



Universidade do Minho
Escola de Engenharia

Ana Cristina Araújo Veloso

**Optimização de Estratégias de Alimentação
para a Identificação de Parâmetros de um
Modelo de *E. coli*. Utilização do Modelo em
Monitorização e Controlo**



Universidade do Minho

Escola de Engenharia

Ana Cristina Araújo Veloso

**Optimização de Estratégias de Alimentação
para a Identificação de Parâmetros de um
Modelo de *E. coli*. Utilização do Modelo em
Monitorização e Controlo**

Tese de Doutoramento
Doutoramento em Engenharia Química e Biológica

Trabalho efectuado sob a orientação do
Professor Doutor Eugénio Manuel de Faria Campos Ferreira

Novembro de 2006

Autora: Ana Cristina Araújo Veloso

Endereço electrónico: anaveloso@deb.uminho.pt

Telefone: +351 253604400

Número do Bilhete de Identidade: 10131794

Título tese: Optimização de Estratégias de Alimentação para Identificação de Parâmetros de um Modelo de *E. coli*. Utilização do Modelo em Monitorização e Controlo

Orientador:

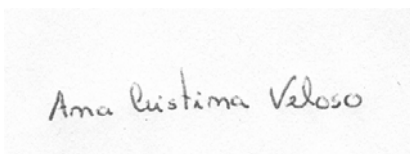
Professor Doutor Eugénio Manuel de Faria Campos Ferreira

Ano de conclusão: 2006

Doutoramento em Engenharia Química e Biológica

É AUTORIZADA A REPRODUÇÃO INTEGRAL DESTA TESE APENAS PARA EFEITOS DE INVESTIGAÇÃO,
MEDIANTE DECLARAÇÃO ESCRITA DO INTERESSADO, QUE A TAL SE COMPROMETE

Universidade do Minho, 27 de Novembro de 2006

A rectangular box containing a handwritten signature in dark ink. The signature reads "Ana Cristina Veloso" in a cursive script.

AGRADECIMENTOS

Chegado este momento é minha intenção expressar os mais sinceros agradecimentos às várias pessoas que, em diferentes momentos, deram o seu contributo para a concretização deste projecto.

Ao Professor Eugénio Ferreira, orientador desta dissertação, gostaria de manifestar o meu reconhecimento pela forma como delineou as linhas deste trabalho e pelos ensinamentos prestados durante a sua realização. Não poderia deixar de reconhecer a confiança que depositou em mim, o seu empenho e a sua disponibilidade bem como o facto de me ter acolhido como membro da sua equipa. Pelo constante incentivo e pela serenidade que sempre me transmitiu, um muito e reconhecido obrigada.

Ao Laboratório de Fermentações do Departamento de Engenharia Biológica da Universidade do Minho, na pessoa do Professor José Teixeira e da Professora Isabel Belo, gostaria de agradecer as condições disponibilizadas.

À Escola Superior Agrária do Instituto Politécnico de Bragança, a cujo corpo docente tenho a honra de pertencer, agradeço o apoio e as facilidades concedidas, sem as quais o caminho percorrido teria sido bem mais difícil.

À Isabel Rocha, pelos momentos partilhados, pelas sugestões, entusiasmos, e interesse que sempre demonstrou. Não poderia deixar de expressar um agradecimento especial pela sua amizade incondicional tanto nos bons como nos maus momentos.

Gostaria ainda de manifestar o meu agradecimento aos colegas do Laboratório, em especial à Nelma Gomes e à Sónia Carneiro, pelo apoio e ajuda sempre manifestados.

Aos meus Amigos, em quem tantas vezes pensei ao longo destes anos, que sempre me apoiaram e entusiasmarão, compreendendo sempre a minha indisponibilidade.

À minha família, em especial à minha irmã Isabel e ao meu Pai, pelo apoio, compreensão e paciência que me têm sempre dedicado. Não posso ainda esquecer o Sr. Elias, a tia Margarida, o Zé e o João pelo carinho e o afecto com que me acolheram e formamos uma nova família e pela sua compreensão pela minha ausência nestes últimos anos.

Uma palavra especial e com muito carinho ao António, que foi a força motriz para a realização deste projecto. Pela cumplicidade, força e entusiasmo que me incutiu desde os primeiros dias, pela companhia que nunca me negou nas horas mais difíceis do trabalho no laboratório e pelos momentos que partilhamos durante a escrita desta tese, que nem sempre foram os mais alegres.

Por fim agradeço ao Programa Prodep III Medida 5 - Acção 5.3 “Formação Avançada de Docentes do Ensino Superior”, pelo apoio financeiro concedido.

A todos, Muito Obrigada.

OPTIMIZAÇÃO DE ESTRATÉGIAS DE ALIMENTAÇÃO PARA IDENTIFICAÇÃO
DE PARÂMETROS DE UM MODELO DE FERMENTAÇÃO DE *E. coli*.
UTILIZAÇÃO DO MODELO EM MONITORIZAÇÃO E CONTROLO

Os principais objectivos desta tese são: o desenho óptimo de experiências para a identificação de coeficientes de rendimento de um modelo não estruturado de um processo de fermentação semi-contínua de *Escherichia coli*; a verificação experimental das trajectórias de alimentação obtidas por simulação; o desenvolvimento de estratégias de monitorização avançada para a estimação em linha de variáveis de estado e parâmetros cinéticos; e por fim o desenvolvimento de uma lei de controlo adaptativo para controlar a taxa específica de crescimento, com base em estratégias de alimentação de substrato com vista à maximização do crescimento e/ou produção.

São apresentadas metodologias para o desenho óptimo de experiências, que visam a optimização da riqueza informativa das mesmas, quantificada por índices relativos à Matriz de Informação de Fisher. Embora, o modelo utilizado para descrever a fermentação semi-contínua de *E. coli* não esteja ainda optimizado em termos cinéticos e de algumas dificuldades encontradas na implementação prática dos resultados obtidos por simulação para o desenho óptimo de experiências, a qualidade da estimativa dos parâmetros, especialmente os do regime respirativo, é promissora. A incerteza das estimativas foi avaliada através de índices relacionados com o modelo de regressão linear múltipla, índices relativos à matriz de Fisher e pelo desenho das correspondentes elipses dos desvios. Os desvios associados a cada coeficiente mostram que ainda não foram encontrados os melhores valores.

Procedeu-se também à investigação do papel do modelo dinâmico geral no desenho de sensores por programação. Foram aplicados três observadores - observador estendido de Kalman, observador assintótico e observador por intervalo - para estimar a concentração de biomassa, tendo sido avaliado e comparado o seu desempenho bem como a sua flexibilidade. Os observadores estudados mostraram-se robustos, apresentando comportamentos complementares. O observador assintótico apresenta, em geral, um melhor desempenho que o observador estendido de Kalman. Os observadores por intervalo apresentam vantagens em termos de implementação prática, apresentando-se bastante promissores embora a sua validação experimental seja necessária.

É apresentada uma lei de controlo adaptativo com modelo de referência que se traduz num controlo por antecipação/retroacção cuja acção de retroacção é do tipo PI, para controlar a taxa específica de crescimento. A robustez do algoritmo de controlo foi estudada por simulação

numérica gerando dados “*pseudo* reais”, por aplicação de um ruído branco às variáveis medidas em linha, por alteração do valor de referência, por alteração do valor da concentração da glucose na alimentação e variando os valores nominais dos parâmetros do modelo. O estudo realizado permite concluir que a resposta do controlador é em geral satisfatória, sendo capaz de manter o valor da taxa específica de crescimento na vizinhança do valor de referência pretendido e inferior a um valor que conduz à formação de acetato, revestindo-se este facto de grande importância numa situação real, em especial, numa fermentação cujo objectivo seja a produção, nomeadamente de proteínas recombinadas. Foram ainda, analisados diferentes métodos de sintonização dos parâmetros do controlador, podendo concluir-se que, em geral, o método de sintonização automática com recurso à regra de adaptação dos parâmetros em função do erro relativo do controlador foi o que apresentou um melhor desempenho global. Este mecanismo de sintonização automática demonstrou capacidade para melhorar o desempenho do controlador ajustando continuamente os seus parâmetros.

Palavras-chave: Identificação de Sistemas; Planificação de Experiências; Matriz de Informação de Fisher; Ferramentas Estocásticas de Optimização; Observadores de Estado; Sensores por Programação; Controlo Adaptativo; *Escherichia coli*; Fermentação Semi-contínua.

ABSTRACT

FEED STRATEGIES OPTIMIZATION FOR PARAMETER IDENTIFICATION IN A *E. coli* FERMENTATION MODEL FOR MONITORING AND CONTROL

The main objectives of this thesis are: the optimal experiment design for yield coefficients estimation in an unstructured growth model for *Escherichia coli* fed-batch fermentation; the experimental validation of the simulated feed trajectories; the development of advanced monitoring strategies for the on-line estimation of state variables and kinetic parameters; and at last the development of an adaptive control law, based on optimal substrate feed strategies in order to increase the growth and/or the production.

Methodologies for the optimal experimental design are presented, in order to optimise the richness of data coming out from experiments, quantified by indexes based on the Fisher Information Matrix. Although the model used to describe the *E. coli* fed-batch fermentation is not optimised from the kinetic properties point of view and the fact that some difficulties were encountered in practical implementation of the simulated results obtained with the optimal experimental design, the estimated parameter quality, especially for the oxidative regimen, is promising. The estimation uncertainty was evaluated by means of indexes related with multiple linear regression model, indexes related to the Fisher matrix as well as by the construction of the related deviation ellipses. The deviations associated to each coefficient show that the best values were not yet found.

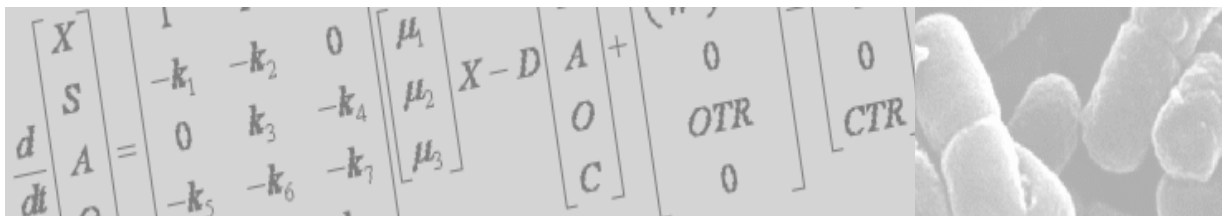
The role of the general dynamical model was also investigated in which concerns the design of state observers, also called software sensors. The performance of three observer classes was compared: Kalman extended observer, asymptotic observer and interval observer. The studied observers showed good performance and robustness, being complementary of each other. Asymptotic observers showed, in general, a better performance than the Kalman extended observer. Interval observers presented advantages concerning practical implementation, showing a promising behaviour although experimental validation is needed.

A model reference adaptive control law is presented and can be interpreted as a PI like feedforward/feedback controller, for specific growth rate control. Algorithm robustness was studied using "pseudo real" data obtained by numerical simulation, by applying a white noise to the on-line measured variables, by modifying the set-point value, by changing the glucose concentration value of the feed rate and varying the nominal model parameter value. The study made allowed to conclude that the controller response is, generally, satisfactory being able to keep the specific growth rate value in the proximity of the desired set-point and lower than the value that permits

acetate formation, which is of major importance namely for real cases, specially, in a fermentation which objective was the production of recombinant proteins. Different tuning devices for controller parameters were analysed being the better performance achieved by the automatic tuning method with an adaptation rate as a function of the controller relative error. This automatic tuning mechanism was able to improve the controller performance adjusting continuously its parameters.

Key words: System identification; Experimental Design; Fisher Information Matrix; Stochastic Optimization Methods; State Observers; Software Sensors; Adaptive Control; *Escherichia coli*; Fed-batch fermentation.

ÍNDICE



ÍNDICE

CAPÍTULO 1

1.	INTRODUÇÃO	3
1.1.	CONTEXTO E MOTIVAÇÃO	3
1.2.	OBJECTIVOS	7
1.3.	ORGANIZAÇÃO DA TESE	7
1.4.	BIBLIOGRAFIA	8

CAPÍTULO 2

2.	MICROORGANISMO, PROCESSO E METODOLOGIAS	13
2.1.	<i>Escherichia coli</i> : CARACTERÍSTICAS FISIOLÓGICAS E BIOQUÍMICAS	13
2.1.1.	Características fisiológicas	14
2.1.2.	Características bioquímicas	15
2.2.	PROCESSO FERMENTATIVO	19
2.3.	PROTEÍNA DE FLUORESCÊNCIA VERDE	25
2.4.	DESCRIÇÃO EXPERIMENTAL	28
2.4.1.	Fermentação	28
2.4.1.1.	Microrganismo	29
2.4.1.2.	Composição do meio de cultura	29
2.4.1.3.	Fermentador e Unidade de Medida e Controlo Digital	31
2.4.1.4.	Condições experimentais	33
2.4.2.	Métodos de análise em linha	34
2.4.3.	Análises em diferido	35
2.4.3.1.	Determinação da concentração de biomassa	35
2.4.3.2.	Determinação da concentração de acetato	36
2.4.3.3.	Determinação da concentração de glucose	37
2.4.4.	Sistema informático	38
2.5.	BIBLIOGRAFIA	38

CAPÍTULO 3

3.	IDENTIFICAÇÃO DO MODELO	47
3.1.	MODELAÇÃO MATEMÁTICA	48
3.1.1.	Classes de modelos matemáticos	48
3.1.2.	Desenvolvimento do modelo	49

3.2. IDENTIFICAÇÃO DE PARÂMETROS: PROBLEMAS E METODOLOGIAS	52
3.2.1. Identificabilidade estrutural.....	53
3.2.2. Identificação numérica	54
3.2.2.1. Planificação de experiências.....	57
3.2.2.2. Funções de sensibilidade	60
3.2.2.3. Avaliação da precisão da estimativa dos parâmetros identificados.....	60
3.3. MODELO DINÂMICO DE REACTORES BIOLÓGICOS	63
3.3.1. Identificação dos coeficientes de rendimento.....	66
3.4. CASO DE ESTUDO: CRESCIMENTO DE UMA ESTIRPE RECOMBINANTE DE <i>Escherichia coli</i> POR OPERAÇÃO SEMI-CONTÍNUA	71
3.4.1. Modelo matemático	71
3.4.2. Planificação de experiências	77
3.4.2.1. Análise de sensibilidade do modelo aos seus parâmetros	78
3.4.2.2. Projecto dos regressores para os modelos parciais a duas reacções.....	84
3.4.2.2.1 Optimização do perfil do caudal de alimentação.....	88
3.4.2.2.2 Diagnóstico da regressão linear múltipla do projecto de regressores	95
3.4.2.2.3 Estabilidade matemática do problema de recuperação dos k 's através dos θ 's	103
3.4.2.2.4 Validação da identificação	104
3.4.3. Identificação dos parâmetros do modelo: implementação prática.....	107
3.5. SÍNTESE.....	114
3.6. BIBLIOGRAFIA	115

CAPÍTULO 4

4. MONITORIZAÇÃO DO PROCESSO	123
4.1. SENSORES POR PROGRAMAÇÃO	124
4.1.1. Observadores de estado	126
4.1.1.1. Observador estendido de Kalman	127
4.1.1.2. Observadores Assimptóticos	129
4.1.1.3. Observadores por Intervalo.....	131
4.1.2. Estimadores de cinética	132
4.2. RESULTADOS	134
4.2.1. Observador estendido de Kalman.....	135
4.2.2. Observadores Assimptóticos	141
4.2.3. Observadores por Intervalo	147
4.3. SÍNTESE.....	155
4.4. BIBLIOGRAFIA	157

CAPÍTULO 5

5.	CONTROLO DO PROCESSO	165
5.1.	ESTRATÉGIAS DE CONTROLO	166
5.1.1.	Controlo por retroacção (<i>feedback</i>)	167
5.1.2.	Controlo por antecipação (<i>feedforward</i>)	168
5.1.3.	Controlo por antecipação/retroacção (<i>feedforward/feedback</i>)	168
5.1.4.	Tipos de controladores	169
5.2.	MODELO DO PROCESSO	170
5.3.	CONTROLO DA TAXA ESPECÍFICA DE CRESCIMENTO	171
5.3.1.	Lei de controlo	173
5.3.2.	Sintonização do controlador	177
5.4.	RESULTADOS	182
5.5.	SÍNTESE	195
5.6.	BIBLIOGRAFIA	196

CAPÍTULO 6

6.	CONCLUSÕES GERAIS E SUGESTÕES PARA TRABALHO FUTURO	203
6.1.	CONCLUSÕES GERAIS	203
6.2.	SUGESTÕES PARA TRABALHO FUTURO	205

ANEXO

A.	CURVAS DE CALIBRAÇÃO	209
A.1.	ESPECTRÓMETRO DE MASSA	209
A.2.	BIOMASSA	211
A.3.	ACETATO	212
A.4.	GLUCOSE	213
A.4.1.	Concentração de glucose determinada por HPLC	213
A.4.2.	Concentração de glucose determinada por <i>kit</i> enzimático	214

LISTA DE TABELAS

CAPÍTULO 2

Tabela 2.1 Resumo das vantagens e desvantagens dos tipos de operação do biorreactor (adaptado de Versyck, 2000)	21
Tabela 2.2 Estratégias de alimentação aplicadas a fermentações semi-contínuas (adaptado de Lee, 1996)	24
Tabela 2.3 Propriedades das proteínas fluorescentes (adaptado de Patterson <i>et al.</i> , 2001)	27
Tabela 2.4 Composição do meio de cultura sintético M9 modificado	30
Tabela 2.5 Composição da solução de Traços de minerais	30
Tabela 2.6 Composição da solução de Vitaminas	31
Tabela 2.7 Descrição dos principais componentes do fermentador	32
Tabela 2.8 Valores de referência para controlo das variáveis ambientais	34
Tabela 2.9 Composição das misturas utilizadas na calibração do MS	35

CAPÍTULO 3

Tabela 3.1 Classes de modelos matemáticos	48
Tabela 3.2 Critérios de optimalidade usados no desenho óptimo de experiências	58
Tabela 3.3 Valores pré-estabelecidos dos coeficientes de rendimento e dos parâmetros cinéticos	78
Tabela 3.4 Identificabilidade dos parâmetros	81
Tabela 3.5 Ângulo entre os vectores de sensibilidade de dois parâmetros: avaliação da colinearidade para uma perturbação de +10% em cada parâmetro	82
Tabela 3.6 Ângulo entre os vectores de sensibilidade de dois parâmetros: avaliação da colinearidade para uma perturbação de -10% em cada parâmetro	83
Tabela 3.7 Partições induzidas nas variáveis de estado (ξ) e de entrada (U) para os modelos parciais RF e R	85
Tabela 3.8 Partições induzidas de K e correspondente matriz H para os modelos parciais RF e R	85
Tabela 3.9 Expressões de cálculo dos coeficientes de rendimento	88
Tabela 3.10 Condições iniciais utilizadas nas simulações e optimização	89
Tabela 3.11 Valores do determinante (\det_{MIF}) e do número de condicionamento (Λ_{MIF}) da MIF obtidos na optimização dos critérios D e E modificado	92
Tabela 3.12 Intervalos de rejeição e de não rejeição de auto-correlação de dados	97
Tabela 3.13 Diagnóstico do modelo $MRLM$ para a partição XA e modelo parcial RF	101

Tabela 3.14 Diagnóstico do modelo <i>MRLM</i> para a partição <i>XA</i> e modelo parcial R	101
Tabela 3.15 Diagnóstico do modelo <i>MRLM</i> para a partição <i>SO</i> e modelo parcial RF	102
Tabela 3.16 Diagnóstico do modelo <i>MRLM</i> para a partição <i>SO</i> e modelo parcial R	102
Tabela 3.17 Valores do número de condição e do determinante da forma escalonada para as duas partições estudadas.....	104
Tabela 3.18 Valores calculados dos coeficientes de rendimento para a partição <i>XA</i> , regime RF, critério <i>D</i>	105
Tabela 3.19 Valores calculados dos coeficientes de rendimento para a partição <i>XA</i> , regime R, critério <i>D</i>	106
Tabela 3.20 Valores de θ e de k e, respectivas incertezas, calculados por <i>MRLM</i> para as experiências Exp_2 e Exp_3	109
Tabela 3.21 Parâmetros associados à Matriz de Fisher usados no estudo da qualidade dos valores de θ	109
 CAPÍTULO 4	
Tabela 4.1 Estudo da observabilidade do modelo para diferentes combinações de variáveis medidas e estimadas usando o OEK	136
Tabela 4.2 Índice de desempenho para o OEK para os casos 1 e 4 da Tabela 4.1.....	138
Tabela 4.3 Índice de desempenho para o OEK para as experiências Exp_4 e Exp_5.....	139
Tabela 4.4 Estudo da observabilidade do modelo para diferentes combinações de variáveis medidas e estimadas usando o OA.....	141
Tabela 4.5 Índices de desempenho para o OA para os casos 1 e 4 da Tabela 4.4.....	144
Tabela 4.6 Índices de desempenho para o OA para as experiências Exp_4 e Exp_5.....	145
 CAPÍTULO 5	
Tabela 5.1 Valores iniciais das variáveis de estado	182
Tabela 5.2 Valores iniciais dos parâmetros γ e β utilizados nas diferentes estratégias de sintonização estudadas	183
 ANEXO A	
Tabela A.1 Valores utilizados na curva de calibração do Azoto.....	209
Tabela A.2 Valores utilizados na curva de calibração do Oxigénio	210
Tabela A.3 Valores utilizados na curva de calibração do Dióxido de Carbono	210
Tabela A.4 Valores utilizados no cálculo da concentração em peso seco da suspensão celular inicial	211
Tabela A.5 Valores da concentração em peso seco das outras suspensões celulares utilizadas para a curva de calibração da biomassa	212
Tabela A.6 Valores utilizados na curva de calibração do Acetato	213

Tabela A.7 Valores utilizados na curva de calibração da Glucose por HPLC	214
Tabela A.8 Valores utilizados na curva de calibração da Glucose por <i>kit</i> enzimático.....	215

LISTA DE FIGURAS

CAPÍTULO 2

- Figura 2.1 Fotografia, obtida por microscopia electrónica, e esquema típico da célula de *E. coli* 14
- Figura 2.2 Metabolismo da glucose e do acetato na *E. coli* (adaptado de Carlson e Srienc, 2004 e de Phue *et al.*, 2005) 18
- Figura 2.3 Representação esquemática da capacidade oxidativa limitada da *E. coli* (adaptado de Cannizzaro *et al.*, 2004) 19
- Figura 2.4 (A) Estrutura terciária da GFP. (B) Tripéptido que constitui o cromóforo da GFP 26
- Figura 2.5 Fotografia da instalação experimental 33

CAPÍTULO 3

- Figura 3.1 Etapas do desenvolvimento do modelo matemático em bioprocessos 50
- Figura 3.2 Efeito das propriedades da *MIF* no tamanho e forma da região de confiança dos parâmetros (retirado de Pauw e Vanrolleghem (2006)) 57
- Figura 3.3 Representação do modelo cinético da *E. coli* (adaptado de Rocha (2003)) 76
- Figura 3.4 Caudal de alimentação usado no estudo das sensibilidades 79
- Figura 3.5 *Scores* de cada parâmetro na dimensão 1 do modelo de componentes principais para uma perturbação de: (A) +10% e (B) -10% 84
- Figura 3.6 Representação gráfica do conjunto de soluções Pareto-ótimo 91
- Figura 3.7 Perfis do caudal de alimentação e dos regressores obtidos para a partição XA_{RF} na optimização do \det_{MIF} (coluna da esquerda), do Λ_{MIF} (coluna do meio) e do $\det_{MIF} + \Lambda_{MIF}$ (coluna da direita) 93
- Figura 3.8 Perfis do caudal de alimentação e dos regressores obtidos para a partição XA_R na optimização do \det_{MIF} (coluna da esquerda), do Λ_{MIF} (coluna do meio) e do $\det_{MIF} + \Lambda_{MIF}$ (coluna da direita) 93
- Figura 3.9 Perfis do caudal de alimentação e dos regressores obtidos para a partição SO_{RF} na optimização do \det_{MIF} (coluna da esquerda), do Λ_{MIF} (coluna do meio) e do $\det_{MIF} + \Lambda_{MIF}$ (coluna da direita) 94
- Figura 3.10 Perfis do caudal de alimentação e dos regressores obtidos para a partição SO_R na optimização do \det_{MIF} (coluna da esquerda), do Λ_{MIF} (coluna do meio) e do $\det_{MIF} + \Lambda_{MIF}$ (coluna da direita) 94

Figura 3.11 Elipses dos desvios dos parâmetros para XA_{RF} (a); XA_R (b); SO_{RF} (c) e SO_R (d): 1- otimização do \det_{MIF} ; 2- otimização do Λ_{MIF} e 3- otimização $\det_{MIF} + \Lambda_{MIF}$	95
Figura 3.12 Estudo da homocedasticidade das variáveis aleatórias residuais (ϵ_i) para a partição XA , modelo parcial RF na otimização do \det_{MIF} (coluna da esquerda), do Λ_{MIF} (coluna do meio) e do $\det_{MIF} + \Lambda_{MIF}$ (coluna da direita): (a) Equação 3.79; (b) Equação 3.80; (c) Equação 3.81	99
Figura 3.13 Estudo da homocedasticidade das variáveis aleatórias residuais (ϵ_i) para a partição XA , modelo parcial R na otimização do \det_{MIF} (coluna da esquerda), do Λ_{MIF} (coluna do meio) e do $\det_{MIF} + \Lambda_{MIF}$ (coluna da direita): (a) Equação 3.79; (b) Equação 3.80; (c) Equação 3.81	99
Figura 3.14 Estudo da homocedasticidade das variáveis aleatórias residuais (ϵ_i) para a partição SO , modelo parcial RF na otimização do \det_{MIF} (coluna da esquerda), do Λ_{MIF} (coluna do meio) e do $\det_{MIF} + \Lambda_{MIF}$ (coluna da direita): (a) Equação 3.79; (b) Equação 3.80; (c) Equação 3.81	100
Figura 3.15 Estudo da homocedasticidade das variáveis aleatórias residuais (ϵ_i) para a partição SO , modelo parcial R na otimização do \det_{MIF} (coluna da esquerda), do Λ_{MIF} (coluna do meio) e do $\det_{MIF} + \Lambda_{MIF}$ (coluna da direita): (a) Equação 3.79; (b) Equação 3.80; (c) Equação 3.81	100
Figura 3.16 Experiência de identificação com caudal de alimentação obtido, para a partição XA_R e otimização do critério D (pontos: valores "pseudo experimentais", linha: valores estimados)	106
Figura 3.17 Funções de sensibilidade relativas dos y 's em relação aos k 's para XA_R	107
Figura 3.18 Elipses dos desvios para as experiências Exp_2 e Exp_3: 1- regime respirativo da Exp_3; 2- regime respirativo da Exp_2 e 3- regime respiro-fermentativo da Exp_3	110
Figura 3.19 Funções de sensibilidade relativas dos y 's em relação aos k 's, para a Exp_3, regime respirativo	111
Figura 3.20 Validação experimental para a Exp_3 (pontos: valores experimentais, linha: valores estimados).....	113
 CAPÍTULO 4	
Figura 4.1 Sensores por programação (adaptado de Farza <i>et al.</i> (1998))	124
Figura 4.2 Desempenho do OEK na estimativa da concentração de biomassa, de glucose e de acetato para o caso 4 estudado por simulação. Os gráficos d), e) e f) representam as variáveis medidas em linha	138
Figura 4.3 Observação de estado numa fermentação semi-contínua de <i>E. coli</i> (Exp_4): dados medidos em linha com um intervalo de amostragem de 18 segundos	139

Figura 4.4	Observação de estado numa fermentação semi-contínua de <i>E. coli</i> (Exp_4): variável estimada (pontos: valores medidos em diferido; linha: valores estimados; RF: regime respiro-fermentativo; R: regime respirativo).....	140
Figura 4.5	Desempenho do OA na estimação da concentração de biomassa, de glucose e de acetato para o caso 4 estudado por simulação. Os gráficos d), e) e f) representam as variáveis medidas em linha	144
Figura 4.6	Observação de estado numa fermentação semi-contínua de <i>E. coli</i> (Exp_5): dados medidos em linha com um intervalo de amostragem de 3 minutos	145
Figura 4.7	Observação de estado numa fermentação semi-contínua de <i>E. coli</i> (Exp_5): a) variável estimada: concentração de biomassa; b) variável estimada: <i>CTR</i> (pontos: valores medidos em diferido no caso da biomassa ou valores medidos em linha para o <i>CTR</i> ; linha: valores estimados; RF: regime respiro-fermentativo; R: regime respirativo)	146
Figura 4.8	Observador por intervalo simulado com incerteza associada à concentração de glucose no caudal de alimentação. (a) observador por intervalo simples; (b) observador por intervalo com ganhos; (c) observador por intervalo com ganhos através do <i>OTR</i>	156
 CAPÍTULO 5		
Figura 5.1	Esquema de controlo por retroacção	167
Figura 5.2	Esquema de controlo por antecipação	168
Figura 5.3	Esquema de controlo por antecipação/retroacção	169
Figura 5.4	Resultados de simulação do controlo da taxa específica de crescimento com parâmetros de sintonização constantes. (a) Variável controlada: taxa específica de crescimento; (b) Parâmetros de sintonização; (c) Concentração de biomassa; (d) Erro médio do controlador; (e) Caudal de alimentação, $F_{e,s}$; (f) Erro de oscilação do controlador. Linha a tracejado corresponde aos valores estimados e linha a cheio aos valores simulados	184
Figura 5.5	Resultados de simulação do controlo da taxa específica de crescimento com parâmetros de sintonização variáveis: regra de adaptação em função do erro relativo do controlador. (a) Variável controlada: taxa específica de crescimento; (b) Parâmetros de sintonização; (c) Concentração de biomassa; (d) Erro médio do controlador; (e) Caudal de alimentação, $F_{e,s}$; (f) Erro de oscilação do controlador. Linha a tracejado corresponde aos valores estimados e linha a cheio aos valores simulados	185

Figura 5.6 Resultados de simulação do controlo da taxa específica de crescimento com parâmetros de sintonização variáveis: regra de MIT recorrendo à sensibilidade do erro. (a) Variável controlada: taxa específica de crescimento; (b) Parâmetros de sintonização; (c) Concentração de biomassa; (d) Erro médio do controlador; (e) Caudal de alimentação, $F_{e,s}$; (f) Erro de oscilação do controlador. Linha a tracejado corresponde aos valores estimados e linha a cheio aos valores simulados	185
Figura 5.7 Resultados de simulação do controlo da taxa específica de crescimento com parâmetros de sintonização variáveis: regra de adaptação em função do modelo de referência. (a) Variável controlada: taxa específica de crescimento; (b) Parâmetros de sintonização; (c) Concentração de biomassa; (d) Erro médio do controlador; (e) Caudal de alimentação, $F_{e,s}$; (f) Erro de oscilação do controlador. Linha a tracejado corresponde aos valores estimados e linha a cheio aos valores simulados	186
Figura 5.8 Resultados de simulação do controlo da taxa específica de crescimento quando alterado o valor de referência de 0,1 para 0,05 h ⁻¹ . (a) Parâmetros de sintonização constantes; (b) Parâmetros de sintonização variáveis de acordo com a regra de adaptação em função do erro relativo; (c) Parâmetros de sintonização variáveis de acordo com a regra de MIT recorrendo à sensibilidade do erro; (d) Parâmetros de sintonização variáveis de acordo com a regra de adaptação em função do modelo de referência	187
Figura 5.9 Resultados de simulação do controlo da taxa específica de crescimento quando o valor da concentração de glucose no caudal de alimentação diminui 20%, após 5 horas de fermentação. (a) Parâmetros de sintonização constantes; Parâmetros de sintonização variáveis: (b) regra de adaptação em função do erro relativo; (c) regra de MIT recorrendo à sensibilidade do erro; (d) regra de adaptação em função do modelo de referência.....	188
Figura 5.10 Variação com o tempo, por aplicação de ruído ±10%, imposta aos coeficientes de rendimento durante a simulação para avaliar a robustez do algoritmo.....	189
Figura 5.11 Variação com o tempo, por aplicação de ruído ±10%, imposta aos coeficientes de rendimento durante a simulação para avaliar a robustez do algoritmo.....	189
Figura 5.12 Resultados de simulação do controlo da taxa específica de crescimento quando os coeficientes de rendimento variam com o tempo com um ruído imposto de ±10%. (a) Parâmetros de sintonização constantes; Parâmetros de sintonização variáveis: (b) regra de adaptação em função do erro relativo; (c) regra de MIT recorrendo à sensibilidade do erro; (d) regra de adaptação em função do modelo de referência.....	190

Figura 5.13 Resultados de simulação do controlo da taxa específica de crescimento quando os coeficientes de rendimento e parâmetros cinéticos variam com o tempo com um ruído imposto de $\pm 10\%$. (a) Parâmetros de sintonização constantes; Parâmetros de sintonização variáveis (b) regra de adaptação em função do erro relativo; (c) regra de MIT recorrendo à sensibilidade do erro; (d) regra de adaptação em função do modelo de referência.....	190
Figura 5.14 Resultados de simulação do controlo da taxa específica de crescimento quando os coeficientes de rendimento sofrem uma variação de $\pm 20\%$ dos seus valores nominais. (a) Parâmetros de sintonização constantes; Parâmetros de sintonização variáveis: (b) regra de adaptação em função do erro relativo; (c) regra de MIT recorrendo à sensibilidade do erro; (d) regra de adaptação em função do modelo de referência.....	191
Figura 5.15 Comparação do desempenho de vários algoritmos de controlo da taxa específica de crescimento. Parâmetros de sintonização constantes: (a) algoritmo de controlo dado pela Equação 5.32; (b) algoritmo de controlo dado pela Equação 5.16. Parâmetros de sintonização variáveis: (c) regra de adaptação em função do erro relativo; (d) regra de MIT recorrendo à sensibilidade do erro; (e) regra de adaptação em função do modelo de referência. (f) controlo adaptativo com modelo de referência de primeira ordem; (g) controlo por antecipação; (h) controlo PI clássico.....	193

ANEXO A

Figura A.1 Curva de calibração para o Azoto.....	209
Figura A.2 Curva de calibração para o Oxigénio.....	210
Figura A.3 Curva de calibração para o Dióxido de Carbono.....	211
Figura A.4 Curva de calibração da Biomassa.....	212
Figura A.5 Curva de calibração do Acetato.....	213
Figura A.6 Curva de calibração da Glucose por HPLC.....	214
Figura A.7 Curva de calibração da Glucose por <i>kit</i> enzimático.....	215

LISTA DE SÍMBOLOS

LETRAS LATINAS MAIÚSCULAS

A	Concentração de acetato	(MM^{-1})
C	Concentração de dióxido de carbono dissolvido	(MM^{-1})
D	Matriz da taxa de diluição; taxa de diluição	(T^{-1})
E(t)	Erro médio	
E_m	Coefficiente de extinção	
F	Valor crítico da distribuição <i>F</i> de <i>Fisher-Snedecor</i>	
F	Vector de caudais mássicos de entradas líquidas no reactor (por unidade de peso do líquido dentro do reactor);	$(MM^{-1}T^{-1})$
F_a	Caudal mássico da solução de ácido adicionada ao reactor	(MT^{-1})
F_{amostra}	Caudal mássico retirado do reactor devido à amostragem	(MT^{-1})
F_b	Caudal mássico da solução de base adicionada ao reactor	(MT^{-1})
F_{e,i}	Caudal mássico de alimentação do reactor na forma líquida, para o componente <i>i</i>	(MT^{-1})
F_{e,S}	Caudal mássico de alimentação	(MT^{-1})
F_{evp}	Caudal mássico de evaporação do reactor	(MT^{-1})
F_{gás}	Caudal mássico retirado do reactor devido a transferências gasosas	(MT^{-1})
GL	Graus de liberdade	
H	Matriz função dos coeficientes de rendimento de dimensão $(n-p) \times p$	
IC_{aprox}(θ_i)	Intervalo de confiança aproximado para o parâmetro θ _i	
J	Matriz Jacobiano	
J	Função perda	
J_i	Função escalar de identificação	
J_{N,i}	Função de identificabilidade do parâmetro <i>i</i>	
K	Matriz de coeficientes de rendimento ou estequiométricos; Vector do conjunto dos parâmetros possíveis do modelo	
K_A	Constante de afinidade pelo acetato	(MM^{-1})
K_{i,A}	Constante de inibição do acetato no seu consumo	(MM^{-1})
K_{i,S}	Constante de inibição do acetato no consumo de glucose	(MM^{-1})
K_S	Constante de <i>Monod</i>	(MM^{-1})
K_a⁺	Inversa generalizada ou pseudo-inversa da matriz K_a	
K₁⁻¹	Pseudo inversa da matriz K₁ , considerando que esta apresenta característica completa	
L	Matriz quadrada; matriz de dimensão $q \times n$	

