

**Universidade Estadual de Campinas**  
**Instituto de Química**  
**Departamento de Físico-Química**

**TESE DE DOUTORADO**

**ANÁLISE MULTIVARIADA EM N-MODOS: METODOLOGIAS  
PARA O MONITORAMENTO DE REAÇÕES QUÍMICAS**

**Autor: Marcos Viero Guterres**

**Orientadora: Márcia Miguel Castro Ferreira**

**Co-orientador: Pedro Luiz Onófrio Volpe**

**Campinas, agosto de 2006**

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

G982a Guterres, Marcos Viero  
Análise multivariada em N- modos: metodologias para o monitoramento de reações químicas / Marcos Viero Guterres. -- Campinas, SP: [s.n] , 2006.

Orientadora: Márcia Miguel Castro Ferreira.

Co-orientador: Pedro Luiz Onófrío Volpe.

Tese - Universidade Estadual de Campinas, Instituto de Química.

1. Quimiometria. 2. Análise multivariada. 3. Análise em multi modos. I. Ferreira, Márcia Miguel Castro. II. Volpe, Pedro Luiz Onófrío. III. Universidade Estadual de Campinas. Instituto de Química. IV. Título.

**Título em inglês:** N- way analysis: methodologies for the monitoring of chemical reactions

**Palavras-chaves em inglês:** Chemometrics, Multivariate analysis, Analysis in multi ways

**Área de concentração:** Físico-Química

**Titulação:** Doutor em Ciências

**Banca examinadora:** Márcia Miguel Castro Ferreira (orientadora), Célio Pasquini, Francisco Benedito Teixeira Pessine, Roberto Kawakami Harrop Galvão, Paulo Gardel Kurka, Ivo Milton Raimundo Júnior, Yuji Takahata.

**Data de defesa:** 23/08/2006

À Tais e Pedro, minha família.

*Dedico com carinho este trabalho.*

Se você não tem oportunidade de fazer grandes coisas, pode fazer pequenas coisas de uma forma grandiosa.

## Agradecimentos

À minha orientadora Márcia Miguel Castro Ferreira, pelo trabalho dedicado a minha formação científica e profissional.

Ao meu co-orientador Pedro Luiz Onófrio Volpe pelo companheirismo e dedicação no desenvolvimento deste trabalho.

Ao professor José Cardoso de Menezes do Instituto Superior Técnico de Lisboa (Portugal) pela contribuição prestada para este trabalho.

Aos amigos do grupo de Quimiometria Teórica e Aplicada pelos anos de convívio.

Aos amigos do Instituto Superior Técnico de Lisboa pela amizade e convívio.

À Denise Aparecida de Oliveira pela dedicação no trabalho experimental das fermentações com a *Chromobacterium violaceum*.

Por fim, à FAPESP pelo apoio financeiro.

## CURRICULUM VITAE

### FORMAÇÃO ACADÊMICA:

#### **Graduação:**

03/1993- 01/1997 - Química Industrial. Universidade Federal de Santa Maria -UFSM, Santa Maria – RS. Brasil.

#### **Pós Graduação:**

03/1997-04/1999 - Mestrado em Química. Universidade Federal de Santa Maria – UFSM. Santa Maria – RS. Brasil. Área: Química Analítica. (Bolsa CAPES).

08/2000–08/2006 - Doutorado em Química. Universidade Estadual de Campinas-UNICAMP. Campinas – SP. Brasil. Área: Físico Química. (Bolsa FAPESP).

### ATIVIDADES ACADÊMICAS:

09/2006-03/2004 Estágio de seis meses realizado no Centro de Biologia e Engenharia Química do Instituto Superior Técnico – Lisboa. Portugal. Orientador: José Cardoso de Menezes.

04/1995-01/1996 Iniciação Científica. Projeto: Complexos metálicos com ligantes bis(feniltriazeno) benzeno/naftaleno e hidrazinacarbotiamida- difração de raios X em cristais. Orientador: Manfredo Horner.

04/1996-03/1997 Iniciação Científica. Projeto: Determinação de sulfeto por microdifusão em fluxo. Orientador: Paulo Cícero do Nascimento.

### PRODUÇÃO CIENTÍFICA:

#### **Trabalhos resumidos em eventos:**

1. GUTERRES, Marcos Viero; FELÍCIO, Candida; LOPES, João Almeida; MENEZES, José Cardoso. A; *Chemometric method for monitoring batch chemical reactions with spectroscopy*. NINTH INTERNATIONAL CONFERENCE ON CHEMOMETRICS IN ANALYTICAL CHEMISTRY, 2004, Lisboa - Portugal.

2. GUTERRES, Marcos Viero; VOLPE, Pedro Luis Onofrio; FERREIRA, Marcia Miguel Castro. *Analysis of Pigments Produced by Chromobacterium Violaceum in Aerobic Fermentations with Parafac Models and 2D Fluorescence*. NINTH INTERNATIONAL CONFERENCE ON CHEMOMETRICS IN ANALYTICAL CHEMISTRY, 2004, Lisboa - Portugal.

3. NASCIMENTO, Paulo Cícero Do; JOST, Cristiane; GUTERRES, Marcos Viero; CARVALHO, Leandro. *Simultaneous Determination of Al (III) and Fe (III) in post-Hemodialysis Fluids by Spectrometry and Multivariate Calibration*. SECOND INTERNATIONAL SYMPOSIUM ON TRACE ELEMENTS AND MINERALS IN MEDICINE AND BIOLOGY, 2004, Munich - Alemanha.

4. GUTERRES, Marcos Viero; VOLPE, Pedro Luis Onofrio; FERREIRA, Márcia Miguel Castro. *Multway Calibration for Creatinine Determination by the Jaffé Method in Human Serum*. EIGHTH INTERNATIONAL CONFERENCE ON CHEMOMETRICS IN ANALYTICAL CHEMISTRY, 2002, Seattle – Estados Unidos.

5. GUTERRES, Marcos Viero; MARQUES, Mariele; NASCIMENTO, Paulo Cícero Do. *Determinação Espectrofotométrica Simultânea de Zn, Cu e Mn com um Agente Complexante em Meio Micelar – Uma contribuição ao controle de fluidos de hemodialise*. 22<sup>a</sup> REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE QUÍMICA, 1999, Poços de Caldas – MG, Brasil.

6. GUTERRES, Marcos Viero; TREVISAN, Marcelo; RAMIRES, Adrian; NASCIMENTO, Denise Boher Do; NASCIMENTO, Paulo Cícero Do. *Determinação Polarográfica Simultânea de Traços de Cu, Zn, Pb e Cd em Sais de Qualidade p.a Após Extração em colunas de Polietileno em Pó condicionadas com um Agente Complexante*. 21<sup>a</sup> REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE QUÍMICA, 1998, Poços de Caldas – MG, Brasil.

7. BOHER, Denise; NASCIMENTO, Paulo Cícero Do; SEIBERT, Edson; GUTERRES, Marcos Viero. *Determination of Pb, Cd, Cu and Zn by EAAS in Concentrates for Hemodialysis after Separation on Polyethylene Powder Columns Conditioned with a Complexing Agent*. FIFTH RIO SYMPOSIUM ON ATOMIC SPECTROMETRY, 1998, Cancun - México.

#### **Artigos completos publicados em periódicos:**

1. GUTERRES, Marcos Viero; VOLPE, Pedro Luiz Onofrio; FERREIRA, Marcia Miguel Castro. *Multway calibration for creatinine determination in human serum using the Jaffé reaction*. Applied Spectroscopy, v.58, n. 1, p. 54-60, 2004.

2. GUTERRES, Marcos Viero; BOHER, Denise; TREVISAN, Marcelo; NASCIMENTO, Paulo Cícero Do. *Quality Control of dialysis fluids by micellar colorimetric and multivariate calibration*. Analytical Letters, v.34, n. 11, p. 1967-1978, 2001.

3. BOHER, Denise; NASCIMENTO, Paulo Cicero Do; GUTERRES, Marcos Viero; SEIBERT, Edson. *Electrothermal atomic absorption spectrometric determination of lead, cadmium, copper and zinc in high-salt content samples after simultaneous separation on polyethylene powder impregnated with PAN: application to the analysis of hemodialysis fluids*. Analyst, v. 124, p. 1345-1350, 1999.

## RESUMO

### ANÁLISE MULTIVARIADA EM N-MODOS: METODOLOGIAS PARA O MONITORAMENTO DE REAÇÕES QUÍMICAS

Neste trabalho de tese a análise multivariada foi usada para desenvolver metodologias para o monitoramento de reações químicas. Duas técnicas de análise foram usadas: a Resolução multivariada de curvas (MCR) para a análise de medidas espectroscópicas e o Controle Estatístico de processos (CEMP) para a análise de variáveis físico-químicas coletadas de um processo industrial. O primeiro trabalho desenvolvido nesta tese usou os métodos da MCR para construir modelos empíricos de reações químicas a partir de medidas de reflectância no espectro de infravermelho médio. Este modelo foi usado para prever o comportamento de novas bateladas da mesma reação química. Os resultados obtidos mostraram que as previsões são mais exatas quando o modelo é adaptado de acordo com o método "Predicted Augmented Classical Least Squares" (PACLS) para incluir uma componente que modela as mudanças da linha base dos espectros de infravermelho. O segundo trabalho apresenta uma nova metodologia desenvolvida para determinar a creatinina no soro humano, baseada no monitoramento da reação de Jaffé com espectroscopia visível e análise dos dados com os métodos de decomposição trilinear da MCR. De acordo com os resultados obtidos, a creatinina pode ser determinada na presença de outras substâncias que interferem com a sua medida espectroscópica. O erro padrão de desempenho (SEP) da metodologia proposta neste trabalho foi de  $2,22 \text{ mg L}^{-1}$  e um respectivo coeficiente de variação em torno de 6 %. O terceiro trabalho apresentado nesta tese descreve uma metodologia desenvolvida para investigar os metabólitos fluorescentes e intracelulares produzidos pela *Chromobacterium violaceum* durante o cultivo desta bactéria em um bioreator sob condições aeróbicas. Os espectros de fluorescência coletados a partir de amostras de biomassa foram analisados com o método PARAFAC (Análise de Decomposição em Fatores Paralelos). Os aminoácidos triptofano e tirosina, além de NADH e violaceína foram identificados através da análise dos parâmetros do modelo trilinear ajustado a partir dos espectros de fluorescência. As estimativas obtidas para a variação da concentração das espécies químicas ao longo do tempo, mostraram que esta bactéria consome o triptofano antes de começar a produzir progressivamente a violaceína. Este resultado está de acordo com o mecanismo proposto na literatura para biosíntese deste pigmento, que é um candidato para o tratamento de tumores e doenças tropicais. O quarto trabalho desta tese apresenta uma metodologia de CEMP desenvolvida para a análise de dados autocorrelacionados. Esta metodologia foi usada para monitorar a etapa de digestão da bauxita usada para extrair o alumínio deste minério. Quatro variáveis físico químicas deste processo foram analisadas com os modelos de séries temporais para modelar a autocorrelação dos dados coletados neste processo. Em uma etapa seguinte da análise, os resíduos dos modelos de séries temporais foram analisados pelo método da análise em componentes principais para obter um modelo matemático que foi usado para construir gráficos de controle multivariado.

## ABSTRACT

### N-WAY ANALYSIS: METHODOLOGIES FOR THE MONITORING OF CHEMICAL REACTIONS

On this work, multivariate analyses were used to develop methodologies for monitoring chemical reactions. Two techniques were used: Multivariate Curve Resolution (MCR) for analysis of spectroscopic measures and the Multivariate Statistical Process Control (MSPC) for the analysis of the physical-chemical variables. The first work developed through this dissertation used the MCR methods to build empirical models of chemical reactions from reflectance measures of infrared spectra. This model was used to predict the behavior of new batches of the same reaction. The obtained results showed that the predictions are more exact when the model is adapted according to the (PACLS) Predicted Augmented Classical Least Squares to include a component in the model which model the baseline changes of the infrared spectrum. The second work presents a new methodology developed for determining the creatinine in human serum, based on the monitoring of Jaffé reaction with visible spectroscopy and analysis of the data with the tree-way MCR methods. According to the obtained results, creatinine can be determined in presence of others substances which interferes on its spectroscopic measure. The Standard Error of Performance (SEP) on this essay was  $2,22\text{mg L}^{-1}$  with a respective coefficient of variation around 6%. The third work presented on this dissertation describes a methodology developed for investigating the fluorescent and intracellular metabolite produced by *Chromobacterium Violaceum* during fermentation conducted in bioreactor under aerobics conditions. The fluorescence spectra collected from biomass samples were analyzed through the PARAFAC method (Parallel Factor Analysis). The tryptophan and tyrosine aminoacids beyond NADH (nicotinamide adenine dinucleotide) and violacein were identified through the analysis of the parameters from trilinear model obtained by analysis of fluorescence spectra. The estimations obtained for the concentration profiles of the chemical species, showed that this bacterium consumes the tryptophan before starting the progressive production of violacein. This result is according to the mechanism proposed on the literature for the biosynthesis of this pigment, which is a candidate to tumors and tropical disease treatment. The forth work on this dissertation presented a methodology of MSPC development for the analysis of autocorrelated data. This methodology was used to monitor the digestion of bauxite, used to extract de aluminum of this ore. Four Physical-chemical variables from this process were analyzed by time series models to model the autocorrelation among measures collected from this process. In a subsequent step of analyses, the residuals of time series models were analysed by Principal Component Analysis (PCA) to obtain a bilinear model that was used to construct multivariate control charts. The obtained results showed that the industrial plant operated out of statistical control.

## ÍNDICE

LISTA DE ABREVIATURAS .....	XXIV
LISTA DE TABELAS .....	XXVI
LISTA DE FIGURAS.....	XXVIII
<b>CAPÍTULO 1 .....</b>	<b>1</b>
<b>INTRODUÇÃO GERAL.....</b>	<b>1</b>
<b>1. INTRODUÇÃO.....</b>	<b>3</b>
<b>A. MATERIAL SUPLEMENTAR: PARTE A .....</b>	<b>7</b>
<b>A.1 Operações matemáticas com arranjos em N-modos.....</b>	<b>7</b>
<b>CAPÍTULO 2 .....</b>	<b>15</b>
<b>RESOLUÇÃO MULTIVARIADA DE CURVAS.....</b>	<b>15</b>
<b>2.1 INTRODUÇÃO.....</b>	<b>17</b>
<b>2.2 MODELOS MATEMÁTICOS MULTIVARIADOS.....</b>	<b>18</b>
<b>2.3 MÉTODOS MATEMÁTICOS EMPREGADOS PARA O AJUSTE DE</b>	
<b>MODELOS MULTIVARIADOS .....</b>	<b>21</b>
<b>2.3.1 Métodos não iterativos .....</b>	<b>21</b>
<i>2.3.1.1 Análise em Fatores Evolucionários (EFA).....</i>	<i>21</i>
<i>2.3.1.2 Método de Decomposição Trilinear Direta (DTLD) .....</i>	<i>26</i>
<b>2.3.2 Métodos de resolução iterativos .....</b>	<b>28</b>
<i>2.3.2.1 Método dos Mínimos Quadrados Alternantes em Dois Modos (ALS-MCR) ..</i>	<i>28</i>
<i>2.3.2.2 Análise de Decomposição em Fatores Paralelos (PARAFAC).....</i>	<i>30</i>
<b>2.4 AMBIGÜIDADE E UNICIDADE DA SOLUÇÃO DOS MÉTODOS DE MCR .....</b>	<b>31</b>
<b>2.5 RESTRIÇÕES DE MODELO .....</b>	<b>33</b>
<b>B. MATERIAL SUPLEMENTAR: PARTE B .....</b>	<b>37</b>
<b>B.1 Decomposição em Valores Singulares (SVD).....</b>	<b>37</b>
<b>B.2 O Problema dos Mínimos Quadrados .....</b>	<b>38</b>
<b>CAPÍTULO 3 .....</b>	<b>41</b>

<b>MONITORAMENTO ESPECTROSCÓPICO DE REAÇÕES QUÍMICAS COM MODELOS EMPÍRICOS ADAPTATIVOS.....</b>	<b>41</b>
<b>3.1 INTRODUÇÃO.....</b>	<b>43</b>
<b>3.2 RESOLUÇÃO MULTIVARIADA DE CURVAS PARA A ANÁLISE DE DADOS ESPECTROSCÓPICOS COLETADOS EM REAÇÕES QUÍMICAS.....</b>	<b>44</b>
<b>3.2.1 O problema da deficiência de posto.....</b>	<b>47</b>
<b>3.3 METODOLOGIA PARA O MONITORAMENTO DE REAÇÕES QUÍMICAS COM MODELOS EMPÍRICOS ADAPTATIVOS .....</b>	<b>48</b>
<b>3.4 PARTE EXPERIMENTAL .....</b>	<b>52</b>
<b>3.4.1 Sistemas dinâmicos em um tanque agitado contínuo.....</b>	<b>53</b>
<b>3.4.2 Experimentos com misturas em tanque agitado contínuo.....</b>	<b>56</b>
<b>3.5 RESULTADOS E DISCUSSÃO .....</b>	<b>58</b>
<b>3.6 CONCLUSÃO .....</b>	<b>69</b>
<b>CAPÍTULO 4 .....</b>	<b>71</b>
<b>DETERMINAÇÃO DE CREATININA EM SORO HUMANO USANDO A REAÇÃO DE JAFFÉ E CALIBRAÇÃO DE SEGUNDA ORDEM.....</b>	<b>71</b>
<b>4.1 INTRODUÇÃO.....</b>	<b>73</b>
<b>4.2 DETERMINAÇÃO DA CREATININA EM SORO HUMANO COM A REAÇÃO DE JAFFÉ.....</b>	<b>74</b>
<b>4.3 CALIBRAÇÃO DE SEGUNDA ORDEM .....</b>	<b>76</b>
<b>4.4 MÉTODO ESPECTRO-CINÉTICO MULTIVARIADO PARA A DETERMINAÇÃO DE CREATININA.....</b>	<b>79</b>
<b>4.5 PARTE EXPERIMENTAL .....</b>	<b>81</b>
<b>4.5.1 Equipamentos.....</b>	<b>81</b>
<b>4.5.2 Reagentes e Soluções .....</b>	<b>81</b>
<b>4.5.3 Experimentos de cinética. ....</b>	<b>82</b>
<b>4.5.4 Experimentos para análise da reação de creatinina com ácido pícrico em meio alcalino.....</b>	<b>82</b>
<b>4.5.5 Experimentos para determinação de creatinina em misturas.....</b>	<b>83</b>

4.5.6 Experimentos para determinação de creatinina em amostras de soro humano .....	84
4.5.7 Análise dos dados .....	85
4.5.7.1 Determinação da ordem da reação da creatinina com ácido pícrico.....	85
4.5.7.2 Decomposição trilinear.....	85
4.6 RESULTADOS E DISCUSSÃO.....	87
4.6.1 Análise da reação de creatinina com ácido pícrico em meio alcalino .....	87
4.6.2 Determinação da creatinina em misturas .....	91
4.6.3 Determinação da creatinina em soro humano.....	98
4.7 CONCLUSÃO.....	101
C. MATERIAL SUPLEMENTAR: PARTE C. ....	103
C.1 Diagnostico de Consistência do Núcleo (CORCONDIA).....	103
<b>CAPÍTULO 5 .....</b>	<b>105</b>
<b>ANÁLISE DOS METABÓLITOS FLUORESCENTES</b>	
<b>PRODUZIDOS PELA <i>CHROMOBACTERIUM VIOLACEUM</i></b>	
<b>COM O MÉTODO PARAFAC.....</b>	<b>105</b>
5.1 INTRODUÇÃO.....	107
5.2 BIOSÍNTESE DA VIOLACEÍNA.....	109
5.3 O PROCESSO DE LUMINESCÊNCIA .....	111
5.4 APLICAÇÕES DA FLUORESCÊNCIA EM BIOTECNOLOGIA .....	112
5.5 ANÁLISE DOS ESPECTROS DE FLUORESCÊNCIA COM OS MÉTODOS DE DECOMPOSIÇÃO TRILINEAR .....	114
5.6 PARTE EXPERIMENTAL.....	116
5.6.1 Experimentos conduzidos no bioreator para a produção de violaceína pela <i>Chromobacterium violaceum</i> .....	116
5.6.2 Coleta das amostras .....	117
5.6.3 Aquisição dos espectros eletrônicos de excitação e emissão.....	117
5.6.4 Tratamento matemático dos dados espectroscópicos .....	118
5.6.4.1 Análise das matrizes de excitação e emissão com o método PARAFAC ...	118
5.6.5 Organização dos dados.....	119

<b>5.7</b>	<b>RESULTADOS E DISCUSSÕES</b> .....	<b>119</b>
5.7.1	Análise das matrizes de excitação e emissão .....	119
5.7.2	Análise dos dados com o método PARAFAC .....	123
<b>5.8</b>	<b>CONCLUSÃO</b> .....	<b>136</b>
<b>D.</b>	<b>MATERIAL SUPLEMENTAR: PARTE D.</b> .....	<b>137</b>
D.1	Espectros de excitação e emissão de moléculas biológicas .....	137
<b>CAPÍTULO 6</b> .....		<b>139</b>
<b>CONTROLE ESTATÍSTICO MULTIVARIADO DE PROCESSOS</b> .....		<b>139</b>
<b>6.1</b>	<b>INTRODUÇÃO</b> .....	<b>141</b>
<b>6.2</b>	<b>PRINCIPIOS BÁSICOS DO CONTROLE ESTATÍSTICO DE PROCESSOS</b> .....	<b>143</b>
<b>6.3</b>	<b>GRÁFICOS DE CONTROLE UNIVARIADOS</b> .....	<b>146</b>
6.3.1	Gráficos de controle univariado para observações individuais.....	148
6.3.2	Gráficos de controle de médias móveis ponderadas exponencialmente. (“EWMA”).....	149
<b>6.4</b>	<b>GRÁFICOS DE CONTROLE MULTIVARIADO</b> .....	<b>151</b>
6.4.1	Gráficos de controle multivariados para observações individuais .....	155
6.4.2	Gráficos de controle multivariado das médias móveis ponderadas exponencialmente (“MEWMA”).....	157
6.4.3	Aplicação da Análise em Componentes Principais (PCA) para a construção dos gráficos de controle multivariado .....	158
6.4.3.1	<i>Limites de controle baseados em métodos não paramétricos</i> .....	161
<b>6.5</b>	<b>CONTROLE ESTATÍSTICO MULTIVARIADO DE PROCESSOS QUÍMICOS INDUSTRIAIS</b> .....	<b>162</b>
<b>E.</b>	<b>MATERIAL SUPLEMENTAR: PARTE E</b> .....	<b>169</b>
E.1	Distribuição de Probabilidade Normal.....	169
<b>CAPÍTULO 7</b> .....		<b>173</b>
<b>CONTROLE ESTATÍSTICO MULTIVARIADO DE PROCESSOS DE OBSERVAÇÕES AUTOCORRELACIONADAS: ANÁLISE DO PROCESSO BAYER PARA A EXTRAÇÃO DE ALUMÍNIO DA BAUXITA.</b> .....		<b>173</b>

<b>7.1</b>	<b>INTRODUÇÃO</b> .....	<b>175</b>
<b>7.2</b>	<b>SÉRIES TEMPORAIS: PROCESSOS AUTOCORRELACIONADOS</b> .....	<b>178</b>
	7.2.1 Identificação de modelos ARIMA ( $p, d, q$ ).....	181
<b>7.3</b>	<b>CONTROLE ESTATÍSTICO MULTIVARIADO DE PROCESSOS AUTOCORRELACIONADOS</b> .....	<b>183</b>
<b>7.4</b>	<b>MONITORAMENTO MULTIVARIADO DO PROCESSO BAYER PARA A EXTRAÇÃO DE ALUMÍNIO DA BAUXITA</b> .....	<b>187</b>
<b>7.5</b>	<b>MATERIAIS E MÉTODOS</b> .....	<b>189</b>
	7.5.1 Amostragem e conjunto de dados .....	189
	7.5.2 Software e Programas .....	189
<b>7.6</b>	<b>RESULTADOS E DISCUSSÃO</b> .....	<b>189</b>
	7.6.1 Análise do correlograma das variáveis de processo.....	190
	7.6.2 Ajuste de modelos de séries temporais das variáveis de processo .....	193
	7.6.3 Análise de correlação entre as variáveis de processo.....	197
	7.6.4 Controle estatístico multivariado de processos .....	199
	7.6.4.1 Fase 1 .....	199
	7.6.5 Nova verificação do estado do processo.....	203
	7.6.5.1 Fase 1 .....	208
	7.6.5.2 Fase 2.....	211
<b>7.7</b>	<b>CONCLUSÃO</b> .....	<b>216</b>
<b>CAPÍTULO 8</b> .....		<b>219</b>
	<b>CONCLUSÃO GERAL</b> .....	<b>219</b>
<b>8.1</b>	<b>CONSIDERAÇÕES FINAIS</b> .....	<b>221</b>
<b>9.</b>	<b>REFÊRENCIAS BIBLIOGRÁFICAS</b> .....	<b>1</b>

## LISTA DE ABREVIATURAS

- PLS - Método de regressão em mínimos quadrados parciais
- PCA - Análise em componentes principais
- HPLC-DAD - Cromatografia Líquida de Alta Eficiência com arranjo de fotodiodos
- MCR - Resolução Multivariada de Curvas
- CEMP - Controle estatístico Multivariado de Processos
- DTLD - Decomposição Trilinear Direta
- EFA - Análise em Fatores Evolucionários
- WFA - “ Windows Factor Analysis”
- ALS - Algoritmo dos Mínimos Quadrados Alternates
- MCR-ALS - Método de resolução multivariada de curvas com o algoritmo dos mínimos quadrados alternates
- SVD - Decomposição em Valores Singulares
- PARAFAC - Análise de Decomposição em Fatores Paralelos
- ATLD - “ Alternating Trilinear Decomposition”
- SWTLD - “ Self-Weighted Alternating Trilinear Decomposition”
- PMF3 - “ Three Way Matrix Positive Factorization”
- PACLS - “Predicted Augmented Classical Least Squares”
- CORCONDIA - “Core consistency diagnostic”
- SEP - “Standard Error of Performance”
- CV - Coeficiente de variação
- LS - Método dos Mínimos Quadrados
- posto(**X**) - Posto da matriz **X**.
- C. violaceum* - *Chromobacterium violaceum*
- PBS - Solução salina de polifosfatos tamponada
- ATP - Trifosfato de adenosina
- NADH - Nicotinamida adenina dinucleotídio
- EWMA - Médias móveis ponderadas exponencialmente
- MEWMA - Médias móveis multivariadas ponderadas exponencialmente

CUSUM - somas acumuladas

RMSEP - Raiz quadrada do erro quadrático médio

ARIMA - Modelo integrado autoregressivo de média móveis

“fac” - Função de autocorrelação amostral

“facp” - Função de autocorrelação amostral parcial

## LISTA DE TABELAS

<b>Tabela 3.1.</b> Valores das constantes usadas para executar os experimentos da figura 3.4.....	57
<b>Tabela 4.1.</b> Planejamento fatorial $2^3$ . Cada experimento corresponde a uma reação com $10 \text{ mg L}^{-1}$ de creatinina.....	83
<b>Tabela 4.2.</b> Conjunto de misturas preparadas para testar a metodologia desenvolvida.....	84
<b>Tabela 4.3.</b> Planejamento Fatorial $2^3$ . Cada experimento corresponde a uma reação com $10 \text{ mg L}^{-1}$ de creatinina.....	89
<b>Tabela 4.4.</b> Erro de previsão “RMSEP” para a análise dos experimentos da tabela 4.2.....	96
<b>Tabela 4.5.</b> Previsão da concentração de creatinina em amostras de soro. ....	98
<b>Tabela 5.1.</b> Comprimentos de onda característicos de emissão e excitação de algumas substâncias autofluorescentes. ....	113
<b>Tabela 5.2.</b> Porcentagem de Variação explicada e Diagnóstico de Consistência do Núcleo (CORCONDIA) calculados para os modelos trilineares de 1 a 5 componentes com o método PARAFAC para o arranjo $\underline{X}_1^*$ .....	124
<b>Tabela 5.3.</b> Porcentagem de Variação explicada e Diagnóstico de Consistência do Núcleo (CORCONDIA) calculados para os modelos trilineares de 1 a 5 componentes com o método PARAFAC para o arranjo $\underline{X}_2^*$ .....	126
<b>Tabela 5.4.</b> Porcentagem de Variação explicada e Diagnóstico de Consistência do Núcleo (CORCONDIA) calculados para os modelos trilineares de 1 a 5 componentes com o método PARAFAC para o arranjo $\underline{X}_3^*$ .....	128
<b>Tabela 5.5.</b> Porcentagem de Variação explicada e Diagnóstico de Consistência do Núcleo (CORCONDIA) calculados para os modelos trilineares de 1 a 5 componentes com o método PARAFAC para o arranjo $\underline{X}_4^*$ .....	131
<b>Tabela 5.6:</b> Substâncias fluorescentes.....	138
<b>Tabela 7.1.</b> Comportamento esperado para os coeficientes das funções “fac” e “facp” de modelos $ARIMA(p,d,q)$ . ....	182

<b>Tabela 7.2.</b> Parâmetros dos modelos AR(1) ajustado pelo método dos mínimos quadrados. ....	194
<b>Tabela 7.3.</b> Parâmetros dos modelos AR(1) ajustado pelo método da máxima probabilidade.....	194
<b>Tabela 7.4.</b> Valores P calculados para os testes de Ljung-Box e Shapiro-Wilk na análise dos resíduos do modelos AR(1) ajustados para cada variável de processo.....	195
<b>Tabela 7.5.</b> Valores P calculados para os testes de Ljung-Box e Shapiro-Wilk na análise dos resíduos do modelos AR(1) ajustados para cada variável de processo.....	195
<b>Tabela 7.6.</b> Coeficientes de correlação linear entre variáveis de controle e processo.....	198
<b>Tabela 7.7.</b> Coeficientes de correlação entre os resíduos dos modelos AR(1).....	198
<b>Tabela 7.8.</b> Parâmetros dos modelos AR(1) calculados para cada propriedade. ....	205
<b>Tabela 7.9.</b> Valores P calculados para os testes de Ljung-Box e Shapiro-Wilk na análise dos resíduos do modelos AR(1) ajustados para cada variável de processo.....	207
<b>Tabela 7.10.</b> Valores P calculados para os testes de Ljung-Box e Shapiro-Wilk na análise da primeira diferença dos dados originais. ....	207

## LISTA DE FIGURAS

<b>Figura 1.1.</b> Estrutura dos dados e modos de monitoramento. A) um modo; B) dois modos; C) três modos.....	6
<b>Figura 1.2.</b> Tipos de desdobramentos do arranjo em três modos: $\mathbf{X}_{iJK}$ , $\mathbf{X}_{jKI}$ e $\mathbf{X}_{kIJ}$ .....	14
<b>Figura 2.1.</b> A) Espectros de absorvância registrados em função do tempo para uma separação cromatográfica. B) Cromatogramas dos três compostos químicos. C) Espectros dos compostos químicos.....	18
<b>Figura 2.2.</b> Representação esquemática de um modelo bilinear. ....	19
<b>Figura 2.3.</b> Representação esquemática de um modelo trilinear. ....	20
<b>Figura 2.4.</b> Esquema da formação das submatrizes $\mathbf{X}_{i \times j}$ . ....	22
<b>Figura 2.5.</b> A) Mapa dos valores singulares para análise no sentido direto. B) Mapa dos valores singulares para análise no sentido inverso. C) e D) Cromatogramas de cada composto químico da separação cromatográfica apresentada na figura 2.1A. ....	24
<b>Figura 2.6.</b> A) Mapa dos valores singulares da análise dos dados apresentados na figura 2.1A. B) Cromatogramas dos compostos químicos da figura 2.1A. (-) Primeiro componente, (-) segundo componente e (-) terceiro componente. ....	25
<b>Figura 3.1.</b> Método para estender a matriz $\mathbf{B}_{ref}$ . A matriz estendida $\mathbf{B}_{est}$ representada acima inclui uma componente que modela uma mudança constante na linha de base.....	52
<b>Figura 3.2.</b> Tanque agitado contínuo.....	53
<b>Figura 3.3.</b> A) Vazão de adição dos compostos em função do tempo para as válvulas $f_R$ , $f_I$ e $f_P$ . B) variação da concentração em função do tempo para os compostos R, I e P, de acordo com as equações 3.10 a 3.12 usando constantes $k_1 = 0,05 \text{ s}^{-1}$ e $k_2 = 0,045 \text{ s}^{-1}$ . ....	55
<b>Figura 3.4.</b> Experimentos com misturas produzidos no tanque agitado contínuo. ....	56
<b>Figura 3.5.</b> Variação da concentração com o tempo dos compostos químicos para as misturas apresentadas na figura 3.4. A) experimento I; B) experimento II; C) experimento III; D) experimento IV. ....	58

<b>Figura 3.6.</b> Espectros de infravermelho registrados para os experimentos: A) I; B) II; C) III; D) IV. ....	59
<b>Figura 3.7.</b> Resultado da análise com o método EFA para a matriz $\mathbf{X}_{ref}^T$ . “Janelas” para a (-) primeira, (-) segunda e (-) terceira componentes. ....	60
<b>Figura 3.8.</b> Autovetores $\mathbf{U}$ obtidos por decomposição em valores singulares da matriz $\mathbf{X}_{ref}$ . Os autovetores apresentados estão multiplicados pelos seus respectivos valores singulares. ....	60
<b>Figura 3.9.</b> Comparação dos sinais resolvidos com o método EFA (+) com os sinais verdadeiros de cada composto químico (-). A) Variação da concentração em função do tempo; B) espectro do composto P; C) espectro do composto R ; D) espectro do composto I.....	62
<b>Figura 3.10.</b> Parâmetros do modelo MCR-ALS resolvidos para a matriz $\mathbf{X}_{ref}$ . A) matriz $\mathbf{A}_{ref}$ , B) matriz $\mathbf{B}_{ref}$ . ....	63
<b>Figura 3.11.</b> Comparação dos sinais resolvidos com o método MCR (+) com os sinais verdadeiros de cada composto químico (-). A) Variação da concentração em função do tempo; B) espectro do composto P; C) espectro do composto R; D) espectro do composto I. ....	64
<b>Figura 3.12.</b> Variação das concentrações em função do tempo previstas pelo modelo bilinear. A) II; B) III; C) IV; (-) Curvas de concentrações verdadeiras, (-) Previsões feitas sem correção de linha base, (-) previsões feitas com correções de linha base. ....	66
<b>Figura 3.13.</b> Variação das concentrações em função do tempo previstas pelo modelo bilinear. A) II; B) III; C) IV; (-) Curvas de concentrações verdadeiras, (-) Previsões feitas sem correção de linha base, (-) previsões feitas com correções de linha base, (-) Previsões feitas com os espectros verdadeiros pelo método dos mínimos quadrados e com correção de linha base. ....	67
<b>Figura 3.14.</b> Resultados obtidos para previsão das novas bateladas (+) com a matriz $\mathbf{B}_{est}$ , comparados com modelo de referência (-). A) batelada II; B) batelada III; C) batelada IV.....	68

<b>Figura 4.1.</b> Arranjo em três modos formado para o conjunto de amostras analisadas com o monitoramento espectro-cinético multivariado .....	80
<b>Figura 4.2.</b> Matriz <b>C</b> do modelo trilinear $\mathbf{A}(\mathbf{B}   \otimes   \mathbf{C})^T$ .....	80
<b>Figura 4.3.</b> Ordem da reação de creatinina com ácido pícrico determinada pelo método das velocidades iniciais.....	88
<b>Figura 4.4.</b> Gráfico $\Delta A \times C_{ini}$ para o comprimento de onda 500 nm. $\Delta A = (A_{20,500} - A_{120,500})$ .....	90
<b>Figura 4.5.</b> Monitoramento multivariado em dois modos para uma das reações do grupo 6 tabela 4.2. ....	91
<b>Figura 4.6.</b> Gráficos do “CORCONDIA” calculados a partir dos parâmetros de modelos PARAFAC de duas (A) e três componentes (B) obtidos para a análise de um experimento do grupo 6, Tabela 4.2. ....	93
<b>Figura 4.7.</b> Parâmetros do modelo trilinear com duas componentes (“o” primeira componente, “+” segunda componente). A) modo espectroscópico; B) modo cinético.....	94
<b>Figura 4.8.</b> Comparação dos sinais verdadeiros do produto da reação da creatinina com ácido pícrico (-), com a primeira componente resolvida com o método PARAFAC (o). A) modo espectroscópico; B) modo cinético.....	95
<b>Figura 4.9.</b> Espectros das substâncias Bilirrubina (o) e Albumina (+).....	95
<b>Figura 4.10.</b> Parâmetros recuperados pelo modelo DTLTD com duas componentes(o), e os sinais “puros” para o produto da reação da creatinina (-). ....	100
<b>Figura 4.11.</b> Parâmetros recuperados pelo modelo DTLTD com três componentes(o), e os sinais esperados para o produto da reação da creatinina (-). ....	100
<b>Figura 5.1.</b> Fórmulas estruturais dos pigmentos violetas produzidos pela <i>C. violaceum</i> : A) violaceína; B) deoxiviolaceína.....	107
<b>Figura 5.2.</b> Esquema da biossíntese da violaceína.....	110
<b>Figura 5.3.</b> Unidades estruturais da molécula da violaceína.....	111
<b>Figura 5.4.</b> Arranjo em três modos $\underline{\mathbf{X}}$ formado por um conjunto de matriz de excitação e emissão. ....	115
<b>Figura 5.5.</b> Matrizes de excitação e emissão de quatro amostras coletadas entre 8 e 14 horas de fermentação aeróbica. ....	120

<b>Figura 5.6.</b> Matrizes de excitação e emissão de quatro amostras coletadas entre 16 e 22 horas de fermentação aeróbica.....	121
<b>Figura 5.7.</b> Matriz de emissão e excitação com valores ausentes adicionados no lugar dos picos de espalhamento Rayleigh.....	123
<b>Figura 5.8.</b> Estimativas dos espectros de excitação e emissão obtidas com o modelo trilinear de três componentes ajustado com o método PARAFAC. ....	125
<b>Figura 5.9.</b> Estimativas dos espectros de excitação e emissão obtidas com o modelo trilinear de três componentes ajustado com o método PARAFAC para o arranjo $\underline{X}_2$ . ....	127
<b>Figura 5.10.</b> Estimativas dos espectros de excitação e emissão obtidas com o modelo trilinear de três componentes ajustado com o método PARAFAC para o arranjo $\underline{X}_3$ . ....	129
<b>Figura 5.11.</b> Estimativas da variação da concentração em função do tempo obtidas do modelo trilinear ajustado a partir do arranjo $\underline{X}_3$ . (o) Primeira componente; (>) segunda componente; (+) terceira componente.....	130
<b>Figura 5.12.</b> Estimativas dos espectros de excitação e emissão obtidas com o modelo trilinear de três componentes ajustado com o método PARAFAC para o arranjo $\underline{X}_4$ . ....	132
<b>Figura 5.13.</b> Estimativas dos espectros de excitação e emissão obtidas com o modelo trilinear de duas componentes ajustado com o método PARAFAC a partir do arranjo $\underline{X}_4$ . ....	133
<b>Figura 5.14.</b> Estimativas da variação da concentração em função do tempo obtidas do modelo trilinear ajustado a partir do arranjo $\underline{X}_4$ . (o) Primeira componente; (>) segunda componente.....	134
<b>Figura 5.15.</b> Espectros de excitação e emissão de moléculas presentes no metabolismo de seres vivos. ....	137
<b>Figura 6.1.</b> Gráfico de controle do tipo Shewhart para monitorar a média de uma propriedade. ....	142
<b>Figura 6.2.</b> Tipos de variáveis presentes nos processos industriais.....	144
<b>Figura 6.3.</b> Exemplo de uma carta de controle do tipo EWMA. ....	151
<b>Figura 6.4.</b> Gráfico de controle $\chi^2_o$ . ....	153

<b>Figura 6.5.</b> Método “Multi-Way PCA” .....	166
<b>Figura 6.6.</b> Esquema do método de monitoramento “Model-Based” PCA.....	168
<b>Figura 6.7.</b> Curva de Gauss ou distribuição normal. ....	170
<b>Figura 6.8.</b> Intervalos de probabilidade sob a área da distribuição normal. ....	171
<b>Figura 7.1.</b> A) Concentração de soda cáustica determinada em amostras do concentrado de aluminato de sódio retirado do tanque de digestão do processo Bayer. B) Correlograma das medidas de soda cáustica.....	176
<b>Figura 7.2.</b> Esquema da metodologia para o monitoramento estatístico de processos com observações autocorrelacionadas.....	184
<b>Figura 7.3.</b> Etapa de Digestão do Processo Bayer. ....	188
<b>Figura 7.4.</b> Variáveis do processo monitoradas ao longo de duzentas horas. ....	190
<b>Figura 7.5.</b> Função de autocorrelação amostral para cada variável do processo. ....	192
<b>Figura 7.6.</b> Função de autocorrelação amostral parcial de cada variável de processo. ....	192
<b>Figura 7.7.</b> Função de autocorrelação amostral da primeira diferença tomada para cada variável de processo. ....	193
<b>Figura 7.8.</b> Correlograma dos resíduos dos modelos AR(1) de cada variável do processo. ....	196
<b>Figura 7.9.</b> Distribuição dos resíduos dos modelos AR(1) para cada variável do processo. ....	197
<b>Figura 7.10.</b> Gráficos de controle <i>D</i> e <i>Q</i> do modelo PCA.....	200
<b>Figura 7.11.</b> Gráficos de controle MEWMA- <i>D</i> do modelo PCA. ....	201
<b>Figura 7.12.</b> Gráfico das contribuições das propriedades para a observação feita em 110 e 112 horas. A barra mais a esquerda é o limite inferior e a mais a direita é o limite superior. ....	202
<b>Figura 7.13.</b> Funções de autocorrelação amostral para cada propriedade entre 130 e 240 horas. ....	204
<b>Figura 7.14.</b> Funções de autocorrelação amostral para a primeira diferença dos dados de 130 e 240 horas. ....	204
<b>Figura 7.15.</b> Funções de autocorrelação amostral parcial de cada propriedade entre 130 e 240 horas. ....	205

<b>Figura 7.16.</b> Funções de autocorrelação amostral dos resíduos dos modelos apresentados na tabela 7.8.....	206
<b>Figura 7.17.</b> Distribuição dos resíduos dos modelos AR(1) apresentados na tabela 7.8. ....	206
<b>Figura 7.18.</b> Gráficos de controle <i>D</i> e <i>Q</i> . Fase1. ....	208
<b>Figura 7.19.</b> Gráficos de controle MEWMA- <i>D</i> . Fase1.....	209
<b>Figura 7.20</b> Gráficos de controle <i>D</i> e <i>Q</i> após a excluir os ponto anômalos. Fase1. ....	210
<b>Figura 7.21.</b> Gráficos de controle MEWMA- <i>D</i> após a excluir os pontos anômalos. Fase1. ....	210
<b>Figura 7.22.</b> Gráficos de controle <i>D</i> e <i>Q</i> . Fase 2. ....	211
<b>Figura 7.23.</b> Gráficos de controle MEWMA- <i>D</i> para Fase2 dos resíduos dos modelos AR(1).....	212
<b>Figura 7.24.</b> Gráficos de controle MEWMA- <i>D</i> para a Fase2 tomados da primeira diferença dos dados originais (ARIMA(0,1,0)).....	213
<b>Figura 7.25.</b> Gráfico das contribuições das propriedades para a observação feita 382 horas. A barra mais a esquerda é o limite inferior e a mais a direita é o limite superior.....	214
<b>Figura 7.26.</b> Medidas de índice de refração coletadas entre 130 e 578 horas. ....	215
<b>Figura 7.27.</b> A) Correlograma calculada para as observações entre 130 e 380 horas. B) Correlograma calculada para as observações entre 130 e 400 horas.....	215

# **CAPÍTULO 1**

## **INTRODUÇÃO GERAL**

## 1. INTRODUÇÃO

A presença maciça de microcomputadores nos laboratórios tornou relativamente simples o trabalho de fazer operações matemáticas que envolvem o processamento de uma grande quantidade de dados. Uma questão importante dentro deste contexto, é o que fazer com os dados e de que modo estes devem ser analisados? Este problema pode parecer simples se dispusermos de um conjunto de duas ou três variáveis que são independentes entre si, mas frente a um banco de dados contendo muitas variáveis, a análise pode ser um problema extremamente complexo.

A presença de variáveis correlacionadas é uma característica marcante nos sistemas químicos, visto que as propriedades destes sistemas normalmente são influenciadas por fatores externos ou internos correlacionados entre si. Um exemplo clássico de correlação entre as variáveis é a relação entre pressão, temperatura, número de mols e o volume de um gás ideal. Todas estas variáveis são correlacionadas entre si através de uma equação matemática, definida como:

$$PV = nRT$$

1.1

onde  $R$  é a constante universal dos gases perfeitos. De acordo com a equação dos gases ideais a pressão ( $P$ ) de um gás é inversamente proporcional ao volume ( $V$ ) ocupado por este, mas diretamente proporcional à sua temperatura ( $T$ ) e ao número de mols ( $n$ ) do mesmo.

Nas reações químicas, a concentração de uma substância que participa da reação geralmente depende do valor da concentração de duas ou mais espécies que também participam da reação. O mecanismo das reações químicas pode ser formulado em termos matemáticos através de um sistema de equações diferenciais como o mostrado na equação 1.2, para uma reação que tem as seguintes etapas:



$$\frac{dC_A}{dt} = -k_1 C_A$$

$$\frac{dC_B}{dt} = k_1 C_A - k_2 C_B$$

$$\frac{dC_C}{dt} = k_2 C_B$$

1.2

De acordo com o sistema de equações diferenciais, apresentado acima, a concentração do produto da reação ( $C_C$ ) em um dado instante depende da concentração do intermediário ( $C_B$ ) da reação, que por sua vez, depende da concentração do reagente ( $C_A$ ). As constantes das equações de velocidade,  $k_1$  e  $k_2$ , dependem de variáveis como a temperatura da reação, pressão e condições de ativação dos catalisadores. Para todas estas variáveis é possível escrever uma relação matemática que expressa a dependência das constantes  $k$  em relação a estas variáveis que por sua vez estão relacionadas também com a composição química da reação em um dado instante.

Os dois exemplos citados mostram que as variáveis dos sistemas químicos são intimamente correlacionadas entre si e infelizmente, o modelo mecanístico que descreve o comportamento das variáveis nem sempre é conhecido, da mesma forma que não é conhecido como elas são correlacionadas entre si. Sendo assim, uma das funções dos químicos é a de monitorar um conjunto de variáveis para encontrar através da análise dos dados, um modelo capaz de fazer uma descrição das propriedades do sistema sob investigação.

A descoberta de um modelo mecanístico para a modelagem de sistemas químicos geralmente é um processo muito complicado e, por razões práticas, muitas vezes é mais interessante ajustar um modelo empírico a partir do conjunto de dados disponíveis que seja capaz de informar as relações existentes entre as variáveis e as propriedades do sistema. A desvantagem do modelo empírico é que este só deve ser usado para fazer previsões das propriedades do sistema que estejam nas mesmas condições experimentais em que o modelo foi derivado.

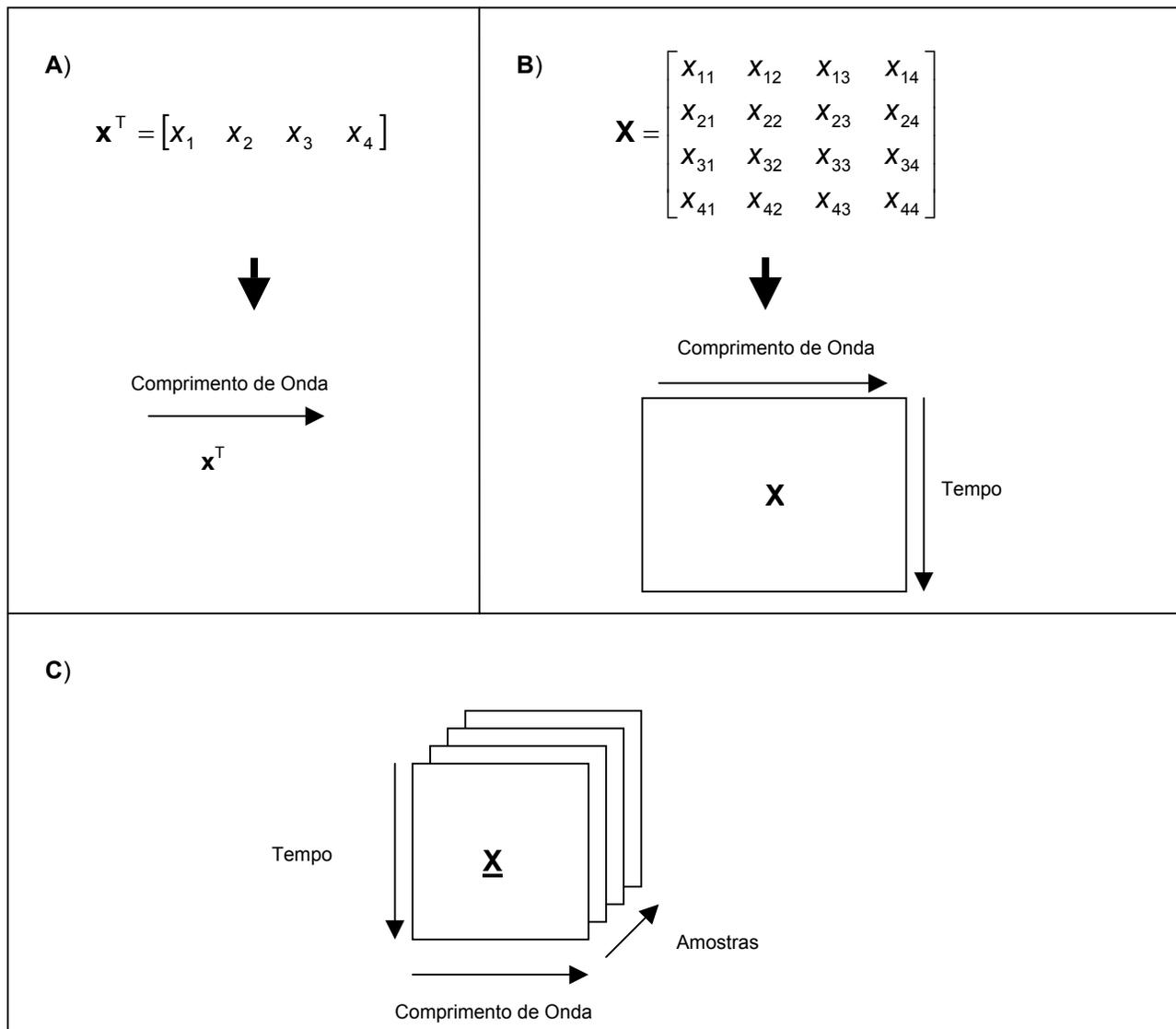
A estatística é uma das ferramentas básicas para a modelagem empírica, porém, os métodos tradicionais desta disciplina, como o método de regressão linear múltipla,<sup>1</sup> considera que as variáveis são independentes entre si e, por isto, muitas vezes estes

métodos não podem ser usados para modelar sistemas químicos. Em contra partida, a quimiometria, definida como a disciplina da química que utiliza métodos matemáticos e estatísticos para a análise dos dados coletados nos sistemas químicos,<sup>2</sup> surgiu da necessidade dos químicos de dispor de métodos de análise multivariada que sejam eficientes para a análise de tais dados.

Os métodos mais importantes da quimiometria correspondem a uma união dos métodos da álgebra linear com os métodos estatísticos, formando o que hoje é conhecido como métodos de análise multivariada. Alguns destes métodos tornaram-se relativamente populares como os métodos de Regressão em Mínimos Quadrados Parciais (PLS)<sup>3</sup> e a Análise em Componentes Principais (PCA).<sup>4</sup> Estes dois métodos são utilizados para as mais variadas tarefas, como para análise classificatória,<sup>5</sup> reconhecimento de padrões,<sup>6</sup> controle estatístico multivariado de processos,<sup>7</sup> análise exploratória,<sup>8</sup> resolução de sinais sobrepostos,<sup>9</sup> calibração de ensaios analíticos<sup>10</sup> e para análise da relação entre a estrutura das moléculas e a atividade biológica.<sup>11</sup>

Atualmente, um tema central para o desenvolvimento dos métodos da quimiometria é a estrutura dos dados multivariados. Uma amostra retirada de um sistema químico gera dados que podem ser organizados em estruturas como vetores, matrizes e cubos como mostra a figura 1.1 ou em estruturas com mais que três dimensões. Por exemplo, um espectro de absorção molecular registrado para uma amostra, pode ser organizado em um vetor conforme mostra a figura 1.1A, porque cada elemento do vetor corresponde a uma medida de absorbância em um comprimento de onda específico. Um equipamento hifenado, tal como um cromatógrafo acoplado a um espectrofotômetro com arranjo de fotodiodos (HPLC-DAD),<sup>12</sup> registra espectros de absorção molecular em intervalos regulares de tempo. Os dados produzidos por este tipo de equipamento podem ser organizados em uma matriz como mostra a figura 1.1B. Para o caso em que mais do que de uma amostra é analisada com um equipamento HPLC-DAD, os dados podem ser organizados em um cubo, tal como mostra a figura 1.1C.

A quimiometria classifica as diferentes estruturas de dados como arranjos em N-modos. Um vetor é um arranjo em um modo, uma matriz é um arranjo em dois modos e um cubo é um arranjo em três modos.



**Figura 1.1.** Estrutura dos dados e modos de monitoramento. A) um modo; B) dois modos; C) três modos.

Neste trabalho de tese, três técnicas de análise multivariada em N-modos foram usadas para o monitoramento de reações químicas em sistemas puramente químicos, biológicos e industriais. Os quatro primeiros capítulos estão relacionados com a aplicação das técnicas de Resolução Multivariada de Curvas (MCR)<sup>12</sup> e Calibração de Ordem Superior,<sup>13</sup> enquanto que os dois últimos capítulos tratam exclusivamente do Controle Estatístico Multivariado de Processos (CEMP).<sup>7</sup> Ao todo, são apresentadas quatro aplicações independentes, desenvolvidas nesta tese e que utilizam uma destas técnicas. Uma discussão rápida de cada uma delas é apresentada no resumo desta dissertação.

## A. MATERIAL SUPLEMENTAR: PARTE A

O texto abaixo foi elaborado com o objetivo específico de fixar a notação matemática e as operações básicas com matrizes e vetores utilizadas nesta tese.

### A.1 Operações matemáticas com arranjos em N-modos

Considere o espectro de uma amostra de um único constituinte de concentração  $c$ . As medidas de intensidade da radiação, feitas em cada um dos  $J$  comprimentos de onda da faixa espectral considerada podem ser organizadas em um vetor coluna  $\mathbf{x}$ , conforme mostra o esquema abaixo:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_J \end{bmatrix} \quad 1.3$$

A operação  $\mathbf{x}^T$  (lê-se transposta de  $\mathbf{x}$ ) organiza os elementos de  $\mathbf{x}$  em um vetor linha conforme mostra o esquema abaixo:

$$\mathbf{x}^T = [x_1 \quad x_2 \quad \cdots \quad x_J] \quad 1.4$$

Por convenção, um vetor escrito na forma  $\mathbf{x}$  é sempre um vetor coluna, enquanto que  $\mathbf{x}^T$  é um vetor linha.

Assumindo uma relação linear entre a concentração  $c$  e as medidas de intensidade da radiação, podemos decompor o vetor  $\mathbf{x}$  no produto de um escalar por um vetor conforme mostra o esquema abaixo:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_J \end{bmatrix} = \mathbf{c} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_J \end{bmatrix} = \mathbf{cb} \quad \mathbf{1.5}$$

Se  $c$ , é a concentração da substância em mol L<sup>-1</sup>, então os elementos do vetor  $\mathbf{b}$  correspondem aos coeficientes de absorvidade molar da substância para cada  $J$  comprimento de onda. Um exemplo bem conhecido é o coeficiente de absorvidade molar da equação da lei de Beer.

A equação 1.5, descreve um modelo linear para um sistema químico formado por uma amostra e um constituinte, mas no caso de haver mais do que uma amostra, o modelo passa a ser escrito na seguinte forma:

$$\mathbf{X} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_K \end{bmatrix} \times [b_1 \quad b_2 \quad \dots \quad b_J] = \mathbf{cb}^T \quad \mathbf{1.6}$$

Note que na equação 1.6, as concentrações da substância química em cada amostra são agrupadas em um vetor  $\mathbf{c}$ , sendo o número de amostras igual a  $K$ . O elemento  $c_1$  de  $\mathbf{c}$  é a concentração da substância na amostra 1.

O produto de dois vetores é possível quando o número de linhas do vetor à direita é igual ao número de colunas do vetor à esquerda. Na equação 1.6,  $\mathbf{c}$  e  $\mathbf{b}^T$  tem respectivamente as seguintes dimensões  $K \times 1$  e  $1 \times J$ . O resultado da equação 1.6 é uma matriz  $\mathbf{X}$  com dimensões  $K \times J$ , onde cada linha desta matriz corresponde ao espectro de uma amostra e cada coluna reúne as medidas de intensidade da radiação feitas para todas as amostras no comprimento de onda  $j$ .

No caso de uma amostra constituída por duas substâncias, seu espectro pode ser decomposto no seguinte modelo que envolve a adição de dois vetores:

$$\mathbf{x} = c_{11} \begin{bmatrix} b_{11} \\ b_{21} \\ \vdots \\ b_{j1} \end{bmatrix} + c_{12} \begin{bmatrix} b_{12} \\ b_{22} \\ \vdots \\ b_{j2} \end{bmatrix} \quad 1.7$$

No exemplo acima são usados dois índices para representar os elementos do vetor  $\mathbf{b}$  e do escalar  $c$ . Os primeiros índices dos elementos  $b$  e  $c$  correspondem respectivamente ao comprimento de onda  $j$  e a amostra  $k$ . O segundo índice desses elementos, identifica a componente  $r$  do modelo. Para sistemas multicomponentes como o apresentado na equação 1.7, define-se o número de componentes do modelo,  $R$ , como o número de substâncias que contribuem para a resposta  $\mathbf{x}$ . A equação 1.7 é formada por dois componentes,  $R = 2$ .

Em situações não ideais o número de componentes pode ser maior que o número de espécies químicas devido a variações da linha de base dos espectros no caso de medidas espectroscópicas ou devido a desvios de linearidade.

A equação 1.7 assume que o espectro da amostra é o resultado da soma dos coeficientes de extinção do espectro de cada substância ponderados pelo valor de concentração de cada uma e que a variação da medida de intensidade da radiação é linear e proporcional às concentrações de cada substância. A equação 1.7 pode ser representada também pela equação 1.8.

$$\mathbf{x} = c_{11} \mathbf{b}_1 + c_{12} \mathbf{b}_2 \quad 1.8$$

Note que o índice do vetor  $\mathbf{b}$  e o segundo do escalar  $c$  na equação 1.8, faz referência aos espectros e as concentrações das substâncias 1 e 2 presentes na amostra.

Considere agora um conjunto de  $K$  amostras formadas por duas substâncias 1 e 2. Neste caso o modelo pode ser escrito das seguintes formas (ambas equivalentes):

$$\mathbf{X} = \begin{bmatrix} c_{11} \\ c_{21} \\ \vdots \\ c_{K1} \end{bmatrix} \times [b_{11} \quad b_{21} \quad \dots \quad b_{J1}] + \begin{bmatrix} c_{12} \\ c_{22} \\ \vdots \\ c_{K2} \end{bmatrix} \times [b_{12} \quad b_{22} \quad \dots \quad b_{J2}] =$$

$$= \mathbf{c}_1 \mathbf{b}_1^T + \mathbf{c}_2 \mathbf{b}_2^T \quad \mathbf{1.9}$$

ou

$$\mathbf{X} = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \\ \vdots & \vdots \\ c_{K1} & c_{K2} \end{bmatrix} \times \begin{bmatrix} b_{11} & b_{21} & \dots & b_{J1} \\ b_{12} & b_{22} & \dots & b_{J2} \end{bmatrix} = \mathbf{CB}^T \quad \mathbf{1.10}$$

Na equação 1.9, os elementos dos vetores  $\mathbf{c}_1$  e  $\mathbf{c}_2$  representam as concentrações das respectivas substâncias 1 e 2 em cada amostra. O modelo descrito pela equação 1.9 também pode ser representado pelo produto de duas matrizes como mostra a equação 1.10.

A equação 1.10 é um exemplo de multiplicação entre duas matrizes, a matriz  $\mathbf{C}$  de dimensão  $K \times 2$  é multiplicada por uma matriz  $\mathbf{B}^T$  de dimensão  $2 \times J$ . Note que o número de colunas de  $\mathbf{C}$  é igual ao número de linhas de  $\mathbf{B}^T$ .

Considere agora o caso em que as medidas feitas para uma amostra envolvam o monitoramento em dois modos, tempos e comprimentos de onda, como no caso das medidas feitas com um instrumento HPLC-DAD,<sup>12</sup> figura 1.1B. Se a amostra tem apenas uma substância com concentração  $c$ , podemos escrever um modelo matemático que descreve a relação linear entre a concentração da substância e seus respectivos sinais nos modos de monitoramento tempo x comprimento de onda, conforme a equação abaixo:

$$\mathbf{X} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_l \end{bmatrix} \times c \times [b_1 \quad b_2 \quad \dots \quad b_J] = \mathbf{acb}^T \quad \mathbf{1.11}$$

Na equação 1.11, o vetor  $\mathbf{a}$  de dimensão  $I \times 1$  corresponde ao perfil de eluição (cromatograma) da substância química presente na amostra. O sub índice  $i$  refere-se ao instante (tempo) em que um espectro foi registrado durante a separação cromatográfica. Logo o índice  $I$ , corresponde ao número total de instantes em que os espectros foram registrados ao longo do tempo.

Para uma amostra contendo duas substâncias, 1 e 2, cada uma com seus respectivos cromatogramas  $\mathbf{a}_1$  e  $\mathbf{a}_2$  e espectros  $\mathbf{b}_1$  e  $\mathbf{b}_2$ , podemos escrever o modelo da seguinte forma:

$$\mathbf{X} = \begin{bmatrix} \mathbf{a}_{11} & \mathbf{a}_{12} \\ \mathbf{a}_{21} & \mathbf{a}_{22} \\ \vdots & \vdots \\ \mathbf{a}_{I1} & \mathbf{a}_{I2} \end{bmatrix} \times \begin{bmatrix} c_1 & 0 \\ 0 & c_2 \end{bmatrix} \times \begin{bmatrix} b_{11} & b_{21} & \cdots & b_{J1} \\ b_{12} & b_{22} & \cdots & b_{J2} \end{bmatrix} = \mathbf{ADB}^T \quad 1.12$$

A equação 1.12 demonstra o produto de três matrizes, onde  $\mathbf{D}$  é uma matriz diagonal, com os elementos da diagonal principal correspondendo às concentrações das duas substâncias, 1 e 2. Note que o modelo da equação 1.12 pode ser representado na forma de:

$$\mathbf{X} = \mathbf{a}_1 c_{11} \mathbf{b}_1^T + \mathbf{a}_2 c_{12} \mathbf{b}_2^T \quad 1.13$$

Considere o caso das medidas feitas com um instrumento HPLC-DAD para um conjunto de amostras contendo duas substâncias, 1 e 2. Para cada amostra  $k$  existe uma matriz  $\mathbf{X}_k$  de dimensão  $I \times J$  para os respectivos modos de monitoramento tempo x comprimento de onda, figura 1C. Um modelo linear que descreva as respostas medidas para este conjunto de amostras deve ser capaz de representar através de operações algébricas as relações de proporcionalidade entre as matrizes  $\mathbf{A}$ ,  $\mathbf{B}$  e  $\mathbf{C}$ . Sendo assim é necessário escrever um modelo envolvendo o produto de três matrizes que resultam no sistema de equações abaixo:



$$\mathbf{X}_{I \times JK} = [\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_K]$$

e a matriz  $\mathbf{X}_{I \times JK}$  é decomposta no modelo abaixo:

$$\mathbf{X}_{I \times JK} = \mathbf{A}(\mathbf{C} \otimes \mathbf{B})^T \quad 1.17$$

onde a matriz  $\mathbf{I}$  é denominada de hiperdiagonal com dimensões  $R \times RR$ , conforme mostra o exemplo abaixo para uma hiperdiagonal de dimensão  $3 \times 9$ , de um arranjo em três modos formado por três componentes:

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad 1.18$$

A equação 1.19 representa o modelo da equação 1.17 quando há duas amostras,  $\mathbf{X}_1$  e  $\mathbf{X}_2$ , contendo duas substâncias químicas,  $R = 2$ .

$$\mathbf{X}_{I \times J2} = \begin{bmatrix} \mathbf{a}_{11} & \mathbf{a}_{12} \\ \mathbf{a}_{21} & \mathbf{a}_{22} \\ \vdots & \vdots \\ \mathbf{a}_{I1} & \mathbf{a}_{I2} \end{bmatrix} \times \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \times \left( \begin{bmatrix} \mathbf{c}_{11} & \mathbf{c}_{12} \\ \mathbf{c}_{21} & \mathbf{c}_{22} \\ \vdots & \vdots \\ \mathbf{c}_{K1} & \mathbf{c}_{K2} \end{bmatrix} \otimes \begin{bmatrix} \mathbf{b}_{11} & \mathbf{b}_{12} \\ \mathbf{b}_{21} & \mathbf{b}_{22} \\ \vdots & \vdots \\ \mathbf{b}_{J1} & \mathbf{b}_{J2} \end{bmatrix} \right)^T \quad 1.19$$

O sistema de equações 1.14 também pode ser representado pelo produto de Kathri-Rao de acordo com o seguinte modelo:

$$\mathbf{X}_{I \times JK} = \mathbf{A}(\mathbf{C} | \otimes | \mathbf{B})^T \quad 1.20$$

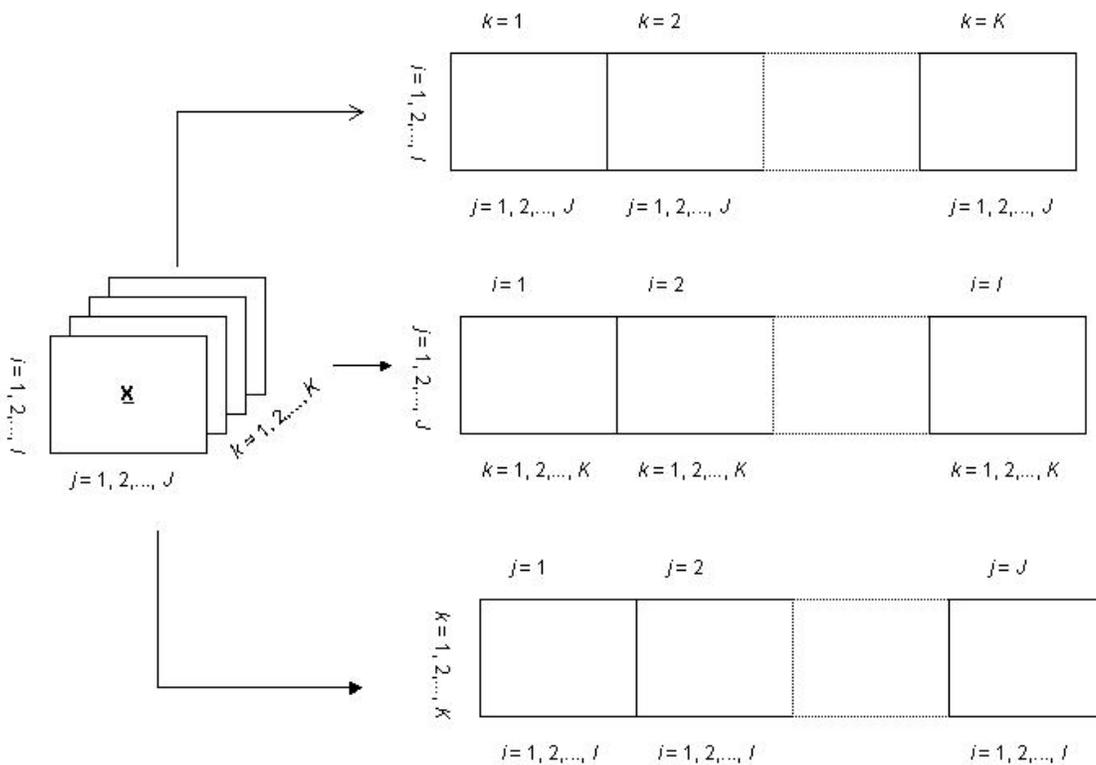
sendo que o símbolo  $| \otimes |$ , refere-se ao produto de Kathri-Rao definido conforme a equação abaixo:

$$\mathbf{C} | \otimes | \mathbf{B} = [\mathbf{c}_1 \otimes \mathbf{b}_1 \quad \mathbf{c}_2 \otimes \mathbf{b}_2 \quad \dots \quad \mathbf{c}_R \otimes \mathbf{b}_R] \quad 1.21$$

note a diferença do produto acima em relação ao produto de Kronecker que envolve a seguinte seqüência de operações:

$$\mathbf{C} \otimes \mathbf{B} = [\mathbf{c}_1 \otimes \mathbf{b}_1 \quad \mathbf{c}_2 \otimes \mathbf{b}_1 \quad \mathbf{c}_1 \otimes \mathbf{b}_2 \quad \mathbf{c}_2 \otimes \mathbf{b}_2 \quad \dots \quad \mathbf{c}_R \otimes \mathbf{b}_R] \quad 1.22$$

O arranjo  $\mathbf{X}_{I \times JK}$  corresponde a uma das formas de desdobrar o cubo  $\mathbf{X}$  em uma matriz. Além desta, existem outras duas formas como mostra a figura 1.2.



**Figura 1.2.** Tipos de desdobramentos do arranjo em três modos:  $\mathbf{X}_{I \times JK}$ ,  $\mathbf{X}_{J \times KI}$  e  $\mathbf{X}_{K \times JI}$ .

## **CAPÍTULO 2**

### **RESOLUÇÃO MULTIVARIADA DE CURVAS**

*Fundamentos Teóricos*

## 2.1 INTRODUÇÃO

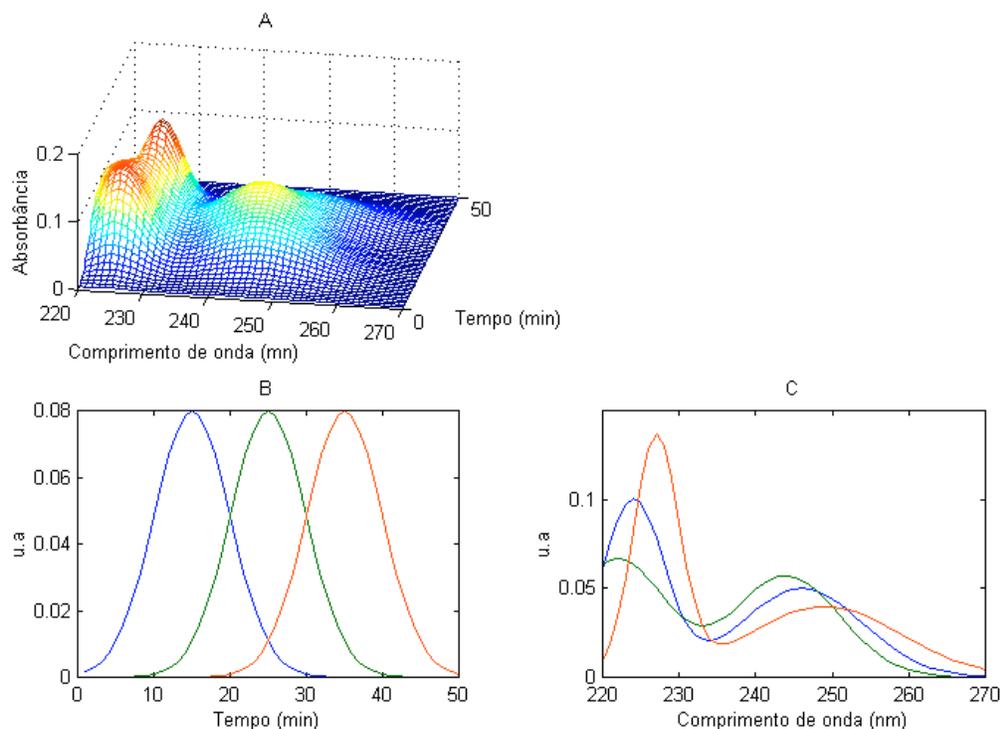
Resolução Multivariada de Curvas (MCR) é uma técnica matemática usada para separar os sinais individuais dos compostos químicos que contribuem para o sinal total registrado em uma mistura destes compostos. Um sinal multivariado pode ser um espectro, um voltamograma, um cromatograma, uma imagem, ondas, etc. Neste texto a resolução multivariada de curvas é tratada como uma técnica composta por diferentes métodos matemáticos.

MCR pode ser utilizada para a separação de espectros sobrepostos de substâncias químicas presentes em uma mistura. Para ilustrar esta aplicação, observe os dados apresentados na figura 2.1A. Estes correspondem a uma simulação computacional dos sinais registrados com um equipamento do tipo HPLC-DAD para uma amostra contendo uma mistura de três substâncias químicas que podem ser detectadas no espectro ultravioleta. Os dados apresentados na figura 2.1A simulam uma análise química em que os cromatogramas e os espectros de três compostos encontram-se sobrepostos. Situações deste tipo ocorrem com certa frequência na análise multiresíduo de pesticidas<sup>16</sup> e durante o desenvolvimento de um método de análise cromatográfica.

Assumindo que o número de espécies químicas presentes nos cromatogramas apresentados na figura 2.1A não é conhecido, podemos agrupar estes dados em uma matriz  $\mathbf{X}$  como aquela mostrada na figura 1.1B. Cada linha desta matriz corresponde a um instante em que o espectro foi registrado e cada coluna corresponde à variação temporal das medidas de absorvância em um comprimento de onda do espectro ultravioleta. A primeira etapa da análise com MCR consiste em determinar o número de componentes do modelo, sendo que no exemplo dado foi determinado três. Sendo assim, a matriz  $\mathbf{X}$  pode ser decomposta na soma de três componentes conforme mostra a equação 2.1:

$$\mathbf{X} = \mathbf{a}_1\mathbf{b}_1^T + \mathbf{a}_2\mathbf{b}_2^T + \mathbf{a}_3\mathbf{b}_3^T \quad 2.1$$

Na equação 2.1, os vetores  $\mathbf{a}_1$ ,  $\mathbf{a}_2$  e  $\mathbf{a}_3$  são estimativas dos cromatogramas (figura 2.1B) e os vetores  $\mathbf{b}_1$ ,  $\mathbf{b}_2$  e  $\mathbf{b}_3$  são estimativas dos espectros (figura 2.1C) de cada uma das três componentes que contribuem para o cromatograma mostrado na figura 2.1A.



**Figura 2.1.** A) Espectros de absorvância registrados em função do tempo para uma separação cromatográfica. B) Cromatogramas dos três compostos químicos. C) Espectros dos compostos químicos.

A aplicação da MCR para a análise de sistemas químicos é bastante ampla, visto que atualmente os químicos dispõem de um arsenal de técnicas instrumentais para o monitoramento multivariado. MCR é usada extensivamente para a análise dos dados espectroscópicos obtidos durante o monitoramento de reações químicas,<sup>17</sup> titulações,<sup>18</sup> misturas envolvendo espécies em equilíbrio químico,<sup>19</sup> ensaios de desnaturação de proteínas,<sup>20</sup> transporte de substâncias através de membranas<sup>21</sup> e para o tratamento de imagens.<sup>22</sup> Além da espectroscopia, encontra-se na literatura uma grande quantidade de aplicações do MCR para a análise de voltamogramas<sup>23</sup> registrados em sistemas químicos semelhantes aos discutidos acima.

## 2.2 MODELOS MATEMÁTICOS MULTIVARIADOS

Um modelo bilinear é definido da seguinte forma:

$$\mathbf{X} = \mathbf{A}\mathbf{B}^T + \mathbf{E}$$

2.2

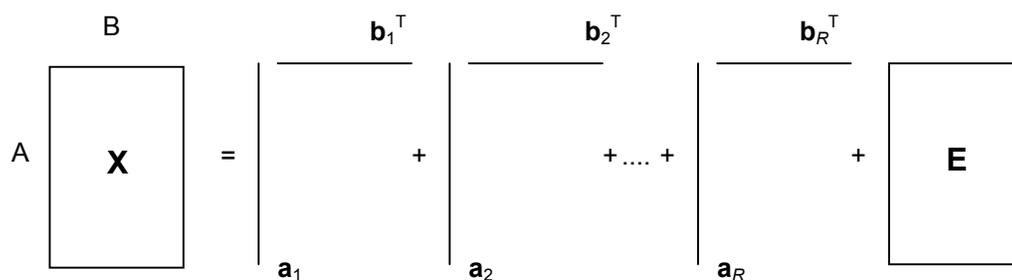
onde  $\mathbf{X}$  é uma matriz contendo os dados registrados para cada variável ou propriedade monitorada no sistema químico. O produto  $\mathbf{A}\mathbf{B}^T$  corresponde à estrutura bilinear e a matriz  $\mathbf{E}$  aos resíduos do modelo, que é a parte de  $\mathbf{X}$  que não pode ser representada pelo modelo bilinear.

