Yc(:,i).^2)*(L1-DF))))*100;
end
w=i*j;
end

%FIGURAS
figure;plot(residuo_c,'ko');
hold on;plot(residuo_un,'m*');
sxm(1:L3)=mean(residuo_c(:));
hold on;plot(sxm*2.326,'-');
if L4~=0
RUIDO(1:L3)=ruído;
hold on;plot(RUIDO*2.326,'-');
end
if ni>0;hold on;plot(residuo_uni,'^');end
title('Resíduos de decomposicao');xlabel('Numero da amostra');ylabel('Resíduos');

% Salva saidas
modelo.yun=yun;
modelo.yc_est=yc_est;
modelo.rmsec=rmsec;
modelo.rep=rep;
modelo.perfil_Be=Be;
modelo.perfil_Ce=Ce;
modelo.k_n=k_n;
modelo.residuo_c=residuo_c;
modelo.residuo_un=residuo_un;
modelo.residuo_uni=residuo_uni;
modelo.SEN=sen;
modelo.SEL=sel;
modelo.SENA=sen_anal;
modelo.LD=LD;
modelo.LQ=LQ;
end
end
```

**B2 Sub-rotina func\_mat\_obj\_rbl: utilizada no cálculo do RBL para o modelo BLLS no sistema Matlab 6.5.**

*% Funcao utilizada na minimização da matriz de erros para o modelo BLLS/RBL  
% pelo sistema Matlab (versao 6.5).*

```
function f = func_mat_obj_rbl(beta1,Xnl)

load tempj
XT=reshape(Ynl,C1,H1);
Yest_gn=beta1;
Res=(XT-B*diag(Yest_gn)*C');
[u0,s0,v0]=svd(Res,0);
Be=[B,(u0(:,1:ni))];
Ce=[C,(v0(:,1:ni)*s0(1:ni,1:ni))];
Sexp=Sc;
for z=1:ni
Sexp=[Sexp,s0(z,z)*kron(v0(:,z),u0(:,z))]; % Expande Scal com os perfis dos interferentes
end
Yest_gn=pinv(Sexp)*Ynl;          %estima a concentração otimizada
XT_est=Be*diag(Yest_gn)*Ce';
f=XT_est(:);
```

**B3 Sub-rotina func\_mat\_obj\_rbl\_rd: utilizada no cálculo do RBL para o modelo BLLS com deficiência de posto no sistema Matlab 6.5.**

```
% Funcao que retorna o valor da funcao objetivo para a minimizacao da
% matriz de erros o RBL no modelo BLLS para sistemas com deficiencia de
% posto. (Sistema Matlab 6.5)

function f = func_mat_obj_rbl_rd(beta0,X)

load temp2
XT=reshape(Ynl,C1,H1);
Res=(XT-B*diag(beta0)*C');
[u0,s0,v0]=svd(Res,0);
Be=[B,(u0(:,1:ni))];
Ce=[C,(v0(:,1:ni)*s0(1:ni,1:ni))];
Sexp=Sc;
for z=1:ni
Sexp=[Sexp,s0(z,z)*kron(v0(:,z),u0(:,z))]; % Expande Scal com os perfis dos interferentes
end
Yest_gn=pinv(Sexp)*Ynl;          %estima a concentração otimizada
Mcalc=Be*diag(Yest_gn)*Ce';
f=Mcalc(:);
```

**B4 Sub-rotina func\_oct\_obj\_rbl: utilizada no cálculo do RBL para o modelo BLLS no sistema Octave 3.0.0.**

*% Funcao que retorna o valor da funcao objetivo para a minimizacao da  
% matriz de erros do RBL no modelo BLLS. (Sistema Octave 3.0.0)*

```
function residuo = func_oct_obj_rbl(beta0)

load temp
XT=reshape(Ynl,C1,H1);
Yest_gn=beta0;
Res=(XT-B*diag(Yest_gn)*C');
[u0,s0,v0]=svd(Res,0);
Be=[B,(u0(:,1:ni))];
Ce=[C,(v0(:,1:ni)*s0(1:ni,1:ni))];
Sexp=Sc;
for z=1:ni
Sexp=[Sexp,s0(z,z)*kron(v0(:,z),u0(:,z))]; % Expande Scal com os perfis dos interferentes
end
Yest_gn=pinv(Sexp)*Ynl; %estima a concentração otimizada
Resl=(XT-Be*diag(Yest_gn)*Ce');
residuo=norm(Resl,'fro')/sqrt((C1-1)*(H1-1));
```

**B5 Sub-rotina func\_mat\_obj\_rbl\_rd: utilizada no cálculo do RBL para o modelo BLLS com deficiência de posto no sistema Octave 3.0.0.**

```
% Funcao que retorna o valor da funcao objetivo para a minimizacao da
% matriz de erros do RBL no modelo BLLS para sistemas com deficiencia de
% posto. (Sistema Octave 3.0.0)

function residuo = func_oct_obj_rbl_rd(beta0)

load temp2gnu
XT=reshape(Ynl,C1,H1);
Res=(XT-B*diag(beta0)*C');
[u0,s0,v0]=svd(Res,0);
Be=[B,(u0(:,1:ni))];
Ce=[C,(v0(:,1:ni)*s0(1:ni,1:ni))];
Sexp=Sc;
for z=1:ni
Sexp=[Sexp,s0(z,z)*kron(v0(:,z),u0(:,z))]; % Expande Scal com os perfis dos interferentes
end
Yest_gn=pinv(Sexp)*Ynl; %estima a concentração otimizada
Resl=(XT-Be*diag(Yest_gn)*Ce');
residuo=norm(Resl,'fro')/sqrt((C1-1)*(H1-1));
```

## B6 Programa upls1\_rbl: construção do modelo UPLS1 utilizando o RBL quando há interferentes.