\mathcal{M}	Estrutura do processo ou do modelo	
N	Número total de pontos experimentais ou de parâmetros	
O	Concentração de oxigénio dissolvido	(MM^{-1})
$O(k)$	Erro de oscilação	
P	Concentração de produto	(MM^{-1})
P	Matriz quadrada dos coeficientes de ponderação	
P_{ij}	Elemento j do componente principal i	
Q	Vector de caudais mássicos na forma gasosa (por unidade de peso do líquido dentro do reactor)	$(\text{MM}^{-1}\text{T}^{-1})$
Q_e	Caudal volumétrico de entrada do reactor na forma gasosa	$(\text{L}^3\text{M}^{-1})$
Q_s	Caudal volumétrico de saída do reactor na forma gasosa	$(\text{L}^3\text{M}^{-1})$
R	Coefficiente de correlação amostral ou R de Pearson	
R	Matriz simétrica quadrada	
R	Regime respirativo	
R	Matriz simétrica de ordem n gerada pela equação de <i>Riccati</i>	
R^2	Coefficiente de determinação	
RF	Regime respiro-fermentativo	
S	Concentração de glucose	(MM^{-1})
S_e	Concentração de glucose na corrente de alimentação	(MM^{-1})
SO	Partição SO	
SP	Matriz de sensibilidade	
SP_i	Vector de sensibilidade do parâmetro i	
SP_j	Vector de sensibilidade do parâmetro j	
T	Temperatura	(θ)
T	Período de amostragem	(T)
U	Vector que representa o balanço entre as entradas e/ou saídas de substratos e produtos gasosos	$(\text{MM}^{-1}\text{T}^{-1})$
VIF	Factor de inflação da variância	
W	Massa de líquido dentro do reactor	(M)
X	Concentração da biomassa	(MM^{-1})
XA	Partição XA	
X_{cultura}	Concentração em peso seco da suspensão celular inicial	(MM^{-1})
X_m	<i>Pseudo</i> medida para a biomassa	(MM^{-1})
$X_{\text{suspensões}}$	Concentração em peso seco das suspensões celulares	(MM^{-1})
Y_Q	Rendimento quântico	
$Y_{\text{sx}}^{\text{ox}}$	Rendimento oxidativo em glucose	
$Y_{\text{sx}}^{\text{red}}$	Rendimento redutivo em glucose	

Y_{Ax}^{ox}	Rendimento oxidativo em acetato	
Z	Matriz resultante de uma transformação linear de estado	(MM^{-1})

LETRAS LATINAS MINÚSCULAS

a		
b	Vector que representa o termo independente de um SEL	
c_a	Representa a constante que relaciona a quantidade de ácido necessária para equilibrar o ph	(MM^{-1})
c_b	Representa a constante que relaciona a quantidade de base necessária para equilibrar o pH	(MM^{-1})
c_e	Representa a velocidade de evaporação	(MM^{-1})
c_{ij}	Coeficiente de correlação aproximado entre dois parâmetros estimados i e j	
d	Taxa de diluição	(T^{-1})
d.i.	Diâmetro interno	(L)
d_{DW}	Estatística de Durbin-Watson	
dif_{ξ}	Diferença quadrática entre os dados experimentais e os estimados para a variável de estado ξ	
e	Erro de observação; Erro da medição	
g	Ganho do observador por intervalo	
k	Vector dos parâmetros do modelo; índice de tempo	
k^*	Vector dos parâmetros estimados do processo	
k_i	Parâmetro i do modelo	
k_{ij}	Coeficientes de rendimento ou estequiométricos do componente i na reacção j	
k_o	Valores dos parâmetros nominais	
\hat{k}	Vector dos parâmetros estimados do modelo	
$k_j^{perturbado}$	Valor do parâmetro j perturbado	
$k_j^{nominal}$	Valor do parâmetro j nominal	
m	Número de reacções	
m	Cociente entre o coeficiente de rendimento k_1 e o valor máximo da taxa específica de consumo de glucose	($MM^{-1}T$)
n	Número de componentes ou variáveis de estado; número de equações; número de parâmetros do modelo	
n_j	Número de componentes envolvidos na reacção j	
np	Número de pontos experimentais	
n_p	Número de parâmetros estimados	

n_t	Número de dados experimentais; número de amostras	
$n_{cond}(A)$	Número de condição da matriz A	
p	Valor de prova; característica da matriz K	
q	Número de variáveis de estado medidas em linha	
q_{Ac}	Valor da taxa específica de consumo de acetato	$(MM^{-1}T^{-1})$
$q_{Ac,máx}$	Taxa específica máxima de consumo de acetato	$(MM^{-1}T^{-1})$
q_S	Valor da taxa específica total de consumo de glucose	$(MM^{-1}T^{-1})$
$q_{S,crit}$	Valor da taxa específica crítica de consumo de glucose	$(MM^{-1}T^{-1})$
$q_{S,máx}$	Valor máximo para a taxa específica de consumo de glucose	$(MM^{-1}T^{-1})$
r	Vector das taxas cinéticas de reacção	$(MM^{-1}T^{-1})$
r	Número de funções de estado desconhecidas	
r_j	Taxa cinética de reacção j	$(MM^{-1}T^{-1})$
s_{aprox}	Desvio padrão aproximado	
s_θ	Desvio padrão associado ao parâmetro θ	
t	Tempo	(T)
t	Valor crítico da distribuição t de <i>Student</i>	
t_0	Tempo inicial	(T)
$t_{branqueamento}$	Tempo relativo de branqueamento (<i>photobleaching</i>)	
t_f	Tempo final da experiência ou da fermentação	(T)
t_i	Tempo i	(T)
u	Variável manipulada	
u	Vector das variáveis de entrada	$(MM^{-1}T^{-1})$
w	Ruído branco de distribuição normal, não correlacionado de média nula e variância unitária	
w_j	Soma ponderada dos componentes principais e dos seus valores próprios associados	
x	Vector das variáveis de estado; vector solução de um SEL	
x_0	Vector das variáveis de estado iniciais	
y	Vector das variáveis de saída; variável controlada	
$y(k)$	Vector das medidas de previsão do modelo	
y_l	Componente escalar de y ($l=1, \dots, n-p$)	
y_m	Vector das medidas experimentais	
$y_{ijz}^{perturbado}$	Valor do dado z , da variável de estado i , calculado pelo modelo utilizando o parâmetro j perturbado	
$y_{ijz}^{nominal}$	Valor do dado z , da variável de estado i , calculados pelo modelo utilizando o parâmetro j nominal	
$\bar{y}_{ijz}^{nominal}$	Média dos valores da variável de estado	

LETRAS GREGAS MAIÚSCULAS

χ^2	Distribuição de Qui-quadrado
Ω	Matriz de ganhos dos estimadores

LETRAS GREGAS MINÚSCULAS

α	Nível de significância	
α_{ij}	Ângulo formado entre os vectores de sensibilidade de dois parâmetros i e j	
β_i	Constantes da série de MacLaurin para a função do erro relativo do controlador	
δ	Grau relativo (ordem da equação diferencial)	
$d\varepsilon/d\gamma$	Sensibilidade do erro	
$\partial y/\partial k ; \partial y/\partial \theta$	Matriz das funções de sensibilidade	
ε	Termo associado ao ruído; erros ou resíduos aleatórios do modelo de regressão linear múltipla	
ε^{rel}	Erro relativo do controlador	
ε_0^{rel}	Valor pré-definido do erro relativo do controlador	
ϕ	Vector regressor	
γ_1	Parâmetro de sintonização do controlador	
γ_2	Parâmetro de sintonização do controlador	
λ_i	Valor próprio i	
λ_{Em}	Comprimento de onda de emissão	(L)
λ_{Ex}	Comprimento de onda de excitação	(L)
$\lambda_{máx}$	Valor próprio máximo	
$\lambda_{mín}$	Valor próprio mínimo	
μ	Taxa específica de crescimento	(T ⁻¹)
$\mu_{crítico}$	Taxa específica crítica de crescimento	(T ⁻¹)
$\mu_{máxima}$	Taxa específica máxima de crescimento	(T ⁻¹)
θ	Vector de parâmetros do modelo	
θ_0	Valor nominal de θ	
ξ	Vector de estado constituído pelas concentrações dos componentes	(MM ⁻¹)
ξ_i	Variável de estado (concentração do componente i)	(MM ⁻¹)
$\xi_{e,i}$	Concentração do componente i na correspondente alimentação	(MM ⁻¹)
ξ_{exp}	Valores experimentais da variável de estado ξ	(MM ⁻¹)
ξ_{est}	Valores estimados da variável de estado ξ	(MM ⁻¹)

$\xi_{e,i}^G$	Concentração do componente i na corrente de entrada gasosa	(MM ⁻¹)
$\xi_{s,i}^G$	Concentração do componente i na corrente de saída gasosa	(MM ⁻¹)
ω_{ij}	Ganho	
δ_C	Norma dos desvios dos parâmetros causados pelos desvios das medições	
δy	Desvios das medições	
$\partial_{y_k}^r$	Função de sensibilidade relativa de y em relativamente a k	
$\delta\theta$	Incerteza associada ao valor θ	
δk	Incerteza associada ao valor de k	
ψ	Matriz transformação de estado	
τ	Período natural de oscilação	(T)
ζ	Coeficiente de amortecimento	

ÍNDICES

a, b	Partição
$crit$	Crítico
e	Entrada
est	Estimado
exp	Experimental
f	Condições finais da fermentação
l	Linha de uma dada matriz
$máx$	Máximo
min	Mínimo
tot	Total

NOTAÇÃO MATEMÁTICA

T	Transposta de uma matriz
-1	Inversa de uma matriz
$+$	Pseudo-inversa de uma matriz; limite superior
$-$	Limite inferior
\wedge	Valor estimado
$*$	Valor nominal, valor de referência
\sim	Erro de estimação
$\arccos(\dots)$	Função trigonométrica inversa da função cosseno
\det	Determinante
$E\{\dots\}$	Operador estatístico esperança

$f(\dots, \dots)$	Vector de funções reais
$g(\dots, \dots)$	Vector de funções reais
I_n	Matriz identidade (de ordem n)
max	Valor máximo
min	Valor mínimo
$sen(\dots)$	Função trigonométrica seno
δ	Varição; incremento infinitesimal
∂	Derivada parcial
Λ	Número de condicionamento de uma matriz
$\ \dots\ $	Norma
$ \dots $	Valor absoluto
$\langle \dots, \dots \rangle$	Produto interno
Σ	Somatório
Π	Produtório
\int	Integral

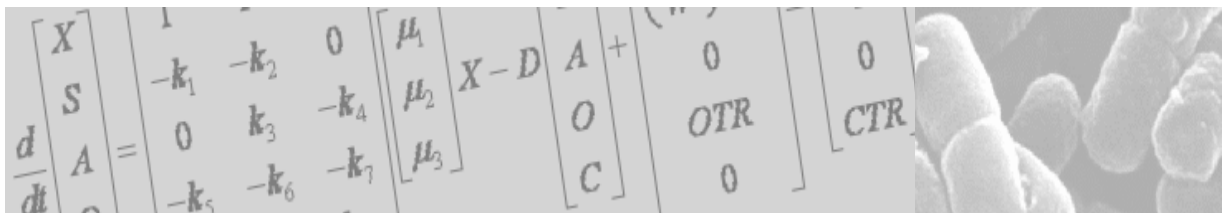
SIGLAS

ACP	Análise de Componentes Principais
ADN	Ácido desoxirribonucleico
ADP	Adenosina difosfato
ANOVA	Análise de Variância
AO	Observador assintótico
ATC	Ciclo dos ácidos tri-carboxílicos, ciclo de <i>Krebs</i> ou ciclo do ácido cítrico
ATP	Adenosina trifosfato
BLUE	<i>Best linear unbiased estimator</i>
CAT	Cloranfenicol-acetyl transferase
CER	Taxa de produção de dióxido de carbono (<i>carbon dioxide evolution rate</i>)
CTR	<i>Carbon Transfer Rate</i>
DCU	Unidade de medida e controlo digital directo (<i>digital control unit</i>)
DDE	<i>Dynamic data Exchange</i>
DO	Densidade óptica
EBFP	Proteína de fluorescência azul melhorada
ECFP	Proteína de fluorescência cinzenta melhorada
EGFP	Proteína de fluorescência verde melhorada

EMQ	Erro dos mínimos quadrados
EQM	Erro quadrático médio residual
EYFP	Proteína de fluorescência amarela esverdeada melhorada
FADH₂	Hidrogeno-flavina adenina dinucleótido
FDA	<i>Food and Drug Administration</i>
GAPDH	Enzima gliceraldaído-3-fosfato desidrogenase
GEAtbx	<i>Genetic and Evolutionary Algorithm Toolbox</i>
GFP	Proteína de fluorecência verde (<i>green fluorescent protein</i>)
IPTG	Isopropil- β -D-tiogalactósido
K-S	Teste de <i>Kolmogorov-Smirnov</i>
MATLAB	<i>Matrix Laboratory</i>
MIF	Matriz de informação de <i>Fisher</i>
MIT	Massachusetts Institute of Technology
MRLM	Modelo de Regressão Linear Múltipla
MS	Espectrómetro de massa
NADH	Hidrogeno-nicotinamida adenina dinucleótido
OEK	Observador estendido de <i>Kalman</i>
OI	Observador por intervalo
OTR	<i>Oxygen Transfer Rate</i>
OUR	Taxa de consumo de oxigénio (<i>Oxygen Uptake Rate</i>)
P	Acção de controlo proporcional
PD	Acção de controlo proporcional derivativo
PI	Acção de controlo proporcional integral
PID	Acção de controlo proporcional integral derivativo
SDS	Dodecilsulfato de sódio
SEL	Sistema de Equações Lineares
SEQ	Soma dos erros quadráticos
SPSS	<i>Statistical Package for the Social Sciences</i>
TNF	Factor de necrose tumoral
VIF	Factor de inflação da variância

CAPÍTULO 1

INTRODUÇÃO



1. INTRODUÇÃO

1.1. CONTEXTO E MOTIVAÇÃO

Nas últimas décadas tem-se assistido a uma grande dinâmica na investigação e desenvolvimento de tecnologias biológicas, resultando no aparecimento de novos produtos e serviços em sectores tão diversos e essenciais como o farmacêutico e a saúde, o agro-alimentar e o ambiente. Essas novas tecnologias, conjugadas com o desenvolvimento em outras áreas como a informática e a engenharia, potenciam um elevado crescimento económico, resultado da transformação da estrutura industrial associada ao aparecimento da indústria biotecnológica.

A Biotecnologia é hoje um sector de enorme importância económica e estratégica, constituindo um projecto integrado de várias áreas do conhecimento das ciências da vida e da engenharia, nomeadamente da biologia, bioquímica, genética e bioprocessos.

Tratando-se de um mercado extremamente dinâmico e inovador, a União Europeia tem pautado por políticas que estimulam o investimento neste sector. Assim sendo, a Comissão Europeia estima que no final da década os mercados globais, incluindo sectores onde a biotecnologia e as ciências da vida constituem a maior parcela da nova tecnologia aplicada, poderão atingir mais de 2000 biliões de euros (Pissarra, 2003).

A Biotecnologia moderna utiliza técnicas de ADN (ácido desoxirribonucleico) recombinado *in vitro* para desenvolver e aumentar a capacidade produtiva dos microrganismos utilizados na indústria. As técnicas de manipulação de ADN desenvolveram-se nos últimos 30 anos, passando da transferência de informação genética apenas entre organismos procariontes, para tecnologias que facilitam a produção eficiente e controlada de proteínas complexas em hospedeiros.

A bactéria *Escherichia coli* (*E. coli*) é o hospedeiro mais utilizado para produzir proteínas recombinadas, não só por ser um dos microrganismos melhor caracterizados mas também devido à sua capacidade de crescer em altas concentrações celulares, permitindo processos fermentativos de elevado rendimento (Schmidt, 2004).

A maioria das proteínas sintetizadas naturalmente em *E. coli* é intracelular, acumulando-se no citoplasma e resultando muitas vezes na formação de corpos de inclusão, que são agregados insolúveis e inactivos de proteínas superproduzidas. Se por um lado, a formação de corpos de inclusão facilita a recuperação da proteína com elevada pureza e concentração, protege-a da acção de proteases e é útil quando a proteína, na sua forma activa, é tóxica e muitas vezes letal para a célula hospedeira, por outro lado, a formação de corpos de inclusão dificulta o processo de renaturação das proteínas precipitadas, para se tornarem novamente activas, o que nem sempre é possível, conduzindo a uma diminuição do rendimento final (Baneyx, 1999; Mergulhão *et al.*, 2005).

Nos meados dos anos 80, do século passado, surgem os sistemas de expressão de proteínas em células animais e vegetais, motivados pela dificuldade do uso de bactérias como células hospedeiras para a síntese de proteínas recombinadas de eucariontes na forma funcionalmente activa. Esta limitação deve-se ao facto dos procariontes serem incapazes de efectuar modificações pós-tradução dos polipéptidos sintetizados, muitas vezes necessárias à obtenção da proteína funcional e que ocorrem no hospedeiro natural. Contudo, apesar das vantagens da utilização de células animais e vegetais, estes sistemas apresentam grandes limitações no desenvolvimento de processos simples, eficientes, de baixo custo e de elevado rendimento, para a produção em escala industrial de proteínas.

Progressos recentes levaram ao melhoramento de estirpes de *E. coli*, permitindo alterar as suas características de produção, e reforçando, por isso, o seu papel como o hospedeiro mais utilizado para a produção de proteínas recombinadas (Richins *et al.*, 1997; Baneyx, 1999; Bird *et al.*, 2004). Assim, existem actualmente no mercado vários produtos farmacêuticos produzidos em *E. coli*, sendo a insulina a primeira proteína humana obtida de uma estirpe recombinante dessa bactéria. Outros casos de sucesso são a produção de outras proteínas humanas como a hormona do crescimento, o factor de necrose tumoral - TNF, o interferão e a interleucina-2, péptidos sintéticos como factores de coagulação sanguínea e factores de crescimento e novos tipos de vacinas (Fieschko e Ritch, 1986; Lee, 1996; Huang *et al.*, 1999; Juan *et al.*, 2003; Jeong *et al.*, 2004; Yuan *et al.*, 2004).

Os bioprocessos podem ser descritos como processos nos quais uma população pura ou mista de microrganismos consome vários substratos que utiliza para crescer e produzir metabolitos. Paralelamente ao melhoramento das estirpes dos microrganismos é aconselhável otimizar as condições do bioprocessos, nomeadamente a temperatura de crescimento e/ou de produção, o pH e a composição do meio, tendo em vista a obtenção de processos de elevado rendimento (Lee, 1996; Sonnleitner, 1998; Zawada e Swartz, 2005).

Actualmente, a capacidade de controlar de uma forma automática e precisa os bioprocessos no seu estado óptimo, é de enorme importância para a maior parte das indústrias, uma vez que contribui para a diminuição dos custos de produção e aumenta o rendimento, mantendo a qualidade dos produtos. Devido ao aumento da competitividade neste sector da indústria, estratégias baseadas apenas no conhecimento empírico e de tentativa erro já não são suficientes nem eficientes. Este facto é realçado pela FDA (*Food and Drug Administration*) que aconselha a utilização de novas estratégias de monitorização e controlo como meio de melhorar e assegurar a qualidade dos produtos, no caso particular da indústria farmacêutica (FDA, 2004).

Nas últimas décadas, vários estudos académicos têm demonstrado a superioridade de estratégias baseadas em modelos matemáticos para o desenho, a optimização e o controlo de bioprocessos. No entanto, a grande maioria dos processos industriais continua a ser controlada e optimizada sem o recurso explícito a esses modelos. O desenvolvimento de modelos matemáticos, que permitam descrever esses processos, tornou-se essencial, uma vez que é normalmente mais barato modelar um sistema e simular as suas condições de operação do que realizar experiências laboratoriais.

Existem ainda situações em que, para além do ponto de vista económico, existem razões de ordem prática, de segurança e éticas que tornam inviáveis a experimentação em sistemas reais (Jacquez, 1998; Versyck e Van Impe, 1999).

A modelação matemática, a monitorização e o controlo em tempo real dos bioprocessos é hoje um enorme desafio para biotecnólogos e engenheiros de controlo, cabendo-lhes a tarefa de criar plataformas de comunicação entre si e com a indústria, de modo a que esta possa usufruir das técnicas desenvolvidas.

O desenvolvimento de estratégias de modelação, monitorização, controlo e optimização é justificado pela necessidade de acompanhar, de modo regular e preciso, o processo, que se pretende que seja reproduzível. Contudo, a principal dificuldade na modelação, monitorização e controlo de sistemas biológicos reside na reduzida oferta ou mesmo falta de sensores baratos e robustos, capazes de permitirem medições directas e em linha das variáveis de estado, na significativa incerteza da estrutura e parâmetros dos modelos utilizados e na natureza não linear e dinâmica desses sistemas.

A descrição de bioprocessos, por modelos matemáticos, envolve normalmente problemas de identificação, quer da estrutura, quer dos parâmetros desses modelos. A planificação de experiências constitui a aplicação de metodologias para seleccionar experiências eficientes e informativas. Assim, é necessário estabelecer uma função matemática que permita determinar a eficiência de uma dada experiência tendo em vista o seu objectivo e seleccionar as condições óptimas (Bernard *et al.*, 2001).

Das várias funções objectivo descritas na literatura, são as funções baseadas na Matriz de Informação de Fisher (*MIF*) as que têm sido mais utilizadas para avaliar a precisão da identificação dos parâmetros do modelo matemático. De facto, a *MIF* é um instrumento teórico que permite avaliar a sensibilidade dos parâmetros e erros experimentais, facilitando a quantificação da qualidade dos parâmetros identificados.

A etapa de monitorização engloba a aquisição e análise de informação recolhida quer por sensores físicos quer por sensores por programação, também designados por observadores de estado, tendo em vista o acompanhamento de variáveis de estado que afectam o bioprocessos. Os observadores têm sido, nos últimos anos, motivo de vários estudos e representam uma alternativa muito interessante aos sensores físicos, permitindo a monitorização em linha de variáveis de estado que não podem ser medidas em tempo real, utilizando modelos matemáticos em conjugação com um conjunto limitado de variáveis de estado passíveis de serem medidas (de Assis e Maciel, 2000; Dochain, 2003; Bogaerts e Vande Wouwer, 2004; Bernard e Gouzé, 2004). Apesar da teoria que fundamenta alguns sensores por programação estar bem consolidada, não existem ainda muitos exemplos documentados onde estes algoritmos sejam aplicados a bioprocessos complexos, descritos por modelos dinâmicos, constituídos por várias equações de balanço e com cinéticas complexas. Assim, o desenvolvimento e validação experimental destes algoritmos poderá ser uma ajuda importante no controlo e optimização de processos industriais, colmatando por um lado a falta de

sensores físicos e por outro permitindo ultrapassar a relutância da indústria na utilização de novos sensores físicos, devido a problemas de esterilização dos mesmos.

A etapa de controlo pode ser vista como a aplicação da informação recolhida e tratada na etapa de monitorização. O objectivo desta etapa é manter algumas variáveis de estado do processo, ou uma função dessas variáveis, próximo de valores de referência pré-estabelecidos, face a perturbações e/ou variações do processo.

No projecto de algoritmos de controlo de bioprocessos destacam-se duas correntes: a teoria de controlo óptimo e a teoria de controlo adaptativo. A primeira assume que o modelo do processo é perfeito e completamente conhecido. Esta hipótese faz com que em geral a implementação prática desses algoritmos no controlo de experiências reais falhe e os resultados obtidos se desviem dos resultados esperados por simulação. A segunda, onde os algoritmos de controlo são desenhados sem o completo conhecimento do modelo do processo, não garante *a priori* que o processo consiga atingir o seu desempenho óptimo.

Assim, o desenvolvimento de novas estratégias de controlo baseadas em modelos matemáticos que possam ser implementadas em linha e assim eliminar as perturbações do processo adquire, cada vez mais, uma relevância primordial. Com efeito, essas metodologias devem ser capazes de melhorar o desempenho do processo, aumentando a eficiência da produção e o lucro.

Apesar da importância reconhecida da problemática abordada verifica-se que, na literatura, apenas se encontra um reduzido número de trabalhos de aplicação de técnicas de desenho experimental e sua implementação prática em bioprocessos descritos por sistemas dinâmicos complexos. Assim sendo, e de modo a cativar uma maior atenção por parte da comunidade académica bem como da indústria é fundamental criar estratégias simples mas, ainda assim, eficazes e robustas.

Deste modo, uma estratégia possível será aquela que combine:

- o desenho óptimo de experiências para a identificação dos parâmetros de um modelo matemático capaz de descrever satisfatoriamente o processo fermentativo;
- a aplicação do modelo matemático no desenho de observadores de estado e sua posterior implementação prática de modo a, em algumas situações, possibilitar a substituição de sensores menos robustos e noutras monitorizar variáveis de estado para as quais não existem métodos de análise em linha;
- e por fim a utilização do modelo do processo no controlo e optimização do processo fermentativo sob estudo.

1.2. OBJECTIVOS

Com o presente trabalho de investigação pretendeu-se:

- (i) Desenvolver estratégias de alimentação de substrato para identificação de parâmetros - estimação de coeficientes de rendimento para uma estrutura pré-estabelecida do modelo - de um modelo do processo de crescimento de uma estirpe recombinante de *Escherichia coli* à escala laboratorial por operação semi-contínua.
- (ii) Verificar experimentalmente as trajectórias de alimentação calculadas por simulação.
- (iii) Desenvolver estratégias de monitorização avançada para a estimação em linha de variáveis de estado e parâmetros cinéticos.
- (iv) Desenvolver e implementar leis do controlo adaptativo com vista à maximização do crescimento e/ou produção.

A descrição dinâmica do processo em estudo assenta numa equação não linear do espaço dos estados denominada de Modelo Dinâmico Geral de Reactores Biológicos (Bastin e Dochain, 1990). As aproximações matemáticas utilizadas no desenvolvimento de algoritmos de identificação, de estimação e controlo têm por base este modelo dinâmico geral e inspiram-se nas propostas apresentadas por esses autores. A aproximação matemática para a identificação dos coeficientes de rendimento baseou-se nos trabalhos desenvolvidos por Chen (1992) e Ferreira (1995).

O trabalho desenvolvido teve sempre presente a necessidade de um cuidado desenho experimental tendo em vista a obtenção de dados com elevado grau de informação e riqueza que permitissem a determinação adequada dos parâmetros do modelo, aumentando deste modo a confiança do mesmo.

1.3. ORGANIZAÇÃO DA TESE

A tese é organizada em seis capítulos onde se descreve o trabalho de investigação realizado. O contexto e âmbito da tese são introduzidos no presente capítulo.

No Capítulo 2 descreve-se o processo em estudo. As principais características fisiológicas e metabólicas do microrganismo utilizado são apresentadas com o propósito de melhor compreender o comportamento do mesmo durante as fermentações. O equipamento usado, a instalação experimental bem como as metodologias e as técnicas de análise adoptadas são descritos em pormenor.

No Capítulo 3 são apresentados os modelos matemáticos de espaço de estados para o processo em estudo. São desenvolvidos algoritmos para a identificação dos coeficientes de rendimento, tendo por base a optimização da riqueza informativa da experiência, quantificada por índices relativos à

matriz de informação de Fisher. A otimização foi realizada recorrendo à rotina *Genetic and Evolutionary Algorithm Toolbox (GEAtbx 3.5)* desenvolvida por Pohlheim (2004) para correr em MATLAB versão 7.1 (*Matrix Laboratory*). São ainda apresentados os resultados da verificação experimental das trajectórias de alimentação obtidas com base na metodologia descrita anteriormente. A precisão dos parâmetros determinados foi investigada recorrendo a funções de sensibilidade e ao cálculo do respectivo erro. Apresenta-se ainda um estudo da estabilidade matemática do método de determinação dos coeficientes de rendimento.

O Capítulo 4 apresenta uma revisão teórica dos sensores por programação para a estimação de variáveis de estado e de parâmetros cinéticos. Apresentam-se algoritmos para a monitorização de variáveis de estado não mensuráveis em linha, através de observadores de Kalman, assintóticos e por intervalo, do processo em estudo. É estabelecida uma comparação do desempenho desses três observadores sendo o seu comportamento validado experimentalmente.

No Capítulo 5, a lei do controlo adaptativo, baseada em estratégias de alimentação, é desenvolvida. Com esta metodologia pretende-se controlar a taxa específica de crescimento com o intuito de maximizar o crescimento e/ou a produção. Este estudo foi realizado por simulações matemáticas, avaliando-se a robustez do algoritmo ao ruído nas variáveis medidas em linha, à variação dos parâmetros do modelo e à existência de perturbações externas.

Por fim, as conclusões assim como sugestões para trabalhos futuros são apresentadas no Capítulo 6.

1.4. BIBLIOGRAFIA

Baneyx, F., (1999). Recombinant protein expression in *Escherichia coli*. *Current Opinion in Biotechnology*, 10 (5), 411-421.

Bastin, G. e Dochain, D., (1990). *On-line estimation and adaptive control of bioreactors*, Elsevier Science Publishers, Amsterdam.

Bernard, O. e Gouzé, J.L., (2004). Closed loop observers bundle for uncertain biotechnological models. *Journal of Process Control*, 14 (7), 765-774.

Bernard, O., Hadj-Sadok, Z., Dochain, D., Genovesi, A. e Steyer, J.P., (2001). Dynamical model development and parameter identification for an anaerobic wastewater treatment process. *Biotechnology and Bioengineering*, 75 (4), 424-438.

Bird, P.I., Pak, S.C., Worrall, D.M. e Bottomley, S.P., (2004). Production of recombinant serpins in *Escherichia coli*. *Methods*, 32 (2), 169-176.

Bogaerts, P. e Vande Wouwer, A., (2004). Parameter identification for state estimation - application to bioprocess software sensors. *Chemical Engineering Science*, 59 (12), 2465-2476.

Chen, L., (1992). *Modelling, identifiability and control of complex biotechnological systems*. Tese de Doutoramento, Universidade Católica de Louvain, Bélgica.

de Assis, A.J. e Maciel, R., (2000). Soft sensors development for on-line bioreactor state estimation. *Computers & Chemical Engineering*, 24 (2-7), 1099-1103.

Dochain, D., (2003). State and parameter estimation in chemical and biochemical processes: a tutorial. *Journal of Process Control*, 13 (8), 801-818.

Ferreira, E.C., (1995). *Identificação e controlo adaptativo de processos biotecnológicos*. Tese de Doutoramento, Universidade do Porto, Portugal.

Fieschko, J. e Ritch, T., (1986). Production of human alpha consensus interferon in recombinant *Escherichia coli*. *Chemical Engineering Communications*, 45, 229-240.

Food and Drug Administration (FDA), (2004). Guidance for industry PAT - a framework for innovative pharmaceutical development, manufacturing and quality assurance. *Guidance Document, U.S. Department of Health and Human Services Food and Drug Administration*.

Huang, R.C., Lam, E., Chen, Y.H., Hackett, J., Lam, T.L., Liu, D., Ma, M.C., Siu, K.L., Sivakesava, S., Xu, Z.N., Wong, R.S.C. e Wong, W.K.R., (1999). Human epidermal growth factor excreted by recombinant *Escherichia coli* K-12 has the correct N-terminus and is fully bioactive. *Process Biochemistry*, 35 (1-2), 1-5.

Jacquez, J.A., (1998). Design of experiments. *Journal of the Franklin Institute-Engineering and Applied Mathematics*, 335B (2), 259-279.

Jeong, K.J., Choi, J.H., Yoo, W.M., Keum, K.C., Yoo, N.C., Lee, S.Y. e Sung, M.H., (2004). Constitutive production of human leptin by fed-batch culture of recombinant rpoS(-) *Escherichia coli*. *Protein Expression and Purification*, 36 (1), 150-156.

Juan, C.C., Kan, L.S., Huang, C.C., Chen, S.S., Ho, L.T. e Au, L.C., (2003). Production and characterization of bioactive recombinant resistin in *Escherichia coli*. *Journal of Biotechnology*, 103 (2), 113-117.

Lee, S.Y., (1996). High cell-density culture of *Escherichia coli*. *Trends in Biotechnology*, 14 (3), 98-105.

Mergulhão, F.J.M., Summers, D.K. e Monteiro, G.A., (2005). Recombinant protein secretion in *Escherichia coli*. *Biotechnology Advances*, 23, 177-202.

Pissarra, P.N., (2003). Biotechnology in Portugal: promoting downstream private initiative and foreign investment should be the next priorities. *Boletim de Biotecnologia*, 75, 14-17.

Pohlheim, H., (2004). Documentation 3.5a for Genetic and Evolutionary Algorithm Toolbox for use with Matlab: GEATbx version 3.5

Richins, R., Htay, T., Kallio, P. e Chen, W., (1997). Elevated Fis Expression Enhances Recombinant Protein Production in *Escherichia coli*. *Biotechnology and Bioengineering*, 56 (2), 138-144.

Schmidt, F.R., (2004). Recombinant expression systems in the pharmaceutical industry. *Applied Microbiology and Biotechnology*, 65, 363-372.

Sonnleitner, B., (1998). Dynamic adaptation of microbes. *Journal of Biotechnology*, 65 (1), 47-60.

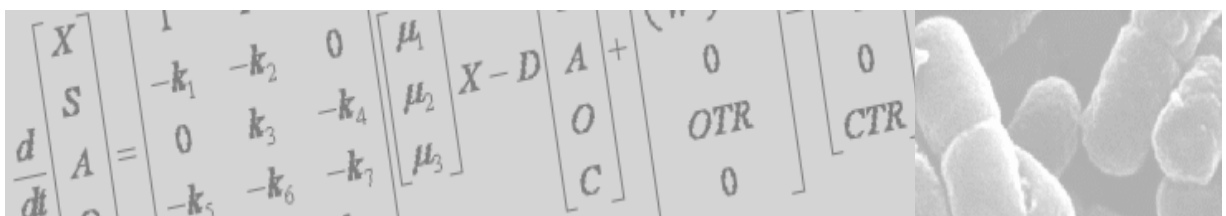
Versyck, K.J. e Van Impe, J.F., (1999). Feed rate optimization for fed-batch bioreactors: from optimal process performance to optimal parameter estimation. *Chemical Engineering Communications*, 172, 107-124.

Yuan, C., Reuland, J.M., Lee, L. e Huang, A.J.W., (2004). Optimized expression and refolding of human keratoepithelin in BL21 (DE3). *Protein Expression and Purification*, 35 (1), 39-45.

Zawada, J. e Swartz, J., (2005). Maintaining rapid growth in moderate-density *Escherichia coli* fermentations. *Biotechnology and Bioengineering*, 89 (4), 407-415.

CAPÍTULO 2

MICROORGANISMO, PROCESSO E METODOLOGIAS



2. MICROORGANISMO, PROCESSO E METODOLOGIAS

Os microrganismos têm sido amplamente utilizados na actividade económica humana, na indústria alimentar, na indústria de produção de antibióticos, vitaminas, aminoácidos, polissacáridos e biopolímeros, ou como agentes biológicos e em biotecnologias de protecção ambiental entre outras.

A combinação da tecnologia de ADN recombinado com processos de produção em larga escala, tem permitido a síntese de proteínas em quantidades que de outra forma seriam difíceis ou mesmo impossíveis de obter. Das várias proteínas sintetizadas por esses processos, as proteínas humanas recombinadas têm-se destacado pela sua importância como produtos biológicos farmacêuticos, sendo a *E. coli* o hospedeiro mais utilizado na sua produção (Lee, 1996; Jeong e Lee, 1999; Huang *et al.*, 1999; Jeong *et al.*, 2004; Yuan *et al.*, 2004; Shiloach e Fass, 2005). Como a maioria das proteínas produzidas em estirpes de *E. coli* recombinantes são intracelulares, a produtividade é proporcional à densidade celular final. Deste modo, têm-se desenvolvido técnicas de cultura de alta densidade celular para otimizar a produtividade, assim como para reduzir os custos de produção e de equipamento. Os processos de fermentação semi-contínua, também designados como *fed-batch*, têm sido utilizados para se obterem culturas de alta densidade celular de *E. coli* (Korz *et al.*, 1995; Prytz *et al.*, 2003; Hu *et al.*, 2004; Kim *et al.*, 2004).

Neste capítulo será feita uma revisão das principais características fisiológicas e metabólicas da *E. coli* assim como dos principais tipos de processos fermentativos. Descreve-se ainda o processo em estudo: crescimento da *E. coli* M15 com produção da proteína EYFP (*Enhanced Yellow-green Fluorescent Protein* - proteína de fluorescência amarela esverdeada melhorada); e da metodologia experimental utilizada para a execução de experiências à escala laboratorial. São também descritos os métodos analíticos e o equipamento utilizado.