Uma outra forma de representar o modelo bilinear é escrevê-lo da seguinte forma:

$$\mathbf{X} = \mathbf{a}_1\mathbf{b}_1^T + \mathbf{a}_2\mathbf{b}_2^T + \dots + \mathbf{a}_R\mathbf{b}_R^T + \mathbf{E} = \sum_{r=1}^R \mathbf{a}_r\mathbf{b}_r^T + \mathbf{E} \quad 2.3$$

A equação 2.3 mostra que a matriz  $\mathbf{X}$  contendo as medidas experimentais pode ser descrita pela soma de  $R$  componentes representados pelo produto de dois vetores  $\mathbf{a}_r\mathbf{b}_r^T$ . Estas componentes estão relacionadas com os sinais das espécies químicas nos modos de monitoramento A e B da matriz (arranjo em dois modos) como mostra a figura 2.2. Dependendo do tipo de decomposição usada, os vetores  $\mathbf{a}_r$  e  $\mathbf{b}_r$  podem ter significado abstrato ou físico. O objetivo da MCR é encontrar a decomposição para a qual os parâmetros do modelo bilinear tenham um significado físico.

O modelo da equação 2.2 é dito bilinear porque tanto os parâmetros de  $\mathbf{A}$  como os de  $\mathbf{B}$  são diretamente proporcionais às medidas em  $\mathbf{X}$ . Apesar disto, nada impede que seja proposto um modelo não linear embora essa prática não tenha sido explorada no contexto da resolução multivariada de curvas.



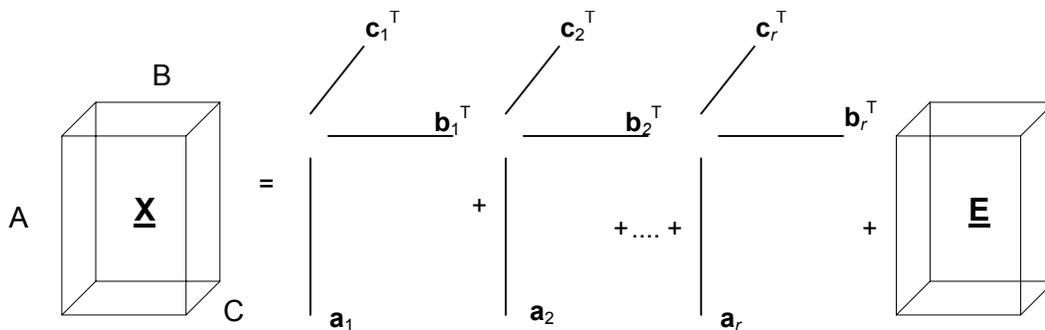
**Figura 2.2.** Representação esquemática de um modelo bilinear.

Outro tipo de modelo muito importante para a descrição de sistemas químicos é o modelo trilinear representado na figura 2.3. Este modelo tem a seguinte forma:

$$\mathbf{X}_{I \times J \times K} = \mathbf{A}(\mathbf{C} | \otimes | \mathbf{B})^T + \mathbf{E} \quad 2.4$$

A matriz  $\mathbf{X}_{I \times J \times K}$  é obtida a partir do desdobramento do cubo de dados conforme mostra a figura 1.2. O modelo trilinear pode ser escrito também como:

$$\begin{aligned} \mathbf{X}_{I \times J \times K} &= \mathbf{a}_1(\mathbf{c}_1 | \otimes | \mathbf{b}_1)^T + \mathbf{a}_2(\mathbf{c}_2 | \otimes | \mathbf{b}_2)^T + \dots + \mathbf{a}_R(\mathbf{c}_R | \otimes | \mathbf{b}_R)^T + \mathbf{E} = \\ &= \sum_{r=1}^R \mathbf{a}_r(\mathbf{c}_r | \otimes | \mathbf{b}_r)^T + \mathbf{E} \end{aligned} \quad 2.5$$



**Figura 2.3.** Representação esquemática de um modelo trilinear.

A equação 2.5 mostra que o cubo de dados  $\mathbf{X}$  contendo as medidas experimentais é formado pela soma de componentes individuais representados pelo produto de três vetores  $\mathbf{a}_r(\mathbf{c}_r | \otimes | \mathbf{b}_r)^T$ . Como no caso dos parâmetros do modelo bilinear, estes vetores estão associados aos sinais dos compostos químicos nos modos de monitoramento A, B e C.

Modelos de ordem superior apresentam três ou mais modos de monitoramento. Para a análise de sistemas químicos já foram relatados na literatura modelos com quatro e cinco modos de monitoramento.<sup>24,25</sup> Um outro modelo de ordem superior muito importante é o modelo Tucker<sup>26</sup> que tem a seguinte forma:

$$\mathbf{X}_{I \times JK} = \mathbf{AG}(\mathbf{C} \otimes \mathbf{B})^T + \mathbf{E}$$

2.6

A matriz  $\mathbf{G}$  é denominada de núcleo e descreve as interações entre aos vetores das matrizes  $\mathbf{A}$ ,  $\mathbf{B}$  e  $\mathbf{C}$ . O modelo Tucker é um modelo linear e genérico, pois através da manipulação dos elementos de  $\mathbf{G}$  podemos gerar diversos modelos com estruturas diferentes.<sup>27</sup> O próprio modelo trilinear corresponde a um caso particular do modelo Tucker.

## 2.3 MÉTODOS MATEMÁTICOS EMPREGADOS PARA O AJUSTE DE MODELOS MULTIVARIADOS

Os métodos de MCR podem ser classificados pelo número de modos que são ajustados no modelo matemático. Um método de análise em dois modos é aquele que ajusta um modelo bilinear enquanto que um método de análise em três modos ajusta um modelo trilinear ou um modelo do tipo Tucker como mostrado na equação 2.6 que têm quatro parâmetros mas refere-se à decomposição de um cubo (arranjo em três modos).

Há vários métodos numéricos para a decomposição de uma matriz ou um cubo nos seus respectivos modelos bi e trilineares. Estes métodos numéricos são classificados em algoritmos iterativos e não iterativos. Fazem parte dos algoritmos não iterativos os métodos “Window Factor Analysis (WFA)”,<sup>28</sup> Análise em Fatores Evolucionários (EFA)<sup>29</sup> e Decomposição Trilinear Direta (DTLD),<sup>30</sup> enquanto que para os métodos iterativos, o método mais conhecido é o do Algoritmo dos Mínimos Quadrados Alternantes (ALS).<sup>31</sup>

O artigo de revisão de Jiang dá uma idéia bastante ampla dos métodos de MCR existentes para a análise em dois modos.<sup>32</sup>

### 2.3.1 Métodos não iterativos

#### 2.3.1.1 Análise em Fatores Evolucionários (EFA)

A Análise em Fatores Evolucionários (EFA) é um método não iterativo de resolução multivariada de curvas para a análise em dois modos. Este método utiliza a Decomposição

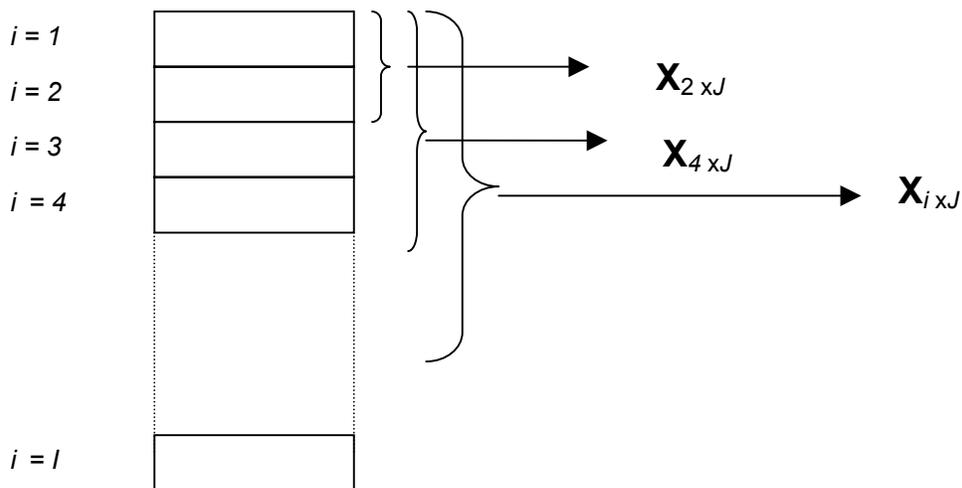
em Valores Singulares (SVD) como ponto de partida para a resolução das curvas e para a determinação do número de componentes que será usado para ajustar o modelo bilinear. Maiores detalhes sobre o método SVD são fornecidos no material suplementar: parte B.

O método EFA, consiste em formar um conjunto de submatrizes  $\mathbf{X}_{i \times J}$ , a partir da matriz  $\mathbf{X}$ , começando da submatriz  $\mathbf{X}_{2 \times J}$  e terminando na matriz  $\mathbf{X}_{i \times J}$  como mostra a figura 2.4. No total são geradas  $I-1$  submatrizes, sendo que cada uma destas é sujeita a uma decomposição em valores singulares. A formação das submatrizes corresponde a um procedimento evolucionário em que as linhas da matriz  $\mathbf{X}$  são varridas com o método SVD. As colunas da matriz  $\mathbf{X}$  podem ser varridas se for usada a matriz  $\mathbf{X}^T$ .

O método SVD consiste em decompor a matriz  $\mathbf{X}$  no seguinte modelo:

$$\mathbf{X} = \mathbf{USV}^T \quad 2.7$$

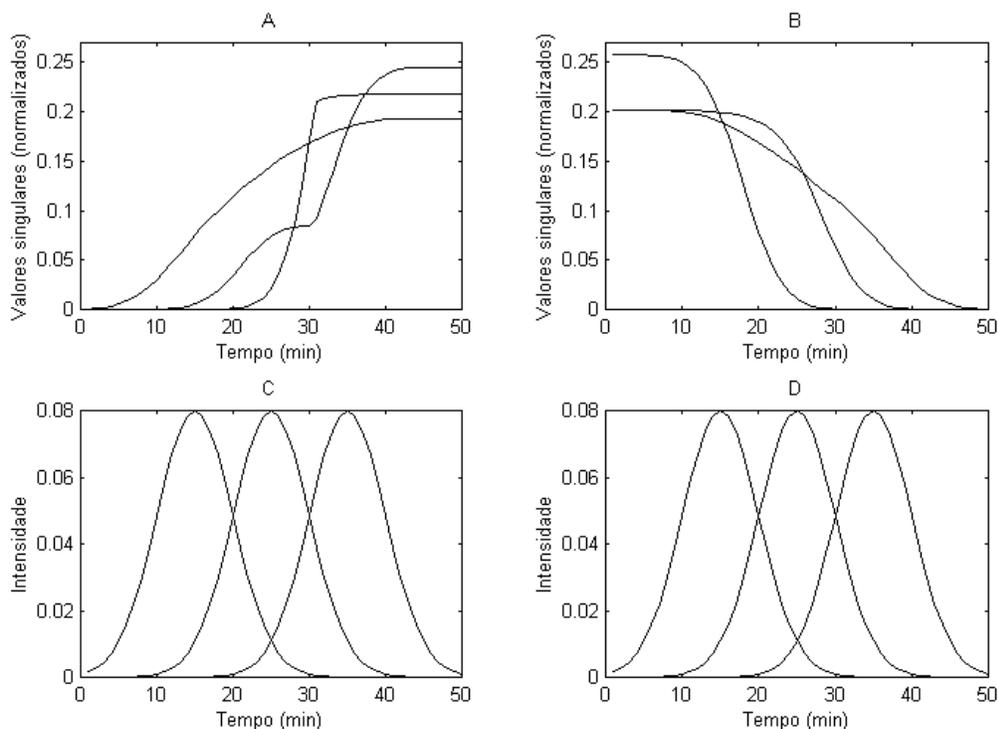
As matrizes,  $\mathbf{U}$  e  $\mathbf{V}$ , são denominadas de autovetores e a matriz  $\mathbf{S}$  de valores singulares (Material Suplementar: parte B.1). O objetivo de varrer as linhas de  $\mathbf{X}$  com o método SVD é monitorar os valores singulares em função do número de linhas desta matriz. Os valores singulares informam o posto da matriz, uma propriedade matemática que está associada ao número de componentes  $R$ , que deve ser usado para o ajuste do modelo bilinear da equação 2.2.



**Figura 2.4.** Esquema da formação das submatrizes  $\mathbf{X}_{i \times J}$ .

Os valores singulares obtidos a partir da análise evolucionária podem ser visualizados em função das linhas como mostram as figuras 2.5A e 2.5B. Estes gráficos foram obtidos a partir da análise dos dados apresentados na figura 2.1A com o método EFA. A varredura das linhas pode ser feita no sentido de 1 a  $l$  (direto), conforme mostra a figura 2.5A, ou no sentido de  $l$  a 1 (inverso), conforme mostra a figura 2.5B. Perceba que na figura 2.5A há três valores singulares maiores que zero que aumentam de magnitude à medida que as linhas da matriz  $\mathbf{X}$  são varridas. Neste exemplo, as linhas da matriz correspondem ao instante em que cada espectro foi registrado (cromatograma da amostra).

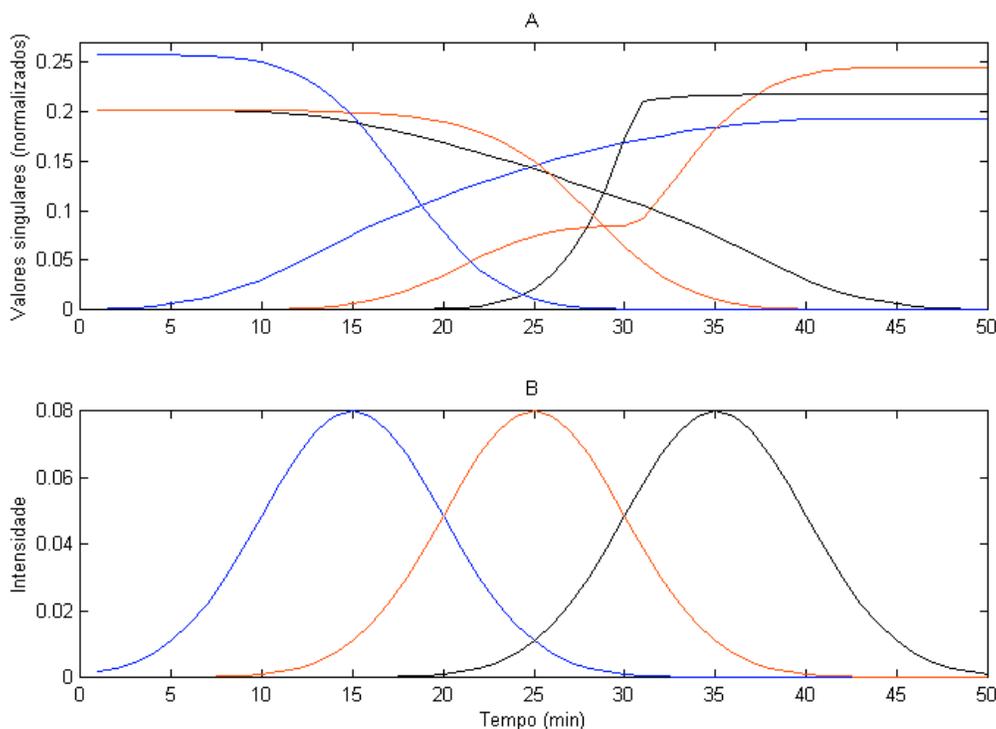
Na figura 2.5A, a partir de 5 min surge o primeiro valor singular diferente de zero. Este coincide aproximadamente com o instante em que é detectada a primeira espécie química da separação cromatográfica conforme mostra a figura 2.5C. O segundo e o terceiro valores singulares diferentes de zero surgem respectivamente após 10 e 20 min, coincidindo com os tempos em que o segundo e o terceiro compostos químicos começam a ser detectados na separação cromatográfica. No sentido inverso da análise, o comportamento é semelhante visto que o primeiro, segundo e o terceiro valores singulares aumentam respectivamente a partir de 50, 40 e 30 min. Em ambos os casos do sentido inverso, o surgimento de cada valor singular diferente de zero corresponde aos instantes em cada que composto químico é detectado da saída da coluna cromatográfica. Há portanto, uma relação dos sinais cromatográficos de cada espécie química com o aumento dos valores singulares em função do número de linhas da matriz  $\mathbf{X}$ .



**Figura 2.5.** A) Mapa dos valores singulares para análise no sentido direto. B) Mapa dos valores singulares para análise no sentido inverso. C) e D) Cromatogramas de cada composto químico da separação cromatográfica apresentada na figura 2.1A.

As análises evolucionárias no sentido direto e inverso podem ser colocadas em um mesmo gráfico denominado de mapa dos valores singulares conforme mostra a figura 2.6A. A combinação das curvas de valores singulares obtidas a partir da análise em ambos sentidos indica a posição em que cada espécie química dá a sua contribuição para as medidas no cromatograma da mistura, figura 2.6B. Esta posição é denominada de “janela”.

A “janela” é uma característica única de cada composto químico e a existência desta em um determinado conjunto de dados é fundamental para que o sinal do composto químico seja resolvido pelo método EFA. Se houver uma “janela” para cada composto químico, o método EFA resolverá os espectros de todos eles.



**Figura 2.6.** A) Mapa dos valores singulares da análise dos dados apresentados na figura 2.1A. B) Cromatogramas dos compostos químicos da figura 2.1A. (-) Primeiro componente, (-) segundo componente e (-) terceiro componente.

O mapa dos valores singulares é uma ferramenta muito usada em MCR, porque dele é deduzido o número de componentes que serão usados para o ajuste do modelo bilinear e se há medidas seletivas para cada componente da matriz. Para a separação dos sinais da mistura é necessário seguir a seguinte seqüência de cálculos:

1. Calcule:  $\mathbf{X} = \mathbf{USV}^T$  **2.8**

2. Defina:  $\mathbf{R}$

3. calcule:  $\mathbf{A} = \overline{\mathbf{U}}\mathbf{R}$  **2.9**

4. Calcule:  $\mathbf{B} = \mathbf{R}^{-1}\overline{\mathbf{V}}^T$  **2.10**

A etapa 1 da seqüência acima, corresponde à decomposição em valores singulares (ver Suplemento B.1). A etapa 2 consiste em definir uma matriz quadrada  $\mathbf{R}$  de posto completo e de dimensão  $R \times R$ , que será usada para aplicar uma transformação não ortogonal nas matrizes truncadas  $\overline{\mathbf{V}}$  de dimensão  $J \times R$  e  $\overline{\mathbf{U}}$  de dimensão  $I \times R$ , a fim de

encontrar novas matrizes **A** e **B** com significado físico. Truncar uma matriz consiste em seleccionar as  $R$  primeiras colunas das matrizes **U** e **V** da equação 2.7.

A matriz **R** pode ser definida da seguinte forma: com base no mapa dos valores singulares evolucionários, figura 2.6A, construa uma matriz **U**<sub>0</sub> com os elementos da matriz  $\bar{\mathbf{U}}$  que estão fora da janela do componente  $r$  e resolva o seguinte sistema linear homogêneo:

$$\mathbf{U}_0 \mathbf{r}_r = \mathbf{0} \quad 2.11$$

Por exemplo, o segundo componente da separação cromatográfica do exemplo da figura 2.1A tem uma janela entre 10 e 40 min como mostra a figura 2.6A. Para calcular o vetor  $\mathbf{r}_2$  de acordo com a equação 2.11, deve-se incluir na matriz **U**<sub>0</sub> as medidas entre 1 e 9 min mais as medidas entre 41 e 50 min.

O vetor  $\mathbf{r}_r$  é uma coluna de **R** que resolve o sinal do composto químico  $r$ , através da seguinte operação:

$$\mathbf{a}_r = \mathbf{U} \mathbf{r}_r \quad 2.12$$

Se forem encontradas “janelas” para cada composto, podemos montar um sistema linear homogêneo como o da equação 2.11 para cada componente e achar todas as colunas de **R**, resolvendo assim os sinais de todos os compostos químicos com a equação 2.9. Para encontrar a matriz **B**, basta resolver a equação 2.10.

O método WFA<sup>28</sup> também utiliza as “janelas” de cada componente para resolver os sinais, mas difere do método EFA quanto ao método matemático empregado.

### 2.3.1.2 Método de Decomposição Trilinear Direta (DTLD)

A Decomposição Trilinear Direta (DTLD) é um método não iterativo de resolução multivariada de curvas para a análise em três modos (um cubo). Neste método, o arranjo em três modos  $\underline{\mathbf{X}}$  é convertido em duas pseudo-amostras **G**<sub>1</sub> e **G**<sub>2</sub>, que posteriormente são introduzidas no problema de autovalores e autovetores generalizados.<sup>33</sup> O método DTLD

inicia com o desdobramento do cubo  $\underline{\mathbf{X}}$  nos três modos que o definem, figura 1.2, e uma decomposição em valores singulares é calculada para cada uma das matrizes formadas no desdobramento,  $\mathbf{X}_{I \times JK}$ ,  $\mathbf{X}_{J \times KI}$ ,  $\mathbf{X}_{K \times JI}$ , conforme mostram as equações abaixo:

$$\mathbf{X}_{I \times JK} = \mathbf{U}_I \mathbf{S}_I \mathbf{V}_I^T \quad \bar{\mathbf{U}} = \mathbf{U}_I \quad (I \times R) \quad 2.13$$

$$\mathbf{X}_{J \times KI} = \mathbf{U}_J \mathbf{S}_J \mathbf{V}_J^T \quad \bar{\mathbf{V}} = \mathbf{U}_J \quad (J \times R) \quad 2.14$$

$$\mathbf{X}_{K \times JI} = \mathbf{U}_K \mathbf{S}_K \mathbf{V}_K^T \quad \bar{\mathbf{W}} = \mathbf{U}_K \quad (K \times 2) \quad 2.15$$

As matrizes  $\bar{\mathbf{U}}$ ,  $\bar{\mathbf{V}}$ ,  $\bar{\mathbf{W}}$  são as matrizes obtidas após truncar as matrizes originais, sendo  $R$  o número de componentes do modelo trilinear que está associado ao número de compostos químicos que contribuem para os dados presentes no cubo  $\underline{\mathbf{X}}$ . Após a SVD, duas pseudo amostras,  $\mathbf{G}_1$  e  $\mathbf{G}_2$  são construídas de acordo com as equações abaixo:

$$\mathbf{G}_1 = \sum_{k=1}^K \mathbf{w}_{k2} \bar{\mathbf{U}}^T \mathbf{X}_k \bar{\mathbf{V}} \quad 2.16$$

$$\mathbf{G}_2 = \sum_{k=1}^K \mathbf{w}_{k1} \bar{\mathbf{U}}^T \mathbf{X}_k \bar{\mathbf{V}} \quad 2.17$$

As pseudo-amostras  $\mathbf{G}_1$  e  $\mathbf{G}_2$ , são introduzidas no problema dos autovalores e autovetores generalizados, equação 2.18 e 2.19, para calcular os vetores que definem uma base comum a ambas pseudo-amostras.

$$\mathbf{G}_1 \mathbf{Z}_a = \mathbf{G}_2 \mathbf{Z}_a \lambda_a \quad 2.18$$

$$\mathbf{G}_2^T \mathbf{Z}_b = \mathbf{G}_2^T \mathbf{Z}_b \lambda_b \quad 2.19$$

Nas equações 2.18 e 2.19,  $\mathbf{Z}_a$  e  $\mathbf{Z}_b$  são os autovetores e  $\lambda_a$ ,  $\lambda_b$  são os autovalores dos respectivos autovetores. Na prática a solução pode levar a autovetores complexos e neste caso é necessário realizar uma transformação de similaridade para transformá-los em vetores reais.<sup>34</sup>

Uma estimativa dos sinais de cada um dos compostos químicos é obtida resolvendo as equações abaixo:

$$\mathbf{A} = \overline{\mathbf{U}}(\mathbf{Z}_a^T)^{-1} \quad \mathbf{2.20}$$

$$\mathbf{B} = \overline{\mathbf{V}}(\mathbf{Z}_b^T)^{-1} \quad \mathbf{2.21}$$

As concentrações de cada espécie presente nas amostras que compõe o cubo  $\mathbf{X}$  podem ser calculadas de acordo com as seguintes operações:

$$\mathbf{Z} = (\mathbf{B} | \otimes | \mathbf{A}) \quad \mathbf{2.22}$$

$$\mathbf{C} = (\mathbf{Z}^{-1} \mathbf{X}_{K \times IJ}^T)^T \quad \mathbf{2.23}$$

Geralmente a decomposição trilinear com o método DTLTD é rápida e esta característica tornou este método bastante adequado para as análises químicas quantitativas (calibração de segunda ordem).

## 2.3.2 Métodos de resolução iterativos

### 2.3.2.1 Método dos Mínimos Quadrados Alternantes em Dois Modos (ALS-MCR)

O método do ajuste dos Mínimos Quadrados Alternantes (ALS)<sup>31</sup> consiste em uma otimização onde os parâmetros do modelo bilinear,  $\mathbf{AB}^T$ , são calculados através de uma seqüência de operações matemáticas que são repetidas em ciclos (iteração) até que um critério bem definido tenha sido atingido. Para cada iteração do algoritmo, é obtida uma nova estimativa dos parâmetros  $\mathbf{A}$  e  $\mathbf{B}$  a partir da solução do problema dos mínimos quadrados (Material suplementar: parte B.2).

O critério usado para o ajuste de um modelo pelo método dos mínimos quadrados ordinários é o de minimizar a soma quadrática dos resíduos do modelo bilinear. O resíduo  $\mathbf{E}$  do modelo bilinear é calculado como:

$$\mathbf{E} = \mathbf{X} - \mathbf{AB}^T \quad 2.24$$

logo o critério dos mínimos quadrados consiste em:

$$\min_{\mathbf{A}} \|\mathbf{X} - \mathbf{AB}^T\|_2^2 \quad 2.25$$

onde o símbolo  $\|\cdot\|_2^2$  indica a norma euclidiana ao quadrado. A soma quadrática dos resíduos do modelo bilinear é definida como:

$$SQR = \|\mathbf{X} - \mathbf{AB}^T\|_2^2 \quad 2.26$$

O algoritmo ALS, como um todo, apresenta a seguinte seqüência de cálculos:

1. Defina um critério de convergência:  $\varepsilon$
2. Inicie uma matriz  $\mathbf{A}_0$  e calcule:

$$\mathbf{A} = \mathbf{A}_0;$$

3. Resolva o problema dos mínimos quadrados para obter  $\mathbf{B}$ :

$$\min_{\mathbf{B}} \|\mathbf{X} - \mathbf{AB}^T\|_2^2$$

4. Com a solução da etapa 3, resolva o problema abaixo para obter  $\mathbf{A}$ :

$$\min_{\mathbf{A}} \|\mathbf{X} - \mathbf{AB}^T\|_2^2$$

5. Calcule:

$$SQR = \|\mathbf{X} - \mathbf{AB}^T\|_2^2 \quad \text{e} \quad \sigma = \frac{SQR_{iter-1} - SQR_{iter}}{SQR_{iter-1}}$$

6. Se  $\sigma \geq \varepsilon$ , retorne ao passo 3.

O método do Ajuste por Mínimos Quadrados (LS)<sup>35</sup> desempenha um papel fundamental no Algoritmo dos Mínimos Quadrados Alternantes, pois é através do LS que

se obtêm as estimativas para os parâmetros **A** e **B**, nas etapas 3 e 4. Maiores detalhes sobre como resolver o método LS é fornecido no material suplementar: parte B.2.

### 2.3.2.2 *Análise de Decomposição em Fatores Paralelos (PARAFAC)*

O método PARAFAC consiste no Algoritmo dos Mínimos Quadrados Alternantes para o ajuste de um modelo trilinear conforme mostra a seqüência de cálculos abaixo:

1. Defina o critério de convergência:  $\varepsilon$

2. Inicie as matrizes **B**<sub>0</sub> e **C**<sub>0</sub> e calcule:

$$\mathbf{B} = \mathbf{B}_0 \text{ e } \mathbf{C} = \mathbf{C}_0$$

3. Calcule **A**:

$$\min_{\mathbf{A}} \left\| \mathbf{X}_{I \times JK} - \mathbf{A} \mathbf{Z}^T \right\|_2^2 \quad \text{onde: } \mathbf{Z} = (\mathbf{C} | \otimes | \mathbf{B})$$

4. Atualize **B**:

$$\min_{\mathbf{B}} \left\| \mathbf{X}_{J \times KI} - \mathbf{B} \mathbf{Z}^T \right\|_2^2 \quad \text{onde: } \mathbf{Z} = (\mathbf{C} | \otimes | \mathbf{A})$$

5. Atualize **C**:

$$\min_{\mathbf{C}} \left\| \mathbf{X}_{K \times IJ} - \mathbf{C} \mathbf{Z}^T \right\|_2^2 \quad \text{onde: } \mathbf{Z} = (\mathbf{B} | \otimes | \mathbf{A})$$

6. Calcule o critério de convergência e SQR:

$$\text{SQR} = \left\| \mathbf{X}_{I \times JK} - \mathbf{A}(\mathbf{C} | \otimes | \mathbf{B})^T \right\|_2^2 \text{ e } \sigma = \frac{\text{SQR}_{iter-1} - \text{SQR}_{iter}}{\text{SQR}_{iter-1}}$$

7. Se  $\sigma \geq \varepsilon$ , retorne ao passo 3.

8. Normalize os fatores **A** e **B** (opcional).

Note que para realizar os cálculos é necessário desdobrar o cubo nas três matrizes,  $\mathbf{X}_{I \times JK}$ ,  $\mathbf{X}_{J \times KI}$ ,  $\mathbf{X}_{K \times IJ}$ , apresentadas na figura 1.2.

Além do algoritmo ALS, existem outros métodos iterativos para a decomposição trilinear, como o método “Three-Way Positive Matrix Factorization” (PMF3),<sup>36</sup> “Alternating Trilinear Decomposition” (ATLD)<sup>37</sup> e o “Self-weighted alternating trilinear decomposition” (SWTLD).<sup>38</sup> O método PMF3, é bastante eficiente para resolver sinais com um nível de

sobreposição muito grande, porém peca pela grande quantidade de memória computacional que consome para fazer os cálculos.

Os métodos ATLD e SWTLD ajustam o modelo trilinear com menos iterações que o método PARAFAC e, além disso, os parâmetros do modelo trilinear calculados por estes métodos são insensíveis ao número de componentes usados.<sup>38</sup> No caso do método PARAFAC é necessário determinar o número exato de componentes que contribuem para os sinais colocados no cubo  $\underline{X}$ . Se este valor for excedido não há garantias de que os parâmetros ajustados com o método PARAFAC tenham um significado físico. Por outro lado com os métodos SWTLD e ATLD, os verdadeiros sinais são resolvidos mesmo que tenha sido ajustado um modelo com mais componentes do que realmente existem em  $\underline{X}$ .

Alguns trabalhos recentes foram feitos com o objetivo de comparar os diversos algoritmos de decomposição trilinear iterativos. Estes trabalhos, defendem que o método PARAFAC é melhor no requisito de exatidão dos sinais resolvidos.<sup>39,40</sup> Mas em termos práticos, recomendam que o analista deve pesar as vantagens e desvantagens de cada método. Quando a exatidão da resolução é um fator preponderante, o método PARAFAC é a melhor escolha, todavia se este não é o caso, o método SWTLD é provavelmente a melhor escolha porque é um método mais prático e rápido.

## 2.4 AMBIGÜIDADE E UNICIDADE DA SOLUÇÃO DOS MÉTODOS DE MCR

Dependendo do método utilizado para o ajuste de um modelo bi ou trilinear, a solução é sujeita a dois tipos de ambigüidades denominadas de “ambigüidade de intensidade” e “ambigüidade de rotação”.<sup>31,41</sup>

Até o momento, neste texto, os parâmetros dos modelos bi e trilinear foram definidos como **A**, **B** e **C**. Na verdade, estes são parâmetros genéricos, visto que representam todas as soluções possíveis com algum método de MCR para o ajuste de um modelo bi ou trilinear. Considerando que existe uma única solução do modelo bilinear que representa os verdadeiros sinais (sinais “puros”) de cada componente na matriz **X**, devemos escrever esta solução como:

$$\mathbf{X} = \mathbf{KS}^T + \mathbf{E}$$

2.27

e para um modelo trilinear como:

$$\mathbf{X}_{I \times J \times K} = \mathbf{F}(\mathbf{K} | \otimes | \mathbf{S})^T + \mathbf{E} \quad 2.28$$

A ambigüidade de intensidade significa que os parâmetros do modelo foram resolvidos para uma escala de intensidade desconhecida em que as matrizes  $\mathbf{K}$  e  $\mathbf{S}$  (modelo bilinear) estão multiplicadas por um valor constante como mostra a relação definida abaixo:

$$\mathbf{X} = \sum_{r=1}^R \frac{1}{\alpha_r} \mathbf{k}_r (\alpha_r \mathbf{s}_r^T) + \mathbf{E} = \sum_{r=1}^R \mathbf{a}_r \mathbf{b}_r^T + \mathbf{E} \quad 2.29$$

onde  $\mathbf{a} = (1/\alpha_r) \mathbf{k}$  e  $\mathbf{b} = \alpha_r \mathbf{s}_r$ . Como se vê, a multiplicação da matriz  $\mathbf{S}$  por uma constante é compensada pela divisão da matriz  $\mathbf{K}$  por esta mesma constante. Em termos práticos a ambigüidade de intensidade impede que possamos identificar a intensidade real dos sinais puros de cada composto, mas esta ambigüidade não altera a forma dos sinais contidos nas matrizes  $\mathbf{K}$  e  $\mathbf{S}$ . Portanto se somente este tipo de intensidade estiver presente na solução do modelo, podemos realizar análises qualitativas como a de identificar as substâncias pela forma dos espectros resolvidos. O valor de  $\alpha$  pode ser determinado se no conjunto de amostras que são usadas para a resolução de curvas, tiver soluções padrões dos compostos que se deseja determinar.

A ambigüidade de rotação refere-se ao caso em que os sinais resolvidos para um modelo bi ou trilinear correspondem a uma combinação linear dos verdadeiros sinais  $\mathbf{F}$ ,  $\mathbf{K}$  e  $\mathbf{S}$ . Isto significa que podemos encontrar uma solução matemática que tem um significado físico abstrato visto que esta solução não corresponde aos sinais “puros” de cada componente da matriz  $\mathbf{X}$ . Ao contrário, os sinais obtidos são uma “mistura” dos sinais puros, combinados de acordo com alguma relação matemática desconhecida. O problema da ambigüidade de rotação pode ser representado matematicamente por:

$$\mathbf{X} = \mathbf{K}\mathbf{S}^T + \mathbf{E} = \mathbf{K}\mathbf{R}\mathbf{R}^{-1}\mathbf{S}^T + \mathbf{E} = (\mathbf{K}\mathbf{R})[\mathbf{S}(\mathbf{R}^{-1})^T]^T + \mathbf{E} = \mathbf{A}\mathbf{B}^T + \mathbf{E} \quad 2.30$$

onde  $\mathbf{A} = \mathbf{KR}$  e  $\mathbf{B} = \mathbf{S}(\mathbf{R}^{-1})^T$ , sendo  $\mathbf{R}$  uma matriz não ortogonal de dimensão  $R \times R$  e posto completo. A equação 2.30 mostra que uma combinação linear das colunas de  $\mathbf{K}$ , pode ser compensada por uma combinação linear apropriada das colunas da matriz  $\mathbf{S}$ , sem que seja alterado os resíduos do modelo bilinear  $\mathbf{E}$ . Por isso, uma solução particular obtida para um modelo bilinear através de um método de MCR pode ser modificada sem que a função de ajuste do modelo como expressa pela equação 2.26 seja alterada. Em outras palavras, a solução obtida do método de MCR pode ser convertida em outra que também é “correta”, segundo o critério matemático usado para o ajuste do modelo.

Freqüentemente, a ambigüidade de rotação é um problema difícil de ser resolvido e esta propriedade impede que o objetivo principal da técnica MCR seja alcançado, que é a de separar os sinais “puros”. O problema tem sido bastante discutido na literatura, e algumas estratégias têm sido sugeridas para resolver ou pelo menos minimizá-lo, sendo uma das principais a utilização de restrições de modelo que serão discutidas a seguir.

A ambigüidade de solução é um problema inerente a muitos métodos de resolução multivariada de curvas, sendo a ambigüidade de rotação a mais problemática. Os métodos de decomposição trilinear PARAFAC e DTLTD, resolvem um modelo sem ambigüidades de rotação se as componentes do cubo  $\mathbf{X}$  seguem razoavelmente bem o comportamento trilinear. Devido a esta propriedade estes métodos têm sido retratados como métodos de solução única<sup>42</sup> pois apresentariam a importante propriedade matemática de “unicidade da solução”. Por unicidade da solução entende-se que a solução obtida não pode ser convertida em outra sem que a função de ajuste do modelo seja alterada. Todavia mesmo para os métodos PARAFAC e DTLTD, permanece a ambigüidade de intensidade na solução do modelo trilinear.

## 2.5 RESTRIÇÕES DE MODELO

Uma restrição é uma imposição na solução do modelo matemático. As restrições de cálculo são extensivamente usadas com o objetivo de minimizar o problema da ambigüidade de rotação pois estas restringem a solução de um método de resolução de curvas a um intervalo pequeno de soluções possíveis.<sup>43,44</sup>

Para ilustrar o uso das restrições de cálculo considere os dados apresentados na figura 2.1A. O ajuste de um modelo bilinear a partir deste conjunto de dados resulta na solução formada pelos parâmetros **A** e **B**, sujeita à ambigüidade de solução, que são estimativas dos perfis de eluição (figura 2.1B) e dos espectros de absorção molecular (figura 2.1C) de cada composto químico presente na amostra. Para o ajuste deste modelo podemos colocar uma restrição no cálculo que estabelece que apenas valores maiores ou iguais a zeros sejam considerados para o ajuste das matrizes **A** e **B**, visto que fisicamente os cromatogramas e os espectros de absorção molecular são positivos. Este tipo de restrição é denominado de não negatividade.<sup>45</sup>

Outro tipo de restrição muito usada é a de fixar os elementos das matrizes **A** e/ou **B** quando estes forem conhecidos com antecedência. No exemplo da figura 2.1A, pode ser que antes de ajustar o modelo bilinear, o analista conheça o espectro de alguma substância e neste caso não há necessidade de calculá-lo pois ele pode ser fixado no ajuste do modelo.<sup>46</sup> A restrição de “closure”<sup>41</sup> é bastante mencionada para a análise de sistemas envolvendo um equilíbrio químico. Neste tipo de restrição os parâmetros da matriz **A**, por exemplo, somam sempre um valor constante tal como ocorre no balanço das massas dos reagentes e produtos de uma reação química.

A seletividade<sup>31</sup> é um tipo de restrição muito importante. As “janelas” discutidas no método EFA são um tipo de seletividade que este método explora para resolver o sinal de cada componente presente no sistema analisado. A “janela” corresponde à região em que uma espécie química responde ao método analítico utilizado e fora desta “janela” sua contribuição é igual a zero. Por isso, a restrição de seletividade pode ser adicionada no método MCR-ALS fixando valores iguais a zero para aqueles elementos dos vetores  $\mathbf{a}_r$  e  $\mathbf{b}_r$  que estão fora da “janela” de contribuição da espécie  $r$ .

Uma revisão sobre as restrições de modelo pode ser encontrada na tese de Bro.<sup>47</sup> Restrições de cálculo são classificadas em restrições de igualdade e de desigualdade. São exemplos de restrições de igualdade usadas para fixar parâmetros: “Closure” e seletividade. A restrição de não negatividade é um exemplo de restrição de desigualdade. O artigo de Van Benthem<sup>48</sup> faz uma ótima explanação dos métodos numéricos computacionais usados para implementar restrições lineares de igualdade de acordo com o critério de ajuste por mínimos quadrados.

A restrição de igualdade no problema dos mínimos quadrados é matematicamente formulada da seguinte forma: Dada uma matriz  $\mathbf{B}$  de dimensão  $J \times R$ , e um vetor  $\mathbf{x}$  de dimensão  $J \times 1$ , encontre o vetor  $\mathbf{a}$  de dimensão  $R \times 1$ , que satisfaça:

$$\min_{\mathbf{a}} \|\mathbf{x} - \mathbf{Ba}\|_2^2 \quad \text{sujeito a: } \mathbf{Ha} = \mathbf{d} \quad \mathbf{2.31}$$

onde  $\mathbf{H}$  é uma matriz de dimensão  $M \times R$  e  $\mathbf{d}$  um vetor de dimensão  $M \times 1$ . A matriz  $\mathbf{H}$  contém coeficientes que definem a combinação linear em que o produto  $\mathbf{Ha}$  é igual ao vetor  $\mathbf{d}$ . Para aqueles elementos da matriz  $\mathbf{H}$  que forem iguais a zero, as respectivas linhas do vetor  $\mathbf{a}$  serão livres de restrição.

A equação 2.31 é conectada ao modelo bilinear dado na equação 2.2 se considerarmos o vetor  $\mathbf{x}$  é uma linha transposta da matriz  $\mathbf{X}$ , o vetor  $\mathbf{a}$  uma linha transposta da matriz  $\mathbf{A}$  e a matriz  $\mathbf{B}$  a mesma matriz para ambas equações. Todos os elementos da matriz  $\mathbf{A}$  podem ser sujeitos a restrições de igualdade, se for resolvido um problema dos mínimos quadrados como o da equação 2.31 para cada uma das linhas da matriz  $\mathbf{A}$ . Para uma descrição mais detalhada do assunto recomenda-se uma consulta ao livro de Lawson.<sup>49</sup>

A restrição de não negatividade é uma restrição de desigualdade, matematicamente formulada conforme a equação abaixo:

$$\min_{\mathbf{a}} \|\mathbf{x} - \mathbf{Ba}\|_2^2 \quad \text{sujeito a: } \mathbf{a} \geq \mathbf{0}, \quad \mathbf{2.32}$$

O artigo de Haskell discute a inclusão simultânea de restrições de igualdade e desigualdade.<sup>50</sup> Um tratamento matemático detalhado deste assunto bem como os algoritmos são encontrados no livro de Lawson.<sup>49</sup>

## B. MATERIAL SUPLEMENTAR: PARTE B

### B.1 Decomposição em Valores Singulares (SVD)

Seja  $\mathbf{X}$  uma matriz ( $I \times J$ ) composta por números reais. Então existe uma matriz ortogonal  $\mathbf{U}$  ( $I \times I$ ), e uma matriz ortogonal  $\mathbf{V}$  ( $J \times J$ ) que satisfaz a seguinte decomposição:

$$\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^T \quad 2.33$$

Os elementos diagonais ( $r,r$ ) da matriz  $\mathbf{S}$  ( $I \times J$ ) são maiores que zero e todos os outros elementos fora desta diagonal são iguais a zero. O esquema abaixo ilustra a estrutura da matriz  $\mathbf{S}$  quando  $I > J$ .

$$\mathbf{S} = \begin{bmatrix} s_{1,1} & 0 & 0 & 0 & 0 & 0 \\ 0 & s_{2,2} & 0 & 0 & 0 & 0 \\ 0 & 0 & s_{3,3} & 0 & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & 0 & 0 & s_{R,R} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Os elementos  $s_{r,r}$  são chamados de valores singulares de  $\mathbf{X}$ , sendo que estes são positivos, constantes e satisfazem a seguinte relação:

$$s_{1,1} \geq s_{2,2} \geq s_{3,3}, \dots \geq s_{R,R} \geq 0$$

O valor singular  $s_{R,R}$ , que é o último valor singular diferente de zero, informa o posto da matriz  $\mathbf{X}$  que é  $R$ . O posto de uma matriz é definido como o número de linhas linearmente independentes da matriz  $\mathbf{X}$ . A determinação do posto da matriz está associada à determinação do número de componentes do modelo bilinear.

A decomposição em valores singulares pode ser vista como uma expansão da matriz  $\mathbf{X}$ , da mesma forma que as expansões de Taylor para um polinômio. Se  $R$  é o posto

da matriz  $\mathbf{X}$ , então esta matriz pode ser representada como a soma da contribuição do produto dos vetores  $\mathbf{u}_r$  por  $\mathbf{v}_r$  para todo  $r = 1, 2, \dots, R \leq \min(I, J)$ , conforme mostra a equação abaixo:

$$\mathbf{X} = \sum_{r=1}^R s_{r,r} \mathbf{u}_r \mathbf{v}_r^T = \bar{\mathbf{U}}_R \bar{\mathbf{S}}_R \bar{\mathbf{V}}_R^T \quad 2.34$$

onde,  $\bar{\mathbf{U}} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_R]$  tem a dimensão de  $I \times R$ , e  $\bar{\mathbf{V}} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_R]$  tem dimensão de  $J \times R$ . A matriz  $\bar{\mathbf{S}}$  é diagonal de dimensão  $R$  com os elementos da diagonal iguais aos valores singulares  $s_{r,r}$ . As matrizes,  $\bar{\mathbf{U}}$ ,  $\bar{\mathbf{V}}$  e  $\bar{\mathbf{S}}$  são denominadas de matrizes truncadas.

A decomposição em valores singulares de uma matriz  $\mathbf{X}$ , pode ser interpretada como uma compressão do número de variáveis necessários para descrever  $\mathbf{X}$ , pois  $R \leq \min(I, J)$ . A informação contida nas variáveis originais, é comprimida em um modelo contendo  $R$  variáveis latentes.

## B.2 O Problema dos Mínimos Quadrados

Seja  $\mathbf{X}$  uma matriz  $I \times J$  formada pelas variáveis independentes e  $\mathbf{y}$  um vetor  $I \times 1$  formado pela variável dependente. Um modelo linear que relaciona  $\mathbf{y}$  com  $\mathbf{X}$  é expresso pelo sistema de equações:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{e} \quad 2.35$$

onde  $\mathbf{b}$  são os coeficientes da função linear que relaciona cada coluna (variável) de  $\mathbf{X}$  com  $\mathbf{y}$  e os elementos do vetor  $\mathbf{e}$  são os resíduos do modelo, ou seja à porção dos dados que não pode ser descrita pelo produto  $\mathbf{Xb}$ .

Um sistema de equações lineares pode apresentar umas das seguintes características:

1. Número de linhas iguais ao número de colunas ( $I = J$ );
2. Número de linhas maior que número de colunas ( $I > J$ );

3. Número de linhas menor que número de colunas ( $I < J$ );

Além disto podemos ter ainda:

4. Posto ( $\mathbf{X}$ )  $\leq \min(I, J)$ .

O problema dos mínimos quadrados (LS) consiste em encontrar um vetor de coeficientes  $\mathbf{b}$  para o qual, o produto  $\mathbf{Xb}$  é a melhor aproximação de  $\mathbf{y}$ . Este problema é enunciado da seguinte forma:

“Dada uma matriz real  $\mathbf{X}$  de dimensão  $I \times J$  com  $\text{posto}(\mathbf{X}) \leq \min(I, J)$  e dado um vetor  $\mathbf{y}$  de dimensão  $J \times 1$ , encontre um vetor  $\mathbf{b}$  que minimiza o comprimento euclidiano do vetor de resíduos  $\mathbf{y} - \mathbf{Xb}$ .”

Note que de acordo com a equação 2.35, o vetor de resíduos é calculado como:

$$\mathbf{e} = \mathbf{y} - \mathbf{Xb} \quad 2.36$$

e o problema dos mínimos quadrado consiste em obter um modelo linear que minimiza a soma quadrática dos resíduos. O enunciado do problema dos mínimos quadrados é simbolicamente representado por:

$$\min_{\mathbf{b}} \|\mathbf{y} - \mathbf{Xb}\|_2^2 \quad 2.37$$

Para um sistema linear com  $I > J$  e  $\text{posto}(\mathbf{X}) = J$ , a solução dos mínimos quadrados pode ser calculada da seguinte forma:

$$\mathbf{X}^T \mathbf{y} = \mathbf{X}^T \mathbf{Xb} \quad 2.38$$

de modo que:

$$\mathbf{b} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad 2.39$$

A equação 2.39 é uma representação da solução dos mínimos quadrados e pode ser implementada diretamente em “pacotes computacionais” como MATLAB e Scilab. Todavia em linguagens de programação como Fortran ou C, a solução dos mínimos quadrados é calculada através do método de decomposição de Cholesky, ou do método de decomposição QR.<sup>35</sup> Estes métodos encontram-se disponíveis também nos softwares Matlab e Scilab. Deve-se salientar ainda que a inversão da matriz de covariância  $\mathbf{X}^T\mathbf{X}$ , não é um método numérico eficiente e dependendo do número de condição desta matriz, a solução obtida pode ser instável, mesmo quando posto  $(\mathbf{X}) = J$ .

Uma das situações mais encontradas em química, é a de resolver um sistema de equações lineares onde  $I < J$  ou  $I > J$  sendo que o posto  $(\mathbf{X}) < \min(I, J)$ . Estes sistemas são conhecidos por sistemas com deficiência de posto. Neste caso, uma forma mais estável de calcular a solução para o problema dos mínimos quadrados é usar o método SVD tal como mostrado abaixo:

$$\mathbf{b} = \overline{\mathbf{V}}\overline{\mathbf{S}}^{-1}\overline{\mathbf{U}}^T \mathbf{y} \qquad \qquad \qquad \mathbf{2.40}$$

onde  $\overline{\mathbf{U}}$ ,  $\overline{\mathbf{V}}$  e  $\overline{\mathbf{S}}$  são as matrizes truncadas definidas no equação 2.34

Um teorema importante é que se o posto da matriz  $\mathbf{X}$  é menor que  $I$  ou  $J$ , então o problema dos mínimos quadrados não tem uma solução única. Mas entre todas as soluções possíveis, há uma única solução que minimiza o comprimento euclidiano de  $\mathbf{b}$ . Esta solução é denominada de “norma mínima”.

## **CAPÍTULO 3**

### **MONITORAMENTO ESPECTROSCÓPICO DE REAÇÕES QUÍMICAS COM MODELOS EMPÍRICOS ADAPTATIVOS**

### 3.1 INTRODUÇÃO

Reatores que operam em modo descontínuo (bateladas) desempenham um importante papel dentro da indústria para a produção de polímeros, fármacos e produtos químicos de alto valor agregado. Nestes processos normalmente é necessário monitorar a composição química dentro do reator a fim controlar a qualidade dos produtos e a segurança das operações industriais.

Espectrofotômetros equipados com fibras óticas podem ser usados para fazer medições remotas, rápidas e multivariáveis da composição química de dentro do reator.<sup>51,52,53</sup> Mas para a análise dos dados obtidos neste tipo de monitoramento são necessários métodos matemáticos sofisticados como os métodos da resolução multivariada de curvas (MCR).<sup>54,55,56,57,58</sup>

Os métodos da MCR podem ser usados para ajustar um modelo empírico bilinear como o da equação 2.2, para descrever a variação das medidas espectroscópicas em função do tempo, decorrente da variação da composição química da reação. Uma das vantagens dos métodos da MCR é que não precisamos saber o número nem as espécies que participam da reação na etapa de modelagem dos dados.

O objetivo da análise com os métodos MCR é resolver os sinais dos compostos que formam uma mistura. No caso específico do monitoramento espectroscópico de reações químicas, os respectivos sinais a serem resolvidos são a variação da concentração em função do tempo e os espectros puros dos compostos que participam da reação. Estes sinais são representados por um modelo bilinear que pode ser usado para diferentes aplicações, como para a caracterização da reação química em bateladas consecutivas produzidas no reator,<sup>57,58</sup> para fazer estimativas das constantes de velocidade das reações químicas,<sup>59,60</sup> para o controle estatístico de processos<sup>61</sup> ou simplesmente para uma análise exploratória do mecanismo da reação.

Para o monitoramento, em tempo real, da composição química da reação, primeiro o modelo bilinear é ajustado usando um método de MCR para a análise dos espectros registrados para uma batelada de referência da reação. Em uma etapa subsequente os espectros puros estimados deste modelo são usados como coeficientes para prever as concentrações dos reagentes e produtos de novas bateladas da mesma reação química.

Entretanto, os espectros registrados entre uma batelada e outra podem sofrer alterações que afetam a exatidão das previsões feitas com o modelo bilinear. O próprio espectrofotômetro pode ser responsável por estas alterações, devido aos deslocamentos da linha de base,<sup>62</sup> especialmente para os equipamentos que monitoram em linha as plantas industriais. Além disto, podem ocorrer alterações nos espectros das substâncias químicas devido a mudanças de temperatura, das interações por ligação hidrogênio e da viscosidade do meio da reação.<sup>63</sup>

A metodologia desenvolvida nesta tese aborda este problema e propõe que o método “Predicted Augmented Classical Least Squares” (PACLS) desenvolvido por Haaland,<sup>64</sup> seja usado para adaptar o modelo bilinear obtido a partir da análise com um método de MCR. O método PACLS consiste em adicionar aos coeficientes do modelo bilinear uma componente a mais, que modela as alterações que possam ocorrer em espectros registrados para novas bateladas da reação química. Esta componente pode ser uma função matemática que modela a variação da linha base, ou o espectro de um novo componente que não estava presente nas medidas espectroscópicas da batelada de referência.

No trabalho original de Haaland,<sup>64</sup> o método PACLS foi desenvolvido para resolver os problemas associados com a previsão a partir de modelos de calibração multivariada. Neste trabalho de tese, será demonstrado como é possível associar o método PACLS aos métodos de MCR. O método MCR é usado para obter um modelo que descreve a variação da concentração ao longo do tempo dos compostos, tendo como única informação os espectros registrados no processo químico, enquanto que o método PACLS é usado para adaptar o modelo quando este sofre uma alteração de uma batelada para outra. A única exigência para utilizar o método PACLS é que seja possível obter o sinal que representa a variação das medidas de uma batelada para outra.

### **3.2 RESOLUÇÃO MULTIVARIADA DE CURVAS PARA A ANÁLISE DE DADOS ESPECTROSCÓPICOS COLETADOS EM REAÇÕES QUÍMICAS**

O primeiro trabalho associado ao MCR para a resolução de espectros foi publicado por Lawton e Sylvestre.<sup>65</sup> Estes pesquisadores mostraram que é possível resolver os

espectros de uma mistura binária partindo dos parâmetros do modelo bilinear ajustado com o método de Análise em Componentes Principais (PCA). Eles denominaram esta técnica de “Self Modeling Curve Resolution” e a usaram para resolver os espectros de uma mistura binária de pigmentos usados para a produção de filmes fotográficos. Esta técnica também foi utilizada pelo próprio Lawton, para investigar o mecanismo das reações químicas monitoradas por espectroscopia e determinar as constantes de velocidade das reações.<sup>66</sup>

O monitoramento espectroscópico em tempo real de uma reação química conduzida em bateladas gera uma matriz de espectros  $\mathbf{X}$ , com os respectivos modos de monitoramento tempo x comprimento de onda. Assumindo que a relação entre as medidas espectroscópicas e a variação da concentração dos compostos químicos é linear, então a matriz  $\mathbf{X}$  pode ser decomposta no modelo bilinear abaixo:

$$\mathbf{X} = \mathbf{a}_1 \mathbf{b}_1^T + \mathbf{a}_2 \mathbf{b}_2^T + \dots + \mathbf{a}_R \mathbf{b}_R^T + \mathbf{E} \quad 3.1$$

Os vetores  $\mathbf{a}_r$  e  $\mathbf{b}_r$  são, respectivamente, estimativas para a variação da concentração em função do tempo e dos espectros “puros” das espécies que podem ser detectadas pelo método espectroscópico utilizado.

Os métodos de MCR são muito usados para resolver os sinais de substâncias intermediárias formadas na reação.<sup>67</sup> Um intermediário da reação tem um tempo de vida bastante curto e, freqüentemente, não pode ser isolado do meio para ser quantificado. Além disso, os métodos de MCR são usados para investigar o mecanismo de reações químicas através da análise das curvas de variação das concentrações em função do tempo resolvidas no modelo bilinear.<sup>68,69,70</sup>

Se as equações de velocidade da reação são conhecidas, estas podem ser inseridas nos métodos de MCR para se fazer estimativas das constantes de velocidade das equações.<sup>71,72</sup> Esta metodologia pode ser implementada usando as funções analíticas ou o sistema de equações diferenciais<sup>69,73</sup> que descrevem uma reação química. As equações diferenciais são usadas nos casos em que o mecanismo de reação é muito complexo e não é possível obter uma solução analítica (as funções).

A decomposição em dois modos não é limitada à análise de uma única matriz  $\mathbf{X}$ , visto que um conjunto de matrizes  $\mathbf{X}_1, \dots, \mathbf{X}_2, \dots, \mathbf{X}_K$  obtidas a partir do monitoramento de bateladas consecutivas, podem ser organizadas em arranjos como os apresentados nas equações 3.2 e 3.3.

$$\mathbf{X}_{I \times K \times J} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_K \end{bmatrix} = \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \\ \vdots \\ \mathbf{A}_K \end{bmatrix} \mathbf{B}^T \quad 3.2$$

$$\mathbf{X}_{I \times J \times K} = [\mathbf{X}_1 \quad \mathbf{X}_2 \quad \dots \quad \mathbf{X}_K] = \mathbf{A} [\mathbf{B}_1 \quad \mathbf{B}_2 \quad \dots \quad \mathbf{B}_K]^T \quad 3.3$$

O arranjo da equação 3.2 assume que os espectros de cada espécie química são os mesmos nas diferentes bateladas, mas que variação das concentrações em função do tempo podem ser diferentes. Por outro lado, o arranjo apresentado da equação 3.3 assume que a variação da concentração em função do tempo de cada espécie é igual em todas as bateladas, enquanto que os espectros são diferentes.

O arranjo apresentado na equação 3.2 tem sido explorado para o desenvolvimento de metodologias de caracterização de reações químicas em processos industriais.<sup>57,58</sup> A solução obtida com este arranjo é um conjunto de matrizes  $\mathbf{A}_K$  que descrevem a variação da concentração de cada composto químico ao longo do tempo em cada batelada. Uma característica importante da solução obtida com o modelo 3.2 é que a escala de intensidade entre as matrizes é a mesma, isto significa que solução obtida para a batelada  $\mathbf{A}_1$  pode ser comparada diretamente com a solução da batelada  $\mathbf{A}_2$ . No monitoramento de processos industriais não há garantias de que os espectros das substâncias sejam iguais em todas as bateladas especialmente para os processos sujeitos a variações de temperatura, pressão e pH. Para estas situações o modelo da equação 3.2 não deve ser usado.

Quando há dois modos de monitoramento paralelos para as reações químicas, é possível organizar os dados em um cubo e fazer a resolução multivariada com os métodos de decomposição trilinear.<sup>74</sup> Estes casos são limitados as reações químicas com

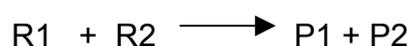
mecanismo de primeira ordem, porque o perfil da curva de variação da concentração em função do tempo de cada composto químico que participa da reação é igual em todos os experimentos consecutivos, independente da concentração inicial dos reagentes.

### 3.2.1 O problema da deficiência de posto.

Há uma série de questões que devem ser analisadas quando os métodos de MCR são usados para a modelagem de uma reação química com um modelo bilinear. Além da “ambiguidade da solução”, um outro problema importante é a deficiência de posto.<sup>75,76</sup>

A determinação do número de componentes de um modelo bilinear geralmente é feita pela determinação do posto da matriz  $X$  com o método de Decomposição em Valores Singulares (Material suplementar: Parte B; B.1), e esta análise é feita antes de ajustar o modelo com o método de MCR. Assumindo que todas as componentes do modelo bilinear estejam associadas apenas com a variação da concentração dos compostos químicos, é possível que na etapa de determinação do posto, o número de componentes detectados seja inferior ao número de espécies que realmente contribuem para as medidas do espectro. Isto ocorre por exemplo, quando duas espécies químicas apresentam espectros “puros” iguais, neste caso existem duas espécies, mas o posto será determinado matematicamente como sendo igual a um.

A deficiência de posto é um problema inerente a muitos tipos de reações químicas devido às próprias leis que regem o balanço de massa destes sistemas. Por exemplo, considere uma reação com mecanismo de segunda ordem da equação química abaixo:



Considerando que todas as espécies são detectadas pelo método espectroscópico usado, teremos uma matriz de espectros  $X$  formada pela contribuição de quatro espécies químicas. Porém o posto desta matriz será igual a dois, porque R1 e R2 apresentam a mesma curva de variação da concentração em função do tempo, e o mesmo ocorre para P1 e P2. Sendo assim, a variação da absorbância em função do tempo pode ser descrita por dois vetores linearmente independentes que correspondem a combinações lineares

dos sinais de R1 e R2, P1 e P2. Conclusões semelhantes podem ser verificadas para outros tipos de mecanismos de reações químicas.<sup>77</sup>

Existem outros fatores que podem levar ao problema da deficiência do posto, como a operação de centrar na média as colunas da matriz  $\mathbf{X}$  para alguns tipos de reações químicas.<sup>78</sup>

Um das formas de resolver o problema da deficiência de posto é preparar diferentes bateladas da reação química alterando a concentração inicial dos reagentes e agrupando as matrizes  $\mathbf{X}_1, \dots, \mathbf{X}_2, \dots, \mathbf{X}_K$  em um dos arranjos mostrados nas equações 3.2 e 3.3. A análise do posto dos respectivos arranjos pode indicar o número correto de espécies químicas presentes na reação. Para maiores detalhes sobre este assunto pode-se consultar os trabalhos de Garrido<sup>79</sup> e Saurina.<sup>80</sup>

### 3.3 METODOLOGIA PARA O MONITORAMENTO DE REAÇÕES QUÍMICAS COM MODELOS EMPÍRICOS ADAPTATIVOS

A metodologia proposta neste trabalho pode ser usada para monitorar ou caracterizar a composição química de uma reação a partir da análise dos dados espectroscópicos registrados em tempo real durante a operação de um reator. A primeira etapa da metodologia consiste em obter um modelo que relaciona a variação dos dados de espectroscopia com a composição química da reação. A segunda etapa, consiste em usar o modelo ajustado na primeira etapa para prever o comportamento de novas bateladas da reação.

O modelo bilinear é ajustado a partir de uma batelada da reação escolhida como referência. A partir desta obtêm-se o modelo abaixo:

$$\mathbf{X}_{\text{ref}} = \mathbf{A}_{\text{ref}} \mathbf{B}_{\text{ref}}^T$$

**3.4**

Cada linha da matriz  $\mathbf{X}_{\text{ref}}$  corresponde a um espectro registrado no instante  $i$  e cada coluna corresponde a um comprimento de onda  $j$ . As matrizes  $\mathbf{A}_{\text{ref}}$  e  $\mathbf{B}_{\text{ref}}$  são respectivamente estimativas da variação das concentrações em função do tempo e dos espectros puros dos compostos químicos que contribuem para as medidas em  $\mathbf{X}_{\text{ref}}$ .

Há três maneiras possíveis de obter o modelo da equação 3.4. Se os espectros de todas as substâncias que participam da reação são conhecidos, então  $\mathbf{A}_{\text{ref}}$  pode ser determinado da seguinte forma:

$$\mathbf{A}_{\text{ref}} = \mathbf{X}_{\text{ref}} \mathbf{B}_{\text{ref}} (\mathbf{B}_{\text{ref}}^T \mathbf{B}_{\text{ref}})^{-1} \quad 3.5$$

por outro lado se os espectros não são conhecidos, mas a concentração de cada espécie que participa da reação pode ser determinada em intervalos regulares de tempo através de um ensaio analítico, então o modelo pode ser resolvido da seguinte forma:

$$\mathbf{B}_{\text{ref}}^T = (\mathbf{A}_{\text{ref}}^T \mathbf{A}_{\text{ref}})^{-1} \mathbf{A}_{\text{ref}}^T \mathbf{X}_{\text{ref}} \quad 3.6$$

Os dois procedimentos apresentados acima correspondem a uma calibração multivariada. Porém, se nem os espectros e nem as concentrações de todas as espécies são conhecidas, então o modelo da equação 3.4 tem que ser obtido através dos métodos de resolução multivariada de curvas, a fim de determinar ambas as matrizes  $\mathbf{A}_{\text{ref}}$  e  $\mathbf{B}_{\text{ref}}$ , a partir da matriz  $\mathbf{X}_{\text{ref}}$ .

Se mais do que uma batelada de referência estiver disponível, o seguinte modelo pode ser ajustado:

$$\begin{bmatrix} \mathbf{X}_{\text{ref1}} \\ \mathbf{X}_{\text{ref2}} \\ \vdots \\ \mathbf{X}_{\text{refk}} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{\text{ref1}} \\ \mathbf{A}_{\text{ref2}} \\ \vdots \\ \mathbf{A}_{\text{refk}} \end{bmatrix} \mathbf{B}_{\text{ref}}^T \quad 3.7$$

A estrutura do modelo definido na equação 3.7 assume que o espectro de cada substância é o mesmo em cada batelada da reação. Este modelo não pode ser usado quando há algum agente capaz de mudar a forma dos espectros das substâncias de uma batelada para outra.

Quando um método de MCR é usado para análise dos dados é necessário prestar atenção na solução obtida. Devido às ambigüidades associadas à solução destes

métodos, pode ser que os elementos de  $\mathbf{A}_{\text{ref}}$  e  $\mathbf{B}_{\text{ref}}$  não correspondam respectivamente às verdadeiras curvas de variação da concentração em função do tempo e dos espectros puros das espécies.

Há uma ambigüidade na escala de intensidade das concentrações resolvidas conforme foi discutido no tópico 2.4, e se a concentração inicial dos reagentes não for conhecida então não há como definir qual a escala real de concentração dos compostos na reação. A ambigüidade rotacional é um problema mais sério, visto que a solução obtida pelo método MCR pode ser na verdade uma combinação linear desconhecida dos verdadeiros sinais de cada espécie. Esta característica dos métodos MCR pode tornar a análise bastante confusa. Por exemplo, se o algoritmo MCR-ALS for iniciado de diferentes formas, podem ser obtidas diferentes soluções com um mesmo ajuste, todas corretas de acordo com o critério matemático usado.

O intervalo de soluções possíveis poderá ser pequeno se forem usadas as restrições de cálculo, mas geralmente permanecerá uma dúvida sobre qual das soluções é a mais próxima dos verdadeiros sinais. A metodologia proposta neste trabalho necessita de um modelo de referência e os parâmetros deste modelo, devem de preferência, corresponder aos verdadeiros sinais das espécies. Se isto não for possível então o analista deve escolher uma das soluções obtidas a partir do método MCR e usar sempre a mesma solução como o modelo de referência para fazer comparações com outras bateladas da reação que serão processadas no futuro. A escolha do modelo de referência é subjetiva quando a ambigüidade de rotação não pode ser evitada e deve levar em conta o conhecimento que o analista tem sobre a reação em estudo.

Para prever o comportamento da variação da concentração dos compostos químicos, nas novas bateladas, a seguinte equação deve ser resolvida:

$$\mathbf{A}_{\text{novo}} = \mathbf{X}_{\text{novo}} \mathbf{B}_{\text{ref}} \left( \mathbf{B}_{\text{ref}}^T \mathbf{B}_{\text{ref}} \right)^{-1} \quad 3.8$$

onde  $\mathbf{X}_{\text{novo}}$  é a matriz dos dados espectroscópicos coletados para a nova batelada que não foi usada para a modelagem da reação química. Os parâmetros da matriz  $\mathbf{A}_{\text{novo}}$  descrevem como varia a concentração dos compostos na nova batelada produzida no reator.

Os parâmetros da matriz  $\mathbf{A}_{\text{nov}}o$  têm a mesma escala de intensidade dos parâmetros da matriz de referência  $\mathbf{A}_{\text{ref}}$ . Portanto, os parâmetros destas duas matrizes podem ser comparados em um gráfico que mostra a variação da concentração dos compostos químicos da reação em função do tempo, para a batelada de referência e para a nova batelada. Este gráfico é bastante instrutivo porque mostra o estado da composição química da nova batelada da reação.

A equação 3.8 pode ser usada para o monitoramento em tempo real pois no instante  $i$  podemos fazer uma previsão da composição da reação química no processo usando os espectros registrados até este instante.

Um agravante surge quando os espectros registrados para a nova batelada apresentam alguma componente que não estava presente nos espectros registrados para a reação de referência ou se ocorreu alguma alteração no espectro dos compostos da batelada de referência. Neste caso, a matriz  $\mathbf{B}_{\text{ref}}$  determinada na primeira etapa não pode ser usada mais para prever o comportamento das novas bateladas da reação.

Uma forma de adaptar o modelo às variações que este possa sofrer é usar o método PACLS desenvolvido por Haaland.<sup>64</sup> Este método foi desenvolvido para adaptar os coeficientes de regressão linear do modelo de calibração multivariada baseado na Lei de Beer que para os quimiometristas é denominado de “*Classical Least Squares (CLS)*”. Neste trabalho de tese, propõe-se que o método PACLS seja usado para adaptar os coeficientes de um modelo que foi derivado do método MCR.

O modelo ajustado com o método CLS é o mesmo do MCR, equação 3.4, mas o método CLS é usado em um procedimento de calibração e para isto é necessário dispor das matrizes  $\mathbf{X}_{\text{ref}}$  e  $\mathbf{C}_{\text{ref}}$  ou  $\mathbf{X}_{\text{ref}}$  e  $\mathbf{A}_{\text{ref}}$ . Mas se apenas a matriz  $\mathbf{X}_{\text{ref}}$  estiver disponível então é necessário resolver o modelo com um método de MCR.

A idéia do método PACLS é estender a matriz  $\mathbf{B}_{\text{ref}}$  a fim de incorporar na etapa de previsão, as componentes do espectro que não fazem parte da batelada de referência. Estender a matriz  $\mathbf{B}_{\text{ref}}$ , significa incluir nesta matriz os parâmetros que descrevem a componente que falta.

A figura 3.1 mostra um esquema de como é possível estender a matriz  $\mathbf{B}_{\text{ref}}$  para corrigir uma mudança na linha base do tipo “offset”. A coluna de valores unitários 1, tem a forma das alterações que ocorrem na linha base e por isto modelam exatamente a

componente que pode mudar de uma batelada para outra. Caso a linha base seja uma reta com uma inclinação ou uma curva semelhante à de um polinômio, estas linhas também podem ser modeladas colocando na matriz  $\mathbf{B}_{\text{ref}}$  componentes que simulam uma linha reta ou a curva de um polinômio.<sup>81</sup>

$$\mathbf{B}_{\text{ref}} = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \\ b_{41} & b_{42} & b_{43} \end{bmatrix} \xrightarrow{\text{Estendendo a matriz}} \mathbf{B}_{\text{est}} = \begin{bmatrix} b_{11} & b_{12} & b_{13} & 1 \\ b_{21} & b_{22} & b_{23} & 1 \\ b_{31} & b_{32} & b_{33} & 1 \\ b_{41} & b_{42} & b_{43} & 1 \end{bmatrix}$$

**Figura 3.1.** Método para estender a matriz  $\mathbf{B}_{\text{ref}}$ . A matriz estendida  $\mathbf{B}_{\text{est}}$  representada acima inclui uma componente que modela uma mudança constante na linha de base.

Se surgir uma nova espécie química na matriz  $\mathbf{X}_{\text{novo}}$ , que não estava presente na matriz  $\mathbf{X}_{\text{ref}}$ , o espectro desta espécie pode ser adicionado na matriz  $\mathbf{B}_{\text{ref}}$  como uma nova coluna, para incluir a contribuição desta nova substância. Nos casos em que o espectro de uma determinada substância é alterado por ação da temperatura, pode-se obter experimentalmente a forma como este espectro é alterado e então adicionar na matriz  $\mathbf{B}_{\text{ref}}$  esta componente que representa a variação do espectro devido a ação da temperatura.<sup>64</sup> O método PACLS é bastante versátil, e não há restrição para o número de componentes que podem ser adicionadas simultaneamente na matriz  $\mathbf{B}_{\text{ref}}$ . Entretanto para que funcione corretamente é preciso saber que tipo de alteração ocorreu e obter o espectro que modela a variação ocorrida.

Após estender a matriz  $\mathbf{B}_{\text{ref}}$ , esta pode ser usada para prever o comportamento das novas bateladas utilizando a equação 3.9.

$$\mathbf{A}_{\text{nova}} = \mathbf{X}_{\text{nova}} \mathbf{B}_{\text{est}} \left( \mathbf{B}_{\text{est}}^T \mathbf{B}_{\text{est}} \right)^{-1} \quad \mathbf{3.9}$$

### 3.4 PARTE EXPERIMENTAL

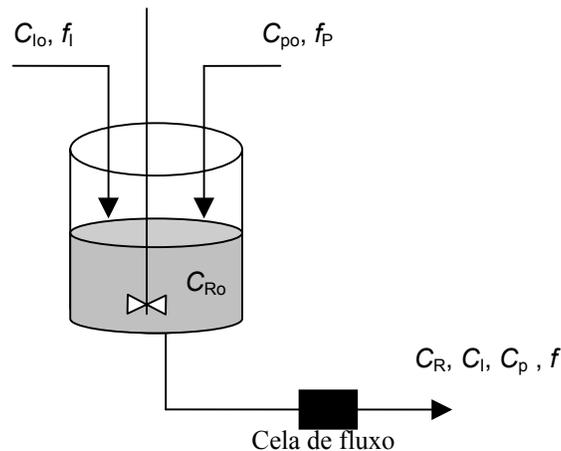
Este trabalho foi desenvolvido em conjunto com o grupo de Biotecnologia e Quimiometria do Instituto Superior Técnico de Lisboa, Portugal, sob orientação do prof. Dr. José Cardoso de Menezes.

### 3.4.1 Sistemas dinâmicos em um tanque agitado contínuo

A metodologia proposta neste trabalho foi aplicada para a análise de um conjunto de experimentos realizados em um equipamento desenvolvido para operar sistemas químicos dinâmicos. O equipamento denominado de tanque agitado contínuo é formado por um tanque de vidro, um agitador mecânico e por válvulas que introduzem soluções de compostos químicos e retiram a mistura destes em regime de operação contínuo e estacionário, conforme mostra a figura 3.2.

As válvulas que operam as vazões de entrada  $f_i$ ,  $f_p$  e a vazão de saída  $f$  são controladas por um programa de computador escrito no software "LABVIEW". As variáveis  $C_{R0}$ ,  $C_{I0}$  e  $C_{P0}$ , indicadas na figura 3.2, são as respectivas concentrações das substâncias R, I e P das soluções adicionadas no tanque e  $C_R$ ,  $C_I$  e  $C_P$  são as concentrações das mesmas substâncias na mistura.

Na linha de saída do tanque há uma cela para medidas em fluxo por reflectância total atenuada. Para este trabalho foi usado um espectrofotômetro de infravermelho médio FTLA 2000 da ABB Bomem. O acessório para medidas de reflectância atenuada em fluxo (com cristal de ZnSe) foi obtido da MIRacle, Pike Technologies, Inc.