```

% UPLS1-RBL (Unfold Partial Least Squares [1] with Residual Bilinearization): Este programa
% constroi o modelos de quadrados minimos parciais desdobrado utilizando o processo de
% bilinearizacao dos residuos [2-5] para a modelas interferentes nao calibrados.
% A utilizacao e distribuicao deste programa e livre e permitida desde que devidamente
% citada a sua fonte [11]. Contudo, os autores nao oferecem nenhuma garantia da sua funcionalidade
% ou ajuste na analise de qualquer conjunto de dados.
%
% Obs: este programa foi desenvolvido como parte do projeto de doutorado de livre
%
% SINTAXE: modelo=upls1_rbl(Xc,yc,A,center,Xun,ni,ruido,sistema,erroY);
%
% ENTRADAS:
% Xc : cubo de dados contendo as amostras de calibracao, onde a
% primeira dimensao deve ser a das amostras;
%
% yc : vetor (ou matriz) contendo os valores de referencias para as
% concentracoes nas amostras de calibracao;
%
% A: numero de variaveis latentes a serem utilizadas na etapa de calibracao;
%
% center: variavel logica que informa se os dados devem ser centrados na media (center=qualquer
% numero diferente de 0) ou nao (center =0);
%
% Xun : cubo de dados contendo as amostras para analise;
%
% ni: numero de interferentes presente nas amostras;
%
% ruido: estimativa do ruido instrumental. Caso nao seja dispoivel, deve-se inserir utilizar [] como
% entrada do programa, que utilizara o residuo obtido pela decomposicao da amostra como uma estimativa.
%
% sistema: variavel de texto informa qual o utilizado. Deve-se inserir 'matlab' ou 'octave', entre apostofos.
%
% erroY: estimativa do desvio padrao dos erros das concentracoes de
% referencia;
%
% SAIDAS: (variavel estruturada modelo)
%
% yun: estimativas da propriedade de interesse para as amostras analisadas;
%
% yc_est: estimativas para a propriedade de interesse para as amostras de calibracao;
%
% nnas_c: norma do sinal analitico liquido para as amostras de
% calibracao [7];
%
% nnas_un: norma do sinal analitico liquido para as amostras analisadas
% [7];
%
% rmsec: raiz quadrada do erro medio quadratico de calibracao com A variaveis latentes;
%

```

% rmsecv: raiz quadrada do erro medio quadratico de validacao cruzada com A variaveis latentes;  
%  
% rep\_cal: estimativa do erro relativo de previsao para as amostras de calibracao (%) [8];  
%  
% SE: estimativa dos erros padroes das amostras de calibracao (cal) e amostras analisadas (un),  
% determinado utilizando a mesma abordagem para calibracao de primeira-ordem [10];  
%  
% lev: leverages dos padroes de calibracao (cal) e amostras (un);  
%  
% reg: vetor que contem os coeficientes de regressao do modelo UPLS  
%  
% residuo: residuos de decomposicao para as amostras de calibracao(residuo\_cal) e para as amostra  
% analisadas (residuo\_un) quando nao ha interferentes nao calibrados;  
%  
% SEN: valores de sensibilidade [6]  
%  
% inv\_SENA: inverso da sensibilidade analitica determinada pela estimativa de ruido ou pelos residuos  
% de decomposicao das amostras caso o ruido nao for fornecido [6]  
%  
% LOD: limite de deteccao pela estimativa de ruido ou pelos residuos de decomposicao das amostras caso  
% o ruido nao for fornecido [9];  
%  
% LOQ: limite de quantificacao pela estimativa de ruido ou pelos residuos de decomposicao das amostras caso  
% o ruido nao for fornecido [9];  
%  
% ruido: estimativa do ruido experimental fornecido;  
%  
% Observacoes, se ni>0:  
% 1)perfis\_1 e perfil\_2 serao os perfis dos analitos e interferentes em cada dimensao estimados  
% pelo RBL para cada amostra;  
% 2)residuo\_uni=residuos de decomposicao para as amostras de analisadas, considerando os interferentes;  
% 3)SEL: valores de seletividade estimados quando ha interferentes nao  
% calibrados;[11]  
% 4)SE\_un\_new: estimativa dos erros padroes das amostras analisadas, que leva em conta o efeito dos  
% interferentes nao calibrados;[11]  
%  
% Referencias (indicadas entre colchetes):  
% 1- B. S. Dayal e J. H. Macgregor; J. Chemom. 1997, 11, 73.  
% 2- J. Ohman, P. Geladi e S. Wold; J. Chemom. 1990, 4, 79.  
% 3- J. Ohman, P. Geladi e S. Wold; J. Chemom. 1990, 4, 135.  
% 4- A. C. Olivieri; J. Chemom. 2005, 19, 253.  
% 5- D. B. Gil, A. M. De La Pena, J. A. Arancibia, G. M. Escandar e A. C. Olivieri; Anal. Chem. 2006, 78, 8051.  
% 6- A. C. Olivieri e N. M. Faber; J. Chemom. 2005, 19, 583.  
% 7- R. Bro e C. M. Andersen; J. Chemom. 2003, 17, 646.  
% 8- N. J. Miller-Ihli, T. C. O'Haver e J. M. Harnly, Spectrochim Acta, 1984, 39(12),1603-1614.  
% 9- L. A. Currie, Anal. Chim. Acta, 1999, 391, 105-126.  
% 10- J. A. Fernandes Pierna, L. Jin, F. Wahl, N. M. Faber, D. L. Massart, Chemom. Intell. Lab. Syst. 2003, 65, 281-  
% 291.  
% 11- J. W. B. Braga; Aplicacao e validacao de modelos de calibracao de segunda ordem em quimica analitica.  
% Tese de Doutorado, Universidade estadual de Campinas, Brasil, 2008.  
%  
% Em caso de duvidas, sugestoes ou qualquer problema na execucao deste programa,  
% por favor, entre em contato com os autores:  
% Jez Willian Batista Braga (email: jezwillian@gmail.com)  
% Ronei Jesus Poppi (e-mail: ronei@iqm.unicamp.br)  
%