2.1. *Escherichia coli*: CARACTERÍSTICAS FISIOLÓGICAS E BIOQUÍMICAS

As bactérias são microrganismos procariotas, unicelulares e reproduzem-se apenas assexuadamente. Caracterizam-se pela sua capacidade de adaptação rápida a alterações no meio ambiente crescendo e dividindo-se rapidamente. Tipicamente, as células procariotas apresentam taxas metabólicas (quantidade de substrato ou oxigénio consumida por hora e por unidade de massa celular) entre 10 a 100 vezes superior às das células eucariotas. Esta capacidade, aliada à sua composição química, estrutura, características bioquímicas e genéticas, faz das bactérias um objecto ideal para o estudo do crescimento celular, desenvolvimento e aplicação de novos métodos

de engenharia genética e metabólica e desenvolvimento de processos industriais de produção de proteínas recombinadas.

A bactéria *Escherichia coli* foi descrita pela primeira vez em 1885 por Theodor Escherich, um bacteriologista alemão, que a designou como *bacterium coli commune*. O nome "*Escherichia coli*", refere-se a um grupo de bactérias constituído por diferentes estirpes que têm em comum várias características (género: *Escherichia*; espécie: *coli*). Existem na literatura e na *internet* várias obras de referência bem como bases de dados onde se pode encontrar informação detalhada sobre a fisiologia, o metabolismo e genoma da *E. coli* (Neidhardt *et al.*, 1990; Holms, 1996; Karp *et al.*, 2002; Martinez-Antonio *et al.*, 2003; Carlson e Srienc, 2004).

2.1.1. Características fisiológicas

A *E. coli* como microrganismo procariota é relativamente pequena e simples. A sua dimensão típica é de 0,5 µm de diâmetro e 1,5 µm de comprimento, apresentando uma forma de bastonete. Na Figura 2.1 é apresentado um esquema típico da célula de *E. coli* e também uma fotografia obtida por microscopia electrónica.

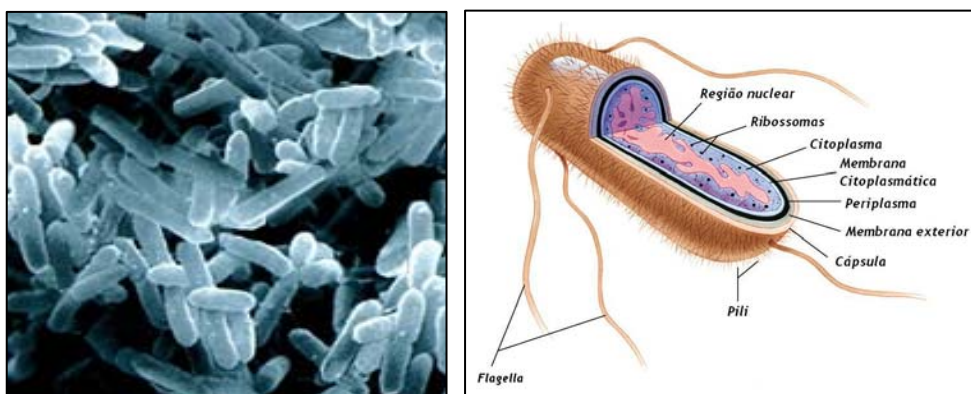


Figura 2.1 Fotografia¹, obtida por microscopia electrónica, e esquema² típico da célula de *E. coli*.

As células da *E. coli* são compostas essencialmente por compostos orgânicos, sendo a sua composição média, obtida por análise elementar em peso seco, de 50% de carbono, 20% de oxigénio, 14% de azoto, 8% de hidrogénio, 3% de fósforo, 2% de potássio, 1% de enxofre, 0,05% de cálcio, magnésio e cloro, 0,2% de ferro e um total de 0,3% de minerais como manganésio, cobalto, cobre, zinco e molibdénio (Neidhardt *et al.*, 1990).

A *E. coli* é uma bactéria Gram-negativa, cuja célula é revestida pela membrana citoplasmática, uma parede fina de mureína e uma membrana exterior. Em determinadas condições pode ainda

¹ Retirado de <http://www.sartorius.de>

² Adaptado de <http://cwx.prenhall.com>

observar-se uma camada viscosa de polímeros, essencialmente polissacáridos, designada por cápsula. Esta estrutura tem um papel preponderante na capacidade da *E. coli* se proteger de condições ambientais adversas facilitando-lhe também a adesão a superfícies. A membrana citoplasmática, com uma espessura típica de 70 Å, tem uma estrutura semelhante à das células eucariontes. Esta membrana é de grande importância pois determina quais as espécies químicas que podem ser transferidas entre a célula e o meio ambiente assim como a velocidade dessa transferência. A região designada por periplasma é característica apenas das bactérias Gram-negativas e situa-se entre a membrana exterior e a membrana citoplasmática. Esta parede de peptidoglicano ou mureína confere à célula equilíbrio estrutural permitindo-lhe preservar a sua integridade numa grande diversidade de condições externas (Bailey e Ollis, 1986; Neidhardt *et al.*, 1990).

Tipicamente, a *E. coli* possui dois tipos de apêndices superficiais, os *flagella* e os *pili*. Ambos surgem da membrana citoplasmática e são essencialmente agregados de proteínas. Apesar da sua composição semelhante, têm funções distintas. Os *flagella* são organelos de locomoção enquanto que os *pili* são organelos de adesão. Existem ainda *pili*, designados como *pili* sexuais, que permitem a transferência do ADN durante a conjugação bacteriana (Neidhardt *et al.*, 1990).

O material genético da bactéria localiza-se na região central ou pré-núcleo, sendo constituído por uma molécula de ADN circular. No entanto, podem também possuir outras pequenas moléculas de ADN circulares, designadas por plasmídeos.

A *E. coli* é uma bactéria anaeróbia facultativa. Sendo uma bactéria mesófila, pode crescer a temperaturas entre os 8 e os 48 °C. Contudo, a sua temperatura óptima de crescimento é de 39 °C, sendo a sua gama normal entre 21 e 37 °C. Cresce bastante bem em ambientes próximos da neutralidade, isto é, a um pH entre 6 e 8, podendo ainda crescer, embora mais lentamente, em condições onde o pH é de aproximadamente uma unidade a menos ou a mais desses valores (Neidhardt *et al.*, 1990).

2.1.2. Características bioquímicas

O crescimento celular é garantido pela oxidação de nutrientes através de uma complexa rede de reacções bioquímicas. Durante esse processo é produzido novo material biológico acompanhado pela produção de energia utilizada para manter a organização celular. Assim, observa-se a combinação de fluxos associados à síntese de biomassa e outros fluxos metabólicos que produzem energia utilizada na manutenção do gradiente iónico, reparação de proteínas e polinucleótidos existentes, assim como para manter as infra-estruturas bioquímicas que permitem à célula uma rápida adaptação a alterações ambientais.

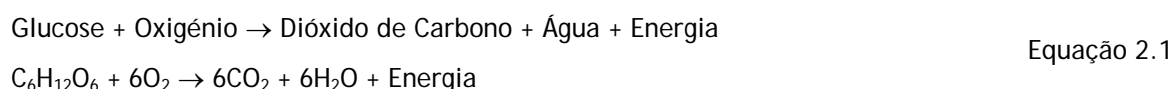
O princípio de conversão de substratos em biomassa e energia é relativamente simples, consistindo no transporte de cada fonte de carbono para a célula e depois na sua degradação a moléculas mais simples e incorporação nas vias metabólicas centrais. Estas contêm cerca de trinta

compostos: ácidos fosforilados e/ou carboxílicos e tiol-ésteres da coenzima A. A conversão de compostos intermediários é catalizada por enzimas organizadas em vias metabólicas como a glicólise ou via *Embden-Meyerhof-Parnas*, das pentoses fosfato, e o ciclo dos ácidos tri-carboxílicos (ATC), também designado como ciclo de *Krebs* ou ciclo do ácido cítrico e variações deste ciclo como a via *Entner-Doudorof* e da fosfoceolase. Estas vias conduzem não só à produção de energia, sob a forma de ATP (adenosina trifosfato) e coenzimas reduzidas, como o NADH (hidrogênio-nicotinamida adenina dinucleótido) e o FADH₂ (hidrogênio-flavina adenina dinucleótido), mas também, de moléculas de baixo peso molecular que são utilizadas na síntese de blocos precursores (aminoácidos, bases púricas e pirimídicas, açúcares fosfatados e outros derivados de açúcares, ácidos orgânicos e diferentes metabolitos) das diferentes macromoléculas que constituem as células (proteínas, ácidos nucleicos, lípidos e polissacáridos) (Varma *et al.*, 1993; Holms, 1996; Lima e Mota, 2003).

O metabolismo da *E. coli* serviu, durante mais de uma década, como objecto base para novos métodos de análise das vias metabólicas. Aceita-se que esta bactéria possua essencialmente quatro vias metabólicas, consideradas mais eficientes, para converter a fonte de carbono e oxigénio em novas células e energia de manutenção sob qualquer nível de limitação de oxigénio (Zhao e Shimizu, 2003; Carlson e Srienc, 2004). A glucose é o substrato mais utilizado experimentalmente, uma vez que pode ser simultaneamente fonte de carbono e energia, é facilmente metabolizado e o seu preço é relativamente baixo.

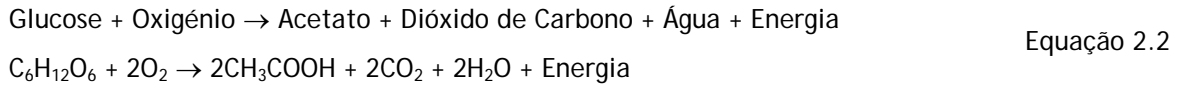
Na ausência de oxigénio ocorre a fermentação da glucose ou crescimento fermentativo com produção essencialmente de etanol, acetato, formato, succinato e lactato. A via anaeróbia mais eficiente produz três moléculas de ATP, uma de acetato, uma de etanol e duas de formato por molécula de glucose fermentada, traduzindo-se num rendimento energético de aproximadamente 0,5 moles de ATP por mole de carbono de glucose consumida. Por outro lado, em condições de microaerobiose obtém-se um rendimento energético de 1,33 moles de ATP por mole de carbono de glucose consumida, produzindo-se oito moléculas de ATP, duas de acetato e duas de formato (Carlson e Srienc, 2004).

Na Equação 2.1 representa-se a reacção mais eficiente para a produção de energia em termos de rendimento de ATP em glucose. Esta via metabólica aeróbia permite a oxidação completa da glucose em dióxido de carbono, consumindo seis moléculas de oxigénio e produzindo vinte e seis de ATP, e apresenta um rendimento energético de 4,33 moles de ATP por mole de carbono de glucose consumida.

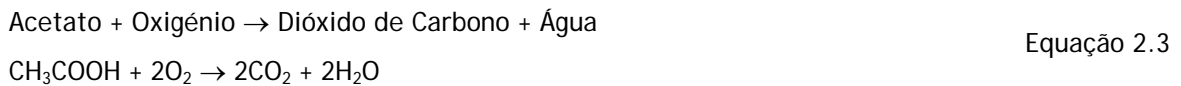


Em condições aeróbias e quando a glucose está em excesso, esta pode seguir uma via fermentativa originando a produção de compostos de baixo peso molecular. O acetato é o composto

mais frequente em fermentações de *E. coli* (Phue e Shiloach, 2004). Esta via apresenta um rendimento energético de 2 moles de ATP por mole de carbono de glucose oxidada e tem a seguinte estequiometria:



O acetato formado pela via anterior pode ser oxidado e usado como fonte de carbono secundária se a sua concentração não for inibitória e se a capacidade do ciclo dos ATC não estiver toda ocupada pelo metabolismo da glucose (Alba e Calvo, 2000). Quando o acetato é a única fonte de carbono os compostos intermediários do ciclo dos ATC são produzidos pelo ciclo do glioxilato que é inibido pela presença da glucose (Carlson e Srienc, 2004). A reacção de oxidação do acetato está representada na Equação 2.3.



Deve ser realçado que nas equações anteriores não é considerada a formação de biomassa, não se incluindo por isso, os coeficientes de rendimento que tomam em linha de conta a síntese celular.

Na Figura 2.2 é representado um esquema simplificado do metabolismo da glucose e do acetato para a *E. coli*. Em resumo, quando a *E. coli* cresce na presença de glucose em condições de aerobiose, esta é metabolizada em várias etapas das seguintes vias: glicólise, ciclo dos ATC e cadeia respiratória. Pela glicólise, a glucose 6-fosfato é convertida em piruvato. No ciclo dos ATC, a acetil-CoA é oxidada a dióxido de carbono e água e a estas reacções está também associada a produção de coenzimas reduzidas. Uma reacção intermédia faz a união entre as duas vias anteriores, onde pela acção da piruvato desidrogenase é produzida a acetil-CoA através do piruvato. Na cadeia respiratória as coenzimas reduzidas são reoxidadas pela transferência de electrões ao longo dessa cadeia que está associada à membrana citoplasmática. O oxigénio é o aceitador final de electrões sendo reduzido a água. Esta transferência de electrões está associada ao transporte de prótons para o exterior da membrana citoplasmática e conduz à produção de um gradiente electroquímico de prótons que é utilizado pelo complexo enzimático ATP sintetase na produção de ATP. Como referido anteriormente, quando a glucose se encontra em excesso há também produção de acetato. A explicação mais comum para a sua formação e acumulação é o fluxo elevado de carbono através da glicólise que excede a capacidade do ciclo dos ATC. Assim, o acetato é produzido através da acetil-CoA pela acção das enzimas fosfotransacetilase e acetato quinase e através do piruvato pela piruvato oxidase. Por outro lado, o acetato formado poderá ser novamente convertido em acetil-CoA pela acetil-CoA sintetase, sendo o reabastecimento do ciclo dos ATC feito via ciclo do glioxilato.

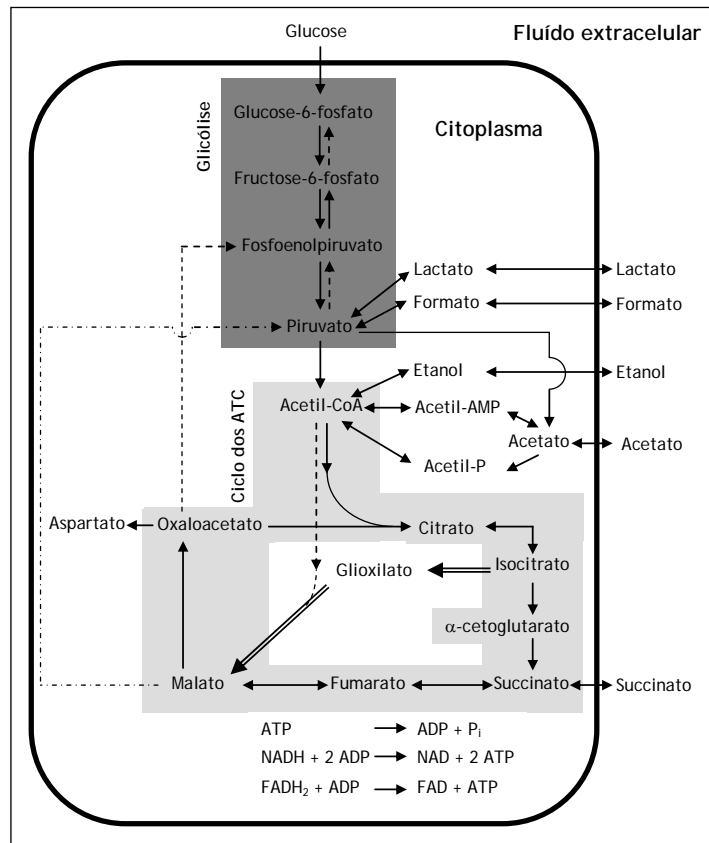


Figura 2.2 Metabolismo da glicose e do acetato na *E. coli* (adaptado de Carlson e Srienc, 2004 e de Phue *et al.*, 2005).

Na Figura 2.3 está representada a capacidade oxidativa limitada da *E. coli* em condições de aerobiose. Tenta-se representar de uma forma esquemática as principais vias metabólicas que podem ocorrer durante um processo fermentativo com esta bactéria. Para valores inferiores à taxa específica crítica de crescimento ($\mu_{\text{crítico}}$) a glicose é completamente oxidada com um elevado rendimento em biomassa por glicose consumida, denominando-se rendimento oxidativo em glicose ($Y_{\text{sx}}^{\text{ox}}$). A Equação 2.1 traduz este metabolismo oxidativo, o qual é representado na Figura 2.3 pelas setas 1 e 2. Para valores superiores a $\mu_{\text{crítico}}$ e inferiores à taxa específica máxima de crescimento ($\mu_{\text{máxima}}$), o excesso de glicose é convertido, através do metabolismo oxido-redutivo, em acetato, obtendo-se um rendimento inferior de biomassa por glicose consumida. Este rendimento é designado por redutivo em glicose ($Y_{\text{sx}}^{\text{red}}$). A Equação 2.1 e a Equação 2.2 traduzem o metabolismo oxido-redutivo representado pela seta 3 da Figura 2.3. Quando o consumo de glicose não utiliza a totalidade da capacidade oxidativa da *E. coli*, ou na ausência desta, a *E. coli* consome simultaneamente o acetato ou apenas este último, respectivamente. As setas 4 e 5 da Figura 2.3 descrevem esta situação, designada por metabolismo oxidativo em glicose e acetato ou oxidativo em acetato, respectivamente (Equação 2.1 e Equação 2.3).

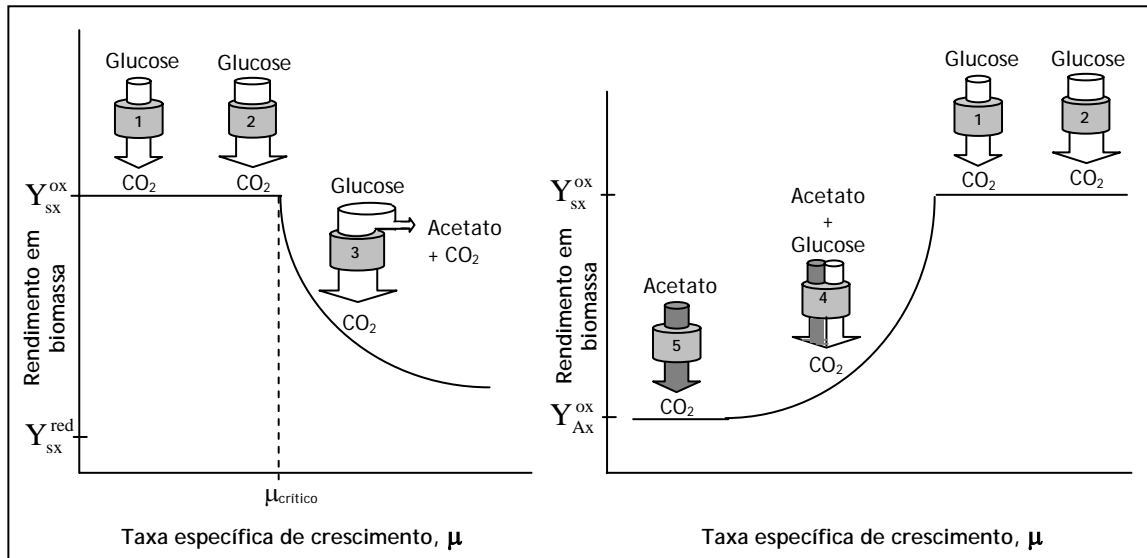


Figura 2.3 Representação esquemática da capacidade oxidativa limitada da *E. coli* (adaptado de Cannizzaro *et al.*, 2004).

Uma vez que a acumulação de acetato afecta o crescimento e a produção de proteínas, têm sido aplicadas várias estratégias para limitar a sua acumulação durante culturas de alta densidade celular de *E. coli*, das quais se destacam a adição gradual de glucose usando técnicas semi-contínuas de alimentação e o desenvolvimento de estirpes mutantes com vias metabólicas modificadas. Embora, o metabolismo da glucose e do acetato para a *E. coli* tenha sido intensamente estudado durante os últimos 50 anos, ainda não há uma justificação unânime para os fenómenos mais importantes que caracterizam o crescimento desta bactéria (Varma e Palsson, 1993a; 1993b; Ko *et al.*, 1994; Van de Walle e Shiloach, 1998; Xu *et al.*, 1999; Edwards e Palsson, 2000; Dittrich *et al.*, 2005).

2.2. PROCESSO FERMENTATIVO

O biorreactor, dependendo do regime de alimentação, pode operar em modo descontínuo (*batch*), semi-contínuo (*fed-batch*) e contínuo. Em reactores descontínuos não há adição de substratos nem nenhum meio de cultura é retirado ao longo da fermentação. Consequentemente, o volume permanece aproximadamente constante. Por outro lado, em reactores com operação semi-contínua, os substratos são adicionados contínua ou intermitentemente sem que nenhum meio de cultura seja retirado até ao final da fermentação originando um aumento de volume. Nos reactores contínuos o volume permanece constante, uma vez que simultaneamente à adição contínua de

substratos ocorre a saída também contínua de meio de cultura. Na Tabela 2.1 são apresentadas as principais vantagens e desvantagens destes modos de operação no que se refere à sua utilização para o crescimento de microrganismos, a produção de metabolitos e a sua aplicação em estudos de identificação de modelos matemáticos de bioprocessos.

As operações descontínua e contínua de biorreactores têm sido as mais utilizadas na identificação e determinação de parâmetros cinéticos de modelos a partir de dados experimentais (Holmberg, 1982). O potencial de aplicação da operação semi-contínua, nesse tipo de estudos, só foi reconhecido recentemente e é motivado pela sua aplicabilidade no desenho óptimo de experiências.

Do ponto de vista experimental, a operação descontínua é a mais atractiva tendo em conta a sua fácil implementação e redução dos custos em consumíveis e tempo de experimentação. A nível industrial, este tipo de operação é também uma configuração comum na produção de proteínas recombinadas. Por um lado, em fermentações descontínuas obtêm-se baixos rendimentos devido essencialmente à concentração limitante de substrato e/ou à concentração de produtos que inibem o crescimento. Por outro lado, neste tipo de fermentação observa-se grande estabilidade nas células recombinadas, traduzindo-se numa maior estabilidade nos plasmídeos (Ryan e Parulekar, 1990; Li e Taylor, 1994; Patnaik, 2002).

A operação contínua de biorreactores tem sido aplicada com sucesso na caracterização do crescimento e produção de várias estirpes de *E. coli*, embora implique um tempo de experimentação bastante longo, elevados gastos de reagentes e maior risco de contaminação (Flickinger e Rouse, 1993).

Han *et al.* (1992) estudaram a formação de acetato em fermentações contínuas de *E. coli*, em vários meios de cultura. Estes autores propõem que a formação de acetato possa ser reduzida através da adição ao meio de cultura de extracto de levedura (suplemento de constituintes necessários à célula) e metionina, reduzindo assim o fluxo da glucose e aumentando a capacidade oxidativa, respectivamente.

Um outro trabalho, desenvolvido por Rothen *et al.* (1998), caracteriza o crescimento da *E. coli* HB101[pGEc47] num biorreactor a operar em modo contínuo. Este estudo centrou-se no desenvolvimento de um meio de cultura definido óptimo para o crescimento desta estirpe, assim como na determinação da taxa de diluição crítica acima da qual há produção de acetato. Concluíram que o crescimento da estirpe de *E. coli* estudada é semelhante ao observado em meio complexo contendo extracto de levedura ($\mu = 0,41 - 0,45 \text{ h}^{-1}$). Para além disso, concluíram que a optimização do meio de cultura e o conhecimento da taxa de diluição crítica permitem aumentar o desempenho deste sistema de biotransformação.

Tabela 2.1 Resumo das vantagens e desvantagens dos tipos de operação do biorreactor (adaptado de Versyck, 2000).

Modo de operação	Vantagens	Desvantagens
Descontínuo (<i>batch</i>)	<ul style="list-style-type: none"> - equipamento simples - adequado para baixa produção - redução do esforço experimental, tempo e consumo de reagentes - fácil implementação 	<ul style="list-style-type: none"> - longo tempo de espera para enchimento e lavagem - variação das condições de reacção com o tempo - inadequado para estudos de identificação de parâmetros de modelos - não permite a transferência de conhecimento de parâmetros para outras condições de operação - inadequado para estudos de desenho óptimo de experiências
Contínuo	<ul style="list-style-type: none"> - permite elevada produção - boa qualidade dos produtos devido às condições constantes de operação - adequado para estudos cinéticos 	<ul style="list-style-type: none"> - requer controlo de fluxo - a cultura poderá ser instável durante o longo período de fermentação - grande esforço experimental, tempo e gastos de consumíveis - pouco adequado para estudos de identificação de parâmetros de modelos - permite com alguma incerteza a transferência de conhecimento de parâmetros para outras condições de operação - pouco adequado para estudos de desenho óptimo de experiências
Semi-contínuo (<i>fed-batch</i>)	<ul style="list-style-type: none"> - permite o controlo de condições ambientais, como por exemplo a concentração do substrato - adequado para estudos de identificação de parâmetros de modelos - permite a transferência de conhecimento de parâmetros para outras condições de operação - adequado para estudos de desenho óptimo de experiências 	<ul style="list-style-type: none"> - necessita de estratégias de alimentação para se obter a concentração desejada de substrato e produtos - necessita de equipamentos e <i>software</i> sofisticados - implica algum esforço experimental, tempo e gastos de consumíveis

Têm sido realizados vários trabalhos onde se aplica a fermentação contínua ao estudo e modelação da estabilidade dos plasmídeos em células recombinadas e a sua consequência na produção de proteínas. Ryan e Parulekar (1990) estudaram e compararam o efeito do pH, da concentração de fosfatos inorgânicos e do nível de oxigénio dissolvido no crescimento, estabilidade dos plasmídeos e produção de proteína em *E. coli* JM103[pUC8] em operação descontínua e contínua. Observaram que em contínuo existe uma maior instabilidade dos plasmídeos, traduzindo-se no aumento, ao longo do tempo, de células que não possuem o plasmídeo.

Mosrati *et al.* (1993) propõem um modelo para a variação da frequência da perda de plasmídeos das células recombinadas em função da taxa específica de crescimento das células que possuem o plasmídeo. O estudo foi realizado, em reactor contínuo, com a *E. coli* C600 gal K(GAPDH), estirpe geneticamente modificada que produz elevadas quantidades da enzima gliceraldeído-3-fosfato desidrogenase (GAPDH). Aqueles autores concluíram que a probabilidade de perda de plasmídeos aumenta com a taxa específica de crescimento das células que possuem o plasmídeo, isto é, com o aumento da taxa de diluição. Sugerem ainda que o modelo proposto permite o controlo em linha da estabilidade do conteúdo de plasmídeos nas células, o que é essencial para a optimização do processo.

A principal motivação para utilizar sistemas em contínuo em estudos de identificação da estrutura e dos parâmetros de modelos é a aplicabilidade da condição de crescimento balanceado. O crescimento balanceado implica que as velocidades de conversão dos diferentes componentes celulares sejam iguais, de tal modo que possam ser representadas por uma velocidade de reacção global. Esta consideração é a base da aplicação de leis de crescimento não estruturadas como a cinética de *Monod* e de *Haldane*. Assim sendo, num biorreactor em operação contínua e em estado estacionário é razoável considerar um crescimento balanceado, uma vez que a concentração dos vários compostos, como o substrato, a biomassa entre outros, assim como a taxa específica de crescimento global permanecem constantes com o tempo.

Uma dificuldade encontrada na aplicação da operação em contínuo em procedimentos de identificação de parâmetros é a ocorrência de lavagem do reactor ou *wash-out*. São disso exemplo, as experiências realizadas para determinar a taxa específica máxima de crescimento, parâmetro normalmente desconhecido, em que se utilizam taxas de diluição crescentes atingindo valores próximos da taxa de *wash-out*. Por outro lado, em geral os parâmetros determinados em condições de estado estacionário não descrevem, de modo adequado, o comportamento dinâmico das operações descontínuas e semi-contínuas, para o mesmo modelo não estruturado.

A fermentação semi-contínua tem sido usada na indústria desde o início do século XX, na produção de cerveja e de levedura para panificação (fermento de padeiro) e mais tarde, desde 1950, na produção de antibióticos. Contudo, é nos últimos vinte anos que esta técnica tem sido mais explorada na produção de proteínas recombinadas. Nestes processos têm sido aplicadas diferentes estratégias de alimentação, com e sem controlo por retroacção (*feedback*), de modo a otimizar a massa celular e a produção de proteínas. Actualmente, a operação semi-contínua em conjunto com

a utilização de meios de cultura otimizados e a aplicação de métodos de biologia molecular, tornaram possível alcançar densidades celulares acima de 190 g/L (em peso seco), em fermentações de *E. coli* recombinada, com produção de várias proteínas (Shiloach e Fass, 2005). Yamanè e Shimizu (1984) e Lee (1996) apresentam uma revisão pormenorizada da teoria e das várias estratégias de alimentação utilizadas em fermentação semi-contínua.

Como referido anteriormente, um dos aspectos importantes durante o crescimento da *E. coli* em culturas de alta densidade celular, em meios de glucose, é a formação e acumulação de acetato. A sua acumulação torna-se particularmente relevante quando há síntese de proteínas, uma vez que se tem constatado que concentrações de acetato superiores a 2,4 g/L afectam negativamente o crescimento e a produção de proteínas (Noronha *et al.*, 2000). Sendo assim, a estratégia de alimentação usada desempenha um papel crítico para o sucesso da fermentação, visto dela depender, em grande parte, o controlo da produção de acetato, a concentração máxima celular e a produtividade alcançadas.

Na literatura encontram-se descritos vários trabalhos onde diferentes estratégias de alimentação têm sido aplicadas em fermentações semi-contínuas de elevada densidade celular de *E. coli* com e sem controlo por retroacção (Tabela 2.2) (Korz *et al.*, 1995; Narciandi, 1996; Wangsa-Wirawan *et al.*, 1997; Suárez e Kilikian, 2000; Johnston *et al.*, 2002; Kleist *et al.*, 2003). Vários estudos mostram que têm sido aplicadas com sucesso quer estratégias simples como as de caudal de alimentação constante ou exponencial, como métodos mais sofisticados que envolvem esquemas de controlo indirecto com retroacção, de parâmetros físicos como o nível de oxigénio dissolvido e o pH.

Kim *et al.* (2004) desenvolveram uma estratégia de alimentação que combina a adição de nutrientes com caudal exponencial com o método *pH-stat*, de modo a evitar a acumulação de substrato no meio de cultura. Com esta estratégia de alimentação conseguiram controlar a taxa específica de crescimento num valor de $0,1 \text{ h}^{-1}$, obtendo uma densidade celular de 101 g/L.

O trabalho desenvolvido por Lin *et al.* (2005) realça a importância do desenvolvimento e aplicação das técnicas da engenharia metabólica em conjunto com estratégias de fermentação para desenvolver novos processos de produção em *E. coli*. Estes autores descrevem um processo de produção de succinato, em condições aeróbias, num reactor semi-contínuo. A estirpe mutante de *E. coli* HL27659k(Pkk313) foi desenvolvida pelos autores e demonstrou ter grande capacidade para produzir succinato em condições de total aerobiose, obtendo uma concentração de 58,3 g/L em 59 horas de fermentação. Os autores concluíram que este novo sistema é mais versátil do que os sistemas de produção anaeróbios convencionais podendo por isso, potenciar o seu desenvolvimento e optimização à escala industrial.

Grande parte dos trabalhos referidos anteriormente concentra-se no estudo do caudal de alimentação de nutrientes da fase pré-indução do processo de produção. Após a indução, o crescimento e a actividade metabólica das células são influenciados pela síntese das proteínas recombinadas, sendo também conhecido, que em certos casos o isopropil- β -D-tiogalactósido (IPTG), frequentemente utilizado como indutor, pode alterar o metabolismo da *E. coli*. Assim, a estratégia

de alimentação da fase de pós-indução poderá adquirir maior importância do que a alimentação da fase pré-indução, tendo em vista uma eficiente produção de proteínas.

Tabela 2.2 Estratégias de alimentação aplicadas a fermentações semi-contínuas (adaptado de Lee, 1996).

Estratégias de alimentação
<p>Sem controlo por retroacção:</p> <ul style="list-style-type: none">- <i>caudal constante</i> - adição de nutrientes a uma velocidade constante pré estabelecida. A taxa específica de crescimento diminui progressivamente;- <i>caudal crescente</i> - alimentação de nutrientes a uma velocidade crescente (linear ou por patamares). A diminuição da taxa específica de crescimento pode ser atenuada;- <i>caudal exponencial</i> - adição de nutrientes a uma velocidade exponencial. Pode-se conseguir uma taxa específica de crescimento constante.
<p>Com controlo por retroacção:</p> <p><i>controlo por retroacção indirecto:</i></p> <ul style="list-style-type: none">- <i>DO-stat</i> - adição de nutrientes quando se observa um aumento da concentração de oxigénio dissolvido, que resulta da diminuição da concentração de substrato;- <i>pH-stat</i> - adição de nutrientes quando se observa um aumento do pH resultado da diminuição da concentração da principal fonte de carbono;- <i>taxa de produção de dióxido de carbono (CER³)</i> - esta taxa é aproximadamente proporcional à taxa de consumo da fonte de carbono. É frequentemente utilizada no controlo da taxa específica de crescimento (pode ser estimada em linha, por exemplo, por espectrometria de massa);- <i>concentração celular</i> - o caudal de adição de nutrientes é determinado a partir da concentração celular (pode ser medida em linha, por exemplo, por turbidimetria); <p><i>controlo por retroacção directo:</i></p> <ul style="list-style-type: none">- <i>concentração de substrato</i> - a adição de nutrientes é directamente controlada pela concentração da principal fonte de carbono (um analisador em linha de glucose é usado para controlar directamente a concentração de glucose no fermentador).

Têm também sido aplicadas várias estratégias de alimentação na fase pós-indução de fermentações semi-contínuas de diferentes estirpes recombinadas de *E. coli*. O trabalho

³ CER - Carbon dioxide Evolution Rate

apresentado por Wong *et al.* (1998) estuda o efeito do tipo de alimentação na fase pós-indução na produção de uma proteína em *E. coli*, usando o IPTG como indutor. Observaram que o crescimento celular na fase pós-indução é independente do caudal de alimentação. No entanto, a síntese da proteína é significativamente afectada pela estratégia de alimentação aplicada nessa fase. O caudal de alimentação linear apresentou o melhor desempenho atingindo-se uma produção de proteína de 5,3 g/L.

O papel da operação semi-contínua tem adquirido maior destaque em estudos de identificação de parâmetros de modelos, visto oferecer melhores oportunidades no desenho óptimo de experiências. Um aspecto limitante na sua aplicação é a violação da condição de crescimento balanceado, que é a base da utilização de modelos não estruturados. Contudo, como ao longo de uma fermentação semi-contínua as condições ambientais se alteram lentamente, poder-se-á considerar que o sistema vai passar por sucessivos estados *quasi*-estacionários, para os quais a condição de crescimento balanceado poderá ser considerada.

Uma análise de trabalhos publicados acerca de técnicas aplicadas em fermentações de alta densidade celular de *E. coli* revela uma variação da concentração final celular entre 20 a 190 g/L (em peso seco). Esta variabilidade significativa poderá ser atribuída ao facto do crescimento de diferentes estirpes poder ser afectado por parâmetros e/ou factores não completamente conhecidos e compreendidos. Assim, um maior conhecimento e uma melhor modelação desses parâmetros e factores poderão contribuir para o desenvolvimento e optimização de métodos de crescimento e produção de proteínas recombinadas e permitir a sua implementação com sucesso a nível industrial.

2.3. PROTEÍNA DE FLUORESCÊNCIA VERDE

A proteína de fluorescência verde (GFP - *green fluorescent protein*) proveniente de medusa (*Aequorea victoria*) tem sido extensivamente aplicada em várias áreas das ciências biológicas. A sua característica única de auto-sintetizar o cromóforo necessitando apenas da presença de oxigénio molecular, isto é, sem necessidade de adicionar qualquer composto (substrato ou cofactor) externo ao organismo que a expressa para que a sua fluorescência seja detectada, faz dessa proteína um excelente marcador *in vivo* de expressão genética (Timmons *et al.*, 1997; Gill *et al.*, 2000) e de localização de proteínas (Gerdes e Kaether, 1996; Margolin, 2000) em vários sistemas biológicos como mamíferos, peixes, insectos, plantas, leveduras e numa grande variedade de bactérias (March *et al.*, 2003; Su, 2005).

A GFP é uma proteína composta por 238 aminoácidos com um peso molecular entre 27 e 30 kDa (Gerdes e Kaether, 1996). A sua estrutura terciária é representada na Figura 2.4 (A). Apresenta uma forma cilíndrica bastante compacta, sendo formada por onze cadeias exteriores não paralelas (a verde) que constituem a sua parede e possui um diâmetro e um comprimento aproximado de 30 e 40

Å, respectivamente. Existem ainda pequenas secções em α -hélice que formam o limite do cilindro (a cinzento) e um segmento α -hélice irregular (a azul) que assegura a estrutura do cromóforo que se localiza no centro geométrico do cilindro (a vermelho), e que é constituído pelo tripéptido Ser₆₅-Tyr₆₆-Gly₆₇, representado na Figura 2.4 (B) (Yang *et al.*, 1996).