**Figura 3.2.** Tanque agitado contínuo.

Através do tanque agitado contínuo podemos misturar soluções de compostos químicos diferentes, fazendo com que a concentração destes na mistura varie ao longo do tempo e que esta variação siga uma função matemática previamente definida. Por exemplo, podemos fazer uma mistura dos compostos R, I e P dentro do tanque, de modo

que as concentrações  $C_R$ ,  $C_I$  e  $C_P$  destes compostos variem em função do tempo de acordo com o comportamento descrito pelas funções 3.10, 3.11 e 3.12, sendo que  $k_1$  e  $k_2$  são constantes previamente definidas e a variável  $t$  é o tempo.

$$C_R = C_{R0} e^{-k_1 t} \quad \mathbf{3.10}$$

$$C_I = \frac{C_{R0} k_1}{k_2 - k_1} (e^{-k_1 t} - e^{-k_2 t}) \quad \mathbf{3.11}$$

$$C_P = C_{R0} \left[ 1 + \frac{1}{k_2 - k_1} (k_1 e^{-k_1 t} - k_2 e^{-k_2 t}) \right] \quad \mathbf{3.12}$$

Para produzir a mistura descrita pelas equações 3.10, 3.11 e 3.12 é necessário programar as válvulas para que estas mudem ao longo do tempo as vazões  $f_I$ ,  $f_P$  e  $f$  de acordo com as equações 3.13, 3.14 e 3.15, deduzidas após introduzir o sistema de equações 3.10, 3.11 e 3.12 nas equações do balanço de massa do tanque agitado contínuo que opera em regime estacionário.

$$f = k_1 V \quad \mathbf{3.13}$$

$$f_I = \frac{V k_1 C_{R0}}{C_{I0}} e^{-k_2 t} \quad \mathbf{3.14}$$

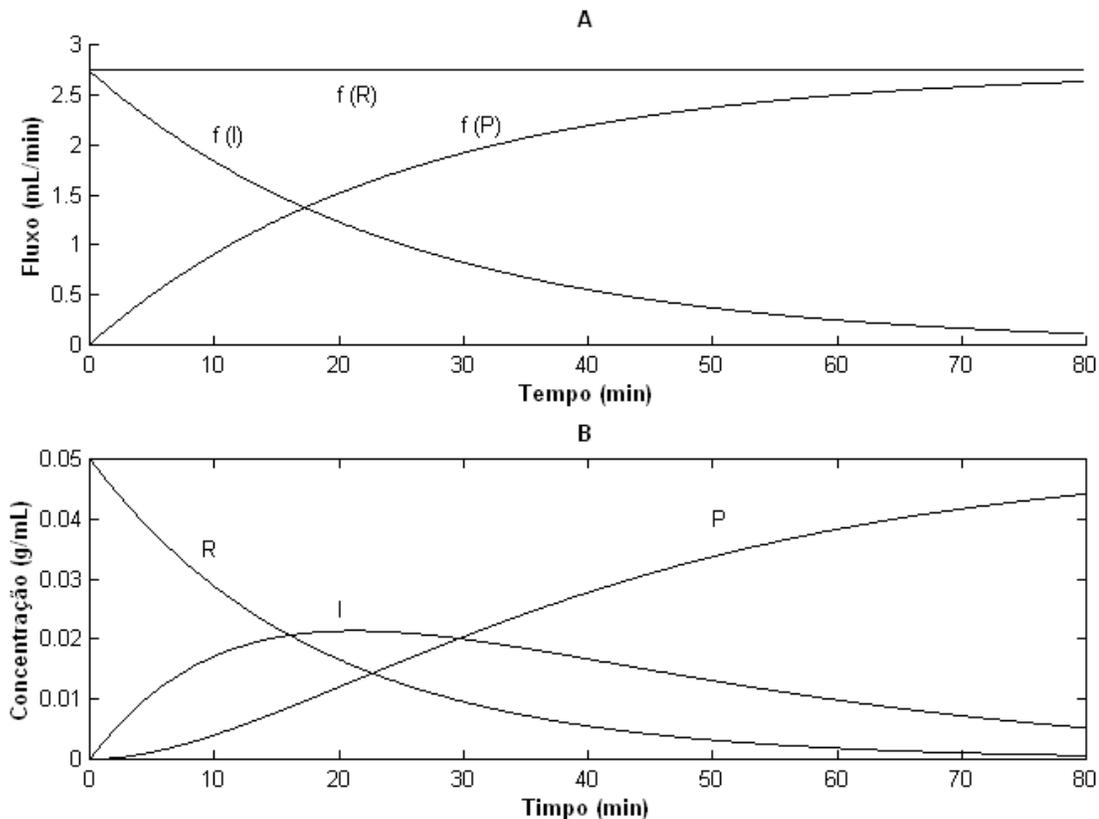
$$f_P = \frac{V k_1 C_{R0}}{C_{P0}} (1 - e^{-k_2 t}) \quad \mathbf{3.15}$$

No sistema de equações 3.13, 3.14 e 3.15, a variável  $V$  é o volume de líquido dentro do tanque que é mantido constante no regime de operação estacionário. As concentrações  $C_{P0}$ ,  $C_{R0}$  e  $C_{I0}$  das soluções adicionadas no tanque devem ser iguais para que as concentrações  $C_R$ ,  $C_I$  e  $C_P$  sigam a variação descrita pelas equações 3.10, 3.11 e 3.12, ao longo do tempo.

A figura 3.3A mostra como as vazões  $f_P$ ,  $f_I$ , e  $f$  devem variar ao longo do tempo para produzir dentro do tanque uma mistura dos compostos R, I e P com as variações de

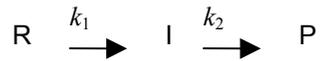
concentrações com o tempo iguais as mostradas na figura 3.3B, que são descritas pelas equações 3.10, 3.11 e 3.12 quando  $k_1 = 0,05 \text{ s}^{-1}$ ,  $k_2 = 0,04 \text{ s}^{-1}$  e  $C_{R0} = C_{I0} = C_{P0} = 50 \text{ g/L}$ .

A mistura de composição variável mostrada na figura 3.3B é produzida ao adicionar inicialmente 50 mL de uma solução aquosa  $50 \text{ g L}^{-1}$  do composto R dentro do tanque e ir adicionando ao longo de 80 min as soluções aquosas  $50 \text{ g L}^{-1}$  dos compostos I e P através das válvulas  $f_p$  e  $f_i$ , programadas para variar a vazão de acordo com as equações 3.13, 3.14 e 3.15. Na saída do tanque, há um espectrofotômetro “FT-MIR” que monitora a composição química da mistura que sai do tanque com vazão  $f$  expressa pela equação 3.13. O “atraso” entre o instante  $i$  de uma determinada composição dentro no tanque e o instante  $i+\Delta i$  no momento da medida, é determinado a fim de corrigir a escala de tempo.



**Figura 3.3.** A) Vazão de adição dos compostos em função do tempo para as válvulas  $f_R$ ,  $f_I$  e  $f_P$ . B) variação da concentração em função do tempo para os compostos R, I e P, de acordo com as equações 3.10 a 3.12 usando constantes  $k_1 = 0,05 \text{ s}^{-1}$  e  $k_2 = 0,045 \text{ s}^{-1}$ .

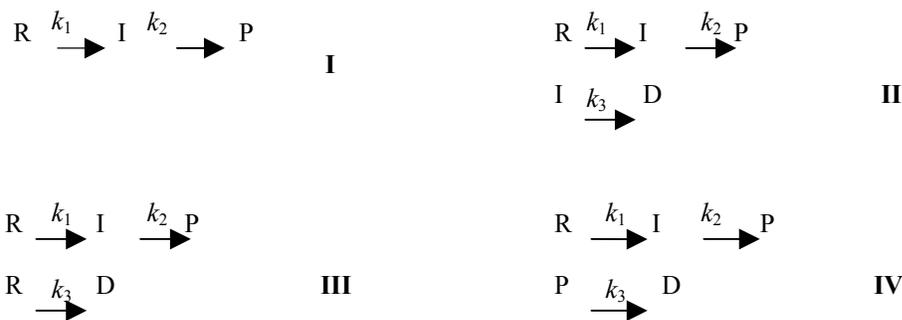
Note que as equações 3.10, 3.11 e 3.12 correspondem as mesmas equações que descrevem a variação da concentração com o tempo de compostos químicos que participam de uma reação química em série e de primeira ordem representada genericamente abaixo:



No caso específico de uma reação química,  $k_1$  e  $k_2$  são constantes de velocidade das reações químicas. Sendo assim, através do tanque agitado contínuo podemos produzir misturas com curvas de variação das concentrações com o tempo, iguais as que ocorrem nas reações químicas.

### 3.4.2 Experimentos com misturas em tanque agitado contínuo

A figura 3.4 mostra um esquema de quatro misturas que foram feitas no tanque agitado contínuo apresentado na figura 3.2 e que foram usados para aplicar a metodologia proposta neste trabalho. A tabela 3.1, mostra os valores das constantes  $k_1$ ,  $k_2$  e  $k_3$  usadas para produzir cada mistura. Note, que os experimentos I, II, III e IV não são reações químicas como sugere o esquema apresentado na figura 3.4, mas sim, misturas de compostos em que a composição varia ao longo do tempo da mesma forma que varia os compostos em uma reação química de primeira ordem com as etapas apresentas nos experimentos I, II, III e IV da figura 3.4.



**Figura 3.4.** Experimentos com misturas produzidos no tanque agitado contínuo.

Nos esquemas apresentados na figura 3.4 a variação de concentração com o tempo do composto R na mistura produzida dentro do tanque é similar a de um composto reagente em uma reação de primeira ordem e em série, da mesma forma que a variação das concentrações dos compostos I, P e D nas misturas são similares a de um composto intermediário (I), produto (P) e produto lateral (D) de reações químicas de primeira ordem em série.

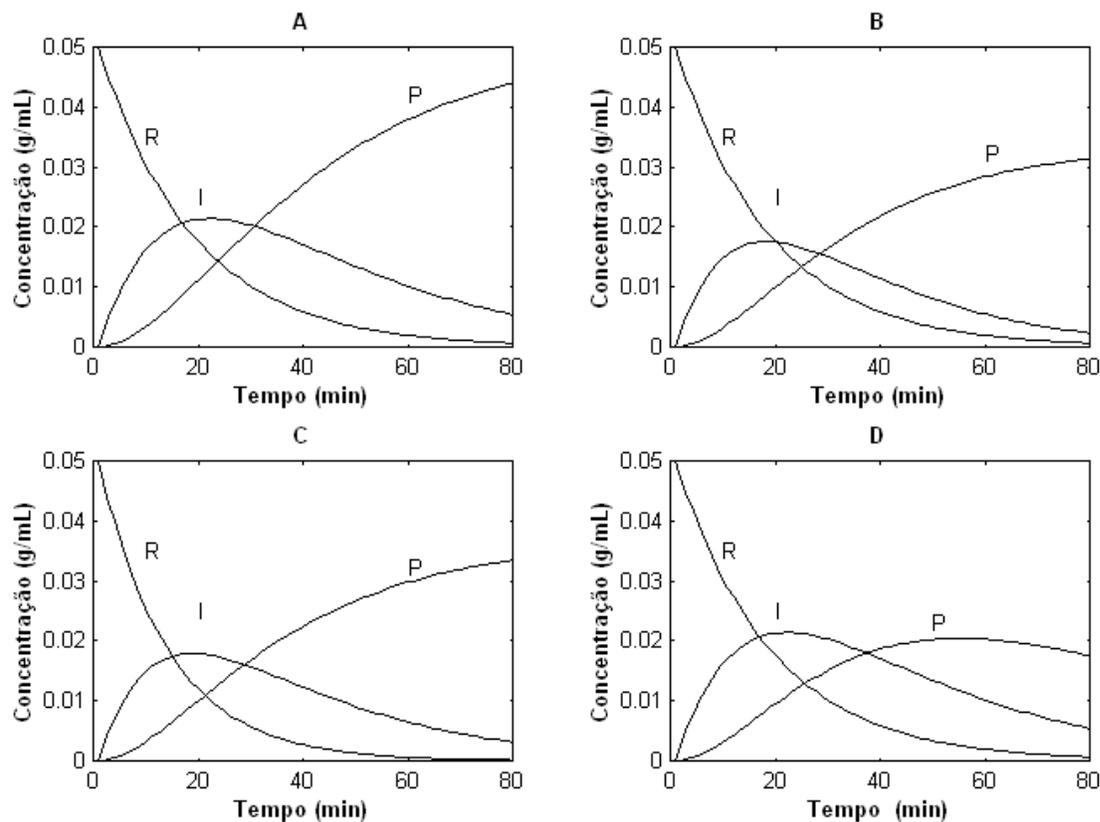
Para fazer os experimentos de misturas da figura 3.4 foram adicionadas soluções aquosas de frutose (R), sacarose (I), glicose (P) com concentrações de 50 g L<sup>-1</sup> e água (D). A figura 3.5 mostra a variação da concentração com o tempo de cada composto químico R, I e P em cada uma das misturas dos experimentos I, II, III e IV.

**Tabela 3.1.** Valores das constantes usadas para executar os experimentos da figura 3.4.

Esquema/Constantes	$k_1$ (s <sup>-1</sup> )	$k_2$ (s <sup>-1</sup> )	$k_3$ (s <sup>-1</sup> )
I	0,055	0,04	0,00
II	0,055	0,04	0,02
III	0,055	0,04	0,02
IV	0,055	0,04	0,02

O experimento I foi escolhido como batelada de referência e os espectros de infravermelho coletados para este experimento foram organizados na matriz  $\mathbf{X}_{ref}$ , equação 3.4. Os outros experimentos simulam o aparecimento de um produto lateral D na mistura que surge da degradação de um dos compostos R, I e P. O produto D não é detectado na região do infravermelho usada neste trabalho.

Embora a metodologia proposta neste trabalho seja testada para a análise de misturas com composição variável ao longo do tempo, esta metodologia pode ser usada diretamente para a análise de espectros registrados em reações químicas visto que o comportamento da composição química ao longo do tempo é similar.



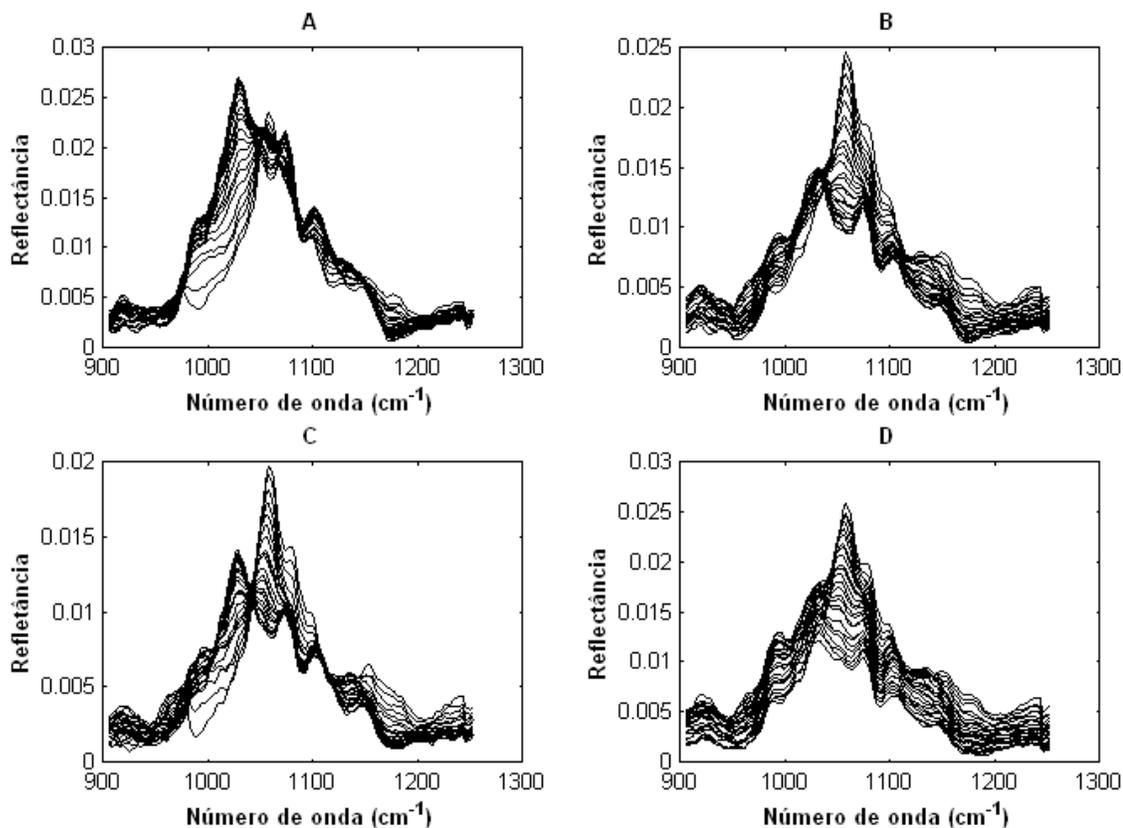
**Figura 3.5.** Variação da concentração com o tempo dos compostos químicos para as misturas apresentadas na figura 3.4. A) experimento I; B) experimento II; C) experimento III; D) experimento IV.

### 3.5 RESULTADOS E DISCUSSÃO

A figura 3.6 mostra os espectros de infravermelho registrados ao longo do tempo para os experimentos I, II, III e IV. Mesmo que pequena, há uma variação da linha de base dos espectros de acordo com um deslocamento vertical constante (“offset”) que é mais pronunciado no experimento IV. Os espectros registrados para a batelada de referência (experimento I) são aqueles mostrados na figura 3.6A. Estes foram organizados em uma matriz  $\mathbf{X}_{\text{ref}}$  de dimensões 312 X 181 para os respectivos modos de monitoramento tempo x comprimento de onda.

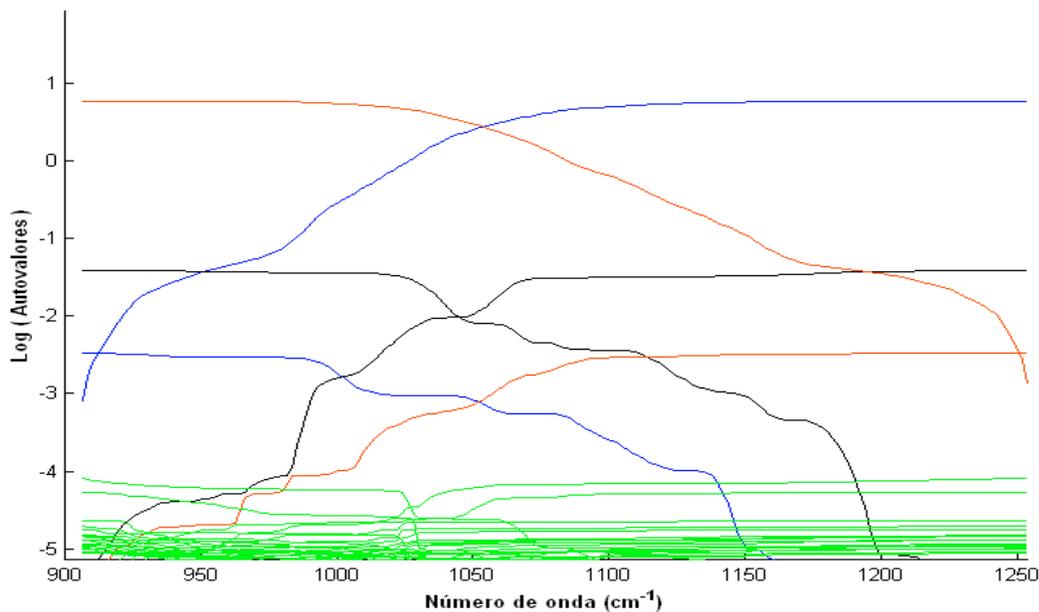
A análise da matriz de referência com o método EFA (tópico 2.3.1.1), nos sentidos direto e inverso da matriz  $\mathbf{X}_{\text{ref}}^T$  produz o resultado mostrado na figura 3.7. A evolução dos valores singulares para ambos os sentidos identifica pelo menos três valores singulares

significativos e estes apresentam um valor superior a - 4 no gráfico apresentado na figura 3.7.

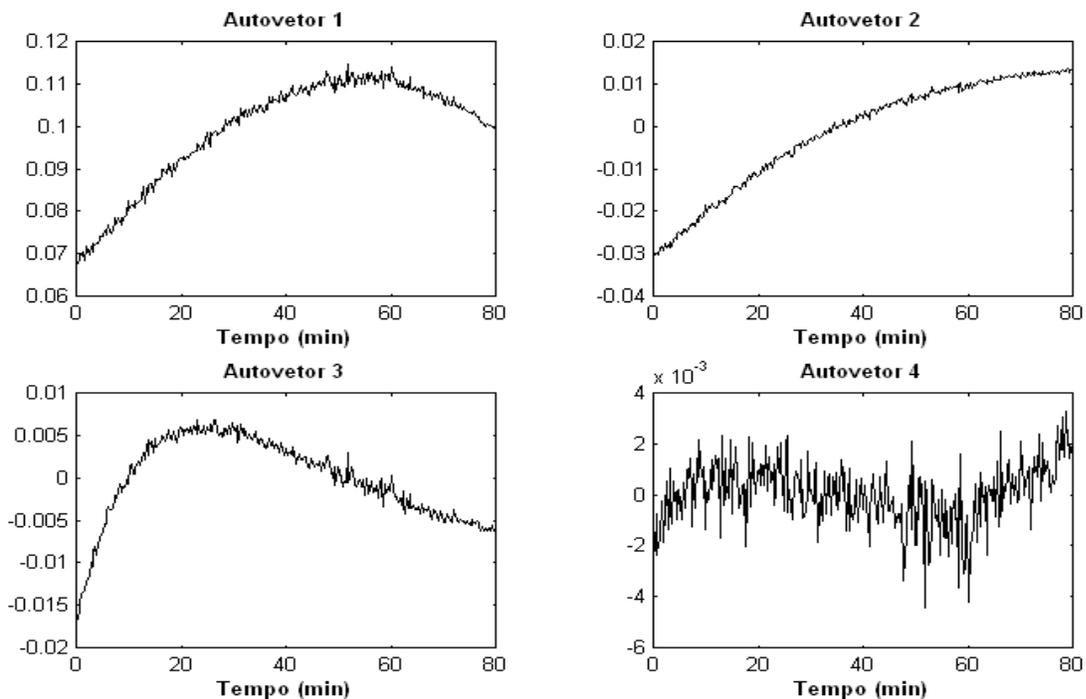


**Figura 3.6.** Espectros de infravermelho registrados para os experimentos: A) I; B) II; C) III; D) IV.

A figura 3.8, mostra os quatro primeiros autovetores da matriz  $\mathbf{U}$ , multiplicados pelos seus respectivos valores singulares, obtidos por SVD (Capítulo 2, material suplementar B.1) a partir da análise da matriz  $\mathbf{X}_{\text{ref}}$ . Os três primeiros autovetores apresentam uma estrutura bem definida e descrevem 99,90% da variação dos dados. O quarto autovetor não tem um comportamento estruturado e contribui apenas com 0,02% para a variação total. Como a variância da quarta componente é muito pequena comparada as das três primeiras componentes, não há razões para esperar que esta componente esteja associada a algum composto químico que contribui para as medidas de reflectância da batelada de referência.



**Figura 3.7.** Resultado da análise com o método EFA para a matriz  $\mathbf{X}_{\text{ref}}^T$ . “Janelas” para a (-) primeira, (-) segunda e (-) terceira componentes.



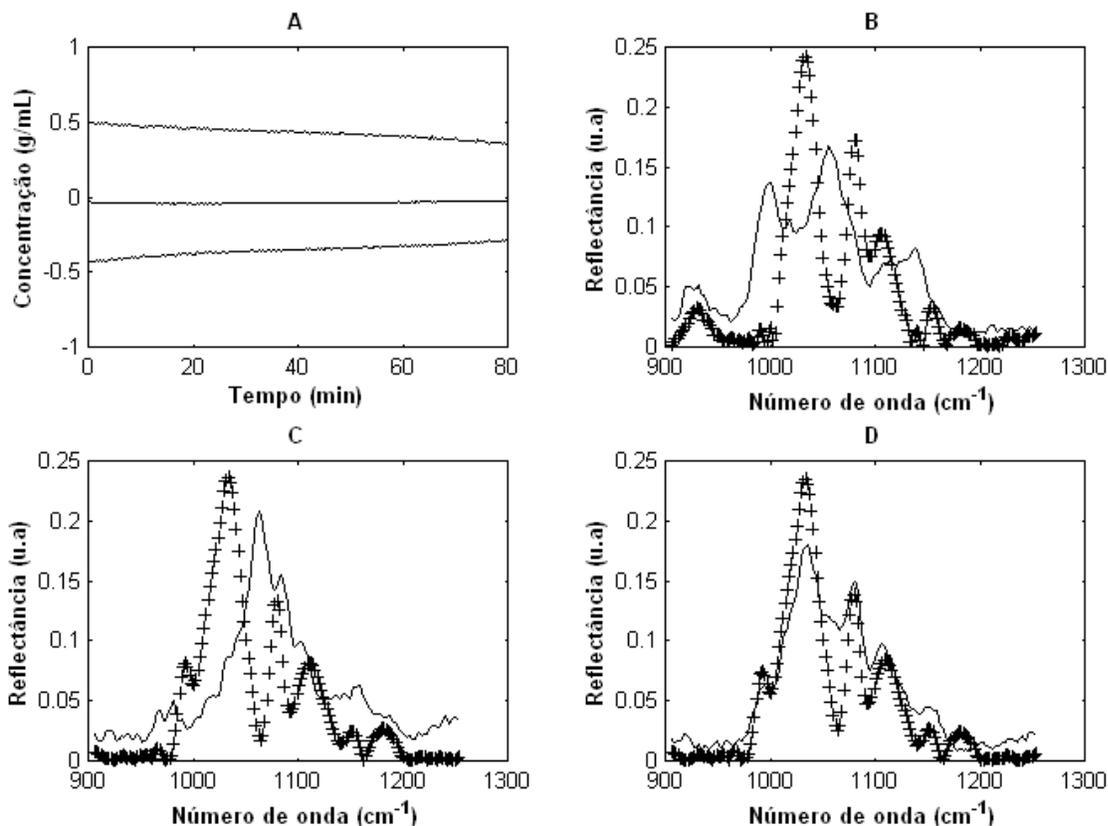
**Figura 3.8.** Autovetores  $\mathbf{U}$  obtidos por decomposição em valores singulares da matriz  $\mathbf{X}_{\text{ref}}$ . Os autovetores apresentados estão multiplicados pelos seus respectivos valores singulares.

Os autovetores da matriz  $\mathbf{U}$  estão associados às concentrações das espécies químicas que contribuem para as medidas dos espectros de reflectância, e os três primeiros vetores  $\mathbf{u}_1$ ,  $\mathbf{u}_2$  e  $\mathbf{u}_3$  que compõem esta matriz, correspondem a uma combinação linear das concentrações dos verdadeiros sinais  $\mathbf{c}_1$ ,  $\mathbf{c}_2$  e  $\mathbf{c}_3$ .<sup>65</sup> Embora os autovetores descrevam informações abstratas, é possível perceber um comportamento dinâmico para a variação da concentração das espécies, ao analisar o comportamento ao longo do tempo dos três primeiros autovetores  $\mathbf{u}_1$ ,  $\mathbf{u}_2$  e  $\mathbf{u}_3$  da figura 3.8.

Com base no mapa dos valores singulares da figura 3.7, foi feita uma tentativa de identificar “janelas” no espectro de infravermelho para os compostos químicos presentes na batelada de referência. A janela da primeira componente foi selecionada no intervalo entre 906 e 1136  $\text{cm}^{-1}$ , enquanto que as janelas das outras duas componentes foram selecionadas entre 970 e 1191, 1000 e 1251  $\text{cm}^{-1}$ .

As figuras 3.9B, C e D mostram as estimativas obtidas para os espectros dos compostos químicos nas misturas, resolvidos com o método EFA a partir das “janelas” indicadas acima. A figura 3.9A, mostra as concentrações previstas usando a equação 3.5 para a batelada de referência a partir dos espectros resolvidos com o método EFA. A metodologia proposta nesta tese assume que o analista não dispõe e não conhece os sinais verdadeiros dos compostos químicos. Porém, para os dados apresentados neste trabalho os verdadeiros sinais são conhecidos de modo que os espectros verdadeiros de cada composto químico também foram colocados nas figuras 3.9B, C e D para fazermos uma comparação com a solução do método EFA.

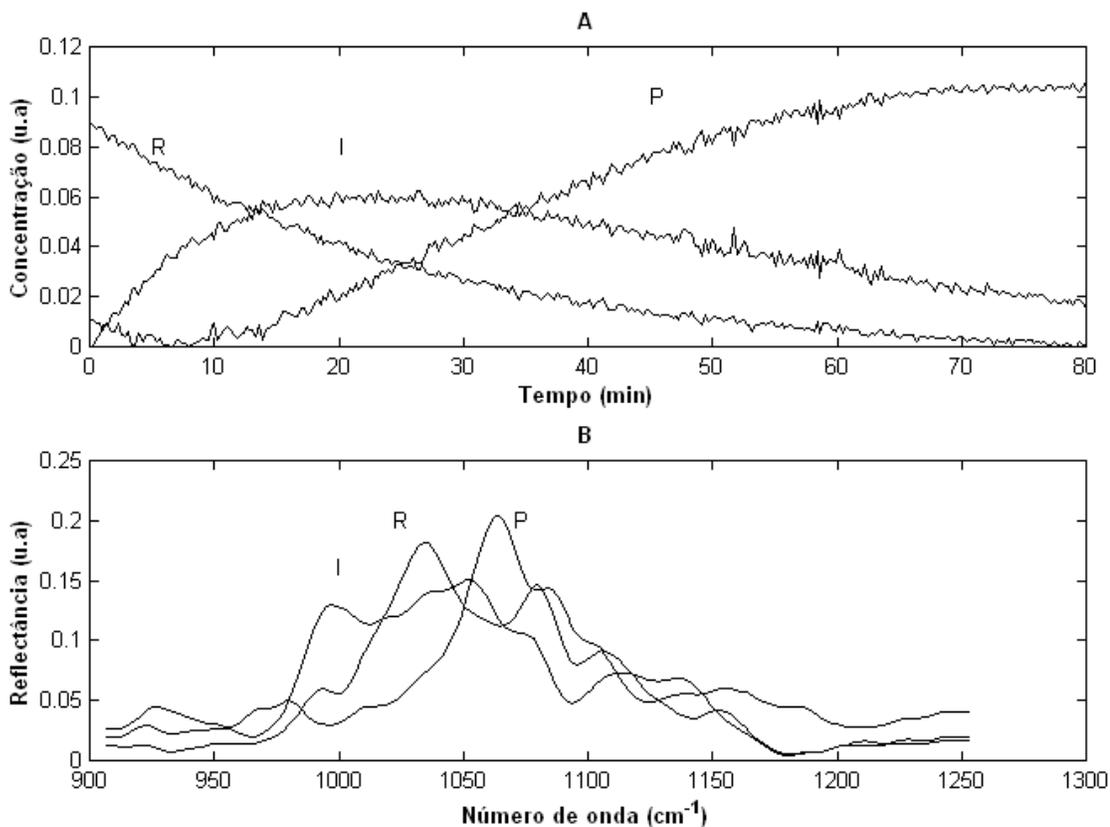
Os espectros resolvidos pelo método EFA não estão de acordo com os verdadeiros espectros dos compostos R, I e P usados para produzir as misturas de composição variável da figura 3.5. Da mesma forma, as concentrações previstas apresentam valores negativos contrário ao que se deve esperar de uma solução real para o qual as concentrações devem ter apenas valores positivos. Por isso, a solução obtida com o método EFA não foi considerada adequada.



**Figura 3.9.** Comparação dos sinais resolvidos com o método EFA (+) com os sinais verdadeiros de cada composto químico (-). A) Variação da concentração em função do tempo; B) espectro do composto P; C) espectro do composto R ; D) espectro do composto I.

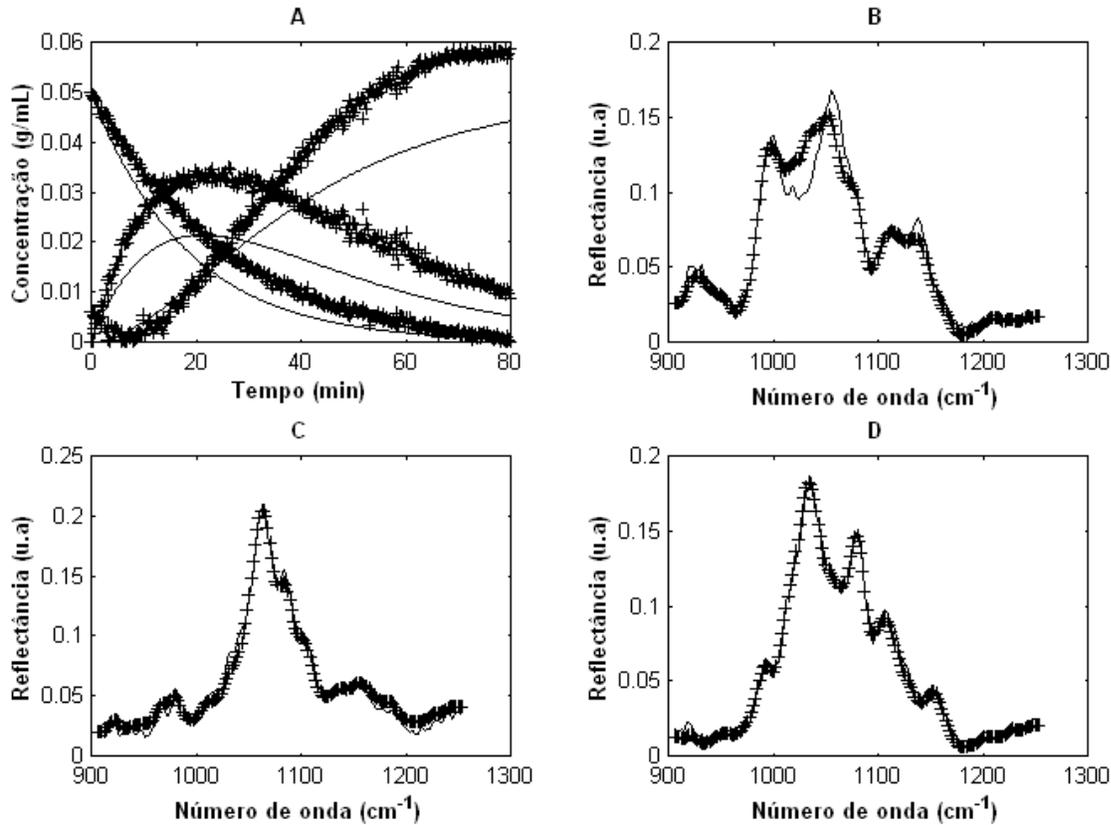
O método MCR-ALS foi usado para ajustar um modelo bilinear com três componentes. Este modelo foi calculado com restrições de não negatividade para ambas as matrizes  $\mathbf{A}_{\text{ref}}$  e  $\mathbf{B}_{\text{ref}}$ , e além da restrição de não negatividade, os parâmetros da matriz  $\mathbf{B}_{\text{ref}}$  foram normalizados para o comprimento unitário. Para iniciar o algoritmo ALS (tópico 2.3.2.1) foram usados valores aleatórios e o ajuste do modelo foi executado várias vezes sempre iniciando o cálculo com uma matriz  $\mathbf{A}_0$  diferente.

Verificou-se que várias soluções podem ser obtidas cada vez que um novo modelo é calculado, ficando evidente que há ambigüidade na solução. Entretanto, as soluções obtidas não foram muito diferentes da solução mostrada na figura 3.10, sugerindo que as restrições de não negatividade reduzem bastante o intervalo de soluções possíveis. Como não houve meios de obter uma única solução, foi escolhida a solução apresentada na figura 3.10 para formar as matrizes  $\mathbf{A}_{\text{ref}}$  e  $\mathbf{B}_{\text{ref}}$ .



**Figura 3.10.** Parâmetros do modelo MCR-ALS resolvidos para a matriz  $\mathbf{X}_{\text{ref}}$ . A) matriz  $\mathbf{A}_{\text{ref}}$ , B) matriz  $\mathbf{B}_{\text{ref}}$ .

A figura 3.11 compara as estimativas feitas com o método MCR e os sinais verdadeiros de cada composto químico. De um modo geral a matriz  $\mathbf{A}_{\text{ref}}$  representa muito bem o comportamento esperado para a variação da concentração em função do tempo de compostos químicos que participam de um sistema dinâmico, embora os perfis resolvidos não sejam perfeitamente iguais aos verdadeiros. Os espectros dos compostos R e I foram bem resolvidos, mas o espectro resolvido para o composto P apresentou um desvio em relação ao verdadeiro, conforme mostra a figura 3.11B para a região do espectro entre 1000 e 1080  $\text{cm}^{-1}$ .



**Figura 3.11.** Comparação dos sinais resolvidos com o método MCR (+) com os sinais verdadeiros de cada composto químico (-). A) Variação da concentração em função do tempo; B) espectro do composto P; C) espectro do composto R; D) espectro do composto I.

Os espectros resolvidos com o método MCR-ALS são bem diferentes daqueles resolvidos com o método EFA. Da mesma forma, as curvas de variação das concentrações em função do tempo resolvidas com o método MCR-ALS (figura 3.11A), apresentam perfis de acordo com o que se espera para um sistema dinâmico onde as concentrações são positivas. Por isso, podemos considerar que as estimativas obtidas com o método MCR-ALS apresentam um significado químico, contrário as estimativas obtidas com o método EFA.

Os espectros do modelo bilinear de referência, matriz  $\mathbf{B}_{\text{ref}}$  (figura 3.10B), foram usados para prever a composição química das outras três bateladas identificadas como experimentos II, III e IV na figura 3.4. Como mencionado no início da discussão, foi verificado que a linha base dos espectros muda de uma batelada para outra segundo um deslocamento do tipo “offset”, de modo que uma coluna de vetores unitários  $\mathbf{1}$  foi

adicionada na matriz  $\mathbf{B}_{\text{ref}}$  como mostra o esquema da figura 3.1. A nova matriz formada  $\mathbf{B}_{\text{est}}$  foi usada para prever a composição das misturas nos experimentos II, III e IV.

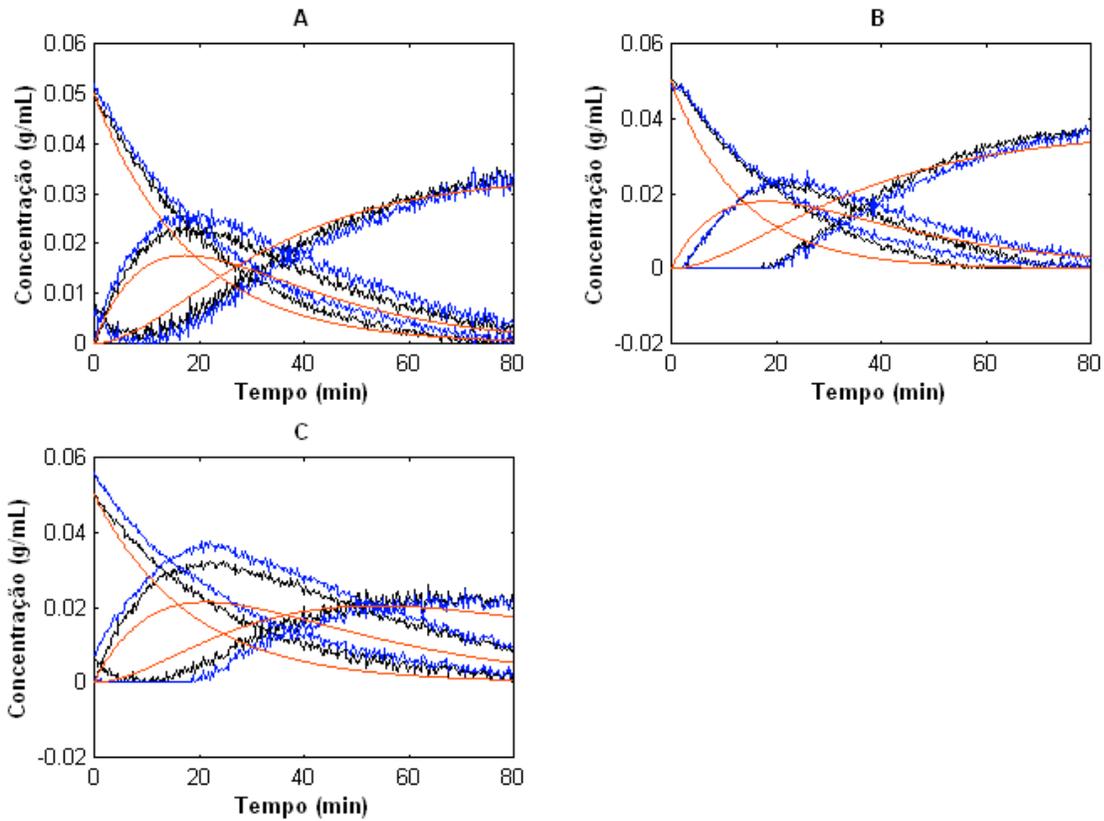
As figuras 3.12A-C mostram três curvas de variação da concentração em função do tempo. Uma destas curvas corresponde à solução matemática das equações usadas para produzir as misturas que portanto são as verdadeiras curvas (linha vermelha). As outras duas curvas são os resultados da previsão com a matriz  $\mathbf{B}_{\text{ref}}$  (linha azul) e  $\mathbf{B}_{\text{est}}$  (linha preta), de modo que são comparadas as previsões com e sem correção de linha base com a verdadeira curva. As previsões feitas para os experimentos II e IV foram estreitamente melhores quando feitas com a matriz  $\mathbf{B}_{\text{est}}$ , enquanto que para o experimento III, não foram observadas diferenças significativas.

A variação de linha base no experimento IV é maior que dos outros experimentos e por isso o efeito da correção é mais facilmente percebido para esta batelada. Pode ser observado ainda que as previsões feitas com as matrizes  $\mathbf{B}_{\text{ref}}$  e  $\mathbf{B}_{\text{est}}$  não são iguais às verdadeiras curvas, principalmente porque o espectro resolvido pelo método MCR para o composto R apresenta um desvio em relação ao verdadeiro espectro desta substância como mostra a figura 3.11B. Sob estas circunstâncias é esperado que as curvas preditas pelas matrizes  $\mathbf{B}_{\text{ref}}$  e  $\mathbf{B}_{\text{est}}$  apresentem um desvio em relação à verdadeira curva.

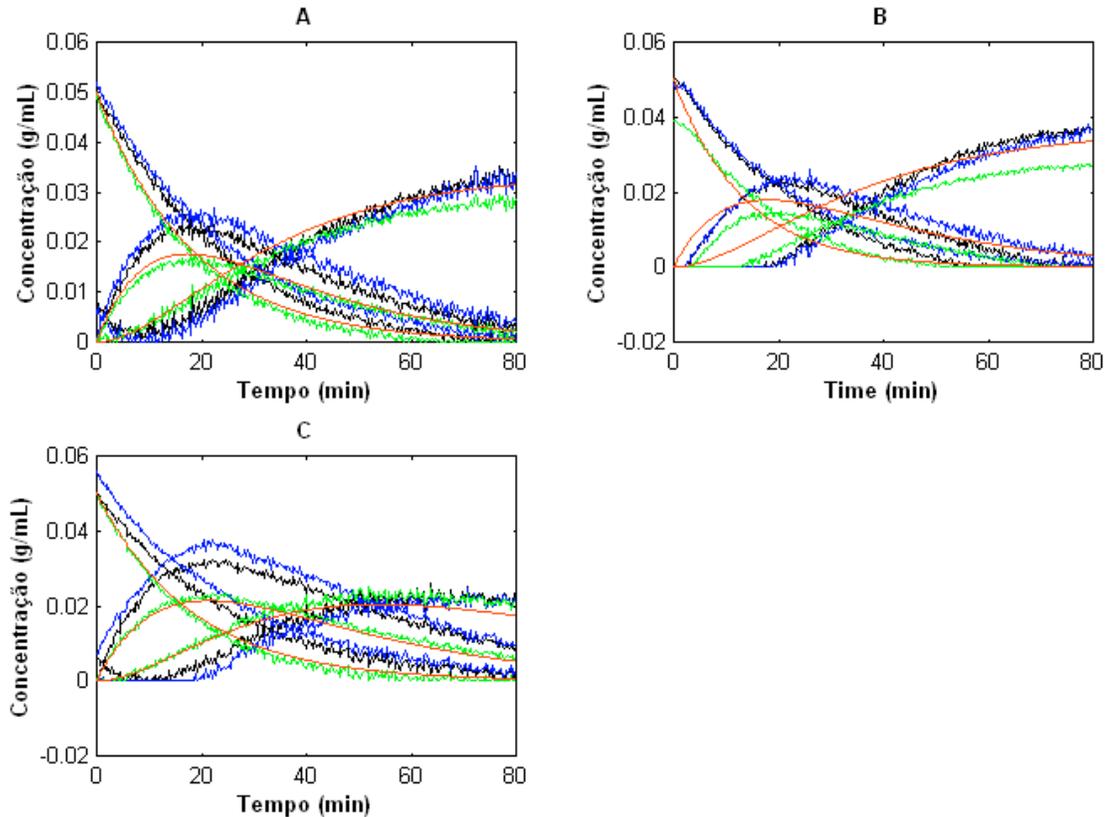
Para demonstrar a hipótese discutida no parágrafo anterior, uma matriz  $\mathbf{S}$  foi preparada com os espectros verdadeiros dos composto R, I e P usados para a produzir as misturas nos experimentos I, II, III e IV mais um vetor coluna  $\mathbf{1}$ . Esta nova matriz foi usada para prever a variação das concentrações em função do tempo de cada composto químico nos experimentos II, III e IV. Os resultados obtidos são mostrados nas figuras 3.13A-C. Se a adição dos compostos no tanque agitado contínuo (figura 3.2) tivesse sido perfeita, as linhas verdes e vermelhas deveriam se sobrepôr nas figuras 3.13A-C, visto que uma dessas é a solução das equações usadas para produzir as misturas (linhas vermelhas) e a outra é uma solução obtida a partir dos verdadeiros espectros das substâncias (linhas verdes). Para os experimentos II e IV a correspondência é muito boa, porém no experimento III certamente houve algum problema experimental na reprodução do experimento.

Note entretanto que se assumirmos que a solução obtida com os espectros verdadeiros é a que de fato corresponde ao verdadeiro perfil das curvas de concentração

das substâncias dentro do tanque, ainda assim, as previsões feitas com a matriz estendida  $\mathbf{B}_{est}$  (linha preta) aproximam-se mais da verdadeira solução (linha verde). Estas duas linhas não se sobrepõem por que a estimativa do espectro do composto P com o método MCR-ALS não é igual ao espectro verdadeiro deste composto.

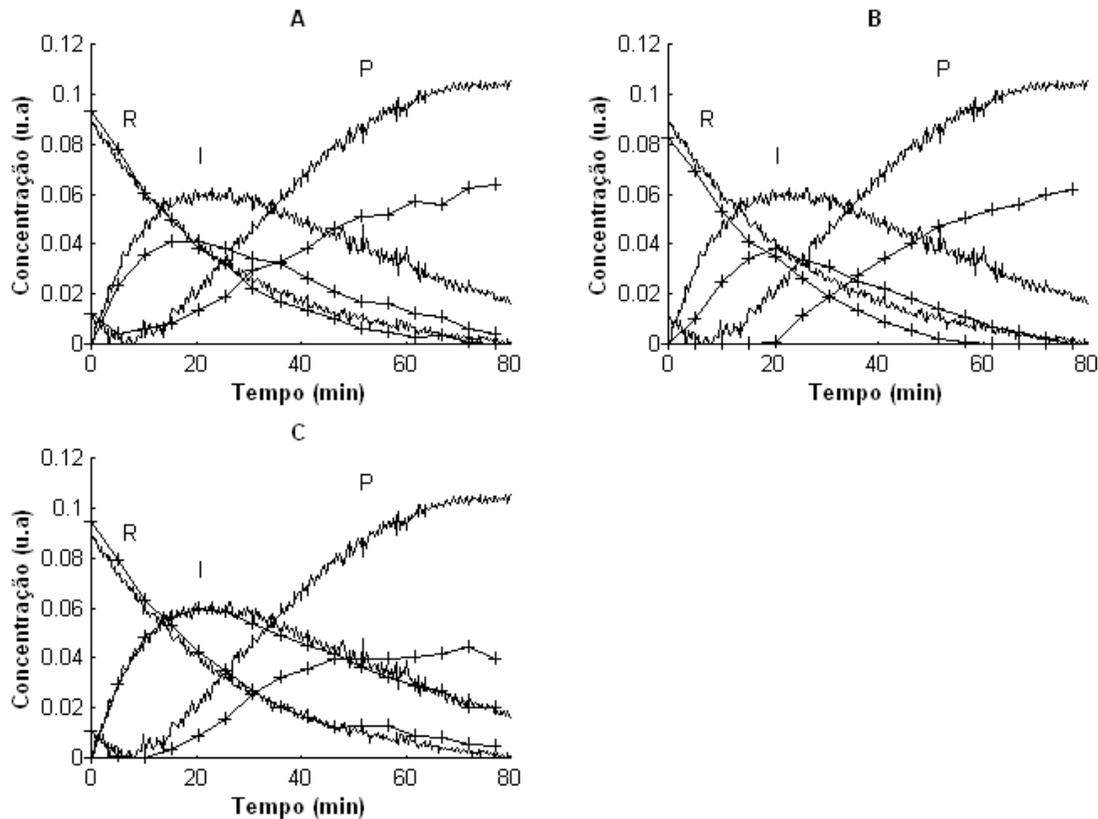


**Figura 3.12.** Variação das concentrações em função do tempo previstas pelo modelo bilinear. A) II; B) III; C) IV; (-) Curvas de concentrações verdadeiras, (-) Previsões feitas sem correção de linha base, (-) previsões feitas com correções de linha base.



**Figura 3.13.** Variação das concentrações em função do tempo previstas pelo modelo bilinear. A) II; B) III; C) IV; (-) Curvas de concentrações verdadeiras, (-) Previsões feitas sem correção de linha base, (-) previsões feitas com correções de linha base, (-) Previsões feitas com os espectros verdadeiros pelo método dos mínimos quadrados e com correção de linha base.

Para comparar cada nova batelada com a de referência, podemos colocar em um mesmo gráfico as curvas de variação das concentrações da batelada de referência e as curvas previstas com a matriz  $\mathbf{B}_{est}$  (linha de base corrigida) para as novas bateladas, como mostra a figura 3.14. Através das curvas de concentração em função do tempo, pode-se observar que nos experimentos II, III e IV, houve a formação de uma quantidade menor do produto. Se os exemplos apresentados neste trabalho fossem uma produção industrial verdadeira onde as misturas fossem na verdade alguma reação química, o operador do reator teria concluído que após 18 min da batelada do experimento II (figura 3.14A), a produção do composto P estaria desviando-se para valores mais baixos de concentração em relação a produção feita com a batelada de referência.



**Figura 3.14.** Resultados obtidos para previsão das novas bateladas (+) com a matriz  $\mathbf{B}_{est}$ , comparados com modelo de referência (-). A) batelada II; B) batelada III; C) batelada IV.

Ao comparar a produção da batelada de referência com a batelada III, o operador iria verificar que todos os compostos químicos nesta nova batelada tendem a valores de concentrações mais baixos. Conseqüentemente a quantidade de produto produzida pela batelada III seria menor que a da batelada de referência. Na batelada IV, as concentrações dos compostos R e I são iguais a da batelada de referência, mas a quantidade produzida de produto P é inferior a quantidade obtida na batelada de referência. Esta informação poderia ser avaliada como uma degradação do produto.

Os experimentos analisados neste trabalho apresentam um deslocamento na forma de uma linha constante, “offset”. Porém a metodologia não é limitada a este caso e pode ser usada para a modelagem de outros tipos de deslocamentos da linha base, bastando para isso adicionar em  $\mathbf{B}_{ref}$  a função que modela o tipo de deslocamento desta linha.<sup>62</sup>

### 3.6 CONCLUSÃO

Neste trabalho foi desenvolvida uma metodologia geral para o monitoramento espectroscópico de sistemas dinâmicos tal como reatores químicos. O método de resolução multivariado de curvas, foi usado para a modelagem do processo. Este método tem como vantagem o fato de que um modelo pode ser deduzido mesmo quando não há nenhum tipo de conhecimento do mecanismo e número de compostos químicos que participam do sistema. Em uma segunda etapa o modelo é usado para prever o comportamento de bateladas consecutivas do processo e pode ser adaptado caso ocorra alguma alteração nos espectros de uma batelada para outra.

O desempenho da metodologia foi avaliado apenas para as alterações na linha base. Do ponto de vista teórico é possível corrigir o modelo sempre que for determinada a forma da alteração dos espectros nas novas bateladas. Em termos práticos nem sempre é possível determinar esta informação e este fato corresponde a uma limitação da metodologia em questão.

## **CAPÍTULO 4**

**DETERMINAÇÃO DE CREATININA EM SORO HUMANO USANDO A REAÇÃO DE JAFFÉ E CALIBRAÇÃO DE SEGUNDA ORDEM.**

## 4.1 INTRODUÇÃO

A concentração de creatinina no soro humano é um parâmetro clínico muito importante para o tratamento dos pacientes submetidos a hemodiálise, visto que as determinações de creatinina têm por objetivo o diagnóstico e o monitoramento da insuficiência renal aguda e crônica e o monitoramento da diálise renal. A determinação de creatinina no soro humano é feita nos laboratórios clínicos por métodos enzimáticos<sup>82,83</sup> ou por métodos baseados na reação de Jaffé.<sup>84</sup>

As reações enzimáticas utilizam reações específicas para converter a creatinina em um composto derivado que pode ser detectado por diferentes métodos analíticos.<sup>85</sup> A reação de Jaffé usa o reagente ácido pícrico em meio alcalino para converter a creatinina em um composto que pode ser detectado na região do espectro visível, em torno do comprimento de onda de 500 nm.<sup>86</sup>

Os métodos enzimáticos têm como vantagem melhor seletividade para determinação de creatinina, mas nos laboratórios clínicos, os métodos baseados na reação de Jaffé são os mais utilizados provavelmente porque apresentam um custo menor para a análise quando comparado com os métodos enzimáticos, além da maior facilidade no manejo dos reagentes, visto que a reação de Jaffé não utiliza reagente bioquímico.

Embora a reação de Jaffé seja muito utilizada para a determinação da creatinina, há muitas substâncias no soro humano que podem interferir com a determinação desta substância, tais como proteínas, bilirrubina, glicose, ácido aceto-acético e antibióticos. Estes compostos reagem com o ácido pícrico em meio alcalino e formam compostos químicos que interferem na medida fotométrica do produto da reação de creatinina com ácido pícrico.<sup>87</sup>

Para eliminar ou minimizar a interferência das outras substâncias, é necessário separar a creatinina ou utilizar algum método de pré-tratamento da amostra. Porém, estes procedimentos nem sempre são adequados para as análises de rotina, e os laboratórios clínicos em sua maioria preferem usar o método cinético baseado na reação de Jaffé para determinação da creatinina no soro humano.

O método cinético baseado na reação de Jaffé torna possível a determinação de creatinina no soro humano na presença de proteínas sem a necessidade de etapas de separação ou pré - tratamento da amostra.<sup>88</sup> No método cinético, a seletividade da medida para a creatinina é consequência da diferença de velocidades das reações da creatinina e albumina com o ácido pícrico. Apesar disto, este método não está livre da interferência de outras substâncias como bilirrubina, ácido aceto-acético e antibióticos.<sup>89,85</sup>

Neste trabalho de tese, uma nova metodologia analítica para a determinação da creatinina é proposta. O procedimento experimental é semelhante ao do método cinético, a diferença é que a reação de Jaffé é monitorada por um espectrofotômetro com arranjo de diodos ao invés de um fotômetro. Para a análise quantitativa é usado um método de decomposição trilinear para a determinação da concentração de creatinina. Esta metodologia permite o monitoramento multivariado em dois modos (espectroscópico x cinético) e apresenta como vantagem o fato de que as interferências podem ser resolvidas matematicamente e a concentração de creatinina determinada sem que sejam necessárias as etapas de separação e pré-tratamento da amostra por procedimentos químicos ou físicos.

## 4.2 DETERMINAÇÃO DA CREATININA EM SORO HUMANO COM A REAÇÃO DE JAFFÉ

De acordo com o trabalho de Vasiliades,<sup>84</sup> a creatinina reage com o ácido pícrico em meio alcalino em duas etapas. Na primeira etapa da reação, ocorre a formação de uma espécie intermediária a partir da reação das moléculas de ácido pícrico com os íons hidróxidos, [ picrato-OH ]<sup>\*</sup>. A espécie intermediária reage com as moléculas de creatinina formando um composto complexo de cor vermelha que pode ser detectado na região do espectro visível de 480 a 600 nm.

Se uma concentração em excesso de ácido pícrico for usada e a concentração de hidróxido mantida fixa, a velocidade da reação é de pseudo primeira ordem para a creatinina, de modo que :

$$\frac{dc}{dt} = -k_1c$$

4.1

onde  $c$  na equação 4.1 é a concentração da creatinina e  $k_1$  a constante de velocidade da reação de primeira ordem.

A reação de Jaffé é a base de várias metodologias que foram desenvolvidas com o intuito de determinar creatinina no soro humano. A diversidade de metodologias baseadas nesta reação é justificada pelo fato de que o ácido pícrico em meio alcalino reage com muitas substâncias presentes no soro humano. As proteínas séricas são um caso problemático visto que estão sempre presentes no soro e a reação destas com ácido pícrico formam compostos que apresentam espectros de absorção molecular que se sobrepõem ao do produto da reação da creatinina com ácido pícrico.

Para minimizar a interferência das proteínas presentes no soro humano foram desenvolvidas metodologias baseadas na extração da creatinina ou na eliminação destas proteínas.<sup>90,91</sup> Uma das mais representativas neste sentido é a metodologia que utiliza o reativo de Loyd,<sup>92</sup> para a separação da creatinina por adsorção. As metodologias baseadas no método cinético são uma outra forma de resolver o problema da interferência causada pelas proteínas.<sup>88,93,94</sup>

Uma substância que pode criar problemas é a bilirrubina. Esta substância costuma estar presente no soro em concentrações abaixo do nível necessário para interferir com a determinação da creatinina. Mas há doenças que atacam principalmente as crianças, que elevam a concentração de bilirrubina para níveis acima do normal no soro. Uma das formas usadas para remover a interferência da bilirrubina é incubar as amostras de soro em temperaturas superiores ao do método analítico por um período de tempo a fim de degradar a bilirrubina.<sup>87</sup>

Pardue,<sup>89</sup> investigou diferentes combinações das concentrações de ácido pícrico, hidróxido de sódio e temperatura da reação, a fim de otimizar a seletividade do método cinético para a determinação da creatinina em soro humano. O tratamento dos dados cinéticos foi feito com um método de ajuste de curvas não lineares e a metodologia foi capaz de minimizar as interferências causadas pelas substâncias albumina, bilirrubina, glicose e ácido aceto-acético.

### 4.3 CALIBRAÇÃO DE SEGUNDA ORDEM

Na década de 1990, Booksh e Kowalski publicaram um artigo em uma prestigiada revista da área de química analítica com um conteúdo bastante original, onde eles formaram uma teoria para a química analítica.<sup>95</sup> Esta teoria é baseada no número de modos que um instrumento ou uma metodologia analítica monitora a mesma amostra. Um instrumento de ordem zero é aquele que mede apenas um único sinal para uma amostra. Um instrumento de ordem um, mede mais do que um sinal para uma amostra, porém, em um único modo de monitoramento, enquanto que um instrumento de ordem dois registra centenas de sinais para uma mesma amostra em dois modos de monitoramento.

De acordo com esta teoria um fotômetro é um instrumento de ordem zero e a medida feita com este equipamento corresponde a um elemento escalar. Um espectrofotômetro é um instrumento de ordem um e as medidas feitas com este instrumento podem ser agrupadas em um vetor  $\mathbf{x}$ . Já um cromatógrafo acoplado a um espectrofotômetro é um equipamento de ordem dois e as medidas feitas com este instrumento podem ser agrupadas em uma matriz  $\mathbf{X}$ .

Para cada tipo de instrumento existe um método de calibração mais adequado. Para um instrumento de ordem zero utiliza-se o método de calibração clássico, para a calibração de um instrumento de ordem um, utiliza-se um método de calibração de primeira ordem (multivariado) e para um instrumento de ordem dois, um método de calibração de segunda ordem (multivariado).

Há uma vantagem associada para cada modo de monitoramento a mais que for incluído na calibração. Com a calibração clássica é necessário que o sinal medido para um analito seja seletivo em relação a outras espécies presentes nas amostras. Para a calibração de primeira ordem os sinais do analito não precisam ser seletivos e este pode ser quantificado na presença de outras espécies interferentes, porém, é necessário que nas soluções padrões estejam presentes todas as espécies que interferem com a medida do analito. Na calibração de segunda ordem, os sinais do analito não precisam ser seletivos e não há necessidade de incluir nas soluções padrões as outras espécies interferentes, pois a espécie de interesse pode ser determinada nas amostras mesmo

quando há outras espécies desconhecidas que não foram modeladas na etapa de calibração.

Há uma imensa vantagem analítica associada à calibração de segunda ordem que foi denominada de “vantagem da segunda ordem”. Esta “vantagem” guiou o desenvolvimento de algumas metodologias analíticas para o desenvolvimento de sensores de segunda ordem capazes de fazer medidas em dois modos de monitoramento (tempo x comprimento de onda), como o sensor desenvolvido para a determinação de metais em águas<sup>96</sup> e para a determinação *in situ* de compostos organoclorados em água.<sup>97</sup>

A calibração de segunda ordem tornou-se uma técnica importante com o advento das técnicas de cromatografia acopladas a espectrofotômetros ou espectrômetros de massa. Estes equipamentos hifenados são instrumentos de segunda ordem, assim como o espectrofluorímetro, para os quais podemos usar os métodos de calibração de segunda ordem e fazer uso da vantagem da segunda ordem.<sup>98,99</sup>

Os métodos de calibração de segunda ordem como foram originalmente propostos apresentam um formalismo matemático baseado na resolução multivariada de curvas, pois o sinal de cada composto químico presente no arranjo  $\underline{X}$  é separado matematicamente para que em seguida seja usado para a análise quantitativa. A vantagem de segunda ordem é uma consequência direta da unicidade da solução dos métodos de decomposição trilinear,<sup>42</sup> uma vez que os parâmetros ajustados para o modelo trilinear correspondem a estimativas para os sinais de cada composto químico presente no arranjo  $\underline{X}$ .

O Método Generalizado de Anulação de Posto (GRAM)<sup>9</sup> foi um dos primeiros métodos propostos para a calibração de segunda ordem. No método GRAM, o arranjo  $\underline{X}$  é formado por duas amostras, uma solução padrão contendo apenas o analito de interesse e a amostra real em que se quer fazer a quantificação. O método de Decomposição Trilinear Direto (DTLD) é uma generalização do método GRAM, desenvolvido para ampliar o número de soluções padrões usados para calibração de segunda ordem. Posteriormente, surgiram outros métodos, sendo que o mais importante destes é o PARAFAC.<sup>100</sup>

Para as análises de rotina em que é necessário realizar muitas determinações, o método GRAM é o preferido porque este método de decomposição trilinear não é iterativo e por isso ajusta um modelo mais rapidamente do que o método iterativo PARAFAC. A

vantagem do método PARAFAC é que este geralmente produz resultados mais precisos e exatos que o método GRAM.

Para algumas aplicações há necessidade de ajustar um modelo diferente do trilinear para a determinação do analíto. Geralmente são aplicações envolvendo a análise de dados com o problema da deficiência de posto. Para estas aplicações utilizam-se os métodos TUCKER com núcleo restrito,<sup>101</sup> Anulação de Posto não linear (NBRA)<sup>102</sup> ou o MCR-ALS.<sup>103</sup>

O método de segunda ordem proposto por Linder e Sundberg,<sup>104</sup> denominado de “Ajuste dos Mínimos Quadrados Bilineares” (BLLS), apresenta um formalismo matemático diferente dos métodos de resolução de curvas. A estratégia adotada por Linder é similar a da análise de regressão. Primeiro, um modelo bilinear é ajustado apenas com as soluções padrões dos analitos de interesse (etapa de calibração), em seguida este modelo é usado para prever a concentração dos analitos nas amostras desconhecidas (etapa de previsão). A vantagem de segunda ordem é mantida com o custo de que a interferência deve ser detectada e o seu efeito corrigido no modelo de calibração com o método “residual bilinearization” (RBL).<sup>76</sup>

Trabalhos recentes<sup>105,106</sup> mostram que a precisão do método BLLS é superior aos métodos GRAM, DTLTD e PARAFAC e apresenta, ainda, a vantagem de que há uma expressão derivada para o cálculo da incerteza das previsões. Uma desvantagem do BLLS é que para fazer a calibração é necessário dispor de um número de amostras iguais ao número de analitos, o que não é uma restrição para os métodos PARAFAC e DTLTD, além disto, a exatidão do resultado obtido com o método BLLS depende da capacidade de detectar eventuais espécies que não estavam presentes nas amostras de calibração.

Algum progresso tem sido feito no desenvolvimento dos métodos para o cálculo das características de método (ou figuras de mérito) dos modelos de segunda ordem. Algumas propostas foram feitas para calcular parâmetros como a sensibilidade, seletividade e limite de detecção.<sup>107,108,109</sup> Uma expressão para o cálculo da incerteza da previsão com a calibração de segunda ordem encontra-se disponível até o momento apenas para os métodos GRAM<sup>110</sup> e o BBLs.<sup>111</sup>

#### 4.4 MÉTODO ESPECTRO-CINÉTICO MULTIVARIADO PARA A DETERMINAÇÃO DE CREATININA

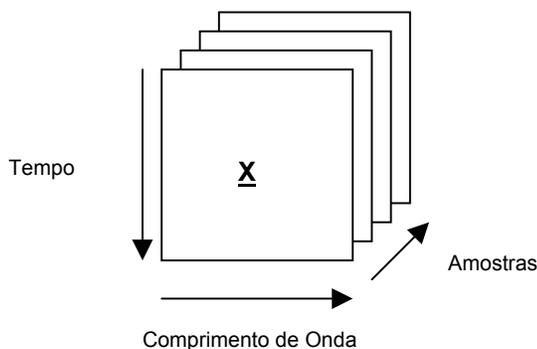
A metodologia proposta nesta tese para a determinação de creatinina em soro humano é denominada de método espectro-cinético multivariado e é baseada no monitoramento multivariado em dois modos da reação de Jaffé. O primeiro modo de monitoramento são os comprimentos de onda na região do espectro visível (480-600 nm) e o segundo modo as medidas feitas ao longo do tempo. Para cada amostra monitorada é obtida uma matriz  $X_{i \times J}$  onde os elementos desta matriz são os dados de absorção molecular.

O método espectro-cinético multivariado faz as seguintes suposições:

- I. A reação da creatinina com ácido pícrico em meio alcalino é de pseudo primeira ordem.
- II. A constante de velocidade da reação é a mesma em todos os experimentos.
- III. A variação da absorbância é proporcional e linear com a variação da concentração de creatinina.

Estas suposições são as mesmas necessárias para a aplicação dos métodos cinéticos univariados. A primeira suposição, garante que a absorbância medida em um instante  $i$  é diretamente proporcional à concentração inicial de creatinina ( $c_{ini}$ ). A segunda suposição é necessária para garantir que a curva da variação da concentração em função do tempo para o produto da reação da creatinina é a mesma em todos os experimentos.

A análise quantitativa baseada na calibração de segunda ordem apresenta duas etapas. A primeira etapa consiste em agrupar as amostras em um arranjo com três modos  $X_{i \times J \times K}$  e ajustar um modelo trilinear para este conjunto de dados. Note, que  $I$  é o número de espectros registrados em função do tempo,  $J$  é o número de comprimentos de onda em que foram medidas as absorbâncias e  $K$  o número de amostras, figura 4.1. Todas as análises realizadas neste trabalho agrupam quatro amostras no arranjo com três modos. As três primeiras amostras são soluções padrões de creatinina e a quarta amostra é aquela em que se deseja determinar a concentração de creatinina.



**Figura 4.1.** Arranjo em três modos formado para o conjunto de amostras analisadas com o monitoramento espectro-cinético multivariado

Na segunda etapa da análise é identificada a componente  $r$  do modelo trilinear que corresponde aos sinais da creatinina. Estes sinais são, a variação da concentração em função do tempo ( $\mathbf{a}_r$ ), o espectro puro ( $\mathbf{b}_r$ ) e a concentração relativa em cada amostra ( $\mathbf{c}_r$ ). Os elementos do vetor  $\mathbf{c}_r$  são as estimativas para as concentrações da creatinina nas  $K$  amostras. Os três primeiros elementos de  $\mathbf{c}_r$ , referem-se às concentrações de creatinina nas soluções padrões e o quarto elemento corresponde a uma estimativa para a concentração na amostra, figura 4.2

$$\mathbf{C} = \begin{bmatrix} c_{11} & c_{12} & \dots & c_{1r} \\ c_{21} & c_{22} & \dots & c_{2r} \\ c_{31} & c_{32} & \dots & c_{3r} \\ c_{41} & c_{42} & \dots & c_{4r} \end{bmatrix} \xrightarrow{\text{Coluna identificada}} \mathbf{c}_r = \begin{bmatrix} c_{1r} \\ c_{2r} \\ c_{3r} \\ c_{4r} \end{bmatrix}$$

$\left. \begin{matrix} c_{1r} \\ c_{2r} \\ c_{3r} \end{matrix} \right\} \text{ Soluções padrões}$   
 $\left. \begin{matrix} c_{4r} \end{matrix} \right\} \text{ Amostra}$

**Figura 4.2.** Matriz  $\mathbf{C}$  do modelo trilinear  $\mathbf{A}(\mathbf{B} | \otimes | \mathbf{C})^T$ .

Os parâmetros resolvidos pelo modelo trilinear correspondem aos sinais de cada espécie na mistura, todavia, existe uma ambigüidade associada com a intensidade dos sinais. Esta ambigüidade existe porque os parâmetros  $\mathbf{A}$ ,  $\mathbf{B}$  e  $\mathbf{C}$  resolvidos por um dos métodos de decomposição trilinear, são multiplicados por uma constante desconhecida. Portanto, as concentrações resolvidas precisam ser escaladas para a escala de concentração das soluções padrões. A conversão da escala é realizada através da equação 4.2, para a qual  $\mathbf{c}$  é um vetor formado pelos três primeiros elementos de  $\mathbf{c}_r$  e o

vetor  $\mathbf{c}_p$  é o respectivo vetor com as concentrações das três soluções padrões de creatinina. O coeficiente  $d$  é resolvido por um ajuste dos mínimos quadrados e a concentração de creatinina na amostra  $K = 4$  é determinada pela equação 4.3.

$$\mathbf{c} = \mathbf{c}_p d \quad 4.2$$

$$c_{pred} = c_{4r} d \quad 4.3$$

## 4.5 PARTE EXPERIMENTAL

### 4.5.1 Equipamentos

Todos os espectros foram registrados com um espectrofotômetro com arranjo de diodos (Hewlett-Packard, Palo Alto, CA, modelo 8453A), equipado com uma cela para controle de temperatura. Os reagentes foram misturados manualmente em cubetas de plástico com um caminho ótico de 1 cm.

Os dados do espectrofotômetro foram transferidos para um microcomputador equipado com um microprocessador AMD 850 MHZ. Rotinas de autoria própria para o cálculo do modelo trilinear com os métodos PARAFAC e DTLTD, foram implementadas no software Matlab, versão 6 (Mathworks).

### 4.5.2 Reagentes e Soluções

Todos reagentes químicos utilizados foram de grau P.A. Creatinina e Albumina (Fração V) foram obtidas da Sigma. Bilirrubina foi adquirida da Acros Organics e Ácido Clorídrico, Hidróxido de Sódio e Glicose da Merck. Todas as soluções aquosas foram preparadas com água ultra pura obtida com o aparelho Milli-Q (Millipore, Bedford, MA).

Uma solução estoque de creatinina  $1 \text{ g L}^{-1}$  foi preparada por dissolução de 1 g de creatinina em 1 L de uma solução de  $0,1 \text{ mol L}^{-1}$  de ácido clorídrico. Soluções analíticas de creatinina foram preparadas por diluição direta de volumes apropriados da solução estoque.

A solução reagente “alcalino picrato” foi preparada por dissolução de 0,75 mmol de ácido pícrico em 25 mL de água e 5 mL de uma solução 5 M de hidróxido de sódio. Após completa dissolução, o volume foi completado a 50 mL com água ultra pura. Soluções do reagente “alcalino picrato” foram preparadas toda vez que uma batelada de análises foi executada. Para os experimentos do tópico 4.5.4, foram preparadas soluções de alcalino picrato com concentrações diferentes conforme exposto na tabela 4.1.

Uma solução estoque de bilirubina foi preparada por dissolução de 20 mg de bilirrubina em 1 mL de dimetilsulfóxido e 3 mL de uma solução de carbonato de sódio 0,1 mol L<sup>-1</sup>. Após completa dissolução, o volume foi completado com 10 mL com água ultra pura.

### **4.5.3 Experimentos de cinética.**

Em uma cubeta de plástico de 1 mL foram adicionados 1 mL de água ultra pura mais 0,5 mL da amostra ou solução padrão de creatinina. A esta solução foi adicionado 1 mL do reagente alcalino picrato, de modo que o volume final de cada ensaio analítico foi sempre de 2,5 mL.

O início da reação é aquele em que a última gota do reagente alcalino picrato foi adicionada. Medidas de absorbância foram feitas em intervalos regulares de 2 s durante 300 s. Os espectros foram registrados em intervalos regulares de 1 nm para o intervalo de comprimentos de onda entre 450 a 600 nm. As medidas de absorbância foram feitas contra um branco constituído por água.

Todas soluções e reagentes usadas nos experimentos foram incubadas durante 30 min a 30 °C.

### **4.5.4 Experimentos para análise da reação de creatinina com ácido pícrico em meio alcalino**

Os experimentos do planejamento fatorial 2<sup>3</sup> apresentado na tabela 4.1, foram executados a fim de investigar o efeito da concentração de hidróxido de sódio,

concentração de ácido pícrico e temperatura na velocidade da reação da creatinina com ácido pícrico.

Os experimentos 5 e 8 da tabela 4.1 foram feitos em duplicata variando a concentração de creatinina nos níveis de 10, 20, 40 e 60 mg L<sup>-1</sup>. Esta seqüência de reações foi usada para determinar a ordem da reação pelo método das velocidades iniciais.<sup>112</sup>

Todos os demais experimentos da tabela 4.1 foram feitos em duplicata para uma concentração de creatinina de 10 mg L<sup>-1</sup>.

**Tabela 4.1.** Planejamento fatorial 2<sup>3</sup>. Cada experimento corresponde a uma reação com 10 mg L<sup>-1</sup> de creatinina.

Experimentos	[OH <sup>-</sup> ] (mol L <sup>-1</sup> )	[Ác. Pícrico] (mol L <sup>-1</sup> )	Temperatura (°C)
		0,2 (-)	6 x 10 <sup>-3</sup> (-)
	0,4 (+)	9 x 10 <sup>-3</sup> (+)	30 (+)
1	-	-	-
2	+	-	-
3	-	+	-
4	+	+	-
5*	-	-	+
6	+	-	+
7	-	+	+
8*	+	+	+

\* Experimentos realizados em duplicatas com concentrações de creatinina de 10, 20, 40 e 60 mg L<sup>-1</sup>.

#### 4.5.5 Experimentos para determinação de creatinina em misturas

Um conjunto de nove grupos de misturas, conforme mostra a tabela 4.2, foi preparado com intuito de testar o método espectro-cinético multivariado. Cada grupo de misturas é formado por cinco experimentos com concentrações diferentes de creatinina de 5, 10, 20, 30 e 50 mg L<sup>-1</sup>. No total foram realizados 45 experimentos.

**Tabela 4.2.** Conjunto de misturas preparadas para testar a metodologia desenvolvida.

Grupo	Creatinina (mg L <sup>-1</sup> )	Albumina (mg L <sup>-1</sup> )	Bilirrubina (mg L <sup>-1</sup> )	Glicose (mg L <sup>-1</sup> )	Aceto-acético (mmol L <sup>-1</sup> )
1	5-50	50000	-	-	-
2	5-50	80000	-	-	-
3	5-50	50000	50	-	-
4	5-50	20000	10	1000	-
5	5-50	50000	10	1000	-
6	5-50	80000	10	1000	-
7	5-50	50000	50	1000	-
8	5-50	50000	10	2000	-
9	5-50	50000	-	-	2,6

#### 4.5.6 Experimentos para determinação de creatinina em amostras de soro humano

Seis amostras de soro humano foram obtidas do laboratório de análises clínicas do Hospital de Clínicas da UNICAMP. As concentrações de creatinina nestas amostras, foram determinadas com o método cinético padrão em um analisador automático Roche/Hitachi 917. A concentração usada de ácido pícrico foi de 25 mmol L<sup>-1</sup> e 200 mmol L<sup>-1</sup> de hidróxido de sódio.

As concentrações de creatinina determinadas no hospital foram usadas como referência para avaliar a metodologia proposta.

## 4.5.7 Análise dos dados

### 4.5.7.1 Determinação da ordem da reação da creatinina com ácido pícrico

O método utilizado para a determinação da ordem de reação foi o das velocidades iniciais.<sup>112</sup> As reações com as condições experimentais definidas nos experimentos 5 e 8 da tabela 4.1 foram feitas variando a concentração inicial de creatinina ( $c_{ini}$ ) nos níveis 10, 20, 40 e 60 mg/L. Nas condições usadas a concentração de ácido pícrico está presente em excesso. Os dados de absorvância medidos do produto da reação de creatinina com ácido pícrico em 500 nm, foram convertidos em valores de concentração dividindo a absorvância pelo seu coeficiente de absorvância e a ordem da reação foi determinada a partir da equação 4.4 pelo ajuste dos mínimos quadrados.

$$\log\left(\frac{dc}{dt}\right) = -\log k + n \log c_{ini} \quad 4.4$$

Na equação 4.4,  $dc/dt$  é a velocidade inicial da reação da creatinina com ácido pícrico medida no instante  $t=0$ ,  $k$  é a constante de velocidade da reação e  $n$  a ordem da reação. A função definida na equação 4.4 é linear, e o coeficiente angular desta função corresponde à ordem da reação. Para uma reação de primeira ordem,  $n$  é igual 1. A velocidade inicial de cada reação foi calculada a partir da diferenciação de um polinômio ajustado para a curva da variação das medidas de concentração em função do tempo.

### 4.5.7.2 Decomposição trilinear

A calibração de segunda ordem foi executada de acordo com o procedimento descrito no tópico 4.4. O cubo de dados  $\underline{X}$  formado para a análise de cada amostra foi sempre composto de três soluções padrões de creatinina (10, 20 e 40 mg L<sup>-1</sup>) mais uma amostra da tabela 4.2, no caso da análise das misturas ou uma amostra de soro humano. Para fins de comparação foram ajustados modelos trilineares com os métodos DTLD e PARAFAC.