```

% LAQQA - Laboratorio de Quimiometria em Quimica Analitica
% Instituto de Quimica - UNICAMP
% 21/01/2008
%

function modelo=upls1_rbl(Xc,yc,A,center,Xun,ni,ruido,sistema,erroY);
[L1,C1,H1]=size(Xc);
[C2]=size(yc,2);
[L3]=size(Xun,1);
[L4]=size(ruido);

% Desdobra as matrizes dos dados experimentais (Lx(CxH))
for i=1:L1;
Xc1(:,i)=reshape(Xc(i,,:),1,C1*H1)';
end
for i=1:L3;
Xun1(:,i)=reshape(Xun(i,,:),1,C1*H1)';
end
Xc=Xc1';Xun=Xun1';clear Xc1 Xun1;

% Centrar os dados na media
if center==1
my=mean(yc);
yc=yc-ones(L1,1)*my;
mx=mean(Xc);
Xc=Xc-ones(L1,1)*mx;
Xun=Xun-ones(L3,1)*mx;
end
%%%% VALIDACAO CRUZADA %%%
%% Modelo U-PLS1 %%%
for j=1:C2
padroes=1:L1;
for z=1:L1
dentro=setdiff(padroes,z);
Xvc=Xc(dentro,:);yvc=yc(dentro,:);
Xi=Xc(z,:);
XY=Xvc'*yvc(:,j);
for i=1:A
if C2==1
w=XY;
else
[C,S,D]=svd(XY'*XY);
q=C(:,1);
w=(XY*q);
end
w=w/sqrt(w'*w);
r=w;
for k=1:i-1
r=r-(P(:,k)*w)*R(:,k);
end
t=Xvc*r;
p=(Xvc*t)/(t'*t);
q=(r'*XY)/(t'*t);
XY=XY-((p*q)'*(t*t))';
W(:,i)=w;
P(:,i)=p;

```

```

Q(:,i)=q';
R(:,i)=r;
T(:,i)=t;
end
b=R*Q';

yc_vc(z,j)=Xi*b;
clear T Q W R P b
end
err=yc(:,j)-yc_vc(:,j);
msecv(1,j)=(err'*err)/L1;
end

%%%%%%%%%%%%%% Modelo U-PLS1 %%%%%%%%%%%%%%%
for j=1:C2
XY=Xc'*yc(:,j);
for i=1:A
if C2==1
w=XY;
else
[C,S,D]=svd(XY'*XY);
q=C(:,1);
w=(XY*q);
end
w=w/sqrt(w'*w);
r=w;
for k=1:i-1
r=r-(P(:,k)*w)*R(:,k);
end
t=Xc*r;
p=(Xc'*t)/(t'*t);
q=(r'*XY)/(t'*t);
XY=XY-((p*q)'*(t'*t))';
W(:,i)=w;
P(:,i)=p;
Q(:,i)=q';
R(:,i)=r;
T(:,i)=t;
end
b(:,j)=R*Q';
SEN0(:,j)=1./diag(sqrt(b(:,j)*b(:,j)));
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
yc_est=Xc*b;
if ni>0
disp(['Especie de Interesse ' num2str(j)]);
for i=1:L3
Peff=[];
Tu=Xun(i,:)*R;
Res1=Xun(i,:)-Tu*P';
sx_un(i,j)=sqrt(sum(Res1.^2)/((C1*H1)-A));
dt=norm(Tu); dtnovo=zeros(1,1); nit=0; % inicializacao das vartiaveis do processo iterativo
while norm(dt-dtnovo)>0.00001*dt
nit=nit+1;
dt=dtnovo;
Res=(Xun(i,:)-Tu*P');
[u0,s0,v0]=svd(reshape(Res,C1,H1),0); % obtem os perfis dos interferentes

```

```

Xmod=Xun(i,:)-reshape((u0(:,1:ni)*s0(1:ni,1:ni)*v0(:,1:ni)'),1,C1*H1);
Tu=Xmod*R;
Res1=Xun(i,:)-Tu*P'-reshape((u0(:,1:ni)*s0(1:ni,1:ni)*v0(:,1:ni)'),1,C1*H1);
dtново=norm(Tu);
if nit==1000
break
end
end
end
Ynl=Xun(i,:);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
if sistema=='matlab';
Xnl=zeros(size(Ynl));
save temp3mat.mat Ynl Xnl P R ni C1 H1
beta0=Tu;
[Tu]=nlinfit(Xnl,Ynl,'func_obj_mat_rbl_upls',beta0);
Tun(i,:)=Tu';
resi=Ynl-(P*Tu);
XT=reshape(resi,C1,H1);
[u0,s0,v0]=svd(XT,0);
perfil_1=u0(:,1:ni);perfil_1=perfil_1';
perfil_2=v0(:,1:ni)*s0(1:ni,1:ni);perfil_2=perfil_2';
Xint=u0(:,1:ni)*s0(1:ni,1:ni)*v0(:,1:ni)';
Xmod=(P*Tu)+ Xint(:);
resi2=Ynl-Xmod';
sx_uni(i,j)=sqrt(sum(resi2.^2)/((C1*H1)-(A+ni)));