A GFP é uma proteína excepcionalmente estável, devido essencialmente à sua estrutura terciária compacta. É bastante resistente à acção de proteases, agentes químicos desnaturantes como a ureia e o dodecilsulfato de sódio (SDS). Para além disso, mantém a sua fluorescência numa grande gama de valores de pH (5,5 a 12,0) e a temperaturas elevadas (até 70 °C).

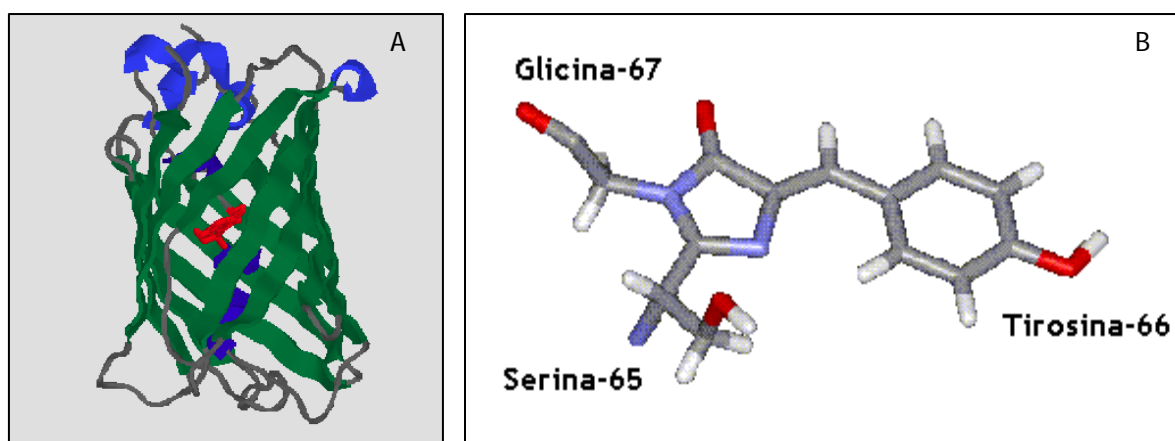


Figura 2.4 (A) Estrutura terciária da GFP⁴. (B) Tripéptido que constitui o cromóforo da GFP⁵.

Algumas variações na GFP deram origem a proteínas que têm fluorescência azul (EBFP), cinzenta (ECFP) e amarela esverdeada (EYFP). Por outro lado, manipulações genéticas possibilitaram a criação de variantes da GFP, como a proteína com fluorescência verde melhorada (EGFP) que fluoresce trinta e cinco vezes mais que a GFP selvagem. Na Tabela 2.3 encontram-se descritas algumas propriedades destas proteínas, tais como: o coeficiente de extinção (E_m), o rendimento quântico (Y_Q), o comprimento de onda de excitação (λ_{EX}), o comprimento de onda de emissão (λ_{EM}), o valor médio de pH ao qual a proteína perde 50% de fluorescência, o tempo relativo de branqueamento (*photobleaching* - $t_{\text{branqueamento}}$).

A monitorização e controlo de bioprocessos têm sido realizados, normalmente, recorrendo à análise de parâmetros como o oxigénio dissolvido, pH, densidade óptica (DO), a concentração de substrato e a concentração dos gases de saída da fermentação. Contudo, nem sempre é possível medir em linha esses parâmetros e nenhum deles reflecte, em tempo real, todas as alterações dos fenómenos que ocorrem no interior das células. Por outro lado, a monitorização da concentração das proteínas recombinadas, as quais são o produto de interesse em grande parte dos bioprocessos,

⁴ Retirado de <http://public-1.cryst.bbk.ac.uk/PPS2/projects/jonda/structur.htm>

⁵ Retirado de <http://dwb.unl.edu/Teacher/NSF/C08/C08Links/pps99.cryst.bbk.ac.uk>

raramente é feita em linha, devido à sua análise ser, quase sempre, um processo demorado e complexo.

A GFP tem demonstrado um enorme potencial como método de monitorização não destrutivo *in situ* de bioprocessos, não só pelas características já referidas anteriormente mas também porque não apresenta, normalmente, riscos de toxicidade para o hospedeiro (Reischer *et al.*, 2004). Esta proteína tem sido usada para monitorizar em linha a concentração de biomassa usando um sensor de fibra óptica em fermentações de *E. coli* de baixa (Randers-Eichhorn *et al.*, 1997; Kang *et al.*, 2002) e alta densidade celular (Delisa *et al.*, 1999; Chae *et al.*, 2000; Delisa *et al.*, 2001).

Tabela 2.3 Propriedades das proteínas fluorescentes (adaptado de Patterson *et al.*, 2001).

Proteína	Aminoácidos substituídos ⁽ⁱ⁾	E_m ($\text{cm}^{-1}\text{M}^{-1}$)	Y_Q (%)	λ_{EX} (nm)	λ_{Em} (nm)	pH	$t_{\text{branqueamento}}$
GFP	--	9500	~80	395(470)	509(540)	--	--
EGFP	Phe ₆₄ Leu Ser ₆₅ Thr	55000	~60	488	507	5,9	100
EBFP	Phe ₆₄ Leu Tyr ₆₆ His Tyr ₁₄₅ Phe	31000	~25	383	445	5,8	3
ECFP	Phe ₆₄ Leu Ser ₆₅ Thr Tyr ₆₆ Trp Asn ₁₄₆ Ile Met ₁₅₃ Thr Val ₁₆₃ Ala	26000	~40	433(453)	475(501)	4,7	85
EYFP	Ser ₆₅ Gly Val ₆₈ Leu Ser ₇₂ Ala Thr ₂₀₃ Tyr	84000	~61	513	527	6,5	35

(i) Ala - alanina; Asn - asparagina; Gly - glicina; His - histidina; Ile - isoleucina; Leu - leucina; Met - metionina; Phe - fenilalanina; Ser - serina; Thr - treonina; Trp - triptofano; Tyr - tirosina; Val - valina.

E_m - coeficiente de extinção; Y_Q - rendimento quântico; λ_{EX} - comprimento de onda de excitação; λ_{Em} - comprimento de onda de emissão; $t_{\text{branqueamento}}$ - tempo relativo de branqueamento (*photobleaching*).

A GFP tem sido também utilizada, com sucesso, como proteína de fusão, uma vez que tem demonstrado não interferir com inúmeras proteínas às quais se liga quer por N- ou C-terminal (Poppenborg *et al.*, 1997; Cha *et al.*, 2000; Jones *et al.*, 2004). Seo *et al.* (2003) conseguiram detectar níveis de *stress* celular fundindo a GFP com elementos do promotor do gene da proteína induzida por *stress* por choque térmico, em várias estirpes de *E. coli*. Albano *et al.* (1998) e Delisa *et al.* (1999) construíram um plasmídeo que contém a fusão de dois genes repórter, a GFP e a

cloranfenicol-acetyl transferase (CAT). Os seus resultados sugerem uma correlação quantitativa directa entre a intensidade da fluorescência da GFP e a actividade da enzima CAT.

No trabalho desenvolvido por Liu *et al.* (2001) o crescimento celular e a expressão da GFP em culturas de células de tabaco transgénico foram monitorizados, tendo utilizado a fluorescência da GFP para determinar o caudal de alimentação em reactor semi-contínuo.

Estes e muitos outros trabalhos desenvolvidos enfatizam o interesse da GFP como repórter genético *in vivo* e como modelo de proteína, para monitorizar e otimizar a produção de proteínas recombinadas em processos fermentativos (Li *et al.*, 2000; Richards *et al.*, 2003; Jones *et al.*, 2004; Zupan *et al.*, 2004).

2.4. DESCRIÇÃO EXPERIMENTAL

Nesta secção descreve-se o trabalho experimental realizado. Simultaneamente fornecem-se informações relativas ao equipamento e instalação experimental utilizada, ao grau de pureza dos reagentes químicos, ao microrganismo seleccionado, aos métodos analíticos, entre outros.

2.4.1. Fermentação

As fermentações foram realizadas em operação descontínua e semi-contínua com um peso inicial de meio de cultura de 3 kg. Durante as fermentações semi-contínuas foram adicionados no máximo 2 kg de solução de alimentação. No decurso da fermentação os pesos do fermentador e da alimentação foram constantemente monitorizados através de uma balança SB 32001 (Mettler Toledo GmbH, Suíça) e duma balança PM 4800 (Mettler Toledo GmbH, Suíça), respectivamente.

O crescimento do inóculo (0,3 kg), preparado em *erlenmeyer*, decorreu até se atingir uma densidade óptica de aproximadamente 1,3, que de um modo geral demorava cerca de 17 horas, a uma temperatura de 37 °C e a uma velocidade de agitação de 150 rpm.

Todas as fermentações semi-contínuas foram precedidas por uma fase descontínua. O início da alimentação ocorreu quando se observou um aumento do oxigénio dissolvido devido ao consumo total da glucose e do acetato do meio.

Foram realizadas várias fermentações contudo, serão apenas utilizados os resultados de 5 fermentações, que ao longo deste trabalho serão denominadas por: Exp_1, Exp_2, Exp_3, Exp_4 e Exp_5. A fermentação Exp_1 foi utilizada para estudar o comportamento da estirpe de *E. coli* utilizada no modo de operação semi-contínua. Os dados obtidos nesta experiência permitiram definir e determinar, de modo empírico, o modelo cinético e respectivos parâmetros, como será apresentado no Capítulo 3. As fermentações Exp_2 e Exp_3 foram utilizadas no estudo do desenho óptimo de experiências, como será descrito também no Capítulo 3. A fermentação Exp_4 e Exp_5

permitiram a validação experimental dos observadores de estado estudados. Esses resultados serão apresentados no Capítulo 4.

2.4.1.1. *Microorganismo*

O microrganismo utilizado foi a estirpe de *E. coli* M15 (Nal^s, Str^s, Rif^s, Thi⁻, Lac⁻, Ara⁺, Gal⁺, Mtl⁻, F⁻, RecA⁺, Uvr⁺, Lon⁺), (QUIAGEN Inc., USA), derivada da *E. coli* K12, contendo os plasmídeos [pREP4] (QUIAGEN Inc., USA) e [pEYFP] (Clontech, USA). O plasmídeo pREP4, de baixo número de cópias, confere resistência à kanamicina e expressa a proteína repressor *lac*, codificada pelo gene *lac I*, derivando do plasmídeo pACYC. Nas células hospedeiras estão presentes várias cópias deste plasmídeo que asseguram a produção de elevados níveis da proteína repressor *lac* que se liga às sequências do promotor no plasmídeo pEYFP e regula rigorosamente a expressão das proteínas recombinadas. O seu objectivo é manter a expressão em níveis mínimos até se ter atingido uma densidade celular elevada, diminuindo assim a instabilidade dos plasmídeos e minimizando os efeitos do fardo metabólico. A estirpe *E. coli* M15[pREP4] é assim capaz de uma produção de proteínas mais controlada que as estirpes de *E. coli* XL1 Blue, JM109 e TG1, pois produz baixos níveis de expressões na fase pré-indução. Este facto é significativo, uma vez que permite a obtenção de culturas mais estáveis.

A expressão da proteína de fluorescência amarela esverdeada melhorada (EYFP) codificada pelo vector pEYFP (que é o plasmídeo pUC19 derivado do plasmídeo pPD16.43) é rapidamente induzida pela adição de 1 mM de IPTG, que se liga às proteínas repressor *lac* inactivando-as. A EYFP, como referido na secção 2.3, é uma variação da GFP com a substituição de quatro aminoácidos que alteram o espectro de emissão de verde para amarelo esverdeado. O seu nível de fluorescência é aproximadamente igual ao da EGFP. O plasmídeo pEYFP, de elevado número de cópias, confere resistência à ampicilina.

2.4.1.2. *Composição do meio de cultura*

O meio de cultura sintético M9 modificado utilizado para as fermentações está descrito nas Tabelas 2.4 a 2.6. Nessas tabelas indica-se ainda o grau de pureza dos reagentes químicos usados bem como o fabricante.

Na preparação do meio de cultura da fase descontínua, os compostos marcados com (1) na Tabela 2.4 foram dissolvidos em água destilada e esterilizados em autoclave, dentro do reactor, durante 30 minutos à temperatura de 121 °C. A água destilada foi obtida usando um destilador Aquatron (Bibby). O peso total desta solução foi de 2,6 kg. Os compostos (2) foram dissolvidos em água destilada num *erlenmeyer* e esterilizados em autoclave, durante 20 minutos à mesma temperatura da solução anterior, tendo um peso total de 0,37 kg. A solução dos compostos (3), com um peso total de 0,03 kg, foi esterilizada por filtração com acrodiscos estéreis de 0,2 µm de porosidade e 25 mm de diâmetro (Gelman Sciences, EUA). O meio do inóculo e a solução de

alimentação da fase semi-contínua foram preparados recorrendo a um procedimento idêntico ao anteriormente descrito.

Tabela 2.4 Composição do meio de cultura sintético M9 modificado.

Nome	Compostos		Fases da fermentação (g/kg)			Modo de esterilização
	Grau de pureza (\geq)	Marca	Inóculo	Descontínua	Semi-contínua	
Glucose monohidratada	comercial	COPAM	5,0	5,0	250	1
Na ₂ HPO ₄	98,5%	Riedel-de-Haën	6,0	6,0	--	2
KH ₂ PO ₄	98%	Riedel-de-Haën	3,0	3,0	--	2
NH ₄ Cl	99,5%	Panreac	1,0	1,0	10	2
NaCl	99,8%	Riedel-de-Haën	0,50	0,50	--	2
MgSO ₄ .7H ₂ O	99%	Riedel-de-Haën	0,12	0,12	4	3
Tiamina	99%	Fluka	0,34	0,34	--	3
CaCl ₂ .2H ₂ O	99%	Riedel-de-Haën	0,015	0,015	--	3
Kanamicina A	92%	Sigma	0,025	0,025	--	3
Ampicilina	91%	Sigma	0,1	0,1	--	3
Traços de minerais ⁽ⁱ⁾	----	----	--	2,0	--	3
Vitaminas ⁽ⁱ⁾	----	----	--	2,0	--	3

(i) As unidades destas soluções são mL/kg.

Nas Tabelas 2.5 e 2.6 indicam-se as composições das soluções de traços minerais e de vitaminas utilizadas. Estas soluções foram preparadas no laboratório a partir dos compostos químicos mencionados nessas tabelas.

Tabela 2.5 Composição da solução de Traços de minerais.

Compostos	Grau de pureza (\geq)	Marca	Concentração (g/L)
FeCl ₃	99%	Riedel-de-Haën	27
ZnCl ₂	98%	Riedel-de-Haën	2,0
CoCl ₂	98%	Riedel-de-Haën	2,0
NaMoO ₄	99,5%	Merck	2,0
CaCl ₂	99%	Riedel-de-Haën	1,0
CuCl ₂	99%	Riedel-de-Haën	1,0
H ₃ BO ₃	99,8%	Merck	0,50
HCl	95%	Merck	100 ⁽ⁱ⁾

(i) unidade: mL/L.

Tabela 2.6 Composição da solução de Vitaminas.

Compostos	Grau de pureza (\geq)	Marca	Concentração (g/L)
Riboflavina	98%	Riedel-de-Haën	0,42
Ácido Pantoténico	99%	Fluka	5,4
Ácido Nicotínico	99%	Riedel-de-Haën	6,1
Piridoxina	99,5%	Merck	1,4
Biotina	98,5%	Merck	0,060
Ácido Fólico	95%	Merck	0,042

2.4.1.3. Fermentador e Unidade de Medida e Controlo Digital

As fermentações descontínuas e semi-contínuas foram efectuadas num fermentador comercial (Biostat MD, B.Braun, Alemanha) com uma capacidade máxima de 5 litros. O biorreactor está equipado com uma camisa de controlo de temperatura, um motor de agitação, um controlador de caudal mássico para arejamento e sondas para monitorizar a concentração de oxigénio dissolvido, a concentração de dióxido de carbono dissolvido, o pH e a temperatura. O reactor está ligado a uma unidade de Medida e Controlo Digital directo (DCU - *Digital Control Unit*) onde algumas dessas variáveis podem ser controladas. Cada variável tem associado um ciclo de controlo, com parâmetros adequados e pré-definidos, que podem ser alterados directamente pelo operador ou por comunicação remota através do computador de supervisão. A bomba peristáltica (modelo 101 U/R, Watson Marlow, Reino Unido) utilizada para adicionar a solução de alimentação na fase semi-contínua da fermentação, está também ligada à DCU por uma porta analógica I/O de 25 *pins* e a sua velocidade de rotação é uma das variáveis manipuladas. Na Tabela 2.7 encontram-se descritos os componentes do fermentador. Tal como referido anteriormente (secção 2.4.1) os pesos do fermentador e da alimentação são monitorizados, recorrendo a balanças apropriadas (Tabela 2.7), e registados a cada 3 minutos, sendo possível, caso necessário, reduzir esse tempo de aquisição.

A DCU permite outras funções como a amplificação dos sinais dos eléctrodos e a sua calibração. Tem também a seu cargo a actuação nos elementos finais de controlo, nomeadamente nas bombas de adição de ácido e base para controlo do pH, no aquecimento eléctrico da água da camisa de controlo da temperatura, na velocidade de agitação e/ou no caudal de arejamento para controlo da concentração de oxigénio dissolvido.

O sistema de fermentação é constituído ainda por uma outra unidade responsável pelo fornecimento de potência para agitação, para aquecimento da água e para bombeamento de fluidos. Inclui bombas para circulação/adição de fluidos como a água da camisa de controlo da temperatura, soluções ácida e alcalina para controlo do pH, agente anti-espuma e ar.

Tabela 2.7 Descrição dos principais componentes do fermentador.

Componente	Descrição
Fermentador BIOSTAT MD (B. Braun, Alemanha)	vaso de cultura em vidro de boro-silicato com camisa externa para controlo de temperatura e várias portas para introdução de eléctrodos; capacidade máxima de 5 litros
Agitação	veio de agitação com 3 impulsores de 6 lâminas cada; motor com potência de 55 W
Sonda de temperatura	termómetro de platina (Pt-100)
Sonda de pH InPro 3100/325 Pt 100 (Mettler Toledo, Suíça)	eléctrodo de vidro esterilizável; medição potenciométrica
Sonda de oxigénio dissolvido (Mettler Toledo, Suíça)	tipo polarográfico; mede a pressão parcial (tensão) de oxigénio
Sonda de dióxido de carbono dissolvido InPro 5000/12/320 (Mettler Toledo, Suíça)	medição potenciométrica
Sistema de arejamento (alimentação de ar comprimido)	filtro 1 µm (Swagelok, EUA); medidor/controlador de caudal de ar Hi-Tec F201C-FB (Bronkhorst, Holanda); filtro de membrana ACRO 50 (Pall, EUA)
Sistema do gás de saída	condensador do reactor; 2 condensadores do tipo <i>Dimroth</i> ; 1 coluna de sílica; 1 coluna de pedra porosa
Controlo de pH	2 bombas peristálticas para adição de solução ácida e alcalina
Sistema de alimentação de substrato	bomba peristáltica de velocidade variável (Wastson Marlow, Reino Unido); balança PM 4800 (Mettler Toledo GmbH, Suíça) com precisão: 0,001 (0-1 kg) ou 0,1 (1-4 kg)
Balança SB 32001 (Mettler Toledo GmbH, Suíça) para monitorizar o peso total do reactor	capacidade 32 kg, precisão 0,1 g

O ar fornecido ao fermentador é obtido pelo sistema de ar comprimido do laboratório e é controlado por um medidor/controlador de caudal mássico (modelo Hi-Tec F201C-FB, Bronkhorst,

Holanda) instalado na unidade descrita anteriormente. O sistema do ar de entrada é ainda constituído por duas unidades de filtração: um filtro de 1 μm (Swagelok, EUA) localizado antes do controlador mássico para remover partículas e um filtro de membrana hidrofóbica autoclavável de 0,2 μm (ARO 50, Pall, EUA) localizado antes da entrada do fermentador para esterilizar o ar.

À saída do fermentador, o gás passa por três condensadores, um do reactor e os outros dois do tipo *Dimroth*, para remoção de humidade. Este sistema inclui ainda uma coluna de sílica e outra de pedra porosa para remoção final da humidade. A tubagem é em polietileno e aço *inox* e tem um comprimento aproximado de 8 metros. O gás é assim encaminhado para o espectrómetro de massa Dymaxion (DM 200M) (Ametek, EUA) onde será analisado.

A Figura 2.5 mostra uma fotografia da instalação experimental utilizada para a realização das experiências descritas anteriormente.

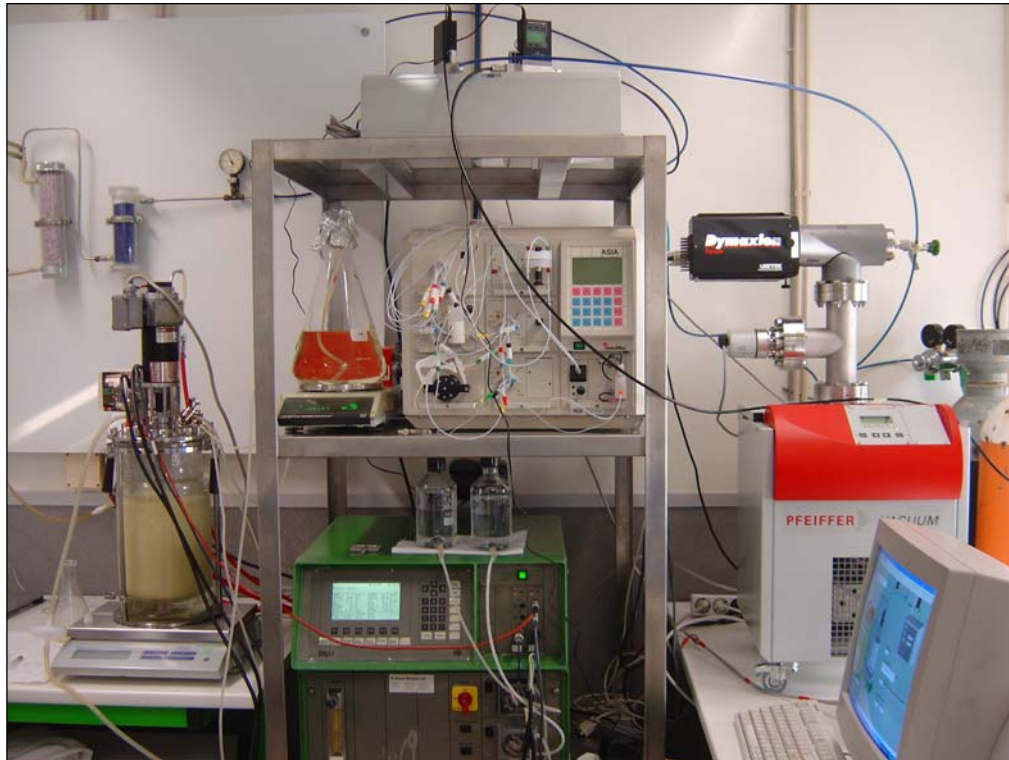


Figura 2.5 Fotografia da instalação experimental.

2.4.1.4. Condições experimentais

As fermentações foram efectuadas com os valores de referência para o controlo das variáveis ambientais indicados na Tabela 2.8. Foram utilizadas soluções aquosas (diluição 1:4) de hidróxido de amónio 25% e de ácido fosfórico 85%, para controlar o pH do meio a $7,0(\pm 0,1)$. A concentração de oxigénio dissolvido foi mantida a 30% do valor de saturação por manipulação do caudal de ar de entrada e da velocidade de agitação. Numa primeira fase, a variável manipulada foi o caudal de ar,

que após atingir o seu valor máximo faz alterar a variável manipulada para a velocidade de agitação. De modo a assegurar um caudal de ar e uma velocidade de agitação mínimos dentro do reactor estipularam-se como valores mínimos de referência 1 L.min⁻¹ e 100 rpm, respectivamente. O controlo das variáveis ambientais foi efectuado automaticamente pela DCU. Para controlar o nível de espuma foi adicionado, quando necessário, um agente anti-espuma (Merck).

Tabela 2.8 Valores de referência para controlo das variáveis ambientais.

Variável	Valor de referência	Acção de controlo ⁽ⁱ⁾
Temperatura	37 °C	PID
pH	7,0	PID
Oxigénio dissolvido	30%	Cascata

(i) P - Proporcional; I - Integral; D - Derivativa.

2.4.2. Métodos de análise em linha

A detecção e quantificação em linha dos gases de saída do fermentador, nomeadamente oxigénio, dióxido de carbono e azoto, e a determinação do *OTR* (*Oxygen Transfer Rate*) e *CTR* (*Carbon Transfer Rate*), foi efectuada através de um espectrómetro de massa (MS) Dymaxion (DM 200M) (Ametek, EUA) do tipo quadripolo, com uma gama de massas de 200 Da (u.m.a). O aparelho é constituído por um sistema de vácuo, um analisador de gás quadripolo e uma entrada capilar.

O gás de saída do fermentador, após desumidificação, é introduzido no espectrómetro sob vácuo (<10⁻⁵ mbar) através de um capilar. A ionização é conseguida por impacto de electrões que são emitidos por um filamento aquecido. A separação dos iões produzidos com diferentes massas é conseguida através do filtro de massas quadripolar e a sua detecção efectuada através de um detector do tipo *taça Faraday* em que os iões são direccionados para um elemento em forma de taça sendo medida a própria corrente iónica que é proporcional à correspondente pressão parcial molecular na fonte. O sistema está ainda equipado com um segundo detector do tipo multiplicador de electrões.

Para controlar o aparelho e aquisição de dados utilizou-se o *software Dycor System 2000*, marca Ametek, versão 1.66, a correr no computador de supervisão em ambiente Windows XPTM. O computador comunica, em série, com a unidade de controlo do espectrómetro via protocolo RS-232. Este programa permite a troca de dados com outro programa que esteja a ser executado paralelamente no mesmo computador, através da implementação do protocolo DDE (*Dynamic Data Exchange*) que constitui um padrão para troca de dados em ambiente *Windows*.

A calibração do MS foi efectuada antes de cada fermentação utilizando misturas de gases de composição conhecida. A composição das misturas gasosas utilizadas é apresentada na Tabela 2.9, as quais foram escolhidas de modo a que a fracção molar englobe a composição do ar de entrada e saída do reactor (Ferreira *et al.*, 1998). As misturas gasosas foram obtidas através de compostos puros contidos em garrafas pressurizadas de aço (Praxair, Inc., Espanha) impondo diferentes caudais

controlados por controladores de caudal mássico (HFC 202, Teledyne Hastings Inc., EUA e MC-100SCCM-D, Alicat Scientific, Inc., EUA).

Tabela 2.9 Composição das misturas utilizadas na calibração do MS.

Misturas	Oxigénio (% molar)	Dióxido de carbono (% molar)	Azoto (% molar)
1	22,00	1,55	76,45
2	21,47	0,00	78,53
3	20,53	3,08	76,39
4	18,40	3,07	78,53
5	10,20	10,20	79,59
6	25,51	5,10	69,39

Para a calibração efectuou-se a análise das misturas gasosas e determinou-se a média dos valores da pressão obtidos para cada massa, em que as massas 28, 32 e 44 foram usadas para quantificar o azoto, oxigénio e dióxido de carbono, respectivamente. Para cada massa efectuou-se uma regressão linear entre a média da pressão obtida e a respectiva composição na mistura gasosa, obtendo-se assim uma correlação entre pressão e composição para cada gás analisado. No Anexo A é apresentado um exemplo da calibração do MS.

2.4.3. Análises em diferido

Durante a fermentação, as amostras foram retiradas manualmente, de hora a hora, através de uma porta *ad hoc* do fermentador, para análise em diferido. A análise em diferido das variáveis de estado foi realizada não só para a sua monitorização e posterior utilização no estudo desenvolvido no Capítulo 3, mas também para validar os resultados obtidos para os sensores por programação descritos no Capítulo 4.

Cada amostra foi dividida, sendo uma parte utilizada na determinação da concentração de biomassa, e a outra, após centrifugação, durante 20 minutos a 3000 rpm numa centrífuga CENCOM (J. P. Selecta s.a., Espanha) e filtração com filtros descartáveis GyroDisc Syr. CA-PC (Orange Scientific, Belgica) de porosidade de 0,2 μm e 30 mm de diâmetro, foi armazenada a -20 °C, até ser analisada, para determinação da concentração de glucose e acetato.

2.4.3.1. Determinação da concentração de biomassa

A determinação da concentração de biomassa das amostras foi efectuada, em duplicado, por uma medida equivalente determinada por leitura da densidade óptica das amostras a 600 nm (DO_{600}) num espectrofotómetro UV/Vis V-560 (Jasco, Inc., EUA) e correlacionando esses valores com o peso seco. Como branco foi usada água destilada. Em geral, as amostras tiveram que ser diluídas com

água destilada para adequar a medição à gama linear de absorvância (0 a 0,6 unidades de absorvância).

A concentração de biomassa foi calculada recorrendo à curva de calibração de densidade óptica *versus* peso seco. Para cada fermentação realizada foi efectuada uma nova calibração, encontrando-se um exemplo descrito no Anexo A.

A curva de calibração de densidade óptica *versus* peso seco é determinada por leitura da densidade óptica de suspensões, em duplicado, obtidas por diluição de uma suspensão de células de densidade óptica e de concentração em peso seco conhecidas. A densidade óptica destas suspensões está compreendida entre 0,1 e 0,6 unidades de absorvância. A determinação da concentração em peso seco da suspensão celular inicial é efectuada gravimetricamente do seguinte modo:

- (1) colocar seis filtros com tamanho de poro de 0,45 µm em placas de petri numeradas, na estufa a 105 °C durante a noite (até peso constante);
- (2) deixar arrefecer os filtros num excicador durante aproximadamente 15 minutos;
- (3) pesar os filtros numa balança analítica;
- (4) recolher aproximadamente 50 mL de meio do fermentador e manter em gelo;
- (5) filtrar um determinado peso certo de cultura em cada filtro e lavar com igual peso de água destilada;
- (6) colocar novamente os filtros na estufa a 105 °C durante 24 horas (até peso constante);
- (7) repetir os pontos (2) e (3).

O peso seco da biomassa obtém-se por diferença entre o valor do peso do filtro com a biomassa e o seu peso inicial. A concentração em peso seco da suspensão celular inicial é determinada a partir do peso seco médio das seis medições efectuadas, pela seguinte expressão:

$$X_{\text{cultura}} (\text{g/kg}) = \frac{\text{peso seco médio de biomassa (g)}}{\text{peso filtrado (g)}} \times 1000 \quad \text{Equação 2.4}$$

A concentração em peso seco das outras suspensões celulares obtém-se por:

$$X_{\text{suspensões}} (\text{g/kg}) = \frac{\text{volume adicionado da suspensão inicial (mL)}}{\text{volume final da suspensão (mL)}} \times X_{\text{cultura}} (\text{g/kg}) \quad \text{Equação 2.5}$$

Com base nos valores determinados para a DO_{600} e $X_{\text{suspensões}}$ construiu-se a curva de calibração.

2.4.3.2. Determinação da concentração de acetato

A concentração de acetato foi determinada por Cromatografia Líquida de Alta Resolução (HPLC). As análises foram realizadas num sistema cromatográfico constituído por um cromatógrafo marca

Jasco (Japão), equipado com uma bomba modelo 880-Pu e um injector automático AS-2057 Plus, provido de um *loop* de 20 µL. A detecção foi feita simultaneamente com um detector UV/Vis Jasco, modelo 870-UV, de comprimento de onda variável e sensibilidade de 0,005 AUFS e um detector por índice de refração Jasco, modelo 830-RI. Na aquisição e tratamento dos dados utilizou-se o *software Star Chromatography Workstation*, marca Varian, versão 6.30.

A separação cromatográfica foi conseguida usando uma coluna cromatográfica de ácidos orgânicos Chrompack (modelo 25 7001) constituída por uma resina de exclusão catiónica H⁺ (300×7,7 d.i.) e uma pré-coluna com o mesmo enchimento da coluna cromatográfica. A coluna foi colocada no interior de um forno Chrompack HCT (Holanda).

Os eluentes foram filtrados usando membranas GN-6 Metrice[®] Grid (Pall Life Sciences, EUA) de porosidade 0,45 µm e diâmetro 47 mm e desgasificados, durante pelo menos 20 minutos, num banho ultra-sons, marca Sonicor SC-52, antes de serem utilizados.

A análise foi efectuada após a injeção de 20 µL de amostra usando como eluente uma solução 0,01 N de ácido sulfúrico. A eluição foi realizada a um caudal constante de 0,5 mL/min, à temperatura de 35 °C. Cada amostra foi injectada em duplicado.

A curva de calibração para a determinação da concentração de acetato foi obtida a partir 6 soluções padrão cujas concentrações variaram entre 0,5 e 10 gramas de acetato por quilograma de solução. Cada uma das soluções padrão foi injectada em triplicado. No Anexo A apresenta-se um exemplo de uma calibração.

2.4.3.3. *Determinação da concentração de glucose*

A determinação da concentração de glucose foi realizada por HPLC, nas mesmas condições descritas na análise do acetato. A curva de calibração para a determinação da concentração de glucose foi obtida a partir de 6 soluções padrão cujas concentrações variavam entre 0,5 e 10 gramas de glucose por quilograma de solução. Cada uma das soluções padrão foi injectada em triplicado. Apresenta-se no Anexo A um exemplo de uma calibração.

A concentração de glucose foi ainda determinada com o *kit* enzimático Glucose Trinder GOD-POD (SPINREACT, Espanha) de acordo com as especificações do fornecedor. As amostras foram previamente cozidas a 80 °C durante 15 minutos. Para a determinação da curva de calibração foram preparadas soluções padrão com concentração compreendida entre 0,1 e 5 g/kg. Posteriormente, e após agitação da microplaca contendo as soluções a analisar, procedeu-se à leitura da densidade óptica das amostras (em duplicado) e dos padrões (em triplicado) num leitor de microplacas ELISA Synergy HT (Bio-Tek Instruments, inc., EUA), a um comprimento de onda de 505 nm. É apresentado no Anexo A um exemplo da calibração deste *kit*.

2.4.4. Sistema informático

O sistema de aquisição de dados e controlo encontra-se descrito em pormenor em Rocha (2003) tendo sido desenvolvido em ambiente *LabVIEW Professional Development System*, versão 6.1. Corre num computador pessoal, com processador Intel®, Celeron™, 1,2 GHz, com 512 MB de memória RAM, em ambiente Windows XP™. A comunicação entre a DCU e o computador é feita através do protocolo RS-232.

Os resultados obtidos por computador e a maior parte dos gráficos apresentados nesta tese foram gerados em MATLAB versão 7.1 (The MathWorks, Inc., EUA). A simulação do modelo matemático do processo em estudo foi realizada por integração das equações diferenciais que o representam (Capítulo 3) através da subrotina ODE23s. A implementação dos observadores e estimadores descritos no Capítulo 4 e da lei de controlo descrita no Capítulo 5 foi efectuada pelo método de integração de Euler. A maior parte das operações matemáticas efectuadas quer no Capítulo 3, na identificação dos coeficientes de rendimento, quer no Capítulo 4, no desenho dos observadores e estimadores, foram realizadas usando a rotina *Symbolic Math Toolbox* que corre em MATLAB. A rotina *Genetic and Evolutionary Algorithm Toolbox* (GEAtbx 3.5) desenvolvida por Pohlheim (2004) foi utilizada como método de optimização para a determinação dos perfis do caudal de alimentação, no desenho óptimo de experiências apresentado no Capítulo 3.

A análise estatística apresentada no Capítulo 3 foi efectuada recorrendo ao *software -Statistical Package for the Social Sciences - SPSS 11.5*.

2.5. BIBLIOGRAFIA

Alba, M.J.G. e Calvo, E.G., (2000). Characterization of bioreaction processes: aerobic *Escherichia coli* cultures. *Journal of Biotechnology*, 84 (2), 107-118.

Albano, C.R., Randers-Eichhorn, L., Bentley, W.E. e Rao, G., (1998). Green fluorescent protein as a real time quantitative reporter of heterologous protein production. *Biotechnology Progress*, 14, 351-354.

Bailey, J.E. e Ollis, D.F., (1986). *Biochemical Engineering Fundamentals*, McGraw-Hill Book Co., Singapore.

Cannizzaro, C., Valentinotti, S. e von Stockar, U., (2004). Control of yeast fed-batch process through regulation of extracellular ethanol concentration. *Bioprocess and Biosystems Engineering*, 26 (6), 377-383.

Carlson, R. e Srienc, F., (2004). Fundamental *Escherichia coli* biochemical pathways for biomass and energy production: Identification of reactions. *Biotechnology and Bioengineering*, 85 (1), 1-19.