Os cálculos com o método PARAFAC foram sempre iniciados com valores aleatórios e o critério de convergência usado foi sempre de  $1 \times 10^{-6}$ . Dois tipos de modelos foram ajustados com o método PARAFAC: um modelo foi calculado com restrições de não negatividade nos três modos. O outro modelo foi calculado com restrições de não negatividade mais as seguintes restrições dos elementos da matriz **C**:

$$\mathbf{C} = \begin{bmatrix} 10 & 0 & \dots & 0 \\ 20 & 0 & \dots & 0 \\ 40 & 0 & \dots & 0 \\ \alpha & \chi & \dots & \beta \end{bmatrix}$$

as três primeiras linhas foram fixadas no cálculo e as estimativas foram calculadas apenas para a quarta linha. As três primeiras linhas referem-se à composição química das soluções padrões de creatinina, que são conhecidas e por isto foram fixadas.

A determinação do número de componentes do modelo trilinear foi feita com base no gráfico ou valor do CORCONDIA (Material Suplementar: parte C) calculado para cada modelo trilinear. Um valor de CORCONDIA superior a 90 % indica bom ajuste dos dados a uma estrutura trilinear. A confirmação final do número de componentes foi feita por exame visual das estimativas dos parâmetros do modelo trilinear, obtidas a partir de cada método.

A previsão da concentração de creatinina nas misturas foi avaliada pelo erro definido como raiz do erro quadrático médio (RMSEP), calculado de acordo com a seguinte equação:

$$RMSEP = \sqrt{\frac{\sum_{k=1}^K (c - \hat{c})^2}{K}} \quad 4.5$$

onde  $c$  é o valor da concentração de creatinina adicionado nas misturas,  $\hat{c}$  é o valor predito pela metodologia proposta e  $K$  o número de experimentos em cada grupo da tabela 4.2.

Para comparar a previsão do método espectro-cinético multivariado com o método cinético nas análises realizadas em amostras de soro humano, foi utilizado o erro padrão

de desempenho (SEP), calculado conforme a equação 4.6 e o coeficiente de variação (CV), calculado de acordo com a equação 4.7.

$$SEP = \sqrt{\frac{\sum_{k=1}^K (\hat{c}_k - y_k)^2 - \frac{\left(\sum_{k=1}^K (\hat{c}_k - y_k)\right)^2}{K}}{K - 1}} \quad 4.6$$

$$CV = \frac{SEP \times 100}{\bar{c}} \quad 4.7$$

Na equação 4.6, a variável  $y_k$  é a concentração de creatinina prevista pelo método de referência na amostra  $k$  e a variável  $\bar{c}$  da equação 4.7 é a média aritmética das previsões feitas com o método espectro cinético multivariado.

## 4.6 RESULTADOS E DISCUSSÃO

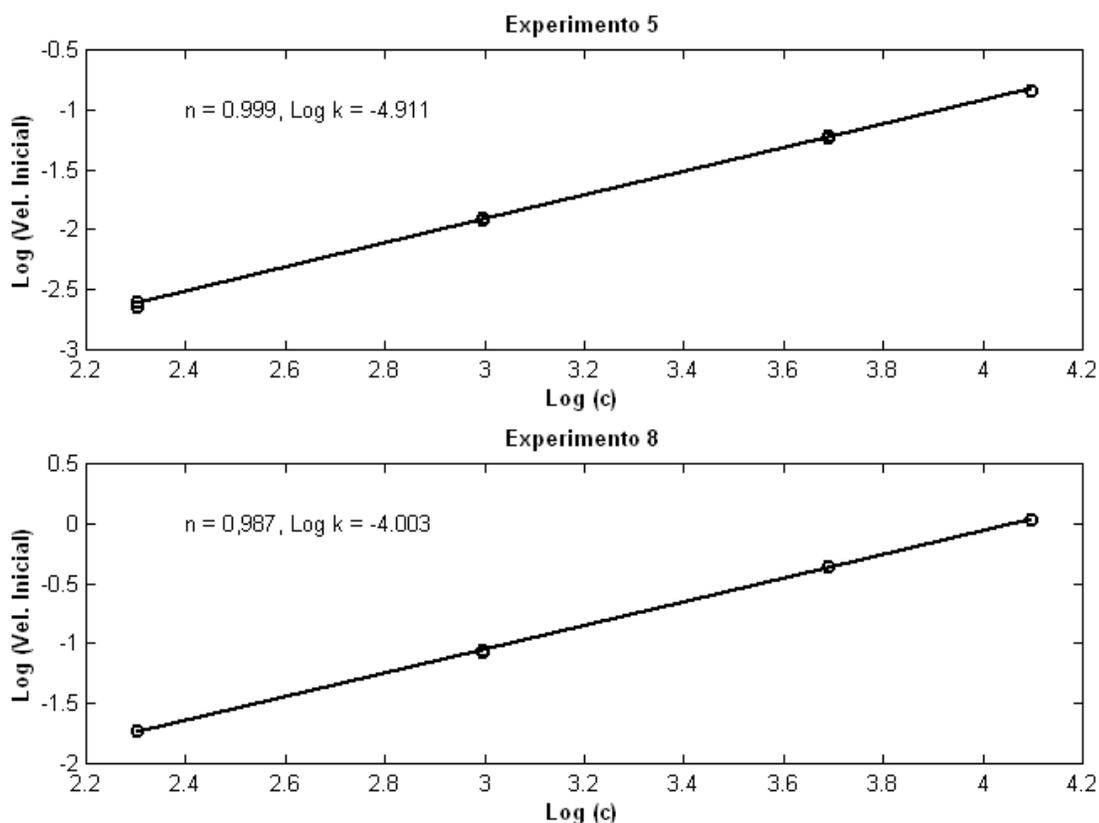
### 4.6.1 Análise da reação de creatinina com ácido pícrico em meio alcalino

Fontes da literatura<sup>87</sup> indicam que a reação da creatinina com o ácido pícrico em meio alcalino é de pseudo-primeira ordem para uma concentração de ácido pícrico entre 1 e 30 mmol L<sup>-1</sup>, e para uma concentração de hidróxido de sódio até 700 mmol L<sup>-1</sup>. Para este trabalho foram escolhidas concentrações abaixo de 10 mmol L<sup>-1</sup> de ácido pícrico, pois de acordo com Bowers,<sup>87</sup> a interferência na determinação da creatinina é menor para concentrações menores de ácido pícrico.

A figura 4.3, mostra o resultado da aplicação do método das velocidades iniciais<sup>112</sup> para os experimentos apresentados no tópico 4.5.4. O coeficiente angular da função do logaritmo das velocidades iniciais em função do logaritmo das concentrações foi de 0,999 para o experimento 5 e 0,987, para o experimento 8. Este resultado confirma o mecanismo

de pseudo-primeira ordem da reação nas condições experimentais definidas para os experimentos 5 e 8 da tabela 4.1.

Para os demais experimentos da tabela 4.1 foi assumido que o mecanismo de pseudo primeira ordem é válido, visto que são condições experimentais semelhantes às dos experimentos 5 e 8. Com base nestes resultados verificou-se que a suposição I (tópico 4.4) do método espectro-cinético multivariado é válida para estas condições experimentais.



**Figura 4.3.** Ordem da reação de creatinina com ácido pícrico determinada pelo método das velocidades iniciais.

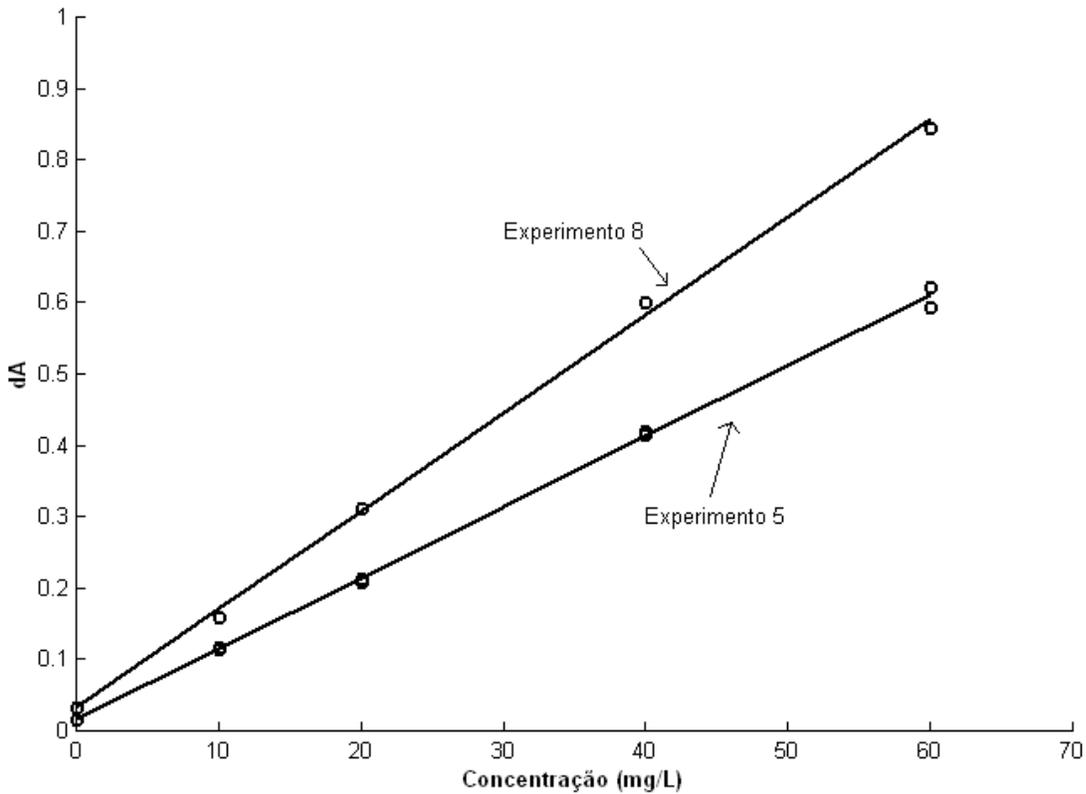
A tabela 4.3, mostra o valor das constantes de pseudo-primeira ordem calculadas pelo método de Guggenheim<sup>113</sup> para cada experimento da tabela 4.1. A concentração de hidróxido de sódio é a variável que mais afeta a velocidade da reação da creatinina com ácido pícrico. Ao dobrar a concentração de hidróxido de sódio, verifica-se um aumento de no mínimo 2 vezes o valor da constante de velocidade. No nível baixo de hidróxido, a constante  $k$  fica entre  $4,5$  e  $8,0 \times 10^{-3} \text{ s}^{-1}$ , enquanto que para o nível alto de hidróxido, os

valores das constantes  $k$  são superiores a  $11 \times 10^{-3} \text{ s}^{-1}$ . Para diminuir o tempo da análise química deve-se usar maiores concentrações de hidróxido de sódio.

**Tabela 4.3.** Planejamento Fatorial  $2^3$ . Cada experimento corresponde a uma reação com  $10 \text{ mg L}^{-1}$  de creatinina.

Experimentos	[OH <sup>-</sup> ]	[Ác. Pícrico]	Temperatura	Constante	
	mol L <sup>-1</sup>	mol L <sup>-1</sup>	°C	$k$	$k$
	0,2 (-)	$6 \times 10^{-3}$ (-)	25 °C (-)	$(10^{-3}) \text{ s}^{-1}$	$(10^{-3}) \text{ s}^{-1}$
	0,4 (+)	$9 \times 10^{-3}$ (+)	30 °C (+)		
1	-	-	-	4,50	4,60
2	+	-	-	11,40	11,70
3	-	+	-	6,60	6,40
4	+	+	-	13,0	12,20
5	-	-	+	5,50	6,00
6	+	-	+	11,40	11,00
7	-	+	+	8,00	7,20
8	+	+	+	14,3	13,00

A figura 4.4, mostra o gráfico  $\Delta A \times C_{ini}$ , calculado para os experimentos 5 e 8 da tabela 4.1. Neste gráfico,  $\Delta A$  é a diferença entre os valores de absorvância medidos nos instantes 20 s e 120 s no comprimento de onda de 500 nm e  $C_{ini}$  é a concentração inicial de creatinina. Conforme mostra a figura 4.4, a variação da absorvância entre dois pontos fixos do tempo<sup>88</sup> é proporcional a concentração inicial de creatinina para as medidas feitas no comprimento de 500 nm. O mesmo resultado foi verificado para os comprimentos de onda de 490, 520, 540, 560, 580 e 600 nm. Logo, a suposição III feita no tópico 4.4 é válida para as condições experimentais dos experimentos 5 e 8 da tabela 4.1.



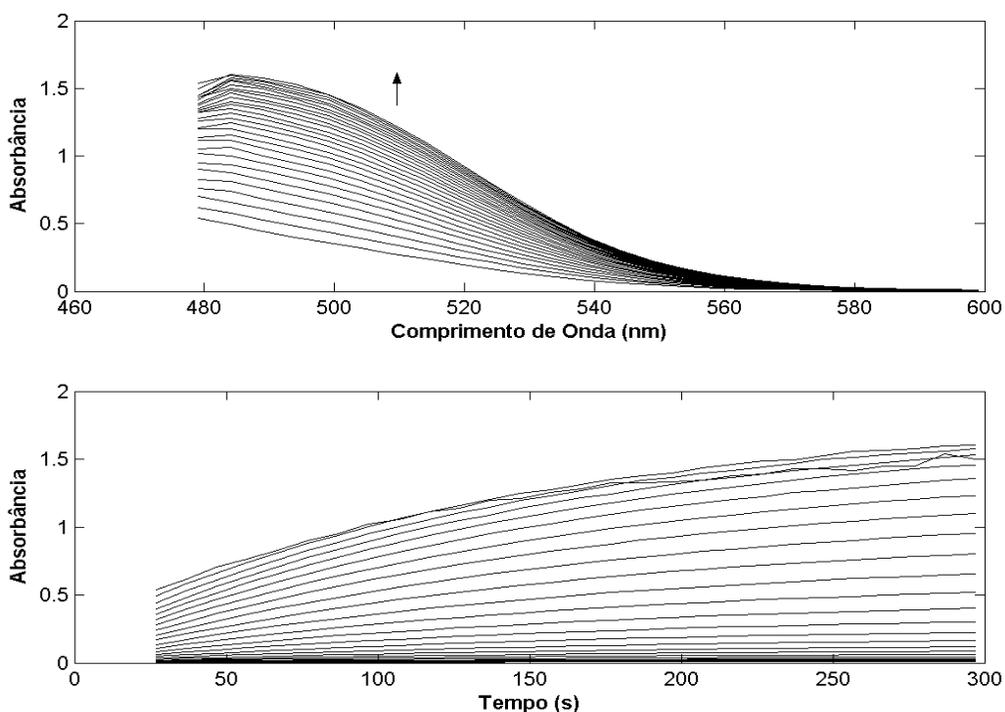
**Figura 4.4.** Gráfico  $\Delta A \times C_{ini}$  para o comprimento de onda 500 nm.  $\Delta A = (A_{20,500} - A_{120,500})$

Conforme demonstrado, as condições experimentais escolhidas para os experimentos 5 e 8 da tabela 4.1, reúnem as propriedades necessárias para o desenvolvimento do método espectro-cinético multivariado. Embora os demais experimentos da tabela 4.1 não tenham sido investigados com tantos detalhes como os experimentos 5 e 8, não há razões para supor que os outros não possuam as características necessárias para o método espectro cinético multivariado.

As condições definidas no experimento 5 da tabela 4.1 foram escolhidas para testar a metodologia proposta, para qual temos uma constante da velocidade de reação entre  $5,5$  e  $6 \times 10^{-3} \text{ s}^{-1}$ , conforme foram calculadas para cada uma das replicatas como mostra a tabela 4.3. Note que as condições experimentais do experimento 8 também podem ser usadas, com a vantagem de que o tempo da análise é reduzido pela metade.

## 4.6.2 Determinação da creatinina em misturas

A figura 4.5 mostra a variação das medidas de absorbância feitas nos dois modos de monitoramento (comprimento de onda x tempo) para uma das reações do grupo 6, tabela 4.2. A absorbância varia em uma estreita faixa de comprimentos de onda compreendida entre 480 e 600 nm. As substâncias creatinina, albumina e ácido acético reagem com ácido pícrico em meio alcalino, produzindo compostos que absorvem nesta faixa de comprimentos de onda.<sup>89</sup> A bilirrubina apresenta um espectro de absorção molecular nesta região, e no meio alcalino esta substância é degradada em um composto que não apresenta um espectro na região do visível. A reação da glicose com ácido pícrico não forma substâncias que absorvem no espectro visível.<sup>114</sup>



**Figura 4.5.** Monitoramento multivariado em dois modos para uma das reações do grupo 6 tabela 4.2.

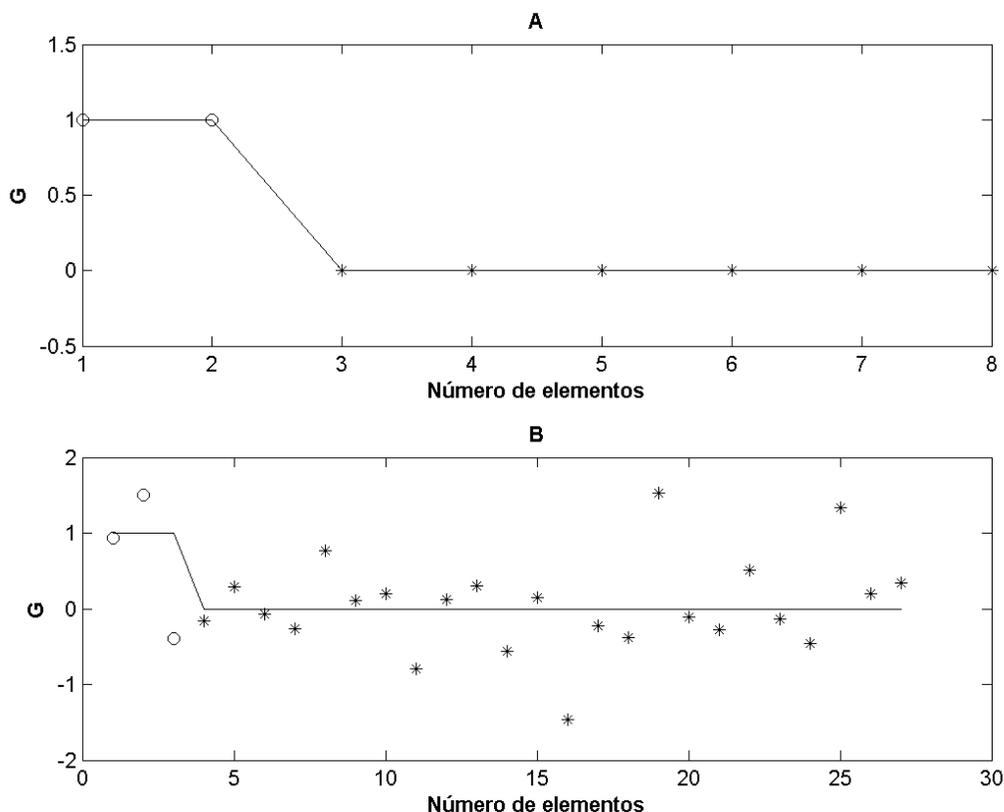
Para a análise de decomposição trilinear, foram usadas somente as medidas feitas no intervalo de comprimentos de onda entre 480 e 600 nm. Abaixo de 480 nm há uma

contribuição residual das medidas de absorvância do ácido pícrico que é descontado do espectro da reação pelo branco. No modo de monitoramento cinético, foram usadas somente as medidas feitas entre 22 s e 300 s. Abaixo de 20 s as medidas apresentam uma variabilidade um pouco maior devido à movimentação do líquido dentro da cubeta, originado do processo de mistura dos reagentes.

A figura 4.6 mostra os gráficos “CORCONDIA”, calculados para dois modelos trilineares distintos, um com duas e outro com três componentes. Estes gráficos referem-se à análise de uma amostra do grupo 6, tabela 4.2. O melhor modelo, é sem dúvida aquele com duas componentes. Resultados iguais a este foram obtidos para a análise das outras quarenta e cinco amostras da tabela 4.2.

Quase todos os modelos com três componentes que foram calculados resolveram pelo menos duas componentes idênticas porém com sinais opostos, além de que foram necessárias mais de 3000 iterações do algoritmo do método PARAFAC para convergir para uma solução. O comportamento observado para modelos com três componentes é típico de uma solução degenerada<sup>115</sup> e não é adequado para as análises quantitativas.

Para os modelos com duas componentes a convergência foi sempre obtida antes de 200 iterações recuperando sempre em torno de 98% da variação dos dados. Todas estas evidências sugerem que modelo trilinear com duas componentes é o que melhor descreve a variação da absorvância em função do tempo para os 45 experimentos da tabela 4.2.



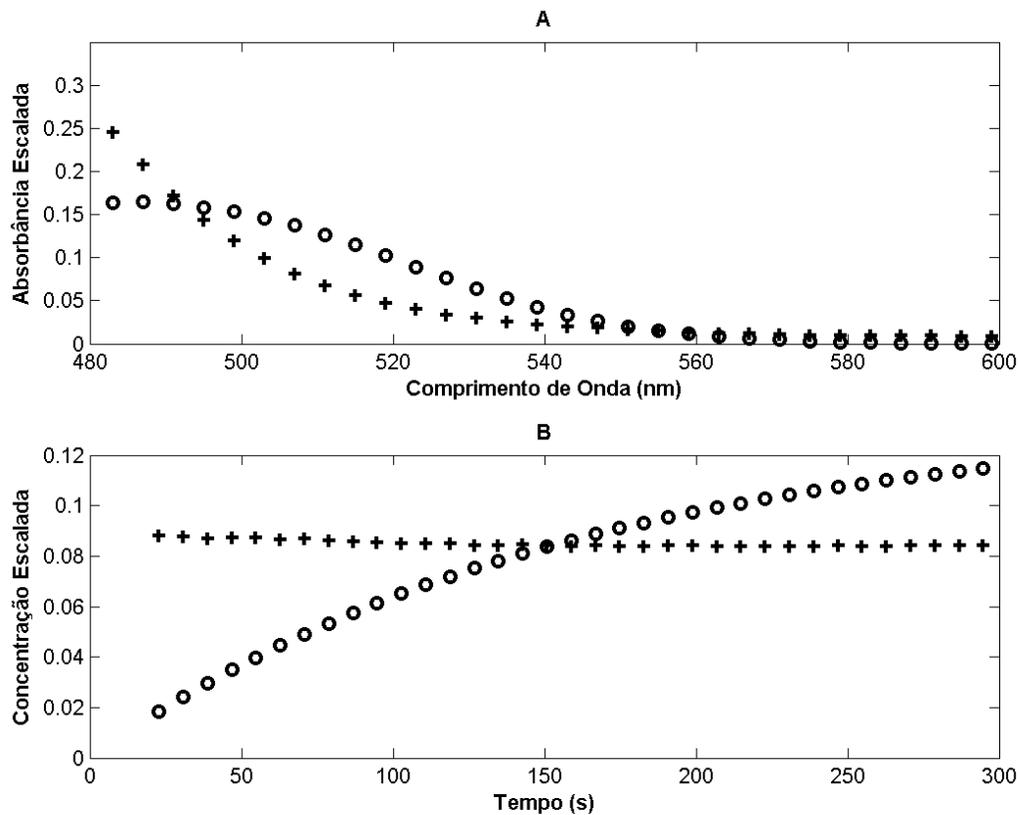
**Figura 4.6.** Gráficos do “CORCONDIA” calculados a partir dos parâmetros de modelos PARAFAC de duas (A) e três componentes (B) obtidos para a análise de um experimento do grupo 6, Tabela 4.2.

A figura 4.7, mostra as duas componentes resolvidas com o método PARAFAC para a análise de uma das amostras do grupo 7 da tabela 4.2. Uma destas componentes resolveu muito bem os sinais do composto formado da reação da creatinina com o ácido pícrico como mostra a figura 4.8. A outra componente resolveu um espectro que é similar ao espectro do produto da reação da albumina com o ácido pícrico e do espectro da bilirrubina, figura 4.9. Resultados similares a estes foram sempre obtidos para a análise de todas as outras amostras.

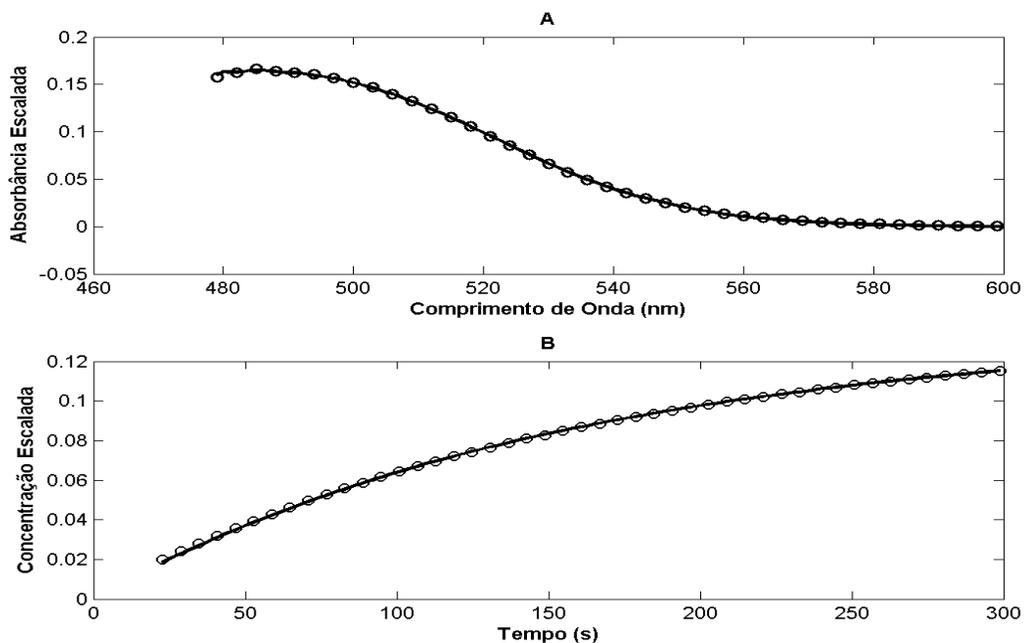
Conforme mostra a figura 4.9, o espectro da bilirrubina é muito parecido com o da albumina. O mesmo foi verificado para o modo de monitoramento cinético. A albumina reage com o ácido pícrico e forma um composto que absorve na região monitorada, entretanto verificou-se que a formação deste composto é muito lenta e o que se observa é um sinal praticamente constante, muito parecido com a da segunda componente da figura

4.7. Como a bilirrubina não reage, o sinal no modo cinético acaba sendo muito parecido com o do produto da reação da albumina.

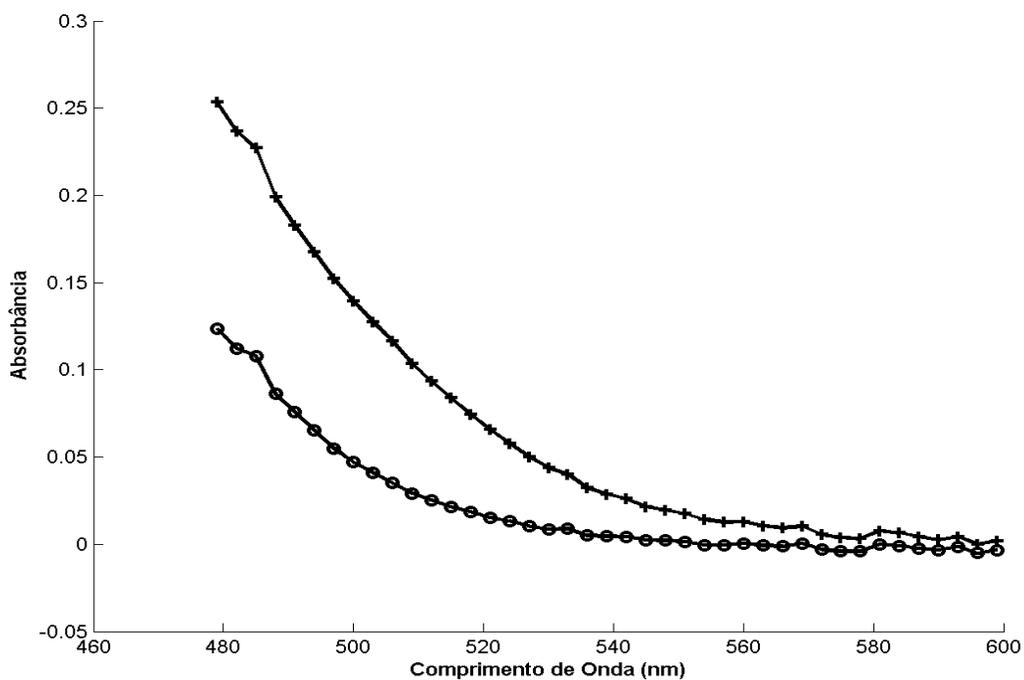
A similaridade dos sinais da albumina e da bilirrubina, explica porque os modelos com três componentes não são capazes de resolver os sinais destas duas substâncias quando ambas estão nas amostras. As componentes que deveriam resolver os sinais de cada uma destes dois compostos formam vetores colineares o que leva a uma solução degenerada durante a decomposição trilinear.



**Figura 4.7.** Parâmetros do modelo trilinear com duas componentes (“o” primeira componente, “+” segunda componente). A) modo espectroscópico; B) modo cinético.



**Figura 4.8.** Comparação dos sinais verdadeiros do produto da reação da creatinina com ácido pícrico (-), com a primeira componente resolvida com o método PARAFAC (o). A) modo espectroscópico; B) modo cinético.



**Figura 4.9.** Espectros das substâncias Bilirrubina (o) e Albumina (+).

A tabela 4.4 mostra os resultados obtidos para a previsão da concentração da creatinina nos diferentes grupos de amostras da tabela 4.2. Os resultados apresentados nesta tabela referem-se às previsões feitas com os modelos de duas componentes. Os resultados obtidos com o método PARAFAC restrito indicados na tabela 4.4, correspondem ao modelo em que as concentrações dos padrões foram fixadas na matriz **C**.

**Tabela 4.4.** Erro de previsão “RMSEP” para a análise dos experimentos da tabela 4.2.

Grupo	TLD	PARAFAC	PARAFAC restrito
1	0,53	0,53	0,36
2	0,80	0,78	0,44
3	0,52	0,48	0,52
4	0,41	0,40	0,15
5	0,50	0,49	0,26
6	0,53	0,55	0,56
7	0,55	0,50	0,77
8	0,47	0,47	0,38
9	1,52	1,55	1,33

De acordo com a tabela 4.4, os três métodos de decomposição trilinear apresentaram erros de previsão semelhantes, exceto que o método PARAFAC restrito fez previsões bem melhores para as amostras dos grupos 1, 2, 4, 5 e 8. Este resultado era esperado visto que as concentrações dos padrões foram fixadas no cálculo com o objetivo de melhorar as estimativas dos sinais do produto da reação da creatinina.

Se o comportamento de cada reação química seguisse perfeitamente o comportamento trilinear, provavelmente não haveria muita diferença para as previsões feitas com os modelos ajustados com os métodos DTLT e PARAFAC em relação às previsões feitas com o modelo ajustado com o método PARAFAC restrito. As condições experimentais para a reação da creatinina com ácido pícrico foram acertadas para que as suposições I e III do método espectro-cinético fossem verdadeiras (tópico 4.4). Entretanto a suposição II, assume que a constante da velocidade de reação é igual em todos os experimentos, mas como mostram os resultados da tabela 4.3, estas constantes não são

exatamente iguais para a mesma reação executada mais de uma vez devido a fatores que não podem ser controlados no experimento.

Estas diferenças das constantes de velocidade da reação entre os experimentos caracterizam desvios de trilinearidade que invariavelmente tem uma influência na exatidão e na precisão dos sinais resolvidos. Apesar disto, este problema não foi tão grave visto que o erro de previsão para o valor de concentração de creatinina nas amostras foi considerado pequeno e plenamente satisfatório.

O RMSEP não foi igual para os diferentes grupos da Tabela 4.2, de modo que a composição das amostras tem alguma influência na precisão das estimativas dos parâmetros do modelo trilinear. Foi possível verificar que o erro de previsão aumenta com o aumento do valor da concentração da albumina na amostra. Este aumento do erro de previsão parece indicar que a concentração da albumina tem alguma influência na reação da creatinina com o ácido pícrico. Este problema foi observado por Falcó que investigou a variação da constante de velocidade em função da concentração da albumina e as conseqüências deste efeito para as determinações com o método cinético com dois pontos de leitura fixos.<sup>116</sup>

O maior erro de previsão obtido foi para as amostras do grupo 9, tabela 4.4. As amostras deste grupo contem ácido aceto acético e o produto da reação desta substância com ácido pícrico, apresenta um espectro de absorção molecular muito parecido com produto da reação da creatinina com ácido pícrico. Embora as velocidades de reações da creatinina e do ácido aceto-acético com ácido pícrico sejam muito diferentes, os modelos trilineares de duas componentes não foram capazes de separar os sinais dessas duas substâncias.

Tal como o método cinético de Jaffé, o ácido aceto-acético interfere na determinação da creatinina pelo método proposto nesta tese. Vale ressaltar que a concentração de ácido aceto-acético presente nas amostras do grupo 9 é bem superior a concentração normal desta substância no soro de pessoas adultas.

### 4.6.3 Determinação da creatinina em soro humano

A metodologia proposta neste trabalho de tese foi testada em seis amostras de soro humano obtidas no Hospital de Clínicas da UNICAMP. Estas amostras foram analisadas no laboratório do hospital que emprega o método cinético de Jaffé com leitura em dois pontos fixos para as determinações quantitativas. O método é automático e apresenta um coeficiente de variação de 2,3 % para as determinações de creatinina no soro humano presente em níveis de concentração de  $16,7 \text{ mg L}^{-1}$ . Os valores de concentração obtidos por este método foram usados como uma referência para avaliar a metodologia proposta.

A metodologia usada para as determinações de creatinina em soro humano foi a mesma usada para a determinação nas misturas do tópico anterior. Exceto para a amostra 5, todos os outros modelos trilineares foram calculados com duas componentes. Este resultado mostra que o comportamento das amostras de soro é similar ao das misturas testadas no tópico anterior. O resultado da previsão das concentrações com os cada um dos métodos de decomposição trilinear são apresentados na tabela 4.5.

**Tabela 4.5.** Previsão da concentração de creatinina em amostras de soro.

Amostras	Valor de Referência $\text{mg L}^{-1}$	DTLD $\text{mg L}^{-1}$	PARAFAC $\text{mg L}^{-1}$	PARAFAC restrito $\text{mg L}^{-1}$
1	7,30	5,89	6,04	7,10
2	74,00	76,04	77,14	79,18
3	9,80	7,47	7,60	8,02
4	16,90	15,64	16,24	16,00
5	97,20	100,12	99,82	97,60
6	11,90	10,86	10,82	10,96
SEP*		2,12	2,22	2,51
CV**		6,16	6,13	6,87

\* Equação 4.6. \*\* Equação 4.7.

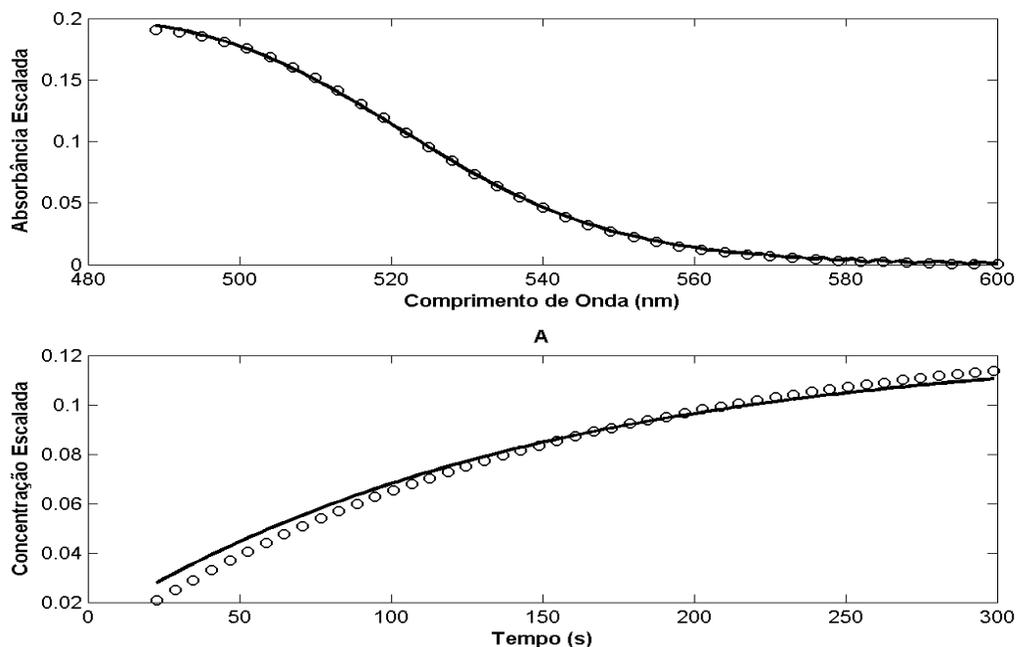
O coeficiente de variação em torno de 6% para ambos os métodos de decomposição trilinear é razoável para a determinação da creatinina em soro humano pois o critério que é aceito para esta variação é abaixo de 10%. Deve-se enfatizar que o método usado como referência não é um método padrão para a determinação de creatinina embora seja um método válido.

Comparando os valores previstos por cada método de decomposição com o valor previsto pelo método de referência, verifica-se que o método PARAFAC restrito só não fez uma previsão melhor para as amostras 2 e 4. Na verdade, o erro de previsão para a amostra 2 com o método PARAFAC restrito foi excepcionalmente alto, e se essa amostra fosse excluída do cálculo do SEP e do CV, obteria-se um SEP de  $0,82 \text{ mg L}^{-1}$  e um CV de 2,95 % para as previsões feitas com o método PARAFAC restrito contra um SEP de  $1,84 \text{ mg L}^{-1}$  e CV de 6,55 % para o método PARAFAC e um SEP de  $2,04 \text{ mg L}^{-1}$  e CV de 7,29% para o método DTLD. Estes resultados mostram que o método PARAFAC restrito apresentou o melhor desempenho comparado aos outros dois métodos para a maioria das amostras, embora o SEP e o CV desse método seja similar ao dos outros dois métodos quando todas as amostras são consideradas.

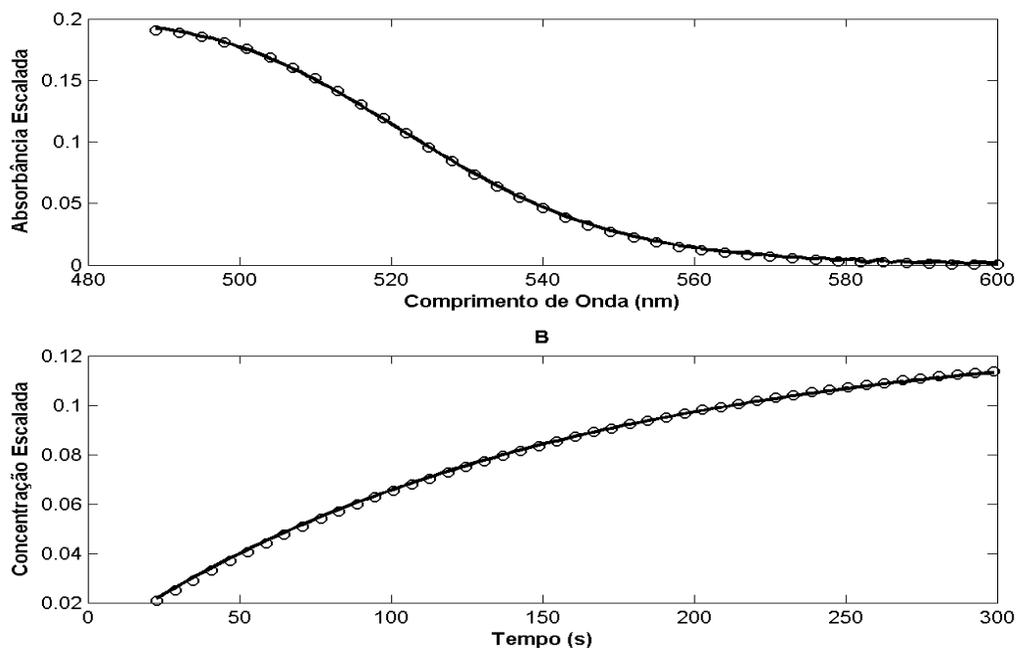
A análise da amostra 5 apresentou problemas, visto que tanto o modelo com duas como o modelo com três componentes apresentaram uma solução degenerada. As figuras 4.10 e 4.11 fazem respectivamente uma comparação dos sinais resolvidos pelo método DTLD para os modelos com duas e três componentes, com os sinais “puros” do produto da reação creatinina com ácido pícrico. No modo espectroscópico ambos modelos recuperam bem o espectro do produto da reação, porém, no modo cinético a estimativa foi melhor para o modelo com três componentes. Conseqüentemente, o modelo com três componentes foi escolhido para a análise quantitativa.

Vale notar que a amostra 5 apresenta um comportamento diferente das demais amostras, pois a concentração de creatinina nesta amostra está muito acima do normal ( $10 \text{ mg L}^{-1}$  para adultos) e caracteriza uma pessoa com sérios problemas renais. É muito provável que nesta amostra de soro, os valores normais de outras substâncias também estejam alterados, e alguma destas substâncias provavelmente está contribuindo para as medidas de absorbância e os sinais não foram separados do sinal da creatinina como no

caso verificado para as misturas contendo ácido aceto-acético. Apesar disto, a recuperação do valor de concentração da creatinina na amostra foi boa.



**Figura 4.10.** Parâmetros recuperados pelo modelo DTLTD com duas componentes(o), e os sinais “puros” para o produto da reação da creatinina (-).



**Figura 4.11.** Parâmetros recuperados pelo modelo DTLTD com três componentes(o), e os sinais esperados para o produto da reação da creatinina (-).

## 4.7 CONCLUSÃO

A metodologia proposta foi capaz de determinar a concentração da creatinina no soro humano com um erro dentro do tolerado em análises clínicas. O método PARAFAC restrito apresentou um excelente desempenho para a exatidão das previsões. Entretanto o método DTLTD também pode ser usado sem comprometer a exatidão das previsões e tem a vantagem de que é um procedimento de cálculo mais rápido que o procedimento iterativo do método PARAFAC.

Trabalhos futuros poderiam explorar melhor as variáveis que alteram a velocidade da reação e escolher uma condição que leve a um tempo de análise menor do que o proposto nesta tese. O planejamento ortogonal de experimentos usado neste trabalho para investigar a reação da creatinina com o ácido pícrico mostrou que a concentração de hidróxido de sódio no reagente alcalino picrato é uma variável que pode ser modificada com o objetivo de reduzir o tempo de análise.

## C. MATERIAL SUPLEMENTAR: PARTE C.

### C.1 Diagnostico de Consistência do Núcleo (CORCONDIA)

A estrutura do modelo trilinear pode ser representada da seguinte forma:

$$\mathbf{X}_{I \times JK} = \mathbf{A}(\mathbf{C} \otimes \mathbf{B})^T \quad 4.7$$

A matriz  $\mathbf{I}$  é uma superdiagonal de dimensões  $R \times RR$ . Os elementos  $i_{111}, i_{222}, \dots, i_{RRR}$ , desta matriz são iguais a 1 e os demais elementos são iguais a zero. Se as matrizes  $\mathbf{A}$ ,  $\mathbf{B}$  e  $\mathbf{C}$ , resolvidas com um método de decomposição trilinear descreverem corretamente uma estrutura trilinear, então a solução do problema dos mínimos quadrados definidos na equação 4.8, deve resultar em uma matriz  $\mathbf{G}$  que é igual a matriz  $\mathbf{I}$  da equação 4.7.

$$\min_{\mathbf{G}} = \|\mathbf{X} - \mathbf{A}\mathbf{G}(\mathbf{B} \otimes \mathbf{C})^T\|_2^2 \quad 4.8$$

O parâmetro CORCONDIA é calculado de acordo com a equação 4.9. Um Valor acima de 90 % para o CORCONDIA indica uma estrutura trilinear, abaixo de 90 %, o modelo apresenta desvios, enquanto que para valores negativos o modelo trilinear é inadequado para descrever a estrutura dos dados.

$$\text{CORCONDIA} = 100 \left( 1 - \frac{\sum_{r=1}^R \sum_{r=1}^R \sum_{r=1}^R (g_{nnn} - i_{nnn})^2}{R} \right) \quad 4.9$$

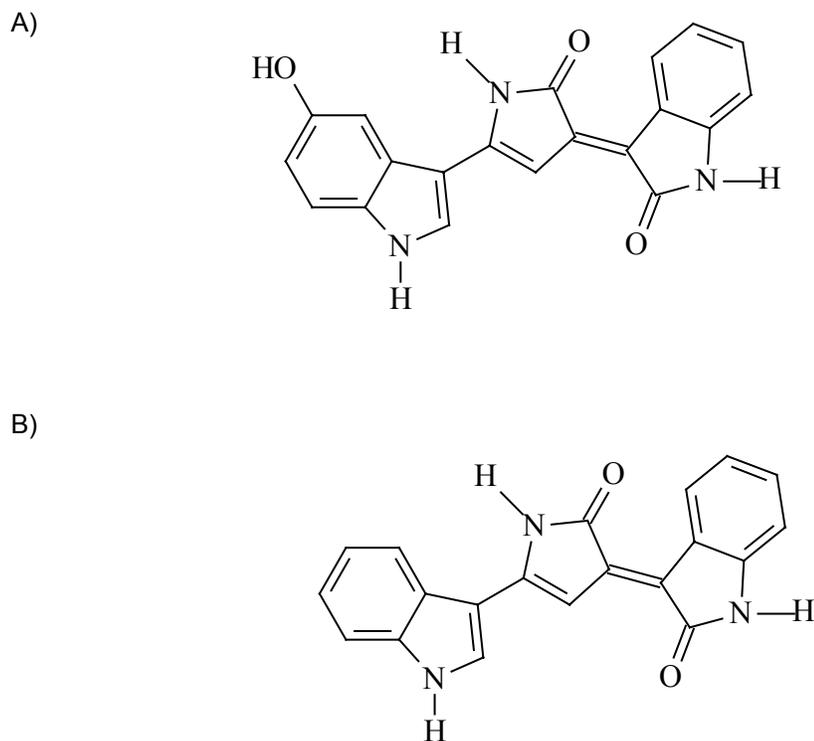
O gráfico CORCONDIA é construído a partir da disposição dos elementos da matriz  $\mathbf{G}$  em função do número de elementos que esta matriz apresenta. Os elementos da superdiagonal devem ser iguais a um e os demais elementos iguais a zero, para que o modelo apresente um bom ajuste à estrutura trilinear.

## **CAPÍTULO 5**

### **ANÁLISE DOS METABÓLITOS FLUORESCENTES PRODUZIDOS PELA *CHROMOBACTERIUM VIOLACEUM* COM O MÉTODO PARAFAC**

## 5.1 INTRODUÇÃO

A *Chromobacterium violaceum* é uma bactéria gram negativa, aeróbica facultativa, encontrada nos solos e nas águas de regiões tropicais e subtropicais do planeta. Este microorganismo faz parte do ecossistema do Rio Negro (bacia do Rio Amazonas) e contribui para a coloração negra das águas porque produz os pigmentos violetas violaceína e deoxiviolaceína (figura 5.1).<sup>117</sup> A violaceína é um microbiocida e quando irradiada com luz visível, decompõe-se em outros pigmentos que também são tóxicos para a microfauna do rio.<sup>118</sup> Por causa desta propriedade fotoquímica, a violaceína já foi apontada como uma das causas da baixa densidade de biomassa do Rio Negro porque este pigmento seria responsável pela mortalidade dos microorganismos que são a principal fonte de alimento dos peixes e insetos.



**Figura 5.1.** Fórmulas estruturais dos pigmentos violetas produzidos pela *C. violaceum*: A) violaceína; B) deoxiviolaceína.

Diversos ensaios realizados *in vitro* demonstraram que a violaceína (figura 5.1A) apresenta atividade antibiótica,<sup>118</sup> antitumoral<sup>119</sup> e antiviral,<sup>120</sup> além de ser letal para protozoários causadores da doença de Chagas<sup>121</sup> e Leishmaniose.<sup>122</sup> Este largo espectro de atividade biológica da violaceína atraiu a atenção dos pesquisadores interessados no desenvolvimento de novos fármacos, levando-os a desenvolver métodos para a produção, extração e purificação do pigmento violeta que pode ser usado também como corante de produtos têxteis.<sup>123</sup>

A *Chromobacterium violaceum* produz vários metabólitos de interesse para a indústria farmacêutica<sup>124,125,126</sup> como o peptídeo FR901228 que está sendo testado para o tratamento do câncer,<sup>127</sup> e o antibiótico aerocavin.<sup>128</sup> Em condições anaeróbicas e num meio de fermentação composto por glicerina como fonte de carbono, a *C. violaceum* produz ácido cianídrico,<sup>129,130,131</sup> de modo que pode ser usada para extração de metais pesados dos minérios e de solos contaminados. A *C. violaceum* acumula polímeros biodegradáveis que apresentam propriedades físico-químicas semelhantes às do polietileno.<sup>132</sup> Em meios de cultivo contendo como única fonte de carbono o ácido valérico, a *C. violaceum* produz o homopoliéster do ácido 3-hidroxisovalérico (3HV). Steinbuchel desenvolveu um processo para a produção do ácido poli(3-hidroxisovalerato) em bioreatores que produz 40 g de massa seca por litro de fermentação.<sup>133</sup>

Devido ao potencial biotecnológico da *Chromobacterium violaceum*, o governo brasileiro financiou um consórcio de laboratórios de âmbito nacional para fazer o seqüenciamento do genoma desta bactéria, finalizado no ano de 2003.<sup>134</sup>

A violaceína é um dos metabólitos mais investigados da *C. violaceum*, dentre os que apresentam características econômicas desejáveis e a biossíntese desta molécula têm sido estudada através de experimentos de laboratório e da análise do genoma desta bactéria. De Moss foi o primeiro pesquisador a propor que a violaceína é sintetizada a partir do L-triptofano em presença de oxigênio molecular,<sup>135</sup> tendo este processo sido desvendado através da marcação dos átomos de carbono da molécula de triptofano com o isótopo <sup>14</sup>C. Utilizando a mesma técnica de marcação com isótopos dos átomos da molécula de triptofano, Hoshino demonstrou que todos os átomos de nitrogênio, hidrogênio e carbono presentes na violaceína são oriundos da molécula de L-triptofano.<sup>136</sup>

Um entendimento mais detalhado da biossíntese da violaceína poderia ser obtido se fossem empregadas técnicas instrumentais capazes de detectar os metabólitos intracelulares da bactéria durante o curso das transformações químicas que ocorrem no bioreator. Esta informação analítica poderia ser útil não só para investigar a biossíntese do pigmento mas, também, para otimizar e controlar a produção da violaceína em tempo real.

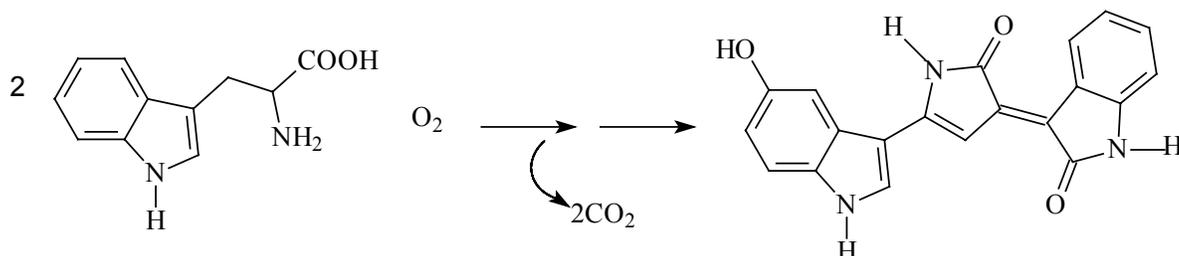
Diversas moléculas que apresentam importante papel no metabolismo dos microorganismos possuem fluorescência própria (autofluorescentes) e podem ser detectadas diretamente após excitadas com a radiação de comprimentos de onda da região ultra violeta do espectro eletromagnético. Como o triptofano e a violaceína são moléculas autofluorescentes, a espectroscopia de fluorescência poderia ser usada para a investigar e monitorar a biossíntese da violaceína.

Neste trabalho propomos a utilização da espectroscopia de fluorescência para monitorar os metabólitos intracelulares da *C. violaceum* durante a fase de crescimento exponencial quando cultivada em meio aeróbico. O método PARAFAC foi usado para resolver os espectros de fluorescência de uma mistura de fluoróforos presentes nas amostras de biomassa coletadas do bioreator. Os resultados obtidos neste trabalho mostram que quatro fluoróforos: triptofano, tirosina, violaceína e a coenzima NADH podem ser detectados, da mesma forma que foi verificado que a bactéria consome o aminoácido triptofano algumas horas antes de começar a produção progressiva do pigmento violaceína, concordando com o mecanismo proposto para a biossíntese desta molécula.

## 5.2 BIOSÍNTESE DA VIOLACEÍNA

A rota metabólica para a produção da violaceína tem sido extensivamente investigada por Hoshino.<sup>137,136</sup> Antes deste, DeMoss demonstrou que a biossíntese da violaceína ocorre a partir de duas moléculas de *L*-triptofano e que o oxigênio molecular é essencial para a formação do pigmento pois em condições de anaerobiose a bactéria não produz a violaceína.<sup>135</sup> Utilizando átomos de <sup>14</sup>C em diferentes posições da molécula de *L*-triptofano, DeMoss percebeu que apenas o carbono carboxílico desta molécula é incorporado na violaceína, deduzindo que para a biossíntese desta molécula ocorre um processo de descarboxilação (figura 5.2). De Moss notou, também, que a quantidade de

pigmento formada é proporcional a quantidade de L-triptofano presente no meio, da mesma forma que a velocidade de síntese da violaceína é diretamente proporcional à concentração de células.<sup>138</sup>

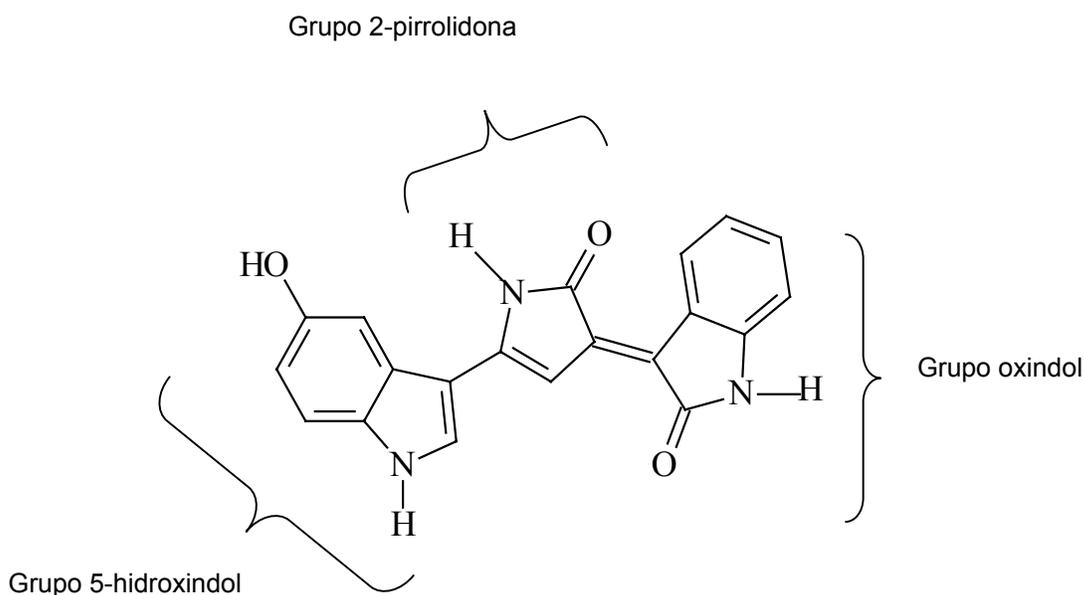


**Figura 5.2.** Esquema da biossíntese da violaceína.

A violaceína é formada por três unidades estruturais: grupo 2-pirrolidona (2PY), grupo 5-hidroxiindol (5HI) e o grupo oxindol (OX), conforme mostra a figura 5.3. Os trabalhos de Hoshino demonstraram que todos os átomos de hidrogênio, nitrogênio e carbono desta substância são provenientes de duas moléculas de triptofano e que o átomo de oxigênio é incorporado a partir do oxigênio molecular. Parte do triptofano presente no meio é convertido também em deoxiviolaceína. Nos estudos conduzidos por Rettori, foi encontrada uma proporção 9:1 de violaceína e deoxiviolaceína respectivamente no extrato obtido após extração com etanol absoluto.<sup>139</sup>

Para a incorporação dos átomos de oxigênio a *C. violaceum* utiliza uma enzima oxigenase. Esta enzima foi isolada e usada para produzir a violaceína na presença do cofator NADH a partir de um precursor da biosíntese do triptofano, sem a presença das células da bactéria.<sup>140</sup> A violaceína pode ser produzida a partir do ácido indol 3-acético, um metabólito precursor do L-triptofano,<sup>121</sup> sendo que neste caso a deoxiviolaceína não é sintetizada.

O papel biológico da violaceína para a bactéria ainda não é conhecido. Há uma hipótese de que o pigmento seja produzido pela *C. violaceum* para protegê-la do estresse oxidativo devido a sua propriedade antioxidante,<sup>141</sup> ou da luz ultra violeta visto que a violaceína tem um alto coeficiente de absorvidade molar e habita principalmente as regiões tropicais do planeta.



**Figura 5.3.** Unidades estruturais da molécula da violaceína.

Trabalhos mais recentes conduzidos por pesquisadores do Instituto de Química da Universidade Estadual de Campinas (UNICAMP) buscaram otimizar a produção do pigmento violaceína. Usando a técnica de planejamento fatorial de experimentos e superfície de respostas, Mendes foi capaz de aumentar a produção de violaceína de  $0,17 \text{ g L}^{-1}$  para  $0,43 \text{ g L}^{-1}$  alterando a composição química do meio de cultivo.<sup>142</sup> Além disso, Rettori desenvolveu uma técnica para extração e purificação da violaceína.<sup>139</sup> Através do monitoramento da produção de violaceína em bioreatores com microcalorimetria em fluxo, Oliveira<sup>143</sup> observou que o pico de produção de calor ocorre antes da fase de crescimento exponencial da bactéria e que este processo é dependente do tipo de carboidrato utilizado como fonte carbono.

### 5.3 O PROCESSO DE LUMINESCÊNCIA

A Luminescência<sup>144</sup> é um fenômeno caracterizado pela emissão de fótons por parte das moléculas como resultado de uma transição eletrônica. Caso os fótons emitidos sejam da região do espectro visível, o fenômeno de luminescência pode ser detectado como um fenômeno luminoso. As transições eletrônicas têm origem na absorção de energia pelas

moléculas, e esta energia pode ser fornecida de diferentes formas na qual caracterizam os diferentes tipos de luminescência:

1. Eletroluminescência: energia fornecida pela passagem de corrente elétrica;
2. Triboluminescência: energia fornecida pela fricção;
3. Quimioluminescência: energia fornecida por uma reação química;
4. Fotoluminescência: energia fornecida pela absorção de fótons;

No processo fotoluminescente uma molécula absorve energia na forma de um fóton pela interação com a radiação eletromagnética e passa do seu estado fundamental  $S_0$  para um estado excitado  $S_1$ . Como o estado excitado é termodinamicamente instável, a molécula retorna ao seu estado fundamental após dissipar a energia recebida. A energia de excitação pode ser dissipada na forma energia térmica, na forma de energia química a partir de uma reação da molécula no seu estado excitado ou pela liberação de um fóton.

A liberação de um fóton para que a molécula retorne do estado excitado ao seu estado fundamental é o processo pela qual caracteriza a fluorescência e a fosforescência. Na fluorescência, o elétron que participa da transição eletrônica tem a mesma multiplicidade de spin nos estados fundamental e excitado e o processo de retorno ao estado fundamental é bastante rápido, da ordem de  $10^{-8}$  s. Na fosforescência a multiplicidade do spin no estado fundamental é diferente do estado excitado e antes de retornar para o estado fundamental deve ocorrer uma inversão do spin, processo este que pode levar minutos ou horas. Para quem observa ambos os fenômenos, a fluorescência se manifesta enquanto as moléculas forem submetidas à fonte luminosa e encerra quando a fonte for removida, já a fosforescência continua a se manifestar mesmo após a fonte ter sido removida por um determinado período de tempo em consequência do tempo necessário para converter os spins.

## 5.4 APLICAÇÕES DA FLUORESCÊNCIA EM BIOTECNOLOGIA

A espectroscopia de fluorescência é utilizada para o monitoramento de substâncias orgânicas que desempenham funções biológicas muito importantes nos seres vivos. Esta

técnica exibe melhor seletividade e sensibilidade quando comparada às técnicas de espectroscopia de infravermelho, ultravioleta e visível. A tabela 5.1, mostra os comprimentos de onda característicos de excitação e emissão de algumas substâncias autofluorescentes e que são de interesse biológico.

**Tabela 5.1.** Comprimentos de onda característicos de emissão e excitação de algumas substâncias autofluorescentes.<sup>145</sup>

Substância	Excitação $\lambda_{\max}$ (nm)	Emissão $\lambda_{\max}$ (nm)	Solvente
Fenilalanina	258	284	água
Tirosina	276	302	água
Triptofano	280	357	água
Vitamina A	346	480	Iso octano
Vitamina B2	270	518	água
Vitamina B6	328	393	água
Vitamina E	298	326	Iso octano
NADH	344	465	água
ATP	292	388	água
Clorofila A	428	663	Iso octano
Hematoporfirina	396	614	Iso octano

A técnica de microscopia ótica de fluorescência é uma importante ferramenta para o monitoramento dos processos biológicos através de imagens que identificam a posição dos metabólitos intracelulares dentro das células. Esta técnica é usada para investigar a interação de moléculas pequenas com proteínas nas células.<sup>146</sup> Nas análises clínicas, a fluorescência é utilizada para detectar diversas substâncias intracelulares sendo esta a base de vários testes imunológicos,<sup>147</sup> e para a análise da expressão de genes em microarranjos de DNA.<sup>148</sup>

O NADH e o triptofano são substâncias autofluorescentes usadas como alvo no monitoramento *in situ* de bioreatores.<sup>149,150,151</sup> A excitação da molécula de NADH no comprimento de onda de 366 nm e a detecção do respectivo sinal de emissão em 460 nm, foram usados por Zabrizkie e Humprey para medir a concentração de biomassa dos

microorganismos *Saccharomyces cerevisiae*, *Streptomyces* e *Thermoactiomyces* durante o cultivo em bioreatores.<sup>149</sup>

Horvath e colaboradores encontraram uma boa correlação entre a fluorescência de triptofano e a concentração de biomassa de *Saccharomyces cerevisiae* e comparou este sinal com o do NADH.<sup>150</sup> De acordo com este pesquisador, o sinal de emissão do triptofano entre 320 e 360 nm é duas vezes mais intenso do que a emissão NADH no intervalo entre 440 e 480 nm, além de que a emissão fluorescente do triptofano sofre menor influência das condições físicas do meio como a variação agitação e concentração de oxigênio.

O monitoramento de bioreatores através da medida de fluorescência feita em um único comprimento de onda é uma aplicação bastante limitada, visto que os meios de cultivo são formados por um grande número de substâncias autofluorescentes que podem ter seus espectros de emissão sobrepostos entre si. Trabalhos mais recentes têm explorado o monitoramento de bioreatores usando medidas feitas em vários comprimentos de excitação, formando um arranjo de dados com três modos de monitoramento como mostra a figura 5.4, devido a aquisição simultânea de um espectro excitação e emissão.<sup>152,153,154</sup>

O monitoramento com espectroscopia de fluorescência em dois modos excitação e emissão, têm sido usado para a detecção simultânea aminoácidos, proteínas, cofatores enzimáticos, antibióticos, substâncias aromáticas e vitaminas.<sup>155,156</sup> Neste tipo de aplicação, são utilizados os métodos de análise multivariada como o método de regressão dos mínimos quadrados parciais (PLS),<sup>157</sup> redes neurais<sup>158</sup> e o método PARAFAC.<sup>159</sup>

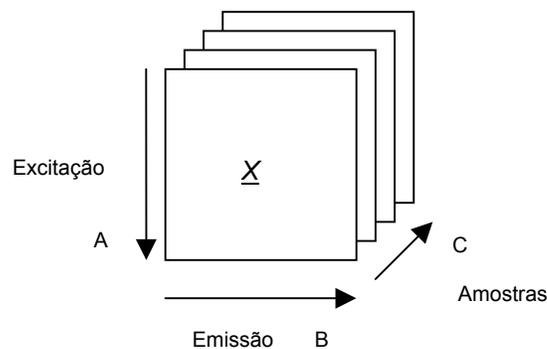
## **5.5 ANÁLISE DOS ESPECTROS DE FLUORESCÊNCIA COM OS MÉTODOS DE DECOMPOSIÇÃO TRILINEAR**

Os dados de fluorescência podem ser organizados em uma matriz **X**, denominada de matriz de excitação e emissão, onde cada linha desta matriz é um espectro de emissão do comprimento de onda de excitação *i* e cada coluna é um espectro de excitação do comprimento de onda de emissão *j*. Para um conjunto de amostras, os dados podem ser organizados em um arranjo de três modos, como mostra a figura 5.4, onde os modos A, B e C são respectivamente os modos de excitação, emissão e das amostras.

Assumindo que a concentração de cada espécie fluorescente é proporcional à variação da intensidade de radiação, podemos usar um modelo trilinear como o da equação 5.1 para descrever a variação das medidas de fluorescência no arranjo  $\underline{\mathbf{X}}$ .

$$\mathbf{X}_{K \times I \times J} = \mathbf{K}(\mathbf{F} | \otimes | \mathbf{S})^T \quad 5.1$$

Na equação 5.1, o vetor coluna  $\mathbf{s}_r$  da matriz  $\mathbf{S}$  é o espectro de excitação da substância  $r$ , o vetor coluna  $\mathbf{f}_r$  da matriz  $\mathbf{F}$  é o espectro de emissão e o vetor coluna  $\mathbf{k}_r$  da matriz  $\mathbf{K}$  informa a concentração da substância  $r$  nas amostras que compõe o arranjo  $\underline{\mathbf{X}}$ .



**Figura 5.4.** Arranjo em três modos  $\underline{\mathbf{X}}$  formado por um conjunto de matriz de excitação e emissão.

O modelo trilinear da equação 5.1 pode ser ajustado com o método PARAFAC para obter o modelo mostrado abaixo:

$$\mathbf{X}_{K \times I \times J} = \mathbf{C}(\mathbf{B} | \otimes | \mathbf{A})^T + \mathbf{E} \quad 5.2$$

onde  $\mathbf{A}$ ,  $\mathbf{B}$  e  $\mathbf{C}$  são as respectivas estimativas dos parâmetros do modelo mecanístico  $\mathbf{K}$ ,  $\mathbf{F}$  e  $\mathbf{S}$  definido na equação 5.1 e a matriz  $\mathbf{E}$  é a parte do arranjo  $\underline{\mathbf{X}}$  que não pode ser modelada pela estrutura trilinear do modelo 5.2.

O método PARAFAC é uma ferramenta extremamente importante para a análise de matrizes de excitação e emissão, entretanto, é necessário ter alguns cuidados quando este método é usado para a análise dos dados de fluorescência. Alguns elementos

presentes na matriz de excitação e emissão podem não seguir o comportamento trilinear descrito pela equação 5.1, como os picos de espalhamento Rayleigh e Raman.<sup>144</sup>

Antes de decompor o arranjo  $X$  com o método PARAFAC é necessário fazer um pré- tratamento nos dados para eliminar os picos associados ao espalhamento Rayleigh e Raman. A forma mais usada é substituir estes picos por valores ausentes,<sup>160,161</sup> onde os dados são simplesmente removidos, não sendo adicionado valor algum em seus lugares. Para análise de um arranjo contendo valores ausentes é necessário modificar o algoritmo ALS, pois além de uma estimativa dos parâmetros é necessário também fazer uma estimativa para os valores que faltam. Uma discussão sobre como tratar um conjunto de dados com valores ausentes é apresentada por Walczak e Massart.<sup>162,163</sup>

Outra forma de eliminar os dados da matriz de excitação e emissão que não seguem um modelo trilinear é substituir estes dados por valores iguais a zero.<sup>164,165</sup> Este procedimento é prático e pode ser usado com o algoritmo ALS convencional, todavia, este procedimento deve ser executado com cautela visto que o procedimento de colocar valores iguais a zero pode alterar a estrutura de correlação das variáveis da matriz de excitação e emissão.<sup>160</sup> Outra forma de resolver o problema é utilizar o método de ajuste ponderado do PARAFAC, de modo que são adicionados pesos pequenos para a região da matriz de excitação e emissão que contém os picos de espalhamento Rayleigh enquanto que para a região contendo os espectros de fluorescência das substâncias é dado um peso maior no ajuste do modelo.<sup>166</sup>

## 5.6 PARTE EXPERIMENTAL

### 5.6.1 Experimentos conduzidos no bioreator para a produção de violaceína pela *Chromobacterium violaceum*

A fermentação foi realizada em um bioreator BIOSTAT B2 (B. Braun Biotech). O meio de cultura foi preparado por dissolução de 7,5 g de glicose, 7,5 g de peptona bacteriológica e 3.2 g de extrato de levedura em 1,5 L de água. O meio de cultura foi esterilizado em autoclave durante 30 minutos a uma temperatura de 121° C e 1 Kgf/cm<sup>2</sup>. Após a esterilização o reator foi mantido em repouso até atingir a temperatura de 30 ° C.

Uma suspensão da *C. violaceum*, contendo aproximadamente  $1,0 \times 10^{11}$  bactérias por  $\text{dm}^3$  em uma solução salina de polifosfatos tamponada (PBS) 10 %, foi adicionada ao meio de cultura.

Durante a fermentação, a temperatura do bioreator foi mantida constante a  $30^\circ \text{C}$  através de um sistema de controle automático. O fluxo de ar comprimido foi mantido a  $1,5 \text{ L min}^{-1}$  e a velocidade de agitação do meio de cultura em 200 rpm. O pH do meio de cultura não foi controlado. Durante as fermentações foram feitas medidas em tempo real do teor de oxigênio dissolvido e do pH do meio de cultura.

### **5.6.2 Coleta das amostras**

As amostras do meio de cultura foram retiradas com o auxílio de seringas plásticas durante a fase exponencial de crescimento da bactéria em intervalos regulares de 2 horas, sendo que a primeira amostra foi coletada após 8 horas da inoculação e a última amostra foi retirada após 22 horas. No total foram coletadas 8 amostras com um volume de 10 mL.

Cada amostra foi centrifugada e a fração líquida separada da biomassa. Uma alíquota de 3 mL de etanol foi adicionada ao tubo de centrifuga contendo a biomassa (separada do sobrenadante), para dissolução dos metabólitos sob agitação vigorosa, com auxílio da própria seringa. Em seguida, a fase orgânica foi filtrada em um filtro de nitrocelulose com um tamanho de poro de  $0,45 \mu\text{m}$  para a separação do material sólido. O filtrado foi usado para as medidas no espectrofluorímetro.

### **5.6.3 Aquisição dos espectros eletrônicos de excitação e emissão**

Todos os espectros eletrônicos de excitação e emissão foram coletados com um espectrofotômetro de luminescência modelo LS 55 da Perkin Elmer. O espectro de excitação foi registrado entre 200 e 700 nm, em intervalos regulares de 10 nm. Os espectros de emissão foram coletados entre 200 e 800 nm em intervalos regulares de 0,5 nm. A velocidade de aquisição dos espectros foi  $1500 \text{ nm min}^{-1}$  e a largura da fenda usada para a resolução dos espectros foi 5 nm.

## 5.6.4 Tratamento matemático dos dados espectroscópicos

Todos os espectros de excitação e emissão foram gravados em arquivos de extensão ASCII para importação dos dados entre diferentes softwares. O software MATLAB 6.5 (Mathworks) foi utilizado para o tratamento matemático dos dados de espectroscopia.

O método PARAFAC foi usado para decompor o arranjo de matrizes de excitação e emissão (figura 5.4) em um modelo trilinear (equação 5.2). O programa usado neste trabalho é de autoria própria e utiliza o algoritmo ALS modificado para o tratamento dos valores ausentes adicionados na matriz de excitação e emissão. As funções utilizadas para incluir restrições de não negatividade nos modelos foram obtidas da página da internet mantida por Bro,<sup>167</sup> e estas funções foram incorporadas no programa PARAFAC.

### 5.6.4.1 Análise das matrizes de excitação e emissão com o método PARAFAC

Para cada matriz de emissão e excitação foram adicionados valores ausentes nas posições correspondentes aos picos de espalhamento Rayleigh e nos comprimentos de onda de emissão abaixo dos comprimentos de onda de excitação. Este procedimento foi executado pela substituição do dado de intensidade de radiação pelo valor “nan” que é o padrão adotado pelo software MATLAB. O método PARAFAC foi adaptado para o tratamento de dados ausentes de acordo com a metodologia proposta por Walczak.<sup>162</sup>

O número de componentes do modelo trilinear foi escolhido a partir da avaliação simultânea das funções CORCONDIA<sup>168</sup> e porcentagem da variação explicada pelo modelo definida na equação abaixo:

$$\%VAR = 100 \times \left( 1 - \frac{\sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K (x_{ijk} - \hat{x}_{ijk})^2}{\sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K (x_{ijk})^2} \right) \quad 5.3$$

onde  $x_{ijk}$  é a medida de intensidade de radiação fluorescente no comprimento de onda de excitação  $i$ , emissão  $j$  na amostra  $k$ . O elemento  $\hat{x}_{ijk}$  é o valor correspondente a  $x_{ijk}$  predito pelo modelo empírico trilinear (equação 5.2) ajustado com o método PARAFAC. Além disto, os resíduos do modelo foram inspecionados para verificar se há ou não algum tipo de variação sistemática não modelada pelo modelo trilinear.

### 5.6.5 Organização dos dados.

Os espectros de excitação e emissão das amostras foram divididos em diferentes arranjos em três modos. O arranjo  $\underline{X}_1$  foi formado pelo conjunto de três amostras coletadas 8, 10 e 12 horas após o começo da reação e os espectros de excitação entre 240 e 300 nm com seus respectivos espectros de emissão entre 290,5 e 400 nm.

Um segundo arranjo denominado de  $\underline{X}_2$  foi formado com as oito amostras coletadas ao longo do experimento de fermentação e espectros de excitação e emissão iguais ao do arranjo  $\underline{X}_1$ . Um outro arranjo denominado de  $\underline{X}_3$  foi formado com as oito amostras, mais os espectros de excitação entre 240 e 290 nm com os respectivos espectros de emissão entre 290,5 e 490,5 nm. Por fim, um arranjo denominado de  $\underline{X}_4$  foi formado pelos espectros de excitação entre 240 e 410 nm e seus respectivos espectros de emissão entre 420 e 550 nm, com as oito amostras coletadas entre oito e vinte duas horas de fermentação.

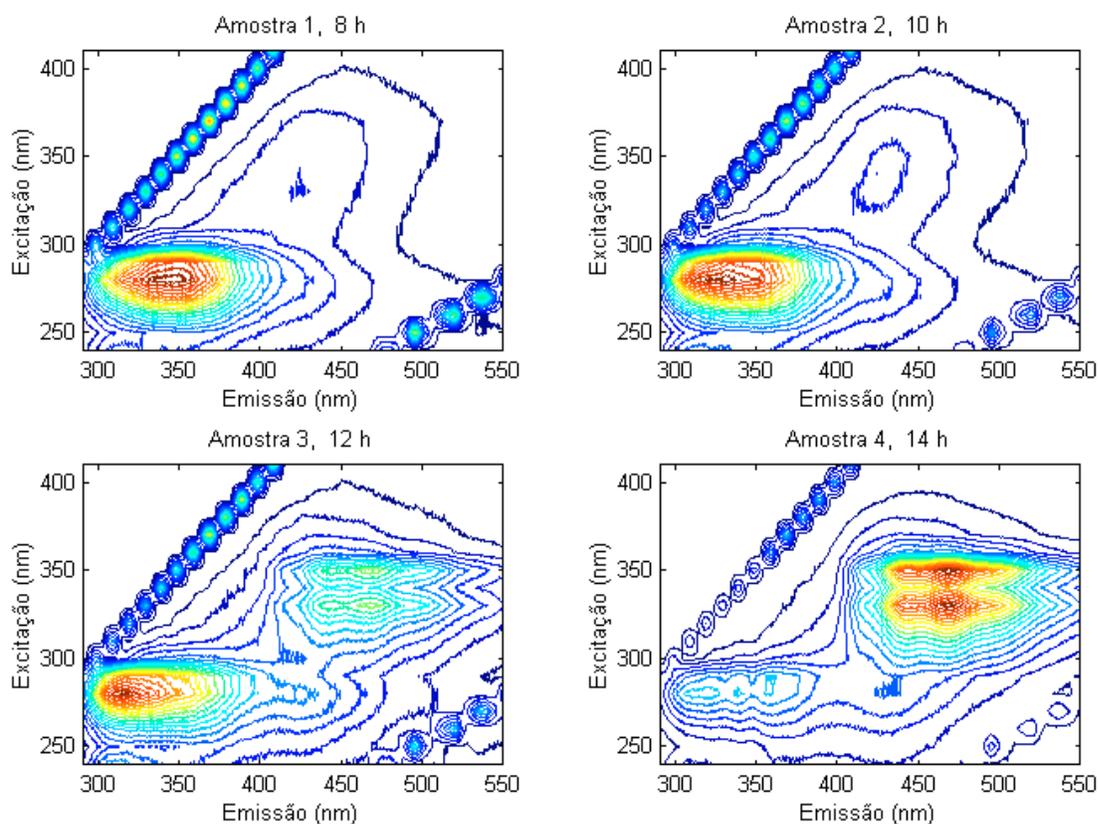
Um modelo trilinear foi ajustado em cada caso usando o método PARAFAC com restrições de não-negatividade para os três modos de monitoramento. Todos os cálculos foram iniciados com valores aleatórios como uma primeira estimativa das matrizes  $\mathbf{C}_o$  e  $\mathbf{B}_o$ .