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
else sistema=='octave';
save -binary temp3 Ynl P R ni C1 H1 A
Xnl=zeros(size(Ynl));
beta0=Tu;
[Tu, obj, info, iter, nf, lambda] = sqp (beta0', @func_obj_oct_rbl_upls, [], []);
Tun(i,:)=Tu';
resi=Ynl-(P*Tu);
XT=reshape(resi,C1,H1);
[u0,s0,v0]=svd(XT,0);
perfil_1=u0(:,1:ni);perfil_1=perfil_1';
perfil_2=v0(:,1:ni)*s0(1:ni,1:ni);perfil_2=perfil_2';
Xint=u0(:,1:ni)*s0(1:ni,1:ni)*v0(:,1:ni)';
Xmod=(P*Tu)+ Xint(:);
resi2=Ynl-Xmod';
sx_uni(i,j)=sqrt(sum(resi2.^2)/((C1*H1)-(A+ni)));
end

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
Peff=[];
for k=1:A
PbP=reshape(P(:,k),C1,H1)-perfil_1'*pinv(perfil_1)*reshape(P(:,k),C1,H1);
PcP=PbP-PbP*perfil_2'*pinv(perfil_2');
Peff=[Peff,PcP(:)];
end

SEN(i,j)=1/norm(pinv(Peff)*Q);

clear Peff

```

```

yun(i,j)=Tun(i,:)*Q';

% Calculo do sinal analítico líquido (escalar)
nmas_c(:,j)=yc_est(:,j).*(SEN(i,j));
nmas(i,j)=yun(i,j).*SEN(i,j);
end

if center ==1
Z(1:L1,1)=1; Z(:,2)=yc_est(:,j)+ones(L1,1)*my(j);
cte=pinv(Z)*nmas_c(:,j);
nmas(:,j)=nmas(:,j)-cte(1,1);
nmas_c(:,j)=nmas_c(:,j)-cte(1,1);
Xun2=Xun+ones(L3,1)*mx;
else
Xun2=Xun;
end

% RESIDUOS DE DECOMPOSICAO p/ Calibração
E=(Xc-T*P');

for i=1:L1
sx_cal(i,j)=sqrt(sum(E(i,:).^2)/(C1*H1-A));
end

% Leverages do conjunto de calibração
for i=1:L1
hc(i,j)=T(i,1:A)*inv(T(:,1:A)'*T(:,1:A))*T(i,1:A);
end

% Leverages do conjunto de validação
for i=1:L3
hv(i,j)=Tun(i,1:A)*inv(T(:,1:A)'*T(:,1:A))*Tun(i,1:A);
end

% FIGURAS DE MERITO
if L4~=0 % Obtidas a partir do ruído instrumental
inv_SENA(:,j)=ruído./SEN(:,j);
LOD(:,j)=3.3*inv_SENA(:,j);
LOQ(:,j)=10*inv_SENA(:,j);
end

for i=1:L3;
inv_SENAx(i,j)=sx_uni(i,j)./SEN(i,j); %obtidas a partir do residuo de decomposicao
end
LODx(:,j)=3.3*inv_SENAx(:,j);
LOQx(:,j)=10*inv_SENAx(:,j);

for i=1:L3
SEL(i,j)=SEN(i,j)/SEN0(1,j);
end

elseif ni==0

% RESIDUOS DE DECOMPOSICAO
% Calibração
E=(Xc-T*P');

```

```

for i=1:L1
sx_cal(i,j)=sqrt(sum(E(i,:).^2)/(C1*H1-A));
end

% Conjunto teste
Tun=Xun*R;
E=(Xun-Tun*P');

for i=1:L3
sx_un(i,j)=sqrt(sum(E(i,:).^2)/(C1*H1-A));
end

% Leverages do conjunto de calibracao
for i=1:L1
hc(i,j)=T(i,1:A)*inv(T(:,1:A)*T(:,1:A))*T(i,1:A);
end

% Leverages do conjunto de validacao
for i=1:L3
hv(i,j)=Tun(i,1:A)*inv(T(:,1:A)*T(:,1:A))*Tun(i,1:A);
end

yun=Xun*b;           %Estimativa das concentrações nas amostras

% FIGURAS DE MERITO
SEN=1./diag(sqrt(b*b));

% Calculo do sinal analitico liquido (escalar)
nmas_c(:,j)=yc_est(:,j).*SEN(j,:);
nmas(:,j)=yun(:,j).*SEN(j,:);

if center ==1
Z(1:L1,1)=1;  Z(:,2)=yc_est(:,j)+ones(L1,1)*my(j);
cte=pinv(Z)*nmas_c(:,j);
nmas(:,j)=nmas(:,j)-cte(1,1);
nmas_c(:,j)=nmas_c(:,j)-cte(1,1);
Xun2=Xun+ones(L3,1)*mx;
else
Xun2=Xun;
end

if L4~=0 % Obtidas a partir do ruido instrumental
inv_SENA=ruido./SEN;
LOD=3.3*inv_SENA;
LOQ=10*inv_SENA;
end

for i=1:L3;
inv_SENAx=sx_un(i,j)./SEN;      %obtidas a partir do residuo de decomposicao
end
LODx=3.3*inv_SENAx;
LOQx=10*inv_SENAx;
end
end

```