Cha, H.J., Wu, C.F., Valdes, J.J., Rao, G. e Bentley, W.E., (2000). Observations of green fluorescent protein as a fusion partner in genetically engineered *Escherichia coli*: Monitoring protein expression and solubility. *Biotechnology and Bioengineering*, 67 (5), 565-574.

- Chae, H.J., Delisa, M.P., Cha, H.J., Weigand, W.A., Rao, G. e Bentley, W.E., (2000). Framework for online optimization of recombinant protein expression in high-cell-density *Escherichia coli* cultures using GFP-fusion monitoring. *Biotechnology and Bioengineering*, 69 (3), 275-285.
- Delisa, M.P., Chae, H.J., Weigand, W.A., Valdes, J.J., Rao, G. e Bentley, W.E., (2001). Generic model control of induced protein expression in high cell density cultivation of *Escherichia coli* using on-line GFP-fusion monitoring. *Bioprocess and Biosystems Engineering*, 24 (2), 83-91.
- Delisa, M.P., Li, J.C., Rao, G., Weigand, W.A. e Bentley, W.E., (1999). Monitoring GFP-operon fusion protein expression during high cell density cultivation of *Escherichia coli* using an on-line optical sensor. *Biotechnology and Bioengineering*, 65 (1), 54-64.
- Dittrich, C.R., Bennett, G.N. e San, K.Y., (2005). Characterization of the acetate-producing pathways in *Escherichia coli*. *Biotechnology Progress*, 21 (4), 1062-1067.
- Edwards, J.S. e Palsson, B.O., (2000). Robustness analysis of the *Escherichia coli* metabolic network. *Biotechnology Progress*, 16 (6), 927-939.
- Ferreira, B.S., van Keulen, F. e Fonseca, M.M.R., (1998). Novel calibration method for mass spectrometers for on-line gas analysis. Set-up for the monitoring of a bacterial fermentation. *Bioprocess Engineering*, 19, 289-296.
- Flickinger, M. e Rouse, M.P., (1993). Sustaining protein synthesis in the absence of rapid cell division: an investigation of plasmid-encoded protein expression in *Escherichia coli* during very slow growth. *Biotechnology Progress*, 9, 555-572.
- Gerdes, H.H. e Kaether, C., (1996). Green fluorescent protein: Applications in cell biology. *Febs Letters*, 389 (1), 44-47.
- Gill, R.T., Valdes, J.J. e Bentley, W.E., (2000). A comparative study of global stress gene regulation in response to overexpression of recombinant proteins in *Escherichia coli*. *Metabolic Engineering*, 2 (3), 178-189.
- Han, K., Lim, H.C. e Hong, J., (1992). Acetic acid formation in *Escherichia coli* fermentation. *Biotechnology and Bioengineering*, 39 (6), 663-671.
- Holmberg, A., (1982). On the Practical Identifiability of Microbial-Growth Models Incorporating Michaelis-Menten Type Nonlinearities. *Mathematical Biosciences*, 62 (1), 23-43.
- Holms, H., (1996). Flux analysis and control of the central metabolic pathways in *Escherichia coli*. *Fems Microbiology Reviews*, 19 (2), 85-116.
- Hu, S.Y., Wu, J.L. e Huang, J.H., (2004). Production of tilapia insulin-like growth factor-2 in high cell density cultures of recombinant *Escherichia coli*. *Journal of Biotechnology*, 107 (2), 161-171.
- Huang, R.C., Lam, E., Chen, Y.H., Hackett, J., Lam, T.L., Liu, D., Ma, M.C., Siu, K.L., Sivakesava, S., Xu, Z.N., Wong, R.S.C. e Wong, W.K.R., (1999). Human epidermal growth factor excreted by recombinant *Escherichia coli* K-12 has the correct N-terminus and is fully bioactive. *Process Biochemistry*, 35 (1-2), 1-5.
- Jeong, K.J., Choi, J.H., Yoo, W.M., Keum, K.C., Yoo, N.C., Lee, S.Y. e Sung, M.H., (2004). Constitutive production of human leptin by fed-batch culture of recombinant rpoS(-) *Escherichia coli*. *Protein Expression and Purification*, 36 (1), 150-156.
- Jeong, K.J. e Lee, S.Y., (1999). High-level production of human leptin by fed-batch cultivation of recombinant *Escherichia coli* and its purification. *Applied and Environmental Microbiology*, 65 (7), 3027-3032.

Johnston, W., Cord-Ruwisch, R. e Cooney, M.J., (2002). Industrial control of recombinant *E. coli* fed-batch culture: new perspectives on traditional controlled variables. *Bioprocess and Biosystems Engineering*, 25 (2), 111-120.

Jones, J.J., Bridges, A.M., Fosberry, A.P., Gardner, S., Lowers, R.R., Newby, R.R., James, P.J., Hall, R.M. e Jenkins, O., (2004). Potential of real-time measurement of GFP-fusion proteins. *Journal of Biotechnology*, 109 (1-2), 201-211.

Kang, D.G., Kim, Y.K. e Cha, H.J., (2002). Comparison of green fluorescent protein expression in two industrial *Escherichia coli* strains, BL21 and W3110, under co-expression of bacterial hemoglobin. *Applied Microbiology and Biotechnology*, 59 (4-5), 523-528.

Karp, P.D., Riley, M., Saier, M., Paulsen, I.T., Collado-Vides, J., Paley, S.M., Pellegrini-Toole, A., Bonavides, C. e Gama-Castro, S., (2002). The EcoCyc Database. *Nucleic Acids Research*, 30 (1), 56-58.

Kim, B.S., Lee, S.C., Lee, S.Y., Chang, Y.K. e Chang, H.N., (2004). High cell density fed-batch cultivation of *Escherichia coli* using exponential feeding combined with pH-stat. *Bioprocess and Biosystems Engineering*, 26 (3), 147-150.

Kleist, S., Miksch, G., Hitzmann, B., Arndt, M., Friehs, K. e Flaschel, E., (2003). Optimization of the extracellular production of a bacterial phytase with *Escherichia coli* by using different fed-batch fermentation strategies. *Applied Microbiology and Biotechnology*, 61, 456-462.

Ko, Y.-F., Bentley, W.E. e Weigand, W.A., (1994). A metabolic model of cellular energetics and carbon flux during aerobic *Escherichia coli* fermentation. *Biotechnology and Bioengineering*, 43, 847-855.

Korz, D.J., Rinas, U., Hellmuth, K., Sanders, E.A. e Deckwer, W.-D., (1995). Simple fed-batch technique for high cell density cultivation of *Escherichia coli*. *Journal of Biotechnology*, 39 (1), 59-65.

Lee, S.Y., (1996). High cell-density culture of *Escherichia coli*. *Trends in Biotechnology*, 14 (3), 98-105.

Li, J.C., Wang, S., VanDusen, W.J., Schultz, L.D., George, H.A., Herber, W.K., Chae, H.J., Bentley, W.E. e Rao, G., (2000). Green fluorescent protein in *Saccharomyces cerevisiae*: Real-time studies of the GAL1 promoter. *Biotechnology and Bioengineering*, 70 (2), 187-196.

Li, X.L. e Taylor, K.B., (1994). Effect of Glucose on the Expression Parameters of Recombinant Protein in *Escherichia coli* During Batch Growth in Complex Medium. *Biotechnology Progress*, 10 (2), 160-164.

Lima, N. e Mota, M., (2003). *Biotecnologia: Fundamentos e Aplicações*, Lidel.

Lin, H., Bennet, G.N. e San, K.-Y., (2005). Fed-Batch culture of a metabolically engineered *Escherichia coli* strain designed for high-level succinate production and yield under aerobic conditions. *Biotechnology and Bioengineering*, 90 (6), 775-779.

Liu, S., Bugos, R.C., Dharmasiri, N. e Su, W.W., (2001). Green fluorescent protein as a secretory reporter and a tool for process optimization in transgenic plant cell cultures. *Journal of Biotechnology*, 87, 1-16.

March, J.C., Rao, G. e Bentley, W.E., (2003). Biotechnological applications of green fluorescent protein. *Applied Microbiology and Biotechnology*, 62, 303-315.

- Margolin, W., (2000). Green fluorescent protein as a reporter for macromolecular localization in bacterial cells. *Methods*, 20, 62-72.
- Martinez-Antonio, A., Salgado, H., Gama-Castro, S., Gutierrez-Rios, R.M., Jimenez-Jacinto, V. e Collado-Vides, J., (2003). Environmental conditions and transcriptional regulation in *Escherichia coli*: A physiological integrative approach. *Biotechnology and Bioengineering*, 84 (7), 743-749.
- Mosrati, R., Nancib, N. e Boudrant, J., (1993). Variation and modeling of the probability of plasmid loss as a function of growth rate of plasmid-bearing cells of *Escherichia coli* during continuous cultures. *Biotechnology and Bioengineering*, 41, 395-404.
- Narciandi, R.E., (1996). High-level production of p36 from HIV-2 in fed-batch culture of recombinant *Escherichia coli*. *Journal of Fermentation and Bioengineering*, 81 (4), 360-362.
- Neidhardt, F.C., Ingraham, J.L. e Schaechter, M., (1990). *Physiology of the bacterial cell - a molecular approach*, Sinauer Associates, Sunderland, USA.
- Noronha, S.B., Yeh, H.J.C., Spande, T.F. e Shiloach, J., (2000). Investigation of the TCA cycle and the glyoxylate shunt in *Escherichia coli* BL21 and JM109 using C-13-NMR/MS. *Biotechnology and Bioengineering*, 68 (3), 316-327.
- Patnaik, P.R., (2002). Comparative evaluation of batch and fed-batch bioreactors for GAPDH production by recombinant *Escherichia coli* with distributed plasmid copy number. *Chemical Engineering Journal*, 87 (3), 357-366.
- Patterson, G., Day, R.N. e Piston, D., (2001). Fluorescent protein spectra. *Journal of Cell Science*, 114 (5), 837-838.
- Phue, J.-N., Noronha, S.B., Hattacharyya, R., Wolfe, A.J. e Shiloach, J., (2005). Glucose metabolism at high density growth of *E. coli* B and *E. coli* K: Differences in metabolic pathways are responsible for efficient glucose utilization in *E. coli* B as determined by microarrays and northern blot analyses. *Biotechnology and Bioengineering*, 90 (7), 805-820.
- Phue, J.-N. e Shiloach, J., (2004). Transcription levels of key metabolic genes are the cause for different glucose utilization pathways in *E. coli* B (BL21) and *E. coli* K (JM109). *Journal of Biotechnology*, 109 (1-2), 21-30.
- Pohlheim, H., (2004). Documentation 3.5a for Genetic and Evolutionary Algorithm Toolbox for use with Matlab: GEATbx version 3.5
- Poppenborg, L., Friehs, K. e Flaschel, E., (1997). The green fluorescent protein is a versatile reporter for bioprocess monitoring. *Journal of Biotechnology*, 58 (2), 79-88.
- Prytz, I., Sanden, A.M., Nystrom, T., Farewell, A., Wahlstrom, A., Forberg, C., Pragai, Z., Barer, M., Harwood, C. e Larsson, G., (2003). Fed-batch production of recombinant β -galactosidase using the universal stress promoters *uspA* and *uspB* in high cell density cultivations. *Biotechnology and Bioengineering*, 83 (5), 595-603.
- Randers-Eichhorn, L., Albano, C.R., Sipior, J., Bentley, W.E. e Rao, G., (1997). On-line green fluorescent protein sensor with LED excitation. *Biotechnology and Bioengineering*, 55 (6), 921-926.
- Reischer, H., Schotola, I., Striedner, G., Potschacher, F. e Bayer, K., (2004). Evaluation of the GFP signal and its aptitude for novel on-line monitoring strategies of recombinant fermentation processes. *Journal of Biotechnology*, 108 (2), 115-125.

Richards, H.A., Halfhill, M.D., Millwood, R.J. e Stewart, C.N., (2003). Quantitative GFP fluorescence as an indicator of recombinant protein synthesis in transgenic plants. *Plant Cell Reports*, 22 (2), 117-121.

Rocha, I., (2003). *Model-based strategies for computer-aided operation of recombinant E. coli fermentation*. Tese de Doutorado, Universidade do Minho, Portugal.

Rothen, S.A., Sauer, M., Sonnleitner, B. e Witholt, B., (1998). Growth characteristics of *Escherichia coli* HB101 [pGEC47] on defined medium. *Biotechnology and Bioengineering*, 58 (1), 92-100.

Ryan, W. e Parulekar, S.J., (1990). Effects of culture conditions on plasmid stability and production of a plasmid-encoded protein in batch and continuous cultures of *Escherichia coli* JM103[pUC8]. *Annals of New York Academy of Sciences*, 589, 91-110.

Seo, J.H., Kang, D.G. e Cha, H.J., (2003). Comparison of cellular stress levels and green-fluorescent-protein expression in several *Escherichia coli* strains. *Biotechnology Applied Biochemistry*, 37 (Pt 2), 103-107.

Shiloach, J. e Fass, R., (2005). Growing *E. coli* to high cell density - A historical perspective on method development. *Biotechnology Advances*, 23 (5), 345-357.

Su, W.W., (2005). Fluorescent proteins as tools to aid protein production. *Microbial Cell Factories*, 4.

Suárez, D.C. e Kilikian, B.V., (2000). Acetic acid accumulation in aerobic growth of recombinant *Escherichia coli*. *Process Biochemistry*, 35 (9), 1051-1055.

Timmons, L., Becker, J., Barthmaier, P., Fyrberg, C., Shearn, A. e Fyrberg, E., (1997). Green fluorescent protein/beta-galactosidase double reporters for visualizing *Drosophila* Gene expression patterns. *Developmental Genetics*, 20, 338-347.

Van de Walle, M. e Shiloach, J., (1998). Proposed mechanism of acetate accumulation in two recombinant *Escherichia coli* strains during high density fermentation. *Biotechnology and Bioengineering*, 57 (1), 71-78.

Varma, A., Boesch, B. e Palsson, B.O., (1993). Biochemical production capabilities of *Escherichia coli*. *Biotechnology and Bioengineering*, 42, 59-73.

Varma, A. e Palsson, B.O., (1993b). Metabolic capabilities of *Escherichia coli*: I. Synthesis of biosynthetic precursors and cofactors. *Journal of Theoretical Biology*, 165, 477-502.

Varma, A. e Palsson, B.O., (1993a). Metabolic capabilities of *Escherichia coli*: II. Optimal growth patterns. *Journal of Theoretical Biology*, 165, 503-522.

Versyck, K.J., (2000). *Dynamic Input Design for Optimal Estimation of Kinetic Parameters in Bioprocess Models*. Tese de Doutorado, Universidade Católica de Leuven, Bélgica.

Wangsa-Wirawan, N., Lee, Y., Falconer, R., Mansell, C., O'Neill, B. e Middelberg, A., (1997). Novel fed-batch strategy for the production of insulin-like growth factor 1 (IGF-1). *Biotechnology Letters*, 19 (11), 1079-1082.

Wong, H.H., Kim, Y.C., Lee, S.Y. e Chang, H.N., (1998). Effect of post-induction nutrient feeding strategies on the production of bioadhesive protein in *Escherichia coli*. *Biotechnology and Bioengineering*, 60 (3), 271-276.

Xu, B., Jahic, M., Blomsten, G. e Enfors, S.O., (1999). Glucose overflow metabolism and mixed-acid fermentation in aerobic large-scale fed-batch processes with *Escherichia coli*. *Applied Microbiology and Biotechnology*, 51 (5), 564-571.

Yamanè, T. e Shimizu, S., (1984). Fed-batch techniques in microbial processes. *Advances in Biochemical Engineering / Biotechnology*, 30, 148-194.

Yang, F., Moss, L.G. e Phillips, G.N.Jr., (1996). The molecular structure of green fluorescent protein. *Nature Biotechnology*, 14, 1246-1251.

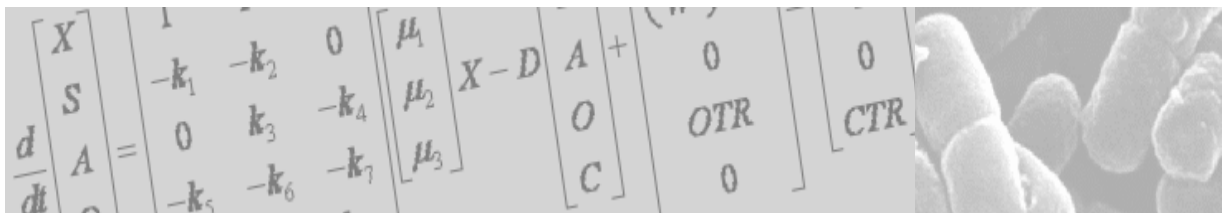
Yuan, C., Reuland, J.M., Lee, L. e Huang, A.J.W., (2004). Optimized expression and refolding of human keratopithelin in BL21 (DE3). *Protein Expression and Purification*, 35 (1), 39-45.

Zhao, J. e Shimizu, K., (2003). Metabolic flux analysis of *Escherichia coli* K12 grown on C¹³-labeled acetate and glucose using GG-MS and powerful flux calculation method. *Journal of Biotechnology*, 101 (2), 101-117.

Zupan, A.L., Trobec, S., Gaberc-Porekar, V. e Menart, V., (2004). High expression of green fluorescent protein in *Pichia pastoris* leads to formation of fluorescent particles. *Journal of Biotechnology*, 109 (1-2), 115-122.

CAPÍTULO 3

IDENTIFICAÇÃO DO MODELO



3. IDENTIFICAÇÃO DO MODELO

Os modelos matemáticos têm adquirido cada vez mais importância em biotecnologia, sendo já uma ferramenta indispensável no estudo de bioprocessos.

O objectivo da modelação matemática consiste em conseguir um modelo que, apesar da sua simplicidade, seja capaz de atingir a finalidade para a qual está a ser formulado, sem no entanto, perder aplicabilidade e assim evitar que o seu processo de identificação e validação seja mais complexo do que o necessário.

A simulação de bioprocessos tem suscitado um grande interesse, quer no meio académico quer industrial, apresentando-se como um instrumento efectivo no projecto e desenvolvimento de algoritmos robustos, na optimização, monitorização, controlo e caracterização desses processos. A formulação do modelo matemático, sendo uma etapa essencial na implementação de um processo, permite modelar o sistema e simular as suas condições de operação com menor esforço experimental, uma vez que a modelação é normalmente um processo mais rápido, em biotecnologia, que as experiências, que são, em geral, demoradas e dispendiosas. Contudo, como grande parte dos sistemas biológicos têm dinâmicas complexas e estão sujeitos a várias perturbações, é necessário recorrer também a dados experimentais para além do conhecimento baseado em leis físicas. Os dados experimentais são por isso, pré-requisitos para o projecto e aplicação de modelos biológicos. De facto, esses dados são necessários em quase todas as etapas de modelação de bioprocessos, as quais se centram na selecção da estrutura do modelo e sua calibração ou identificação dos seus parâmetros.

A aplicação dos conceitos de identificação de sistemas, baseados na Teoria de Sistemas, no campo da biotecnologia, remonta aos anos oitenta do século passado (Holmberg, 1982; Munack, 1984; Godfrey e DiStefano III, 1985; Espie e Macchietto, 1989; Baltes *et al.*, 1994; Chouakri *et al.*, 1994a; Ejiófor *et al.*, 1994b; Jacquez, 1998; Grijpspeerdt e Vanrolleghem, 1999). A metodologia do desenho óptimo de experiências (baseada na optimização de funções da Matriz de Informação de Fisher) tem como objectivo encontrar um conjunto de dados experimentais suficientemente informativos, dos quais se possa extrair a informação mais pertinente para a identificação de parâmetros do modelo, que caracteriza o sistema em estudo (Posten e Munack, 1990; Schneider e Munack, 1995; Berkholz *et al.*, 2000; Asprey e Macchietto, 2002; Renard *et al.*, 2004).

Neste capítulo será apresentado o modelo matemático de espaço de estados para o processo em estudo. Assume-se que a estrutura do modelo utilizado neste trabalho é a correcta. Consequentemente, o problema de modelação resumiu-se ao problema de estimação de parâmetros baseado em dados experimentais. É estudado o problema de identificação de bioprocessos e são apresentados os procedimentos, sugeridos na literatura, mais relevantes para a sua resolução. A metodologia do desenho óptimo de experiências baseada na Matriz de Informação de Fisher é

descrita mais extensamente, uma vez que a contribuição desta tese assenta nesta metodologia. Estabelece-se uma estratégia de identificação dos coeficientes de rendimento do modelo em estudo, cuja aproximação matemática se baseia nos trabalhos desenvolvidos por Chen (1992) e Ferreira (1995).

3.1. MODELAÇÃO MATEMÁTICA

A capacidade de descrever um objecto, uma acção ou um processo em termos matemáticos é uma das aspirações mais contundentes dos engenheiros de processos, de modo a melhorar a monitorização e controlo de bioprocessos. Tem-se verificado, por isso um crescente interesse na utilização de estratégias baseadas em modelos, quer no meio académico quer industrial, para o projecto, monitorização, controlo e optimização de processos, com vista a ganhar vantagens competitivas.

É desejável estabelecer-se a estrutura mais adequada do modelo e identificar o melhor valor para os seus parâmetros, de modo a conseguir-se a melhor aproximação aos dados experimentais. Assim, os dados experimentais têm um papel de destaque na modelação, uma vez que permitem identificar áreas de deficiência do modelo, levando à formulação de novas hipóteses e assim contribuir para a obtenção de um modelo mais robusto e para a sua validação.

3.1.1. Classes de modelos matemáticos

Os modelos matemáticos podem ser classificados de acordo com diferentes critérios, podendo um dado processo ser descrito por modelos matemáticos pertencentes a mais de uma das categorias indicadas na Tabela 3.1.

Tabela 3.1 Classes de modelos matemáticos.

não verbal	verbal
descritivo	preditivo
determinístico	estocástico
contínuo no tempo	discreto no tempo
não estruturado	estruturado
distribuído	segregado
parâmetros concentrados	parâmetros distribuídos

No livro de Bailey e Ollis (1986) e no artigo de Bailey (1998) são apresentadas várias considerações gerais sobre a modelação e as classes de modelos matemáticos aplicados aos bioprocessos.

Nesta tese é investigada e estudada a estimação de parâmetros de um modelo não estruturado e determinístico.

Nos modelos não estruturados a população de microrganismos é considerada numa única variável, não havendo a preocupação de descrever a composição e a qualidade da biomassa. Estes modelos permitem, normalmente, uma descrição satisfatória dos processos numa variada gama de condições. Contudo, estas condições devem estar bem definidas pois estes modelos consideram que as reacções em todas as células ocorrem à mesma velocidade.

Tem sido demonstrado que os modelos não estruturados podem ser integrados com sucesso, em algoritmos de monitorização avançada e de controlo de bioprocessos, sem a necessidade de equipamento muito sofisticado (Bastin e Dochain, 1990). Além disso, estes modelos estão na base do procedimento do controlo óptimo adaptativo para a optimização de processos (Van Impe, 1993).

Por outro lado, existem os modelos estruturados nos quais um componente, como a biomassa, é subdividido em compartimentos. O comportamento dinâmico do sistema pode assim ser simulado de um modo mais efectivo, visto que é representado um maior número de fenómenos. No entanto, em geral, estes modelos são muito difíceis de identificar, devido ao grande número de parâmetros e uma vez que todas as variáveis de estado necessárias são, em geral, de difícil medição. Neste trabalho, optou-se por não considerar este tipo de modelo.

Nos modelos determinísticos as variáveis de saída e, conseqüentemente, os seus parâmetros são considerados variáveis determinísticas, isto é, com um significado de ausência de incerteza. Contudo, na prática, os parâmetros podem apenas ser identificados com um certo nível de precisão, devido aos erros resultantes de perturbações, tais como, dinâmicas do sistema que não estão modeladas e o ruído dos dados experimentais. Nos modelos determinísticos estes erros não são incorporados no modelo. Grande parte do trabalho desenvolvido na presente tese assenta no projecto de experiências que tem como objectivo aumentar a qualidade dos parâmetros identificados.

Em oposição aos modelos determinísticos, nos modelos estocásticos as variáveis e os parâmetros são considerados quantidades às quais se associa uma distribuição de probabilidades. Nestes modelos são explicitamente tidas em conta as perturbações do sistema, estando as incertezas normalmente associadas às variáveis de entrada e saída do sistema. Uma limitação destes modelos é o relativo esforço computacional necessário para a sua formulação e identificação.

3.1.2. Desenvolvimento do modelo

O desenvolvimento de um modelo matemático capaz de simular o comportamento de um dado processo é uma tarefa iterativa, que decorre em várias etapas fortemente dependentes umas das

outras: (1) formulação do modelo; (2) identificação do modelo e (3) validação do modelo. Na Figura 3.1 são representadas, de uma forma esquemática, as etapas do desenvolvimento do modelo matemático.

A etapa de formulação do modelo consiste na definição e caracterização da sua estrutura. Nesta fase são definidos os limites do sistema, as variáveis de entrada e saída assim como o tipo e o nível de complexidade do modelo a utilizar. São também definidas as várias relações entre as variáveis.

Durante a etapa de identificação pretende-se estimar o valor numérico dos parâmetros envolvidos nas funções que relacionam as variáveis do modelo. Esta é normalmente a etapa crítica do desenvolvimento do modelo matemático, devido essencialmente às dificuldades experimentais e ao tempo que exige.

A etapa final do desenvolvimento do modelo matemático consiste na sua validação, sendo feito nesta etapa o teste à sua capacidade de previsão. As previsões do modelo são comparadas com os resultados experimentais que não foram utilizados na identificação dos seus parâmetros. Se o modelo descrever satisfatoriamente os resultados experimentais conclui-se que o modelo está validado, caso contrário, se as previsões discordarem significativamente desses resultados, não conseguindo também descrever as suas tendências, a estrutura do modelo deverá ser alterada e considerar novamente as etapas anteriores.

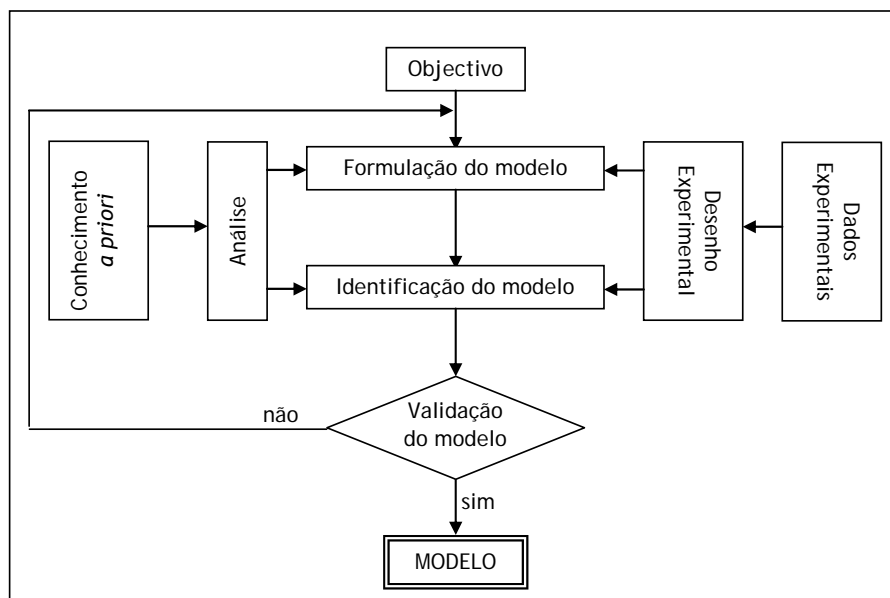


Figura 3.1 Etapas do desenvolvimento do modelo matemático em bioprocessos.

Na modelação matemática de bioprocessos são exploradas essencialmente três fontes de informação: o objectivo do modelo, ou seja, o fim a que se destina deverá ser definido claramente à partida; o conhecimento *a priori*, que pode ser baseado no conhecimento empírico, físico, bioquímico e/ou microbiológico do sistema; e os dados experimentais obtidos durante operações

normais ou durante experiências planejadas para a identificação do modelo (Rao e Iyengar, 1984; Vanrolleghem e Dochain, 1998; Hodgson *et al.*, 2004).

A modelação de processos biotecnológicos envolve normalmente um maior número de iterações do que a dos processos químicos. Na maior parte das aplicações físicas e químicas o conhecimento já adquirido previamente do sistema possui informação de elevada qualidade, permitindo que grande parte da estrutura do modelo seja deduzida através desse conhecimento. A validação experimental permite a identificação dos seus parâmetros e ultrapassar algumas incertezas que ainda possam existir na estrutura do modelo, recorrendo eventualmente a algumas iterações do procedimento de modelação.

Contrariamente, as características inerentes dos bioprocessos, como a sua não linearidade, conjugadas com técnicas analíticas pouco adequadas e robustas, fazem com que o procedimento de modelação matemática não possa ser aplicado sem alteração, implicando muitas vezes um número de iterações relativamente elevado. Por outro lado, os dados experimentais pouco precisos, afectados de alguma incerteza, são também um problema na definição da estrutura e identificação dos parâmetros do modelo, pois aumentam as hipóteses quer de se obter um modelo inválido, quer de não se conseguir estimar um valor único para os seus parâmetros. Este problema é ainda mais evidente na identificação em linha, onde se utilizam técnicas que têm associadas mais incertezas para se obter a informação em tempo real. Quando a calibração do modelo é feita em diferido permite o uso de várias técnicas analíticas complementares, que apesar de exigirem um maior tempo de análise, conduzem a dados experimentais mais precisos (Vanrolleghem e Dochain, 1998).

O desenho óptimo de experiências, que tem como objectivo otimizar o conteúdo de informação dos dados experimentais, de modo a permitir a identificação do modelo matemático do processo em estudo, é um instrumento valioso no processo de modelação. A planificação de experiências poderá permitir não só realizar fermentações em condições em que o sistema seja excitado de modo adequado, como escolher períodos dessas fermentações onde a sensibilidade a cada um dos parâmetros do modelo seja maior e assim conseguir a sua identificação com a maior precisão possível (Merkel *et al.*, 1996; Versyck *et al.*, 1998; Bernaerts *et al.*, 2000; Keesman e Stigter, 2002). A análise prévia da sensibilidade do modelo aos vários parâmetros permite por um lado, determinar quais os que têm mais influência nas previsões do modelo e por isso, deverão ser identificados com maior exactidão, e por outro, os parâmetros e fenómenos menos importantes, podendo introduzir-se algumas simplificações no modelo ou uma estimativa menos precisa do seu valor (Degenring *et al.*, 2004).

3.2. IDENTIFICAÇÃO DE PARÂMETROS: PROBLEMAS E METODOLOGIAS

O objectivo da identificação de um sistema é calibrar um modelo matemático, de modo a que este seja adequado ao fim a que se destina, utilizando para isso dados experimentais. A precisão dos parâmetros do modelo depende não só da quantidade, mas essencialmente da qualidade dos dados experimentais que irão ser utilizados nos algoritmos de identificação.

A estrutura do modelo desempenha também um papel importante na identificação dos parâmetros, uma vez que a sua identificabilidade teórica é determinada por essa estrutura, que deverá permitir a identificação de todos os parâmetros (Walter e Pronzato, 1990).

O processo de identificação dos parâmetros do modelo matemático em biotecnologia é limitado, como já referido, não só pela qualidade da informação disponível, proveniente dos resultados experimentais, mas também devido ao facto desses modelos serem altamente não lineares, traduzindo-se muitas vezes numa considerável correlação entre os parâmetros a identificar. Sendo assim, por vezes, apenas um subconjunto desses parâmetros poderá ser identificado mesmo que seja verificada a identificabilidade teórica do modelo.

O modelo geral de um sistema, caracterizado pelas variáveis correspondentes a entradas, saídas e perturbações pode ser representado, recorrendo à notação utilizada em Teoria de Sistemas, como:

$$\frac{dx(t)}{dt} = f(x(t), k, u(t)), \quad x(t_0) = x_0 \quad \text{Equação 3.1}$$

em que x é o vector das variáveis de estado, k o vector dos parâmetros do modelo, u o vector das variáveis de entrada, f um vector de funções reais que, para bioprocessos, são tipicamente não lineares, e t o tempo. As variáveis de saída $y(t)$ são dadas por:

$$y(t) = g(x(t), k, u(t)) \quad \text{Equação 3.2}$$

O conhecimento empírico, assim como o baseado em leis físicas, bioquímicas e microbiológicas e as variáveis medidas de saída, facilitam a escolha das variáveis de estado x e a definição das equações diferenciais da Equação 3.1, que originam um modelo fenomenológico. Na prática, o vector das variáveis de saída y consiste nas variáveis do processo que podem ser medidas ou estimadas por sensores físicos ou por programação, respectivamente. O vector das variáveis de entrada u é constituído pelas variáveis de controlo, as quais são utilizadas para direccionar o comportamento do sistema do modo mais adequado. Na secção 3.4 será descrita a estrutura do modelo para o caso em estudo neste trabalho.

3.2.1. Identificabilidade estrutural

A identificabilidade estrutural ou teórica deve ser verificada antes de realizar a identificação prática ou numérica, que se baseia em dados experimentais para identificar os parâmetros do modelo. Verificar se um determinado modelo é ou não teoricamente identificável consiste em estudar, para uma estrutura conhecida e considerada perfeita do modelo, a possibilidade de se obter um valor único para cada um dos seus parâmetros. Desse estudo poder-se-á concluir que apenas uma combinação dos parâmetros é identificável, sendo por isso necessário recorrer, por exemplo, a relações teoricamente conhecidas entre os parâmetros para permitir a identificação de todos.

O problema da identificabilidade teórica pode ser inferido do seguinte modo (Walter e Pronzato, 1997): assumindo que o processo e o modelo apresentam uma estrutura \mathcal{M} idêntica, que os dados não têm ruído e que as variáveis de entrada (vector u) e os tempos de medição podem ser escolhidos, dever-se-á obter uma identificação única dos parâmetros do modelo e não várias soluções. De facto, um comportamento idêntico de entrada-saída do processo e do modelo, deverá conduzir à obtenção dos mesmos parâmetros, sendo os estimados pelo modelo (\hat{k}) iguais aos do processo (k^*). Assim sendo, o parâmetro k_i é global e estruturalmente (ou unicamente) identificável, se para *quase* todos os valores de k^* ($\in \mathcal{K}$) a condição seguinte for aplicável:

$$\mathcal{M}(\hat{k}) = \mathcal{M}(k^*) \Rightarrow \hat{k}_i = k_i^* \quad \forall i \quad \text{Equação 3.3}$$

A estrutura \mathcal{M} é global e estruturalmente identificável se todos os seus parâmetros são global e estruturalmente identificáveis.

A utilização da palavra *quase*, na definição anterior, advém do facto de uma propriedade do modelo, como a identificabilidade, ser estrutural se for válida para *quase* todos os valores dos parâmetros, e possivelmente falsa num sub-espaço do espaço paramétrico de medida zero. Logo, uma propriedade que é verdadeira para qualquer valor k que não pertence a nenhuma hipersuperfície atípica poderá ser considerada estrutural, uma vez que a probabilidade de seleccionar aleatoriamente um valor atípico é nula (Walter e Pronzato, 1997).

As técnicas utilizadas para testar a identificabilidade estrutural de sistemas não lineares, como é o caso dos bioprocessos, são muito complexas de implementar, exigindo em geral o recurso a cálculos com matemática simbólica. Em geral, este problema pode ser resolvido quer pela aplicação do método de expansão de séries de *Taylor*, quer por métodos de linearização que transformam o modelo não linear num modelo linear nos parâmetros, embora, muitas vezes sem sucesso.

A aproximação por séries de *Taylor* é a metodologia mais utilizada para testar a identificabilidade estrutural de modelos não lineares (Jacquez, 1998; Noykova *et al.*, 2002; Muller *et al.*, 2002). Contudo, nesta metodologia não existe uma regra para determinar *a priori* o número de derivadas de ordem superior que é necessário incluir na aproximação para se obterem resultados

conclusivos. Por outro lado, é sabido que quanto maior for a ordem do polinómio de *Taylor* utilizado na aproximação maior será o esforço computacional exigido e conseqüentemente o tempo requerido. Mais ainda, não é totalmente evidente como escolher a combinação de parâmetros para a qual a identificabilidade deve ser testada (Vanrolleghem e Dochain, 1998; Grijspeerd e Vanrolleghem, 1999; Versyck, 2000).

Uma vez que para sistemas lineares a identificação estrutural é bem entendida e se encontra bem estabelecida, a linearização do modelo não linear permite a utilização de vários testes para verificar a sua identificabilidade, como por exemplo, o método das transformadas de *Laplace* e a aproximação pela matriz dos parâmetros de *Markov* (Godfrey e DiStefano III, 1985; Vanrolleghem e Dochain, 1998; Versyck, 2000). No entanto, esta última metodologia poderá conduzir a um menor número de combinações de parâmetros identificáveis que as existentes no modelo não linear. Assim, alguns dos parâmetros do modelo não linear poderão não aparecer no modelo linearizado (Versyck, 2000).