## 5.7 RESULTADOS E DISCUSSÕES

### 5.7.1 Análise das matrizes de excitação e emissão

As figuras 5.5 e 5.6 mostram os espectros de excitação e emissão de oito amostras coletadas durante a fase exponencial de crescimento da bactéria no bioreator. A primeira amostra (figura 5.5) coletada após oito horas do começo da fermentação, apresenta um

espectro de fluorescência bem diferente da última amostra coletada após vinte e duas horas de fermentação (figura 5.6). Nesta fase entre 8 e 22 horas, a *C. violaceum* começa a produção do pigmento de cor violeta, um fenômeno que é facilmente identificado pela variação da cor cinza da biomassa para a coloração violeta.

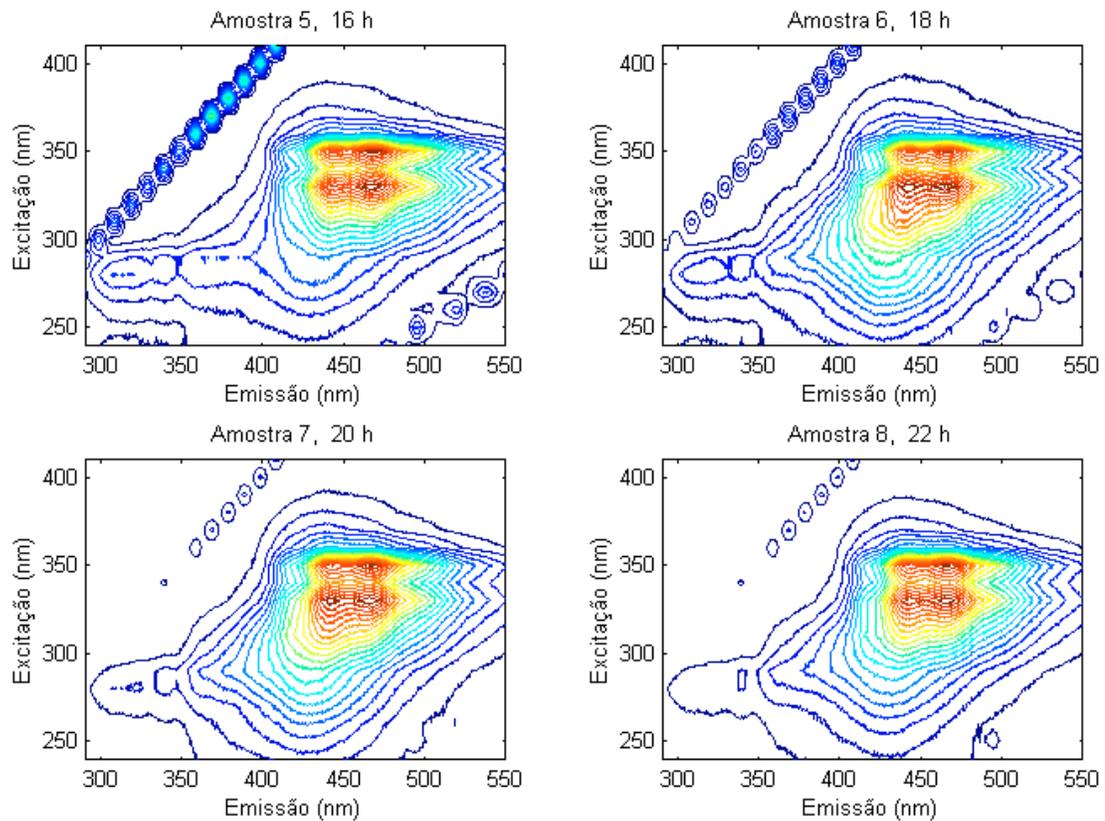


**Figura 5.5.** Matrizes de excitação e emissão de quatro amostras coletadas entre 8 e 14 horas de fermentação aeróbica.

Após a centrifugação das amostras coletadas do bioreator, o sobrenadante apresenta coloração amarela pálida, enquanto que a biomassa depositada no fundo do tubo de centrifugação apresenta coloração violeta forte, indicando que o pigmento produzido é um metabólito intracelular.

O comportamento dinâmico do metabolismo desta bactéria é facilmente observado pelo padrão de bandas que surgem e desaparecem ao longo do tempo nos espectros de fluorescência mostrados nas figuras 5.5 e 5.6. A amostra coletada em doze horas de fermentação (figura 5.5) apresenta banda no espectro de fluorescência situada entre os

comprimentos de onda de excitação de 300 a 370 nm e emissão entre 410 e 550 nm, que não está presente nas duas primeiras amostras coletadas em oito e dez horas de fermentação. Esta banda está associada ao pigmento produzido pela bactéria porque corresponde ao instante em que o meio de cultivo dentro do bioreator começa a adquirir coloração violeta.



**Figura 5.6.** Matrizes de excitação e emissão de quatro amostras coletadas entre 16 e 22 horas de fermentação aeróbica.

As matrizes de excitação e emissão das amostras um, dois e três coletadas respectivamente em oito, dez e doze horas de fermentação (figura 5.5) apresentam banda de fluorescência situada entre 250 e 300 nm no espectro de excitação e entre 290 a 420 nm para o respectivo espectro de emissão. Estas bandas provavelmente sejam os espectros dos aminoácidos tirosina, triptofano e fenilalanina porque são substâncias autofluorescentes com intensidade máxima de excitação e emissão nessa região do espectro de fluorescência (tabela 5.1).

A banda de fluorescência associada aos aminoácidos sofre um deslocamento do pico máximo de emissão situado em torno de 345 nm na amostra coletada em oito horas para um pico próximo a 320 nm na amostra coletada em doze horas. Este comportamento dinâmico sugere que algum dos aminoácidos citados acima está sendo consumido ou produzido neste intervalo de tempo. Considerando que o pico máximo de emissão do triptofano é de 350 nm e o da tirosina de 302 nm (Tabela 5.1) é provável que entre oito e doze horas de fermentação a bactéria esteja consumindo o triptofano. Isso explicaria a diminuição da intensidade de emissão próxima a 345 nm e o deslocamento para a direção dos comprimentos de onda de 320 nm que são mais próximos dos máximos de emissão dos espectros da tirosina e fenilalanina (tabela 5.1) que não estariam sendo consumidos ao mesmo tempo que o triptofano. Esta hipótese estaria de acordo com o mecanismo de biossíntese da violaceína onde duas moléculas de triptofano são consumidas para produzir a violaceína (Tópico 5.2).

Além dos espectros de fluorescência dos aminoácidos, as amostras um e dois apresentam outra banda de fluorescência com um pico máximo de emissão próximo a 430 nm, localizada entre 300 e 360 nm do espectro de excitação. Esta banda possivelmente esteja associada à presença da molécula de NADH ou de vitaminas como a vitamina B6 ou vitamina E (tabela 5.1). Após doze horas de fermentação, surge o espectro do pigmento que se sobrepõe a esta banda de fluorescência.

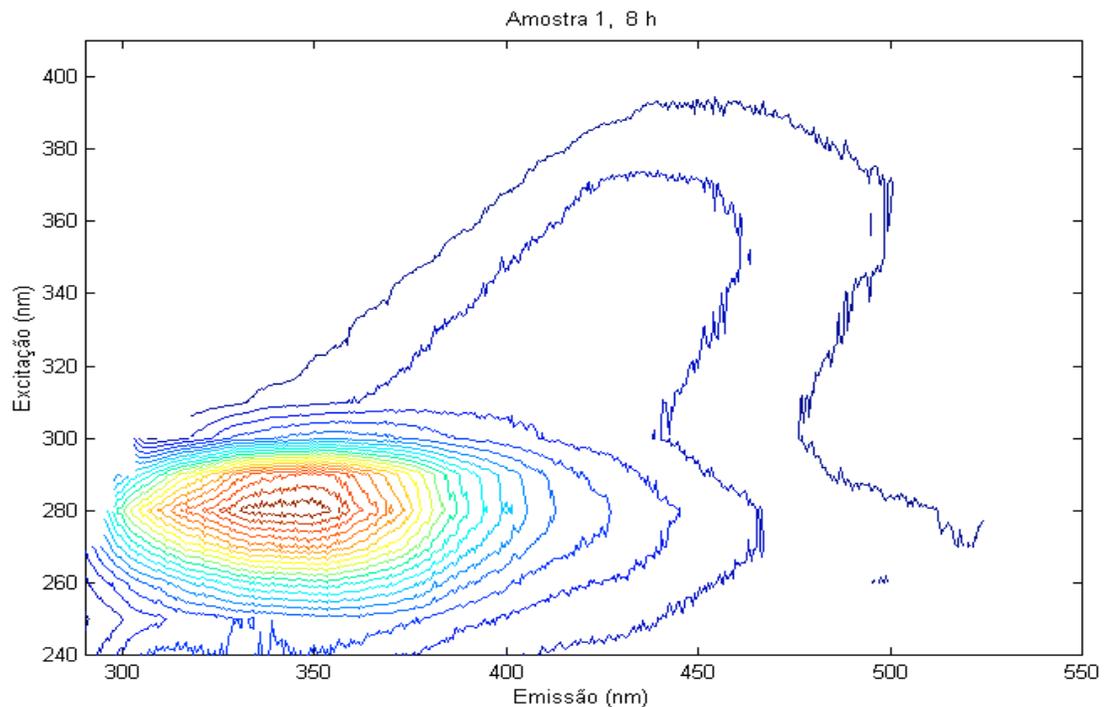
Os picos estreitos que formam uma linha diagonal nos espectros excitação e emissão correspondem aos picos de espalhamento Rayleigh. A origem destes picos é atribuída a interação da radiação eletromagnética com as moléculas do solvente usado para extrair os metabólitos (etanol) e por isso não trazem informação sobre o metabolismo da bactéria.

De acordo com os espectros apresentados nas figuras 5.5 e 5.6, as amostras coletadas do bioreator são formadas por uma mistura de substâncias químicas autofluorescentes que não podem ser identificadas com exatidão devido à sobreposição das bandas de fluorescência. Os métodos de resolução de curvas podem ser usados neste caso para obtermos informações mais claras a respeito da natureza das bandas de fluorescência destes espectros, visto que a resolução de curvas proporciona separação dos espectros de excitação e emissão de cada espécie química presente na amostra.

## 5.7.2 Análise dos dados com o método PARAFAC

Para a análise dos dados com o método PARAFAC, as matrizes de excitação e emissão apresentadas nas figuras 5.5 e 5.6 foram divididas em diferentes intervalos para formar os arranjos  $\underline{X}_1$ ,  $\underline{X}_2$ ,  $\underline{X}_3$  e  $\underline{X}_4$  conforme discutido no tópico 5.6.5. Estas divisões foram feitas com o objetivo de incluir dentro do mesmo arranjo apenas as medidas de fluorescência de comprimentos de onda de emissão maiores que os comprimentos de onda de excitação, além de reduzir a complexidade do número de bandas dentro do mesmo arranjo para facilitar a interpretação dos dados.

A figura 5.7 mostra a matriz de excitação e emissão de uma amostra após substituição dos picos do espalhamento Rayleigh por valores ausentes. Estes picos foram retirados dos dados porque não seguem um comportamento trilinear.



**Figura 5.7.** Matriz de emissão e excitação com valores ausentes adicionados no lugar dos picos de espalhamento Rayleigh.

A tabela 5.2 mostra alguns parâmetros de ajuste de cada modelo trilinear calculado com o método PARAFAC para a análise do arranjo  $\underline{X}_1$ . Este arranjo foi formado com as

três primeiras amostras, selecionando a região dos espectros de excitação e emissão que são característicos da fluorescência dos aminoácidos fenilalanina, tirosina e triptofano. De acordo com o critério da porcentagem de variação explicada, o modelo de três componentes é provavelmente o que descreve toda a variação sistemática dos dados de fluorescência no arranjo  $\underline{X}_1$ . Esta afirmação baseia-se na observação de que o modelo com quatro componentes não aumenta muito a porcentagem variação explicada em relação ao modelo de três componentes. Por outro lado, o critério CORCONDIA com um valor de 30,68 % para o modelo de três componentes é um valor baixo, e de acordo com este critério o modelo que melhor ajusta uma variação sistemática trilinear, seria o modelo de uma componente que tem um CORCONDIA de 100 %, visto que o CORCONDIA de 60,98% do modelo de duas componentes também é considerado um valor baixo.

Os dois critérios de ajuste mencionados acima não estão de acordo quanto ao melhor modelo trilinear que descreve o arranjo  $\underline{X}_1$ . Isso é freqüentemente observado nos casos em que a variação dos dados desvia-se do comportamento esperado para um conjunto de medidas que seguem a estrutura do modelo trilinear. Embora o modelo de uma componente esteja de acordo com o comportamento trilinear, a porcentagem de variação explicada (tabela 5.2) sugere que esse modelo não descreve toda a variação sistemática das medidas de fluorescências presentes no arranjo  $\underline{X}_1$ . Neste arranjo provavelmente haja três componentes que estão contribuindo para a variação da fluorescência, mas esta variação desvia-se da estrutura trilinear como mostra o CORCONDIA de 30,68 % do modelo de três componentes.

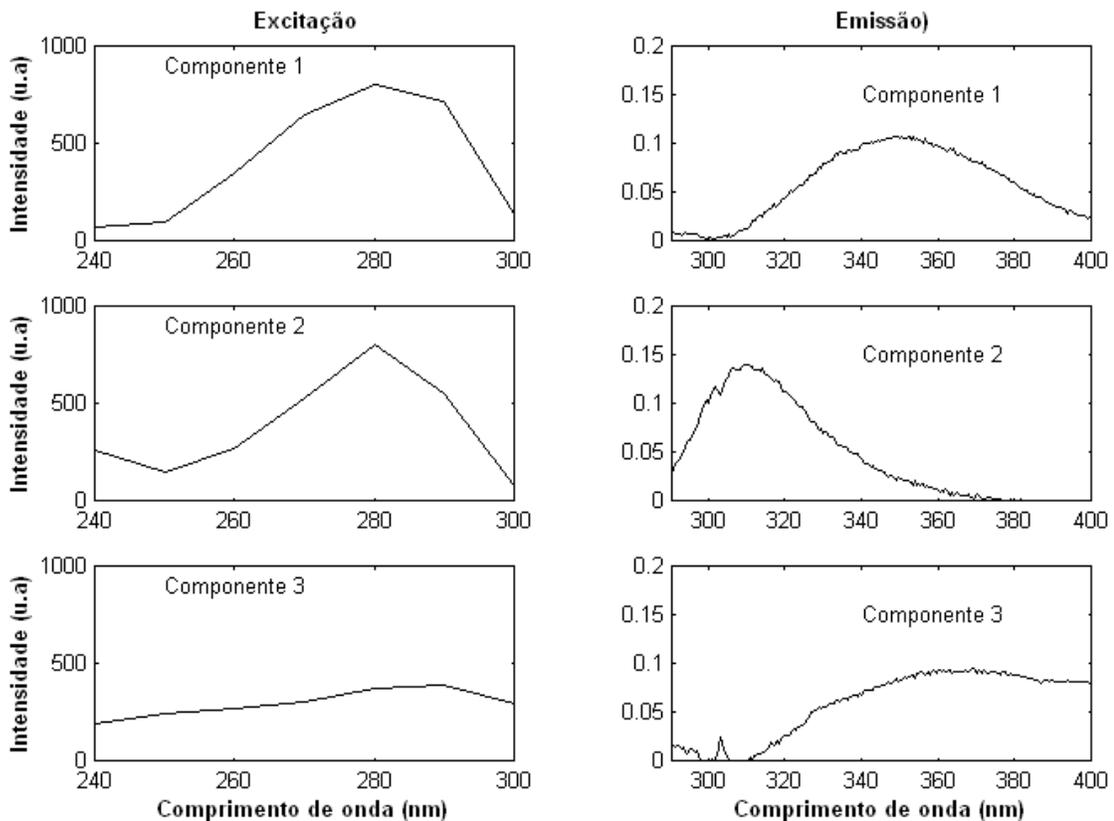
**Tabela 5.2.** Porcentagem de Variação explicada e Diagnóstico de Consistência do Núcleo (CORCONDIA) calculados para os modelos trilineares de 1 a 5 componentes com o método PARAFAC para o arranjo  $\underline{X}_1$ .\*

<b>Número de componentes</b>	1	2	3	4	5
<b>Varição explicada (%)</b>	98,75	99,74	99,92	99,94	99,97
<b>CORCONDIA (%)</b>	100	60,98	30,68	98,12	93,98

\* Ver tópico 5.6.5.

A figura 5.8 mostra os espectros de excitação e emissão resolvidos com o modelo de três componentes que correspondem respectivamente aos vetores coluna das matrizes **A**

e **B** do modelo trilinear (equação 5.2). A primeira componente do modelo tem um máximo de excitação a 280 nm e um máximo de emissão a 349 nm, sendo que estes espectros apresentam uma semelhança na forma e nos máximos de intensidade com os verdadeiros espectros de excitação e emissão do triptofano (Material Suplementar; Parte D, e tabela 5.1). A segunda componente têm os respectivos picos máximos de excitação e emissão nos comprimentos de onda de 280 nm e 310 nm. Os espectros resolvidos por esta componente aproximam-se dos verdadeiros espectros de excitação e emissão da tirosina (Material Suplementar: Parte D.1, e tabela 5.1). Os espectros resolvidos com a terceira componente não foram identificados porque os espectros de excitação e emissão desta componente não apresentam uma estrutura bem definida.



**Figura 5.8.** Estimativas dos espectros de excitação e emissão obtidas com o modelo trilinear de três componentes ajustado com o método PARAFAC.

De acordo com as estimativas dos espectros resolvidos pelo modelo de três componentes, os aminoácidos triptofano e tirosina são as moléculas que contribuem para a fluorescência da banda de emissão posicionada entre 290 e 400 nm na região entre 240

e 300 nm do espectro de excitação, como mostra a figura 5.5 para as amostras 1, 2 e 3. A presença de uma terceira componente neste modelo indica que pode haver ainda uma outra substância que foi mal resolvida, talvez porque contribua muito pouco para as medidas de fluorescência quando comparada as contribuições da tirosina e do triptofano.

Baseado nos resultados obtidos com o arranjo  $\underline{X}_1$ , uma outra análise foi feita a partir do arranjo  $\underline{X}_2$  (tópico 5.6.5) a fim de incluir todas as amostras para verificar o que acontece com os aminoácidos tirosina e triptofano ao longo de 22 horas de fermentação. A tabela 5.3 mostra os dois parâmetros de ajuste de cada modelo trilinear calculado com o método PARAFAC para a análise do arranjo  $\underline{X}_2$ .

De acordo com a porcentagem de variação explicada, o modelo de três componentes deve explicar praticamente toda a variação sistemática das medidas de fluorescência visto que entre o modelo de três e quatro componentes, a quantidade de variação explicada não aumenta muito. O CORCONDIA de 99,9% do modelo de duas componentes é excelente, mas apesar disso o aumento de porcentagem de variação explicada entre o modelo de duas e três componentes parece ser significativo, o que sugere um modelo de três componentes. Da mesma forma, o CORCONDIA de 74,74% do modelo de três componentes indica uma concordância razoável entre o comportamento dos dados e a estrutura do modelo trilinear.

**Tabela 5.3.** Porcentagem de Variação explicada e Diagnóstico de Consistência do Núcleo (CORCONDIA) calculados para os modelos trilineares de 1 a 5 componentes com o método PARAFAC para o arranjo  $\underline{X}_2$ \*

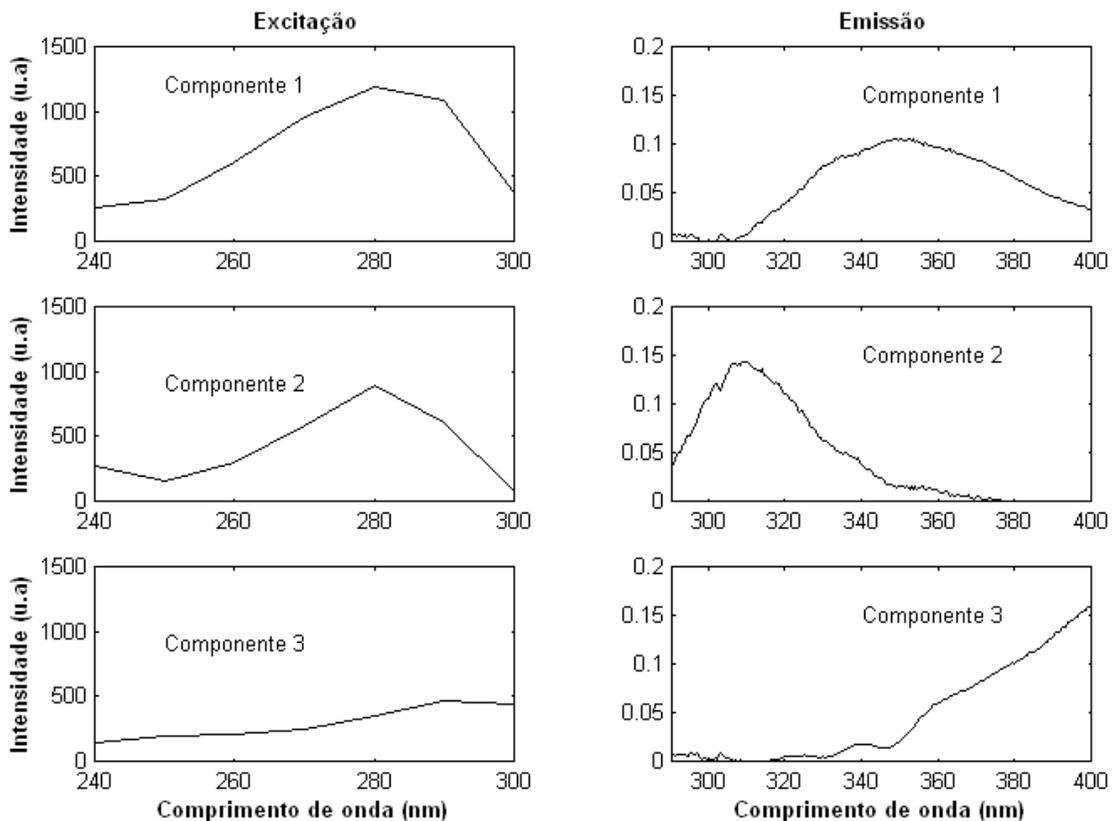
<b>Número de componentes</b>	1	2	3	4	5
<b>Varição explicada (%)</b>	94,78	99,00	99,78	99,87	99,91
<b>CORCONDIA (%)</b>	100	99,99	74,74	16,05	-0,67

\* Ver tópico 5.6.5.

A figura 5.9 mostra os espectros resolvidos com o modelo de três componentes a partir do arranjo  $\underline{X}_2$ . Da mesma forma que na análise anterior (arranjo  $\underline{X}_1$ ), verifica-se que a primeira e segunda componentes desse modelo são respectivamente uma estimativa dos espectros de excitação e emissão dos aminoácidos triptofano e tirosina. Entretanto, a terceira componente desse novo modelo, apresenta um espectro de emissão com o

formato de uma cauda, como se o espectro de alguma substância tivesse sido cortado. O mesmo não se pode concluir do espectro de excitação, porque esse não apresenta uma forma bem definida.

Este novo modelo ajustado a partir do arranjo  $\underline{X}_2$  sugere que existe realmente uma terceira substância além dos aminoácidos tirosina e triptofano, e que a sua contribuição aumenta ao longo do tempo visto que após incluir mais amostras no arranjo dos dados foi obtida uma estimativa melhor para o seu espectro de emissão que se encontra truncado.



**Figura 5.9.** Estimativas dos espectros de excitação e emissão obtidas com o modelo trilinear de três componentes ajustado com o método PARAFAC para o arranjo  $\underline{X}_2$ .

Baseado nesse resultado, um novo modelo foi calculado a partir do arranjo  $\underline{X}_3$ , no qual foram incluídas as emissões até os comprimentos de onda de 490,3 nm, com espectros de excitação entre 240 e 290 nm. Neste arranjo preferimos não incluir os espectros de emissão excitados a 300 nm para evitar uma possível contribuição da banda de fluorescência do pigmento em emissões acima de 400 nm.

A tabela 5.4, mostra os parâmetros de ajuste calculados para os modelos trilineares de um a cinco componentes ajustados com o método PARAFAC a partir do arranjo  $X_3$ . A porcentagem de variação explicada do modelo de três componentes explica a maior parte da variação sistemática dos espectros de fluorescência, apesar desse modelo apresentar um CORCONDIA de 44,27%, o que indica problemas de desvio do comportamento dos dados em relação ao esperado para um modelo trilinear.

**Tabela 5.4.** Porcentagem de Variação explicada e Diagnóstico de Consistência do Núcleo (CORCONDIA) calculados para os modelos trilineares de 1 a 5 componentes com o método PARAFAC para o arranjo  $X_3$ .

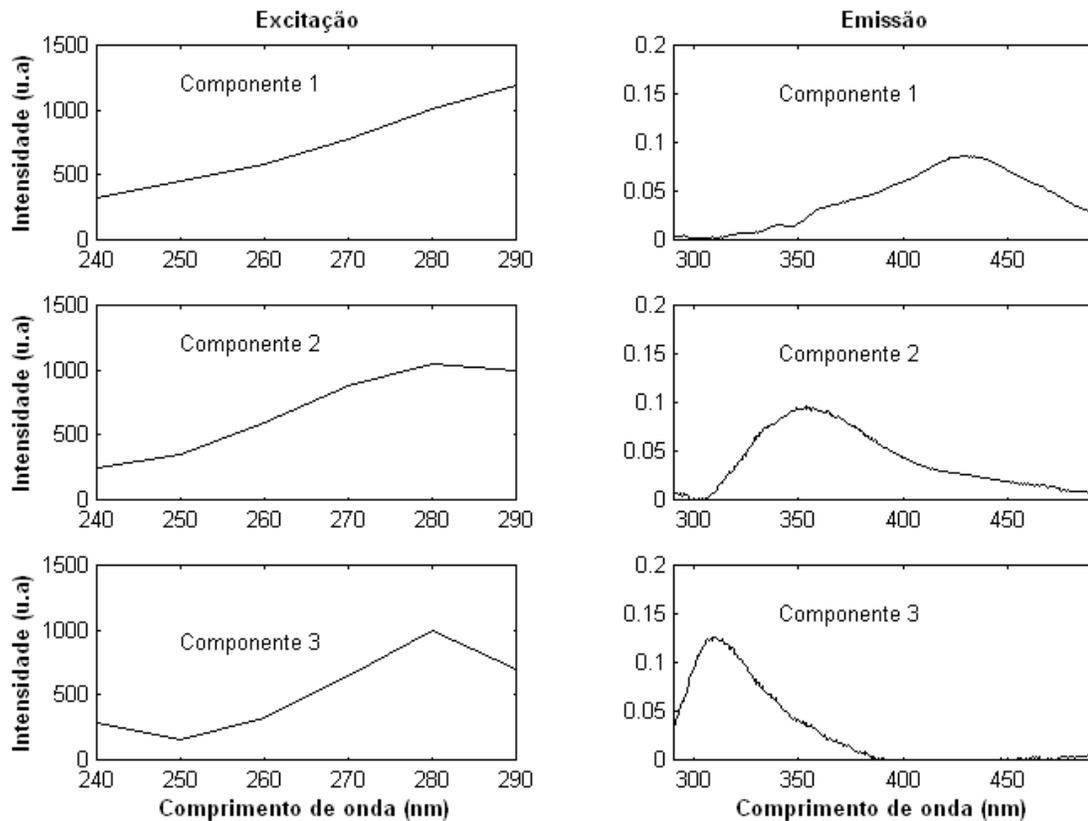
Número de componentes	1	2	3	4	5
Varição explicada (%)	83,92	98,99	99,58	99,81	99,91
CORCONDIA (%)	100,00	99,99	44,27	1,95	-1,43

\* Ver tópico 5.6.5.

A figura 5.10, mostra o modelo de três componentes ajustado para o arranjo  $X_3$ . Nesse modelo a segunda e a terceira componentes são estimativas para os espectros de excitação e emissão dos aminoácidos triptofano e tirosina, respectivamente. Por outro lado, a primeira componente desse modelo resolve um espectro de emissão com uma forma muito bem definida, mostrando que de fato há uma outra substância presente na região dos espectros de excitação entre 240 e 290 nm, cuja emissão máxima ocorre num comprimento de onda próximo a 430 nm. O espectro de excitação resolvido para a primeira componente não apresenta um máximo visto que parece estar truncado.

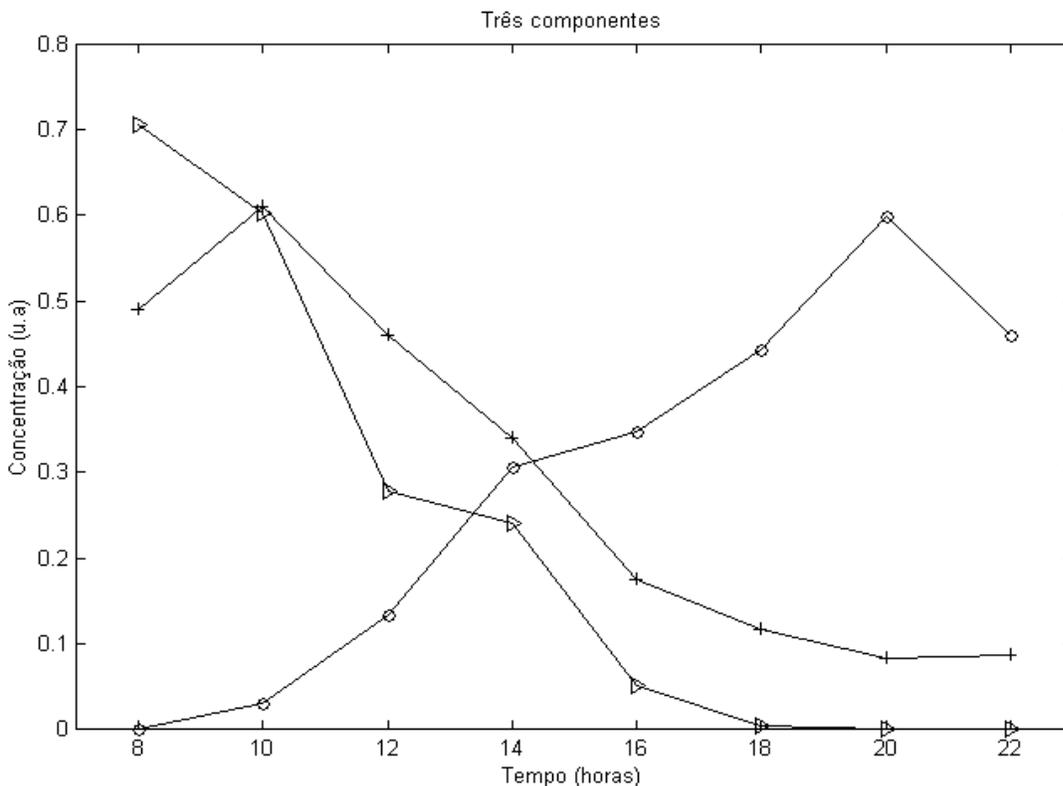
De acordo com a estimativa feita pela primeira componente para o espectro de emissão, essa poderia estar associada à molécula de ATP, NADH ou alguma vitamina como B6 ou E (Material Suplementar: Parte D, e tabela 5.1). Entretanto, se analisarmos a figura 5.5 verificamos que nas amostras 1 e 2, há uma banda de emissão com máximo de intensidade em 430 nm para a excitação no comprimento de onda de 340 nm, que após a amostra 3 (12 horas de fermentação) tem seu espectro sobreposto pela banda de fluorescência do pigmento. Se cruzarmos esta informação com as estimativas feitas pela primeira componente apresentada na figura 5.10, podemos supor que o espectro de excitação desta componente se estenda até 400 nm, estando ela associada a substância

com um pico de emissão em 430 nm, verificada nas amostras 1 e 2. Se isto é verdadeiro, o mais provável é que a molécula em questão seja NADH (Material suplementar: parte D.1).



**Figura 5.10.** Estimativas dos espectros de excitação e emissão obtidas com o modelo trilinear de três componentes ajustado com o método PARAFAC para o arranjo  $\underline{X}_3$ .

A figura 5.11, mostra as estimativas para a variação de concentração de cada substância em função do tempo obtidas a partir da matriz  $\mathbf{C}$  do modelo trilinear de três componentes. A concentração dos aminoácidos triptofano (segunda componente) e tirosina (terceira componente) diminuem ao longo do tempo, sendo que a concentração de triptofano ao final de 22 horas é muito próxima de zero, sugerindo que o triptofano foi completamente consumido. A concentração da espécie química estimada pela primeira componente aumenta ao longo do tempo até 20 horas. Note que os valores de concentração estão em uma escala de unidades arbitrárias, porque a solução obtida a partir do método PARAFAC apresenta ambiguidade de intensidade.



**Figura 5.11.** Estimativas da variação da concentração em função do tempo obtidas do modelo trilinear ajustado a partir do arranjo  $\underline{X}_3$ . (o) Primeira componente; (>) segunda componente; (+) terceira componente.

Após a análise do arranjo  $\underline{X}_3$  na qual foram detectadas três substâncias, um novo arranjo denominado  $\underline{X}_4$  foi analisado para investigar a outra região dos espectros de excitação e emissão onde se encontram os espectros do pigmento que começa a ser produzido após doze horas de fermentação (figura 5.5). O arranjo  $\underline{X}_4$  foi formado pelos espectros de excitação entre 240 e 410 nm e seus respectivos espectros de emissão entre 420 e 550 nm, com as oito amostras coletadas entre oito e vinte duas horas de fermentação.

A tabela 5.5 mostra os parâmetros de modelos trilineares de um a cinco componentes ajustados a partir do arranjo  $\underline{X}_4$  (tópico 5.6.5) com o método PARAFAC. O modelo de duas componentes explica 99,42% da variação dos dados de fluorescência com um CORCONDIA de 99,99 %. A porcentagem de variação explicada tem aumento significativo ao passar de dois para três componentes, enquanto que o CORCONDIA cai do patamar de 90%.

**Tabela 5.5.** Porcentagem de Variação explicada e Diagnóstico de Consistência do Núcleo (CORCONDIA) calculados para os modelos trilineares de 1 a 5 componentes com o método PARAFAC para o arranjo  $\underline{X}_4$ \*

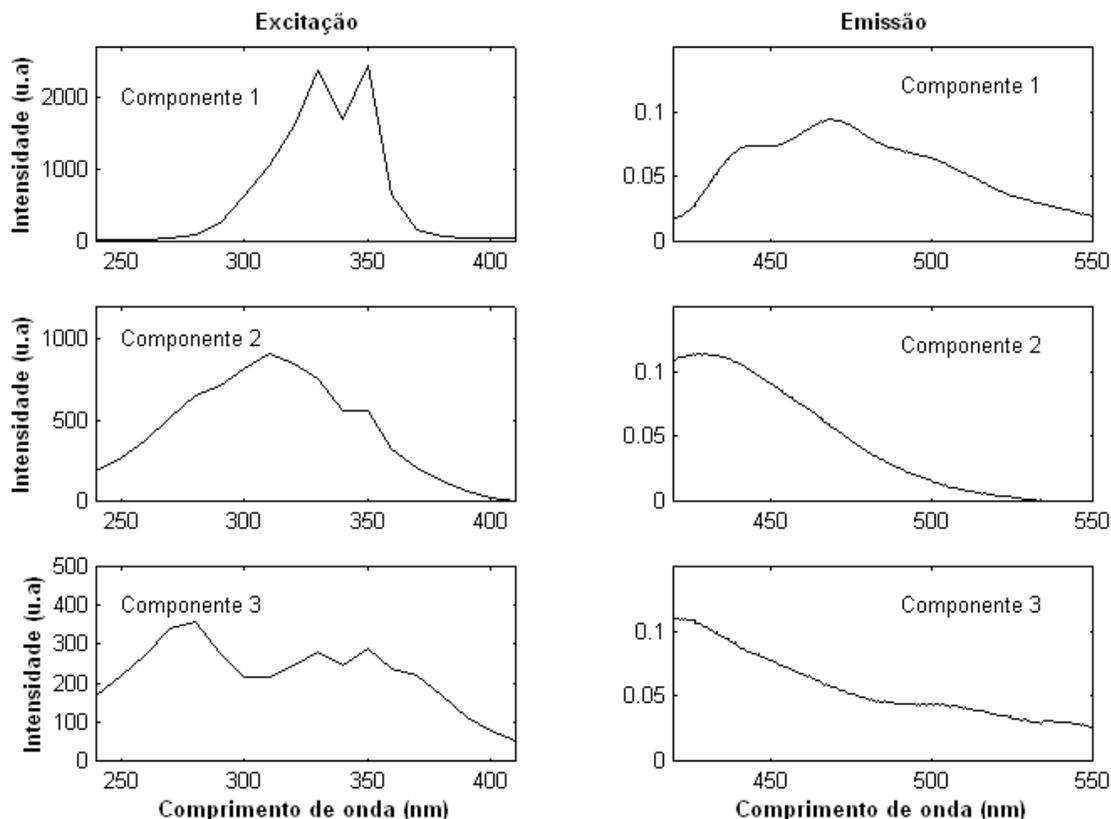
<b>Número de componentes</b>	1	2	3	4	5
<b>Variação explicada (%)</b>	97,12	99,42	99,85	99,96	99,96
<b>CORCONDIA (%)</b>	100,00	99,99	65,06	27,60	0,95

\* Ver tópico 5.6.5.

A figura 5.12 mostra as componentes do modo de emissão e excitação resolvidas com o modelo de três componentes. Os espectros resolvidos pela primeira componente foram identificados como sendo os espectros de excitação e emissão do pigmento devido ao aspecto bimodal das curvas resolvidas para os modos de excitação e emissão. O espectro de excitação resolvido pela primeira componente apresenta um pico em 330 e outro em 350 nm, enquanto que o espectro de emissão tem um pico bem definido em 468 nm e um outro menor em 442 nm.

O espectro de emissão da segunda componente apresenta uma curva unimodal com um pico máximo em 430 nm e o respectivo pico máximo do espectro de excitação em 310 nm. Esta componente parece estar associada a mesma espécie química resolvida pela componente um, do modelo trilinear ajustado a partir do arranjo  $\underline{X}_3$  (figura 5.10). Mas neste novo modelo ajustado a partir do arranjo  $\underline{X}_4$ , o espectro de excitação desta substância foi totalmente resolvido porque foram incluídas as medidas de excitação até 410 nm. Conforme mencionado anteriormente esta substância é provavelmente o NADH.

A terceira componente mostrada na figura 5.12 apresenta um pico no espectro de emissão em 425 nm, enquanto que o seu respectivo espectro de excitação apresenta uma forma bimodal com um dos picos em 280 nm e o outro que parece desdobrar-se em dois pequenos picos, sendo um em 330 nm e outro em 350 nm. A forma bimodal desse espectro de excitação é similar à do verdadeiro espectro de excitação da vitamina B6 (Material suplementar: parte D.1), e o pico máximo de emissão em 425 nm para essa componente é relativamente próximo do máximo de emissão da vitamina B6 que é de 393 nm (tabela 5.1).



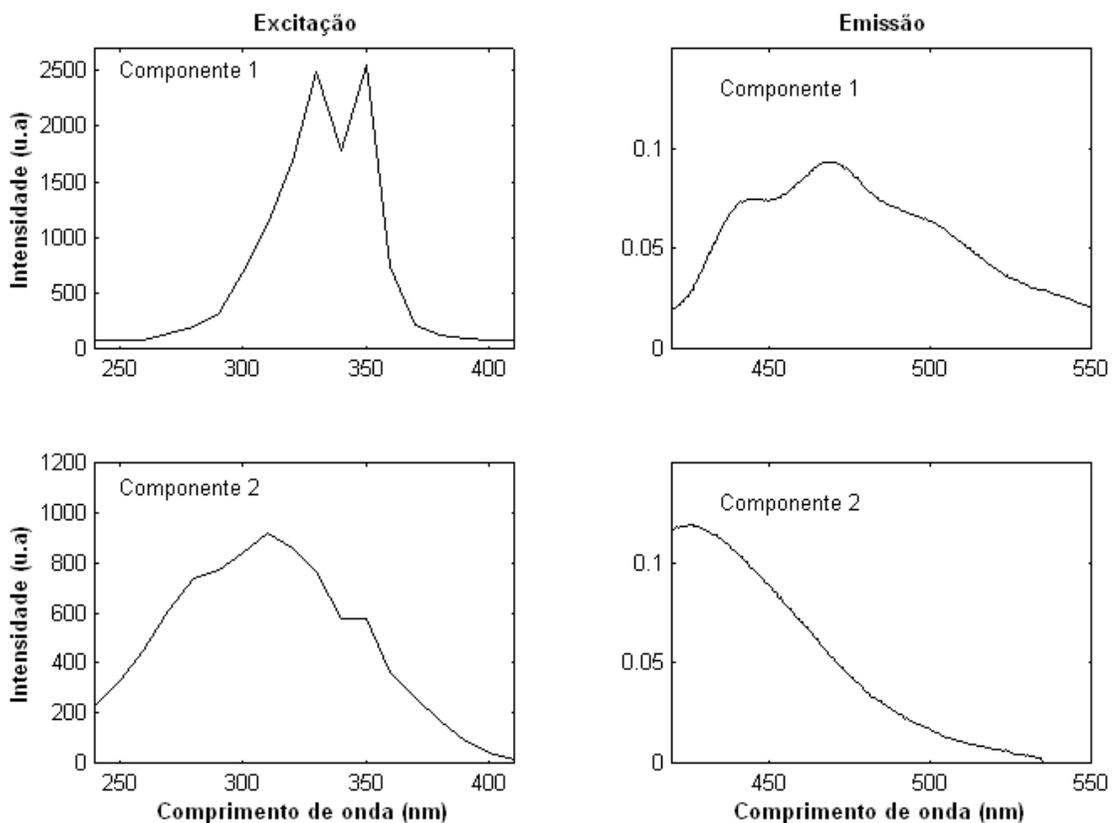
**Figura 5.12.** Estimativas dos espectros de excitação e emissão obtidas com o modelo trilinear de três componentes ajustado com o método PARAFAC para o arranjo  $\mathbf{X}_4$ .

Antes do ajuste dos modelos apresentados acima, foi verificado que as bandas de fluorescência do pigmento sofrem pequenas alterações no perfil de uma amostra para a outra, o que indica um desvio de trilinearidade. Quando os dados apresentam um desvio do comportamento trilinear, a variação sistemática nas medidas de fluorescência pode ser atribuída também a uma variação na forma do espectro de uma mesma espécie química e não devido a espécies químicas diferentes.

Devido ao efeito do desvio de trilinearidade descrito acima, é possível que o espectro resolvido pela terceira componente da figura 5.12 não esteja associado a uma verdadeira espécie química, mas apenas a uma informação que está associada a uma mudança da forma da banda de fluorescência do pigmento. Há uma forte evidência para essa observação, visto que um dos picos de excitação resolvidos pela terceira componente da figura 5.12 desdobra-se em dois picos iguais aqueles resolvidos para o pigmento (330 e 350 nm), além de um outro que corresponde ao máximo de excitação do triptofano (280

nm). Essa terceira componente, na verdade pode estar adicionando no modelo apenas a variação associada a mudanças na forma do espectro do pigmento e a alguma outra variação na forma dos espectros situados na região dos aminoácidos.

A figura 5.13 mostra as estimativas obtidas para os espectros de excitação e emissão com o modelo de duas componentes ajustado a partir do arranjo  $\underline{X}_4$ . A primeira componente é uma curva bimodal igual aquela que foi resolvida pela primeira componente do modelo de três componentes (figura 5.12). A segunda componente apresenta uma máximo de emissão em 430 nm e um máximo de excitação em 310 nm, e é igual a segunda componente do modelo de três componentes (Figura 5.12).

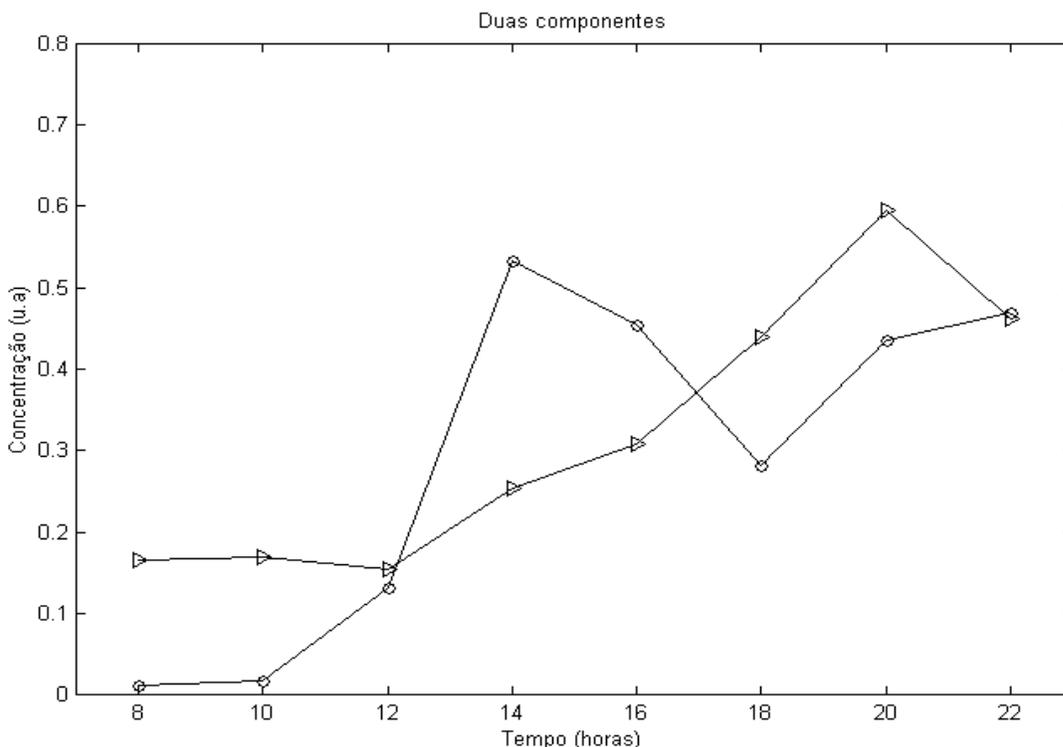


**Figura 5.13.** Estimativas dos espectros de excitação e emissão obtidas com o modelo trilinear de duas componentes ajustado com o método PARAFAC a partir do arranjo  $\underline{X}_4$ .

Ao analisar as estimativas feitas com os modelos de duas e três componentes, pode-se afirmar que, de fato, existem duas substâncias químicas que contribuem para os dados de fluorescência do arranjo  $\underline{X}_4$ , sendo essas o pigmento e o NADH. A terceira componente do modelo de três componentes captura uma quantidade significativa de

variação sistemática como mostra a tabela 5.5 mas, devido aos desvios de trilinearidade e a forma dos espectros resolvidos pela terceira componente, não é possível afirmar que de fato haja uma terceira substância, que poderia ser a vitamina B6, com base na análise dos dados disponíveis neste trabalho.

A figura 5.14 mostra a variação da concentração das duas espécies químicas resolvidas a partir da análise do arranjo  $X_4$  com o modelo de duas componentes. Embora não seja mostrado, a estimativa da variação de concentração em função do tempo feita com o modelo de três componentes para estas duas substâncias são muito parecidas com as estimativas apresentadas na figura 5.12. Este gráfico mostra que a concentração do pigmento tem um pico máximo de produção entre 10 e 14 horas e em seguida sofre uma queda de concentração mantendo-se, aproximadamente, constante até 22 horas com exceção de que a concentração em dezoito horas apresenta um valor surpreendentemente baixo.



**Figura 5.14.** Estimativas da variação da concentração em função do tempo obtidas do modelo trilinear ajustado a partir do arranjo  $X_4$ . (o) Primeira componente; (>) segunda componente.

A explicação mais provável para o ponto em dezoito horas apresentar um valor de concentração tão baixo é que tenha ocorrido uma extração incompleta do pigmento devido a algum erro experimental ou devido a um fenômeno de “Quenching”<sup>144</sup> mais pronunciado neste instante. O “Quenching” é um fenômeno que promove uma supressão na intensidade de radiação emitida. Este efeito também contribui para os desvios de trilinearidade observados na maioria dos modelos ajustados, e identificados pelo baixo valor do CORCONDIA.

De acordo com os resultados obtidos, os espectros de excitação e emissão apresentados nas figuras 5.5 e 5.6, são formados pela fluorescência de, no mínimo, quatro espécies químicas, tendo sido identificado o triptofano, tirosina, NADH e o pigmento. Há suspeitas de que o modelo trilinear ajustado a partir do arranjo  $X_4$  tenha resolvido espectros de uma quinta substância, que seria a vitamina B6. Entretanto, a presença desta substância não pode ser confirmada neste trabalho.

A presença da molécula de NADH no metabolismo da *C. violaceum* é explicada pela via metabólica Entner-Doudoroff<sup>169</sup> utilizada por esta bactéria para a produção de energia a partir do consumo de carboidratos. Além disso, o NADH é utilizado na biossíntese da violaceína como um cofator da enzima oxigenase.<sup>140</sup> Da mesma forma, há muitos outros trabalhos que demonstram que a molécula de NADH pode ser detectada no monitoramento de bioreatores usando fluorescência.<sup>151</sup>

As estimativas obtidas para os espectros de excitação e emissão das espécies químicas, foram suficientes para identificar os aminoácidos e o pigmento, porém apenas qualitativamente. Os modelos são apenas aproximações dos espectros e concentrações verdadeiros, visto que foi verificado um desvio de trilinearidade como acusa os baixos valores de CORCONDIA para todos os modelos ajustados.

Apesar das estimativas não serem tão exatas, as informações qualitativas obtidas mostram que a concentração do triptofano decresce em função do tempo (figura 5.11) horas antes de iniciar a produção do pigmento violeta. Embora não tenha sido apresentado, nesta mesma etapa do crescimento de biomassa, a concentração de oxigênio molecular detectada no meio de cultivo com um sensor apropriado foi mínima. Estas evidências concordam com o mecanismo de biossíntese da violaceína (tópico 5.3)

sugerido por outros pesquisadores e demonstra que a metodologia desenvolvida neste trabalho pode ser usada para investigar em maiores detalhes a produção de violaceína pela *C. violaceum*.

## 5.8 CONCLUSÃO

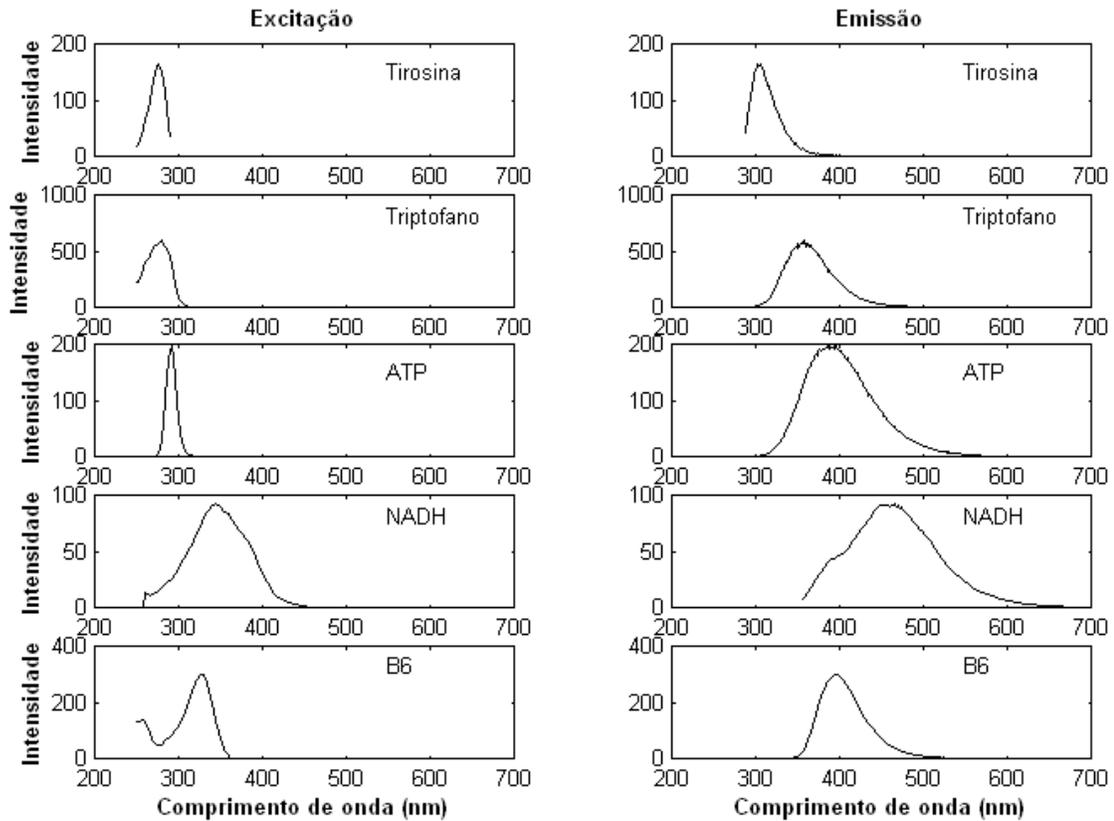
A metodologia utilizada neste trabalho permitiu identificar os metabólitos fluorescentes produzidos pela *C. violaceum* durante o crescimento de biomassa. Os resultados obtidos mostraram que a metodologia é útil para investigar a biossíntese da violaceína, um importante metabólito desta bactéria que apresenta um enorme potencial para o desenvolvimento de antibióticos e remédios que atuam contra o câncer.

Embora tenha sido observado desvios do comportamento dos dados de fluorescência em relação ao modelo proposto para a descrição do sistema, foi possível identificar qualitativamente a presença de espécies químicas como triptofano, tirosina, pigmento e o NADH. Todas estas substâncias são importantes para a descrição do metabolismo da bactéria e o modelo proposto pode ser usado para investigar o comportamento de cada uma dessas substâncias durante o cultivo da *Chromobacterium violaceum* em meio aeróbico.

Neste trabalho foi observado qualitativamente que o pigmento é produzido horas após o consumo de triptofano pela bactéria, concordando com as evidências experimentais de fontes da literatura. Devido a esse resultado, acredita-se que a metodologia desenvolvida nesse trabalho possa ser estendida em trabalhos futuros para investigar quantitativamente a relação entre o consumo e a produção de violaceína. Entretanto, para esta tarefa deveriam ser testados outros métodos de calibração de segunda ordem diferentes do método PARAFAC, que sejam menos sensíveis aos problemas de desvios de trilinearidade.

## D. MATERIAL SUPLEMENTAR: PARTE D.

### D.1 Espectros de excitação e emissão de moléculas biológicas



**Figura 5.15.** Espectros de excitação e emissão de moléculas presentes no metabolismo de seres vivos.

Os espectros apresentados na figura 5.15 foram obtidos da Food Fluorescence Library, na página da internet: “[www.models.kvl.dk](http://www.models.kvl.dk)”.<sup>145</sup> A tabela 5.6 mostra as concentrações de cada substância nas soluções em que foram medidos os espectros da figura 5.15 bem como o tipo de solvente usado e os comprimentos de onda com picos máximos de excitação e emissão.

**Tabela 5.6:** Substâncias fluorescentes.

<b>Substância</b>	<b>Concentração (M)</b>	<b>Excitação <math>\lambda_{\max}</math> (nm)</b>	<b>Emissão <math>\lambda_{\max}</math> (nm)</b>	<b>Solvente</b>
Fenilalanina	$10^{-3}$	258	284	água
Tirosina	$10^{-5}$	276	302	água
Triptofano	$10^{-5}$	280	357	água
Vitamina A	$10^{-5}$	346	480	Iso octano
Vitamina B2	$10^{-5}$	270	518	água
Vitamina B6	$10^{-5}$	328	393	água
Vitamina E	$10^{-4}$	298	326	Iso octano
NADH	$10^{-4}$	344	465	água
ATP	$10^{-3}$	292	388	água
Clorofila A	$10^{-6}$	428	663	Iso octano
Hematoporfirina	$10^{-5}$	396	614	Iso octano

## **CAPÍTULO 6**

### **CONTROLE ESTATÍSTICO MULTIVARIADO DE PROCESSOS**

#### **Fundamentos Teóricos**

## 6.1 INTRODUÇÃO

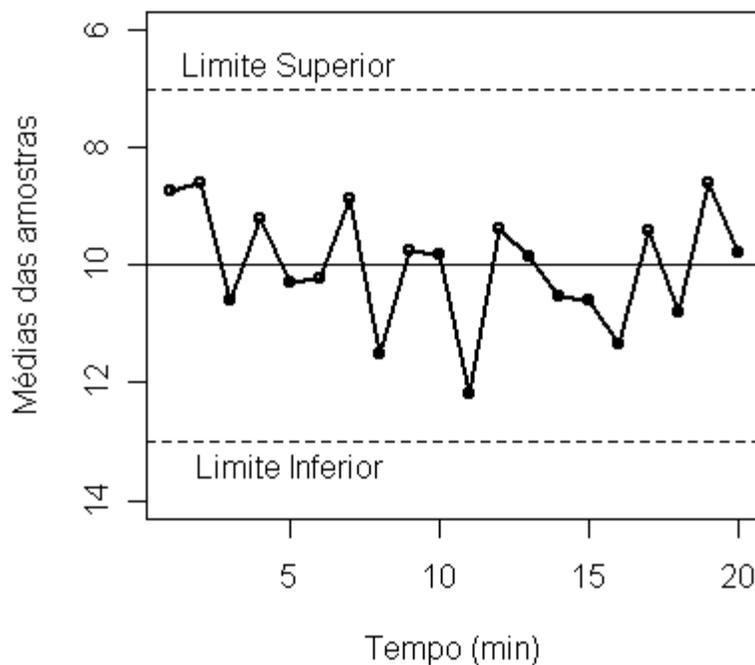
Durante a década de 1980 a indústria mundial de eletrodomésticos e equipamentos eletrônicos foi abalada pela invasão dos produtos provenientes da indústria japonesa que durante um período de tempo passaram a dominar boa parte do mercado consumidor. Os produtos japoneses eram considerados de melhor qualidade em relação aos produtos de seus concorrentes de outros países. Para não perder espaço, a indústria mundial teve que mudar a sua filosofia de trabalho a fim de poder competir com a indústria japonesa. Nesse processo, foi descoberto que o “segredo” dos japoneses estava em uma metodologia de trabalho focada em resultados e ações que visavam sempre melhorar a qualidade dos seus produtos e reduzir as perdas na produção industrial.

Para conseguir seus objetivos, os japoneses tinham uma filosofia de trabalho voltada para a melhoria contínua da qualidade, desenvolvida por Dr. William Edwards Deming, um americano que na década de 1950 foi ao Japão a convite da “Japanese Union of Scientists and Engineers (JUSE)” para ministrar palestras e cursos sobre técnicas de Controle Estatístico de Processos e Técnicas de Planejamento e Otimização de Experimentos. Deming foi responsável pelo treinamento direto de gerentes das indústrias japonesas e, de acordo com os seus ensinamentos sintetizados em 14 pontos, a melhoria da qualidade deve ser uma ação planejada e controlada com o apoio de métodos estatísticos usados para os seguintes propósitos:

1. Promover o aumento de produtividade;
2. Prevenir que os produtos sejam enviados ao mercado com defeitos;
3. Prevenir ajustes desnecessários na operação industrial;
4. Obter informações importantes tal como o índice de capacidade do processo.<sup>170</sup>

Em 1924, Dr. Walter A. Shewhart então funcionário da Western Electric uma divisão da “Bell Telephone Laboratories”, criou a técnica de controle estatístico de processos (CEP) baseada no monitoramento do processo com os gráficos de controle,<sup>171</sup> que mais tarde seria usada extensivamente por Deming. Segundo Shewhart um processo está em

controle estatístico quando apresenta apenas uma variabilidade que lhe é inerente, devido às “Causas Naturais”. Nesta condição, o processo pode ser modelado de acordo uma distribuição estatística de probabilidade caracterizada por parâmetros como o valor médio e a variância da distribuição. Estes dois parâmetros descrevem o comportamento de uma propriedade do processo em controle estatístico e podem ser monitorados ao longo do tempo através de gráficos de controle como o gráfico apresentado na figura 6.1.



**Figura 6.1.** Gráfico de controle do tipo Shewhart para monitorar a média de uma propriedade.

Além dos pontos que representam o comportamento estatístico de uma propriedade do processo, o gráfico de controle de Shewhart (figura 6.1) apresenta três linhas horizontais, uma para indicar o valor médio e outras duas para indicar os limites de controle estatísticos superior e inferior dentro dos quais os parâmetros da distribuição de probabilidade devem variar se o processo está em controle estatístico. Se uma amostra é assinalada fora dos limites de controle então o processo está fora de controle estatístico e o motivo pode ser devido a uma “Causa Especial” que provoca uma alteração do valor médio ou da variância da propriedade que está sendo monitorada.

A utilização dos gráficos de controle tem um longo histórico na indústria que provam a eficácia desta ferramenta da qualidade. O monitoramento dos processos com os gráficos de controle foi uma das ferramentas responsáveis pela melhoria dos processos e produtos oferecidos pela indústria mundial, principalmente no segmento industrial ligado a produção de peças mecânicas e equipamentos eletrônicos.

Durante a segunda guerra mundial, o pesquisador Harold Hotteling<sup>172</sup> desenvolveu o método de controle estatístico multivariado da qualidade e usou esta metodologia para aumentar a precisão das miras das bombas. O trabalho de Hotteling estendeu a metodologia criada por Shewhart para o monitoramento simultâneo de propriedades correlacionadas entre si. Apesar disto somente o gráfico de shewhart tornou-se um método popular após a década de 1950. O controle estatístico multivariado de processos (CEMP) ficou praticamente esquecido até o começo da década de 1990. Um dos motivos seria a dificuldade prática de realizar manualmente todos os cálculos necessários para a construção do gráfico  $T^2$  que será discutido mais adiante neste texto.

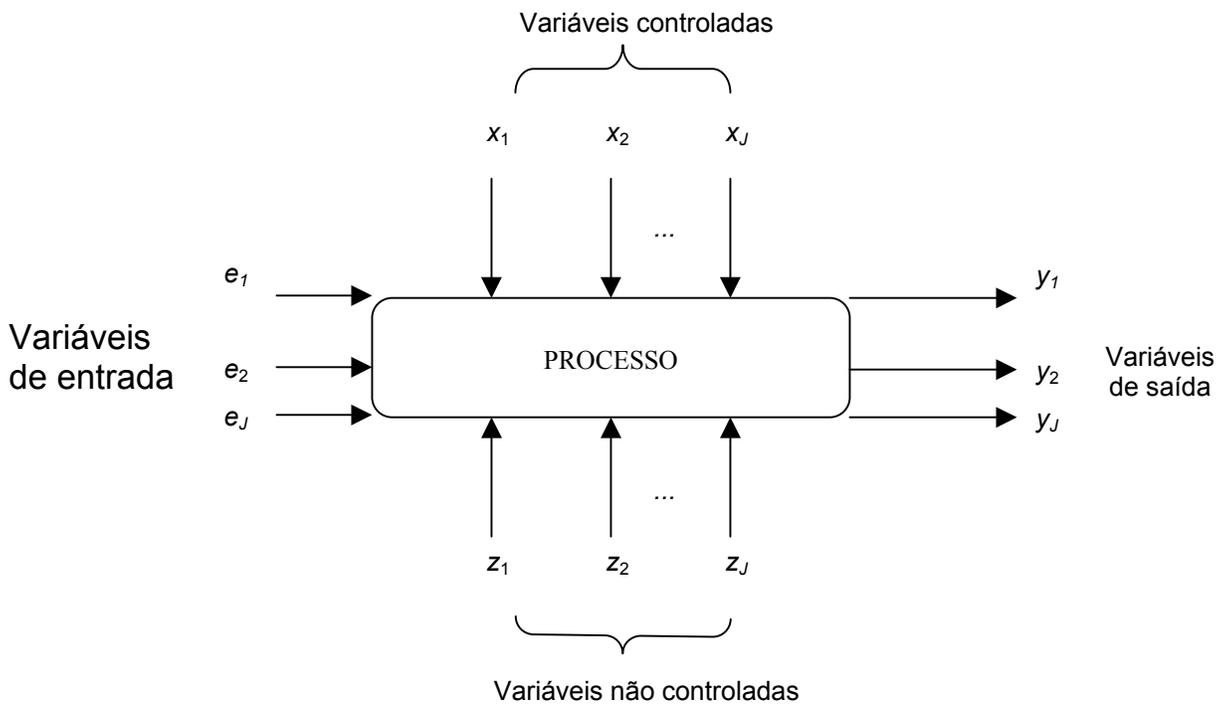
A presença maciça de microcomputadores nas plantas industriais mudou a indústria no final do século XX e com isto os procedimentos de aquisição e processamento matemático de centenas de dados provenientes de sensores passaram a ser uma tarefa relativamente simples e rápida. Neste novo panorama, os métodos de CEMP passaram novamente a serem implementados para as aplicações industriais, sendo que a indústria química tem manifestado muito interesse por esta metodologia de monitoramento nos últimos anos. Desde o começo da década de 1990 tem ocorrido a divulgação sistemática de trabalhos que relatam o desenvolvimento e aplicação do CEMP para o monitoramento das linhas de produção industrial.<sup>173,174,175,176</sup> No texto que segue apresentamos as bases teóricas dos métodos de CEMP e como esta técnica está sendo utilizada atualmente para o monitoramento multivariado de processos químicos.

## **6.2 PRINCIPIOS BÁSICOS DO CONTROLE ESTATÍSTICO DE PROCESSOS**

O controle estatístico de processos é uma atividade que tem por objetivo garantir a qualidade dos produtos industriais e melhorar o desempenho dos processos. A implementação do CEP consiste na utilização de algumas ferramentas de análise como o

gráfico de controle, o gráfico de Pareto e o diagrama de causa e efeito.<sup>170</sup> O gráfico de controle mostrado na figura 6.1, é a ferramenta básica do CEP e é usado para caracterizar e monitorar em tempo real uma propriedade que mede a qualidade do produto.

Nos processos industriais atuam diversas variáveis que podem ou não afetar a qualidade do produto. A figura 6.2 faz uma distinção entre as variáveis existentes. Algumas destas podem ser manipuladas durante a operação como a temperatura e pressão interna do reator e são ditas variáveis controladas, enquanto que outras variáveis não podem ser controladas. O produto do processo industrial normalmente é caracterizado por uma ou mais variáveis de saída que são usadas como medida da qualidade deste produto. As variáveis de entrada costumam ter grande influência na qualidade do produto como a presença de impurezas ou a composição da matéria prima de entrada.



**Figura 6.2.** Tipos de variáveis presentes nos processos industriais

No CEP convencional, que neste texto definimos como sendo o método univariado, somente uma variável de qualidade é monitorada, geralmente uma variável da saída  $y_j$ . Quando mais de uma variável é necessária para a caracterização do produto, o CEP

convencional trata cada uma destas isoladamente sem levar em conta as interações entre elas. No CEP multivariado, mais de uma variável de qualidade é analisada simultaneamente e a possível correlação entre as variáveis é modelada implicitamente.

O gráfico de controle tal como aquele mostrado na figura 6.1 é uma representação gráfica de um modelo estatístico que descreve a distribuição de probabilidade (Material suplementar: Parte E) ajustada a partir de um conjunto de amostras coletadas do processo industrial. O mais comum é usar uma distribuição de probabilidade normal para modelar o processo de modo que a variável monitorada pode ser caracterizada por uma média e uma variância.

Se o processo está em controle, as medidas da variável de qualidade apresentam valor médio  $\mu_0$  e variância  $\sigma^2$  constantes ao longo do tempo. A variabilidade de um processo em controle é atribuída somente à causas aleatórias que, frequentemente, não podem ser controladas. Se, além das causas aleatórias o processo estiver sujeito a causas especiais, ele está fora de controle.

Se o processo está em controle, podemos fazer uma série de previsões ou análises das características operatórias do processo. Para um processo em controle, podemos determinar o número de produtos defeituosos que serão produzidos durante um período de tempo e também determinar o índice de capacidade do processo que informa a capacidade de um dado processo de fabricar produtos dentro da faixa de especificação estipulada.<sup>170</sup>

Para o monitoramento de um processo em controle estatístico utilizam-se dois gráficos de controle denominados de Gráfico de médias ou Gráfico  $\bar{x}$  e Gráfico de amplitudes ou Gráfico  $R$ , estes são usados, respectivamente, para monitorar a média e a variância do processo. Se um ponto for assinalado fora dos limites do Gráfico  $\bar{x}$  significa que a amostra em questão apresenta média  $\mu_1$  diferente da média da população  $\mu_0$  quando em controle estatístico. Nesta condição existe algum tipo variável atuando no processo que desvia este de sua condição normal de operação (CNO) e dizemos que o processo está fora de controle estatístico devido à ação de uma causa especial. Conclusões semelhantes são feitas para um ponto assinalado fora dos limites do Gráfico  $R$ , porém, neste caso é a variância do processo que foi alterada.

Quando um processo é assinalado fora de controle estatístico, a próxima tarefa consiste em identificar a variável responsável pela causa especial. Para esta tarefa normalmente são utilizados o diagrama de causa e efeito e o Gráfico de Pareto.<sup>170</sup> Quando é identificada a variável responsável pela causa especial, esta é removida ou modificada para que o processo retorne ao estado de controle.

O princípio do método de controle estatístico multivariado de processos é igual ao do método univariado, a única diferença é que é usada a distribuição de probabilidade normal multivariada para modelar o processo. Portanto, no método multivariado ao invés de termos um único parâmetro  $\mu_0$  e  $\sigma^2$ , dispomos de dois vetores  $\mu_0$  e  $\sigma^2$  para as médias e variâncias – covariâncias das variáveis que compõe uma distribuição normal multivariada (Material Suplementar: Parte E).

### 6.3 GRÁFICOS DE CONTROLE UNIVARIADOS

Os métodos usados para a construção dos gráficos de controle univariados são bem documentados na literatura, e existem vários livros que tratam deste assunto.<sup>170,171</sup> Por isso, neste texto será feita apenas uma breve explanação.

A construção do Gráfico  $\bar{x}$  consiste em coletar amostras de tamanho  $N$  ao longo do tempo. Sendo assim, uma amostra é um conjunto de medidas ou observações  $x_1, x_2, \dots, x_n, \dots, x_N$  coletadas no instante  $i$ , que medem uma variável da qualidade do produto ou do processo. A média amostral, representada por  $\bar{x}_i$ , é calculada conforme a equação abaixo:

$$\bar{x}_i = \frac{1}{N} \sum_{n=1}^N x_{in} \quad i = 1, 2, \dots, I \quad 6.1$$

onde o índice  $N$  refere-se ao tamanho do conjunto de observações e o índice  $i$  refere-se ao instante em que o conjunto foi coletado. Assumindo que estas amostras são provenientes de uma população com distribuição normal de média  $\mu_0$  e variância  $\sigma^2$ , então há uma probabilidade  $1-\alpha$  da média  $\bar{x}_i$  cair dentro dos limites inferior (LI) e superior (LS) calculados de acordo com as equações:

$$LS = \mu_0 + z_{\alpha/2} \frac{\sigma}{\sqrt{N}}$$

$$\text{media} = \mu_0$$

**6.2**

$$LI = \mu_0 - z_{\alpha/2} \frac{\sigma}{\sqrt{N}}$$

O índice  $\alpha$  refere-se a um valor de probabilidade de modo que:  $0 \leq \alpha \leq 1$ . Na equação 6.2,  $Z_{\alpha/2}$  é uma constante que mede a distância dos limites de controle em relação à média. O famoso limite 3-sigma é obtido substituindo o valor  $Z_{\alpha/2}$  por 3 nas equações 6.2. O gráfico de controle apresentado na figura 6.1 é formado pelas médias amostrais  $\bar{x}_i$ , a linha central composta pela média  $\mu_0$ , mais as linhas que definem o limite superior (LS) e inferior (LI) calculadas de acordo com equação 6.2.

Quando os valores de  $\mu_0$  e de  $\sigma^2$  não são conhecidos ou não podem ser determinados, é possível usar estimativas destas estatísticas calculadas a partir das amostras retiradas do processo que opera em controle. Neste caso, a melhor estimativa de  $\mu_0$  é aquela calculada a partir da equação 6.3.

$$\bar{\bar{x}} = \frac{1}{I} \sum_{i=1}^I \bar{x}_i$$

**6.3**

onde  $\bar{\bar{x}}$  é a “grande média” usada como a linha central da carta de controle e  $\bar{x}_i$  é média amostral para uma amostra coletada no instante  $i$ , calculada de acordo com a equação 6.1. O índice  $I$  é o número total de amostras coletadas ao longo do tempo.

O intervalo, denominado de  $R$ , é a diferença entre o maior e o menor valor das observações que compõe o conjunto de uma amostra de tamanho  $N$ :

$$R = X_{\max} - X_{\min}$$

**6.4**

Seja  $R_1, R_2, \dots, R_i, \dots, R_I$  os intervalos calculados para as  $I$  amostras, então o intervalo médio é definido como:

$$\bar{R} = \frac{1}{R} \sum_{i=1}^I R_i \quad 6.5$$

O intervalo é uma estatística usada para obter uma estimativa do desvio padrão da distribuição de probabilidade normal que modela a variável da qualidade. Após calcular as médias amostrais e os intervalos, o Gráfico de controle  $\bar{x}$  podem ser construído de acordo com a equação 6.6, com base nas estatísticas  $\bar{x}$  e  $\bar{R}$ .

$$\begin{aligned} LS &= \bar{\bar{x}} + A_2 \bar{R} \\ \text{média} &= \bar{\bar{x}} \\ LI &= \bar{\bar{x}} - A_2 \bar{R} \end{aligned} \quad 6.6$$

Os valores da constante  $A_2$  são encontrados em tabelas para vários tamanhos de amostras,  $N$ .

O gráfico  $R$  pode ser construído de acordo com a equação 6.7:

$$\begin{aligned} LS &= D_4 \bar{R} \\ \text{linha central} &= \bar{R} \\ LI &= D_3 \bar{R} \end{aligned} \quad 6.7$$

Os valores das constantes  $D_4$  e  $D_3$  são tabulados para diversos valores de  $N$ .

### 6.3.1 Gráficos de controle univariado para observações individuais

Os processos químicos apresentam uma série de características que costumam afetar o desempenho dos gráficos de controle  $\bar{x}$  e  $R$ . Uma destas características é o fato de que o número de observações que podem ser obtidas para uma amostra é geralmente um, ou seja,  $N = 1$ . Neste caso, há  $I$  amostras de tamanho  $N = 1$  ou seja as amostras são formadas por uma observação individual e o gráfico de controle de médias amostrais é construído de acordo com a equação 6.8.

$$\begin{aligned}
 LS &= \bar{x} + 3 \frac{MR}{d_2} \\
 \text{linha central} &= \bar{x} \\
 LI &= \bar{x} - 3 \frac{MR}{d_2}
 \end{aligned}
 \tag{6.8}$$

onde o parâmetro  $MR$  é a média dos valores da amplitude móvel,  $MR_i$ , calculadas de acordo com a equação 6.9.

$$MR_i = |x_i - x_{i-1}| \quad i = 2, 3, \dots, l \tag{6.9}$$

O operador  $| |$  é o módulo. Se a amplitude móvel é calculada entre duas observações individuais de acordo com a equação 6.9, então:  $d_2 = 1.128$ .

### 6.3.2 Gráficos de controle de médias móveis ponderadas exponencialmente. (“EWMA”)

Gráficos de controle do tipo Shewhart usam somente a informação da observação feita em um instante específico e ignoram as medidas das observações realizadas em instantes anteriores. Em função disto, os gráficos de controle do tipo Shewhart são relativamente insensíveis a pequenos deslocamentos da média do processo com uma ordem de magnitude inferior a  $1.5\sigma$ .

Para detectar deslocamentos pequenos da média ou nos casos em que existe apenas uma observação individual, utilizam-se os gráficos de médias móveis ponderadas exponencialmente (EWMA),<sup>177</sup> ou os gráficos de somas acumuladas (“CUSUM”)<sup>170</sup> visto que estes gráficos apresentam desempenho melhor nestas situações específicas.