```

% Estimativa das concentracoes
yc_est=Xc*b;
ycc=yc;
%volta a media
if center==1
yc_est=yc_est+ones(L1,1)*my;
yun=yun+ones(L3,1)*my;
yc=yc+ones(L1,1)*my;
end

% Numero de graus de liberdade perdidos na construcao do modelo
neff1(1,1:C2)=A;
for i=1:C2
res=yc(:,i)-yc_est(:,i);
msec0(1,i)=(res'*res)/L1;
neff2(1,i)=L1*(1-sqrt(msec0(1,i)/msecv(1,i)));neff2(1,i)=round(neff2(1,i));
if neff2(1,i)>neff1(1,i);
neff(1,i)=neff2(1,i);
else
neff(1,i)=neff1(1,i);
end
end

if center ==1;DF=neff+1;else;DF=neff;end
for i=1:C2
res=yc(:,i)-yc_est(:,i);
msec(1,i)=(res'*res)/(L1-DF(1,i));
rep(1,i)=sqrt(sum((res.^2)./(yc(:,i).^2*(L1-DF(1,i)))))*100;
end
if nargin >8
vy=erroY.^2;
else
vy(1,C2)=0;
end
clear SE_un SE_cal

if L4==0
ruído=mean(sx_cal(:));
end

if ni==0

% Calculo do erro padrao de previsao [igual a calibracao de primeira ordem]
if center==1
for i=1:L1
for j=1:C2
SE_cal(i,j)=sqrt((1+(1/L1)+hc(i,j))*msec(1,j));
end
end
for i=1:L3
for j=1:C2
SE_un(i,j)=sqrt((1+(1/L1)+hv(i,j))*msec(1,j));
end
end
else
for i=1:L1

```

```

for j=1:C2
SE_cal(i,j)=sqrt((1+hc(i,j))*msec(1,j));
end
end
for i=1:L3
for j=1:C2
SE_un(i,j)=sqrt((1+hv(i,j))*msec(1,j));
end
end
end
else
if center==1;DF=A+1;else; DF=A;end
nb0=1./(SEN0.^2);
if center == 1
for j=1:C2
for i=1:L1
SE_cal(i,j)=sqrt((1+(1/L1)+hc(i,j))*msec(1,j));
end
for i=1:L3
nb=1./(SEN(i,j)^2);
Verr(1,j)=msec(1,j)-( nb0(1,j)*(ruído^2) + vy(1,j));
if Verr(1,j)<0; Verr(1,j)=0;end
SE_un_new(i,j)=sqrt((hv(i,j)+(1/L1))*msec(1,j) + Verr(1,j) + nb*(ruído^2));
SE_un(i,j)=sqrt((1+(1/L1)+hv(i,j))*msec(1,j));
end
end
else
for j=1:C2
for i=1:L1
SE_cal(i,j)=sqrt((1+hc(i,j))*msec(1,j));
end
for i=1:L3
nb=1./(SEN(i,j)^2);
Verr(1,j)=msec(1,j)-( nb0(1,j)*(ruído^2) + vy(1,j));
if Verr(1,j)<0; Verr(1,j)=0;end
SE_un_new(i,j)=sqrt((hv(i,j))*msec(1,j) + Verr(1,j) + nb*(ruído^2));
SE_un(i,j)=sqrt((1+hv(i,j))*msec(1,j));
end
end
end
end

% FIGURAS
figure;plot(sx_cal,'ko');
hold on;plot(sx_un,'m*');
sxm(1:L3)=mean(sx_cal(:));
hold on;plot(sxm*2.326,'--');
if nargin>=7
RUIDO(1:L3)=ruído;
hold on;plot(RUIDO*2.326,'-');
end
if ni>0;hold on;plot(sx_uni,'^');end
title('Resíduos de decomposicao');xlabel('Numero da amostra');ylabel('Resíduos');
figure
for i=1:C2
hold on;subplot(1,C2,i);plot(nnas_c(:,i),yc(:,i),'o');

```

```
hold on;subplot(1,C2,i);plot(nnas(:,i),yun(:,i),'*k');  
xlabel('Sinal analitico liquido');ylabel('Valores de referencia (o) e estimados (*)');  
end
```

```
% Salva saidas  
modelo.yun=yun;  
modelo.yc_est=yc_est;  
modelo.nnas_c=nnas_c;  
modelo.nnas_un=nnas;  
modelo.rmsec=sqrt(msec);  
modelo.rmsecv=sqrt(msecv);  
modelo.rep_cal=rep;  
modelo.SE_cal=SE_cal;  
modelo.SE_un=SE_un;  
if ni>0;modelo.SE_un_new=SE_un_new;end  
modelo.lev_cal=hc;  
modelo.lev_un=hv;  
modelo.reg=b;  
modelo.residuo_cal=sx_cal;  
modelo.residuo_un=sx_un;  
if ni>0  
modelo.residuo_uni=sx_uni;  
modelo.perfil_1=perfil_1;  
modelo.perfil_2=perfil_2;  
modelo.SEL=SEL;  
end  
modelo.SEN=SEN;  
if L4~=0  
modelo.inv_SENA=inv_SENA;  
modelo.LD=LOD;  
modelo.LQ=LOQ;  
modelo.ruido=ruido;  
else  
modelo.inv_SENA=inv_SENAx;  
modelo.LD=LODx;  
modelo.LQ=LOQx;  
end  
%modelo.W=W;  
%modelo.P=P;  
%modelo.Q=Q;  
%modelo.R=R;
```

**B7 Sub-rotina func\_obj\_mat\_rbl\_upls: utilizada no cálculo do RBL para o modelo UPLS no sistema Matlab 6.5.**