Atendendo à complexidade das duas metodologias anteriormente referidas e aos problemas que lhes podem estar associados, neste estudo optou-se por recorrer à metodologia proposta por Chen (1992) e Ferreira (1995). Este método usa uma transformação linear de estado baseada na estrutura geral do modelo tendo sido usado com sucesso na identificação de coeficientes de rendimento (Chen, 1992; Ferreira, 1995; Chen e Bastin, 1996; Rocha e Ferreira, 1996). Esta metodologia será descrita em pormenor na secção 3.3.1.

3.2.2. Identificação numérica

O modelo matemático de um dado processo é efectivamente identificado, após ser verificada a sua identificabilidade estrutural, procedendo-se à identificação numérica ou prática que consiste em estimar os parâmetros do modelo (k) através dos dados experimentais, por minimização de uma função escalar de identificação (J):

$$\hat{k} = \arg \min_{k \in K} J_i(k) \quad \text{Equação 3.4}$$

sendo \hat{k} os parâmetros estimados do modelo e K o vector do conjunto dos parâmetros possíveis do modelo (k).

A inferência dos parâmetros \hat{k} pode ser obtida por estimação dos mesmos recorrendo a diferentes critérios como o erro quadrático médio, o erro absoluto médio e o método da máxima verosimilhança (Walter e Pronzato, 1997).

A identificação prática, contrariamente à teórica, que depende apenas da estrutura do modelo, está relacionada com a qualidade dos dados experimentais e com o seu conteúdo de informação que deve ser suficiente para permitir a identificação dos parâmetros do modelo com precisão.

A identificação numérica nem sempre é garantida quando a identificação estrutural foi demonstrada. Com efeito, o reduzido número de dados experimentais normalmente utilizado neste processo, aliado ao facto destes serem afectados de ruído, dificulta a estimação dos parâmetros. Por outro lado, a existência de parâmetros estimados fortemente correlacionados poderá resultar numa aproximação do modelo aos dados experimentais semelhante, visto que a variação de um dos parâmetros pode ser compensada, quase totalmente, por uma alteração proporcional noutro. Outra dificuldade inerente ao processo de identificação pode ser atribuída ao facto de existirem um ou mais parâmetros que pouco influenciam os resultados de simulação obtidos através do modelo. Para além disso, os algoritmos numéricos utilizados para a estimação não linear dos parâmetros apresentam dificuldades de convergência quando confrontados com problemas de mau condicionamento, com não linearidades do modelo e com restrições impostas durante a optimização, sendo também os resultados sensíveis aos valores iniciais pré-estabelecidos para os parâmetros a estimar. Consequentemente, os parâmetros estimados podem apresentar uma grande variabilidade conduzindo a valores de difícil ou impossível interpretação física (Vanrolleghem e Dochain, 1998; Versyck, 2000).

As dificuldades anteriores podem ser ultrapassadas por utilização de dados altamente informativos decorrentes do desenho óptimo de experiências (Munack, 1989; Ejiófor *et al.*, 1994a; Versyck *et al.*, 1999; Versyck e Van Impe, 1999; Asprey e Macchietto, 2000; Bernaerts *et al.*, 2002; Pauw e Vanrolleghem, 2006).

A estimação de parâmetros pode ser formulada pela minimização da seguinte função de identificação, que representa o estimador de máxima verosimilhança:

$$J_i(k) \equiv \int_0^{t_f} (y(k, t) - y_m(t))^T P(t) (y(k, t) - y_m(t)) dt \quad \text{Equação 3.5}$$

em que y_m e $y(k)$ são os vectores das medidas experimentais e de previsão do modelo usando o vector dos parâmetros k , respectivamente, P é a matriz quadrada de coeficientes de ponderação e t_f é o tempo final da experiência.

As seguintes hipóteses são consideradas (Versyck, 2000):

i) O modelo é estruturalmente correcto e teoricamente identificável, pelo que os desvios (ou erros) entre o modelo e os dados experimentais são devidos exclusivamente ao ruído.

ii) As medidas experimentais são afectadas de ruído aleatório, aditivo e independente de média zero, variância unitária e distribuição normal. Deste modo, as variáveis de saída do processo no tempo t_i podem ser escritas como a soma do valor real do processo e um termo associado ao ruído (ε):

$$y_m(t_i) = y(k', t_i) + \varepsilon(t_i) \quad \forall i = 1, \dots, n_t \quad \text{Equação 3.6}$$

iii) A matriz P é escolhida como a matriz dos inversos das variâncias dos erros de medida permitindo assim que, quanto maior for a corrupção dos dados experimentais de uma variável pelo ruído, menor será a sua importância no critério de identificação.

O critério utilizado no desenho óptimo de experiências para a estimação de parâmetros baseia-se no comportamento da função de identificação $J_I(k)$ na região próxima do conjunto dos parâmetros verdadeiros do processo (k^*).

O valor esperado da função objectivo para um conjunto de parâmetros ligeiramente diferentes dos verdadeiros é dado pelo operador estatístico E (Vanrolleghem e Dochain, 1998; Pauw e Vanrolleghem, 2006):

$$E[J_I(k + \delta k)] \cong E[J_I(k)] + \delta k^T \left[\int_0^{t_f} \left(\frac{\partial y}{\partial k}(t) \right)^T P(t) \left(\frac{\partial y}{\partial k}(t) \right) dt \right] \delta k \quad \text{Equação 3.7}$$

Da equação anterior resulta que a optimização do processo de identificação numérica conduz à maximização da diferença entre $J_I(k + \delta k)$ e $J_I(k)$, ou seja, maximizar o termo

$\left[\int_0^{t_f} \left(\frac{\partial y}{\partial k}(t) \right)^T P(t) \left(\frac{\partial y}{\partial k}(t) \right) dt \right]$ da Equação 3.7. Esse termo é designado por Matriz de Informação de

Fisher (*MIF*) e expressa o conteúdo informativo dos dados experimentais:

$$MIF \equiv \int_0^{t_f} \left(\frac{\partial y}{\partial k}(t) \right)^T P(t) \left(\frac{\partial y}{\partial k}(t) \right) dt \quad \text{Equação 3.8}$$

Esta matriz é real, não singular, simétrica, semi-positiva, quadrada de dimensão k , pertencendo a um conjunto de matrizes de informação convexo e fechado (Mehra, 1974; Ferreira, 1995; Hidalgo e Ayesa, 2001).

A Matriz de Informação de Fisher contém informação relativa aos erros de medida das variáveis de saída, através da matriz de ponderação P , e das sensibilidades das variáveis de saída relativamente aos parâmetros do modelo, por intermédio da matriz das funções de sensibilidade

$\frac{\partial y}{\partial k}$. Assim, a *MIF* permite a quantificação aproximada da qualidade da estimativa dos parâmetros na vizinhança do conjunto dos verdadeiros valores dos parâmetros (Goodwin, 1987; Versyck, 2000; Hidalgo e Ayesa, 2001; Smets *et al.*, 2002b).

A *MIF* é a matriz inversa da matriz covariância dos erros da estimação dos parâmetros do melhor estimador linear não tendencioso - estimador BLUE (*Best Linear Unbiased Estimator*) (Vanrolleghem e Dochain, 1998):

$$MIF^{-1} \equiv \left[\int_0^{t_f} \left(\frac{\partial y}{\partial k} (t) \right)^T P(t) \left(\frac{\partial y}{\partial k} (t) \right) dt \right]^{-1}$$

Equação 3.9

3.2.2.1. Planificação de experiências

O desenho de experiências para a identificação de parâmetros deve ter em conta a precisão da estimativa dos parâmetros, a qual pode ser melhorada pelo aumento do número de dados experimentais ou alternativamente pela optimização de uma função escalar da *MIF* (Versyck *et al.*, 1997; Grijspeerd e Vanrolleghem, 1999; Bernaerts *et al.*, 2002).

O desenho óptimo de experiências, baseado na optimização de funções escalares da *MIF*, é analisado neste trabalho. Foram desenvolvidas diferentes estratégias para planificar experiências de tal modo que os dados experimentais permitissem obter uma solução única dos parâmetros ou combinações de parâmetros que demonstrassem ser teoricamente identificáveis, isto é, delinear experiências informativas.

Os valores verdadeiros dos parâmetros do processo k^* são desconhecidos no início da planificação de experiências. Sendo assim, a *MIF* é determinada recorrendo a valores de parâmetros nominais k_o , que são estimativas iniciais obtidas normalmente da literatura e/ou de experiências preliminares. Assim, as sensibilidades das saídas relativamente aos parâmetros do modelo são avaliadas ao longo das trajectórias de saída nominais, as quais são calculadas com o conjunto de parâmetros nominais k_o (para os quais se assume implicitamente que os seus valores são próximos dos valores verdadeiros dos parâmetros do processo k^*).

No caso da identificação de dois parâmetros, a Equação 3.7 permite uma interpretação gráfica da função de identificação J_I e da Matriz de Informação de Fisher, uma vez que as linhas que representam valores constantes de J_I são elipses, como se pode observar na Figura 3.2.

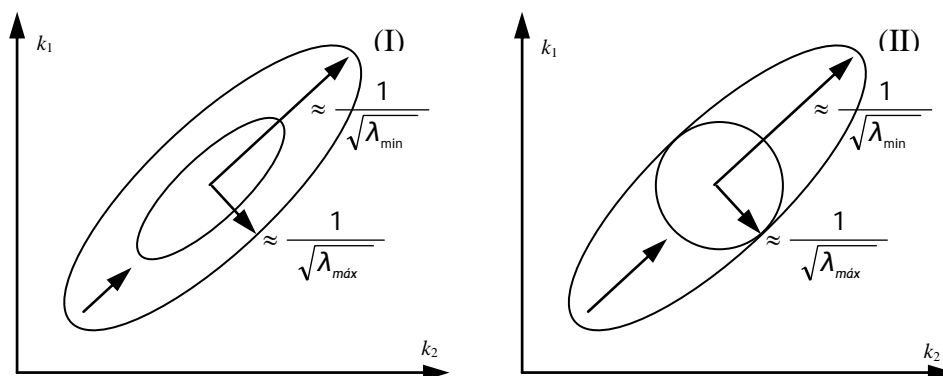


Figura 3.2 Efeito das propriedades da *MIF* no tamanho e forma da região de confiança dos parâmetros (retirado de Pauw e Vanrolleghem (2006)).

O maior e menor diâmetro da elipse são inversamente proporcionais à raiz quadrada do menor (λ_{\min}) e maior (λ_{\max}) valor próprio da *MIF*, respectivamente. Estes valores relacionam-se com os desvios padrão relativos à precisão da estimação dos parâmetros. Dessa forma, poder-se-á concluir que as propriedades da *MIF* determinam as propriedades das elipses dos desvios podendo-se assim, avaliar a precisão dos parâmetros estimados (Pauw e Vanrolleghem, 2006).

Dependendo das exigências do processo, é utilizada como medida do desempenho para o desenho óptimo de experiências, uma função escalar específica da *MIF*. Na literatura são apresentados diferentes critérios de optimalidade (Walter e Pronzato, 1990; Munack, 1991; Renko e Sirvio, 1997; Berkholz e Guthke, 2001). A Tabela 3.2 resume alguns desses critérios. Convém realçar que os resultados obtidos no desenho óptimo dependem do critério escolhido sendo por isso conveniente proceder a uma escolha adequada ao fim pretendido.

Tabela 3.2 Critérios de optimalidade usados no desenho óptimo de experiências.

Critério de optimalidade	Interpretação
Critério A: $\min\{\text{traço}(MIF^{-1})\}$	Minimização da média aritmética do erro de estimação, isto é, minimização da variância média.
Critério D: $\max\{\det(MIF)\}$	Minimização da média geométrica do erro de estimação, isto é, minimização do erro associado à precisão global, que equivale à diminuição do volume das elipsóides dos desvios.
Critério E: $\max\{\lambda_{\min}(MIF)\}$	Minimização do erro máximo de estimação, isto é, minimização do diâmetro máximo da elipsóide dos desvios.
Critério E modificado: $\min\left\{\frac{\lambda_{\max}(MIF)}{\lambda_{\min}(MIF)}\right\}$	Minimização do erro associado à precisão individual de estimação de cada parâmetro, que equivale à transformação de uma elipsóide numa esfera.

Os critérios apresentados podem ser usados como funções objectivo do problema de optimização relativamente às variáveis de projecto como as variáveis de entrada $u(t)$, o tempo de amostragem t_f e o número de amostras n_t . O desenho óptimo de experiências é tipicamente um problema de optimização não linear (Smets *et al.*, 2002b). Estes problemas requerem algoritmos de optimização iterativos para os quais não é garantido, à partida, a obtenção do óptimo global.

Neste trabalho é dada particular atenção aos critérios D e E modificado, bem como à optimização simultânea dos dois critérios.

Na planificação de experiências utilizando os critérios D e E modificado, a função objectivo é definida como o determinante e o número de condicionamento da *MIF*, respectivamente:

$$J_{II} \equiv \det(MIF) \quad \text{Equação 3.10}$$

$$J_{III} \equiv \Lambda(MIF) = \frac{\lambda_{\max}(MIF)}{\lambda_{\min}(MIF)} \quad \text{Equação 3.11}$$

A motivação da utilização destes dois critérios advém do facto de serem não só critérios alternativos mas também complementares, permitindo a optimização da precisão dos parâmetros estimados, sendo a interpretação geométrica e visualização gráfica dos resultados, uma medida que faculta a avaliação rápida da sua qualidade.

A maximização do determinante da *MIF* permite optimizar a precisão global dos parâmetros estimados, traduzindo-se na minimização do volume das elipsóides dos desvios ou de modo equivalente da região de confiança, como é representado na Figura 3.2 (I). Como referido anteriormente, a matriz de informação é uma matriz definida semi-positiva, sendo o valor do seu determinante estritamente positivo.

A minimização do número de condicionamento está relacionada com a optimização da precisão individual de cada parâmetro estimado e permite reduzir a sua correlação. Este critério tende a aproximar a dimensão dos valores próprios máximo (λ_{\max}) e mínimo (λ_{\min}) da *MIF*, resultando também na aproximação dos valores do maior e menor diâmetro das elipsóides dos desvios, traduzindo-se, no ponto óptimo, em regiões de confiança esféricas (Figura 3.2 (II)). Por outro lado, este critério tem a vantagem de ser conhecido, à partida, o seu valor óptimo, isto é, o seu valor mínimo é por definição um. Como para o critério anterior, quando se obtém uma matriz de informação semi-positiva e, por isso, os seus valores próprios são positivos, a experiência diz-se informativa.

Deve ser referido que a optimização individual dos critérios *D* e *E* modificado pode levar à diminuição da precisão global ou individual dos parâmetros estimados, respectivamente. Geometricamente esta situação implica que, no caso da optimização do critério *D*, o hipervolume obtido no espaço dos parâmetros poderá ser uma hipersuperfície muito longa e estreita, ou seja, o maior e menor diâmetro das hiper-elipsóides têm dimensões muito diferentes. Para o caso do critério *E* modificado, poder-se-á obter hipersuperfícies muito largas, resultando em grandes valores do maior e menor diâmetro das hiper-elipsóides, apesar de semelhantes. Esta problemática motiva a optimização simultânea dos dois critérios.

A vantagem desta metodologia de planificação de experiências resulta da possibilidade de testar várias escolhas *ad hoc* das variáveis de projecto antes de conduzir uma experiência. Este facto é realçado no trabalho de Kutalik *et al.* (2004) onde se descreve a aplicação do desenho óptimo de experiências, baseado na optimização do determinante da matriz de Fisher, na selecção de tempos óptimos de amostragem para a identificação de parâmetros de modelos dinâmicos em Biologia de Sistemas. Estes autores concluem que a planificação de experiências tem um enorme impacto na

precisão dos parâmetros estimados e simultaneamente reduz substancialmente os custos experimentais e melhora a informação retirada dos dados experimentais.

3.2.2.2. Funções de sensibilidade

As funções de sensibilidade têm um papel central na avaliação da identificação numérica, uma vez que são o principal componente da Matriz de Informação de Fisher. Se as funções de sensibilidade são linearmente dependentes, a matriz das variâncias torna-se singular e o modelo é não identificável. Com efeito, em muitos dos modelos usados para descrever fenómenos biológicos, as funções de sensibilidade são *quasi*-linearmente dependentes, conduzindo à obtenção de parâmetros altamente correlacionados, em que várias combinações dos mesmos podem descrever de igual modo os dados experimentais.

Assim, uma maneira fácil de estudar a identificação numérica é representar graficamente as funções de sensibilidade. Podem ser encontrados, na literatura, diversos trabalhos (Chouakri *et al.*, 1994b; Claes e Van Impe, 2000; Bernard *et al.*, 2001; Lei e Jorgensen, 2001; Smets *et al.*, 2002a; Pratt *et al.*, 2004; Stigter e Keesman, 2004) que referem um estudo deste tipo, principalmente para o modelo da cinética de *Monod*, medindo a concentração de biomassa e substrato.

Uma outra maneira de avaliar a dependência das funções de sensibilidade é através do cálculo da característica da *MIF*. A ausência de dependência linear deverá conduzir a uma matriz de característica completa. Por outro lado, o valor do número de condicionamento da *MIF*, isto é, a razão entre o maior ($\lambda_{\text{máx}}$) e o menor ($\lambda_{\text{mín}}$) dos seus valores próprios, é também uma medida da dependência linear das funções de sensibilidade, sendo que, quanto maior for o número de condicionamento menor será a possibilidade de identificação numérica dos parâmetros.

Intuitivamente, as funções de sensibilidade expressam a dependência das variáveis de estado com a alteração dos parâmetros. Assim, estas funções permitem determinar as condições onde a dependência é maior e por isso, identificar a(s) condição(ões) que mais informação pode(m) reunir sobre os parâmetros a estimar. Por exemplo, poder-se-á escolher o tempo de amostragem de modo a aumentar o conteúdo informativo da experiência, isto é, numa zona onde a sensibilidade seja mais elevada.

3.2.2.3. Avaliação da precisão da estimativa dos parâmetros identificados

A precisão da estimativa dos parâmetros identificados deve ser determinada após o processo de identificabilidade estrutural e numérica do modelo matemático. Assumindo a estrutura do modelo como correcta, o desvio entre os valores estimados dos parâmetros \hat{k} e os seus verdadeiros valores k^* , é devido à diferença entre os resultados obtidos pelo modelo e os dados experimentais do sistema, ou seja, devido ao ruído associado às medições. Mais ainda, sendo o ruído uma variável aleatória, aditiva e independente de média zero, variância unitária e distribuição normal (ruído

branco) e as variáveis de saída linearizadas relativamente aos parâmetros, pode-se assumir que a distribuição dos valores \hat{k} segue uma distribuição normal, de média igual aos valores verdadeiros k^* e variância MIF^{-1} . No entanto, atendendo ao Teorema do Limite Central, mesmo que as observações não estejam normalmente distribuídas, a distribuição dos parâmetros estimados tende para uma distribuição normal desde que o número de dados seja suficientemente grande (Versyck, 2000).

A MIF^{-1} (Equação 3.9), designada como limite inferior de *Cramer-Rao*, é uma medida da precisão da estimação dos parâmetros do modelo (Ferreira, 1995; Versyck, 2000):

$$MIF^{-1} \leq E \left\{ (\hat{k} - k^*) (\hat{k} - k^*)^T \right\} \quad \text{Equação 3.12}$$

Um ponto importante na identificação numérica é o cálculo de parâmetros que traduzem a qualidade da estimação, como por exemplo, os valores aproximados do desvio padrão e do coeficiente de correlação, que podem ser determinados a partir da MIF (calculada a partir dos melhores valores ajustados do modelo com base nos dados experimentais) (Walter e Pronzato, 1997; Vanrolleghem e Dochain, 1998; Versyck, 2000). A variância constante (homocedasticidade) dos ruídos das medições pode ser aproximada pelo erro quadrático médio residual (EQM):

$$EQM = \frac{EMQ}{n_t - n_p} \quad \text{Equação 3.13}$$

onde EMQ representa o erro dos mínimos quadrados (calculado pela Equação 3.5 com base nos melhores valores ajustados do modelo, sendo a matriz P igual à matriz identidade), n_t é o número de dados experimentais e n_p representa o número de parâmetros estimados.

O desvio padrão (s_{approx}) individual, para cada um dos parâmetros i estimados (\hat{k}_i), pode ser aproximado pela raiz quadrada do elemento da diagonal principal da matriz MIF^{-1} correspondente (MIF_{ii}^{-1}):

$$s_{approx}(\hat{k}_i) = \sqrt{MIF_{ii}^{-1}} \quad \text{Equação 3.14}$$

No entanto, os valores assim obtidos são estatisticamente otimistas, uma vez que a Equação 3.14 implica uma aproximação linear de modelos não lineares na vizinhança da melhor estimativa dos parâmetros. Existem métodos mais robustos que conduzem a valores dos desvios padrão dos parâmetros mais realistas que exigem contudo, um maior esforço computacional e um gasto de tempo mais acentuado (Vanrolleghem e Dochain, 1998; Versyck, 2000).

É ainda possível, com base na MIF^{-1} , calcular o coeficiente de correlação aproximado (c_{ij}) entre dois parâmetros estimados i e j (Walter e Pronzato, 1997; Versyck, 2000), que varia entre -1 e 1 :

$$c_{ij} = \frac{MIF_{ij}^{-1}}{\sqrt{MIF_{ii}^{-1}} \sqrt{MIF_{jj}^{-1}}} \quad \text{Equação 3.15}$$

Assumindo que os erros de estimação seguem uma distribuição normal pode-se estabelecer um intervalo de confiança aproximado a $(1-\alpha) \times 100\%$, em que α é o nível de significância pré-definido (em geral 0,05) para cada parâmetro i estimado:

$$\left[\hat{k}_i - t_{GL} \left(\frac{\alpha}{2} \right) \times \sqrt{MIF_{ii}^{-1}}, \hat{k}_i + t_{GL} \left(\frac{\alpha}{2} \right) \times \sqrt{MIF_{ii}^{-1}} \right] \quad \text{Equação 3.16}$$

em que t representa o valor crítico da distribuição t de Student com GL graus de liberdade:

$$GL = n_t - n_p \quad \text{Equação 3.17}$$

Convém referir que no estudo da incerteza dos parâmetros estimados pelo modelo, a MIF (Equação 3.14 à Equação 3.16) é calculada usando em substituição da matriz de ponderação P um escalar constante igual ao inverso do EMQ (Equação 3.13) (Walter e Pronzato, 1997; Versyck, 2000). Como anteriormente referido, a precisão da estimativa dos parâmetros pode ser melhorada aumentando o número de dados experimentais ou alternativamente pela optimização de uma função escalar da MIF , como se pode inferir pela observação da Equação 3.13 e Equação 3.14, respectivamente.

A incerteza associada à estimação dos parâmetros do modelo pode ainda ser avaliada por dois métodos alternativos ou, possivelmente, complementares: desenho de regiões de confiança assintóticas, baseadas no estimador de máxima verosimilhança, e de regiões de confiança mútuas não assintóticas, baseadas no estimador dos mínimos quadrados, sendo esta mais aconselhável quando o número de dados experimentais disponíveis é reduzido, embora não seja a mais adequada, se o sistema tiver várias variáveis de saída ou apresentar variâncias não constantes para os erros das medidas (ruído). A região de confiança pode ser descrita como um hipervolume limitado por uma hipersuperfície no espaço dos parâmetros de dimensão n_p .

A região de confiança assintótica (hiper-elipsóides) pode ser descrita pela seguinte inequação:

$$\left(\hat{k} - k^* \right)^T \times MIF \times \left(\hat{k} - k^* \right) \geq \alpha \quad \text{Equação 3.18}$$

em que o primeiro termo da expressão anterior segue uma distribuição χ^2 com n_p graus de liberdade, sendo a forma, o tamanho e orientação das elipsóides determinada pelos vectores e valores próprios da MIF .

Convém referir que, quando os ruídos das medições são aditivos apresentando média nula, variância unitária e distribuição gaussiana normal (ruídos brancos), as regiões de confiança assintóticas são proporcionais às elipsóides dos desvios. Estas últimas representam a região de todos os desvios dos parâmetros causados pelos desvios das medições, $\delta y(t)$, que teoricamente, está centrada no valor verdadeiro de k (Chen, 1992). Assim, para uma variação $\delta y(t)$ medida em termos de norma-2, tem-se:

$$\|\delta y(t)\| \equiv \sqrt{\int_0^{t_f} (\delta y(t))^2 dt} = \delta_c \quad (\text{constante}) \quad \text{Equação 3.19}$$

em que a correspondente elipsóide no espaço de parâmetros é dada por:

$$(\hat{k} - k^*)^T \times MIF \times (\hat{k} - k^*) = \delta_c^2 \quad \text{Equação 3.20}$$

A análise das elipsóides dos desvios fornece informação qualitativa equivalente à das regiões de confiança dos parâmetros estimados.

A região de confiança não-assintótica pode ser descrita por:

$$SEQ \leq EMQ \times \left(1 + \frac{n_p}{n_t - n_p} F_{GL_1, GL_2}(\alpha) \right) \quad \text{Equação 3.21}$$

onde SEQ representa a soma dos erros quadráticos e F representa o valor crítico da distribuição F de Fisher-Snedecor, com GL_1 e GL_2 graus de liberdade (iguais a n_p e $n_t - n_p$, respectivamente), a um nível de significância α .

3.3. MODELO DINÂMICO DE REACTORES BIOLÓGICOS

O desenvolvimento de um modelo matemático é normalmente um processo iterativo. A descrição do bioprocessamento recorre a sistemas de equações diferenciais e algébricas que envolvem frequentemente modelos altamente não lineares e *stiff*, que representam princípios de conservação (como balanços de massa) e relações cinéticas ou de equilíbrio (como leis de velocidade de reacção ou de equilíbrio gás-líquido) (Banga *et al.*, 2002).

Em sistemas biológicos a dinâmica da variação com o tempo da quantidade ou concentração de um dado componente no reactor pode ser explicada por dois tipos de fenómenos distintos: o fenómeno de transformação ou conversão, devido às reacções químicas, bioquímicas e biológicas

que permitem a transformação de alguns componentes em outros; e o fenómeno de transporte ou de transferência de massa que ocorre devido a trocas de líquido e/ou gás do reactor com o meio exterior. Sendo assim, o modelo dinâmico de tais sistemas caracteriza-se pela descrição matemática destes dois tipos de fenómenos.

O modelo dinâmico geral para reactores biológicos com agitação perfeita e mistura completa (assume-se que a composição de um componente no meio líquido é homogénea) é obtido através de um balanço macroscópico aos componentes envolvidos num dado esquema reaccional. Dado esse esquema reaccional do processo, a derivação do modelo dinâmico é sistemática baseando-se num sistema de equações não lineares do espaço de estados, denominado Modelo Dinâmico Geral de Reactores Biológicos (Bastin e Dochain, 1990). Para um dado esquema reaccional, composto por m reacções e n componentes, o modelo geral é constituído por um sistema de n equações diferenciais ordinárias representando o balanço a cada componente, designado por variável de estado. A dinâmica da concentração de cada componente ξ_i num reactor com operação semi-contínua, é então escrita da seguinte forma:

$$\frac{d\xi_i}{dt} = \sum_{j \& i} (\pm k_{ij} r_j - \frac{F_{tot}}{W} \xi_i + \frac{F_{e,i}}{W} \xi_{e,i} + \frac{Q_e}{W} \xi_{e,i}^G - \frac{Q_s}{W} \xi_{s,i}^G) \quad \text{Equação 3.22}$$

com $i = 1, \dots, n$ e $j = 1, \dots, m$;

onde,

ξ_i	variável de estado (concentração do componente i);
k_{ij}	coeficientes de rendimento ou estequiométricos do componente i na reacção j ;
r_j	taxa cinética de reacção j (por exemplo: velocidade de crescimento do microrganismo);
F_{tot}	caudal mássico total de alimentação ao reactor na forma líquida;
$F_{e,i}$	caudal mássico de alimentação ao reactor na forma líquida, para o componente i ;
$\xi_{e,i}$	concentração do componente i na correspondente alimentação;
Q_e	caudal volumétrico de entrada do reactor na forma gasosa;
Q_s	caudal volumétrico de saída do reactor na forma gasosa;
$\xi_{e,i}^G$	concentração do componente i na corrente de entrada gasosa;
$\xi_{s,i}^G$	concentração do componente i na corrente de saída gasosa;
W	massa de líquido dentro do reactor.

A notação $j \& i$ significa que o somatório é efectuado nas reacções de índice j envolvendo o componente de índice i .

Os coeficientes k_{ij} são valores estritamente positivos. O termo correspondente apresenta o sinal “-” quando i é reagente e o sinal “+” quando i é produto da reacção. Estes coeficientes são adimensionais (unidade de massa/unidade de massa) em consequência de se exprimir a velocidade de uma reacção em relação à velocidade de um componente dito nominal (Bastin e Dochain, 1990) ou normalizador (Chen, 1992). Nessa reacção o coeficiente associado ao componente nominal é,

naturalmente, igual à unidade. Os componentes não intervenientes numa dada reacção apresentam coeficientes nulos (designados por zeros estruturais).

O modelo geral pode ainda ser escrito usando a seguinte representação matricial:

$$\frac{d\xi}{dt} = Kr(\xi, t) - D\xi + F + Q \quad \text{Equação 3.23}$$

em que,

- ξ vector de estado constituído pelas concentrações dos componentes: $\xi^T = [\xi_1, \xi_2, \dots, \xi_n]$;
- K matriz de coeficientes de rendimento ou estequiométricos: $K = [k_{ij}]$;
- r vector de taxas cinéticas de reacção: $r^T = [r_1, r_2, \dots, r_n]$;
- F vector de caudais mássicos de alimentação do reactor na forma líquidas (por unidade de peso do líquido dentro do reactor): $F^T = [F_1, F_2, \dots, F_n] = \left[\frac{F_{e,1}}{W}, \frac{F_{e,2}}{W}, \dots, \frac{F_{e,n}}{W} \right]$;
- Q vector de caudais mássicos na forma gasosa (por unidade de peso do líquido dentro do reactor): $Q^T = [Q_1, Q_2, \dots, Q_n] = \left[\frac{Q_e}{W} \xi_{e,1}^G - \frac{Q_s}{W} \xi_{s,1}^G, \dots, \frac{Q_e}{W} \xi_{e,n}^G - \frac{Q_s}{W} \xi_{s,n}^G \right]$;
- D matriz taxa de diluição: $D = dI_n$ em que d é definido como o quociente entre o caudal mássico total resultante do líquido das várias correntes líquidas de entrada e saída do reactor (F_{tot}) e o peso do líquido dentro do reactor (W); I_n é a matriz identidade de dimensão ($n \times n$) a qual será omitida nas representações que se seguem e assume-se que:

$$D = d \equiv \frac{F_{tot}}{W} \quad \text{Equação 3.24}$$

Os vectores ξ , F , $Q \in \mathfrak{R}^n$, a matriz $K \in \mathfrak{R}^{n \times m}$, o vector $r \in \mathfrak{R}^m$ e $D \geq 0$ poderá variar com o tempo. A dimensão ($n \times m$) da matriz K é tal que n é maior ou igual a m , isto é, o número de componentes é igual ou superior ao número de reacções.

Nesta equação estão evidenciados os dois fenómenos que ocorrem num reactor biológico: o termo " Kr " que descreve a cinética das reacções químicas, biológicas ou bioquímicas, e, os termos " $-D\xi + F + Q$ " que descrevem a dinâmica dos fenómenos de transporte no reactor.

Adicionalmente, é necessário, uma vez que se descreve um reactor com operação semi-contínua, utilizar uma outra equação para a variação do peso do líquido dentro do reactor, traduzida da seguinte forma:

$$\frac{dW}{dt} = F_{tot} \quad \text{Equação 3.25}$$

3.3.1. Identificação dos coeficientes de rendimento

A identificação de modelos para sistemas biológicos não é uma tarefa fácil, podendo-se realçar duas dificuldades principais: (i) na maior parte dos casos, a determinação de expressões analíticas para descrever a cinética das reacções é um processo empírico, estando, a maior dificuldade, inerente à complexidade e não linearidade da cinética; (ii) os métodos existentes, descritos na literatura (secção 3.2.1), para analisar a identificabilidade teórica de sistemas dinâmicos não lineares são, na prática, de difícil implementação porque normalmente requerem um grande número de manipulações algébricas não lineares. Quanto a esta última dificuldade, embora, a sua utilização se torne mais promissora devido ao desenvolvimento de novos programas de matemática moderna que permitem manipulação simbólica, até ao momento ainda não fornecem propriedades genéricas para esta classe específica de modelos (Chen, 1992). Deste modo, torna-se interessante encontrar métodos alternativos que facilitem a verificação da identificabilidade e todo o processo de identificação.

O problema da identificação de modelos de bioprocessos é, normalmente, duplo:

- Estimação dos coeficientes de rendimento da matriz K ;
- Determinação de uma estrutura apropriada para o modelo das velocidades de reacção $r(\xi, t)$ e a estimação dos seus coeficientes cinéticos.

Neste trabalho, estudar-se-á somente o primeiro problema, resumindo-se à estimação dos coeficientes de rendimento. A metodologia utilizada e que será descrita de seguida, foi parcialmente proposta por Chen (1992) e Ferreira (1995). Esta metodologia baseia-se no facto de ser possível estimar esses parâmetros sem a modelação da cinética das reacções. De facto, numa primeira etapa, é possível identificar os coeficientes de rendimento sem modelar as velocidades de reacção. Numa segunda etapa, poder-se-á modelar essas velocidades de reacção e identificar os seus parâmetros cinéticos, utilizando os coeficientes já identificados na primeira etapa. É interessante realçar que estes dois tipos de parâmetros aparecem no modelo de forma diferente: o modelo é linear relativamente aos coeficientes de rendimento enquanto que, é normalmente, não linear em relação aos coeficientes cinéticos.

Sendo assim, a identificação dos coeficientes de rendimento da matriz K pode ser separada da identificação das velocidades de reacção $r(\xi, t)$ e dos respectivos parâmetros, aplicando uma transformação de estado apropriada atendendo às seguintes condições (Chen, 1992; Bogaerts *et al.*, 2003):

- Conhecem-se os dados de uma experiência onde as variáveis de estado e todos os termos de transporte foram medidos com uma frequência razoável;
- O número de componentes n é maior que o número de reacções m .

O modelo dinâmico geral apresentado na secção anterior pela Equação 3.23 representa uma classe particular de modelos não lineares de espaço de estados. No estudo que se segue, substitui-se, nessa equação, o termo " $F+Q$ " pelo vector U , designado de "entradas" no sistema:

$$\frac{d\xi}{dt} = Kr(\xi, t) - D\xi + U \quad \text{Equação 3.26}$$

A não linearidade do modelo, patente no termo $Kr(\xi, t)$, está associada às velocidades de reacção $r(\xi, t)$ que são funções não lineares das variáveis de estado. Sendo K a matriz dos coeficientes de rendimento considerados constantes, para a dinâmica de cada variável de estado, a não linearidade está sempre associada a uma combinação linear das mesmas funções não lineares $r_1(\xi, t)$, $r_2(\xi, t)$, ..., $r_m(\xi, t)$. Esta característica particular pode ser explorada para separar a parte não linear do modelo da linear, por aplicação de uma transformação de estado linear e não singular adequada, como será apresentado a seguir.

Definindo uma partição não singular da matriz K :

$$\begin{bmatrix} K_a \\ K_b \end{bmatrix} = LK \quad \text{Equação 3.27}$$

em que L representa uma matriz quadrada ($L \in \mathfrak{R}^n$) de permutação de linhas tal que, K_a apresente característica completa (*full rank*). Sendo p a característica da matriz K , então $K_a \in \mathfrak{R}^{p \times m}$ e $K_b \in \mathfrak{R}^{(n-p) \times m}$, em que n é o número de variáveis de estado e m o número de reacções. As correspondentes partições induzidas de ξ e U são escritas como se segue:

$$\begin{bmatrix} \xi_a \\ \xi_b \end{bmatrix} = L\xi, \quad \begin{bmatrix} U_a \\ U_b \end{bmatrix} = LU \quad \text{Equação 3.28}$$

com $\xi_a, U_a \in \mathfrak{R}^p$ e $\xi_b, U_b \in \mathfrak{R}^{(n-p)}$.