O gráfico de controle “EWMA” foi introduzido por Roberts<sup>177</sup> em 1959, que definiu a estatística de médias móveis exponencialmente ponderadas conforme a equação abaixo:

$$z_i = Lx_i + (1-L)z_{i-1} \tag{6.10}$$

onde  $x_i$  é a observação feita no instante  $i$  e  $L$  é o fator de ponderação que varia entre zero e um. O valor inicial da média ponderada é o valor alvo da propriedade, ou  $z_0 = \mu_0$ .

Com a estatística  $z$ , as medidas mais próximas de  $x_i$  são priorizadas porque recebem pesos maiores enquanto que as mais afastadas apresentam menos influência. Para melhorar a detecção de pequenos desvios da média deve-se utilizar fatores de ponderação pequenos sendo uma escolha típica:  $L = 0,2$ .

Se as observações  $x_i$  são independentes ao longo do tempo e com variância  $\sigma^2$ , então a variância de  $z_i$  é expressa por:

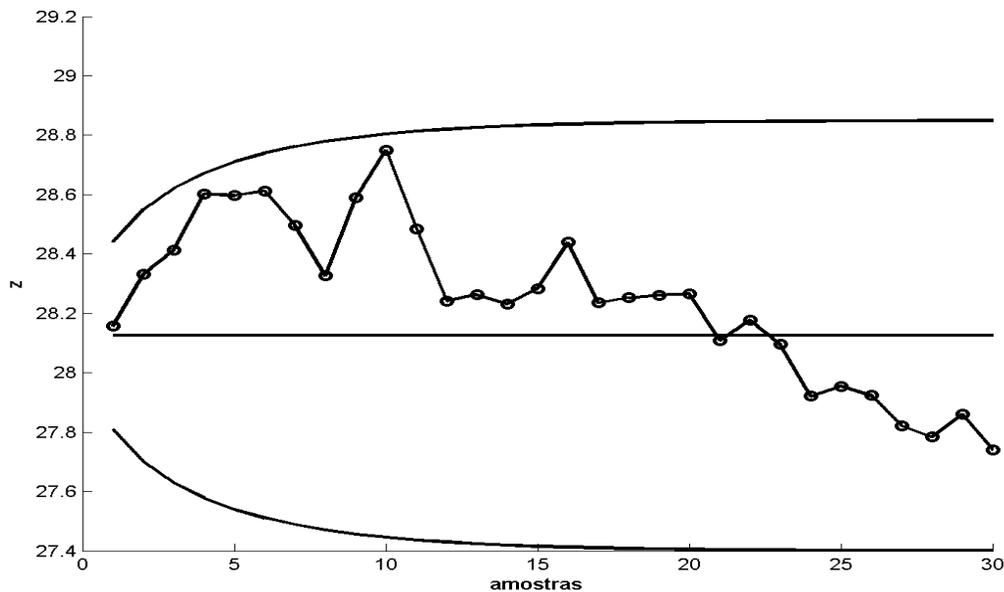
$$\sigma_{z_i}^2 = \sigma^2 \left( \frac{L}{2-L} \right) [1 - (1-L)^{2i}] \quad \mathbf{6.11}$$

de modo que os limites de controle são definidos por:

$$\begin{aligned} LS &= \mu_0 + W\sigma \sqrt{\left( \frac{L}{2-L} \right) [1 - (1-L)^{2i}]} \\ LI &= \mu_0 - W\sigma \sqrt{\left( \frac{L}{2-L} \right) [1 - (1-L)^{2i}]} \end{aligned} \quad \mathbf{6.12}$$

O fator  $W$  é a distância dos limites de controle em relação à média do processo. Ambos parâmetros  $W$  e  $L$  são escolhidos de acordo com o índice  $NMA$  que é uma medida do número médio de observações que serão feitas até uma observação ser assinalada como fora de controle, independente do processo estar fora ou em controle estatístico. Uma das escolhas mais comuns é o  $NMA = 200$ .

De acordo com a equação 6.12, os limites de controle da carta EWMA variam em função do número de amostras, e tendem a um valor constante à medida que aumenta tempo. A figura 6.3 mostra uma carta EWMA.



**Figura 6.3.** Exemplo de uma carta de controle do tipo EWMA.

## 6.4 GRÁFICOS DE CONTROLE MULTIVARIADO

O controle estatístico multivariado trata dos casos em que é necessário monitorar simultaneamente duas ou mais variáveis da qualidade ou processo correlacionadas entre si. Os primeiros trabalhos relacionados ao CEMP foram desenvolvidos por Hotelling<sup>172,178</sup> na década de 1930, posteriormente Jackson ampliou a metodologia de Hotelling incorporando a técnica de análise em componentes principais (PCA) para o desenvolvimento das cartas de controle multivariado.<sup>179</sup>

Um dos gráficos de controle multivariado é gráfico  $\chi_o^2$ . Este gráfico assume que as variáveis que medem a qualidade do produto seguem uma distribuição multivariada normal quando o processo opera em controle estatístico. A construção do gráfico  $\chi_o^2$  é feita com base nas  $I$  amostras coletadas ao longo do tempo, sendo que cada amostra coletada no instante  $i$  é formada por um conjunto de amostras de tamanho  $N$ , onde foram realizadas medições em  $J$  variáveis de qualidade. Um conjunto de observações multivariadas do instante  $i$  é representado por:  $X_{i11}, X_{i21}, \dots, X_{iN1}; X_{i12}, X_{i22}, \dots, X_{iN2}, \dots, X_{i1J}, X_{i2J}, \dots, X_{iNJ}$ .

A média amostral multivariada pode ser calculada de acordo com a equação 6.2.

$$\bar{x}_{ij} = \frac{1}{N} \sum_{n=1}^N x_{inj} \quad j = 1, 2, \dots, J \text{ e } i = 1, 2, \dots, I \quad \mathbf{6.13}$$

onde o índice  $i$  refere-se ao instante em que a amostra foi coletada, o índice  $N$  ao número de observações do conjunto que forma a amostra e o índice  $j$  uma variável de qualidade. As médias amostrais de cada variável de qualidade  $j$  são organizadas em um vetor de médias  $\bar{\mathbf{x}}_j$  conforme mostra a equação 6.13.

$$\bar{\mathbf{x}}_j = \begin{bmatrix} \bar{X}_{1j} \\ \bar{X}_{2j} \\ \vdots \\ \bar{X}_{lj} \end{bmatrix} \quad \mathbf{6.14}$$

Considerando que haja um conjunto de  $J$  variáveis de qualidade monitoradas ao longo do tempo, estes dados podem ser organizados em uma matriz  $\bar{\mathbf{X}}$  onde cada coluna corresponde a uma variável de qualidade conforme mostra o esquema abaixo:

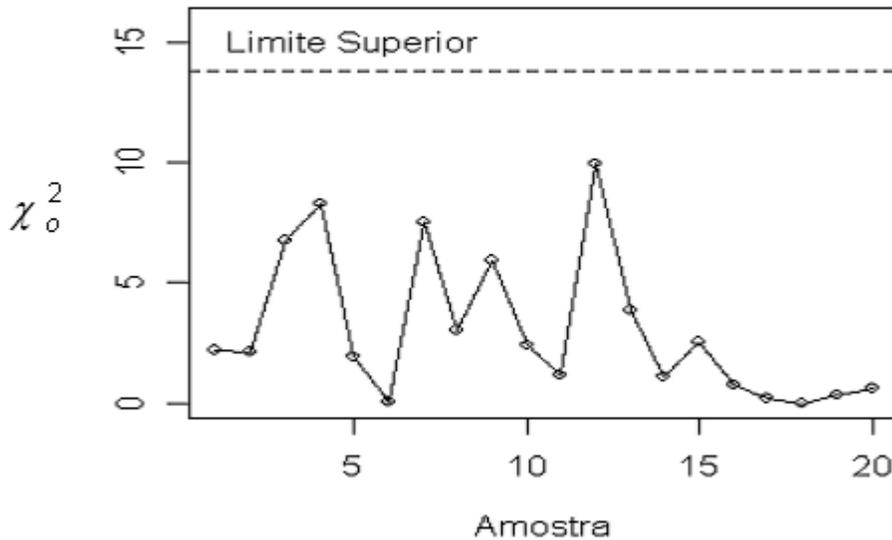
$$\bar{\mathbf{X}} = \begin{bmatrix} \bar{X}_{11} & \bar{X}_{12} & \cdots & \bar{X}_{1J} \\ \bar{X}_{21} & \bar{X}_{22} & \cdots & \bar{X}_{2J} \\ \vdots & \vdots & \vdots & \vdots \\ \bar{X}_{l1} & \bar{X}_{l2} & \cdots & \bar{X}_{lJ} \end{bmatrix} \quad \mathbf{6.15}$$

A estatística  $\chi_o^2$  é calculada de acordo com a equação 6.16, onde  $\boldsymbol{\mu}$  e  $\boldsymbol{\Sigma}$  são respectivamente o vetor da média e a matriz de variância-covariância da distribuição dos dados. O limite de controle superior para a carta  $\chi_o^2$  é calculado a partir da distribuição qui-quadrado com um nível de confiança  $\alpha$  e  $\nu$  graus de liberdade conforme mostra a equação 6.16.

$$\chi_{oi}^2 = N(\bar{\mathbf{x}}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\bar{\mathbf{x}}_i - \boldsymbol{\mu}) \quad i = 1, 2, \dots, I \quad \mathbf{6.16}$$

$$LS = \chi_{\alpha, \nu}^2$$

O gráfico de controle  $\chi_o^2$  consiste em representar a estatística  $\chi_{oi}^2$  em função do tempo ou das amostras, com uma linha horizontal que descreve o limite de controle superior como mostra a figura 6.4.



**Figura 6.4.** Gráfico de controle  $\chi_o^2$ .

Na prática as estatísticas  $\mu$  e  $\Sigma$  não são conhecidas e uma estimativa destes parâmetros estatísticos tem que ser feita a partir do conjunto de amostras coletadas do processo ao longo do tempo quando ele está em controle. Uma estimativa de  $\mu$ , denominada de “grande média” multivariada, pode ser calculada a partir da equação 6.17, usando as médias amostrais calculadas para cada variável.

$$\bar{\bar{x}}_j = \frac{1}{I} \sum_{i=1}^I \bar{x}_{ij} \quad j = 1, 2, \dots, J \quad 6.17$$

As “grande médias” calculadas a partir da equação 6.17 podem ser agrupadas em um vetor  $\bar{\bar{\mathbf{x}}}$  conforme mostra o esquema abaixo:

$$\bar{\mathbf{x}}^T = [\bar{x}_1, \bar{x}_1, \dots, \bar{x}_J] \quad \mathbf{6.18}$$

Para obter um estimativa da matriz de variância-covariância  $\Sigma$ , primeiro calcula-se as variâncias e covariâncias de cada amostra coletada no instante  $i$  conforme a equação abaixo:

$$s_{iuj}^2 = \frac{1}{N-1} \sum_{n=1}^N (x_{inu} - \bar{x}_{iu})(x_{inj} - \bar{x}_{ij}) \quad i=1, 2, \dots, I; \quad u=1, 2, \dots, U; \\ ; j=1, 2, \dots, J; \quad \mathbf{6.19}$$

As variâncias e covariâncias de cada amostra calculadas com a equação 6.19 são os elementos da matriz  $\mathbf{S}_i$  como mostra a equação 6.20. Finalmente, a estimativa da matriz de variância e covariância  $\Sigma$  é obtida calculando a matriz  $\bar{\mathbf{S}}$  como mostra a equação 6.21.

$$\mathbf{S}_i = \begin{bmatrix} s_{i11}^2 & s_{i12}^2 & \dots & s_{i1J}^2 \\ & s_{i22}^2 & \dots & s_{i2J}^2 \\ & & \ddots & \vdots \\ & & & s_{iJJ}^2 \end{bmatrix} \quad \mathbf{6.20}$$

$$\bar{\mathbf{S}} = \frac{1}{I} \sum_{i=1}^I \mathbf{S}_i \quad \mathbf{6.21}$$

A partir das estimativas de  $\mu$  e  $\Sigma$ , podemos calcular a estatística  $T^2$ , conforme mostra a equação 6.22, que é usada para a construção do gráfico de Hottelling, denominado também de gráfico  $T^2$ .

$$T_i^2 = N(\bar{\mathbf{x}}_i - \bar{\bar{\mathbf{x}}})^T \bar{\mathbf{S}}^{-1} (\bar{\mathbf{x}}_i - \bar{\bar{\mathbf{x}}}) \quad i=1, 2, \dots, I \quad \mathbf{6.22}$$

A construção do gráfico de controle deve levar em conta as duas fases distintas do monitoramento estatístico. A primeira fase é a etapa que tem por objetivo verificar se o processo estava em controle estatístico quando as observações foram coletadas. O objetivo desta fase é definir o conjunto de dados que será usado para calcular os limites de controle para a segunda fase. Na primeira fase, os limites da carta  $T^2$  são calculados de acordo com a equação 6.23.

$$LS = \frac{J(I-1)(N-1)}{IN-I-J+1} F_{\alpha, J, IN-I-J+1} \quad \mathbf{6.23}$$

$$LI = 0$$

onde o símbolo  $F$  é a distribuição de probabilidade  $F$ .

Na segunda fase, os gráficos de controle são usados para monitorar as novas observações coletadas do processo, a fim de detectar se estas estão em ou fora de controle estatístico. Nesta fase são utilizados os limites de controle calculados de acordo com a equação 6.24.

$$LS = \frac{J(I+1)(N-1)}{IN-I-J+1} F_{\alpha, J, IN-I-J+1} \quad \mathbf{6.24}$$

$$LI = 0$$

#### **6.4.1 Gráficos de controle multivariados para observações individuais**

As amostras coletadas em um processo químico normalmente são constituídas por uma observação individual ( $N = 1$ ). Neste caso, uma observação multivariada consiste nas medidas  $x_{i1}, x_{i2}, \dots, x_{iJ}$  para cada amostra  $i$ , onde o índice  $j$  indica uma variável de qualidade. O conjunto de medidas obtidos para todas as  $I$  amostras como denifinado na equação 6.15 se reduz a:

$$\mathbf{X} = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1J} \\ X_{21} & X_{22} & \cdots & X_{2J} \\ \vdots & \vdots & \vdots & \vdots \\ X_{I1} & X_{I2} & \cdots & X_{IJ} \end{bmatrix} \quad \mathbf{6.25}$$

A matriz de variância-covariância da matriz  $\mathbf{X}$  neste caso se reduz a:

$$\mathbf{S} = \frac{1}{I-1} \sum_{i=1}^I (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T \quad i = 1, 2, \dots, I \quad \mathbf{6.26}$$

onde os elementos do vetor  $\bar{\mathbf{x}}$  são as médias das respectivas colunas de  $\mathbf{X}$ .

A estatística  $T^2$  para um conjunto de observações individuais é calculada de acordo com a equação 6.27 (de modo semelhante à equação 6.22):

$$T_i^2 = (\mathbf{x}_i - \bar{\mathbf{x}})^T \mathbf{S}^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}) \quad i = 1, 2, \dots, I \quad \mathbf{6.27}$$

Para um conjunto de observações multivariadas individuais ( $N = 1$ ) a estatística  $T^2$  segue uma distribuição beta<sup>180</sup> na fase 1, e os limites de controle devem ser calculados de acordo com a equação 6.28.

$$LS = \frac{(I-1)^2}{I} \beta_{J/2, (I-J)/2} \quad \mathbf{6.28}$$

$$LI = 0$$

Na fase 2, os limites de controle da carta  $T^2$  são calculados com a equação 6.29.

$$LS = \frac{J(I+1)(I-1)}{I(I-J)} F_{1-\alpha/2; J, I-J} \quad \mathbf{6.29}$$

Uma estimativa robusta da matriz de covariância para observações individuais pode ser obtida por um método de cálculo semelhante ao das amplitudes móveis conforme

mostra a equação 6.30.<sup>181</sup> Além disto, esta estimativa da matriz de covariância apresenta um desempenho melhor para a detecção de deslocamentos da média multivariada quando comparada a estimativa calculada com a equação 6.27.

$$\mathbf{S} = \frac{1}{2} \frac{\mathbf{G}^T \mathbf{G}}{l-1} \quad \mathbf{6.30}$$

onde os elementos da matriz  $\mathbf{G}$  são calculados de acordo com a equação 6.31:

$$g_{ij} = x_{ij} - x_{(i-1)j} \quad i = 2, 3, \dots, l \text{ e } j = 1, 2, \dots, J \quad \mathbf{6.31}$$

#### 6.4.2 Gráficos de controle multivariado das médias móveis ponderadas exponencialmente (“MEWMA”)

A carta MEWMA,<sup>182</sup> é uma extensão da carta univariada EWMA para a análise multivariada. A estatística  $\mathbf{Z}$  é calculada conforme a equação abaixo:

$$\mathbf{Z}_i = L\mathbf{X}_i + (1-L)\mathbf{Z}_{i-1} \quad \mathbf{6.32}$$

onde  $0 < L < 1$  e  $\mathbf{z}_0 = \boldsymbol{\mu}_0$ . Os elementos da matriz  $\mathbf{Z}$  correspondem à estatística das médias móveis ponderadas exponencialmente. Na equação 6.32, todas as variáveis são ponderadas de acordo com o mesmo peso, mas no caso de se usar pesos diferentes, o parâmetro  $L$  passa a ser uma matriz diagonal  $\mathbf{L}$ , e o número 1 da equação 6.32 passa a ser a matriz identidade  $\mathbf{I}$ , como mostra a equação 6.33.

$$\mathbf{Z}_i = L\mathbf{X}_i + (\mathbf{I}-L)\mathbf{Z}_{i-1} \quad i = 1, 2, \dots, l \quad \mathbf{6.33}$$

A matriz de variância-covariância da matriz  $\mathbf{Z}$  é calculada conforme a equação abaixo:

$$\mathbf{S}_z = \frac{L}{2-L} [1 - (1-L)^{2i}] \mathbf{S} \quad 6.34$$

onde  $\mathbf{S}$  é a matriz de variância-covariância dos dados originais. À medida que  $i$  aumenta, o fator de ponderação torna-se constante de modo que a matriz de covariância  $\mathbf{S}_z$  tende assintoticamente para:

$$\mathbf{S}_z = \frac{L}{2-L} \mathbf{S} \quad 6.35$$

A estatística MEWMA- $T^2$  é calculada conforme:

$$T_i^2 = \mathbf{Z}_i^T \mathbf{S}_z^{-1} \mathbf{Z}_i > h \quad 6.36$$

onde  $h$  é um limite de controle escolhido de modo a obter um *NMA* específico.<sup>183</sup>

### 6.4.3 Aplicação da Análise em Componentes Principais (PCA) para a construção dos gráficos de controle multivariado

A análise em componentes principais (PCA)<sup>184,185</sup> é um dos métodos mais representativos da quimiometria, visto que esta é geralmente um “meio para um fim”. PCA é usada para a análise de regressão, para a análise classificatória, resolução multivariada de curvas, para a compressão de imagens e para o controle estatístico multivariado de processos.

Matematicamente, o método PCA corresponde a uma decomposição da matriz  $\mathbf{X}$  no modelo bilinear definido abaixo:

$$\mathbf{X} = \mathbf{t}_1 \mathbf{p}_1^T + \mathbf{t}_2 \mathbf{p}_2^T + \dots + \mathbf{t}_R \mathbf{p}_R^T = \mathbf{TP}^T + \mathbf{E} \quad 6.37$$

onde o produto  $\mathbf{T}_R \mathbf{P}_R^T$  é a estrutura do modelo com  $R$  componentes principais e  $\mathbf{E}$  é a matriz dos resíduos. O ajuste do modelo bilinear pode ser feito por diversos métodos como pela decomposição de autovetores e autovalores,<sup>186</sup> decomposição em valores singulares,<sup>186</sup> ou com o algoritmo NIPALS.<sup>187</sup>

Os vetores  $\mathbf{p}_r$  são os autovetores da matriz de covariância, equação 6.27, tal que:

$$\mathbf{S} \mathbf{p}_r = \lambda_r \mathbf{p}_r \quad \mathbf{6.38}$$

Na equação 6.38, há um autovalor  $\lambda_r$  associado a cada autovetor que mede a variância capturada pelo respectivo autovetor. Os autovetores  $\mathbf{p}_1, \dots, \mathbf{p}_R$  são linearmente independentes e ortonormais, pois  $\mathbf{P}^T \mathbf{P} = \mathbf{I}$ .

Em PCA, os autovetores  $\mathbf{p}_r$  são denominados de “loadings” e estão relacionados com as variáveis originais. A matriz  $\mathbf{T}$ , denominada de “scores”, é formada pela projeção dos vetores originais  $\mathbf{x}_i^T$  no novo conjunto de ordenadas representado pelos autovetores  $\mathbf{p}_r$ , o que significa que as medidas realizadas para cada amostra são projetadas em um novo sistema de coordenadas definido pelos autovetores. Cada vetor  $\mathbf{t}_r$  é calculado de acordo com:

$$\mathbf{t}_r = \mathbf{X} \mathbf{p}_r \quad r = 1, 2, \dots, R \quad \mathbf{6.39}$$

Cada componente do modelo PCA é definida por um produto  $\mathbf{t}_r \mathbf{p}_r^T$  e estas são organizadas em ordem decrescente de variância, medida pelo autovalor associado.

Se as colunas de  $\mathbf{X}$  são correlacionadas, o número de autovetores linearmente independentes é menor que o número de variáveis originais e neste caso o método PCA realiza uma compressão do número de variáveis necessários para descrever a matriz  $\mathbf{X}$ . Esta compressão de variáveis é uma das características mais importantes do método PCA e traz algumas vantagens quando utilizado para o CEMP. Ela é importante porque reduz o número de variáveis que será usado para a construção dos gráficos de controle multivariado.

O primeiro pesquisador a mencionar o uso do método PCA para a construção dos gráficos de controle multivariado foi Jackson no final da década de 1950.<sup>179</sup> Ele usou os

“scores” derivados do modelo bilinear (equação 6.41) ajustado pelo método PCA para o desenvolvimento dos gráficos  $T^2$ , que atualmente são denominados de gráficos  $D$ . A estatística  $D$  é calculada com a equação 6.40 assumindo que os dados da matriz  $\mathbf{X}$  foram autoescalados antes da decomposição no modelo bilinear expresso pela equação 6.37 com o método PCA.

$$D_i = \mathbf{x}_i^T \left( \frac{\mathbf{P}\boldsymbol{\lambda}_R^{-1}\mathbf{P}^T}{I-1} \right) \mathbf{x}_i \quad i = 1, 2, \dots, I \quad \mathbf{6.40}$$

Na equação 6.40, o vetor  $\mathbf{x}_i^T$  é uma linha da matriz  $\mathbf{X}$ , a matriz  $\mathbf{P}$  são os loadings do modelo PCA e a matriz diagonal  $\boldsymbol{\lambda}_R$  é formada pelos  $R$  autovalores do modelo PCA. A operação de autoescalar uma matriz é feita de acordo com a equação 6.41:

$$x_{aut} = \frac{x_{ij} - \bar{x}_j}{\sqrt{s_j^2}} \quad i = 1, 2, \dots, I \quad \text{e } j = 1, 2, \dots, J \quad \mathbf{6.41}$$

onde  $x_{aut}$  é um elemento  $x_{ij}$  da matriz  $\mathbf{X}$  autoescalado,  $\bar{x}_j$  e  $s_j^2$  são respectivamente a média e a variância da variável de qualidade  $j$  da matriz  $\mathbf{X}$ .

Na primeira fase do CEMP, o limite de controle do gráfico  $D$  é calculado de acordo com a equação 6.42, e para a segunda fase o limite é calculado com a equação 6.43. Geralmente é representado apenas o limite superior nos gráficos  $D$ .

$$LS = \frac{(I-1)^2}{I} \beta_{R/2, (I-R-1)/2} \quad \mathbf{6.42}$$

$$LS = \frac{R(I+1)(I-1)}{I(I-R)} F_{1-\alpha/2; R, I-R} \quad \mathbf{6.43}$$

No método PCA, é necessário monitorar os resíduos do modelo bilinear, através da estatística  $Q$ . De acordo com a equação 6.37, os resíduos são obtidos a partir de:

$$\mathbf{E} = \mathbf{X} - \mathbf{TP}^T \quad 6.44$$

e a estatística Q conforme a equação abaixo:

$$Q = \sum_{i=1}^I \sum_{j=1}^J (e_{ij})^2 \quad 6.45$$

onde  $e_{ij}$  são os elementos da matriz  $\mathbf{E}$ .

Se os valores de Q seguem uma distribuição de probabilidade normal, o limite de controle da carta Q pode ser aproximado pela equação de Jackson e Mudholkar,<sup>188</sup> definida abaixo:

$$Q_\alpha = \theta_1 \left[ c_\alpha \frac{\sqrt{2\theta_2 h_o^2}}{\theta_1} + 1 + \frac{\theta_2 h_o (h_o - 1)}{\theta_1^2} \right] \frac{1}{h_o} \quad 6.46$$

onde:

$$\theta_a = \sum_{o=R+1}^J \lambda_o^a \quad a = 1, 2, 3 \quad \text{e} \quad h_o = 1 - \frac{(2\theta_1 \theta_3)}{3\theta_2^2} \quad 6.47$$

Nas equações 6.46 e 6.47,  $Q_\alpha$  é o limite de controle para uma probabilidade  $\alpha$ , o índice ou expoente  $a$  é um número que varia de um a três,  $\lambda_o$  é um autovalor da matriz  $\mathbf{X}$ .  $c_\alpha$  é o parâmetro que corresponde ao intervalo  $(1-\alpha)$  da distribuição normal Chi-quadrado.

### 6.4.3.1 Limites de controle baseados em métodos não paramétricos

A questão dos limites de controle, tem sido alvo de estudos. A distribuição normal nem sempre é uma suposição válida para os dados oriundos de processos químicos, e por isto alguns pesquisadores sugerem que sejam utilizados os métodos não paramétricos

para a estimativa dos limites de controle. Um destes, é o método “Kernel Density estimation” (KDE), usado com o objetivo de encontrar uma distribuição adequada para o conjunto de dados a partir de uma análise empírica.<sup>189,190</sup>

## 6.5 CONTROLE ESTATÍSTICO MULTIVARIADO DE PROCESSOS QUÍMICOS INDUSTRIAIS

Desde o surgimento, em 1947, da idéia de utilizar os gráficos de controle multivariado até o final da década de 1980, um número reduzido de trabalhos foram publicados. Da mesma forma não há muitos relatos mencionando a utilização desta técnica na indústria, com algumas exceções como o trabalho de Edward Jackson que combinou o CEMP com o método PCA e aplicou a metodologia para a análise dos processos de produção de filmes fotográficos.<sup>191</sup>

Esta situação mudou radicalmente no início da década de 1990 quando Kresta e MacGregor publicaram um trabalho sobre o CEMP tendo como base os métodos PCA e o método dos mínimos quadrados parciais (PLS) e aplicaram esta metodologia para a análise da simulação do monitoramento de um reator químico e de uma coluna de destilação.<sup>192</sup> No ano seguinte Skagerberg e MacGregor publicaram um outro trabalho na forma de um tutorial para aplicar o CEMP com o método PLS para a análise dos dados de uma simulação de um reator de produção de polietileno de baixa densidade em modo contínuo.<sup>193</sup> Embora estes trabalhos tenham sido desenvolvidos para a análise de dados simulados, estes foram importantes por terem sido os primeiros trabalhos a introduzir os modelos para previsões como uma metodologia para o CEMP.

A proposta de Kresta<sup>192</sup> era de que as variáveis de processos monitoradas de modo automático e *em linha* fossem utilizadas para prever as propriedades que estivessem ligadas à qualidade do produto obtida do processo químico. Segundo os autores, a metodologia consiste em coletar durante um intervalo de tempo medidas de diferentes sensores instalados na linha dos processos químicos, que realizam medidas automáticas, e na saída do processo contínuo coletar durante o mesmo intervalo de tempo amostras do produto e analisar as suas propriedades de qualidade. Desta forma, o conjunto de dados obtidos pode ser organizado em uma matriz **X** com os dados medidos para as variáveis de processo e em uma matriz **Y** formada pelos dados medidos para as

variáveis de qualidade (propriedades) do produto. A correlação entre  $\mathbf{X}$  e  $\mathbf{Y}$  pode ser obtida através de um modelo bilinear tal como mostrado abaixo:

$$\mathbf{y} = \mathbf{Xb} \quad 6.48$$

onde o vetor  $\mathbf{y}$  corresponde as medidas para uma única variável da qualidade e o vetor  $\mathbf{b}$  corresponde ao vetor de regressão que descreve como as variáveis de processo  $\mathbf{X}$  são relacionadas com a qualidade do produto  $\mathbf{y}$ .

PLS<sup>194</sup> foi o método proposto por Kresta<sup>192</sup> para o ajuste do modelo da equação 6.48. O método PLS ajusta o modelo através de uma decomposição das matriz  $\mathbf{X}$  e do vetor  $\mathbf{y}$  em componentes principais, e calcula uma estimativa de  $\mathbf{b}$  a partir dos parâmetros dos modelos obtidos desta decomposição conforme mostram as equações abaixo:

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E}_x \quad 6.49$$

$$\mathbf{y} = \mathbf{Tq} + \mathbf{e}_y \quad 6.50$$

$$\mathbf{b} = (\mathbf{WPW}^T)^{-1}\mathbf{q} \quad 6.51$$

A matriz  $\mathbf{T}$  corresponde aos “scores” do modelo PLS, enquanto que  $\mathbf{P}$  e  $\mathbf{q}$  são os “loadings”.  $\mathbf{W}$  é uma matriz ortonormal calculada durante a decomposição da matriz  $\mathbf{X}$  que é necessária para calcular os “scores” com a restrição de que estes sejam ortogonais. O algoritmo mais conhecido para o ajuste de um modelo PLS é o algoritmo NIPALS,<sup>194</sup> mas atualmente existem outros que tem o mérito de serem mais eficientes para o tratamento de matrizes contendo uma grande quantidade de dados.<sup>195,196</sup> Tal como no método PCA, os parâmetros do modelo PLS também podem ser usados para a construção dos gráficos de controle.

Entre 1994 e 1995 o grupo de pesquisadores orientados por John MacGregor publicou uma seqüência de artigos<sup>197,198,199,200</sup> que firmaram definitivamente o CEMP como uma nova área de pesquisa da química e uma atividade promissora para a então moderna indústria química que encontrava-se bastante carente de métodos de análise empírica que fossem capazes de analisar a quantidade enorme de dados que as plantas

industrias geravam rotineiramente a partir do monitoramento de dezenas de variáveis correlacionadas entre si.

MacGregor e Jaeckle,<sup>197</sup> reconheceram que um processo químico industrial pode ser extremamente complexo devido a presença de muitas operações unitárias diferentes com muitas variáveis sendo monitoradas em cada operação individual. Esta característica torna muitas vezes a modelagem impraticável quando os dados são organizados em uma única matriz  $\mathbf{X}$  e  $\mathbf{y}$ , visto que o modelo que relaciona as variáveis em  $\mathbf{X}$  com a qualidade em  $\mathbf{y}$ , pode ser muito diferente de uma operação unitária para outra. O mesmo é válido para uma operação composta por diversos estágios com características operatórias diferentes. Para estes casos, MacGregor<sup>197</sup> desenvolveu o método PLS em blocos múltiplos a fim de separar o monitoramento global e de cada unidade ou estágio do processo dentro de um mesmo modelo PLS.

O método PLS com blocos múltiplos<sup>201</sup> é capaz de detectar em que estágio ou operação unitária do processo ocorreu um desvio responsável pelo sinal fora de controle. No método com blocos múltiplos, os dados coletados em cada estágio ou operação unitária são organizados em matrizes diferentes  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_K$ , uma para cada estágio ou operação unitária. Estas matrizes são denominadas de blocos, e cada um destes blocos tem a sua própria estrutura bilinear  $\mathbf{T}_R \mathbf{P}_R^T$ , que corresponde a um modelo que pode ser usado para construir gráficos de controle que monitoram cada estágio. Note que no caso dos blocos múltiplos, o índice  $K$  não se refere a uma terceira dimensão como do arranjo em três modos (cubo de dados) mas sim a diferentes matrizes que podem ou não conter modos de monitoramento em comum.

Para tornar mais claro como funciona o método de blocos múltiplos, considere o caso apresentado por MacGregor,<sup>197</sup> formado por um reator de produção de polietileno de baixa densidade que apresenta dois estágios de operação ( $K=1, 2$ ), cada um com as suas próprias variáveis de processo. Os dados que se referem às variáveis de processo de cada estágio podem ser organizados nas matrizes  $\mathbf{X}_1$  e  $\mathbf{X}_2$ . Além destes dados existem também os dados coletados para as propriedades do polietileno que indicam a qualidade deste produto. Estes dados são organizados em uma matriz  $\mathbf{Y}$ . Cada matriz  $\mathbf{X}_k$  pode ser decomposta nos seus respectivos modelos bilineares conforme as equações abaixo:

$$\begin{aligned}\mathbf{X}_1 &= \mathbf{T}_1 \mathbf{P}_1^T + \mathbf{E}_1 \\ \mathbf{X}_2 &= \mathbf{T}_2 \mathbf{P}_2^T + \mathbf{E}_2\end{aligned}\tag{6.52}$$

Os “scores” do modelo de cada estágio podem ser agrupados em uma matriz global  $\mathbf{T}$  da seguinte forma:

$$\mathbf{T} = [\mathbf{T}_1 : \mathbf{T}_2]\tag{6.53}$$

e esta matriz pode ser usada para modelar a relação de cada bloco com a matriz  $\mathbf{Y}$ , conforme o modelo:

$$\mathbf{T} = \bar{\mathbf{T}} \bar{\mathbf{P}}^T + \bar{\mathbf{E}}_x\tag{6.54}$$

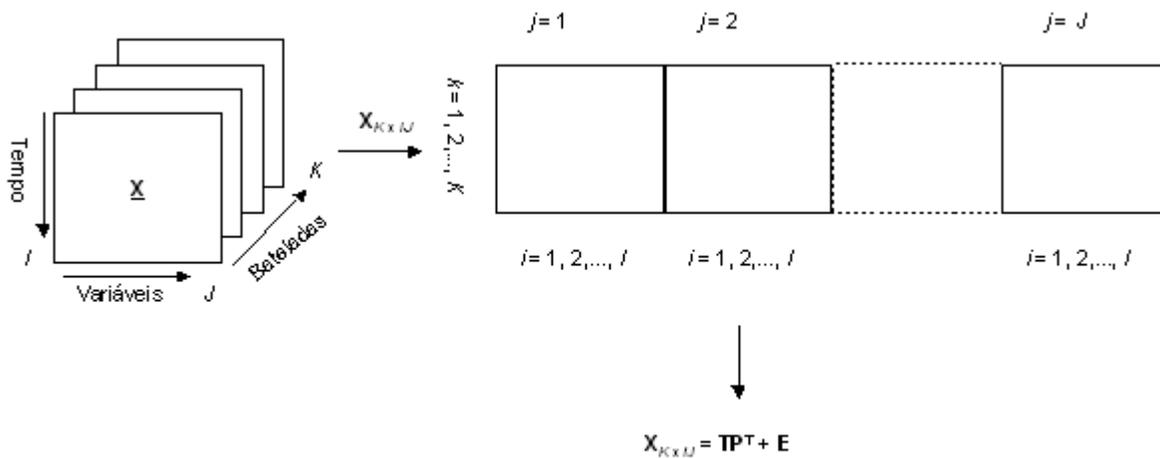
$$\mathbf{Y} = \bar{\mathbf{T}} \bar{\mathbf{Q}}^T + \bar{\mathbf{E}}_y\tag{6.55}$$

$$\mathbf{B} = (\mathbf{W} \bar{\mathbf{P}} \mathbf{W}^T)^{-1} \bar{\mathbf{Q}}^T\tag{6.56}$$

Existem vários algoritmos para o ajuste de modelos PLS e PCA com blocos múltiplos, uma revisão geral do assunto pode ser encontrada nos trabalhos de Smilde<sup>202</sup> e uma aplicação da metodologia a processos industriais no trabalho publicado por Westerhuis.<sup>203</sup>

Em 1994, Nomikos e MacGregor publicaram um artigo original propondo que os dados multivariados coletados para processos químicos industriais que operam de modo discreto (bateladas) deveriam ser organizados em arranjos multivariados com três modos de monitoramento como o mostrado na figura 6.5.<sup>198</sup> A operação em bateladas gera um conjunto de matrizes  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_K$ , uma para cada batelada com os respectivos modos tempo x variáveis de processos. Estes dados podem ser organizados em um arranjo tridimensional  $\underline{\mathbf{X}}$  conforme mostra a figura 6.5. Após o desdobramento do cubo  $\underline{\mathbf{X}}$  na matriz  $\mathbf{X}_{K \times IJ}$ , a compressão do número de variáveis de arranjos de ordem superior, pode ser feita com o método “Multi-Way PLS” ou “Multi-way PCA”,<sup>204</sup> além dos métodos PARAFAC e Tucker.<sup>205,206</sup>

O método “Multi-Way PCA” consiste em desdobrar o cubo  $\underline{X}$  em uma matriz  $\mathbf{X}_{K \times IJ}$ , figura 6.5, para decompor esta matriz no modelo bilinear usando o método PCA. Os “scores” obtidos a partir deste tipo de desdobramento corresponde a projeção das bateladas no espaço definido pelas novas variáveis latentes. Sendo assim os gráficos de controle  $D$  e  $Q$  podem ser construídos a partir deste modelo PCA para verificar se a batelada está sob ou fora de controle estatístico.



**Figura 6.5.** Método “Multi-Way PCA”.

O monitoramento das bateladas pode ser feito após a finalização da batelada, ou durante a operação, sendo que neste último caso é denominado de monitoramento “*on line*”. O monitoramento “*on line*” requer algumas práticas especiais porque durante a operação, dispõe-se apenas das medidas feitas até o instante  $i$ , enquanto que o modelo “Multi way PCA” ou PLS é formado por  $I$  medidas. Sendo assim é necessário prever o comportamento das variáveis para o intervalo de tempo  $I-i$ , ou seja o intervalo que resta entre a medida e a finalização da batelada.

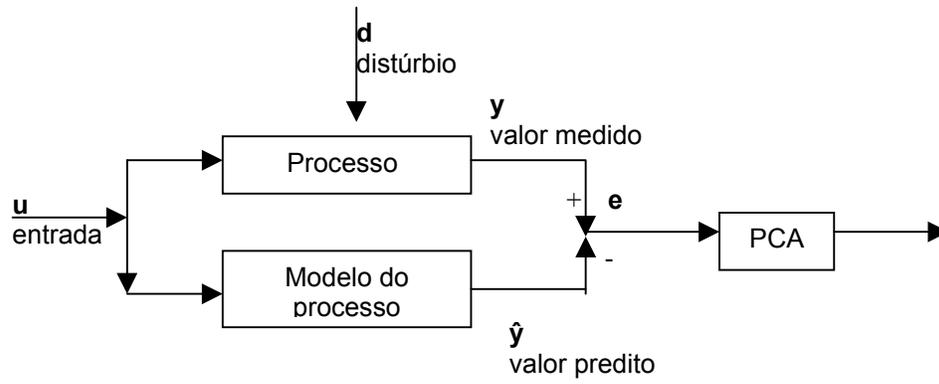
Para preencher os dados que faltam Nomikos<sup>199</sup> sugeriu três metodologias: uma considera que os dados que faltam deveriam ser tratados como valores ausentes, a segunda metodologia consiste em preencher os dados que faltam com zeros, e a terceira consiste em calcular o desvio das medidas feitas para cada variável no instante  $i$ , em

relação a média, e preencher os dados que faltam com este desvio. As duas últimas metodologias são as mais usadas na prática devido a facilidade de implementá-las.

A modelagem empírica de processos químicos com o método PCA pode ser limitada devido ao fato de que este método assume que as componentes principais são variáveis lineares, o que nem sempre corresponde às relações presentes nos sistemas químicos. Para resolver este problema, Dong e MacAvoy sugeriram a modelagem com curvas principais<sup>207</sup> desenvolvidas por Hastie, para o desenvolvimento de um método de análise em componentes principais não linear.<sup>208</sup> Esta técnica foi explorada em aplicações envolvendo processos químicos industriais, e forneceu bons resultados.<sup>209</sup> A versão não linear do método PCA apresenta um desempenho melhor no que diz respeito à eficiência da compressão do número de variáveis. Apesar disto, o método PCA linear tem sido o mais utilizado por se tratar de um método mais simples.

O desenvolvimento de métodos de decomposição em variáveis latentes não lineares tem sido alvo de investigação. Por exemplo, o método PLS não linear combinado a redes neurais<sup>210</sup> ou o método PLS com ajuste de polinômios.<sup>211</sup> O método proposto por Roten,<sup>212</sup> denominado “Model-Based PCA” foi desenvolvido com a intenção de solucionar os problemas associados com a não linearidade e a autocorrelação das observações. No “model-based PCA”, um modelo mecanístico do processo (derivado dos balanços de massa e energia), é usado para prever o comportamento das variáveis de processo. A diferença entre a previsão do modelo mecanístico e as medidas feitas na linha de processo, formam uma matriz de resíduos que é utilizada para análise com o método PCA, figura 6.4. Neste método, o comportamento não linear e/ou a autocorrelação são filtrados pelo modelo mecanístico antes da análise com o modelo PCA.

A desvantagem do método “Model Based PCA” é que para a maioria dos processos químicos não é possível obter um modelo mecanístico suficientemente bom que descreva o comportamento de todas as variáveis de interesse. Cheng propõe a aplicação do algoritmo “Just-in time learning” (JITL)<sup>213</sup> como uma forma de obter o modelo do processo. Utilizando uma base de dados, o método JITL prevê o valor de cada variável do processo para o tempo  $i$  sem ter que usar modelos mecanísticos. O método JITL-PCA foi testado para o controle estatístico em casos simulados, onde mostrou um desempenho melhor que o método PCA linear.



**Figura 6.6.** Esquema do método de monitoramento “Model-Based” PCA.

## E. MATERIAL SUPLEMENTAR: PARTE E

### E.1 Distribuição de Probabilidade Normal

A análise estatística consiste em obter parâmetros numéricos que descrevem uma população de objetos. Definimos como população todos os valores possíveis para uma coleção de objetos finita ou infinita que compõe o sistema sob investigação.

Na prática, quando tratamos de medições em sistemas químicos, estamos quase sempre lidando com objetos que apresentam uma variabilidade para os valores de suas propriedades. Sendo assim os valores  $x_1, x_2, \dots, x_l$  de uma variável aleatória qualquer são diferentes um dos outros, e geralmente flutuam em torno de um valor médio denominado  $\mu$  que constitui o valor médio da população de objetos. Podemos ainda definir uma medida de dispersão dos valores em torno de  $\mu$ , denominada de desvio padrão, representada por  $\sigma$ , enquanto que o quadrado do desvio padrão é a variância representada por  $\sigma^2$ .

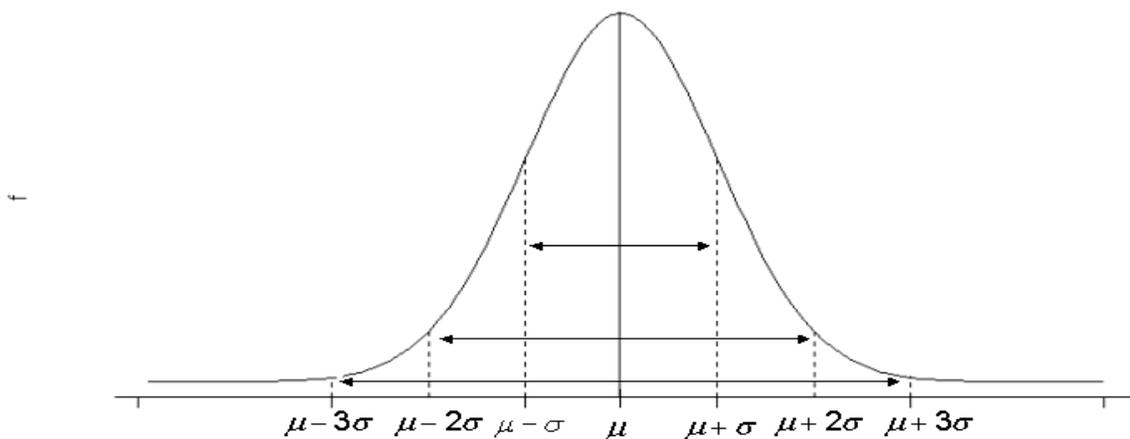
Na análise estatística é comum fazermos uso de uma distribuição de probabilidades para descrever o comportamento da população. Uma distribuição de probabilidades é um modelo matemático que descreve a frequência de ocorrência dos valores de uma variável aleatória. A distribuição normal é um dos modelos de distribuição de probabilidades mais usados, e é descrito por uma função contínua denominada de densidade de probabilidade  $F(x)$ , equação 6.57.

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad -\infty < x < \infty \quad \mathbf{6.57}$$

onde  $x$  é uma variável aleatória,  $\mu$  é a média e  $\sigma^2$  a variância da distribuição. Uma distribuição normal é representada por  $x \approx N(\mu, \sigma^2)$  o que implica que a variável  $x$  apresenta distribuição normal com média  $\mu$  e variância  $\sigma^2$ .

O gráfico da função de distribuição normal produz a bem conhecida curva de Gauss mostrada na figura 6.7. Esta curva é simétrica e unimodal em torno da média  $\mu$ . Há uma

simples interpretação para a variância da distribuição normal conforme mostra a figura 6.7. Tomando o desvio padrão  $\sigma$ , o intervalo dentro da curva definido pelos limites  $\mu - \sigma$  e  $\mu + \sigma$  corresponde a 68,26% da população de valores da curva. Da mesma forma, os intervalos  $\mu \pm 2\sigma$  e  $\mu \pm 3\sigma$  correspondem respectivamente a 95,46% e 99,73% dos valores da distribuição. Estas relações são válidas para qualquer distribuição normal independente da média e da variância.



**Figura 6.7.** Curva de Gauss ou distribuição normal.

A distribuição normal acumulada é definida como:

$$P\{a \leq x \leq b\} = f(a) = \int_{-\infty}^a \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \quad -\infty < x < \infty \quad \mathbf{6.58}$$

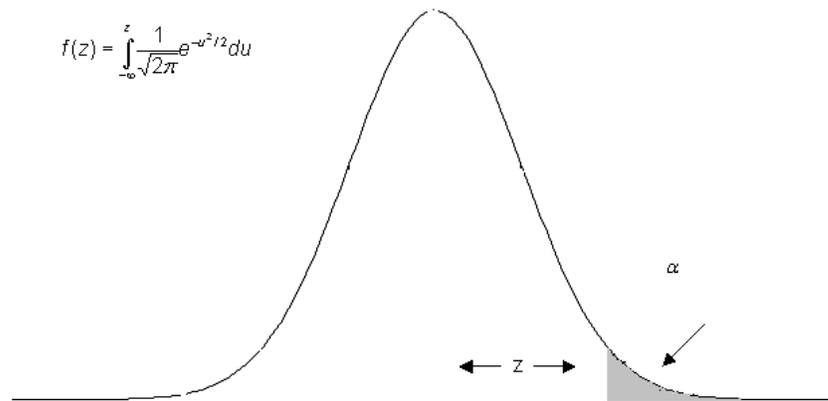
onde  $F(a)$  é a probabilidade de que o valor da variável aleatória  $x$ , de densidade de probabilidade  $F(x)$ , seja observada no intervalo  $[a, b]$ . Esta função tem importante aplicação prática na análise estatística, mas ao invés de integrar  $F(x)$  usando a variável

aleatória  $x$ , normalmente transformamos esta em uma outra variável normal padronizada representada pela letra  $z$ . Esta transformação é feita de acordo com a equação abaixo:

$$z = \frac{X - \mu}{\sigma}$$

**6.59**

Os valores de  $z$  são encontrados em uma tabela disponível na maioria dos livros textos de análise estatística. A variável normal padronizada pertence a uma distribuição normal padrão, com média igual a zero e variância igual a 1, tal que  $z \sim N(0,1)$ . A distribuição normal acumulada mede a área sobre a curva de Gauss correspondente ao intervalo  $[a,b]$ . Para uma distribuição normal padrão, a área total da curva de Gauss é igual a 1, e o intervalo  $1-\alpha$  corresponde a região em branco na figura 6.8 enquanto que a região com uma probabilidade  $\alpha$  corresponde a região da curva indicada pela cor cinza.



**Figura 6.8.** Intervalos de probabilidade sob a área da distribuição normal.

A distribuição de probabilidade normal multivariada é uma generalização da distribuição univariada para mais dimensões, com  $J \geq 2$ . A densidade de probabilidade normal multivariada é expressa pela equação abaixo:

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{J/2} |\boldsymbol{\Sigma}|^{1/2}} e^{-\frac{(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x}-\boldsymbol{\mu})}{2}} \quad -\infty < x_j < \infty \quad j = 1, 2, \dots, J \quad \mathbf{6.60}$$

onde o vetor  $\boldsymbol{\mu}$  de dimensão  $J \times 1$  e matriz  $\boldsymbol{\Sigma}$  de dimensão  $J \times J$ , representam respectivamente o vetor de médias e a matriz de variância - covariância populacionais da matriz  $\mathbf{X}$ , e o vetor  $\mathbf{x}$  é uma amostra  $i$  da população. O operador matemático  $|\cdot|$ , é o determinante da matriz. Um grupo de variáveis aleatórias que seguem a distribuição normal multivariada são representadas por:  $\mathbf{x} \approx N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ .

O expoente da função  $f(\mathbf{x})$ , equação 6.64, é uma constante  $c^2$  que mede o quadrado da distância da amostra  $\mathbf{x}$  a média  $\boldsymbol{\mu}$  em unidade de desvio padrão, tal que:

$$(\mathbf{x} - \boldsymbol{\mu})^T (\boldsymbol{\Sigma})^{-1} (\mathbf{x} - \boldsymbol{\mu}) = c^2$$

A constante  $c^2$  define uma elipse em torno da superfície da densidade de probabilidade normal multivariada, indicando todos os valores de  $\mathbf{x}$  que apresentam a mesma distância quadrática em relação a  $\boldsymbol{\mu}$ . A escolha  $c^2 = \chi_J^2(\alpha)$ , onde  $\alpha$  é a probabilidade da distribuição chi-quadrado com  $J$  graus de liberdade, define uma elipse em torno da superfície da densidade normal que representa um intervalo de valores com probabilidade  $1 - \alpha$  de ocorrência.

## **CAPÍTULO 7**

**CONTROLE ESTATÍSTICO MULTIVARIADO DE PROCESSOS DE  
OBSERVAÇÕES AUTOCORRELACIONADAS: ANÁLISE DO PROCESSO  
BAYER PARA A EXTRAÇÃO DE ALUMÍNIO DA BAUXITA.**

## 7.1 INTRODUÇÃO

O método convencional de controle estatístico de processos assume que as medidas (ou observações) feitas em amostras coletadas de um processo seguem distribuição normal e são independentes entre si. Se estas propriedades são válidas, a variável em questão pode ser usada para monitorar o estado do processo, se em controle ou fora de controle. Neste tipo de monitoramento são usados os gráficos de controle que registram o comportamento da média e desvio padrão das observações ao longo do tempo.

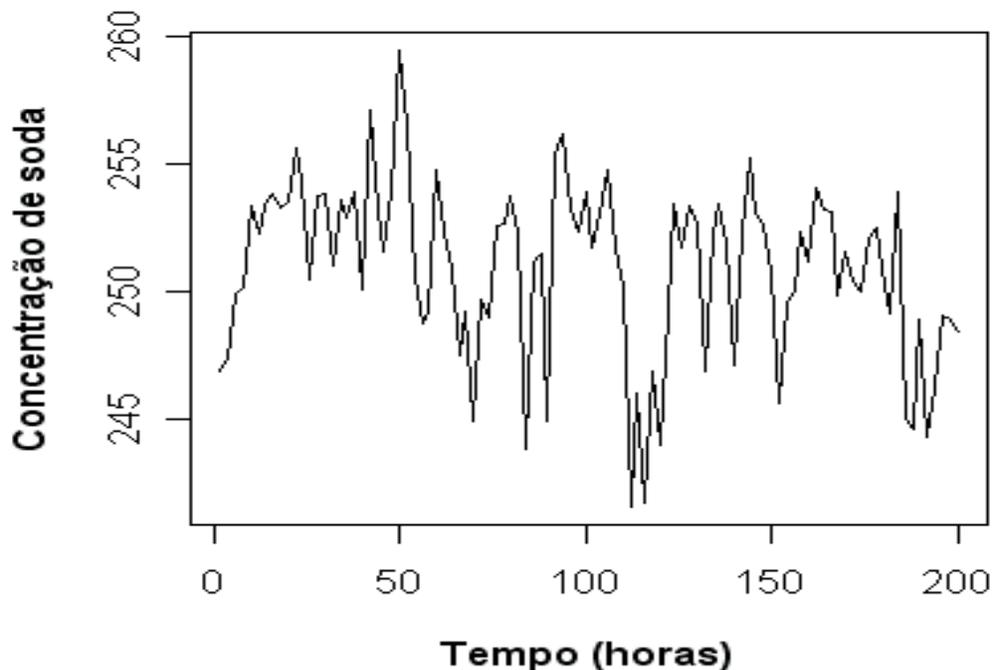
Em muitas situações práticas, os dados não seguem uma distribuição normal exata e pequenos desvios deste comportamento podem ocorrer. Este problema não é tão grave para a análise com os gráficos de controle que são relativamente robustos a pequenos desvios da distribuição normal. O próprio pioneiro da área de controle estatístico Dr. Walter Shewhart não mencionou em seu trabalho original a necessidade dos dados seguirem uma distribuição de probabilidade normal.

Um problema mais grave para o controle estatístico surge quando as observações coletadas ao longo do tempo não são independentes, o que caracteriza a autocorrelação. Nestas circunstâncias, os gráficos de controle convencionais perdem a eficácia porque assinalam muitos pontos fora de controle estatístico quando na verdade o processo encontra-se em controle estatístico.

As variáveis dos processos químicos freqüentemente exibem autocorrelação como conseqüência da própria natureza destes processos. Por exemplo, numa reação química a concentração dos reagentes e produtos em um dado instante dependem da sua concentração em instantes anteriores, de modo que ao longo do tempo teremos uma seqüência de observações autocorrelacionadas. Outro fator que contribui para a autocorrelação é a capacidade de fazer medidas em curtos intervalos de tempo com os atuais procedimentos analíticos que são partes integrantes de plantas industriais.<sup>214</sup>

A figura 7.1A mostra medidas de concentração de soda cáustica feitas ao longo do tempo em amostras do concentrado de aluminato de sódio que foram retiradas de um tanque de digestão do processo Bayer.<sup>215</sup> A correlação entre medidas consecutivas calculada com as observações até duzentas horas foi de 0,45, mostrando que as medidas

de soda cáustica são autocorrelacionadas e por isso não preenchem a exigência de independência entre as medidas, que é necessária para a construção de gráficos convencionais do controle estatístico de processos.



**Figura 7.1.** A) Concentração de soda cáustica determinada em amostras do concentrado de aluminato de sódio retirado do tanque de digestão do processo Bayer. B) Correlograma das medidas de soda cáustica

O efeito negativo da autocorrelação para a análise estatística dos processos com gráficos de controle é um problema bem conhecido mas freqüentemente ignorado, especialmente nas aplicações de controle estatístico multivariado de processos, uma vez que são encontrados muitos trabalhos na literatura relatando a aplicação de gráficos de controle multivariados convencionais para a análise de processos que são autocorrelacionados. Se o problema da autocorrelação é ignorado, o desempenho de um gráfico de controle fica comprometido e com isso esses gráficos tornam-se ferramentas desacreditadas.

Alguns métodos de análise multivariada foram desenvolvidos para a análise de processos autocorrelacionados. Wold<sup>216</sup> desenvolveu o método “MEWMA-PCA” e aplicou

este para a análise de uma planta de purificação do cobre.<sup>217</sup> Ku e outros<sup>218</sup> desenvolveram o método “Dynamic PCA” e testou sua metodologia em dados simulados enquanto que Negiz e Cinar<sup>219</sup> desenvolveu uma metodologia baseada nos modelos de autoregressivos de métricas móveis multivariados (VARMA) para a análise de um processo de pasteurização do leite. Outros trabalhos desenvolvidos dentro deste contexto são os métodos de atualização recursiva dos modelos PLS e PCA desenvolvidos em trabalhos independentes por Dayal,<sup>220</sup> Qin<sup>221</sup> e Li.<sup>222</sup>

Modelos de séries temporais<sup>223</sup> têm sido usados para o controle estatístico univariado de processos autocorrelacionados. Este método, consiste em modelar a estrutura de autocorrelação das observações com um modelo de série temporal apropriado, para obter um conjunto de observações (os resíduos dos modelos de séries temporais) que são independentes e apresentam distribuição normal.

Neste trabalho de tese apresentamos uma metodologia para o controle estatístico multivariado de uma planta industrial do processo Bayer usada para a extração de alumínio da bauxita. O processo Bayer opera em modo contínuo e é monitorado ao longo do tempo com medidas para a temperatura, condutividade, densidade e índice de refração da solução do concentrado de aluminato de sódio processado nos tanques de digestão. As observações feitas ao longo do tempo para cada variável são autocorrelacionadas além de que as quatro variáveis são correlacionadas umas com as outras (correlação cruzada). Por causa da natureza deste processo foi necessário desenvolver uma metodologia de análise multivariada adequada às características de estado deste processo.

O objetivo do trabalho foi o monitoramento estatístico em tempo real do processo com gráficos de controle multivariado. Os modelos de séries temporais foram usados para modelar a estrutura de autocorrelação das observações, a partir dos quais obteve-se um novo conjunto de dados (resíduos dos modelos de séries temporais) que são independentes entre si e que foram introduzidos na análise em componentes principais (PCA) para modelar a correlação cruzada entre as quatro variáveis descritas no parágrafo acima. Os parâmetros do modelo PCA foram usados para construir gráficos de controle do tipo  $D$ ,  $Q$  e  $MWMA-D$ .<sup>224</sup>

Este trabalho é relevante não só no contexto da sua aplicação para o monitoramento da planta de extração de alumínio, mas também porque trata de uma metodologia de CEMP para a análise de medidas autocorrelacionadas, uma característica presente em muitos processos químicos e que tem sido muito pouco investigada. A metodologia proposta neste trabalho é similar a utilizada pelos métodos de controle estatístico univariado para processos autocorrelacionadas no que diz respeito a utilização de séries temporais e é semelhante a metodologia desenvolvida por Spanos<sup>225</sup> para o monitoramento estatístico multivariado de um processo de fabricação de circuitos eletrônicos integrados.

## 7.2 SÉRIES TEMPORAIS: PROCESSOS AUTOCORRELACIONADOS

Uma série temporal é uma coleção de medidas obtidas seqüencialmente ao longo do tempo. Nesta definição, o tempo pode ser substituído por outra variável como pH, temperatura, espaço, etc. Uma característica marcante das séries temporais é que as medidas adjacentes são correlacionadas uma com as outras o que definimos como autocorrelação. As técnicas convencionais de análise estatística assumem que as medidas são independentes umas das outras, contrário ao que ocorre nas séries temporais. Para a análise de séries temporais é necessário utilizar técnicas de análise estatística específicas.

A correlação entre pares de observações adjacentes pode ser calculada pelo coeficiente de autocorrelação amostral  $h_w$ . Para as observações  $x_1, x_2, \dots, x_l$  coletadas ao longo do tempo  $l$  de uma série temporal discreta, calculamos a autocorrelação amostral defasada de  $w$  períodos de tempo conforme a equação abaixo:

$$h_w = \frac{\sum_{i=1}^{l-w} (x_i - \bar{x})(x_{i-w} - \bar{x})}{\sum_{i=1}^l (x_i - \bar{x})^2}, \quad w = 0, 1, \dots, W \quad 7.1$$

onde  $\bar{x}$  é valor médio das observações da série temporal. Os coeficientes de autocorrelação amostral são adimensionais com valores  $-1 < h_w < 1$ . O coeficiente  $h_1$

mede a correlação entre os pares de observações  $x_i$  e  $x_{i-1}$  enquanto que o coeficiente  $h_2$  mede a correlação entre  $x_i$  e  $x_{i-2}$ .

O conjunto de coeficientes de autocorrelação amostral  $h_1, h_2, \dots, h_W$  como função da defasagem  $w$  formam a função de autocorrelação amostral representada por “fca”. O gráfico da “fca” é denominado de correlograma e é uma importante ferramenta usada para a identificação dos modelos de séries temporais (tópico 7.2.1).

Um modelo de série temporal é uma equação matemática que descreve o comportamento das medidas feitas para uma variável ao longo do tempo. Os Modelos de séries temporais são empíricos porque estes são sempre deduzidos de um conjunto de observações experimentais. Há muitas aplicações possíveis para um modelo de série temporal, sendo uma das mais importantes a utilização destes modelos para fazermos previsões futuras das medidas de uma propriedade do sistema sob investigação (“forecasting”). Neste trabalho, além das previsões, interessa-nos a utilização dos modelos de séries temporais para o controle estatístico de processos.

Um modelo de série temporal muito comum é o modelo autoregressivo de ordem  $p$ , representado pela sigla  $AR(p)$ . Suponha que as observações  $x_1, x_2, \dots, x_i$  são oriundas de um processo autoregressivo de ordem  $p$ , então o modelo que descreve o comportamento destas observações é representado pela equação abaixo:

$$x_i = \xi + \phi_1 x_{i-1} + \phi_2 x_{i-2} + \dots + \phi_p x_{i-p} + e_i \quad 7.2$$

onde  $\xi = (1 - \sum_{n=1}^p \phi_n) \mu$  é o intercepto do modelo,  $\phi_p$  são os coeficientes angulares do modelo linear,  $\mu$  é a média das observações,  $x_i$  é a observação feita no tempo  $i$  e  $x_{i-p}$  é a observação feita no instante  $i - p$ . O termo  $e_i$  é um erro aleatório de média  $\mu_0$  e variância  $\sigma^2$ . A equação 7.2 define uma família de modelos, onde os de primeira (equação 7.3) e segunda ordem (equação 7.4) são particularmente muito importantes porque são frequentemente encontrados na prática.

$$x_i = \xi + \phi_1 x_{i-1} + e_i \quad 7.3$$

$$x_i = \xi + \phi_1 x_{i-1} + \phi_2 x_{i-2} + e_i \quad 7.4$$

Outro modelo freqüentemente encontrado na prática é o das médias móveis conforme mostra a equação 7.5:

$$x_i = \mu + e_i - \theta_1 e_{i-1} - \theta_2 e_{i-2} - \dots - \theta_q e_{i-q} \quad 7.5$$

Onde  $\theta_q$  são constantes do modelo linear,  $e_i$  é uma variável aleatória com distribuição normal de média  $\mu_0$  e variância  $\sigma^2$ . O modelo 7.5 é denominado de Médias Móveis de Ordem  $q$ , MA ( $q$ ). Tal como para os modelos autoregressivos, os modelos de média móvel de primeira (equação 7.6) e segunda ordem (equação 7.7) são muito importantes para as aplicações práticas:

$$x_i = \mu + e_i - \theta_1 e_{i-1} \quad 7.6$$

$$x_i = \mu + e_i - \theta_1 e_{i-1} - \theta_2 e_{i-2} \quad 7.7$$

Ambos modelos AR( $p$ ) e MA( $q$ ), podem ser combinados em um modelo misto ARMA( $p,q$ ), conforme mostra a equação 7.8.

$$x_i = \xi + \phi_1 x_{i-1} + \dots + \phi_p x_{i-p} + e_i - \theta_1 e_{i-1} + \dots + \theta_q e_{i-q} \quad 7.8$$

Um modelo ARMA, descreve uma série temporal estacionária. Um processo estacionário é aquele em as propriedades estatísticas das observações  $x_1, x_2, \dots, x_i$ , tal como a média e a variância, são constantes para diferentes períodos de tempo.

Quando a série temporal não é estacionária, ela pode ser diferenciada para obtermos uma nova série temporal que é estacionária. Esta operação é realizada de acordo com a equação abaixo:

$$\nabla x_i = x_i - x_{i-1} \quad 7.9$$

Se a série temporal obtida da primeira diferença não for estacionária, então o resultado desta pode ser usado novamente para calcular a segunda diferença da série, de acordo com a equação 7.9. Geralmente são necessárias no máximo duas diferenças para obter séries temporais estacionárias. Um processo autoregressivo de médias móveis integrado, é representado pela sigla  $ARIMA(p,d,q)$ , onde  $d$  corresponde ao número de diferenças calculadas para tornar a série temporal original em uma nova série temporal estacionária.

### 7.2.1 Identificação de modelos $ARIMA(p,d,q)$

A identificação do modelo que descreve uma série temporal é um processo empírico executado a partir da análise e interpretação de parâmetros estatísticos calculados a partir do conjunto de observações que formam uma série temporal.

A primeira etapa da identificação consiste em verificar se a série temporal é um processo estacionário. Uma boa forma de fazer isto é inspecionar o gráfico das observações  $x_i$  em função do tempo  $t$  e verificar se existem tendências nas medidas que deslocam a média das observações em algum sentido. Caso a série temporal apresente tendências ao longo do tempo, ela é não estacionária e deve ser transformada em uma série estacionária calculando tantas diferenças quanto forem necessárias com a equação 7.9.

Após obter uma série temporal estacionária de índice  $d$ , pode-se calcular a função de autocorrelação amostral “ $\hat{r}_w$ ” definida no tópico 7.2. O comportamento dos coeficientes de autocorrelação amostral em função da defasagem  $w$  dá indicações do modelo de série temporal que seguem as observações  $x_1, x_2, \dots, x_t$ . Por exemplo, se os coeficientes  $h_1, h_2, \dots, h_w$  decrescem em função de  $w$  de acordo com uma curva de decaimento exponencial, significa que as observações  $x_1, x_2, \dots, x_t$  seguem o comportamento de um processo autoregressivo de ordem  $p = 1$ , como indica a tabela 7.1. Para o caso em que  $d = 0$ , temos que o modelo é  $ARIMA(1,0,0)$  e este corresponde a um modelo  $AR(1)$ . Por outro lado se  $d = 1$ , temos um modelo  $ARIMA(1,1,0)$ .

Se somente o coeficiente  $h_1$  é significativo e diferente de zero, então provavelmente a série temporal em questão é um processo de médias móveis de ordem  $q$

= 1 (tabela 7.1). Neste caso se  $d = 0$ , o modelo é ARIMA(0,0,1) que corresponde ao modelo MA(1) e se  $d = 1$  o modelo é ARIMA(0,1,1).

Além da “fac”, costuma-se utilizar também a função de autocorrelação amostral parcial “facp” para identificar os modelos de séries temporais.<sup>223</sup> A “facp” corresponde ao conjunto de coeficientes de autocorrelação amostrais parciais  $g_w$  em função da defasagem  $w$ . Da mesma forma que para a “fac”, o comportamento dos coeficientes  $g_1, g_2, \dots, g_w$  em função de  $w$  fornece informações do tipo de série temporal que segue um conjunto de observações  $x_1, x_2, \dots, x_I$ .

Ambas funções “fac” e “facp” dão informações complementares para identificação do modelo de série temporal. A tabela 7.1, resume o comportamento esperado para os coeficientes da “fac” e “facp”,<sup>223</sup> de modo que a informação dessa tabela pode ser usada como um guia para a identificação de um modelo de série temporal apropriado para a descrição de um determinado conjunto de observações  $x_1, x_2, \dots, x_I$ .

**Tabela 7.1.** Comportamento esperado para os coeficientes das funções “fac” e “facp” de modelos ARIMA( $p,d,q$ ).<sup>223</sup>

Função	ARIMA (1,d,0)	ARIMA (0,d,1)	ARIMA (2,d,0)	ARIMA (0,d,2)	ARIMA (1,d,1)
“fac”	Decaimento exponencial de $h_1, h_2, \dots, h_w$ em função de $w$ .	Somente $h_1$ diferente de zero.	decaimento oscilatório de $h_1, h_2, \dots, h_w$ em função de $w$ .	Somente $h_1$ e $h_2$ diferente de zero.	Decaimento exponencial de $h_1, h_2, \dots, h_w$ em função de $w$ , a partir do primeiro período $w = 1$ .
“facp”	Somente $g_1$ diferente de zero	Decaimento exponencial de $g_1, g_2, \dots, g_w$ em função de $w$ .	Somente $g_1$ e $g_2$ diferente de zero	Decaimento oscilatório de $g_1, g_2, \dots, g_w$ em função de $w$ .	Decaimento exponencial de $g_1, g_2, \dots, g_w$ em função de $w$ , a partir do primeiro período $w = 1$ .

### 7.3 CONTROLE ESTATÍSTICO MULTIVARIADO DE PROCESSOS AUTOCORRELACIONADOS

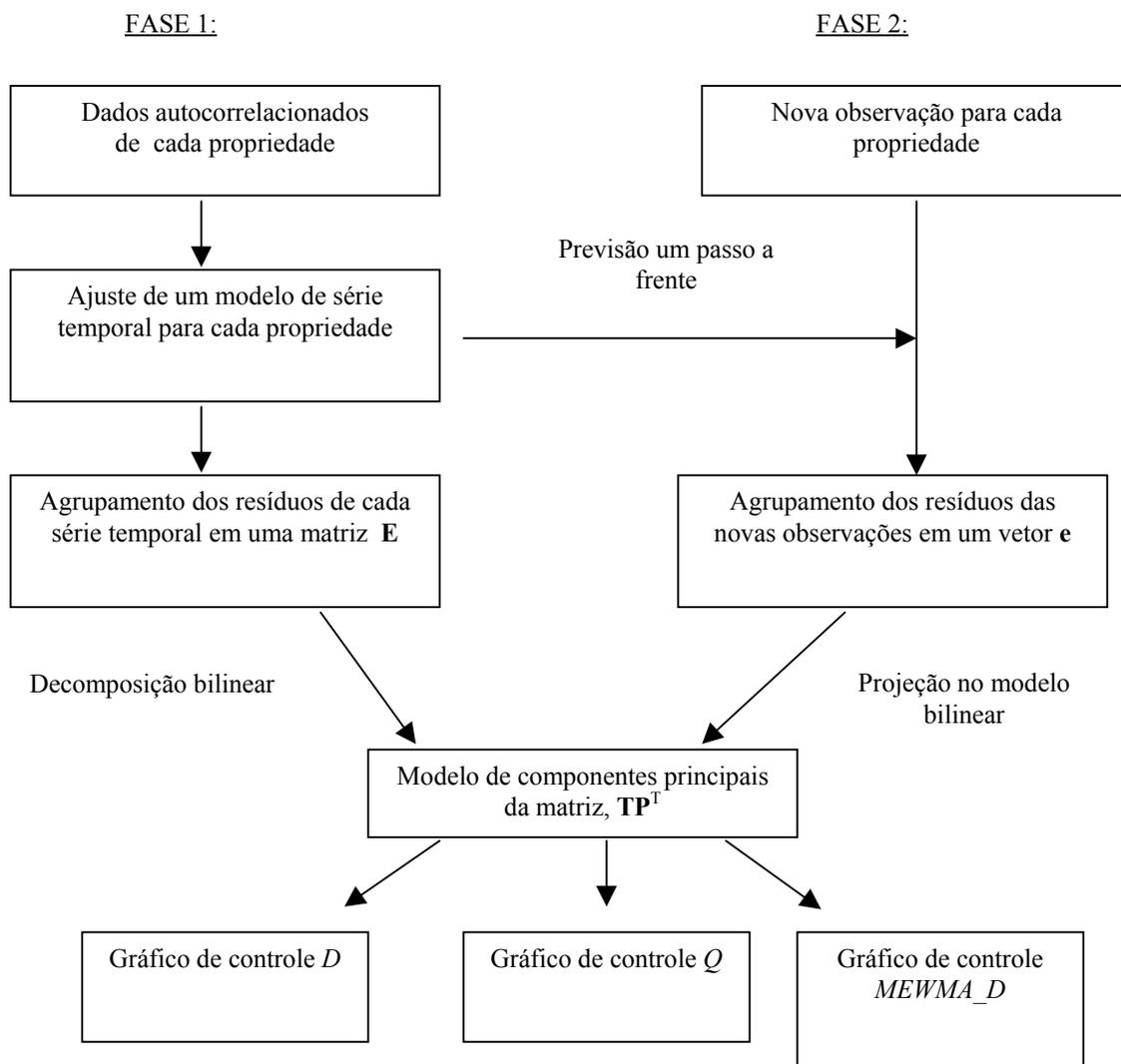
A técnica apresentada neste trabalho de tese para o controle estatístico de processos autocorrelacionados segue a mesma estratégia usada para a construção de gráficos de controle univariados a partir dos resíduos de um modelo de série temporal.<sup>226</sup> A idéia é assumir que o processo segue o comportamento de um modelo ARIMA ( $p,d,q$ ). Com isto, assumimos que as observações são autocorrelacionadas e o processo é visto de uma forma diferente das técnicas convencionais de SPC, que declaram que as observações variam aleatoriamente em torno de um valor médio e são independentes. Na versão das séries temporais cada observação mais atual apresenta uma memória que está associada ao estado anterior do processo, de modo que as observações não apresentam mais um comportamento aleatório mas sim, um comportamento que segue uma tendência que está relacionada ao que aconteceu no processo em instantes anteriores.

Para um determinado modelo ARIMA( $p,d,q$ ) ajustado de um conjunto de observações  $x_1, x_2, \dots, x_i$ , os resíduos  $e_i$  são independentes e com distribuição normal. Na presença de uma causa especial, as características de estado deste processo (média ou variância da série temporal) podem ser afetadas de modo que os resíduos do modelo apresentarão também uma mudança na sua média ou variância. Por isso, gráficos de controle convencionais podem ser construídos a partir dos resíduos dos modelos de séries temporais para detectar se o processo está fora ou em controle estatístico.

Em um processo químico podem ser monitoradas muitas variáveis que são correlacionadas entre si (correlação cruzada) e formadas por medidas autocorrelacionadas ao longo do tempo (autocorrelacionadas). Se estas variáveis pertencem a um mesmo sistema químico, os resíduos dos modelos das séries temporais ajustados para cada variável, apresentam uma correlação cruzada também, porque são partes de um mesmo processo físico. Neste caso, os resíduos dos modelos devem ser analisados com os gráficos de controle multivariado.

A figura 7.2, mostra um esquema da metodologia desenvolvida neste trabalho para o controle estatístico multivariado de dados autocorrelacionados. Durante a fase 1 um modelo ARIMA( $p,d,q$ ) é ajustado para cada variável e os resíduos destes modelos são

agrupados em uma matriz  $\mathbf{E}$  que será autoescalada para a análise em componentes principais. Os parâmetros do modelo bilinear ajustados por PCA são usados para a construção dos gráficos de controle  $D$ ,  $Q$ , e  $MEWMA-D$ . Se o processo estiver em controle estatístico, então os gráficos podem ser utilizados para o monitoramento (Fase 2).



**Figura 7.2.** Esquema da metodologia para o monitoramento estatístico de processos com observações autocorrelacionadas.