```
% Sub-rotina que calcula a função objetivo para a minimização da matriz de  
% erros pelo RBL no modelo UPLS1, onde os erros de decomposição são minimizados  
% por  $X_{un} = P * T_{un} + \text{interferentes}$ . (Sistema Matlab 6.5)
```

```
function f = func_obj_mat_rbl_upls(beta0,Xnl)
```

```
load temp3mat  
Tnovo=beta0;  
resi=Ynl-(P*Tnovo)';  
XT=reshape(resi,C1,H1);  
[u0,s0,v0]=svd(XT,0);  
Xint=u0(:,1:ni)*s0(1:ni,1:ni)*v0(:,1:ni)';  
f=(P*Tnovo)+ Xint(:);
```

**B8 Sub-rotina func\_obj\_oct\_rbl\_upls.m: utilizada no cálculo do RBL para o modelo UPLS no sistema Octave 3.0.0.**

```
% Sub-rotina que calcula a função objetivo para a minimização da matriz de  
% erros pelo RBL no modelo UPLS1, onde os erros de decomposição são minimizados  
% por  $X_{un}=P*T_{un} + \text{interferentes}$ . (Sistema Octave 3.0.0)
```

```
function err = func_obj_oct_rbl_upls(beta0)
```

```
load temp3  
Tnovo=beta0;  
residuo=Ynl-(P*Tnovo)';  
Xres=reshape(residuo,C1,H1);  
[u0,s0,v0]=svd(Xres,0);  
Xint=u0(:,1:ni)*s0(1:ni,1:ni)*v0(:,1:ni)';  
f=(P*Tnovo)+ Xint(:);  
resi2=Ynl-f';  
err=sqrt(sum(resi2.^2)/((C1*H1)-(A+ni)));
```

## B9 Programa parafacpred.m: estima a concentração e figuras de mérito de uma amostra que é analisada.

```

% PARAFACPREP: programa que estima a concentracao e figuras de merito de uma amostra
% de previsao atraves de um modelo PARAFAC {1,2} previamente construido atraves da rotina parafac.m
% distribuida no nway-toobox2.11(http://www.models.kvl.dk/source/). Onde a primeira matriz do
% cubo de dados deve ser a amostra de previsao e as demais amostras as que sao utilizadas na
% etapa de calibracao.
% A utilizacao e distribuicao deste programa e livre e permitida desde que devidamente
% citada a sua fonte {8}. Contudo, os autores nao oferecem nenhuma garantia da sua funcionalidade
% ou ajuste na analise de qualquer conjunto de dados.
%
% SINTAXE : result = parafacpred(Factors,Ycal,analyte,inter,err,scal,ruido);
%
% ENTRADAS:
%
% Factors = matriz celular que contem os fatores obtidos pelo modelo parafac
% (programa parafac.m). Onde a primeira dimensao deve corresponder a relacionada a
% concentracao e a primeira matriz do cubo de dados deve ser a amostra
% de previsao e as demais amostras as que sao utilizadas na etapa de calibracao..
%
% Ycal = vetor (ou matriz) contendo os valores de referencias para as concentracoes nas amostras
% de calibracoes.
%
% analyte = vetor que informa quais componentes do modelo correspondem aos analitos calibrados
% Se em um modelo de 3 componentes os fatores 1 e 3 são os analitos calibrados, o vetor
% analyte =[1 3].
%
% inter = identifica o componente que diz respeito a um composto não calibrado. No exemplo anterior,
% inter =[3]. Se nao ha compostos nao calibrados inter =[].
%
% err = erro de decomposicao, obtido atraves da rotina parafac.m
%
% scal = estimativa do desvio padrao dos erros das concentracoes de referência. Caso nao
% for disponivel utilizar [].
%
% ruido = estimativa do ruido instrumental dos dados.Caso nao for disponivel omitir entrada.
%
% SAIDAS: variavel estruturada result
%
% yc_est :concentracoes previstas para as amostras de calibracao
%
% rmsec: raiz quadrada do erro medio quadratico de calibracao
%
% rep: estimativa do erro relativo de previsao das amostras de
% calibracao (%) {6}
%
% yun: concentracoes previstas para as amostras de validacao
%
% SE: estimativa dos erros padroes das amostras de calibracao (cal) e amostras analisadas (un),
% determinado utilizando a mesma abordagem para calibracao de ordem zero por quadrados
% minimos classico [10];
%

```