Introduzindo a seguinte transformação linear de estado:

$$Z \equiv H\xi_a + \xi_b \quad \text{Equação 3.29}$$

em que a matriz H de dimensão $(n-p) \times p$ constitui a única solução da seguinte equação matricial:

$$HK_a + K_b = 0 \quad \text{Equação 3.30}$$

isto é,

$$H = -K_b K_a^+ \quad \text{Equação 3.31}$$

em que K_a^+ representa a inversa generalizada ou a pseudo-inversa da matriz K_a . Usando a Equação 3.29 e a Equação 3.31, poder-se-á reescrever o modelo inicial (Equação 3.26) da seguinte forma:

$$\frac{d\xi_a}{dt} = K_a r(\xi_a, Z - H\xi_a) - D\xi_a + U_a \quad \text{Equação 3.32}$$

$$\frac{dZ}{dt} = -DZ + HU_a + U_b \quad \text{Equação 3.33}$$

Observa-se, assim, que o modelo inicial (Equação 3.26) é transformado em dois submodelos. Um, descrito pela Equação 3.32, contém as não linearidades do sistema inicial, o outro, Equação 3.33, sendo linear é apenas composto pela dinâmica de transporte do sistema, não envolvendo explicitamente as velocidades das reacções. Sendo assim, esta equação pode ser usada para estimar os coeficientes de rendimento através da identificação da matriz H , independentemente do tipo de leis cinéticas que possam descrever as velocidades das reacções (Chen, 1992). No entanto, na Equação 3.33 pode-se ainda observar a dependência da dinâmica de Z relativamente à matriz H (consequentemente dependente dos coeficientes de rendimento). Se U_a é um vector de zeros, a integração da equação referida anteriormente pode ser usada para estimar os coeficientes de rendimento. Em alternativa poder-se-á fazer surgir a matriz H numa equação de saída pela seguinte decomposição do vector Z :

$$Z \equiv HZ_a + Z_b \quad \text{Equação 3.34}$$

em que as novas variáveis $Z_a \in \mathfrak{R}^p$ e $Z_b \in \mathfrak{R}^{(n-p)}$ são governadas pelas seguintes dinâmicas designadas, por Chen (1992), como *modelo auxiliar*.

$$\frac{dZ_a}{dt} = -DZ_a + U_a \quad \text{Equação 3.35}$$

$$\frac{dZ_b}{dt} = -DZ_b + U_b \quad \text{Equação 3.36}$$

$$\xi_b = Z_b - H(\xi_a - Z_a) \quad \text{Equação 3.37}$$

A Equação 3.35 e a Equação 3.36 são assintoticamente equivalentes à Equação 3.33 se D for positivo. A equação para o vector ξ_b , que pode ser entendido como uma saída, é obtida por substituição da Equação 3.29 na Equação 3.34. Definindo o vector de saída y e o vector regressor ϕ do seguinte modo:

$$y \equiv Z_b - \xi_b \quad \text{Equação 3.38}$$

$$\phi \equiv \xi_a - Z_a \quad \text{Equação 3.39}$$

A Equação 3.37 toma a forma de uma regressão linear normal:

$$y(t) = H\phi(t) \quad \text{Equação 3.40}$$

com $t \in [0, t_f]$ em que t_f é o tempo final da experiência. Neste *modelo auxiliar* intervêm somente a dinâmica de transporte do sistema e pode ser considerado como um modelo linear, variante no tempo, com estado Z , entradas U_a , U_b e ξ_a e saída ξ_b . Trata-se de um modelo não linear relativamente aos coeficientes de rendimento, mas linear em relação aos elementos da matriz H . Para além disso, os parâmetros desconhecidos intervêm somente na equação de saída - Equação 3.37.

Designando os componentes da matriz H por θ_i , a Equação 3.40 toma a seguinte forma:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n-p} \end{bmatrix} = \begin{bmatrix} \theta_1 & \theta_2 & \cdots & \theta_p \\ \theta_{p+1} & \theta_{p+2} & \cdots & \theta_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ \theta_{(n-p-1) \times p+1} & \theta_{(n-p-1) \times p+2} & \cdots & \theta_{(n-p) \times p} \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix} \quad \text{Equação 3.41}$$

Para cada componente escalar y_l ($l= 1, \dots, n-p$) de y pode escrever-se:

$$y_l(t) = \phi^T(t)\theta_l \quad \text{Equação 3.42}$$

em que θ_l representa o vector constituído pelos elementos θ_i ($i= 1, \dots, p$) da linha l da matriz H . Esta equação de regressão constitui a base de cálculo para a estimação dos coeficientes de rendimento.

Como referido anteriormente, Chen (1992) estudou extensivamente vários problemas ligados à identificabilidade dos coeficientes de rendimento. Esta autora estabelece que as propriedades de identificabilidade dos coeficientes de rendimento, para um dado sistema caracterizado pela matriz K , são independentes da escolha da partição e demonstra a condição necessária para a identificação

desses parâmetros de uma dada reacção (elementos de uma coluna de K) através da matriz H : os coeficientes de rendimento da reacção j são identificáveis através de H somente se $n_j - 1 \leq n - p$, sendo n_j o número de componentes envolvidos na reacção j , p a característica da matriz K , n o número de componentes ou variáveis de estado. Esta condição significa que pode identificar-se pelo menos $(n-p)$ coeficientes em cada reacção sem o conhecimento das velocidades de reacção $r(\xi, t)$.

Assumindo que o vector θ representa, genericamente, os parâmetros θ_i que serão determinados por cada equação de regressão, e que substitui o vector θ_l representativo da linha l da matriz H , assim como, o escalar y substituirá o escalar y_l , a equação de regressão (Equação 3.42) tomará a seguinte forma linear:

$$y(t) = \phi^T(t)\theta \quad \text{Equação 3.43}$$

Considerando uma pequena perturbação nos parâmetros θ na vizinhança de θ_o (valor nominal) escrita por $\delta\theta \equiv \theta - \theta_o$ e que origina uma perturbação ou desvio na trajectória de $y(t)$, dada por:

$$\delta y(t) \equiv y(\theta, t) - y(\theta_o, t) = \left(\frac{\partial y}{\partial \theta} \right) \delta\theta \quad \text{Equação 3.44}$$

com $\left(\frac{\partial y}{\partial \theta} \right)$ a matriz de funções de sensibilidade, representando a sensibilidade da saída $y(\theta, t)$ ($i= 1, \dots, l$) aos parâmetros θ_j ($j= 1, \dots, p$).

Com esta metodologia a precisão dos valores dos coeficientes de rendimento estimados é otimizada usando a análise dos regressores, resultando numa equivalente matriz de informação de Fisher:

$$MIF \equiv \int_0^{t_f} \phi(t)P\phi^T(t)dt \quad \text{Equação 3.45}$$

onde a matriz de funções de sensibilidades é facilmente determinada pela Equação 3.43, sendo dada por:

$$\frac{\partial y}{\partial \theta} = \phi^T(t) \quad \text{Equação 3.46}$$

Considerando esta formulação é necessário, então determinar as trajectórias do regressor $\phi(t, \theta, \xi, U)$, as quais são função do vector dos parâmetros a estimar para posterior determinação dos coeficientes de rendimento.

3.4. CASO DE ESTUDO: CRESCIMENTO DE UMA ESTIRPE RECOMBINANTE DE *Escherichia coli* POR OPERAÇÃO SEMI-CONTÍNUA

Neste trabalho, o crescimento de uma estirpe recombinante de *Escherichia coli* por operação semi-contínua é usado como caso de estudo, para introduzir, analisar e validar a metodologia de desenho óptimo de experiências para a identificação dos coeficientes de rendimento descrita na secção anterior. Numa primeira fase, o desenho óptimo de experiências é analisado e discutido através de considerações analíticas e por simulações numéricas. Assume-se, como já referido, que a estrutura do modelo matemático utilizado é a correcta. Por fim, é descrita a implementação prática dos resultados da planificação de experiências obtidos por simulação. É discutido o desenho e a implementação dos perfis otimizados do caudal de alimentação e avaliada a precisão dos parâmetros identificados através da construção das regiões de confiança e da determinação de coeficientes relacionados com a Matriz de Fisher.

3.4.1. Modelo matemático

O crescimento aeróbio da *E. coli* em glucose pode ser caracterizado por três vias metabólicas principais, como descrito no Capítulo 2. Poder-se-á assim, descrever essas vias metabólicas através da seguinte formulação que constitui o chamado esquema reaccional:



em que S , O , X , C e A representam respectivamente glucose, oxigénio, biomassa, dióxido de carbono e acetato. No seguimento, estas variáveis designarão as concentrações no meio de fermentação de glucose, oxigénio dissolvido, biomassa, dióxido de carbono dissolvido e acetato, respectivamente. μ_1 , μ_2 e μ_3 representam as taxas específicas de crescimento e os k_i são os coeficientes de rendimento, expressos em relação à produção de uma unidade de massa de bactéria (o componente nominal) em cada reacção.

As três reacções anteriores constituem uma descrição qualitativa do esquema reaccional do metabolismo da glucose. Esta formulação, comparativamente às equações estequiométricas apresentadas no Capítulo 2, considera a formação de biomassa e inclui coeficientes de rendimento

(vistos como coeficientes estequiométricos) que tomarão em linha de conta a síntese celular e de produtos secundários eventualmente não considerados no esquema reaccional.

O modelo dinâmico associado a este esquema reaccional pode ser descrito pelas seguintes equações:

$$\frac{dX}{dt} = (\mu_1 + \mu_2 + \mu_3)X - DX \quad \text{Equação 3.50}$$

$$\frac{dS}{dt} = (-k_1\mu_1 - k_2\mu_2)X + \frac{F_{e,S}}{W}S_e - DS \quad \text{Equação 3.51}$$

$$\frac{dA}{dt} = (k_3\mu_2 - k_4\mu_3)X - DA \quad \text{Equação 3.52}$$

$$\frac{dO}{dt} = (-k_5\mu_1 - k_6\mu_2 - k_7\mu_3)X + OTR - DO \quad \text{Equação 3.53}$$

$$\frac{dC_{tot}}{dt} = (k_8\mu_1 + k_9\mu_2 + k_{10}\mu_3)X - CTR - DC_{tot} \quad \text{Equação 3.54}$$

Na última equação, a variável C_{tot} designa o dióxido de carbono dissolvido total, quer na forma de dióxido de carbono quer como bicarbonato. A dinâmica da concentração dos gases dissolvidos, oxigénio e dióxido de carbono, adquire uma grande importância em fermentações aeróbias de *E. coli*, uma vez que aqueles são substratos e produtos importantes sendo a sua dinâmica descrita pela Equação 3.53 e Equação 3.54, respectivamente. Uma discussão mais pormenorizada sobre a dinâmica dos gases dissolvidos em fermentações de *E. coli*, poderá ser encontrada em Rocha (2003). Na equação de balanço da glucose utilizou-se a variável S_e para designar a concentração de açúcar na alimentação.

O modelo dinâmico de espaço de estados pode ser representado pela seguinte equação:

$$\frac{d}{dt} \begin{bmatrix} X \\ S \\ A \\ O \\ C_{tot} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ -k_1 & -k_2 & 0 \\ 0 & k_3 & -k_4 \\ -k_5 & -k_6 & -k_7 \\ k_8 & k_9 & k_{10} \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} X - D \begin{bmatrix} X \\ S \\ A \\ O \\ C_{tot} \end{bmatrix} + \begin{bmatrix} 0 \\ \left(\frac{F_{e,S}}{W}\right)S_e \\ 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ OTR \\ -CTR \end{bmatrix} \quad \text{Equação 3.55}$$

Esta estrutura matricial está subjacente ao modelo dinâmico geral de reactores biológicos com cinco variáveis de estado ($n=5$) e três taxas de reacção ($m=3$). A correspondente equivalência é feita como se segue:

$$\xi = \begin{bmatrix} X \\ S \\ A \\ O \\ C_{tot} \end{bmatrix}; K = \begin{bmatrix} 1 & 1 & 1 \\ -k_1 & -k_2 & 0 \\ 0 & k_3 & -k_4 \\ -k_5 & -k_6 & -k_7 \\ k_8 & k_9 & k_{10} \end{bmatrix} \quad \text{Equação 3.56}$$

$$F = \begin{bmatrix} 0 \\ \left(\frac{F_{e,S}}{W}\right)S_e \\ 0 \\ 0 \\ 0 \end{bmatrix}; Q = \begin{bmatrix} 0 \\ 0 \\ 0 \\ OTR \\ -CTR \end{bmatrix} \quad \text{Equação 3.57}$$

$$r(\xi, t) = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} X \quad \text{Equação 3.58}$$

Contudo, em fermentações de *E. coli* verifica-se a ocorrência de quatro regimes metabólicos distintos: (i) crescimento oxidativo e fermentativo simultâneo em glucose ($\mu_1, \mu_2 > 0$ e $\mu_3 = 0$); (ii) crescimento oxidativo em glucose ($\mu_1 > 0$ e $\mu_2 = \mu_3 = 0$); (iii) crescimento oxidativo simultâneo em acetato e glucose ($\mu_1, \mu_3 > 0$ e $\mu_2 = 0$) e (iv) crescimento oxidativo em acetato ($\mu_3 > 0$ e $\mu_1 = \mu_2 = 0$). No entanto, uma vez que a bactéria pode crescer em duas vias simultaneamente, pode efectuar-se uma simplificação considerando-se apenas dois regimes isto é, respiração e fermentação da glucose com produção de acetato, designado como regime respiro-fermentativo (RF) e respiração conjunta de glucose e acetato, designado por regime respirativo (R). Assim, o modelo dinâmico para o crescimento da *E. coli* pode ser dividido em dois modelos parciais como se segue:

$$\frac{d}{dt} \begin{bmatrix} X \\ S \\ A \\ O \\ C_{tot} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ -k_1 & -k_2 \\ 0 & k_3 \\ -k_5 & -k_6 \\ k_8 & k_9 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} X - D \begin{bmatrix} X \\ S \\ A \\ O \\ C_{tot} \end{bmatrix} + \begin{bmatrix} 0 \\ \left(\frac{F_{e,S}}{W}\right)S_e \\ 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ OTR \\ -CTR \end{bmatrix} \quad \text{Equação 3.59}$$

Regime respiro-fermentativo (RF)

$$\frac{d}{dt} \begin{bmatrix} X \\ S \\ A \\ O \\ C_{tot} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ -k_1 & 0 \\ 0 & -k_4 \\ -k_5 & -k_7 \\ k_8 & k_{10} \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_3 \end{bmatrix} X - D \begin{bmatrix} X \\ S \\ A \\ O \\ C_{tot} \end{bmatrix} + \begin{bmatrix} 0 \\ \left(\frac{F_{e,S}}{W}\right)S_e \\ 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ OTR \\ -CTR \end{bmatrix} \quad \text{Equação 3.60}$$

Regime respirativo (R)

Note-se que através da divisão do modelo se reduziu o número de reacções a duas ($m = 2$).

Normalmente, é tida em conta uma equação adicional que descreve a variação da massa do líquido no interior do reactor (volume). No caso em que a alimentação é composta apenas por um substrato, pode-se considerar que a variação do peso depende apenas do caudal de alimentação (considera-se $F_{tot} = F_{e,s}$ e $\frac{F_{e,s}}{W} = D$). De facto, alguns autores (Lee e Ramirez, 1992; Akesson *et al.*, 1999; Patnaik, 2002) utilizam esta consideração na simulação destes processos. No entanto, foi verificado por outros autores e no decorrer deste trabalho, que em fermentações de elevada densidade celular e reactores de pequena escala, a quantidade de cultura removida durante a amostragem, a adição de ácido e base e a evaporação, não podem ser considerados negligenciáveis. Assim, a equação mais completa para determinar a variação do peso do meio de cultura no reactor é (adaptado de Galvanauskas *et al.*, 1998):

$$\frac{dW}{dt} = F_{tot} = F_{e,s} + F_b + F_a - F_{evp} - F_{gás} - F_{amostra} \quad \text{Equação 3.61}$$

onde :

F_b	caudal mássico da solução de base adicionada ao reactor;
F_a	caudal mássico da solução de ácido adicionada ao reactor;
F_{evp}	caudal mássico de evaporação do reactor;
$F_{gás}$	caudal mássico retirado do reactor devido a transferências gasosas;
$F_{amostra}$	caudal mássico retirado do reactor devido à amostragem.

Alguns autores (Cockshott e Bogle, 1999; Xu *et al.*, 1999) consideram apenas alguns destes caudais. De acordo com a estrutura do modelo utilizado neste trabalho (Rocha, 2003), as equações utilizadas na simulação dessas variáveis são as seguintes:

$$F_b = c_b \mu_2 XW; F_a = c_a \mu_3 XW; F_{evp} = c_e; F_{gás} = (CTR - OTR)W \quad \text{Equação 3.62}$$

em que c_b representa a constante que relaciona a quantidade de base necessária para equilibrar o pH devido à produção de acetato por unidade de biomassa formada por essa via metabólica; c_a tem o mesmo significado que a constante anterior, mas para a quantidade de ácido e produção de acetato; c_e representa a velocidade de evaporação que é considerada constante.

Modelos cinéticos para a *E. coli*

Assume-se que a velocidade de crescimento de um microrganismo é, de uma forma genérica, uma função do tipo:

$$r_X = \mu(X, S, P, O, T, \text{pH}, \dots)X$$

Equação 3.63

em que,

μ	taxa específica de crescimento
X	concentração de microrganismos (biomassa)

Considera-se que a taxa específica de crescimento μ é uma função complexa de vários factores como a concentração de biomassa X , a concentração de substrato S , a concentração de produto P , a concentração de oxigénio dissolvido no reactor O , a temperatura T , o pH, etc. Na literatura são apresentados diversos modelos cinéticos que descrevem esta dependência (Bastin e Dochain, 1990; Chen, 1992; Bogaerts, 1999).

A equação proposta por *Monod* (1942), em que a taxa específica de crescimento é modelada apenas como função da concentração de substrato limitante, é talvez o modelo cinético mais conhecido e mais utilizado em trabalhos científicos na área da microbiologia e biotecnologia. Tem a forma de uma equação hiperbólica de saturação semelhante a outras utilizadas em cinética enzimática e em fenómenos de adsorção:

$$\mu(S, t) = \frac{\mu_{\text{máx}} S(t)}{K_s + S(t)}$$

Equação 3.64

em que $\mu_{\text{máx}}$ é a taxa específica máxima de crescimento e K_s , designado como constante de *Monod*, é um parâmetro que traduz a afinidade do microrganismo pelo substrato e corresponde à concentração de substrato para a qual μ atinge metade do seu valor máximo.

Têm sido também aplicados, em modelos matemáticos não estruturados de bioprocessos, modelos cinéticos baseados em expressões modificadas da equação de *Monod*. Nessas expressões são introduzidos factores que pretendem descrever mecanismos que originam a diminuição da velocidade de crescimento, como a repressão metabólica e outros efeitos de inibição e limitação. Estes factores afectam significativamente, quer a velocidade de crescimento dos microrganismos, quer a velocidade de formação de produto e por isso, têm um grande impacto no desempenho do reactor (Xu *et al.*, 1999; Lin *et al.*, 2001; Alagappan e Cowan, 2001; Wirtz, 2002; Haag *et al.*, 2005).

No presente trabalho foi utilizado, como modelo cinético para simular os principais fenómenos observados durante a fermentação aeróbia da *E. coli* em operação semi-contínua, descritos com mais pormenor no Capítulo 2, um conjunto de equações derivadas da equação de *Monod*, adaptado do modelo cinético desenvolvido por Rocha (2003). O modelo cinético utilizado é sintetizado na Figura 3.3 e pretende ser um modelo relativamente simples, capaz de poder ser utilizado em algoritmos de identificação, monitorização e controlo, mas simultaneamente completo conseguindo descrever as principais vias metabólicas que podem ocorrer numa fermentação com esta bactéria.

<p>(I) $S > 0, A > 0$</p> $q_s = q_{s,m\acute{a}x} \frac{S}{K_s + S} \frac{K_{i,s}}{K_{i,s} + A}$	$\left\{ \begin{array}{l} q_s < q_{s,crit} \Rightarrow q_{Ac} = q_{Ac,m\acute{a}x} \frac{A}{K_A + A} \frac{K_{i,A}}{K_{i,A} + A} \\ q_s > q_{s,crit} \end{array} \right. \left\{ \begin{array}{l} \mu_1 = \frac{q_{s,crit}}{k_1} \\ \mu_2 = \frac{(q_s - q_{s,crit})}{k_2} \\ \mu_3 = 0 \end{array} \right.$	$\left\{ \begin{array}{l} \mu_1 = \frac{q_s}{k_1} \\ \mu_2 = 0 \\ \mu_3 = \frac{q_{Ac}}{k_4} \end{array} \right.$
<p>(II) $S > 0, A = 0$</p> $q_s = q_{s,m\acute{a}x} \frac{S}{K_s + S} \frac{K_{i,s}}{K_{i,s} + A}$ $q_{Ac} = 0 \Rightarrow \mu_3 = 0$	$\left\{ \begin{array}{l} q_s < q_{s,crit} \\ q_s > q_{s,crit} \end{array} \right. \left\{ \begin{array}{l} \mu_1 = \frac{q_s}{k_1} \\ \mu_2 = 0 \\ \mu_1 = \frac{q_{s,crit}}{k_1} \\ \mu_2 = \frac{(q_s - q_{s,crit})}{k_2} \end{array} \right.$	
<p>(III) $S = 0, A > 0$</p> $q_s = 0 \left\{ \begin{array}{l} \mu_1 = 0 \\ \mu_2 = 0 \end{array} \right.$ $q_{Ac} = q_{Ac,m\acute{a}x} \frac{A}{K_A + A} \frac{K_{i,A}}{K_{i,A} + A} \Rightarrow \mu_3 = \frac{q_{Ac}}{k_4}$		

Figura 3.3 Representação do modelo cinético da *E. coli* (adaptado de Rocha (2003)).

Este modelo tem em conta as três vias metabólicas de crescimento da *E. coli*, assim como a sua capacidade oxidativa limitada em condições de aerobiose. q_s - taxa específica total de consumo de glucose; $q_{s,crit}$ - taxa específica crítica de consumo de glucose (considerado constante); q_{Ac} - taxa específica de consumo de acetato; $K_{i,s}$ - constante de inibição do acetato no consumo de glucose; $K_{i,A}$ - constante de inibição do acetato no seu consumo.

Têm sido descritos na literatura, vários exemplos da implementação de algoritmos que consideram a estrutura das reacções cinéticas desconhecida, na monitorização e controlo de bioprocessos. É disso exemplo a aplicação de observadores assintóticos, os quais permitem estimar variáveis de estado não medidas em linha através de outras variáveis medidas, sem a

modelação das reacções cinéticas (Bastin e Dochain, 1990; Chen, 1992; Bogaerts, 1999; Claes e Van Impe, 1999). Este tema será abordado no Capítulo 4.

Contudo, deverá ser realçado o importante papel do modelo completo como ferramenta indispensável para estudos de simulação usados, por exemplo, no planeamento de experiências, como a seguir se demonstrará.

A formulação do modelo utilizado neste trabalho seguiu as seguintes considerações:

- Os coeficientes de rendimento e os parâmetros cinéticos são considerados constantes;
- A glucose é o único substrato considerado limitante, uma vez que se considera que a concentração de oxigénio nunca atinge valores críticos;
- Existe uma limitação oxidativa no ciclo dos ATC que limita a oxidação completa da glucose. Assim sendo, quando o consumo de glucose excede um determinado limite, parte da glucose segue a via fermentativa, onde o acetato e dióxido de carbono são os principais produtos. A restante glucose segue a via oxidativa, onde o dióxido de carbono é o principal produto. Os rendimentos em biomassa destas duas vias metabólicas diferem significativamente;
- A velocidade de evaporação é constante e não depende da agitação ou qualquer outra variável;
- Considera-se que o reactor apresenta apenas duas fases (sistema gás-líquido), e não se considera que os microrganismos sejam uma fase distinta devido ao seu diminuto tamanho e densidade semelhante à da água;
- O reactor é considerado um sistema homogéneo quer axial quer radialmente;
- A transferência de massa entre a fase gasosa e líquida é explicada pela teoria do filme (Schumpe e Quicker, 1982; Onken e Liefke, 1989; Aiba e Furuse, 1990; Heinzle e Dunn, 1990; Arrua *et al.*, 1990).

3.4.2. Planificação de experiências

O problema da identificação do modelo reduz-se ao problema de identificação dos coeficientes de rendimento do mesmo. Assim sendo, numa etapa inicial procedeu-se ao estudo da sensibilidade do modelo relativamente aos seus parâmetros com o intuito de identificar aqueles que apresentavam uma maior influência no comportamento do modelo. De seguida, estudou-se a identificabilidade teórica ou estrutural dos parâmetros do modelo. Por fim, o desenho óptimo de experiências foi formulado numericamente como um problema de controlo óptimo, cuja finalidade foi determinar o perfil do caudal de alimentação que permitisse maximizar a riqueza informativa da experiência, quantificada por índices relativos à Matriz de Informação de Fisher. A optimização desse perfil foi efectuada recorrendo a algoritmos genéticos, usando uma parametrização dos perfis do caudal de alimentação como uma sequência de funções lineares simples (em rampa). Nesta secção os resultados apresentados dizem respeito apenas a simulações numéricas.

3.4.2.1. Análise de sensibilidade do modelo aos seus parâmetros

Nesta secção foi analisada a sensibilidade do modelo à incerteza associada à estimativa dos coeficientes de rendimento e parâmetros cinéticos. A análise baseou-se no método indicado por Mhaskar *et al.* (2002), com o objectivo de inferir, de um modo sistemático, quais os parâmetros que maior efeito têm no modelo estudado. Esta metodologia pretende seleccionar um subconjunto de parâmetros, de entre os parâmetros desconhecidos do modelo, que apresenta maior influência no desempenho do mesmo, permitindo que no estudo subsequente sejam esses os parâmetros cuja precisão na estimação deverá ser a maior possível.

O estudo foi realizado impondo uma variação de $\pm 10\%$ em valores pré-estabelecidos dos parâmetros. Estes parâmetros foram determinados efectuando cálculos simples no modelo usando para isso um conjunto de dados experimentais (Exp_1). Na Tabela 3.3 apresentam-se esses valores.

Tabela 3.3 Valores pré-estabelecidos dos coeficientes de rendimento e dos parâmetros cinéticos.

Coeficientes de rendimento	Valor inicial	Parâmetros cinéticos	Valor inicial
k_1	1,649	$q_{S,Max}$	0,8769
k_2	29,80	K_S	0,2135
k_3	17,27	$K_{i,S}$	14,32
k_4	31,91	$q_{S,crit}$	0,4938
k_5	0,3036	$q_{Ac,Max}$	0,3913
k_6	36,49	K_A	1,977
k_7	17,11	$K_{i,A}$	2,064
k_8	0,2761		
k_9	52,48		
k_{10}	17,24		

Na Figura 3.4 indica-se o caudal da alimentação utilizado nas simulações efectuadas no estudo das sensibilidades. As condições iniciais para a concentração de biomassa, glucose, acetato e peso do meio de fermentação foram 1,9 g/kg, 0 g/kg, 0 g/kg e 3 kg, respectivamente. As simulações foram efectuadas através da subrotina ODE23s do MATLAB (versão 7.1).

O método de selecção de parâmetros proposto por Mhaskar *et al.* (2002) tem em conta que:

- Os parâmetros devem ter um efeito significativo nas variáveis de estado, calculadas pelo modelo, ou seja, pequenas variações nos parâmetros deverão traduzir-se em grandes alterações nos valores dessas variáveis;

- O efeito de cada parâmetro nas variáveis medidas deverá ser único, de modo a que os seus efeitos não sejam colineares, evitando a obtenção de soluções não únicas quando o estudo é baseado num limitado número de dados experimentais.

Assim sendo, o método adoptado tem em conta, simultaneamente, as sensibilidades do modelo à variação dos seus parâmetros e a colinearidade entre os efeitos dos parâmetros.

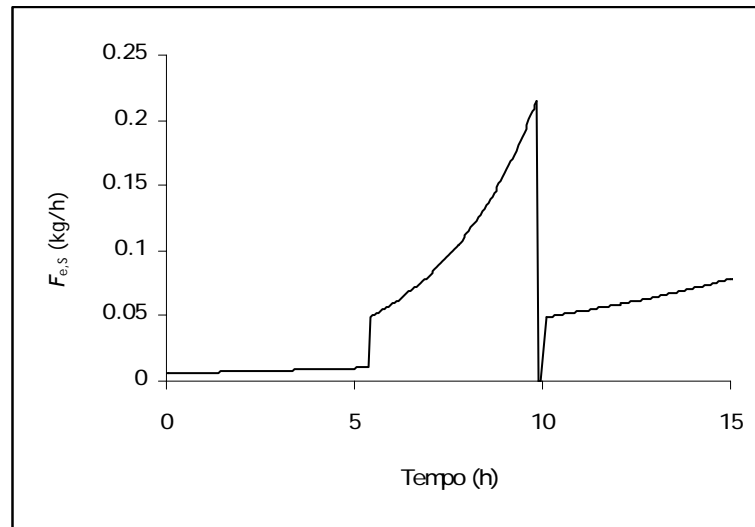


Figura 3.4 Caudal de alimentação usado no estudo das sensibilidades.

O modelo é usado para calcular valores das variáveis de estado (X , A , S , C_{tot} e O , sendo as duas últimas estudadas a partir dos efeitos no CTR e no OTR , respectivamente) com base em valores nominais (Tabela 3.3) e perturbados ($\pm 10\%$) de cada parâmetro estudado. Os elementos da matriz de sensibilidade SP são calculados por:

$$SP_{ij} = \frac{\sum_{z=1}^N \frac{(y_{ijz}^{perturbado} - y_{ijz}^{nominal})^2}{(\bar{y}_{ijz}^{nominal})^2}}{\frac{k_j^{perturbado} - k_j^{nominal}}{k_j^{nominal}}} \quad \text{Equação 3.65}$$

em que,

- i variável de estado em estudo (X , A , S , CTR e OTR);
- j parâmetro do modelo;
- z dado experimental, sendo N o número total de pontos experimentais;
- $y_{ijz}^{perturbado}$ e $y_{ijz}^{nominal}$ valores da variável de estado i , calculados pelo modelo utilizando o parâmetro j perturbado ou nominal, respectivamente;
- $\bar{y}_{ijz}^{nominal}$ média dos valores da variável de estado;
- $k_j^{perturbado}$ e $k_j^{nominal}$ valores do parâmetro j perturbado e nominal, respectivamente.

A Análise de Componentes Principais (ACP) foi aplicada à matriz de sensibilidade SP , para avaliar o efeito de cada parâmetro nos valores de cada variável de estado estudada. A ACP é uma técnica

de análise exploratória multivariada que transforma um conjunto de variáveis correlacionadas num conjunto menor de variáveis independentes, combinações lineares das variáveis originais, designadas por componentes principais (Maroco, 2003).

Os componentes principais são os vectores próprios da matriz covariância $[SP][SP]^T$. A soma ponderada dos componentes principais e os seus valores próprios associados é uma medida do efeito global de cada parâmetro (Mhaskar *et al.*, 2002):

$$w_j = \sum_{i=1}^n |\lambda_i P_{ij}| \quad \text{Equação 3.66}$$

em que, n é o número de parâmetros do modelo sob estudo; P_{ij} é o elemento j do componente principal i (ou seja do vector próprio i); e λ_i é o valor próprio i .

Convém salientar que se usam valores absolutos uma vez que o sinal de cada efeito individual não é relevante. O parâmetro mais influente será aquele que possui o maior efeito global w_j . Os restantes parâmetros são ordenados de modo a providenciarem um compromisso adequado entre o seu efeito global e a sua colinearidade com os parâmetros mais influentes (Mhaskar *et al.*, 2002).

O ângulo formado pelos vectores de sensibilidade de dois parâmetros é uma medida da sua colinearidade:

$$\alpha_{ij} = \arccos\left(\frac{\langle SP_i, SP_j \rangle}{\|SP_i\| \cdot \|SP_j\|}\right) \quad \text{com } \alpha \in [0, \pi] \quad \text{Equação 3.67}$$

em que, SP_i e SP_j são os vectores de sensibilidade dos parâmetros i e j respectivamente; $\langle \cdot, \cdot \rangle$ refere-se ao produto interno; $\|\cdot\|$ refere-se à norma do vector de sensibilidade respectivo.

Assim sendo, se os vectores de sensibilidade de dois parâmetros do modelo forem ortogonais então o seno do ângulo entre eles será igual a 1, pelo que o efeito desses dois parâmetros será facilmente diferenciado. Pelo contrário, se o referido seno for nulo então os dois vectores serão paralelos, ou seja colineares, pelo que o efeito dos parâmetros que lhes estão associados não poderá ser distinguido (Mhaskar *et al.*, 2002).

A identificabilidade de um dado parâmetro i , pode ser definida como o seu efeito global (w_i) ponderado pela sua colinearidade (dada pela Equação 3.67) com os restantes parâmetros:

$$J_{N,i} = w_i \prod_{j=1}^N \text{sen}(\alpha_{ij}) \quad \text{com } i = 1, \dots, N; \quad i \neq j \quad \text{Equação 3.68}$$

onde N é o número total de parâmetros; $J_{N,i}$ traduz a identificabilidade do parâmetro i relativamente aos restantes parâmetros.

Os valores de $J_{N,i}$ bem como os valores dos ângulos entre os vectores de sensibilidade associados aos parâmetros analisados, relativamente a perturbações de $\pm 10\%$ dos valores nominais dos referidos parâmetros, encontram-se na Tabela 3.4, Tabela 3.5 e Tabela 3.6.

Tabela 3.4 Identificabilidade dos parâmetros.

Parâmetro	$J_{N,i} \times 10^{-3}$		Parâmetro	$J_{N,i} \times 10^{-3}$	
	Perturbação +10%	Perturbação -10%		Perturbação +10%	Perturbação -10%
k_1	8,166	11,793	$q_{S,max}$	2,649	4,005
k_2	0,114	0,199	K_S	0,013	0,017
k_3	0,155	0,239	$K_{i,S}$	0,061	0,111
k_4	0,002	0,006	$q_{S,crit}$	6,942	11,724
k_5	0,004	0,009	$q_{Ac,max}$	0,006	0,033
k_6	0,015	0,030	K_A	0,002	0,005
k_7	0,002	0,003	$K_{i,A}$	0,030	0,076
k_8	0,004	0,006			
k_9	0,034	0,056			
k_{10}	0,002	0,003			

A análise da tabela anterior permite verificar que os parâmetros k_1 , $q_{S,crit}$ e $q_{S,max}$ são os que mais influenciam o desempenho do modelo, uma vez que a estes parâmetros estão associados elevados valores de $J_{N,i}$. Tal como referido anteriormente, este índice tem em conta simultaneamente o efeito global do parâmetro e a sua colinearidade com os restantes parâmetros. Deste modo, estes parâmetros deverão ser estimados, dentro do possível, com maior precisão. De entre os coeficientes de rendimento, objecto do estudo desta tese, convém salientar, para além de k_1 , a importância também atribuída pelo estudo a k_2 e k_3 , embora em menor grau.

Na Tabela 3.5 e na Tabela 3.6 apresentam-se os valores dos ângulos entre os vectores de sensibilidade relativos aos diferentes parâmetros estudados. Ângulos iguais a 0° indicam colinearidade entre os vectores de sensibilidade e por conseguinte entre os parâmetros que lhes estão associados. Ângulos iguais a 90° indicam ortogonalidade entre os parâmetros. A análise dos resultados obtidos, independentemente da perturbação provocada, indica que apenas se verifica colinearidade entre os coeficientes de rendimento, tal não se constatando para os parâmetros cinéticos. Mais ainda, podem identificar-se dois grupos de coeficientes de rendimento colineares, cujos efeitos individuais de cada parâmetro dificilmente poderão ser diferenciados: k_5 , k_6 e k_7 ; k_8 , k_9 e k_{10} . Este resultado poderá estar relacionado com o modo como foi estudada a influência desses

parâmetros nas variáveis, isto é, através do cálculo do *OTR* e do *CTR* por aproximação ao *OUR* (*Oxygen Uptake Rate*) e *CER* (*Carbon Dioxide Evolution Rate*), respectivamente⁶. A existência de colinearidade entre aqueles parâmetros reveste-se de grande importância atendendo ao modo como os referidos parâmetros são calculados. Com efeito, regra geral, esses parâmetros são estimados pelo Modelo de Regressão Linear Múltipla (*MRLM*), sendo que, como se verá posteriormente, a ausência de colinearidade é um dos pressupostos do mesmo.

Tabela 3.5 Ângulo entre os vectores de sensibilidade de dois parâmetros: avaliação da colinearidade para uma perturbação de +10% em cada parâmetro.