Na fase 2, o modelo da série temporal de cada variável é usada para fazer uma previsão um passo a frente e a diferença entre a previsão e a medida mais atual é calculada para dar os resíduos do instante mais atual. Estes resíduos são agrupados em um vetor  $\mathbf{e}$  que é autoescalado com a média e o desvio padrão das observações coletadas

na fase 1. O vetor de resíduos das novas observações é projetado no modelo PCA e as respectivas estatísticas  $D$ ,  $Q$ , e  $MEWMA-D$  associadas a ele são calculadas para verificar se o processo continua em controle ou fora de controle.

o procedimento matemático apresentado na figura 7.2 é detalhado nas etapas abaixo:

Fase1:

1. Ajuste um modelo  $ARIMA(p,d,q)$  para cada propriedade monitorada e calcule os resíduos de cada modelo.
2. Agrupe os resíduos de cada modelo de série temporal em uma matriz  $\mathbf{E}$  e autoescale esta matriz.
3. Calcule a média móvel ponderada exponencialmente  $\mathbf{Z}$  para cada propriedade da matriz  $\mathbf{E}$  (equação 6.32).
4. Decomponha a matriz autoescalada  $\mathbf{E}$  pelo método da análise em componentes principais para obter o modelo abaixo:

$$\mathbf{E} = \mathbf{TP}^T + \mathbf{F} \quad 7.10$$

5. Calcule a estatística  $D$  ( equação 6.40).
6. Calcule os limites de controle do gráfico  $D$  (equação 6.42).
7. Calcule a estatística  $Q$  ( equação 6.45).
8. Calcule os limites de controle do gráfico  $Q$  (equação 6.46).
9. Calcule a estatística  $MEWMA-D$  conforme a equação abaixo:

$$MEWMA - D = \mathbf{z}_i^T \left[ \mathbf{P} \left( \frac{L-2}{L} \boldsymbol{\lambda}_R^{-1} \right) \mathbf{P}^T \right] \mathbf{z}_i \quad i = 1, 2, \dots, I \quad 7.11$$

10. Obtenha os limites do gráfico  $MEWMA$  em tabelas para um  $NMA$  específico ou a partir do programa de Rigdom.<sup>227</sup>

Fase 2:

11. Calcule os novos limites para o gráfico  $D$  (equação 6.42).
12. Faça uma previsão um passo a frente com o modelo de série temporal e calcule o resíduo do instante  $i+1$  pela diferença da medida mais atual e o valor previsto.
13. Agrupe os resíduos de cada previsão em um vetor  $\mathbf{e}$ , faça o escalamento e calcule a média móvel ponderada exponencialmente  $\mathbf{z}$  com o mesmo fator de ponderação usado na fase 1.
14. Faça a projeção das novas observações no modelo bilinear e calcule as estatísticas  $D$ ,  $Q$  e MEWMA- $D$  conforme as equações abaixo:

$$D = \mathbf{e}_{novo}^T \left( \frac{\mathbf{P} \boldsymbol{\lambda}_R^{-1} \mathbf{P}^T}{I - 1} \right) \mathbf{e}_{novo} \quad 7.12$$

$$MEWMA - D = \mathbf{z}_{novo}^T \left[ \mathbf{P} \left( \frac{L - 2}{L} \boldsymbol{\lambda}_R^{-1} \right) \mathbf{P}^T \right] \mathbf{z}_{novo} \quad 7.13$$

$$\mathbf{f}_{novo}^T = \mathbf{e}_{novo}^T - \mathbf{e}_{novo}^T \mathbf{P} \mathbf{P}^T \quad 7.14$$

$$Q_{novo} = \mathbf{f}_{novo}^T \mathbf{f}_{novo} \quad 7.15$$

15. Introduza as estatísticas  $D$ ,  $Q$  e MEWMA- $D$  nos seus respectivos gráficos e verifique o estado do processo.

Uma etapa muito importante da metodologia acima é a previsão com os modelos de séries temporais. Para um modelo AR(1), equação 7.3, a previsão um passo a frente para o instante  $i+1$  é feita da seguinte forma:

$$\hat{x}_{i+1} = \xi + \phi x_i \quad 7.16$$

e o resíduo  $e_{i+1}$  é calculado a partir da observação  $x_{i+1}$  coletada do processo conforme abaixo:

$$\mathbf{e}_{i+1} = \mathbf{x}_{i+1} - \hat{\mathbf{x}}_{i+1} \quad 7.17$$

As previsões com outros modelos ARIMA seguem o mesmo procedimento demonstrado acima. O livro de Box e colaboradores é uma boa referência para uma consulta dos métodos matemáticos empregados para a previsão com modelos de séries temporais.<sup>223</sup>

Se uma observação é assinalada fora de controle pelo gráfico, a etapa a seguir consiste em descobrir qual foi a variável ou conjunto de variáveis responsáveis por tal evento. Uma maneira de descobrir isto, é calcular a contribuição de cada variável em um dado instante para a estatística  $D$  conforme mostra a equação abaixo:

$$t_{ir} = x_{i1}p_{1r} + x_{i2}p_{2r} + x_{i3}p_{3r} + \dots + x_{ij}p_{jr} \quad r = 1, 2, \dots, R \quad 7.18$$

o produto  $x_{jr}p_{jr}$ , mede a contribuição de cada variável original  $x_j$  para o cálculo dos “scores”  $t_r$  do modelo bilinear, visto que os loadings  $p_r$  são parâmetros fixos. Esta contribuição portanto está relacionada à estatística  $D$ , visto que:

$$D = \sum_{r=1}^R \frac{\mathbf{t}_r^T \mathbf{t}_r}{\lambda_{rr}^2} \quad 7.19$$

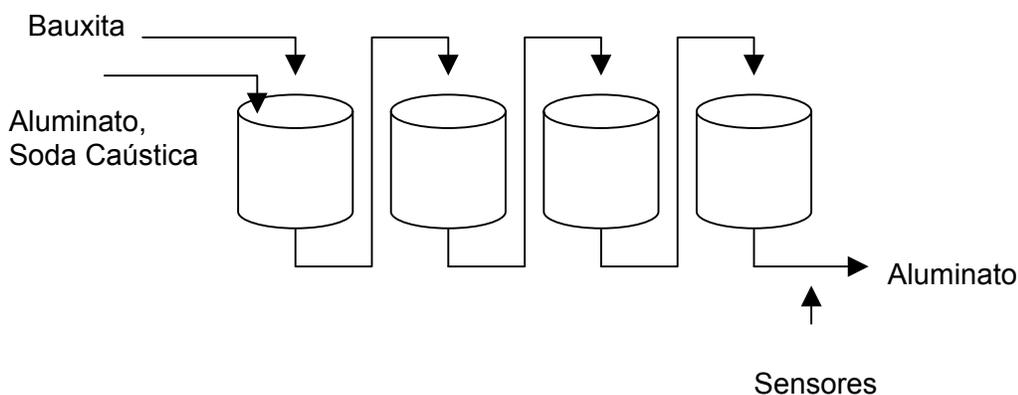
corresponde à equação 7.12. Outros métodos para o cálculo de contribuições existem e podem ser verificados nos trabalhos de Westerhuis<sup>228</sup> e Miller.<sup>229</sup>

## 7.4 MONITORAMENTO MULTIVARIADO DO PROCESSO BAYER PARA A EXTRAÇÃO DE ALUMÍNIO DA BAUXITA

No Processo Bayer há uma etapa denominada de digestão representada na figura 7.3, em que a bauxita e a soda cáustica são introduzidas em um conjunto de quatro tanques de digestão, para fazer a dissolução do óxido de alumínio tri-hidratado presente na bauxita, produzindo o aluminato de sódio. Neste processo é necessário manter a solução de aluminato de sódio na concentração máxima limite para suportar as condições

do processo de clarificação em uma etapa seguinte, sem causar a autoprecipitação do óxido de alumínio tri-hidratado.

A cada duas horas são feitas análises titulométricas no laboratório da empresa para determinar as concentrações de soda cáustica e alumina na solução saturada de aluminato de sódio. No laboratório também foram realizadas medidas do índice de refração. Da mesma forma, são realizadas as medidas em linha (“on line”) da densidade, condutividade e temperatura da solução de aluminato supersaturada no instante em que a amostra é retirada do processo e enviada para o laboratório.



**Figura 7.3.** Etapa de Digestão do Processo Bayer.

O controle automático da planta é feito através da manipulação da carga de soda cáustica adicionada no processo. A quantidade de soda a ser adicionada é calculada a partir da concentração de alumina presente na solução supersaturada de aluminato na saída do processo.

Para o monitoramento estatístico desta planta foram utilizadas as medidas de densidade, condutividade, temperatura e índice de refração, denominadas de variáveis de processo. As concentrações de alumina e soda cáustica são denominadas de variáveis de qualidade. As variáveis de processo foram escolhidas para o monitoramento estatístico porque são medidas que podem ser obtidas em tempo real enquanto que as variáveis de qualidade são informadas a cada duas horas após a amostragem realizada no processo.

## 7.5 MATERIAIS E MÉTODOS

### 7.5.1 Amostragem e conjunto de dados

Cada medida ou análise foi executada em intervalos regulares de duas horas. Ao todo foram coletadas 273 amostras do processo para este estudo, todas compostas por uma observação individual.

Os dados foram divididos em dois grupos, onde as cem primeiras observações foram usadas para a fase 1 do controle estatístico. Posteriormente foram incluídas mais vinte amostras para uma nova avaliação do processo. As outras 153 observações foram usadas para testar a metodologia durante a fase 2, ou seja, o monitoramento do processo.

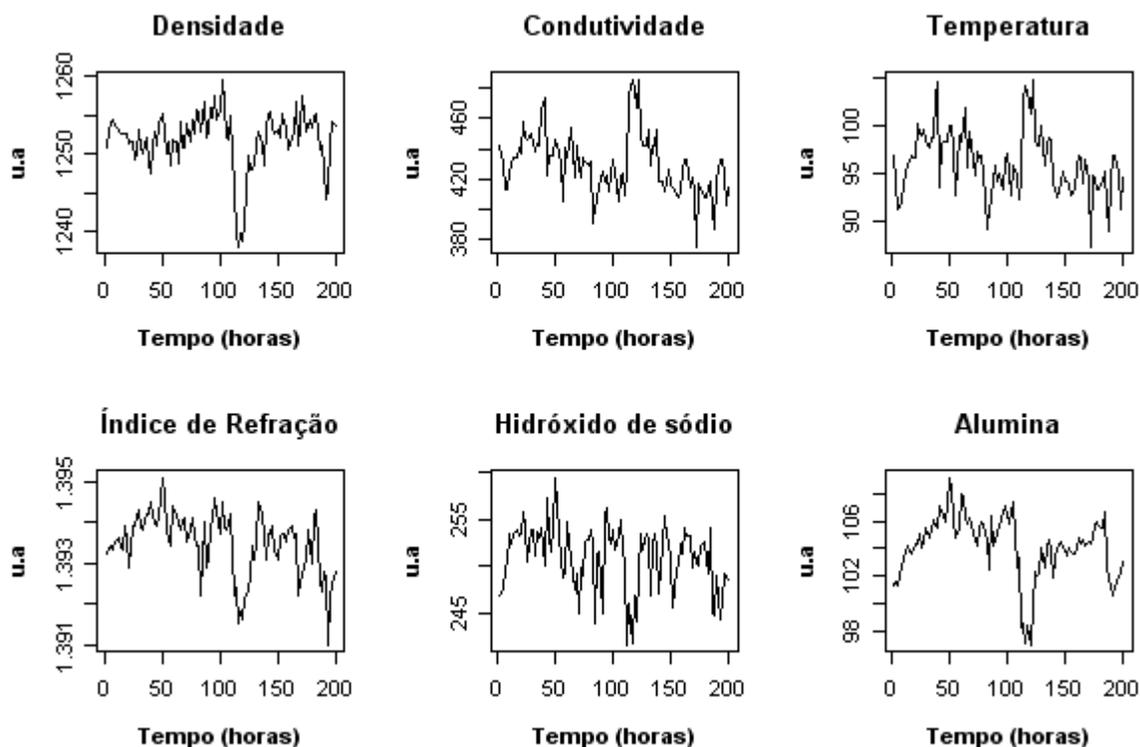
### 7.5.2 Software e Programas

Todos os cálculos deste trabalho foram feitos em um computador COMPAQ, equipado com um microprocessador ATLHON 1 GHz e 128 Mb de memória RAM. Os cálculos foram realizados com o Software R. Os programas usados para os cálculos são de autoria própria com exceção das funções para o ajuste de séries temporais, funções “fac” e “facp”, e as funções de distribuição beta e  $F$ , que são do próprio software R.

## 7.6 RESULTADOS E DISCUSSÃO

A figura 7.4, mostra as medidas coletadas ao longo do tempo para todas as variáveis monitoradas durante a fase 1. Através do exame visual dos gráficos é possível verificar que estas variáveis não apresentam um comportamento aleatório ao longo do tempo. Ao contrário, a observação feita em um dado instante parece seguir a mesma tendência de observações anteriores por um determinado período de tempo. Esta característica é a primeira evidência de que as observações sucessivas para cada variável são autocorrelacionadas.

Levando em conta as características operatórias do processo Bayer, podemos esperar que as variáveis monitoradas neste processo sejam autocorrelacionadas. Isto porque a composição da solução de aluminato de sódio no processo depende da composição da bauxita, a matéria prima que entra nos tanques de digestão e também da quantidade de hidróxido de sódio adicionada nos tanques de digestão. Estas variáveis podem ser alteradas em função dos objetivos de produção e da qualidade da bauxita. Além disso, a composição da bauxita varia de acordo com a procedência deste minério inclusive dentro da mesma jazida.



**Figura 7.4.** Variáveis do processo monitoradas ao longo de duzentas horas.

### 7.6.1 Análise do correlograma das variáveis de processo

A figura 7.5 mostra o correlograma de cada variável de processo (densidade, condutividade, temperatura e índice de refração). As linhas horizontais tracejadas nos correlogramas são os limites que indicam se um coeficiente de correlação amostral é diferente de zero. Em todos os correlogramas da figura 7.5 o coeficiente de correlação amostral  $h_1$  é maior que 0.6, o que significa que as series temporais analisadas são

compostas por observações sucessivas correlacionadas. Da mesma forma, verifica-se que são significativos os coeficientes de correlação amostrais  $h_2$ ,  $h_3$  e  $h_4$ , de modo que a correlação entre as observações sucessivas estende-se para períodos de defasagens maiores que um.

A análise mais importante dos correlogramas é quanto a forma da função de autocorrelação amostral porque este é o principal critério para a identificação dos modelos  $ARIMA(p,d,q)$ . Todos os correlogramas apresentados na figura 7.5 apresentam um comportamento semelhante, onde se observa um decaimento progressivo dos valores dos coeficientes de correlação em função do período de defasagem  $w$  que é similar ao da curva de decaimento exponencial. Os correlogramas das variáveis condutividade e temperatura apresentam um decaimento mais lento, enquanto que os correlogramas das variáveis densidade e índice de refração, apresentam um decaimento brusco para valores próximos a zero.

A figura 7.6 mostra os correlogramas da função de autocorrelação amostral parcial (“facp”) de cada variável do processo. Em todos os casos, o primeiro coeficiente de autocorrelação parcial é significativo e superior a 0,4, enquanto que os demais coeficientes são bem menores e apresentam um comportamento aleatório. Este comportamento sugere que estas variáveis analisadas são processos autoregressivos de primeira ordem. A figura 7.7, mostra o correlograma da função de autocorrelação amostral (“fac”) para a primeira diferença dos dados ( $d = 1$ ). Nestes correlogramas não foram observados coeficientes significativos de modo que a primeira diferença dos dados também remove a correlação entre observações sucessivas.

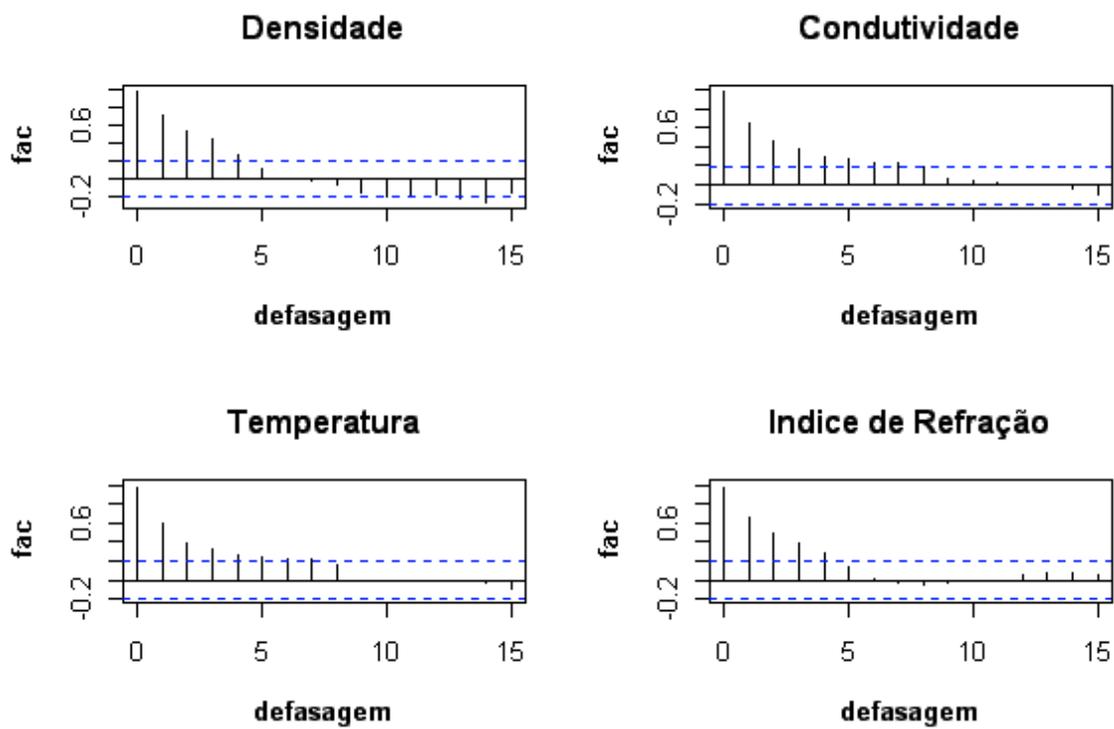


Figura 7.5. Função de autocorrelação amostral para cada variável do processo.

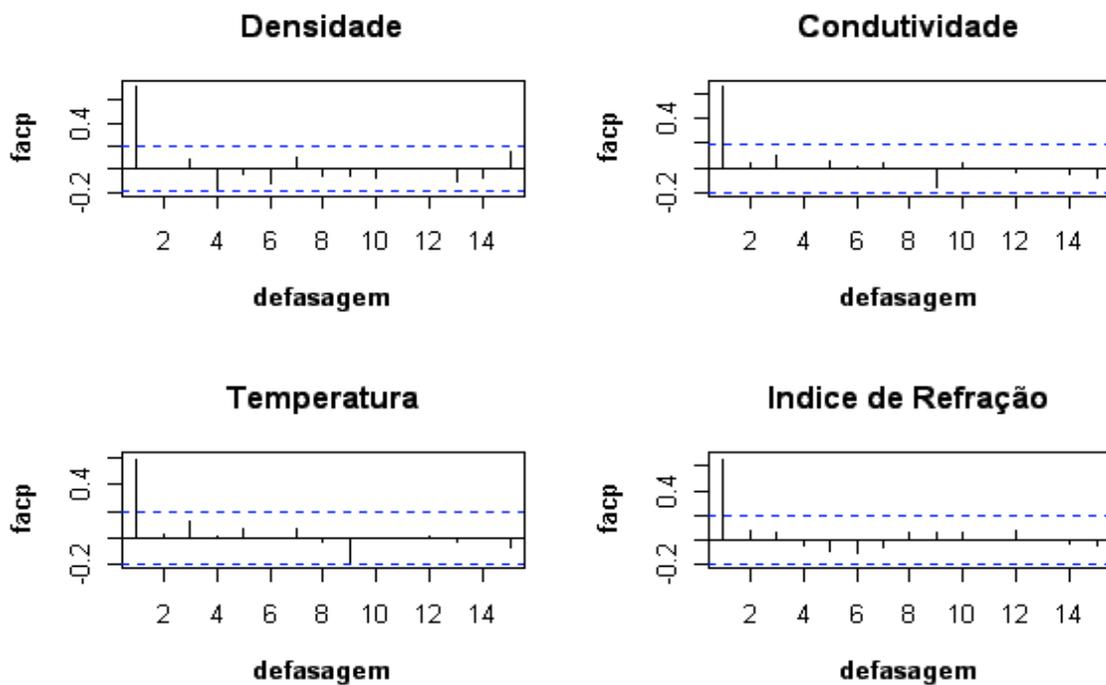
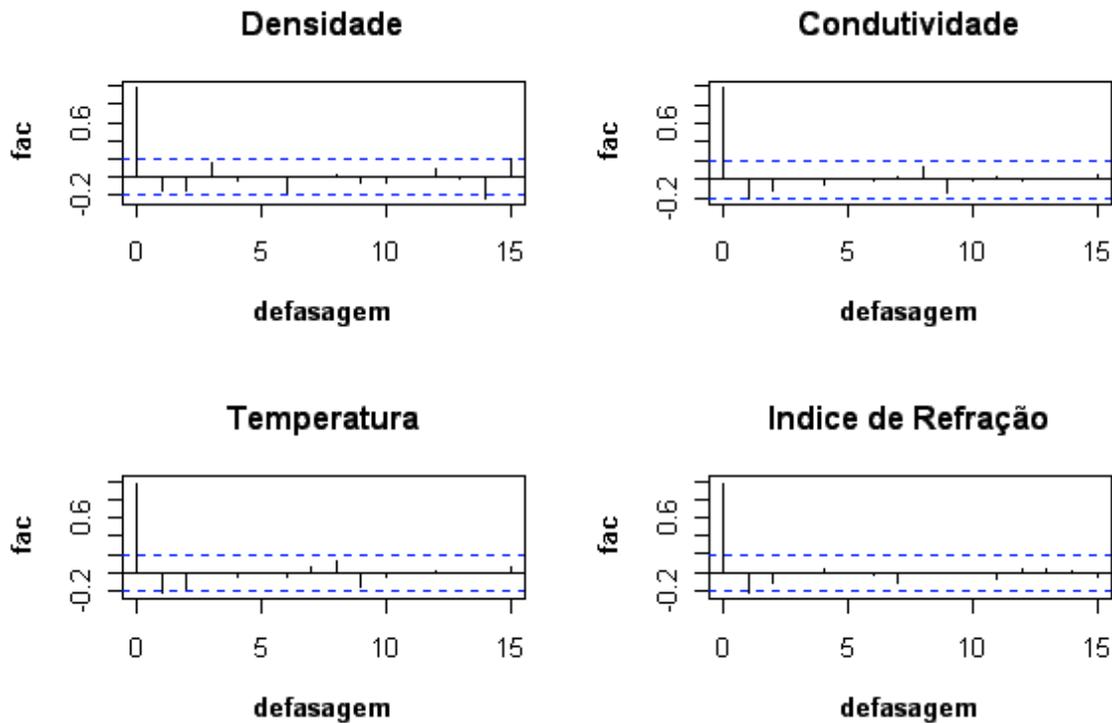


Figura 7.6. Função de autocorrelação amostral parcial de cada variável de processo.



**Figura 7.7.** Função de autocorrelação amostral da primeira diferença tomada para cada variável de processo.

### 7.6.2 Ajuste de modelos de séries temporais das variáveis de processo

Os correlogramas mostrados no tópico anterior indicam que as variáveis analisadas podem ser processos autoregressivos de primeira ordem. Sendo assim, para cada uma das quatro variáveis de processo podem ser ajustados modelos AR(1). A tabela 7.2 mostra os parâmetros dos modelos AR(1) ajustados pelo método dos mínimos quadrados para cada uma destas variáveis, enquanto que a tabela 7.3 mostra os parâmetros de cada modelo AR(1) ajustado pelo princípio da máxima verossimilhança. Ambos métodos de ajuste forneceram resultados equivalentes.

Os modelos AR(1) foram avaliados através da análise dos resíduos que devem ser independentes e com distribuição normal. A figura 7.8 mostra os correlogramas dos resíduos de cada um dos modelos AR(1) da tabela 7.3. Nestes correlogramas não há coeficientes de correlação maiores que zero sugerindo que os resíduos dos modelos são independentes. O teste de Ljung-Box<sup>230</sup> foi usado para confirmar a hipótese nula de que

os resíduos de cada modelo são independentes. A tabela 7.4 mostra os valores P calculados para o teste feito com os resíduos de cada modelo, onde verifica-se que todos são maiores que a probabilidade de 95% e portanto podemos aceitar a hipótese nula.

**Tabela 7.2.** Parâmetros dos modelos AR(1) ajustado pelo método dos mínimos quadrados.

Propriedade	Coefficiente angular	Intercepto	Erro padrão
Densidade	0,7334	333,7179	6,90
Condutividade	0,6744	139,7874	214,50
Temperatura	0,6108	37,4096	6,92
Índice de Refração	0,6801	0,4458	$3,11 \times 10^{-7}$

**Tabela 7.3.** Parâmetros dos modelos AR(1) ajustado pelo método da máxima probabilidade.

Propriedade	Coefficiente angular	Intercepto	Erro padrão
Densidade	0,7266	342,2298	6,833
Condutividade	0,6703	141,5436	213,2
Temperatura	0,6051	37,4096	6,86
Índice de Refração	0,6744	0,4537	$3,09 \times 10^{-7}$

A hipótese de normalidade dos resíduos foi testada pelo teste de Shapiro-Wilk<sup>231</sup> com a hipótese nula de que os resíduos são normais. A tabela 7.4 mostra os resultados deste teste para os quais verifica-se com probabilidade de 95% que os resíduos dos modelos AR(1) das variáveis densidade, condutividade, temperatura são normais, porém a hipótese nula é rejeitada para os resíduos do índice de refração. A figura 7.9 mostra a distribuição dos resíduos de cada modelo AR(1) da tabela 7.2, onde verifica-se que a distribuição das medidas de índice de refração apresentam uma similaridade com a distribuição normal apesar desta hipótese ter sido rejeitada pelo teste de Shapiro-Wilk.

A análise dos correlogramas de cada variável de processo mostra que a primeira diferença dos dados originais também remove a correlação entre observações e que este modelo poderia ser usado para a construção do gráfico de controle. A primeira diferença corresponde a um modelo ARIMA(0,1,0) representado pela equação 7.20:

$$x_i = x_{i-1} + e_i$$

**7.20**

**Tabela 7.4.** Valores P calculados para os testes de Ljung-Box e Shapiro-Wilk na análise dos resíduos do modelos AR(1) ajustados para cada variável de processo.

Modelos AR(1)	Teste de Ljung-Box	Teste de Shapiro-Wilk
	Valor P*	Valor P*
Densidade	0,1804	0,4533
Condutividade	0,8948	0,0882
Temperatura	0,8123	0,1390
Índice de Refração	0,9700	7,513x10 <sup>-5</sup>

\*O teste foi executado assumindo um nível de confiança  $\alpha$  de 0,05 (95%).

O modelo ARIMA(0,1,0) é mais simples que o modelo AR(1) visto que não há parâmetros a serem ajustados pois os resíduos são calculados a partir de primeira diferença conforme mostra a equação 7.9. A tabela 7.5 mostra os resultados dos testes de Ljung-Box e Shapiro-Wilk para análise dos resíduos dos modelos ARIMA(0,1,0). A hipótese nula de que os resíduos são independentes foi rejeitada para os modelos da condutividade, temperatura e índice de refração mas foi aceita para os resíduos do modelo da densidade. Quanto ao teste de Shapiro-Wilk verificou-se que apenas os resíduos do modelo ARIMA(0,1,0) da densidade são normais.

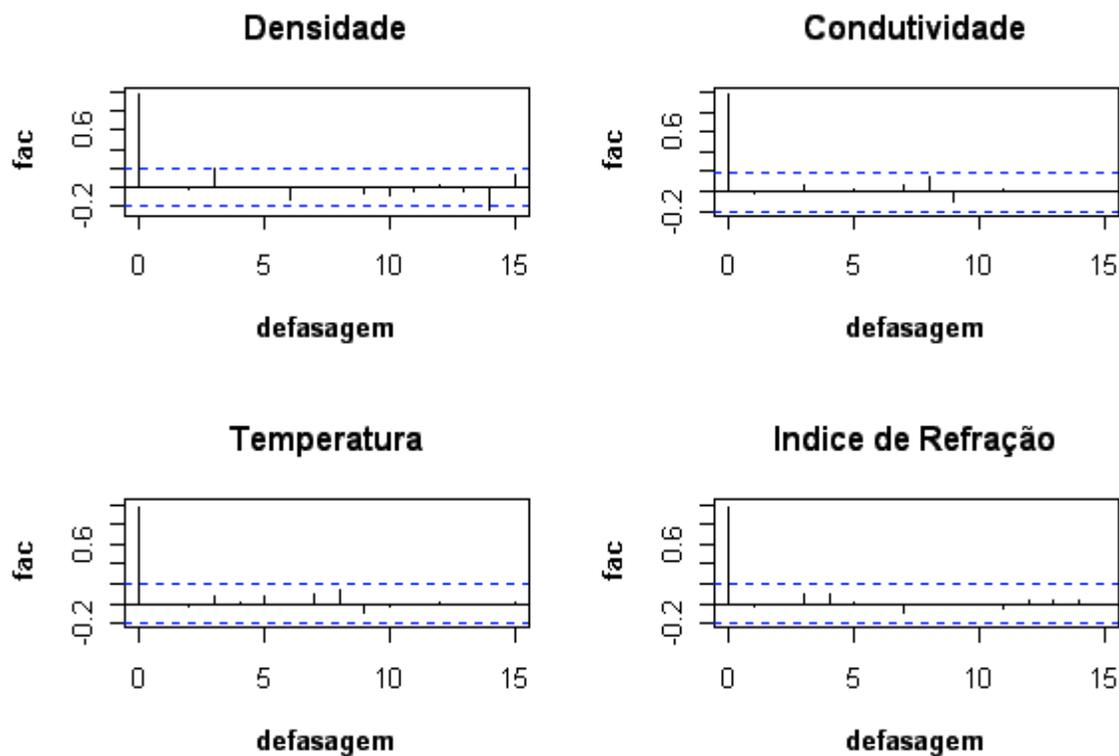
**Tabela 7.5.** Valores P calculados para os testes de Ljung-Box e Shapiro-Wilk na análise dos resíduos do modelos AR(1) ajustados para cada variável de processo.

Modelos AR(1)	Teste de Ljung-Box	Teste de Shapiro-Wilk
	Valor P*	Valor P*
Densidade	0,1103	0,5724
Condutividade	0,0295	0,0108
Temperatura	0,01794	0,0186
Índice de Refração	0,0133	0,00217

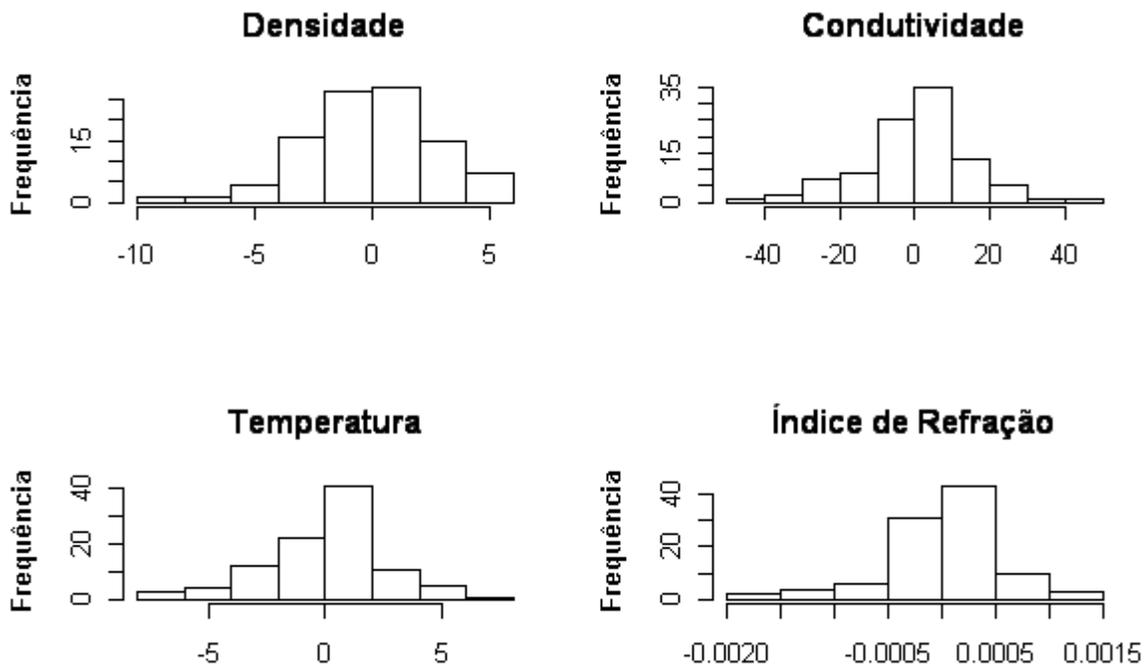
\*O teste foi executado assumindo um nível de confiança  $\alpha$  de 0,05 (95%).

Os resultados obtidos mostram que os modelos AR(1) são mais adequados para a descrição das séries temporais analisadas visto que o modelo ARIMA(0,1,0) não remove totalmente a autocorrelação dos dados originais. A hipótese de normalidade só não foi confirmada para os resíduos do modelo AR(1) do índice de refração. Entretanto como

mostra a figura 7.9, a distribuição dos resíduos do modelo AR(1) do índice de refração, apresentam uma forma muito próxima a da distribuição normal e, por isso, podemos usá-los para a construção dos gráficos de controle que são relativamente robustos à pequenos desvios de normalidade, mas sensíveis a violação da hipótese de independência das medidas.



**Figura 7.8.** Correlograma dos resíduos dos modelos AR(1) de cada variável do processo.



**Figura 7.9.** Distribuição dos resíduos dos modelos AR(1) para cada variável do processo.

### 7.6.3 Análise de correlação entre as variáveis de processo.

Até o momento foram investigadas apenas as correlações entre as observações medidas ao longo do tempo (autocorrelação) e foram ignoradas as correlações entre as variáveis. A tabela 7.6 mostra os coeficientes de correlação linear calculados para cada par de variáveis. O coeficiente de correlação igual a 0,97 entre a temperatura e a condutividade é bastante alto de modo que estas duas variáveis contribuem com a mesma informação do processo (redundância). A densidade é correlacionada com todas as outras variáveis enquanto que o índice de refração é a variável que apresenta a maior correlação com as concentrações de hidróxido e alumina na solução supersaturada de aluminato.

Somente o densidade e o índice de refração são correlacionados com as concentrações de hidróxido e alumina, de modo que estas duas variáveis de processo podem ser usadas para inferir o estado de controle das variáveis da qualidade usadas para o controle automático do processo Bayer.

**Tabela 7.6.** Coeficientes de correlação linear entre variáveis de controle e processo.

	Densidade	Condut.	Temp.	Índ. De Refração	Soda Caústica	Alumina
Densidade	1	-0,714	-0,654	0,564	0,482	0,668
Condut.	0	1	0,976	-0,057	0,023	-0,230
Temp.	0	0	1	0,035	0,068	-0,0587
Índ. Refração	0	0	0	1	0,629	0,787
Soda Caústica	0	0	0	0	1	0,690
Alumínio	0	0	0	0	0	1

A tabela 7.7 mostra os coeficientes de correlação linear calculados para cada par de resíduos dos modelos de séries temporais AR(1). Há uma correlação cruzada entre os resíduos de cada modelo autoregressivo, sendo que esta correlação é similar à respectiva correlação cruzada calculada para cada par de variáveis de processo mostrado na tabela 7.6. Logo, os resíduos dos modelos de séries temporais são correlacionados entre si, da mesma forma que são as suas respectivas variáveis de processo.

**Tabela 7.7.** Coeficientes de correlação entre os resíduos dos modelos AR(1).

	Densidade	Condutividade	Temperatura	Índice de Refração
Densidade.	1	-0,604	-0,582	0,471
Condutividade	0	1	0,973	0,003
Temperatura	0	0	1	0,019
Índice de Refração	0	0	0	1

A análise de correlação mostra que a estrutura do modelo autoregressivo de primeira ordem filtra a autocorrelação entre as observações sucessivas mas não filtra a correlação entre as variáveis e, por isso, é necessário usar um método de controle estatístico multivariado para a análise dos resíduos dos modelos de séries temporais com os gráficos de controle.

## 7.6.4 Controle estatístico multivariado de processos

### 7.6.4.1 Fase 1

Na seqüência das análises, os resíduos de cada um dos quatro modelos de séries temporais AR(1) foram agrupados em uma matriz **E** para a análise em componentes principais. A decomposição da matriz **E** em quatro componentes, mostra que as primeira, segunda, terceira e quarta componentes capturam respectivamente 61,66, 29,95, 7,69 e 0,7 % de variância. Portanto, as duas primeiras componentes principais recuperam 92,27 % da variação dos dados enquanto que o modelo de três componentes recupera 99,30 %.

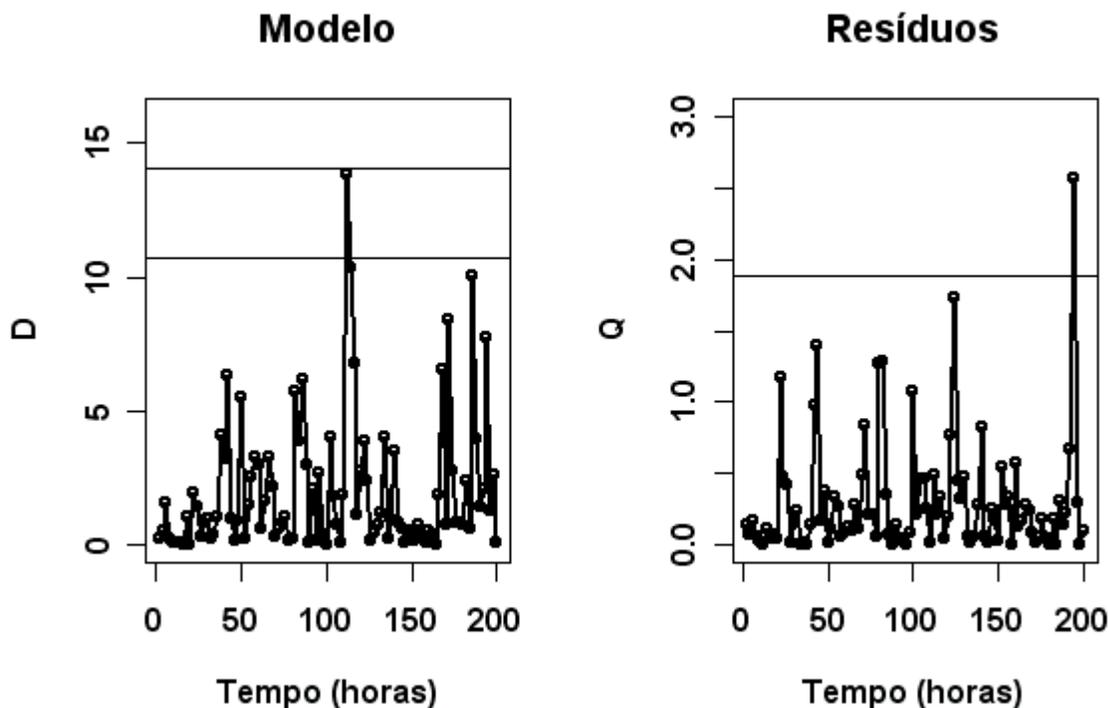
De acordo com a matriz de correlação calculada anteriormente, tabela 7.6, a correlação entre a temperatura e a condutividade é de 0,98 e estas duas variáveis podem ser modeladas por uma componente, o que sugere um modelo com no máximo três variáveis latentes. Entretanto, a temperatura e a condutividade são correlacionadas também com a densidade e esta, por sua vez, é correlacionada com o índice de refração. Por isso, para este conjunto de dados um modelo de duas componentes principais é uma boa escolha para modelar a variabilidade do conjunto de dados e corresponde ao modelo escolhido para a análise estatística do processo.

Considerando que a presente análise corresponde à fase 1 do controle estatístico, gráficos de controle multivariado *D* e *Q* foram construídos a partir do modelo PCA com duas componentes. A figura 7.10 mostra estes gráficos, onde a primeira linha horizontal de baixo para cima é o limite de controle com probabilidade de 95 % e a linha horizontal superior o limite de 99 %. Para o gráfico *Q*, apenas o limite de 99 % é considerado.

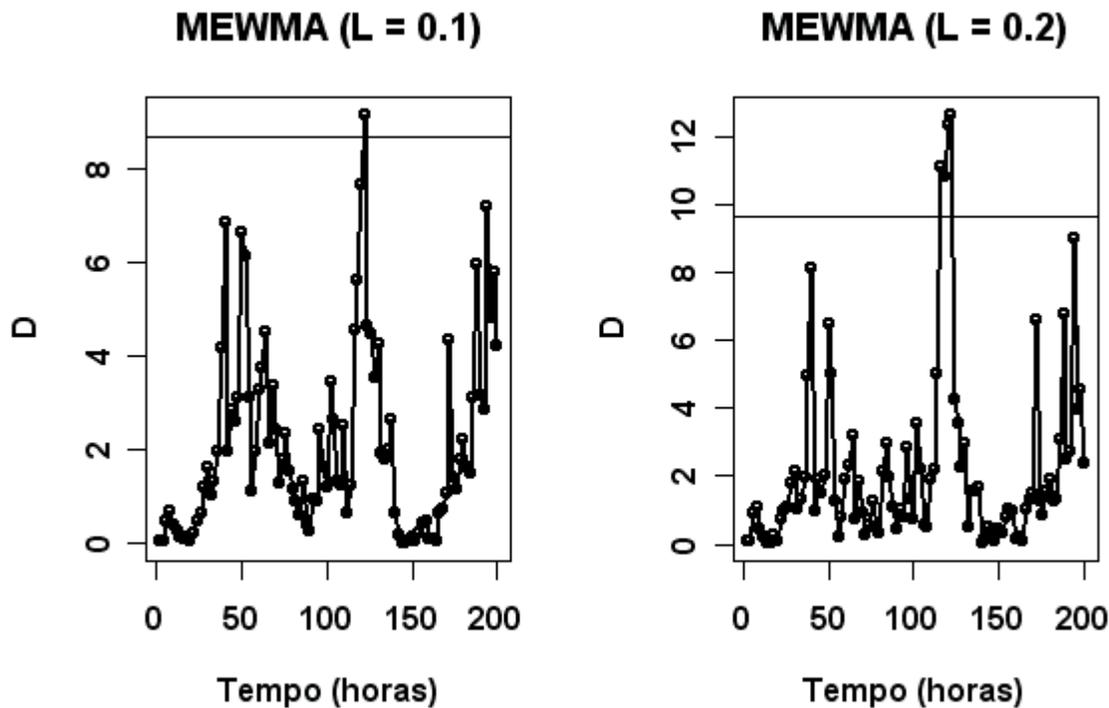
O gráfico *D* assinala a observação coletada em 112 horas fora do limite de controle de 95 % e as observações a seguir em 114 e 116 horas retornam a uma condição normal de operação. O gráfico *Q* assinala uma amostra fora de controle em 197 horas. Cada um dos gráficos assinala uma única observação fora dos limites de controle para um conjunto de cem amostras, de modo que é possível que estas observações assinaladas não sejam devido a causas especiais mas sim devido à probabilidade estatística de uma observação ser assinalada fora dos limites de 99 % para um total de cem amostras, estando o processo fora ou em controle estatístico.

A figura 7.11 mostra dois gráficos MEWMA- $D$ , um calculado com um fator de ponderação  $L = 0,1$ , e o outro gráfico com um fator de ponderação  $L = 0,2$ . O primeiro gráfico MEWMA-PCA assinala a observação feita em 122 horas fora dos limites de controle. Note entretanto que o segundo gráfico assinala quatro pontos fora dos limites de controle entre 116 e 122 horas. Este gráfico torna evidente que o processo encontra-se fora de controle estatístico.

A diferença de detecção dos gráficos MEWMA para diferentes fatores de ponderação é uma propriedade bem conhecida destes gráficos. Fatores de ponderação pequenos, menores que 0,1 favorecem a detecção de deslocamentos pequenos, mas são insensíveis a grandes deslocamentos. Um aumento do fator de ponderação tende a melhorar a detecção dos gráficos MEWMA- $D$  para deslocamentos maiores da média do processo. Para a análise univariada por exemplo, é recomendado em muitos casos monitorar o processo com gráficos de Shewhart e EWMA simultaneamente, um para detectar grandes deslocamentos e outro para detectar pequenos deslocamentos.



**Figura 7.10.** Gráficos de controle  $D$  e  $Q$  do modelo PCA.



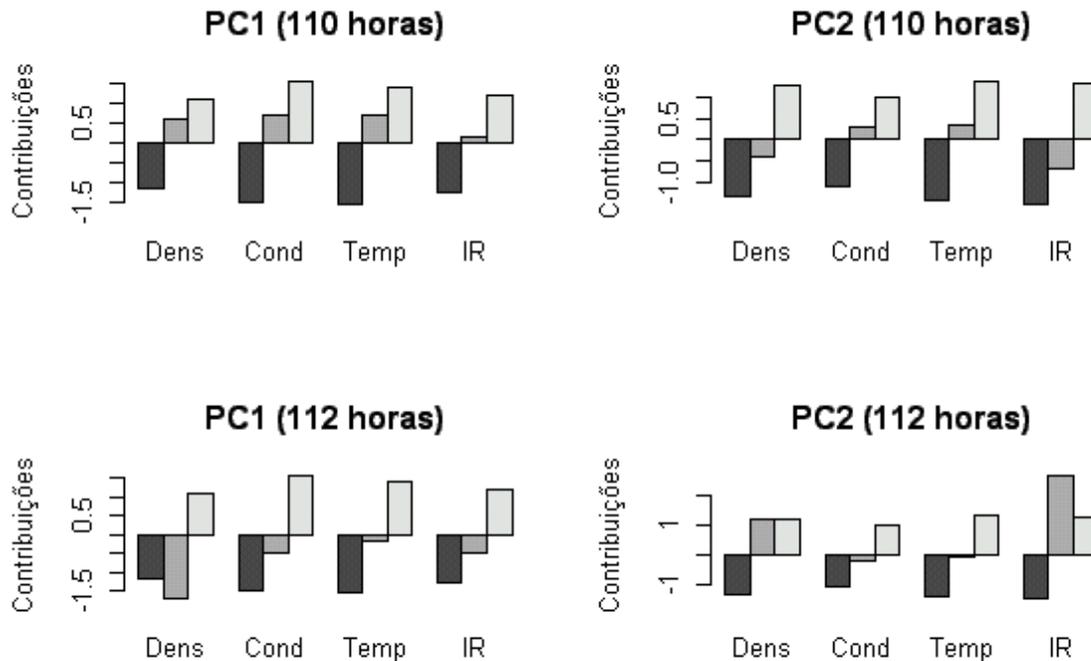
**Figura 7.11.** Gráficos de controle MEWMA- $D$  do modelo PCA.

A presença de quatro observações seguidas fora dos limites de controle estatístico é uma indicação clara de que o processo está fora de controle estatístico. O gráfico MEWMA ( $L= 0.2$ ) assinala um sinal fora de controle a partir da observação feita em 116 horas, mas a carta  $D$  convencional assinala uma observação já em 112 horas.

A figura 7.12 mostra as contribuições das variáveis originais para a estatística  $D$  em 110 horas e 112 horas. A contribuição de cada variável é a barra situada no meio enquanto que a barra à esquerda e à outra a direita são os limites de contribuição calculados como três vezes o desvio padrão das contribuições das cinquenta primeiras observações antes do ponto assinalado como fora de controle (em 112 horas).

Em 110 horas de processo, as contribuições de cada propriedade encontram-se dentro dos limites conforme o esperado visto que esta observação não foi assinalada fora de controle. Porém para a observação em 112 horas assinalada como um ponto fora de controle pela carta  $D$  (figura 7.10), a densidade tem uma contribuição mais elevada em relação às cinquenta observações anteriores, porque ultrapassa os limites de contribuição nas duas componentes principais. Da mesma forma, a contribuição do índice de refração

ultrapassa o seu limite superior na segunda componente principal. Sendo assim, conclui-se que as variáveis de processo densidade e índice de refração são as responsáveis pela observação em 110 horas assinalada fora dos limites de controle.



**Figura 7.12.** Gráfico das contribuições das propriedades para a observação feita em 110 e 112 horas. A barra mais a esquerda é o limite inferior e a mais a direita é o limite superior.

A densidade e o índice de refração são correlacionados com a concentração de alumina na solução saturada de aluminato de sódio. Um ponto assinalado fora de controle devido a estas duas variáveis pode indicar um deslocamento do valor médio da concentração de alumina no processo. De acordo com a figura 7.4 que mostra as observações originais, a concentração de alumínio em 112 horas cai de valores em torno de 106 u.a para 98 u.a mostrando que, de fato, há um deslocamento da média das observações feitas para a concentração de alumina no processo.

A “causa especial” responsável pelo sinal fora de controle ocorre apenas por um determinado tempo, e a planta retorna ao estado normal de operação. Isto pode ser justificado pelo fato de que a concentração de alumina nos tanques de digestão é controlada a partir da quantidade de soda adicionada no processo. Após detectar que o

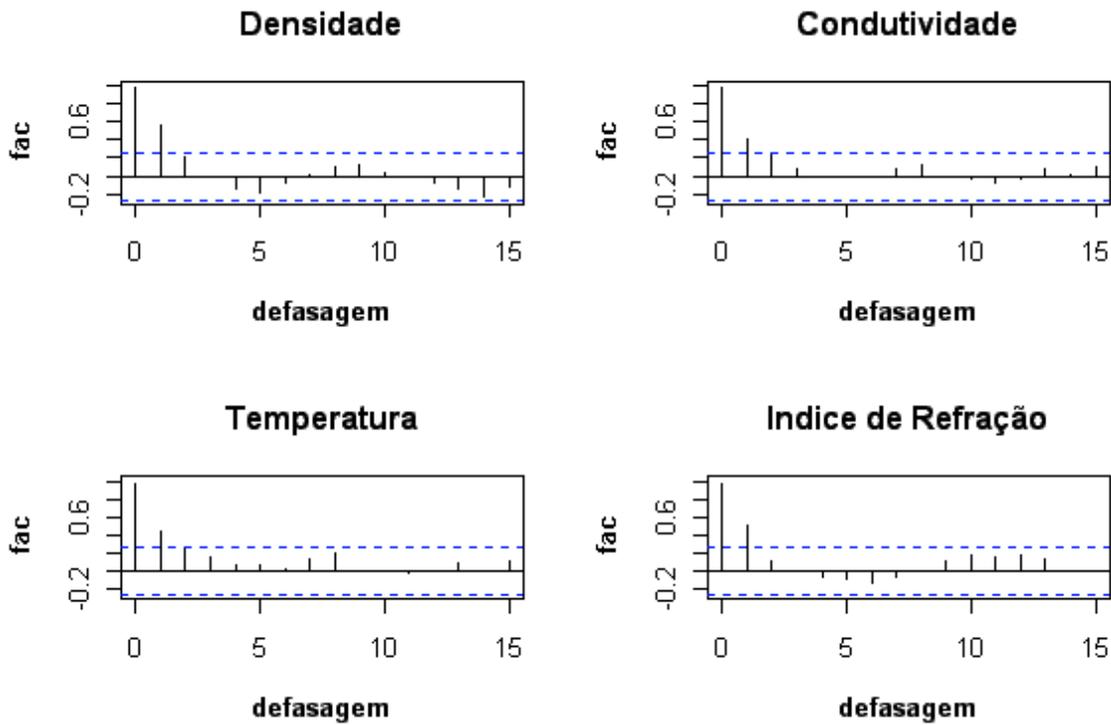
valor de concentração de alumina na solução supersaturada está baixo, pode-se manipular a quantidade de soda adicionada para que a concentração de alumina retorne ao ponto desejado. Entretanto, para fazer a correção no processo é necessário aguardar duas horas até que se obtenha o teor de alumina na solução supersaturada de aluminato de sódio.

### 7.6.5 Nova verificação do estado do processo

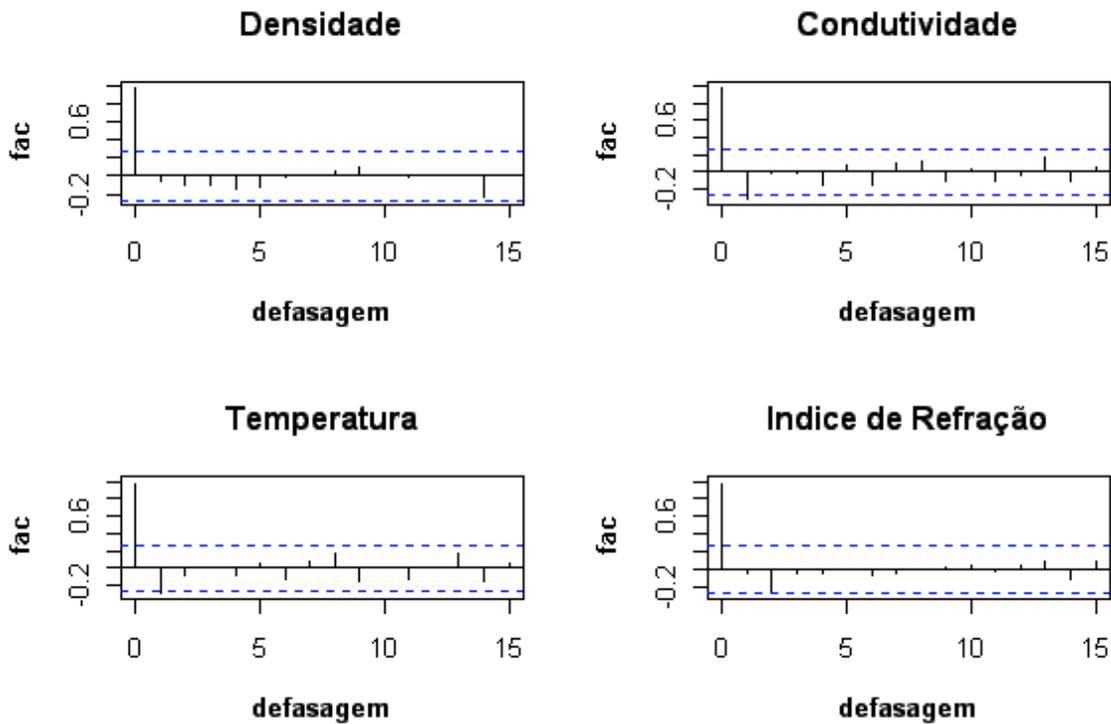
A análise apresentada no tópico anterior mostrou que no período de 112 a 122 horas, o processo apresenta um desvio da sua condição normal de operação. Assumindo que o controle automático realizado nos tanques de digestão foi capaz de recuperar o estado normal de operação após 130 horas, um novo grupo de observações foi coletado entre 130 e 240 horas para uma nova verificação do estado de controle estatístico do processo.

A figura 7.13 mostra os correlogramas “fac” de cada propriedade, calculados para o período de tempo de 130 a 240 horas e a figura 7.14 mostra os correlogramas “fac” calculados para a primeira diferença destes dados.

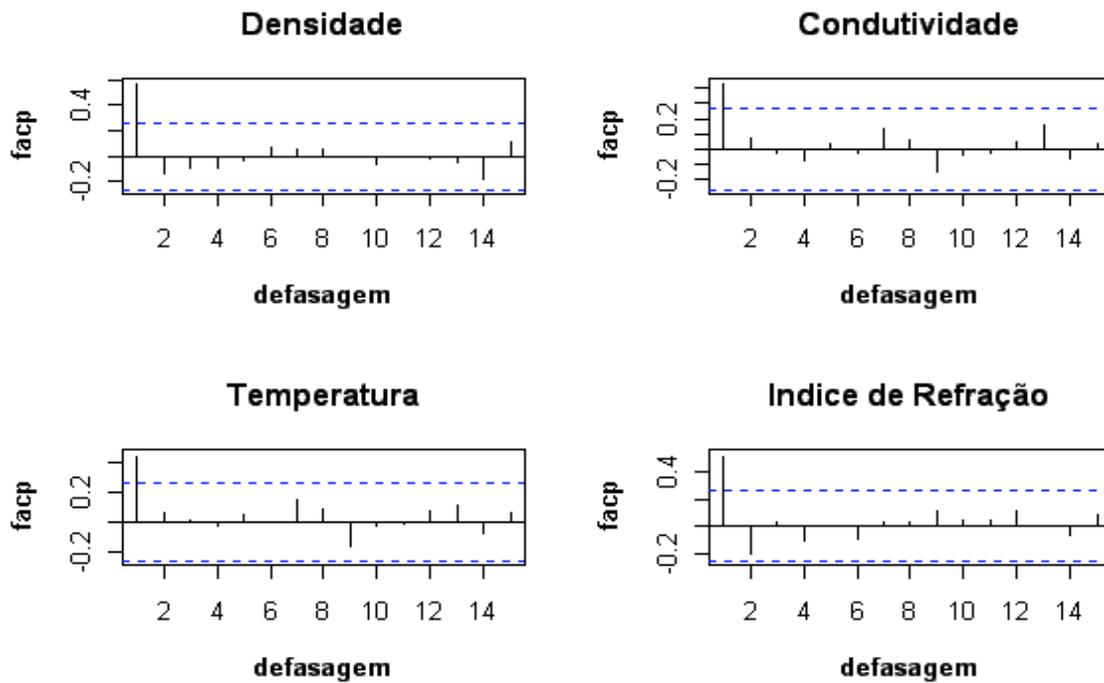
Os correlogramas “fac” apresentados na figura 7.13 mostram um decaimento exponencial brusco para valores próximos a zero. De acordo com este padrão de decaimento dos coeficientes do correlograma, podemos assumir que as variáveis seguem o comportamento de um processo autoregressivo de ordem  $p$ . Já os correlogramas “facp” mostrados na figura 7.15 para cada variável mostram que apenas o primeiro índice de correlação é diferente de zero. Sendo assim podemos esperar que o modelo AR(1) deva produzir uma descrição satisfatória do comportamento dos dados medidos para cada variável no período de 130 a 240 horas. O correlograma da primeira diferença dos dados (figura 7.14), sugere que a primeira diferença dos dados originais também remove a autocorrelação e que o processo pode ser descrito por um modelo ARIMA(0,1,0).



**Figura 7.13.** Funções de autocorrelação amostral para cada propriedade entre 130 e 240 horas.



**Figura 7.14.** Funções de autocorrelação amostral para a primeira diferença dos dados de 130 e 240 horas.

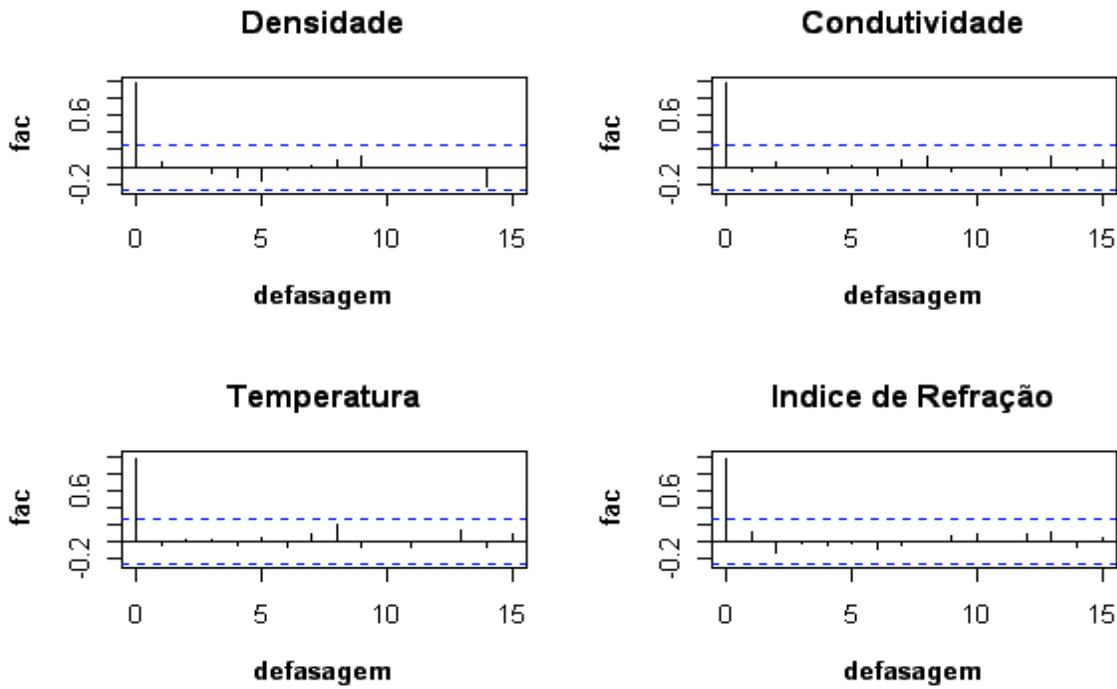


**Figura 7.15.** Funções de autocorrelação amostral parcial de cada propriedade entre 130 e 240 horas.

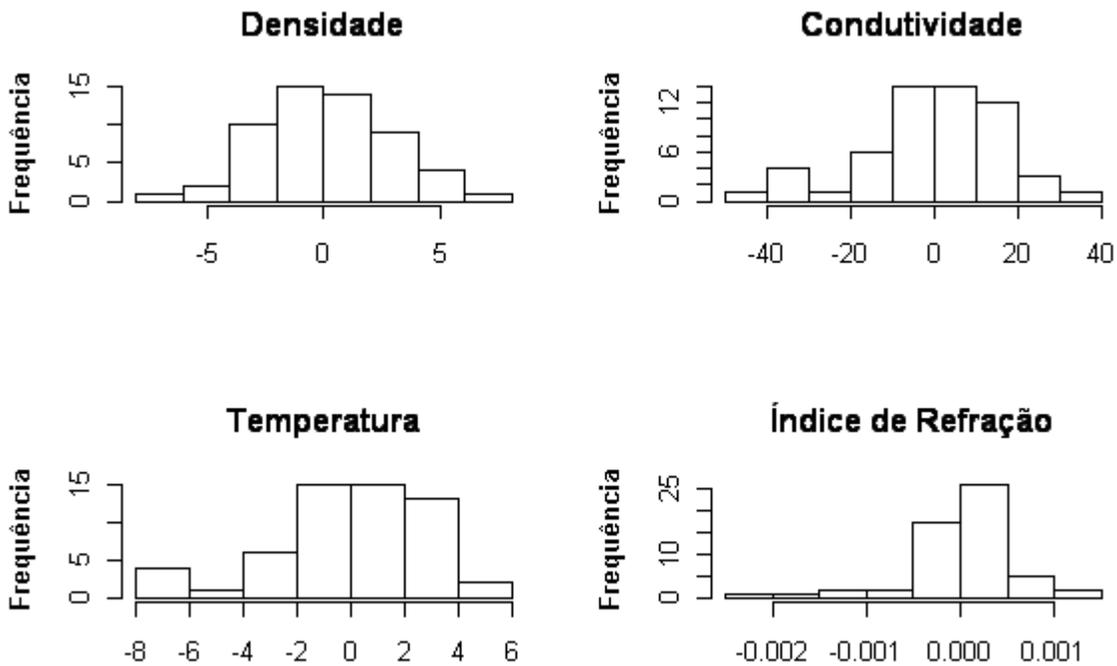
A tabela 7.8 mostra os parâmetros dos modelos AR(1) ajustados para cada uma das variáveis. As figuras 7.16 e 7.17 mostram respectivamente os correlogramas e a distribuição dos resíduos de cada um destes modelos. A hipótese de independência dos resíduos dos modelos AR(1) foi confirmada pelo teste de Ljung-Box para um nível de confiança de 95% conforme mostra os valores P da tabela 7.9, enquanto que a hipótese de normalidade (teste de Shapiro-Wilk) foi aceita somente para os resíduos do modelo AR(1) da densidade.

**Tabela 7.8.** Parâmetros dos modelos AR(1) calculados para cada propriedade.

Propriedade	Coefficiente angular	Intercepto	Erro padrão
Densidade	0,5881	516,2094	7,569
Condutividade	0,4676	220,6677	255,1
Temperatura	0,4909	47,5317	8,128
Índice de Refração	0,5247	0,6622	$3,2162 \times 10^{-7}$



**Figura 7.16.** Funções de autocorrelação amostral dos resíduos dos modelos apresentados na tabela 7.8.



**Figura 7.17.** Distribuição dos resíduos dos modelos AR(1) apresentados na tabela 7.8.

**Tabela 7.9.** Valores P calculados para os testes de Ljung-Box e Shapiro-Wilk na análise dos resíduos do modelos AR(1) ajustados para cada variável de processo.

<b>Modelos AR(1)</b>	<b>Teste de Ljung-Box</b>	<b>Teste de Shapiro-Wilk</b>
	<b>Valor P*</b>	<b>Valor P*</b>
Densidade	0,9336	0,9749
Condutividade	0,8383	0,0464
Temperatura	0,8849	0,0121
Índice de Refração	0,8153	0,0001

A tabela 7.10 mostra os resultados dos testes de Ljung-Box e Shapiro-Wilk para a primeira diferença dos dados originais. A hipótese nula de independência dos resíduos foi aceita para todas as variáveis enquanto que a hipótese nula de normalidade foi rejeitada apenas para o índice de refração. Sendo assim, o modelo ARIMA(0,1,0) é adequado para a descrição do comportamento das séries temporais analisadas.

Apesar de ter sido constatado desvios de normalidade nos resíduos dos modelos AR(1) para as variáveis condutividade, temperatura e índice de refração, estes modelos também foram usados para a construção dos gráficos de controle. Os gráficos são relativamente robustos a desvios de normalidade e conforme será mostrado mais adiante somente um dos gráficos obtidos a partir dos resíduos do modelo AR(1) foi capaz de detectar uma causa especial no processo durante a fase 2 do monitoramento.

**Tabela 7.10.** Valores P calculados para os testes de Ljung-Box e Shapiro-Wilk na análise da primeira diferença dos dados originais.

<b>Modelos AR(1)</b>	<b>Teste de Ljung-Box</b>	<b>Teste de Shapiro-Wilk</b>
	<b>Valor P*</b>	<b>Valor P*</b>
Densidade	0,9810	0,6172
Condutividade	0,3069	0,3552
Temperatura	0,3035	0,6612
Índice de Refração	0,7260	0,0021

### 7.6.5.1 Fase 1

Os resíduos dos modelos AR(1) apresentados na tabela 7.8 foram agrupados em uma matriz **E** para a análise em componentes principais. A decomposição da matriz **E** em quatro componentes principais, mostra que a primeira segunda, terceira e quarta componentes, descrevem respectivamente 68,46, 26.64, 4.44 e 0,04 % da variância total dos dados. As duas primeiras componentes descrevem juntas 95.33 % da variação dos dados e o modelo de duas componentes foi usado para a construção dos gráficos de controle multivariado.

A figura 7.18, mostra os gráficos *D* e *Q* para a fase 1 do CEMP. A carta *Q* assinala a observação em 210 horas fora dos limites. O gráfico MEWMA-*D* com fator de ponderação  $L = 0,2$  assinala a observação em 196 horas fora dos limites conforme mostra a figura 7.19.

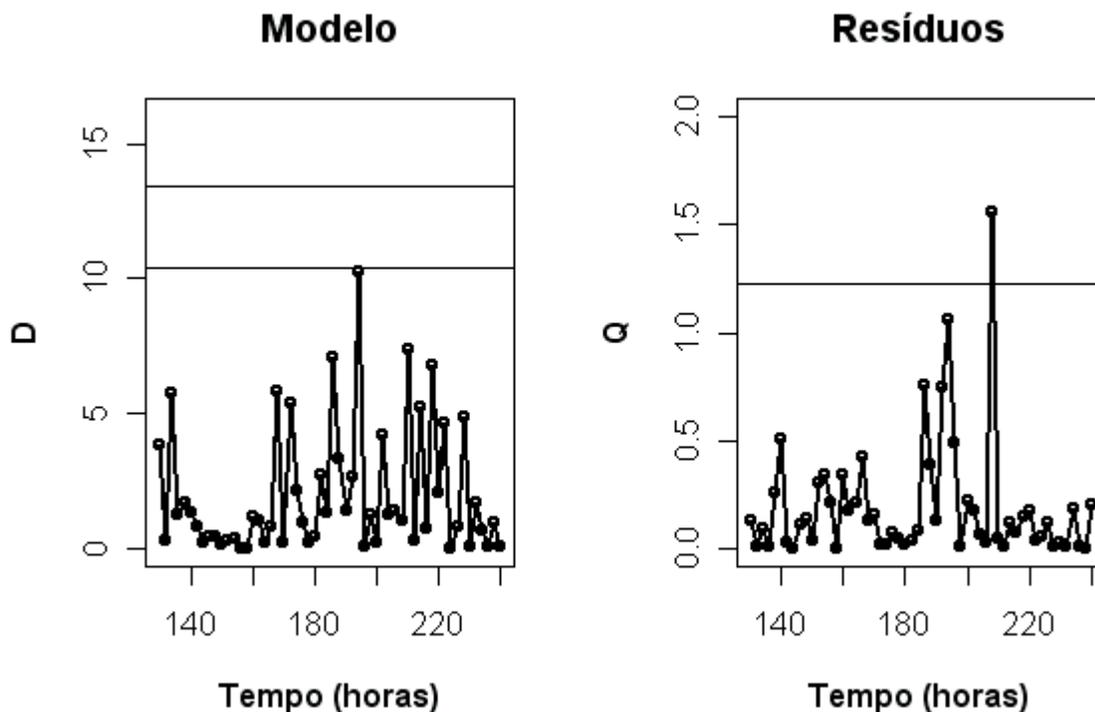
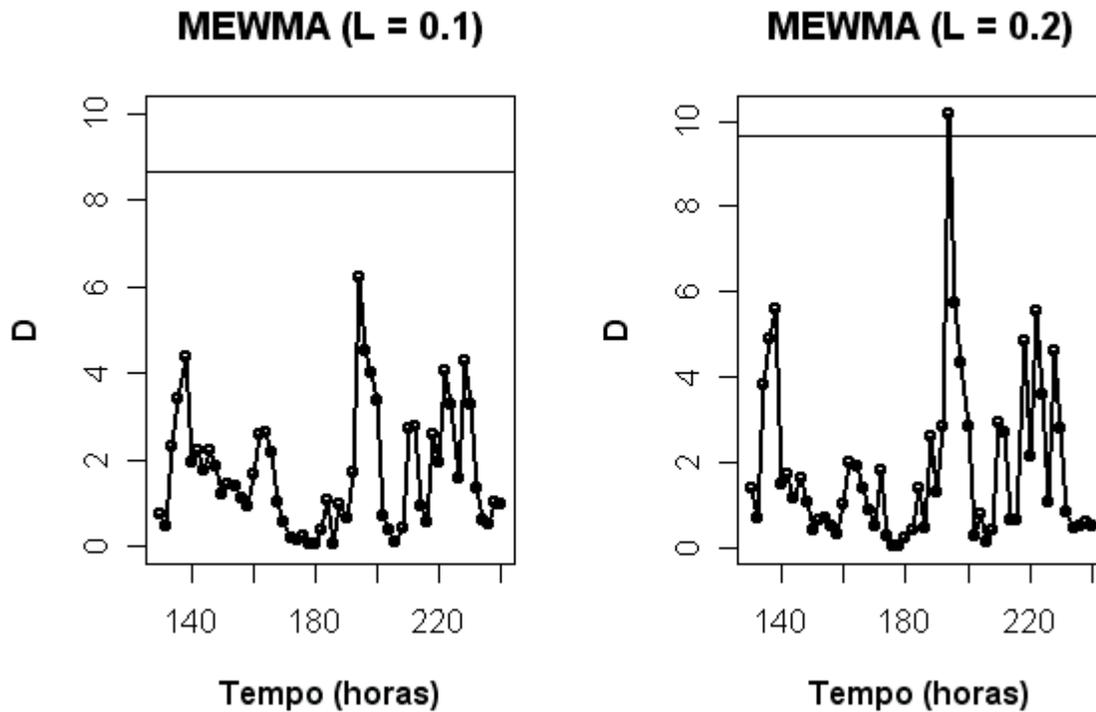


Figura 7.18. Gráficos de controle *D* e *Q*. Fase1.



**Figura 7.19.** Gráficos de controle MEWMA- $D$ . Fase1.

As observações que excederam os limites dos gráficos mencionados acima foram pontos isolados e estes foram considerados pontos anômalos. Estas duas observações foram removidas do conjunto de dados, e um novo modelo PCA com duas componentes foi calculado. Os gráficos  $D$ ,  $Q$  e MEWMA- $D$  para este novo modelo não assinalaram nenhuma observação fora dos limites de controle como mostram as figuras 7.20 e 7.21 e por isso o processo foi considerado em controle neste período de tempo.

A análise com os gráficos de controle construídos a partir do modelo ARIMA(0,1,0) deram resultados equivalentes aos obtidos com os resíduos dos modelos AR(1), sendo que foi concluído ao final da análise que o processo está em controle.

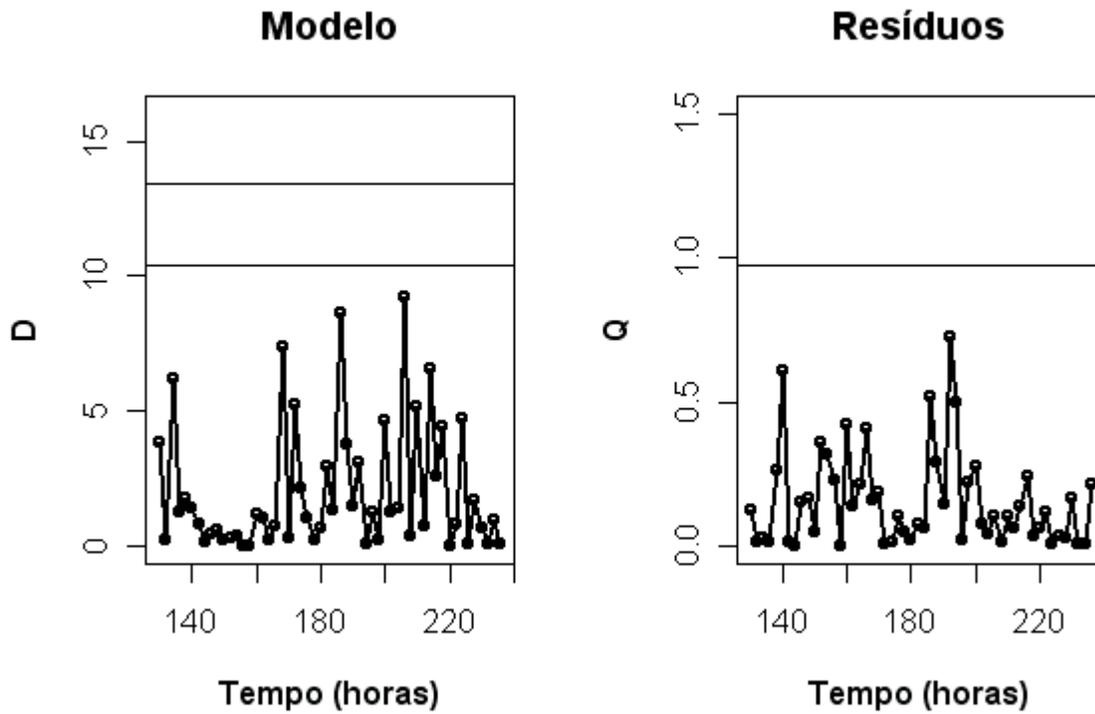


Figura 7.20 Gráficos de controle  $D$  e  $Q$  após a excluir os ponto anômalos. Fase1.

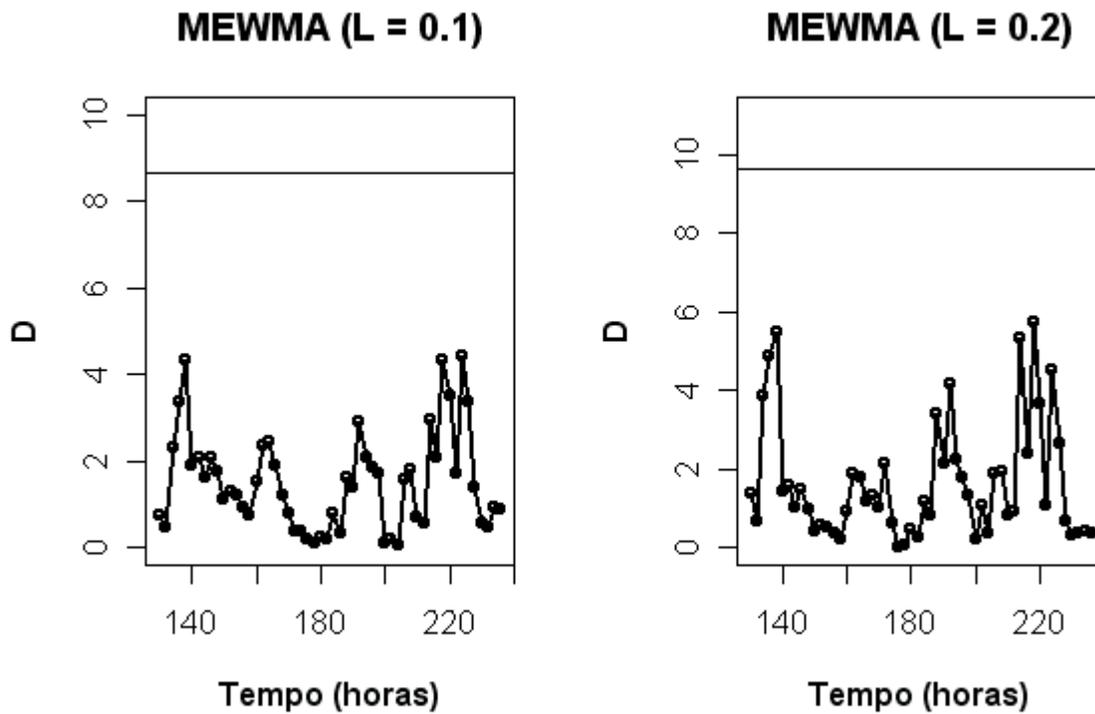


Figura 7.21. Gráficos de controle MEWMA- $D$  após a excluir os pontos anômalos. Fase1.

### 7.6.5.2 Fase 2

As figuras 7.22 e 7.23, mostram os resultados do monitoramento estatístico do processo com os gráficos  $D$ ,  $Q$  e MEWMA- $D$ , para o intervalo de tempo entre 242 e 578 horas de operação, assumindo que as variáveis do processo seguem um modelo AR(1) com os coeficientes apresentados na tabela 7.8. De acordo com os gráficos  $D$  e  $Q$  mostrados na figura 7.22, o processo está em controle até aproximadamente 350 horas. Em 276 horas, o gráfico  $Q$  detecta um ponto isolado fora dos limites de 99%, mas não foi encontrada nenhuma causa especial neste ponto de modo que o processo foi considerado em controle.