```

% leverage_un: leverages ("influencia") das amostras analisadas;
%
% SEN = valores de sensibilidade {3}
%
% SEL = valores de seletividade
%
% SENA= sensibilidade eanalitica {4}
%
% LD = limite de deteccao {5}
%
% LQ = limite de quantificacao {5}
%
% ruído = estimativa do ruído dos dados que foi fornecido.
%
% Referencias (indicadas entre chaves: {}):
% 1- R. Bro, Chemom. Intell. Lab. Syst., 1997, 38, 149.
% 2- N. M. Faber, R. Bro e P. K. Hopke; Chemom. Intell Lab. Syst. 2003, 65, 119.
% 3- A. C. Olivieri e N. M. Faber; J. Chemom. 2005, 19, 583.
% 4- A. Munoz de la Pena, A. E. Mansilla, D. G. Gomes, A. C. Olivieri, H. C. Goicoechea, Anal. Chem. 2003,
% 75, 2640-2646
% 5- L. A. Currie, Anal. Chim. Acta, 1999, 391, 105-126.
% 6- N. J. Miller-Ihli, T. C. O'Haver e J. M. Harnly, Spectrochim Acta, 1984, 39(12),1603-1614.
% 7- J. N. Miller e J. C. Miller; Statistics and chemometrics for analytical chemistry.
% 4ª ed. Prentice Hall, New York, 2000, cap.5.
% 8- J. W. B. Braga; Aplicacao e validacao de modelos de calibracao de segunda ordem em quimica analitica.
% Tese de Doutorado, Universidade estadual de Campinas, Brasil, 2008.
%
% Em caso de duvidas, sugestoes ou qualquer problema na execucao deste programa,
% por favor, entre em contato com os autores:
% Jez Willian Batista Braga (email: jezwillian@gmail.com)
% Ronei Jesus Poppi (e-mail: ronei@iqm.unicamp.br)
%
% LAQQA - Laboratorio de Quimiometria em Quimica Analitica
% Instituto de Quimica - UNICAMP
% 21/01/2008
%
function result = parafacpred(Factors,Ycal,analyte,inter,err,scal,ruído);

% Obtem loadings nos tres modos
A=Factors{1};
B=Factors{2};
C=Factors{3};

nanal=size(Ycal,2);
[nsam,ncomp]=size(A);
dim1=size(B,1);
dim2=size(C,1);
int=size(inter,2);
YY(1:(nsam-1),1)=1;

% Calculo da concentracao da amostra
for i=1:nanal
YY(:,2)=Ycal(:,i);
coefang(:,i)=pinv(YY)*A(2:nsam,analyte(i));
Ae=YY*coefang(:,i);
syx2(i)=sum((A(2:nsam,analyte(i))-Ae(:,1)).^2)/(nsam-2);

```

```

Yt(:,i)=(A(:,analyte(i))-coefang(1,i))./coefang(2,i);
end
yc_est=Yt(2:nsam,:);
yun=Yt(1,:);

% Perfis nos modos 2 e 3 dos analitos calibrados
Bcal=[];Ccal=[];
for i=1:nanal
Bcal(:,i)=B(:,analyte(i));
Ccal(:,i)=C(:,analyte(i));
end

% Perfis nos modos 2 e 3 dos interferentes
Bint=[];Cint=[];
if int~=0
for i=1:int
Bint(:,i)=B(:,inter(i));
Cint(:,i)=C(:,inter(i));
end
end

% Calculo de figuras de merito
for i=1:nanal
h(i)=yun(i).^2/sum(Ycal(:,i).^2);
res=Ycal(:,i)-yc_est(:,i);
rmsec(i)=sqrt((res*res)/(nsam-1-ncomp));
rep(i)=sqrt(sum((res.^2)/((Ycal(:,i).^2)*(nsam-1-ncomp))))*100;
end

if size(scal,1)==0;sc=rmsec;else; sc=scal.^2;end
ruidox=sqrt(err/((nsam-1)*(dim1)*(dim2)));

if length(Bint)==0
Pb_int=eye(dim1);
Pc_int=eye(dim2);
else
Pb_int=eye(dim1)-Bint*pinv(Bint);
Pc_int=eye(dim2)-Cint*pinv(Cint);
end

ii=inv((Bcal'*Pb_int*Bcal).*(Ccal'*Pc_int*Ccal));

for i=1:nanal
sen(i)=coefang(2,i)/sqrt(ii(i,i));
sel(i)=1/sqrt(ii(i,i));
if nargin==7
sen_anal(1,i)=sen(i)/ruido;
LOD(1,i)=3.3/sen_anal(1,i);
LOQ(1,i)=10/sen_anal(1,i);
else
sen_anal(i)=sen(i)/ruidox;
LOD(i)=3.3/sen_analx(i);
LOQ(i)=10/sen_analx(i);
end
SD_uni(1,i)=(sqrt(syx2(i))/coefang(2,i))*sqrt(1+(1/(nsam-1))+((A(1,i)-mean(A(2:nsam,i)))^2)/((coefang(2,i).^2)*sum((Ycal(:,i)-mean(Ycal(:,i))).^2))));
end

```

```
% Salva os resultados
result.yc_est=yc_est;
result.rmsec=rmsec;
result.rep=rep;
result.yun=yun;
result.SE_uni=SD_uni;
result.leverage_un=h;
result.SEN=sen;
result.SEL=sel;
result.SENA=sen_anal;
result.LD=LOD;
result.LQ=LOQ;
if nargin==7;
result.ruido=ruido;
end
```

## B10 Programa ortimef.m: realiza o alinhamento dos tempos de retenção de matrizes de dados.