O gráfico  $Q$  da figura 7.22 assinala onze pontos fora dos limites de 99% no intervalo de tempo de 350 a 550 horas sugerindo que o processo está fora de controle após 348 horas. O gráfico  $D$  também assinala pontos fora de controle a partir de 348 horas, porém em menor quantidade que o gráfico  $Q$ .

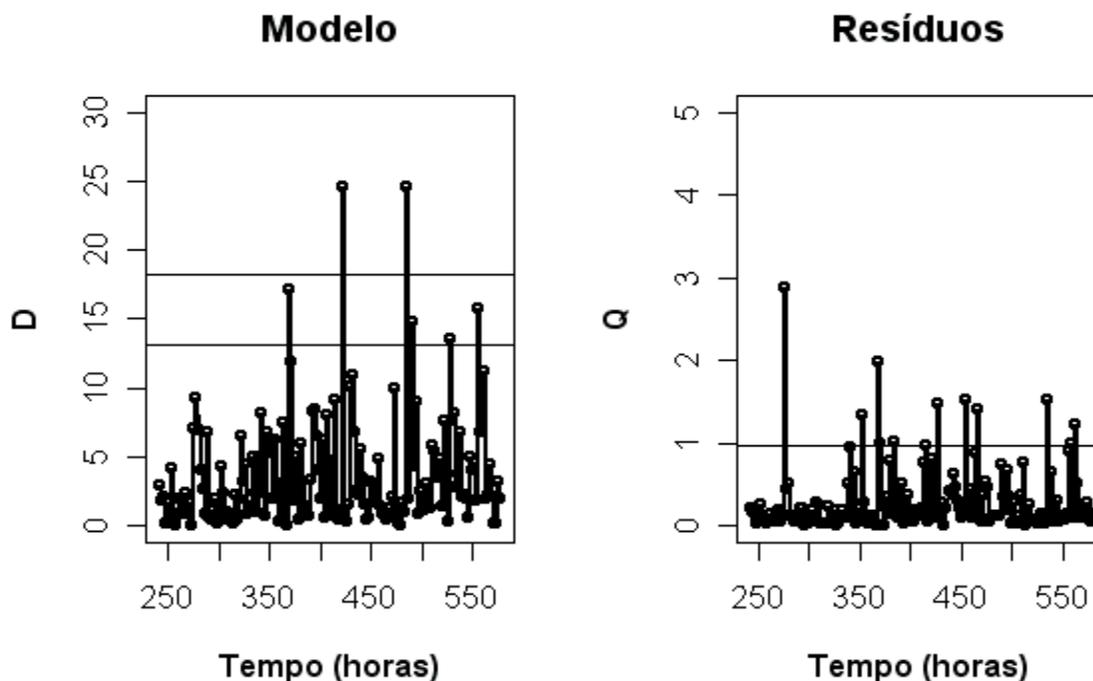
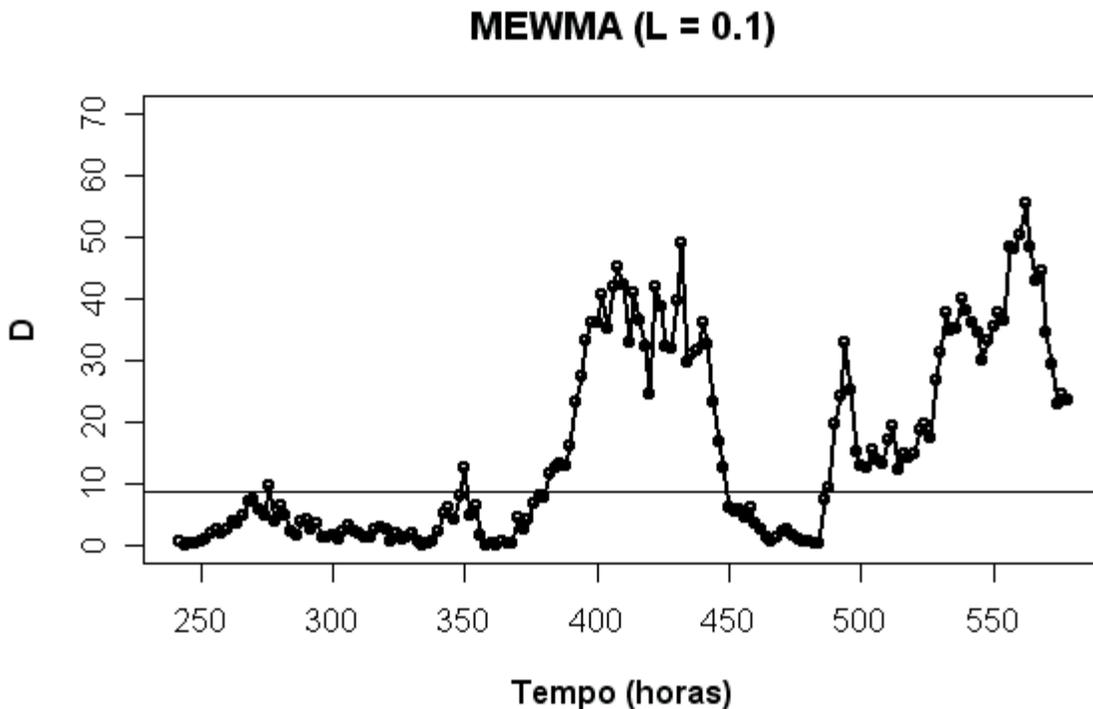


Figura 7.22. Gráficos de controle  $D$  e  $Q$ . Fase 2.

Mais interessante é a informação derivada do gráfico MEWMA- $D$  com o fator de ponderação  $L = 0,1$ , conforme apresentado na figura 7.23. Este gráfico mostra que houve um deslocamento sistemático da média multivariada do processo após 382 horas retornando a condição normal próximo de 450 horas para em seguida sofrer um novo deslocamento sistemático até o final do monitoramento.

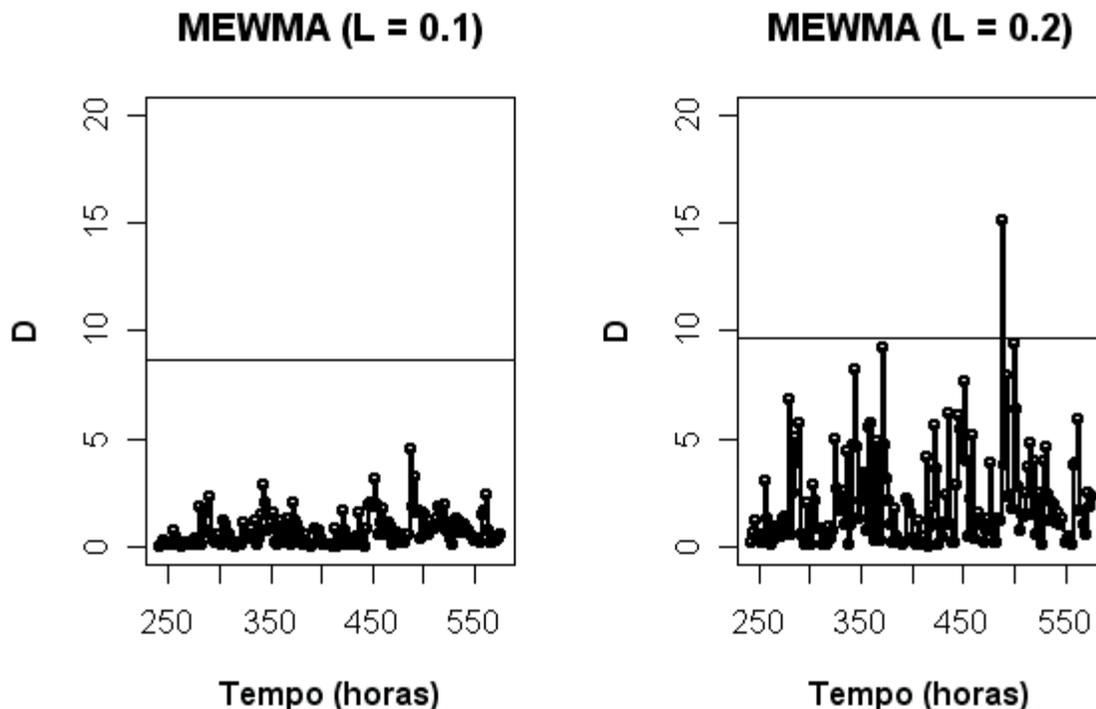


**Figura 7.23.** Gráficos de controle MEWMA- $D$  para Fase2 dos resíduos dos modelos AR(1).

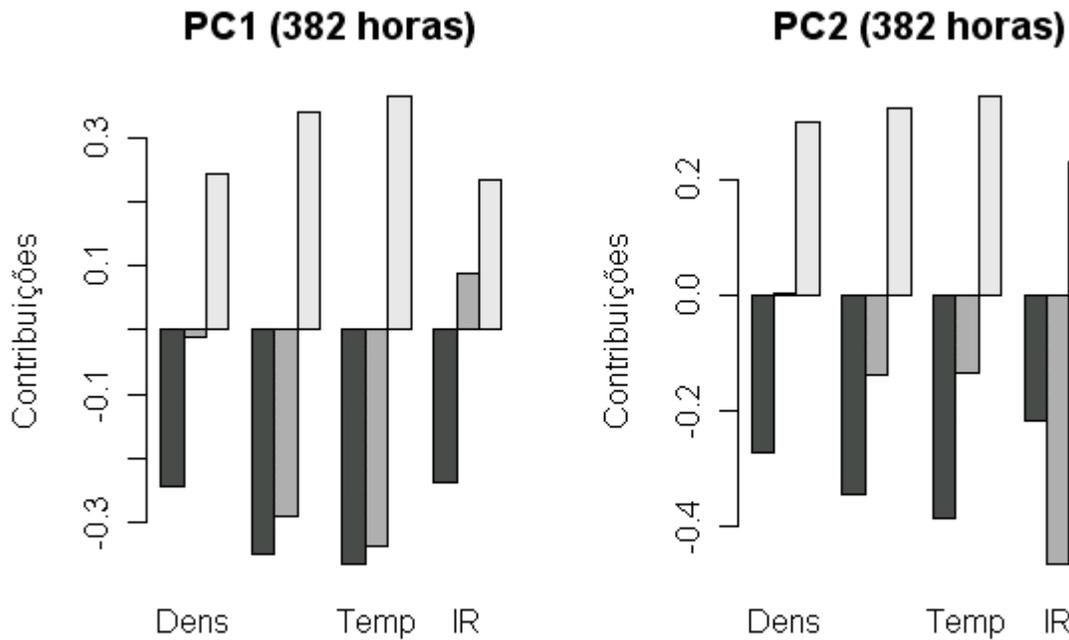
Embora não sejam mostrados, os gráficos de controle multivariado  $D$  e  $Q$  calculados a partir dos resíduos de modelos ARIMA(0,1,0) não detectaram nenhum sinal fora de controle durante o período da fase 2. O gráfico MEWMA- $D$  derivado do modelo ARIMA (0,1,0) também não detectou nenhum sinal fora de controle conforme mostra a figura 7.24. Neste ponto nota-se que os dois modelos propostos para remover a autocorrelação das variáveis originais forneceram informações contrárias na fase 2 do monitoramento estatístico, visto que o gráfico MEWMA- $D$  do modelo AR(1) assinala pontos fora de controle e o gráfico do modelo ARIMA(0,1,0) não assinala pontos fora de controle.

O fato de um mesmo processo poder ser descrito por diferentes modelos  $ARIMA(p,d,q)$  é uma característica citada no livro de Box<sup>223</sup> que menciona as possíveis ambigüidades associadas ao processo de identificação e modelagem das séries temporais. Ambos os modelos  $AR(1)$  e  $ARIMA(0,1,0)$  foram considerados adequados para a descrição do processo no que diz respeito ao critério de independência dos resíduos, mas apesar disto houve uma diferença na previsão do estado do processo.

A figura 7.25 mostra as contribuições de cada propriedade para as observações assinaladas fora de controle em 382 horas detectadas pelo gráfico MEWMA- $D$  dos resíduos dos modelos  $AR(1)$ . De acordo com estas contribuições, o índice de refração é a propriedade responsável pelo ponto assinalado fora dos limites de controle neste instante.



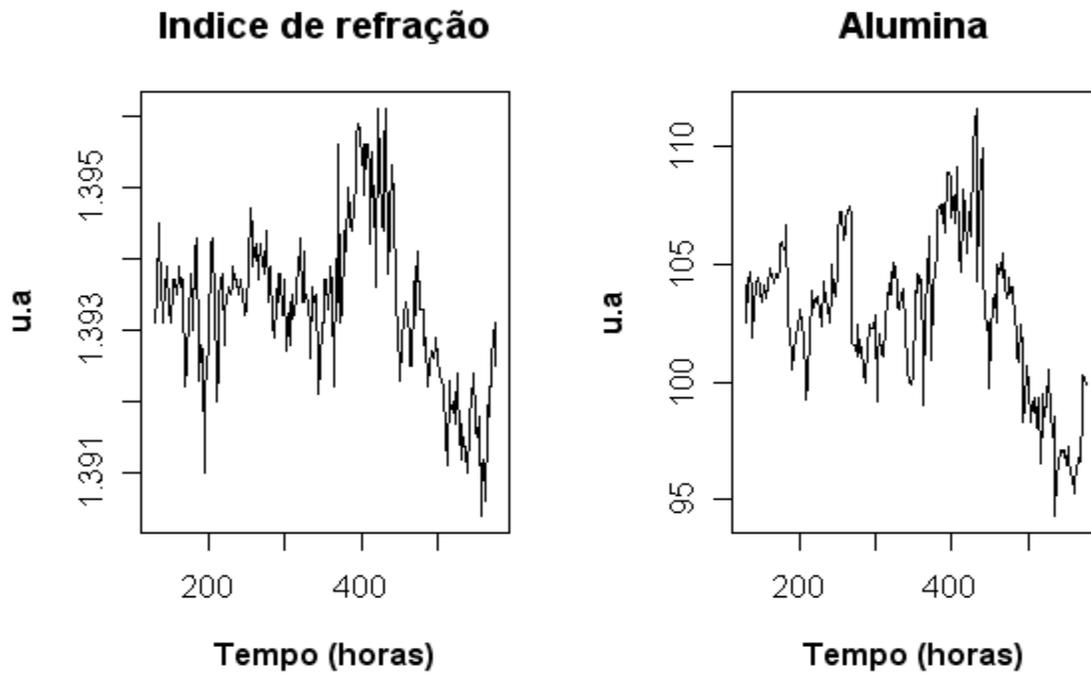
**Figura 7.24.** Gráficos de controle MEWMA- $D$  para a Fase2 tomados da primeira diferença dos dados originais ( $ARIMA(0,1,0)$ ).



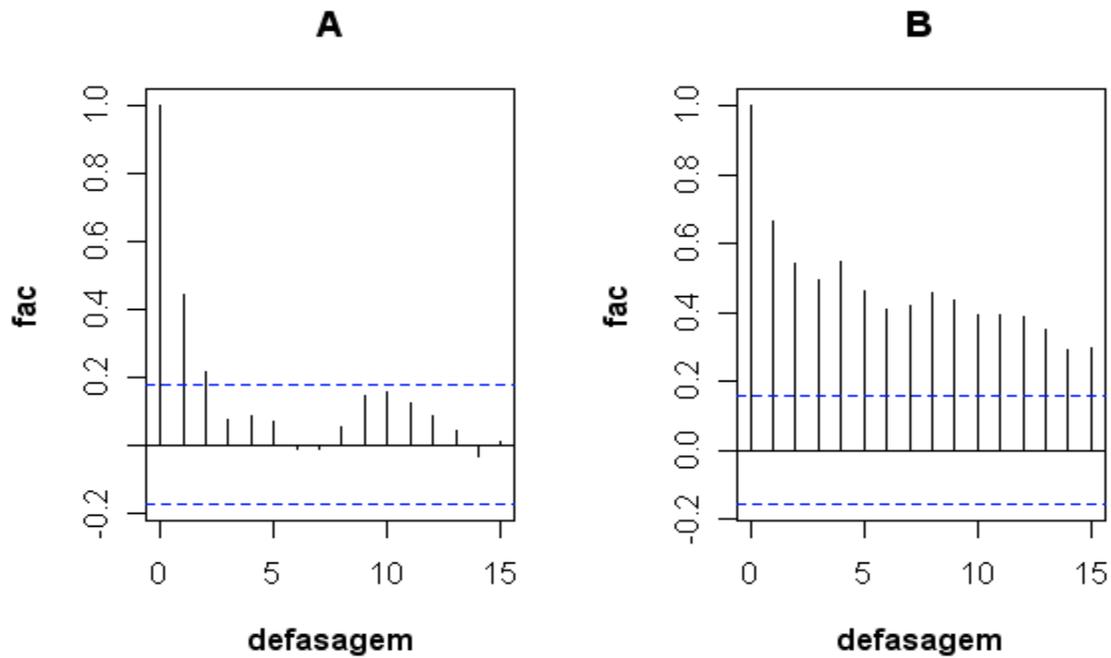
**Figura 7.25.** Gráfico das contribuições das propriedades para a observação feita 382 horas. A barra mais a esquerda é o limite inferior e a mais a direita é o limite superior.

A figura 7.26 mostra os valores das medidas do índice de refração e de alumina na solução de aluminato de sódio coletados durante a fase 1 e 2 do monitoramento. Perceba que até aproximadamente 382 horas, a média dos valores de índice de refração localiza-se em um mesmo patamar mas muda abruptamente para um valor maior, após 382 horas e para um valor menor após 500 horas, sendo o mesmo observado para a concentração de alumina.

O correlograma apresentado na figura 7.27 da série de medidas de índice de refração até 382 horas é similar ao calculado entre 130 e 240 horas no item 7.6.5 (figura 7.13). Isto mostra que o modelo AR(1) é válido até 382 horas de operação. Por outro lado, o correlograma calculado entre 130 e 400 horas mostra um comportamento totalmente diferente, característico de uma série temporal não estacionária, onde o decaimento dos coeficientes de correlação para valores próximo a zero é bastante lento.



**Figura 7.26.** Medidas de índice de refração coletadas entre 130 e 578 horas.



**Figura 7.27.** A) Correlograma calculada para as observações entre 130 e 380 horas. B) Correlograma calculada para as observações entre 130 e 400 horas.

O comportamento do índice de refração na solução de aluminato de sódio sugere que de fato há uma causa especial responsável pela alteração no valor médio das medidas de índice de refração que estão correlacionadas com uma alteração no valor médio da concentração de alumina no processo. Ao se investigar uma das possíveis causas para este deslocamento, observou-se que a média da concentração de carbonato de sódio ao longo do tempo foi alterada próximo a 382 horas, primeiro para valores maiores e após 420 horas para valores menores, mostrando uma correlação com o comportamento do índice de refração e da concentração de alumina na solução supersaturada de aluminato de sódio.

A variação da concentração de carbonato de sódio na solução de aluminato do processo Bayer ocorreu devido a uma variação da quantidade de cal adicionada ao processo para eliminar as impurezas como oxalatos e fosfatos que interferem no rendimento deste. Esta variação da concentração de carbonato alterou a concentração média de alumina da solução de aluminato de sódio que é diretamente correlacionada com o índice de refração. Sendo assim, somente o gráfico MEWMA-*D* do modelo AR(1) foi capaz de detectar este evento.

## 7.7 CONCLUSÃO

Neste trabalho apresentamos uma metodologia de controle estatístico multivariado para a análise de processos autocorrelacionados. Embora a metodologia proposta tenha sido usada para o monitoramento de uma planta industrial do processo Bayer, ela é geral e pode ser aplicada para análise de outros processos químicos.

O princípio básico da metodologia é o de ajustar um modelo de série temporal para cada propriedade do processo e combinar os resíduos de cada um desses modelos em uma matriz para a análise em componentes principais (PCA). As séries temporais são usadas para modelar a autocorrelação entre medidas ao longo do tempo e a análise em componentes principais é usada para modelar a correlação entre os resíduos de cada modelo de série temporal.

Os resultados obtidos neste trabalho mostraram que os resíduos dos modelos das séries temporais ajustados para as propriedades do processo Bayer são formados por

observações independentes que seguem uma distribuição de probabilidade normal. Por isso, estes resíduos puderam ser usados para construir os gráficos de controle multivariado do tipo  $D$ ,  $Q$  e MEWMA- $D$ . Além disso foi verificado que o resíduo do modelo de uma determinada variável é correlacionado com o resíduo do modelo de outra variável, da mesma forma que são as duas variáveis originais.

O gráfico MEWMA- $D$  foi considerado o mais efetivo porque foi capaz de assinalar com maior clareza os pontos fora de controle. Na prática, o processo químico deve ser monitorado com os três gráficos apresentados no trabalho, porque o gráfico MEWMA- $D$  foi projetado para detectar pequenos deslocamentos e os gráficos  $D$  e  $Q$  para detectar grandes deslocamentos do valor médio das observações em relação a média do modelo estatístico ajustado para o processo.

Neste trabalho utilizamos um modelo de série temporal para cada propriedade, porém trabalhos futuros podem explorar modelos de séries temporais multivariadas para a análise das propriedades do processo. Além disso é importante desenvolver uma metodologia que guie a escolha do fator de ponderação para o cálculo das médias móveis multivariadas ponderadas exponencialmente (matriz  $\mathbf{Z}$ ) e a influência que este fator tem na capacidade do gráfico MEWMA- $D$  de detectar os sinais fora dos limites estatísticos.

## **CAPÍTULO 8**

### **CONCLUSÃO GERAL**

## 8.1 CONSIDERAÇÕES FINAIS

Neste trabalho de tese foram desenvolvidas metodologias de análise multivariada em  $N$ -modos para o monitoramento quantitativo, qualitativo e estatístico de reações químicas. Duas técnicas de análise foram exploradas para o desenvolvimento dessas metodologias: A primeira foi a técnica de resolução multivariada de curvas para a análise de dados espectroscópicos coletados em reações químicas e bioquímicas. A segunda técnica investigada foi o controle estatístico multivariado de processos para a análise de um conjunto de variáveis monitoradas ao longo do tempo em uma reação química usada para extrair alumínio do minério bauxita.

Ao final do trabalho verificamos que a técnica de resolução multivariada de curvas é uma ferramenta extremamente poderosa, porque com ela podemos deduzir um modelo que descreve a composição química das reações dispondo apenas de dados espectroscópicos. Com a MCR podemos deduzir o número de espécies químicas e resolver a partir de um conjunto de espectros sobrepostos os sinais individuais de cada um dos compostos químicos que tomam parte da reação.

Os métodos da MCR foram classificados de acordo com o número de modos de monitoramento para tornar mais fácil o planejamento da análise de dados. Por exemplo, a metodologia de monitoramento espectro-cinético multivariado desenvolvida para a determinação de creatinina em soro humano (Capítulo 4) foi planejada em cima da vantagem de segunda ordem que os métodos de decomposição trilinear da MCR oferecem. O objetivo da metodologia proposta foi eliminar ou pelo menos diminuir o efeito da interferência causada por outros compostos químicos na determinação da creatinina em soro humano. Este objetivo foi alcançado para a maioria dos compostos químicos testados e a metodologia desenvolvida apresentou um erro padrão de desempenho SEP em torno de  $2,22 \text{ mg L}^{-1}$  com um respectivo coeficiente de variação em torno de 6 % para todos os métodos de decomposição trilinear testados (ler Capítulo 4 para maiores detalhes desta metodologia).

O método espectro cinético-multivariado desenvolvido para a determinação da creatinina em soro humano tem a vantagem de que não é necessário executar procedimentos de separação do analito e dos compostos químicos interferentes, uma vez

que a seletividade da medida é obtida a partir do método da MCR que resolve o sinal associado a concentração da creatinina de forma unequivoca. Em trabalhos futuros a técnica de planejamento e otimização de experimentos poderia ser usada para encontrar uma condição experimental que minimiza o tempo de análise necessário para as determinação da creatinina. Outro trabalho importante é fazer a validação do método.

O método de decomposição trilinear PARAFAC foi uma excelente ferramenta para a explorar os metabólitos intracelulares e autofluorescentes da *Chromobacterium violaceum* cultivada em bioreator e em condições aeróbicas. Esta bactéria é responsável pela biossíntese de um pigmento violeta denominado de violaceína que apresenta atividades antibiótica e antitumoral.

A decomposição dos espectros de fluorescência no modelo trilinear permitiu identificar os amino ácidos triptofano, tirosina e o pigmento como metabólitos intracelulares deste microorganismo. Neste trabalho, foi verificado que a *C. violaceum* consome todo o triptofano intracelular antes de iniciar a produção do pigmento. A informação obtida do modelo trilinear foi qualitativa e infelizmente não foi possível fazer determinações quantitativas da taxa de consumo de triptofano devido aos desvios de trilinearidade que os dados apresentaram, sendo que esta é uma característica que afeta a exatidão dos sinais resolvidos pelo método PARAFAC.

Para trabalhos futuros, a metodologia desenvolvida para a análise dos metabólitos da *Chromobacterium violaceum* por espectroscopia de fluorescência pode ser ampliada para as análise quantitativas que visam determinar a taxa de consumo de triptofano e de produção da violaceína. Mas para esta tarefa provavelmente seja melhor usar os métodos de calibração multivariada visto que estes são menos sensíveis aos desvios de trilinearidade. Neste caso, o método PARAFAC pode ser adotado para identificação dos compostos químicos nas amostras como um método de análise complementar.

A metodologia desenvolvida no capítulo 2 foi concebida para o monitoramento de reatores químicos por espectroscopia tendo como objetivo principal caracterizar a composição química da reação ao longo do tempo. Os métodos de decomposição bilinear da MCR foram muito úteis para esta tarefa visto que com eles foi possível ajustar um modelo que descreve a composição do processo químico ao longo do tempo usando apenas os dados de espectroscopia.

A metodologia proposta para a correção dos espectros do modelo de referência é bastante versátil e trabalhos futuros poderiam testar esta metodologia para a correção do modelo quando os espectros dos compostos sofrem mudanças devido as alterações de temperatura ou algum outro tipo de agente físico-químico capaz de modificar os espectros dos compostos químicos.

Além do MCR, apresentamos uma outra técnica para monitorar reações químicas em processos industriais. Para a análise de variáveis físico-químicas tão distintas como condutividade, temperatura, índice de refração e densidade, a técnica de resolução multivariada de curvas não pode ser aplicada, e por isso foi explorada a técnica de controle estatístico multivariado de processos (CEMP). Diferente da MCR, no CEMP são usados modelos estatísticos para descrever o comportamento das variáveis, de modo que a informação obtida neste tipo de análise é de natureza diferente da informação obtida com os métodos da MCR.

No capítulo 7 o CEMP foi usado para analisar um processo industrial de extração do alumínio da bauxita. Devido a autocorrelação das medidas neste processo, foi necessário desenvolver uma metodologia de CEMP adequada o tratamento dos dados com essa característica. Nós usamos modelos de séries temporais, mais especificamente o modelo AR(1) para modelar a autocorrelação dos dados. Os resíduos desses modelos seguem a distribuição normal e são independentes entre si e, por isso, apresentam as características necessárias a construção dos gráficos de controle multivariado.

Embora a metodologia apresentada no capítulo 7 tenha sido satisfatória para os objetivos propostos, ele é apenas uma das abordagens possíveis para o tratamento de dados autocorrelacionadas. Os trabalhos futuros poderiam explorar os modelos de séries temporais multivariados, como os modelos VARMA, ou o método “Dynamic PCA” para comparar o desempenho destas metodologias com o da metodologia proposta neste trabalho de tese. Uma outra linha de pesquisa poderia explorar os modelos preditivos multivariados para a análise de processos autocorrelacionados. Este último caso é extremamente importante para a análise do processo Bayer investigado nesta tese, visto que as variáveis de processo são medidas obtidas em tempo real e poderiam ser usadas para inferir as variáveis de qualidade em intervalos de tempo muito menor que as medidas feitas no laboratório.

Por fim, verificamos que os trabalhos desenvolvidos nesta tese abrem muitas outras oportunidades de pesquisa de interesse tecnológico e científico usando como base a análise multivariada em N-modos. Esta última é uma excitante área de desenvolvimento que apresenta muitos problemas que ainda precisam ser resolvidos principalmente para a análise de sistemas químicos tão complexos como bioreatores e plantas de processos químicos industriais.

## 9. REFÊRENCIAS BIBLIOGRÁFICAS

1. Johnson, R. A.; Wichern, D. W. *Applied Multivariate Statistical Analysis*. Fourth edition, Prentice Hall. New Jersey, USA. **1998**, p. 377.
2. Geladi, P. Chemometrics in spectroscopy. Part 1. classical chemometrics. *Spectrochimica Acta; Part. B: Atomic Spectroscopy*, **2003**, 58, 767-782.
3. Wold, S.; Sjostrom, M.; Eriksson, L. PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, **2001**, 58, 109-130.
4. Jackson, J. E. Principal components and factors-analysis: Principal components. *Journal of Quality Technology*, **1980**, 12, 201-213.
5. Gurden, S. P.; Monteiro, V. F.; Longo, E.; Ferreira, M. M. C. Quantitative analysis and classification of AFM images of human hair. *Journal of Microscopy*, **2004**, 215, 13-23.
6. Cardoso, D. R.; Andrade-Sobrinho, L. G.; Leite-Neto, A. F.; Reche, R. V.; Isique, W. D.; Ferreira, M. M. C; Lima-Neto, B. S.; Franco, D. W. Comparison between cachaca and rum using pattern recognition methods. *Journal of Agricultural and Food Chemistry*, **2004**, 52, 3429-3433.
7. MacGregor, J. F.; Kourti, T. Statistical process-control of multivariate process. *Control Engineering Practice*, **1995**, 3, 403-414.
8. Morgano, M.; Queiroz, S. C. N.; Ferreira, M. M. C. Aplicação da análise exploratória na diferenciação de vegetais. *Brazilian Journal of Food Technology*, **1999**, 2, 73-79.
9. dos Reis, M. M.; Ferreira, M. M. C. Curve resolution of simulated and total luminescence spectra by Generalized Rank Annihilation Method (GRAM). *Quimica Nova*, **1999**, 22, 11-17.
10. Fardim, P.; Ferreira, M. M. C.; Duran, N. Multivariate calibration for quantitative analysis of eucalypt kraft pulp by NIR spectrometry. *Journal of Wood Chemistry and Technology*, **2002**, 22, 67-81.
11. Ferreira, M. M. C. Multivariate QSAR. *Journal of Brazilian Chemical Society*, **2002**, 13, 742-753.
12. Tauler, R.; Barcelo, D. Multivariate curve resolution applied to liquid-chromatography diode-array detection. *Trac-Trends in Analytical Chemistry*, **1993**, 12, 319-327.
13. Guterres, M. V.; Volpe, P.; Ferreira, M. M. C. Multiway calibration for creatinine determination in human serum using the Jaffe reaction. *Applied Spectroscopy*, **2004**, 58, 54-60.

- 
14. Van Loan, C. F. The ubiquitous Kronecker product. *Journal of Computational and Applied Mathematics*, **2000**, 123, 85-100.
15. Khatri, C. G.; Rao, C.R. Solutions to some functional equations and their applications to characterization of probability distributions. *Sankhyā: The Indian Journal of Statistics*, **1968**, 30A, 167-180.
16. Rodrigues-Cuesta, M. J.; Boqué, R.; Rius, F. X.; Martinez, V. J. L.; Frensch, G. A. Development and validation of a method for determining pesticides in groundwater from complex overlapped HPLC signals and multivariate curve resolution. *Chemometrics and Intelligent Laboratory Systems*, **2005**, 77, 251-260.
17. Garrido, M.; Larrechi, M. S.; Rius, F. X. Multivariate curve resolution-alternating least squares and kinetic modeling applied to near-infrared data from curing reactions of epoxy resins: Mechanistic approach and estimation of kinetic rate constants. *Applied Spectroscopy*, **2006**, 60, 174-181.
18. Izquierdo Ridorsa, A.; Saurina, J.; Hernandez Cassou, S.; Tauler, R. Second-order multivariate curve resolution applied to rank-deficient data obtained from acid-base spectrophotometric titrations of mixtures of nucleic bases. *Chemometrics and Intelligent Laboratory Systems*, **1997**, 38, 183-196.
19. Saurina, J.; Hernandez-Cassou, S.; Tauler, R.; Izquierdo-Ridorsa, A. Spectrophotometric determination of pK(a) values based on a pH gradient flow-injection system. *Analytica Chimica Acta*, **2000**, 408, 135-143.
20. Navea, S.; de Juan, A.; Tauler, R. Three-way data analysis applied to multispectroscopic monitoring of protein folding. *Analytica Chimica Acta*, **2001**, 446, 187-197.
21. Antunes, A. M.; Ferreira, M.; Volpe, P. A chemometric study of amino acid transport through liquid membranes using UV-vis spectroscopy. *Journal of Chemometrics*, **2002**, 16, 111-116.
22. de Juan, A.; Tauler, R.; Dyson, R.; Marcolli, C.; Rault, M.; Maeder, M. Spectroscopic imaging and chemometrics: a powerful combination for global and local sample analysis. *Trac-Trends in Analytical Chemistry*, **2004**, 23, 70-79.
23. Grabaric, B. S.; Grabaric, Z.; Tauler, R.; Esteban, M.; Casassas, E. Application of multivariate curve resolution to the voltammetric data - Factor analysis ambiguities in the study of weak consecutive complexation of metal ion with ligand. *Analytica Chimica Acta*, **1997**, 341, 105-120.
24. Stanimirova, I.; Simeonov, V. Modeling of environmental four-way data from air quality control. *Chemometrics and Intelligent Laboratory Systems*, **2005**, 77, 115-121.

- 
25. Bro, R.; Heimdal, H. Enzymatic browning of vegetables. Calibration and analysis of variance by multiway methods. *Chemometrics and Intelligent Laboratory Systems*, **1996**, 34, 85-102.
26. Reis, M. M.; Ferreira, M.; Sarmiento, S. A multi-way analysis of starch cassava properties. *Chemometrics and Intelligent Laboratory Systems*, **2002**, 64, 123-135.
27. Smilde, A. K.; Wang, Y. D.; Kowalski, B. R. Theory of medium-rank 2nd-order Calibration with Restricted-tucker models. *Journal of Chemometrics*, **1994**, 8, 21-36.
28. Malinowski, E. R. Window factor-analysis - theoretical derivation and application to flow-injection analysis data. *Journal of Chemometrics*, **1992**, 6, 29-40.
29. Maeder, M. Evolving factor analysis for the resolution of overlapping chromatographic peaks. *Analytical Chemistry*, **1987**, 59, 527-530.
30. Sanches, E.; Kowalski, B. R. Tensorial resolution: a direct trilinear decomposition. *Journal of Chemometrics*, **1990**, 4, 29-45.
31. Tauler, R.; Smilde, A.; Kowalski, B. Selectivity, local rank, 3-way data-analysis and ambiguity in multivariate curve resolution. *Journal of Chemometrics*, **1995**, 9, 31-58.
32. Jiang, T. H.; Liang, Y.; Ozaki, Y. Principles and methodologies in self-modeling curve resolution. *Chemometrics and Intelligent Laboratory Systems*, **2004**, 71, 1-12.
33. Golub, G. H.; Van Loan, C. F. *Matrix Computations*. Third edition, Johns Hopkins University Press. Maryland, USA. **1996**, p. 375.
34. Li, S. S.; Gemperline, P. J. Eliminating Complex Eigenvectors and Eigenvalues in multi way analyses using the direct trilinear decomposition method. *Journal of Chemometrics*, **1993**, 7, 77-88.
35. Björck, A. *Numerical methods for least squares problems*. Society for Industrial and Applied Mathematics (SIAM). Philadelphia, USA. **1996**.
36. Paatero, P.; Tapper, U. Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. *Environmetrics*, **1994**, 5, 111-126.
37. Wu, H. L.; Shibukawa, M.; Oguma, K. An alternating trilinear decomposition algorithm with application to calibration of HPLC-DAD for simultaneous determination of overlapped chlorinated aromatic hydrocarbons. *Journal of Chemometrics*, **1998**, 12, 1-26.
38. Chen, Z. P.; Wu, H. L.; Yu, R. Q. On the self-weighted alternating trilinear decomposition algorithm - the property of being insensitive to excess factors used in calculation. *Journal of Chemometrics*, **2001**, 15, 439-453.

- 
39. Faber, N. M.; Bro, R.; Hopke, P. Recent developments in CANDECOP/ PARAFAC algorithms: A critical review. *Chemometrics and Intelligent Laboratory Systems*, **2003**, 65, 119-137.
40. Tomasi, G.; Bro, R. A comparison of algorithms for fitting the PARAFAC model. *Computational Statistical & Data Analysis*, **2006**, 50, 1700-1734.
41. Smilde, A. K.; Hoefsloot, H. C. J.; Kiers, H. A. L.; Bijlsma, S.; Boelens, H. F. M. Sufficient conditions for unique solutions within a certain class of curve resolution models. *Journal of Chemometrics*, **2001**, 15, 405-411.
42. Ten Berge, J. M. F.; Sidiropoulos, N. D. On Uniqueness in Candecomp/Parafac. *Psychometrika*, **2002**, 67, 399-409.
43. Gemperline, P. J. Computation of the range of feasible solutions in self-modeling curve resolution algorithms. *Analytical Chemistry*, **1999**, 71, 5398-5404.
44. Garrido, M.; Larrechi, M. S.; Rius, F. X.; Tauler, R. Calculation of band boundaries of feasible solutions obtained by Multivariate Curve Resolution Alternating Least Squares of multiple runs of a reaction monitored by NIR spectroscopy. *Chemometrics and Intelligent Laboratory Systems*, **2005**, 76, 111-120.
45. Bro, R.; Jong, S. A fast non-negativity-constrained least squares algorithm. *Journal of Chemometrics*, **1997**, 11, 393-401.
46. Gurden, S. P.; Westerhuis, J. A.; Bijlsma, S.; Smilde, A. K. Modelling of spectroscopic batch process data using grey models to incorporate external information. *Journal of chemometrics*, **2001**, 15, 101-121.
47. Bro, R. *Multi-way Analysis in the Food Industry: Models, Algorithms, and Applications*. Ph. D. Thesis, University of Amsterdam. The Netherlands. **1998**, p. 135.
48. Van Benthem, M. H.; Keenan, M. R.; Haaland, D. M. Application of equality constraints on variables during alternating least squares procedures. *Journal of Chemometrics*, **2002**, 16, 613-622.
49. Lawson, C. L.; Hanson, R. J. *Solving least squares problems*. Prentice-Hall, Inc. Englewood Cliffs. **1974**.
50. Haskell, K. H.; Hanson, R. J. An algorithm for linear least squares problems with equality and non-negativity constraints. *Mathematical Programming*, **1981**, 21, 98-118.
51. Miller, C. E. Chemometrics for on-line spectroscopy applications - theory and practice. *Journal of Chemometrics*, **2000**, 14, 513-528.

- 
52. Hassell, D. C.; Bowman, E. M. Process Analytical Chemistry for Spectroscopists. *Applied Spectroscopy*, **1998**, 52, 18A-29A.
53. Callis, J. B.; Illman, D. L.; Kowalski, B. R. Process analytical chemistry. *Analytical Chemistry*, **1987**, 59, 624A-637A.
54. Ma, B.; Gemperline, P. J.; Cash, E.; Bosserman, M.; Comas, E. Characterizing batch reactions with *in situ* spectroscopic measurements, calorimetry and dynamic modeling. *Journal of Chemometrics*, **2003**, 17, 470-479.
55. Gourvenec, S.; Lamotte, C.; Pestiaux, P.; Massart, D. L. Use of the orthogonal projection approach (OPA) to monitor batch processes. *Applied Spectroscopy*, **2003**, 57, 80-87.
56. Van Sprang, E. N. M.; Ramaker, H. J.; Westerhuis, J. A.; Smilde, A. K.; Gurden, S. P.; Wienke, D. Near-infrared spectroscopic monitoring of a series of industrial batch processes using a bilinear grey model. *Applied Spectroscopy*, **2003**, 57, 1007-1019.
57. Gemperline, P. J.; Zhu, M.; Cash, E.; Walker, D. S. Chemometric characterization of batch reactions. *ISA Transactions*, **1999**, 38, 211-216.
58. Tauler, R.; Kowalski, B. R.; Fleming, S. Multivariate curve resolution applied to spectral data from multiple runs of an industrial process. *Analytical Chemistry*, **1993**, 65, 2040-2047.
59. Thurston, T. J.; Brereton, R. G. Estimation of second order rate constants using chemometric methods with kinetic constraints. *Analyst*, **2002**, 127, 659-668.
60. Zhu, Z. L.; Xia, J.; Zhang, J.; Li, T. H. Determination of rate constants from two-way kinetic-spectral data by using rank annihilation factor analysis. *Analytica Chimica Acta*, **2002**, 454, 21-30.
61. Westerhuis, J. A.; Gurden, S.P.; Smilde, A. K. Spectroscopic monitoring of batch reactions for *on-line* fault detection and diagnosis. *Analytical Chemistry*, **2000**, 72, 5322-5330.
62. Salomaa, I. K.; Kauppinen, J. K. Origin of and compensation for the baseline errors in Fourier transform spectra. *Applied Spectroscopy*, **1998**, 52, 579-586.
63. Wulfert, F.; Kok, W. T.; Smilde, A. K. Influence of temperature on vibrational spectra and consequences for the predictive ability of multivariate models. *Analytical Chemistry*, **1998**, 70, 1761-1767.
64. Haaland, D. M.; Melgaard, D. K. New prediction-augmented classical least-squares (PACLS) methods: Application to unmodeled interferents. *Applied Spectroscopy*, **2000**, 54, 1303-1312.

- 
65. Lawton, W. H.; Sylvestre, E. A. Self modeling curve resolution. *Technometrics*, **197**, 13, 617-633.
66. Sylvestre, E. A.; Lawton, W. H.; Maggio, M. S. Curve resolution using a postulated chemical reaction. *Technometrics*, **1974**, 16, 353-368.
67. Furusjo, E.; Danielsson, L. G. A method for the determination of reaction mechanisms and rate constants from two-way spectroscopic data. *Analytica Chimica Acta*, **1998**, 373, 83-94.
68. Quinn, A. C.; Gemperline, P. J.; Baker, B.; Zhu, M.; Walker, D. S. Fiber-optic UV / Visible composition monitoring for process control of batch reactions. *Chemometrics and Intelligent Laboratory Systems*, **1999**, 45, 199-214.
69. Bezemer, E.; Rutan, S. C. Multivariate curve resolution with non-linear fitting of kinetic profiles. *Chemometrics and Intelligent Laboratory Systems*, **2001**, 59, 19-31.
70. Zogg, A.; Fischer, U.; Hungerbuhler, K. A new approach for a combined evaluation of calorimetric and online infrared data to identify kinetic and thermodynamic parameters of a chemical reaction. *Chemometrics and Intelligent Laboratory Systems*, **2004**, 71, 165-176.
71. Bijlsma, S.; Boelens, H.; Hoefsloot, H.; Smilde, A. K. Constrained least squares methods for estimating reaction rate constants from spectroscopic data. *Journal of Chemometrics*, **2002**, 16, 28-40.
72. Bijlsma, S.; Louwerse, D. J.; Smilde, A. K. Rapid estimation of rate constants of batch processes using on-line SW-NIR. *AIChE Journal*, **1998**, 44, 2713-2723.
73. Bezemer, E.; Rutan, S. C. Three-way alternating least squares using three-dimensional tensors in MATLAB (R). *Chemometrics and Intelligent Laboratory Systems*, **2002**, 60, 239-251.
74. Bijlsma, S.; Louwerse, D. J.; Smilde, A. K. Estimating rate constants and pure UV-vis spectra of a two-step reaction using trilinear models. *Journal of Chemometrics*, **1999**, 13, 311-329.
75. Saurina, J.; Hernandez-Cassou, S.; Tauler, R.; Izquierdo-Ridorsa, A. Multivariate resolution of rank-deficient spectrophotometric data from first-order kinetic decomposition reactions. *Journal of Chemometrics*, **1998**, 12, 183-203.
76. Reis, M. M.; Gurden, S. P.; Smilde, A. K.; Ferreira, M. M. C. Calibration and detailed analysis of second-order flow injection analysis data with rank overlap. *Analytica Chimica Acta*, **2000**, 422, 21-36.

- 
77. Amrhein, M.; Srinivasan, B.; Bonvin, D.; Schumacher, M. M. On the rank deficiency and rank augmentation of the spectral measurement matrix. *Chemometrics and Intelligent Laboratory Systems*, **1996**, 33, 17-33.
78. Bro, R.; Smilde, A. K. Centering and scaling in component analysis. *Journal of Chemometrics*, **2003**, 17, 16-33.
79. Garrido, M.; Lazaro, I.; Larrechi, M. S.; Rius, F. X. Multivariate resolution of rank-deficient near-infrared spectroscopy data from the reaction of curing epoxy resins using the rank augmentation strategy and multivariate curve resolution alternating least squares approach. *Analytica Chimica Acta*, **2004**, 515, 65-73.
80. Saurina, J.; Hernandez-Cassou, S.; Izquierdo-Ridorsa, A.; Tauler, R. pH-Gradient spectrophotometric data files from flow-injection and continuous flow systems for two- and three-way data analysis. *Chemometrics and Intelligent Laboratory Systems*, **2000**, 50, 263-271.
81. Haaland, D. A.; Easterling, R. G. Application of new Least-Squares methods for the quantitative infrared analysis of multicomponent samples. *Applied Spectroscopy*, **1982**, 36, 665-673.
82. Sakslund, H.; Hammerich, O. Effects of pH, temperature and reaction products on the performance of an immobilized creatininase-creatinase-sarcosine oxidase enzyme system for creatinine determination. *Analytica Chimica Acta*, **1992**, 268, 331-45.
83. Nguyen, V. K; Wolff, C. M.; Seris, J. L.; Schwing, J. P. Immobilized enzyme electrode for creatinine determination in serum. *Analytical Chemistry*, **1991**, 63, 611-14.
84. Vasialides, J. Reaction of alkaline sodium picrate with creatinine: 1. Kinetics and mechanism of formation of the mono-creatinine picric acid complex. *Clinical Chemistry*, **1976**, 22, 1664 – 1671.
85. Weber, J. A.; Van Zanten, A. P. Interferences in current methods for measurements of creatinine. *Clinical Chemistry*, **1991**, 37, 695-700.
86. Butler, A. R. The Jaffé reaction, identification of the coloured species. *Clinica Chimica Acta*, **1975**, 59, 227-232.
87. Bowers, L. D.; Edward, T.W.; Kinetic serum creatinine assays. II. A critical evaluation and review. *Clinical Chemistry*, **1980**, 26, 555-561.
88. Lustgarten, J. A.; Wenk, R. E.; Simple, Rapid, Kinetic method for serum creatinine measurement. *Clinical Chemistry*, **1972**, 18, 1419 – 1422.

- 
89. Pardue, H. L.; Bacon, L. B.; Nevius, L. O.; Skoug, J. W. Kinetic study of the Jaffé reaction for Quantifying creatinine in serum: 1. Alkalinity controlled with NaOH. *Clinical Chemistry*, **1987**, 33, 278 – 285.
90. Taussky, H. A. procedure increasing the specificity of the Jaffé reaction method. *Clinica Chimica acta*, **1956**, 1, 210-224.
91. Haeckel, R. Assay of creatinine in serum, with use of Fuller's earth to remove interferences. *Clinical Chemistry*, **1981**, 27, 179-183.
92. Mitchell, R. J. Improved method for specific determination of creatinine in serum and urine. *Clinical Chemistry*, **1973**, 19, 408-410.
93. Falcó, C. P.; Estellés, L. M.; Cabeza, S. M. Kinetic and chemometric studies of the determination of creatinine using Jaffé reaction: Part 2. Application to human serum samples: kinetic behaviour and chemometric evaluation of the determination. *Analyst*, **1989**, 114, 603 – 607.
94. Schoucri, R. M.; Pouliot, M. Some observations on the kinetics of the Jaffé reaction for creatinine. *Clinical Chemistry*, **1977**, 23, 1527 – 1530.
95. Booksh, K. S.; Kowalski, B. R. Theory of analytical chemistry. *Analytical Chemistry*, **1994**, 66, 782A-791A.
96. Lin, Z. H.; Booksh, K. S.; Burgess, L. W.; Kowalski, B. R. 2nd-order fiber optic heavy-metal sensor employing 2nd-order tensorial calibration. *Analytical Chemistry*, **1994**, 66, 2552-2560.
97. Henshaw, J. M.; Burgess, L. W.; Booksh, K. S.; Kowalski, B. R. Multicomponent determination of chlorinated hydrocarbons using a reaction-based chemical sensor. 1. Multivariate calibration of Fujiwara reaction products. *Analytical Chemistry*, **1994**, 66, 3328-3336.
98. Bezemer, E.; Rutan, S. Evaluation of synthetic liquid chromatography-diode array detection-mass spectrometry data for the determination of enzyme kinetics. *Analytica Chimica Acta*, **2003**, 490, 17-29.
99. Jalali-Heravi, M.; Vosough, M. Characterization and determination of fatty acids in fish oil using gas chromatography-mass spectrometry coupled with chemometric resolution techniques. *Journal of Chromatography A*, **2004**, 1024, 165-176.
100. Harshman, R. A.; Lundy, M. E. PARAFAC - Parallel Factor-Analysis. *Computational and Statistical Data Analysis*, **1994**, 18, 39-72.
101. Smilde, A. K.; Tauler, R.; Henshaw, J. M.; Burgess, L. W.; Kowalski, B. R. Multicomponent determination of chlorinated hydrocarbons using a reaction-based

---

chemical sensor. 3. Medium-rank second-order calibration with restricted Tucker models. *Analytical Chemistry*, **1994**, 66, 3345-3351.

102. Reis, M. M.; Gurden, S. P.; Smilde, A. K.; Ferreira, M. M. C. Calibration and detailed analysis of second-order flow injection analysis data with rank overlap. *Analytica Chimica Acta*, **2000**, 422, 21-36.

103. Saurina, J.; Hernandez-Cassou, S.; Tauler, R.; Izquierdo-Ridorsa, A. Procedure for the quantitative determination of mixtures of nucleic acid components based on multivariate spectrophotometric acid-base titrations. *Analytical Chemistry*, **1999**, 71, 126-134.

104. Linder, M.; Sundberg, R. Second-order calibration: bilinear least squares regression and a simple alternative. *Chemometrics and Intelligent Laboratory Systems*, **1998**, 42, 159-178.

105. Faber, N. M. The price paid for the second-order advantage when using the generalized rank annihilation method (GRAM). *Journal of Chemometrics*, **2001**, 15, 743-748.

106. Damiani, P. C.; Nepote, A. J.; Bearzotti, M.; Olivieri, A. C. A test field for the second order advantage in Bilinear Least Squares and Parallel Factor Analyses: Fluorescence determination of ciprofloxacin in human urine. *Analytical Chemistry*, **2004**, 76, 2798-2806.

107. Messick, N. J.; Kalivas, J.H.; Lang, P. M. Selectivity and related measures for nth order data. *Analytical Chemistry*, **1996**, 68, 1572-1579.

108. Rodriguez-Cuesta, M.; Boque, R.; Rius, F. X. Influence of selectivity and sensitivity parameters on detection limits in multivariate curve resolution of chromatographic second-order data. *Analytica Chimica Acta*, **2003**, 476, 111-122.

109. Boque, R.; Ferre, J.; Faber, N. M.; Rius, F. X. Limit of detection estimator for second-order bilinear calibration. *Analytica Chimica Acta*, **2002**, 451, 313-321.

110. Faber, N. M.; Buydens, L. M. C.; Kateman, G. Generalized rank annihilation method. II: bias and variance in the estimated eigenvalues. *Journal of Chemometrics*, **1994**, 8, 181-203.

111. Linder, M.; Sundberg, R. Precision of prediction in second-order calibration, with focus on bilinear regression methods. *Journal of Chemometrics*, **2002**, 16, 12-27.

112. Fogler, H. S. *Elements of Chemical Reaction Engineering*. Second edition, Prentice Hall, Inc. Edwood Cliffs, New Jersey, USA. **1986**, p. 190.

113. Moore, J. W.; Pearson, R. G. *Kinetics and mechanism*. Third edition, John Wiley & Sons. New York, USA. **1981**, p. 70.

- 
114. Lo, S. G.; Tsai, K. S. Glucose interference in Jaffé creatinine method. Effect of calcium from peritoneal dialysate. *Clinical Chemistry*, **1994**, 40, 2326-2327.
115. Rayens, W. S.; Mitchell, B. C. Two-factor degeneracies and a stabilization of PARAFAC. *Chemometrics and Intelligent Laboratory Systems*, **1997**, 38, 173-181.
116. Falcó, C. P.; Estellés, L. M.; Cabeza, S. M. Kinetic and chemometric studies of the determination of creatinine using Jaffé reaction: Part 1. Kinetics of the reaction: analytical conclusions. *Analyst*, **1989**, 114, 597 – 602.
117. Caldas, L. R. Um pigmento nas águas negras. *Ciência Hoje*, **1990**, 11, 56-67.
118. Duran, N. Violaceína: a descoberta de um antibiótico. *Ciência Hoje*, **1990**, 11, 58-59.
119. Melo, P. D.; Maria, S. S.; Vidal, B. D.; Haun, M.; Duran, N. Violacein cytotoxicity and induction of apoptosis in V79 cells. *In Vitro Cellular & Developmental Biology-Animal*, **2000**, 36, 539-543.
120. Andrighetti-Frohner, C. R.; Antonio, R. V.; Creczynski-Pasa, T. B.; Barardi, C.; Simões, C. Cytotoxicity and potential antiviral evaluation of violacein produced by *Chromobacterium violaceum*. *Memórias do Instituto Oswaldo Cruz*, **2003**, 98, 843-848.
121. Duran, N.; Antonio, R. V.; Haun, M.; Pilli, R. A. Biosynthesis of trypanocide by *Chromobacterium violaceum*. *World Journal of Microbiology & Biotechnology*, **1994**, 10, 686-690.
122. Leon, L. L.; Miranda, C. C.; De Souza, A. O.; Duran, N. Antileishmanial activity of the violacein extracted from *Chromobacterium violaceum*. *Journal of Antimicrobial Chemotherapy*, **2001**, 48, 449-450.
123. Shirata, A.; Tsukamoto, T.; Tasui, H.; Hata, T.; Hayasaka, S.; Kojima, A.; Kato, H. Isolation of bacteria producing bluish-purple pigment and use for dyeing. *Japan Agricultural Research Quarterly*, **2000**, 34, 131-140.
124. Duran, N.; Menck, C. F. M. *Chromobacterium violaceum*: A review of pharmacological and industrial perspectives. *Critical Reviews in Microbiology*, **2001**, 27, 201-222.
125. Ueda, H.; Nakajima, H.; Hori, Y.; Fujita, T.; Nishimura, M.; Goto, T.; Okuhara, M. FR901228, a novel antitumor bicyclic depsipeptide produced by *Chromobacterium violaceum* no-968 .1. Taxonomy, fermentation, isolation, physicochemical and biological properties, and Antitumor-activity. *Journal of Antibiotic*, **1994**, 47, 301-310.
126. Duran, N.; Rettori, D.; Menck, C. F. M. Quem é a *Chromobacterium Violaceum*?. *Biotecnologia Ciência & Desenvolvimento*, **2001**, 20, 38-43.

- 
127. Marshall, J. L.; Rizvi, N.; Kauh, J.; Dahut, W.; Figuera, M.; Kang, M. H.; Figg, W. D.; Wainer, I.; Chaissang, C.; Li, M. Z.; Hawkins, M. J. A phase I trial of depsipeptide (FR901228) in patients with advanced cancer. *Journal of Experimental Therapeutics and Oncology*, **2002**, 2, 325-332.
128. Singh, P. D.; Liu, W. C.; Gougoutas, J. Z.; Malley, M. F.; Porubcan, M. A.; Trejo, W. H.; Wells, J. S.; Sykes, R. B. Aerocavin, a new antibiotic produced by *Chromobacterium violaceum*. *Journal of Antibiotic*, **1988**, 41, 446-453.
129. Smith, A. D.; Hunt, R. J. Solubilization of gold by *Chromobacterium violaceum*. *Journal of Chemical Technology and Biotechnology B- Biotechnology*, **1985**, 35, 110-116.
130. Rodgers, P. B.; Knowles, C. J. Cyanide production and degradation during growth of *Chromobacterium violaceum*. *Journal of General Microbiology*, **1978**, 108, 261-267.
131. Faramarzi, M. A.; Stagars, M.; Pensini, E.; Krebs, W.; Brandl, H. Metal solubilization from metal-containing solid materials by cyanogenic *Chromobacterium violaceum*. *Journal of Biotechnology*, **2004**, 113, 321-326.
132. Steinbuchel, A.; Debzi, E. M.; Marchessault, R. H.; Timm, A. Synthesis and production of poly(3-hydroxyvaleric acid) homopolyester by *Chromobacterium violaceum*. *Applied Microbiology and Biotechnology*, **1993**, 39, 443-449.
133. Steinbuechel, A.; Schmack, G. Large-scale production of poly(3-hydroxyvaleric acid) by fermentation of *Chromobacterium violaceum*, processing, and characterization of the homopolyester. *Journal of Environmental Polymer Degradation*, **1995**, 3, 243-258.
134. de Almeida, D. F.; Hungria, M.; Guimaraes, C. T.; Antonio, R. V.; Almeida, F. C.; de Almeida, L.; de Almeida, R.; Alves-Gomes, J. A.; Andrade, E. M.; Araripe, J.; de Araujo, M.; Astolfi, S.; Azevedo, V.; Baptista, A. J.; Bataus, L.; Batista, J. D. The complete genome sequence of *Chromobacterium violaceum* reveals remarkable and exploitable bacterial adaptability. *Proceedings of the National Academy Sciences U. S. A*, **2003**, 100, 11660-11665.
135. DeMoss, R. D.; Evans, N. R. Incorporations of C14-labeled substrates into violacein. *Journal of Bacteriology*, **1960**, 79, 729-735.
136. Hoshino, T.; Takano, T.; Hori, S.; Ogasawara N. Studies on the biosynthesis of violacein. 2. Biosynthesis of violacein – Origins of hydrogen, nitrogen and oxygen – atoms in the 2- pyrrolidone nucleus. *Agricultural and Biological Chemistry*, **1987**, 51, 2733-2741.
137. Hoshino T.; Kondo, T.; Uchiyama, T.; Ogasawara N. Studies on the biosynthesis of violacein. 1. Biosynthesis of violacein – A novel rearrangement in tryptophan-metabolism with a 1,2 shift of the indole ring. *Agricultural and Biological Chemistry*, **1987**, 51, 965-968.

- 
138. DeMoss, R. D.; Evans, N. R. Physiological aspects of violacein biosynthesis in nonproliferating cells. *Journal of Bacteriology*, **1959**, 78, 583-588.
139. Rettori, D.; Duran, N. Production, extraction and purification of violacein: an antibiotic pigment produced by *Chromobacterium violaceum*. *World Journal Microbiology Biotechnology*, **1998**, 14, 685-688.
140. Hoshino, T.; Yamamoto, M. Studies on the biosynthesis of violacein, part VIII. Conversion from tryptophan precursor into violacein pigment by a cell-free system from *Chromobacterium violaceum*, *Bioscience, Biotechnology, and Biochemistry*, **1997**, 61, 2134-2136.
141. Rettori, D. *Estudos físico-químicos e biológicos da violaceína: Um pigmento produzido pela Chromobacterium Violaceum*. Tese de Doutorado, Universidade Estadual de Campinas, Departamento de Físico-Química. Campinas (SP), Brasil. **2000**.
142. Mendes, A. S.; Carvalho, J. E.; Duarte, M. C. T.; Durán, N.; Bruns, R. Factorial design and response surface optimization of crude violacein for *Chromobacterium violaceum* production. *Biotechnology Letters*, **2001**, 23, 1963-1969.
143. Oliveira, D. A. *Estudo microcalorimétrico em tempo real do metabolismo da bactéria Chromobacterium violaceum*. Tese de Doutorado, Universidade Estadual de Campinas, Departamento de Físico-Química. Campinas (SP), Brasil. **2003**.
144. Ingle, J.D.; Crouch, S.R. *Spectrochemical Analysis*. Prentice Hall, Inc. New Jersey, USA. **1988**, pp. 438-491.
145. Christensen, J.; *Autofluorescence of intact food an exploratory multi way study*. PhD Thesis, The Royal Veterinary and Agricultural University, Quality and Technology Department of Food Science. Denmark . **2005**.
146. Wang, Y. L.; Taylor, D. L. *Fluorescence Microscopy of Living Cells in Culture. Part A: Fluorescent Analogs, Labeling Cells, and Basic Microscopy*. Methods in Cell Biology, Vol. 29. Academic Press, Inc. San Diego (California), USA, **1989**.
147. Reis, M. M. *Testes Imunológicos: manual ilustrado para profissionais da saúde*. Editora SENAC. São Paulo, Brasil. **1999**, p. 25.
148. Timlin, J.A.; Haaland, D.M.; Sincalir M.B.; Aragon, A.D.; Martines, M.J.; Werne-Washburne. Hyperspectral microarray scanning: impact on the accuracy and reliability of gene expression data. *BMC Genomics*, **2005**, 6, 1-11.
149. Zabrizkie, D. W.; Humprey, A. E. Biomass concentration by measuring culture fluorescence. *Applied Environmental Microbiology*, **1978**, 35, 337-343.

- 
150. Horvath, J. J.; Glazier, S. A.; Spangler, C. J. In situ fluorescence cell mass measurements of *Saccharomyces cerevisiae* using cellular tryptophan. *Biotechnology Progress*, **1993**, 9, 666 – 670.
151. Siano, S. A.; Mutharasan, R. NADH and flavin fluorescence responses of starved yeast cultures to substrate additions, *Biotechnology Bioengineering*, **1989**, 34, 660–670.
152. Hisiger, S.; Jolicoeur, M. Plant cell culture monitoring using an in situ multiwavelength fluorescence probe. *Biotechnology Progress*, **2005**, 21, 580-589.
153. Lindemann, C.; Marose, S.; Nielsen H.O.; Scheper, T. 2-Dimensional fluorescence spectroscopy for on-line bioprocess monitoring. *Sensors and Actuators B: Chemistry*, **1998**, 51, 273–277.
154. Marose, S.; Lindemann, C.; Shepr, T. Two dimensional fluorescence spectroscopy: A new tool for on line bioprocess monitoring. *Biotechnology Progress*. **1998**, 14, 63-74.
155. Boehl, D.; Solle, D.; Hitsmann, B.; Sheper, T. Chemometric modeling with two-dimensional fluorescence data for *claviceps purpurea* bioprocess characterization. *Journal of Biotechnology*, **2003**, 105, 179-188.
156. Hagedorn, A.; Levadoux, W.; Groleau, D.; Tartakovsky, B. Evaluation of multiwavelength culture fluorescence for monitoring the aroma compound 4-hydroxy-2(or 5)-ethyl-5(or 2)-methyl-3(2H)-furanone (HEMF) production. *Biotechnology Progress*, **2004**, 20, 361-367.
157. Solle, D.; Geissler, D.; Stark, E.; Scheper, T.; Hitzmann, B. Chemometric modelling based on 2D-fluorescence spectra without a calibration measurement. *Bioinformatics*, **2003**, 19, 173-177.
158. Wolf, G.; Almeida, J. S.; Pinheiro, C.; Correia, V.; Rodrigues, C.; Reis, M.; Crespo, J. G. Two-dimensional fluorometry coupled with artificial neural networks: A novel method for on-line monitoring of complex biological processes. *Biotechnology and Bioengineering*, **2001**, 72, 297-306.
159. Haack, M.B.; Eliasson, A.; Olsson, L. On-line cell mass monitoring of *Saccharomyces cerevisiae* cultivations by multi-wavelength fluorescence. *Journal of Biotechnology*. **2004**, 114, 199-208.
160. Andersen, C. M.; Bro, R. Practical aspects of PARAFAC modelling of fluorescence excitation-emission data. *Journal of Chemometrics*, **2003**, 17, 200-215.
161. Bro, R. Exploratory study of sugar production using fluorescence spectroscopy and multi-way analysis. *Chemometrics and Intelligent Laboratory Systems*, **1999**, 46, 133-147.

- 
162. Walczak, B.; Massart, D. L. Dealing with missing data: Part I. *Chemometrics and Intelligent Laboratory Systems*, **2001**, 58, 15-27.
163. Walczak, B.; Massart, D. L. Dealing with missing data: Part II. *Chemometrics and Intelligent Laboratory Systems*, **2001**, 58, 29-42.
164. Thyngesen, L.G.; Rinnan, A.; Barberg, S.; Moller, J. Stabilizing the PARAFAC decomposition of fluorescence spectra by insertion of zeros outside the data area. *Chemometrics and Intelligent Laboratory Systems*, **2006**, 71, 97-106.
165. Rinnan, A.; Andersen, C. M. Handling of first-order Rayleigh scatter in PARAFAC modelling of fluorescence excitation-emission data. *Chemometrics and Intelligent Laboratory Systems*, **2005**, 76, 91-99.
166. Bro, R.; Sidiropoulos, N. D.; Smilde, A. K. Maximum likelihood fitting using ordinary least squares algorithms. *Journal of Chemometrics*, **2002**, 16, 387-400.
167. <http://www.models.kvl.dk/source/>
168. Bro, R. Kiers, H. A. L. A new efficient method for determining the number of components in PARAFAC models. *Journal of Chemometrics*, **2003**, 17, 274-286.
169. Creczynski-Pasa, T. B.; Antônio, R. V. Energetic metabolism of *Chromobacterium violaceum*. *Genetics and Molecular Research*, **2004**, 3, 162-166.
170. Montgomery, D. C. *Introduction to Statistical Quality Control*. Second Edition, John Wiley & Sons Inc. Singapore. **1991**.
171. Costa, A. F. B.; Epprecht, E. K.; Carpinetti, L. C. R. *Controle Estatístico da Qualidade*. Editora ATLAS S.A. São Paulo, Brasil. **2004**.
172. Hotteling, H. Eisenhart C.; Hastay, M.W.; Wallis, W. A. *Multivariate quality control: illustrated by air testing of sample bombsights*. *Techniques in Statistical Analysis*. MacGraw-Hill. New York, USA. **1947**, p.111-184.
173. Simoglou, A.; Georgieva, P.; Martin, E. B.; Morris, A. J.; Foyo de Azevedo, S. On-line monitoring of a sugar crystallization process. *Computers & Chemical Engineering*, **2005**, 29, 1411-1422.
174. Yin, K. K.; Yang, H.; Cramer, F. On-line monitoring of papermaking processes. *Chemical Engineering Communications*, **2002**, 189, 1242-1261.
175. Wikstrom, C.; Albano, C.; Eriksson, L.; Friden, H.; Johansson, E.; Nordahl, A.; Rannar, S.; Sandberg, M.; Kettaneh, W. N.; Wold, S. Multivariate process and quality monitoring applied to an electrolysis process. Part I. Process supervision with multivariate control charts. *Chemometrics and Intelligent Laboratory Systems*, **1998**, 42, 221-231.

- 
176. Morud, T. E. Multivariate statistical process control: example from the chemical process industry. *Journal of Chemometrics*, **1996**, 10, 669-675.
177. Roberts, S. W. Control charts testes based on geometric moving averages. *Technometrics*, **1959**, 1, 239-250.
178. Hotelling, H. The generalization of student's ratio. *Annals of Mathematical Statistics*, **1931**, 2, 360-378.
179. Jackson, J. E. Quality control methods for several related variables, *Technometrics*, **1959**,1, 359-377.
180. Tracy, N. D.; Young, J. C.; Mason, M. R. Multivariate control charts for individuals observations. *Journal of Quality Technology*, **1992**, 24, 88-95.
181. Sullivan, J. H.; Woodall, W. H. A comparison of multivariate control chart for individual observations. *Journal of Quality Technology*, **1996**, 28, 398-408.
182. Lowry, C.A.; Woodall, W. H.; Champ, C. W.; Rigdon, S. E. Multivariate exponentially weighed moving average control chart. *Technometrics*, **1992**, 34, 46-53.
183. Rigdon, S. E. An integral equation for the in-control average run length of a multivariate exponentially weighted moving average control chart. *Journal of Statistical Computation and Simulation*, **1995**, 52, 351-361.
184. Jackson, J. D. Principal Components and Factors Analysis: Part I- Principal Components. *Journal of Quality Technology*, **1980**, 12, 201-213.
185. Wold, S.; Esbensen, K.; Geladi, P. Principal Components Analysis. *Chemometrics and Intelligent Laboratory Systems*, **1987**, 2, 37-52.
186. Leon, S. J. *Álgebra Linear com Aplicações*. LTC-Livros Técnicos e Científicos Editora S.A. Rio de Janeiro, Brasil. **1999**, p. 212.
187. Wold, H. *Estimation of principal components and related models by iterative least squares*. *Multivariate Analysis*. Academic Press, New York, USA. **1966**, p. 391-420.
188. Jackson, J. E.; Mudholkar, G. S. Control procedures for residuals associated with principal component analysis. *Technometrics*, **1979**, 21, 341-349.
189. Chen, Q.; Wynne, R. J.; Goulding, P.; Snadoz, D. The application of principal component analysis and kernel density estimation to enhance process monitoring. *Control Engineering Practice*, **2000**, 8, 531-543.
190. Martin, E. B.; Morris, A. J. Non parametric confidence bounds for process performance monitoring charts. *Journal of Process Control*, **1996**, 6, 349-358.

- 
191. Jackson, J. E.; Morris, R. H. Application of Multivariate Quality Control to Photographic Processing. *Journal of the American Statistical Association*, **1957**, 52, 186-199.
192. Kresta, J. V.; MacGregor, J. F.; Marlin, T. E. Multivariate Statistical Monitoring of Process Operating Performance. *The Canadian Journal of Chemical Engineering*, **1991**, 69, 35-47.
193. Skagerberg, B.; MacGregor, J. F.; Kiparissides, C. Multivariate data analysis applied to low-density polyethylene reactors. *Chemometrics and Intelligent Laboratory Systems*, **1992**, 14, 341-356.
194. Geladi, P.; Kowalski, B. Partial Least Square Regression: A tutorial. *Analytica Chimica Acta*, **1986**, 185, 1-17.
195. Milidiu, R. L.; Renteria, R. P. DPLS and PPLS: two PLS algorithms for large data sets. *Computational Statistics & Data Analysis*, **2005**, 48, 125-138.
196. Dayal, B. S.; MacGregor, J. F. Improved PLS Algorithms. *Journal of Chemometrics*, **1997**, 11, 73-85.
197. MacGregor, J. F.; Jaeckle, C.; Kiparissides, C.; Koutoudi, M. Process Monitoring and Diagnosis by Multiblock PLS Methods. *AIChE Journal*, **1994**, 40, 826-838.
198. Nomikos, P.; MacGregor, J. F. Monitoring Batch Process Using Multiway Principal Components Analysis. *AIChE Journal*, **1994**, 40, 1361-1375.
199. Nomikos, P.; MacGregor, J. F. Multivariate SPC Charts for Monitoring Batch Process. *Technometrics*, **1995**, 37, 41-59.
200. Kourti, T.; MacGregor, J. F. Process analysis, monitoring and diagnosis, using multivariate projections methods. *Chemometrics and Intelligent Laboratory Systems*, **1995**, 28, 3-21.
201. Westerhuis, J. K.; Kourti, T.; MacGregor, J. F. Analysis of Multiblock and Hierarchical PCA and PLS Models. *Journal of Chemometrics*, **1998**, 12, 301-321.
202. Smilde, A. K.; Westerhuis, J. A.; Sijmen, J. A framework for sequential multiblock component methods. *Journal of Chemometrics*, **2003**, 17, 323-337.
203. Westerhuis, J. A.; Coenegracht, P. M. J. Multivariate modelling of the pharmaceutical two-step process of wet granulation and tableting with multiblock partial least squares. *Journal of Chemometrics*, **1997**, 11, 379-392.
204. Wold, S.; Geladi, P.; Esbensen, K.; Ohman, J. Multi Way Principal Components and PLS analysis. *Journal of Chemometrics*, **1987**, 1, 41-56.

- 
205. Louwerse, D. J.; Smilde, A. K. Multivariate statistical process control of batch process based on the three-way models. *Chemical Engineering Science*, **2000**, 55, 1225-1235.
206. Meng, X.; Morris, A. J.; Martin, E. B. On-line Monitoring of batch processes using a PARAFAC representation. *Journal of chemometrics*, **2003**, 17, 65-81.
207. Hastie T.; Stuetzle, W. Principal Curves. *Journal of the American Statistical Association*, **1989**, 84, 502-516.
208. Dong, D.; McAvoy, T. J. Nonlinear principal component analysis-based on principal curves and neural networks. *Computers in Chemical Engineering*, **1996**, 20, 65-78.
209. Zhang, J.; Martin, E. B.; Morris, A. J. Process monitoring using non-linear statistical techniques. *Chemical Engineering Journal*, **1997**, 67, 181-189.
210. Baffi, G.; Martin, E. B.; Morris, A. J. Non linear projection to latent structures revisited (the neural network PLS algorithm). *Computers and Chemical Engineering*, **1999**, 23, 1293-1307.
211. Baffi, G.; Martin, E. B.; Morris, A. J. Non linear projection to latent structures revisited (the quadratic PLS algorithm). *Computers and Chemical Engineering*, **1999**, 23, 395-411.
212. Rotem, Y.; Wachs, A.; Lewin, D. R. Ethylene compressor monitoring using model-based PCA. *AIChE Journal*, **2000**, 46, 1825-1836.
213. Cheng, C.; Chiu, M. S. Non linear process monitoring using JITL-PCA. *Chemometrics and Intelligent Laboratory Systems*, **2005**, 76, 1-13.
214. Zhang, N. F.; Pollard, J. F. Analysis of autocorrelation in dynamic processes. *Technometrics*, **1994**, 36, 354-368.
215. Escobar, E. M. *Determinação simultânea, em linha, da concentração de soda cáustica e alumina em solução de aluminato de sódio do processo Bayer de produção de óxido de alumínio*. Tese de Mestrado, Universidade Estadual de Campinas, Departamento de Físico-Química. Campinas (SP), Brasil. **2004**.
216. Wold, S. Exponentially weighted moving principal components analysis and projections to latent structures. *Chemometrics and Intelligent Laboratory Systems*, **1994**, 23, 149-161.
217. Wisktrön, C.; Albano, C.; Eriksson, L.; Fridén, H.; Johansson, E.; Nordhall, A.; Rånarar, S.; Sandeberg, M.; Kettneh-Wold, N.; Wold, S. Multivariate process and quality monitoring applied to an electrolysis process. Part 1: Process supervision with multivariate control charts. *Chemometrics and Intelligent Laboratory Systems*, **1998**, 42, 221-231.

- 
218. Ku, Wenfu.; Storer, R. H.; Georgakis, C. Disturbance detection and isolation by dynamic principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, **1995**, 30, 179-196.
219. Negiz, A.; Cinar, A. Statistical Monitoring of Multivariable Dynamic Processes with State Space Models. *AIChE Journal*, **1997**, 43, 2002-2020.
220. Dayal, B. S.; MacGregor, J. F. Recursive exponentially weighted PLS and its applications to adaptative control and predictions. *Journal of Process Control*, **1997**, 7, 169-179.
221. Qin, S. J. Recursive PLS algorithms for adaptative data modelling. *Computers and Chemical Engineering*, **1998**, 22, 503-514.
222. Li, Wi.; Yue, H. H.; Valle-cervantes, S.; Qin, S. J. Recursive PCA for adaptative process monitoring. *Journal of Process Control*, **2000**, 10, 471-486.
223. Box, G. E.; Jenkins, G. M.; Reinsel, G. C. *Time Series Analysis: forecasting and control*. Third Edition, Prentice Hall, Inc. New Jersey, USA. **1994**.
224. Chen, J.; Chien-Mao, L.; Lin, F. R. J.; Lu, M. J. Principal component analysis based control charts with memory effect for process monitoring. *Industrial and Engineering Chemistry Research*, **2001**, 40, 1516-127.
225. Spanos, J. C.; Guo, H. F.; Miller, A.; Parril, L. P. Real-time statistical process control using tool data. *IEE transactions on semiconductor manufacturing*, **1992**, 5, 308-318.
226. Alwan, L. C.; Roberts, H. V. Time series modeling for statistical process control. *Journal of Business & Economic Statistics*, **1998**, 6, 87-95.
227. Bodden, K. M.; Rigdon, S. E. A program for approximating the in control ARL for the MEWMA chart. *Journal of Quality Technology*, **1999**, 31, 120-123.
228. Westerhuis, J. A.; Gurden, S. P.; Smilde, A. K. Generalized contribution plots in multivariate statistical process monitoring. *Chemometrics and Intelligent Laboratory Systems*, **2000**, 51, 95-114.
229. Miller, P.; Swanson, R. E.; Heckler, C. F. Contribution plots: the missing link in multivariate statistical process monitoring. *Applied Mathematics and Computer Science*, **1998**, 8, 775-792.
230. Ljung, G. M.; Box, G. E. P. On a measure of lack of fit in time series models. *Biometrika*, **1978**, 65, 297-303.
231. Shapiro, S. S.; Wilk, M. B. An analysis if variance test for normality. *Biometrika*, **1965**, 52, 591-611.