```

% Este programa realiza o alinhamento dos tempos de retencao de matrizes de dados com base na menor
% percentagem de variancia residual obtida por decomposicao de valores
% singulares segundo o metodo proposto por Prazen et al.
% A utilizacao e distribuicao deste programa e livre e permitida desde que devidamente
% citada a sua fonte [3]. Contudo, os autores nao oferecem nenhuma garantia da sua funcionalidade
% ou ajuste na analise de qualquer conjunto de dados.
%
% Sintaxe: saidas=ortimef(N,X,janela);
%
% Entradas:
%   N: matriz de referencia;
%   X: tensor correspondente a uma amostra;
%   JANELA: vetor que contem o intervalo do numero de pontos que o algoritmo tentara corrigir os desvios.
%   Por exemplo: janela =0:1:10, implica que o algoritmo procurara corrigir desvios de 0 a 10
%   pontos para a direita.
%   POSTO (imput): e o pseudo-posto da matriz, que idealmente deve ser igual ao numero de especies
%   quimicas independentes presentes na amostra;
% saidas (variavel estruturada):
%   res: matriz que contem as percentagens de variancia residual para o intervalo da janela para cada amostra;
%   Xoo e Xo: tensores alinhadas, com a mesma dimensao e com dimensao menor que os dados originais.
%   auto: autovalores das amostras corrigidas
%   autoS: autovalores da amostra tomada como referencia
%   posto: numero de autovalores utilizado na correcao
%   janela: janela utilizada na correcao
%
% REFERENCIAS:
% 1- B.J.Prazen, R.E. Synovec E R.B. Kowalski; Anal. Chem.,1998, 70, 218.
% 2- C. G. Fraga, B. J. Prazen e R. E. Synovec; Anal. Chem. 2001, 73, 5833.
% 3- J. W. B. Braga; Aplicacao e validacao de modelos de calibracao de segunda ordem em quimica analitica.
% Tese de Doutorado, Universidade estadual de Campinas, Brasil, 2008.
%
% Em caso de duvidas, sugestoes ou qualquer problema na execucao deste programa,
% por favor, entre em contato com os autores:
% Jez Willian Batista Braga (email: jezwillian@gmail.com)
% Ronei Jesus Poppi (e-mail: ronei@iqm.unicamp.br)
%
% LAQQA - Laboratorio de Quimiometria em Quimica Analitica
% Instituto de Quimica - UNICAMP
% 21/01/2008
%

function saidas=ortimef(N,X,janela);

[L2,C2,H2]=size(X);
[L1,C1]=size(N);
j2=janela;
[Lj2,Cj2]=size(j2);

[u,s,v]=svd(N);S=s.^2;autoS=diag(S)';

```

```

for i=1:L2
[u,s,v]=svd(squeeze(X(i,:,:)));
S=s.^2;
auto(i,:)=diag(S)';
end
disp(['Autovalores na matriz de referencia:'])
disp(autoS(1,1:10))
disp(['Autovalores nas amostras a serem corrigidas:'])
disp(auto(:,1:10))
posto=input('Informe o numero de autovalores significativos (POSTO):');

% Inicializacao da matrizes utilizadas no calculo
M=zeros(L1,C1);
Mi=zeros(L1,C1);
Ni=zeros(L1,C1);
R=[Mi Ni];
Xo=zeros(L2,C2,H2);
u=zeros(C2,C2);
s=zeros(C2,H2);
v=zeros(H2,H2);

tic
% Inicio do algoritmo
for k=1:L2
M=squeeze(X(k,:,:));
j=1;
for i=1:Cj2

if j2(i)<0                                % Cortar o inicio de M e adicionar no final
Ni=N(1:L1+j2(i),:);                       % Mantem-se N constante
Mi=M(1-j2(i):L1,:);                       % Desloca-se M
R=[Mi Ni];
[L,C]=size(R);
[u,s,v]=svd(R);
S=s.^2;
dS=diag(S);
err(j)=100*(sum(dS(posto:min(L,C)))/sum(dS))*((L*C)/((L-posto)*(C-posto)));    % Calculo de residuos

elseif j2(i)== 0                          % Nao fazer nada
Ni=N(1:L1,:);                             % N original
Mi=M(1:L1,:);                             % M original
R=[Mi Ni];
[L,C]=size(R);
[u,s,v]=svd(R);
S=s.^2;
dS=diag(S);
err(j)=100*(sum(dS(posto:min(L,C)))/sum(dS))*((L*C)/((L-posto)*(C-posto)));
else                                       % Cortar no final e adicionar no inicio
Ni=N(1+j2(i):L1,:);                       % Desloca-se N
Mi=M(1:L1-j2(i),:);                       % Mantem-se M constante
R=[Mi Ni];
[L,C]=size(R);
[u,s,v]=svd(R);
S=s.^2;
dS=diag(S);
err(j)=100*(sum(dS(posto:min(L,C)))/sum(dS))*((L*C)/((L-posto)*(C-posto)));

```

```

end

j=j+1;
end
res(k,:)=err;
j=1;
disp(k);
opt=find(err==min(err)); % Encontra o deslocamento que proporciona o residuo minimo
d=j2(opt);
di(k)=d;

% Corrige os desvios nos tempos de retencao de acordo com cada caso

if d<0 % Corta ni inicio e acrescenta no final
Xoo(k,1:L1+d,:)=M(1-d:L1,:);
Xoo(k,L1+d+1:L1,:)=0;
elseif d==0 % Nao ha alteracao
Xoo(k,:)=M;
elseif d>0 % acrescenta no inicio e corta no final
Xoo(k,1:d,1:C1)=0;
Xoo(k,d+1:L1,:)=M(1:L1-d,:);
end
% plot(log,err)
end
mdi=abs(max(di));
Xo=Xoo(:,1+mdi:C2-(mdi+1),:);

%monta variavel estruturada de saida
saidas=struct('Xo',Xo, 'Xoo',Xoo,'res',res,'auto',auto,'autoS',autoS,'posto',posto,'janela',janela);
toc

```