Parâmetro	α_{ij}° (em gaus), perturbação +10%																
	k_1	k_2	k_3	k_4	k_5	k_6	k_7	k_8	k_9	k_{10}	$q_{S,máx}$	K_S	$K_{i,S}$	$q_{S,crit}$	$q_{Ac,máx}$	K_A	$K_{i,A}$
k_1	0	51	49	80	83	83	83	82	82	82	9	10	29	19	62	59	60
k_2	51	0	4	70	76	76	76	75	75	75	43	60	34	33	28	16	24
k_3	49	4	0	73	78	78	78	78	78	78	42	58	34	31	30	19	27
k_4	80	70	73	0	35	35	35	55	55	55	74	79	57	77	43	56	47
k_5	83	76	78	35	0	0	0	90	90	90	80	83	69	83	54	63	57
k_6	83	76	78	35	0	0	0	90	90	90	80	83	69	83	54	63	57
k_7	83	76	78	35	0	0	0	90	90	90	80	83	69	83	54	63	57
k_8	82	75	78	55	90	90	90	0	0	0	77	81	64	78	65	71	67
k_9	82	75	78	55	90	90	90	0	0	0	77	81	64	78	65	71	67
k_{10}	82	75	78	55	90	90	90	0	0	0	77	81	64	78	65	71	67
$q_{S,máx}$	9	43	42	74	80	80	80	77	77	77	0	17	20	12	53	50	52
K_S	10	60	58	79	83	83	83	81	81	81	17	0	34	28	68	67	67
$K_{i,S}$	29	34	34	57	69	69	69	64	64	64	20	34	0	21	35	35	35
$q_{S,crit}$	19	33	31	77	83	83	83	78	78	78	12	28	21	0	48	43	46
$q_{Ac,máx}$	62	28	30	43	54	54	54	65	65	65	53	68	35	48	0	12	4
K_A	59	16	19	56	63	63	63	71	71	71	50	67	35	43	12	0	8
$K_{i,A}$	60	24	27	47	57	57	57	67	67	67	52	67	35	46	4	8	0

Complementarmente, a ACP foi ainda efectuada recorrendo ao *software -Statistical Package for the Social Sciences - SPSS 11.5*. Esta análise foi realizada para a matriz *SP* calculada para perturbações de $\pm 10\%$ dos valores nominais de cada parâmetro. Os resultados obtidos permitem concluir que, para o caso em estudo, apenas é necessária uma dimensão, ou seja um componente principal, para explicar as variações observadas quer para parâmetros com uma perturbação de +10% quer para os parâmetros perturbados de -10%. Este componente principal apresenta valores

⁶ $OTR \cong OUR = (k_5\mu_1 + k_6\mu_2 + k_7\mu_3)X$; $CTR \cong CER = (k_8\mu_1 + k_9\mu_2 + k_{10}\mu_3)X$

próprios, para as perturbações de +10% e -10%, iguais a 4,697 e 4,745; o que corresponde a 93,94% e 94,89% da variância total, respectivamente. Mais ainda, a análise realizada conduziu a valores de α de Cronbach de 0,984 e 0,987 para as perturbações de +10% e -10%, respectivamente, o que sendo próximo de 1, permite inferir da adequada fidelidade de cada dimensão e do modelo em geral.

Tabela 3.6 Ângulo entre os vectores de sensibilidade de dois parâmetros: avaliação da colinearidade para uma perturbação de -10% em cada parâmetro.

Parâmetro	α_{ij}° (em graus), perturbação -10%																
	k_1	k_2	k_3	k_4	k_5	k_6	k_7	k_8	k_9	k_{10}	$q_{S,máx}$	K_S	$K_{i,S}$	$q_{S,crit}$	$q_{Ac,máx}$	K_A	$K_{i,A}$
k_1	0	55	54	67	75	75	75	71	71	71	13	14	24	44	60	58	58
k_2	55	0	5	70	76	76	76	75	75	75	45	62	34	18	10	32	19
k_3	54	5	0	74	78	78	78	79	79	79	43	60	34	13	15	36	23
k_4	67	70	74	0	35	35	35	55	55	55	73	79	57	79	62	39	52
k_5	75	76	78	35	0	0	0	90	90	90	79	83	69	83	68	51	60
k_6	75	76	78	35	0	0	0	90	90	90	79	83	69	83	68	51	60
k_7	75	76	78	35	0	0	0	90	90	90	79	83	69	83	68	51	60
k_8	71	75	79	55	90	90	90	0	0	0	76	81	64	82	74	63	69
k_9	71	75	79	55	90	90	90	0	0	0	76	81	64	82	74	63	69
k_{10}	71	75	79	55	90	90	90	0	0	0	76	81	64	82	74	63	69
$q_{S,máx}$	13	45	43	73	79	79	79	76	76	76	0	17	20	32	51	55	52
K_S	14	62	60	79	83	83	83	81	81	81	17	0	36	48	69	70	69
$K_{i,S}$	24	34	34	57	69	69	69	64	64	64	20	36	0	30	37	36	35
$q_{S,crit}$	44	18	13	79	83	83	83	82	82	82	32	48	30	0	27	45	34
$q_{Ac,máx}$	60	10	15	62	68	68	68	74	74	74	51	69	37	27	0	24	10
K_A	58	32	36	39	51	51	51	63	63	63	55	70	36	45	24	0	13
$K_{i,A}$	58	19	23	52	60	60	60	69	69	69	52	69	35	34	10	13	0

Na Figura 3.5 pode visualizar-se a contribuição de cada parâmetro para a dimensão única do modelo de componentes principais, com base nos *scores* dos factores para cada caso. Os *scores* são combinações lineares das variáveis originais estandardizadas (Pestana e Gageiro, 2005). A análise dos resultados permite verificar que 3 dos 17 parâmetros analisados apresentam um comportamento bastante diferente dos restantes: $q_{S,crit}$, k_1 e $q_{S,máx}$. Mais ainda, pelo afastamento relativamente ao valor zero que apresentam, pode concluir-se que são aqueles que mais contribuem para a variação das respostas das variáveis de estado estudadas (e representadas todas pela dimensão 1),

independentemente da perturbação provocada. Este facto está de acordo com as conclusões extraídas anteriormente, com base no índice $J_{N,i}$.

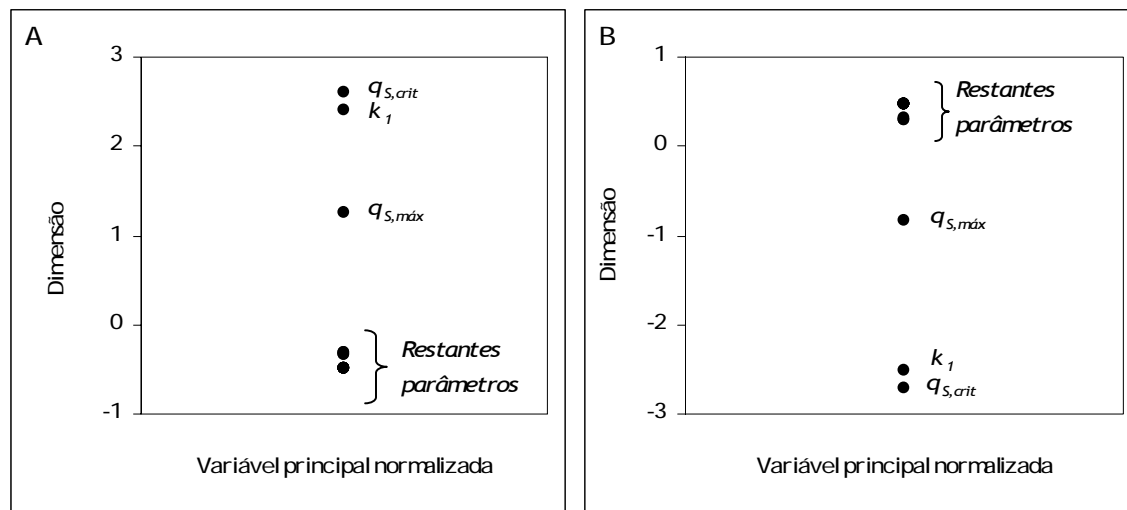


Figura 3.5 Scores de cada parâmetro na dimensão 1 do modelo de componentes principais para uma perturbação de: (A) +10% e (B) -10%.

3.4.2.2. Projecto dos regressores para os modelos parciais a duas reacções

Como já referido, a estrutura do modelo, assim como as variáveis de estado que podem ser medidas experimentalmente, limita o número de parâmetros que podem ser estimados. Quando o modelo global é considerado (Equação 3.55) a característica da matriz K é igual a três e, de acordo com a condição necessária estabelecida por Chen (1992), é possível somente a identificação de dois coeficientes de rendimento em cada uma das três reacções. Assim, só poderão ser identificados seis dos dez coeficientes de rendimento, sendo por isso necessário introduzir relações algébricas entre dois coeficientes em cada reacção ou então considerar como conhecidos quatro desses coeficientes. As relações utilizadas poderão ser baseadas em relações estequiométricas conhecidas. Alternativamente, o procedimento mais adequado será proceder à identificação utilizando os modelos parciais a duas reacções.

Como referido anteriormente, partiu-se do princípio que as propriedades de identificabilidade dos coeficientes de rendimento, para um dado sistema caracterizado pela matriz K , são independentes da escolha da partição (Chen, 1992). Das várias partições que se poderiam estabelecer para o presente caso de estudo, adoptaram-se duas tendo em conta o cálculo posterior dos k 's a partir dos valores dos θ 's.

Na Tabela 3.7 e Tabela 3.8 apresenta-se o projecto dos regressores para as partições consideradas dos modelos parciais dados pela Equação 3.59 (regime RF) e pela Equação 3.60 (regime R).

Tabela 3.7 Partições induzidas nas variáveis de estado (ξ) e de entrada (U) para os modelos parciais RF e R.

Partição XA	Partição SO
$\xi_a = \begin{bmatrix} X \\ A \end{bmatrix}$	$\xi_a = \begin{bmatrix} S \\ O \end{bmatrix}$
$\xi_b = \begin{bmatrix} S \\ O \\ C_{tot} \end{bmatrix}$	$\xi_b = \begin{bmatrix} X \\ A \\ C_{tot} \end{bmatrix}$
$U_a = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$U_a = \begin{bmatrix} F_{e,s}S_e/W \\ OTR \end{bmatrix}$
$U_b = \begin{bmatrix} F_{e,s}S_e/W \\ OTR \\ -CTR \end{bmatrix}$	$U_b = \begin{bmatrix} 0 \\ 0 \\ -CTR \end{bmatrix}$

Pela tabela anterior verifica-se que com a partição XA , que é designada como “*nice partition*”, a dinâmica de Z já não depende da matriz H , uma vez que, como referido na secção 3.3.1, o vector U_a , é um vector nulo. Para a partição SO isso já não acontece o que levará à utilização do modelo auxiliar (Equação 3.35 à Equação 3.37).

Tabela 3.8 Partições induzidas de K e correspondente matriz H para os modelos parciais RF e R.

Partição	Modelo parcial	K_a	K_b	H
XA	RF	$\begin{bmatrix} 1 & 1 \\ 0 & k_3 \end{bmatrix}$	$\begin{bmatrix} -k_1 & -k_2 \\ -k_5 & -k_6 \\ k_8 & k_9 \end{bmatrix}$	$\frac{1}{k_3} \begin{bmatrix} k_1k_3 & k_2 - k_1 \\ k_5k_3 & k_6 - k_5 \\ -k_8k_3 & k_8 - k_9 \end{bmatrix}$
	R	$\begin{bmatrix} 1 & 1 \\ 0 & -k_4 \end{bmatrix}$	$\begin{bmatrix} -k_1 & 0 \\ -k_5 & -k_7 \\ k_8 & k_{10} \end{bmatrix}$	$\frac{1}{k_4} \begin{bmatrix} k_1k_4 & k_1 \\ k_5k_4 & k_5 - k_7 \\ -k_8k_4 & k_{10} - k_8 \end{bmatrix}$
SO	RF	$\begin{bmatrix} -k_1 & -k_2 \\ -k_5 & -k_6 \end{bmatrix}$	$\begin{bmatrix} 1 & 1 \\ 0 & k_3 \\ k_8 & k_9 \end{bmatrix}$	$\frac{1}{k_1k_6 - k_2k_5} \begin{bmatrix} k_6 - k_5 & k_1 - k_2 \\ -k_3k_5 & k_1k_3 \\ k_6k_8 - k_5k_9 & k_1k_9 - k_2k_8 \end{bmatrix}$
	R	$\begin{bmatrix} -k_1 & 0 \\ -k_5 & -k_7 \end{bmatrix}$	$\begin{bmatrix} 1 & 1 \\ 0 & -k_4 \\ k_8 & k_{10} \end{bmatrix}$	$\frac{1}{k_1k_7} \begin{bmatrix} k_7 - k_5 & k_1 \\ k_4k_5 & -k_1k_4 \\ k_7k_8 - k_5k_{10} & k_1k_{10} \end{bmatrix}$

Para os modelos parciais a característica da matriz K é dois. Assim, verifica-se que é teoricamente possível determinar seis coeficientes em cada um dos regimes. Pela tabela anterior

pode-se constatar que para o modelo R, é possível determinar teoricamente todos os coeficientes de rendimento, enquanto que para o modelo RF só será possível identificar seis dos sete coeficientes. Para ultrapassar essa limitação considerou-se que k_6 era um valor conhecido com base em experiências anteriormente realizadas. Esta condição é uma condição necessária mas não suficiente para a identificabilidade teórica. Sendo assim, os coeficientes podem ser teoricamente identificáveis e, no entanto, numericamente não ser possível determiná-los, uma vez que na prática os erros experimentais e o ruído dos dados experimentais podem causar uma acentuada propagação de erros conduzindo a valores de parâmetros com grandes incertezas. Assim sendo, procedeu-se à análise das duas partições escolhidas com o intuito de estudar a influência da sua escolha na propagação de erros e na estabilidade do método de recuperação dos valores de k através dos valores de θ estimados. De seguida descreve-se, a título de exemplo, o projecto dos regressores para a partição XA .

Considera-se genericamente a seguinte configuração para a matriz H :

$$H = \begin{bmatrix} \theta_1 & \theta_2 \\ \theta_3 & \theta_4 \\ \theta_5 & \theta_6 \end{bmatrix} \quad \text{Equação 3.69}$$

Fazendo a transformação de estado ($Z \equiv H\xi_a + \xi_b$), obtém-se:

$$Z = \begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \end{bmatrix} = \begin{bmatrix} \theta_1 X + \theta_2 A + S \\ \theta_3 X + \theta_4 A + O \\ \theta_5 X + \theta_6 A + C_{tot} \end{bmatrix} \quad \text{Equação 3.70}$$

com a correspondente dinâmica:

$$\frac{dZ}{dt} = \frac{d}{dt} \begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \end{bmatrix} = -D \begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \end{bmatrix} + \begin{bmatrix} \left(\frac{F_{e,S}}{W}\right) S_e \\ OTR \\ -CTR \end{bmatrix} \quad \text{Equação 3.71}$$

cujas condições iniciais são:

$$Z_1(0) = \theta_1 X(0) + \theta_2 A(0) + S(0),$$

$$Z_2(0) = \theta_3 X(0) + \theta_4 A(0) + O(0),$$

$$Z_3(0) = \theta_5 X(0) + \theta_6 A(0) + C_{tot}(0).$$

A Equação 3.70 permite escrever as seguintes equações do modelo de regressão linear múltipla (MRLM):

$$y_1(t) = \theta_1 \phi_1(t) + \theta_2 \phi_2(t) \quad \text{Equação 3.72}$$

$$y_2(t) = \theta_3 \phi_1(t) + \theta_4 \phi_2(t) \quad \text{Equação 3.73}$$

$$y_3(t) = \theta_5 \phi_1(t) + \theta_6 \phi_2(t) \quad \text{Equação 3.74}$$

com $y_1(t)=Z_1(t)-S(t)$, $y_2(t)=Z_2(t)-O(t)$, $y_3(t)=Z_3(t)-C_{tot}(t)$ e os regressores dados por $\phi_1(t)=X(t)$, $\phi_2(t)=A(t)$.

Os parâmetros de regressão θ_i são determinados através do método dos mínimos quadrados. Na Tabela 3.9 apresentam-se as fórmulas de cálculo para recuperação dos valores dos coeficientes de rendimento a partir dos θ_i estimados. Essas expressões são deduzidas por resolução das relações da Equação 3.30, usando para cada regime metabólico e para cada partição estudada, as matrizes H definidas na Tabela 3.8.

Uma vez que a respiração da glucose ocorre em ambos os regimes metabólicos, os coeficientes de rendimento associados à via oxidativa da glucose (k_1 , k_5 e k_8) terão naturalmente a mesma expressão de cálculo.

A validação da identificação é feita por comparação das estimativas de S , O e C_{tot} calculadas através da Equação 3.70, com os valores experimentais destas variáveis. Têm-se assim as equações seguintes:

$$\hat{S}(t) = \hat{Z}_1(t) - \theta_1 \phi_1(t) - \theta_2 \phi_2(t) \quad \text{Equação 3.75}$$

$$\hat{O}(t) = \hat{Z}_2(t) - \theta_3 \phi_1(t) - \theta_4 \phi_2(t) \quad \text{Equação 3.76}$$

$$\hat{C}_{tot}(t) = \hat{Z}_3(t) - \theta_5 \phi_1(t) - \theta_6 \phi_2(t) \quad \text{Equação 3.77}$$

A matriz de informação de Fisher, calculada através da Equação 3.45, pode ser escrita numa forma discreta no tempo, como se segue:

$$MIF = \frac{1}{n_t} \begin{bmatrix} \sum_{i=1}^{n_t} \phi_1^2(t_i) & \sum_{i=1}^{n_t} \phi_1(t_i) \phi_2(t_i) \\ \sum_{i=1}^{n_t} \phi_1(t_i) \phi_2(t_i) & \sum_{i=1}^{n_t} \phi_2^2(t_i) \end{bmatrix} \quad \text{Equação 3.78}$$

Tabela 3.9 Expressões de cálculo dos coeficientes de rendimento.

Partição	Coeficientes de rendimento	Modelo parcial	
		RF	R
XA	k_1	θ_1	θ_1
	k_2	$(\theta_1\theta_4 + \theta_2k_6 - \theta_2\theta_3)/\theta_4$	----
	k_3	$(k_6 - \theta_3)/\theta_4$	----
	k_4	----	θ_1/θ_2
	k_5	θ_3	θ_3
	k_6	36,49 ⁽ⁱ⁾	----
	k_7	----	$(\theta_2\theta_3 - \theta_1\theta_4)/\theta_2$
	k_8	$-\theta_5$	$-\theta_5$
	k_9	$(\theta_3\theta_6 - \theta_4\theta_5 + \theta_6k_6)/\theta_4$	----
	k_{10}	----	$(\theta_1\theta_6 - \theta_2\theta_5)/\theta_2$
SO	k_1	$\theta_4/(\theta_1\theta_4 - \theta_2\theta_3)$	$\theta_4/(\theta_1\theta_4 - \theta_2\theta_3)$
	k_2	$(1 - \theta_2k_6)/\theta_1$	----
	k_3	$(\theta_3 + \theta_1\theta_4k_6 - \theta_2\theta_3k_6)/\theta_1$	----
	k_4	----	$-\theta_4/\theta_2$
	k_5	$-\theta_3/(\theta_1\theta_4 - \theta_2\theta_3)$	$-\theta_3/(\theta_1\theta_4 - \theta_2\theta_3)$
	k_6	36,49 ⁽ⁱ⁾	----
	k_7	----	$1/\theta_2$
	k_8	$(\theta_4\theta_5 - \theta_3\theta_6)/(\theta_1\theta_4 - \theta_2\theta_3)$	$(\theta_4\theta_5 - \theta_3\theta_6)/(\theta_1\theta_4 - \theta_2\theta_3)$
	k_9	$(\theta_1\theta_6k_6 - \theta_2\theta_5k_6 + \theta_5)/\theta_1$	----
	k_{10}	----	θ_6/θ_2

(i) no regime respiro-fermentativo, o valor de k_6 foi fixado num valor conhecido.

3.4.2.2.1. Optimização do perfil do caudal de alimentação

Na planificação de experiências para a identificação numérica dos parâmetros do modelo, a escolha da frequência e do tempo de amostragem, da localização dos sensores, das variáveis de estado medidas assim como das variáveis de entrada controladas, é de grande relevância devendo ser realizada criteriosamente. Apesar da importância reconhecida, esta tarefa não é de fácil implementação devido quer a limitações de ordem prática quer à dificuldade de ter em conta múltiplos factores em simultâneo. De entre esses factores, a escolha das variáveis de entrada é o de

maior importância, uma vez que estas podem ser utilizadas para direccionar o comportamento do sistema do modo mais adequado.

O desenho óptimo de experiências foi formulado como um problema de controlo óptimo: determinar o perfil óptimo do caudal de alimentação que permita obter experiências de elevado conteúdo informativo com o intuito de estimar os coeficientes de rendimento a partir das variáveis de estado medidas. Como método de optimização foram utilizados os algoritmos genéticos e evolutivos GEATbx 3.5 (Pohlheim, 2004) e como funções objectivo o determinante (critério *D*) e o número de condicionamento (critério *E* modificado) da *MIF*.

Este método de optimização funciona com vários operadores genéticos sendo as variáveis utilizadas reais. Numa primeira fase o algoritmo gera um conjunto de possíveis soluções designado por população. A solução potencial (indivíduo), é codificada como um vector denominado cromossoma, representando uma solução possível no espaço de procura multidimensional. A qualidade de cada indivíduo da população é avaliada utilizando um critério de selecção baseado nos valores da função objectivo. Após verificação da qualidade de todos os cromossomas da população uma nova população de indivíduos é gerada a partir da população actual recorrendo a três operadores genéticos: reprodução, *crossover* e mutação. O número de indivíduos avaliado em cada iteração foi de 200. Cada indivíduo consistia numa sequência de valores de $F_{e,s}$ compreendidos entre 0 e 0,4 kg/h (valor mínimo e máximo permitido pela bomba de alimentação utilizada).

Os valores dos parâmetros e as condições iniciais usadas durante as simulações e optimização apresentam-se na Tabela 3.3 e na Tabela 3.10, respectivamente.

Tabela 3.10 Condições iniciais utilizadas nas simulações e optimização.

Condições iniciais	
$X(0)$	1,5 g/kg
$S(0)$	0 g/kg
$A(0)$	0 g/kg
$C(0)$	0,272 g/kg
$O(0)$	0,0027 g/kg
$W(0)$	3 kg
S_e	250 g/kg

A optimização do caudal de alimentação foi sujeita às seguintes restrições:

- O tempo inicial da alimentação foi fixado em $t_0=0$ horas;
- O tempo final da fermentação era livre, não podendo, no entanto ser superior a 25 horas;
- As concentrações iniciais de biomassa, substrato, acetato, oxigénio e dióxido de carbono dissolvidos assim como a concentração de glucose na alimentação foram fixadas;
- O peso inicial do meio de cultura, no interior do fermentador, foi fixado;

- O peso máximo do meio de cultura foi fixado em 5 kg, em virtude de ser a capacidade útil máxima do fermentador usado;
- O caudal máximo de alimentação foi limitado pela capacidade da bomba peristáltica usada, tendo sido fixado em 0,4 kg/h;
- Não foi permitida a ocorrência de alternância de regimes mais do que uma vez no decorrer da fermentação, isto é, apenas era permitida uma de duas situações: regime respirativo seguido de regime respiro-fermentativo ou o inverso.

O processo de optimização foi interrompido sempre que o valor da função objectivo permanecia constante durante 500 iterações, sendo que em cada optimização foram efectuadas pelo menos 2000 iterações.

Em virtude de numa fermentação poderem ocorrer consecutivamente quer o regime respirativo quer o regime respiro-fermentativo, realizaram-se diversas tentativas de optimização simultânea de ambos os regimes. Para tal, procedeu-se à optimização do produto dos determinantes da *MIF* para cada regime (Velo *et al.*, 2004). Nessas optimizações verificou-se uma das seguintes situações: pese embora se optimizassem ambos os regimes o caudal de alimentação óptimo conduzia à ocorrência de apenas um dos regimes ou a um número de horas de fermentação no regime respiro-fermentativo muito superior ao do regime respirativo, independentemente da partição escolhida (*XA* ou *SO*). Constatou-se ainda que, na mesma optimização, o valor do determinante da *MIF* para o regime respirativo era muito superior ao valor do determinante da *MIF* para o regime respiro-fermentativo, facto que não se traduzia num maior conteúdo informativo para o regime respirativo. Para tentar atenuar a diferença de grandeza entre os dois determinantes procedeu-se ainda à optimização da soma dos logaritmos decimais de cada determinante. Contudo, esta opção não se revelou satisfatória, continuando a verificarem-se as situações anteriormente descritas.

Assim sendo, optou-se pela optimização individual de cada regime, tendo-se avaliado, para além, do determinante da *MIF*, outro critério de optimalidade: o número de condicionamento da *MIF*. Uma vez que no caso deste último critério é conhecido o valor óptimo (igual a um) a interpretação dos resultados poderia ser facilitada.

Estes dois critérios, como mencionado anteriormente, são complementares e de certo modo competitivos, estando o primeiro relacionado com a precisão global dos parâmetros e o segundo com a sua precisão individual. A optimização atendendo a cada um destes critérios individualmente poderá não conduzir ao aumento da precisão que lhe está associada, o que geometricamente pode ser traduzido pela obtenção de elipsóides longas e estreitas, no caso do critério *D*, ou de elipsóides esféricas mas muito largas, no caso do critério *E* modificado (secção 3.2.2.1). Por outro lado, a optimização de acordo com cada uma dos dois critérios individuais poderá ainda favorecer um tipo de precisão em detrimento do outro. Para tentar ultrapassar esta dificuldade optou-se também pela optimização simultânea dos dois critérios, recorrendo à optimização multiobjectivo.

A otimização Pareto é uma medida da eficiência em problemas multicritérios, isto é, em situações multiobjectivo, que em geral conduzem a um conjunto de soluções óptimas, isto é, soluções Pareto-óptimo. Nestes casos existem dois ou mais critérios (ou funções objectivo) medidos em diferentes unidades não existindo uma conversão que permita exprimir todos os critérios nas mesmas unidades. Uma solução pode ser considerada como um Pareto-óptimo se não existe nenhuma outra solução cujo resultado seja pelo menos tão bom em todos os critérios avaliados e estritamente melhor em pelo menos um desses critérios (Zhou e Titchener-Hooker, 2003; Sarkar e Modak, 2005). Ou seja, uma solução Pareto-óptimo não pode ser melhorada sem prejudicar pelo menos um dos critérios tidos em conta. As soluções que são Pareto-óptimo são também designadas, na literatura, por não-dominadas, não-inferiores ou Pareto-eficientes. Uma solução não é um Pareto-óptimo se um dos critérios pode ser melhorado sem degradar os restantes. Estas soluções são conhecidas como soluções dominadas ou inferiores. A optimalidade de Pareto, para um problema com apenas dois critérios, como é o caso em estudo, pode ser facilmente visualizada graficamente na Figura 3.6. Cada critério (ou função objectivo) é representado num eixo, sendo que se pretende minimizar ambos os critérios. Nesta figura, as soluções Pareto-óptimo são aquelas que não têm qualquer ponto abaixo ou à sua esquerda.

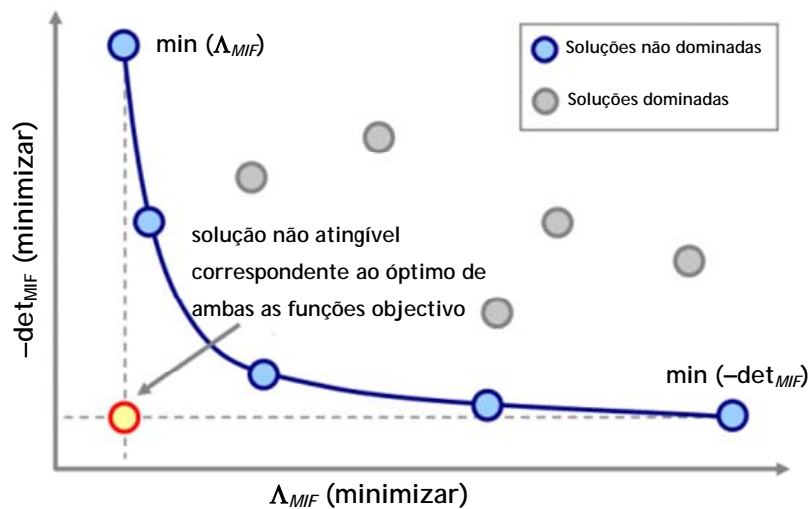


Figura 3.6 Representação gráfica do conjunto de soluções Pareto-óptimo.

Os algoritmos genéticos evolutivos são reconhecidos como particularmente aplicáveis na resolução deste tipo de otimizações multiobjectivo, uma vez que são capazes de lidar simultaneamente com um conjunto de possíveis soluções, permitindo que um conjunto de soluções Pareto-óptimo seja alcançado numa única corrida do algoritmo (Sarkar e Modak, 2005). Encontram-se referenciadas na literatura diversas aplicações para a otimização de fermentações semi-contínuas (Wang e Cheng, 1999; Wang *et al.*, 2001; Na *et al.*, 2002; Nougues *et al.*, 2002; Banga *et al.*, 2002; Rocha e Ferreira, 2004).

Na Tabela 3.11 apresenta-se o valor da função objectivo para os diferentes casos estudados.

Tabela 3.11 Valores do determinante (\det_{MIF}) e do número de condicionamento (Λ_{MIF}) da *MIF* obtidos na optimização dos critérios *D* e *E* modificado.

Critério	Função objectivo	Casos analisados				
		<i>XA_RF</i>	<i>XA_R</i>	<i>SO_RF</i>	<i>SO_R</i>	
<i>D</i>	\det_{MIF}	\det_{MIF}	1139	42260	10658	17445
		$\Lambda_{MIF}^{(i)}$	174	258	305	560
<i>E</i> modificado	Λ_{MIF}	Λ_{MIF}	85	71	257	12
		$\det_{MIF}^{(ii)}$	5	292	484	0,154
<i>D + E</i> modificado	\det_{MIF}		23	19879	7893	0,455
		Λ_{MIF}	200	274	258	14

(i) Λ_{MIF} calculado para o perfil alimentação optimizado com o critério *D*

(ii) \det_{MIF} calculado para o perfil alimentação optimizado com o critério *E* modificado

Os perfis dos caudais de alimentação obtidos para cada regime, partição e critério de optimização encontram-se representados na Figura 3.7 à Figura 3.10. Para cada regime e partição, verifica-se que o perfil optimizado depende do critério de optimalidade utilizado, embora para os critérios de optimalidade *E* modificado e *D* e *E* modificado em simultâneo, os perfis obtidos sejam semelhantes ainda que em alguns casos com diferentes magnitudes.

As correspondentes elipses dos desvios encontram-se representadas na Figura 3.11. Os dados da Tabela 3.11 conjuntamente com a Figura 3.11 permitem avaliar a riqueza informativa das experiências, que é traduzida pela independência dos regressores. Existem dois factores que podem influenciar o grau de independência dos regressores: a sua colinearidade e a diferença entre as suas magnitudes. Convém realçar que a observação das elipses dos desvios permite uma rápida avaliação da qualidade dos resultados obtidos, sendo que, graficamente, a maior riqueza informativa de uma experiência é traduzida por elipses que tendam para círculos e cuja área seja mínima.

Para as duas partições estudadas, quer para o regime respirativo quer para o regime respiro-fermentativo, verifica-se que as optimizações realizadas com o critério *D* conduzem a elipses de menor área. Para o regime RF verifica-se que, para ambas as partições, existe uma certa colinearidade entre os regressores bem como proporcionalidade, embora mais acentuada para a partição *SO*. Para o regime R e para a partição *XA* os regressores não apresentam colinearidade, embora tenham diferentes magnitudes, o que é traduzido por elipses quase circulares. Para esse mesmo regime e para a partição *SO* os regressores apresentam maior proximidade de magnitudes traduzindo-se em elipses mais estreitas. No entanto, em virtude de em alguns pontos os regressores coincidirem, em especial para as optimizações realizadas segundo o critério *E* modificado e segundo o critério *D* e *E* modificado em simultâneo, faz com que as elipses alonguem acentuadamente.

Com base nos resultados obtidos pode afirmar-se que, em geral, o critério D é o que conduz a otimizações mais satisfatórias, para o caso em estudo.

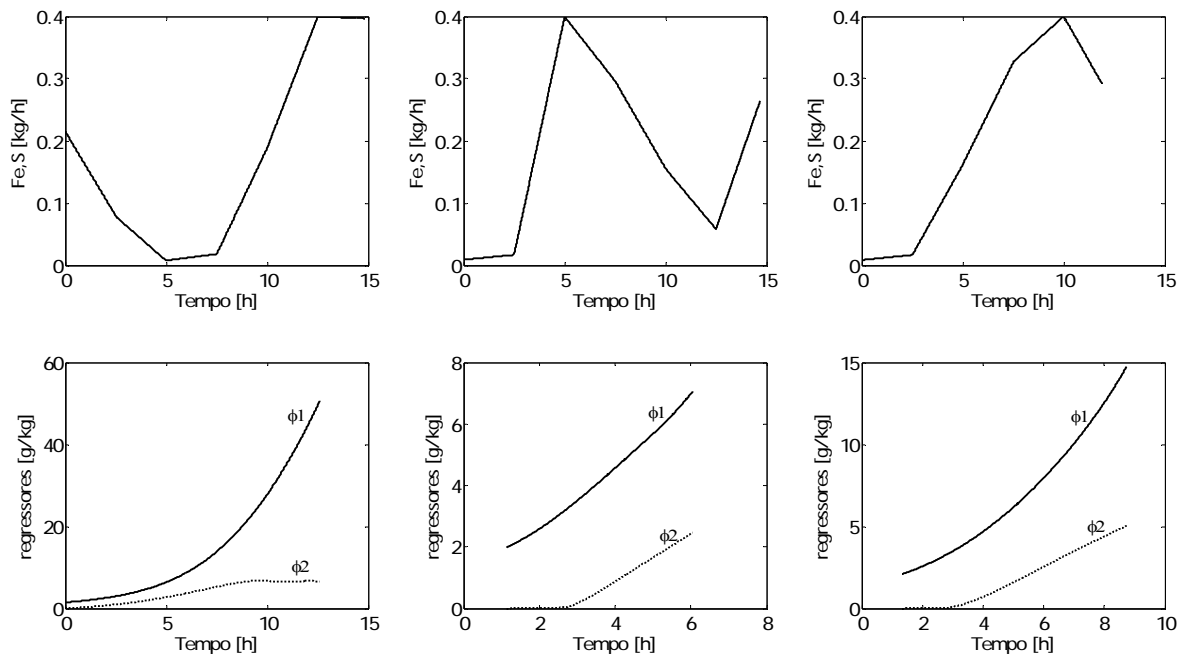


Figura 3.7 Perfis do caudal de alimentação e dos regressores obtidos para a partição XA_RF na otimização do \det_{MIF} (coluna da esquerda), do Δ_{MIF} (coluna do meio) e do $\det_{MIF} + \Delta_{MIF}$ (coluna da direita).

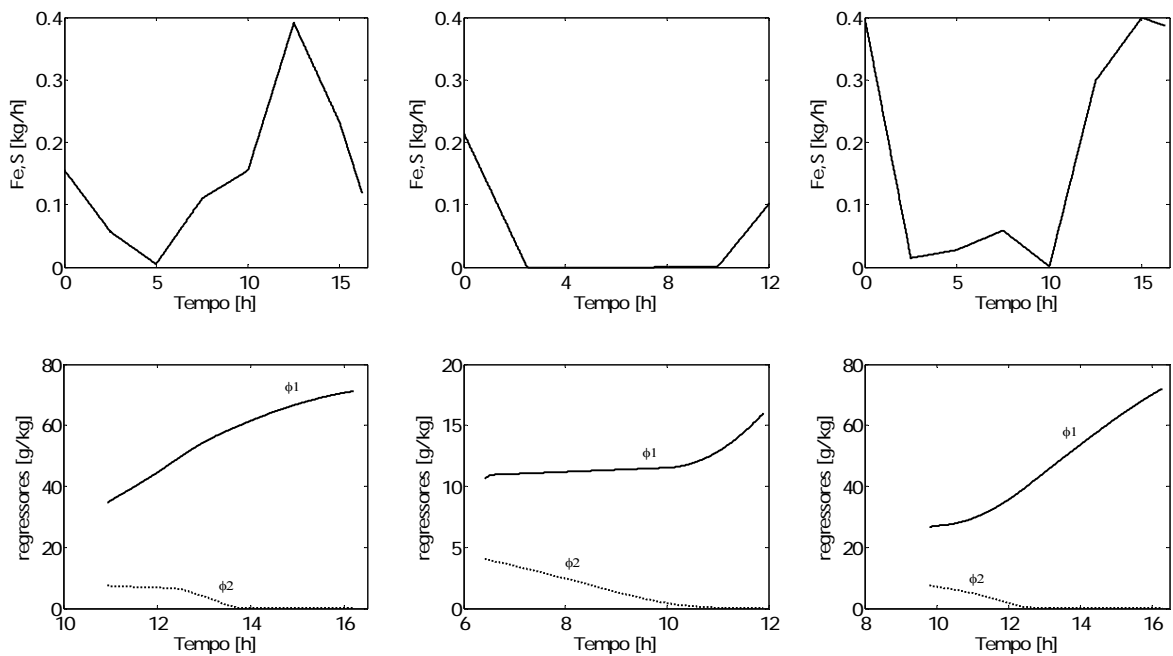


Figura 3.8 Perfis do caudal de alimentação e dos regressores obtidos para a partição XA_R na otimização do \det_{MIF} (coluna da esquerda), do Δ_{MIF} (coluna do meio) e do $\det_{MIF} + \Delta_{MIF}$ (coluna da direita).

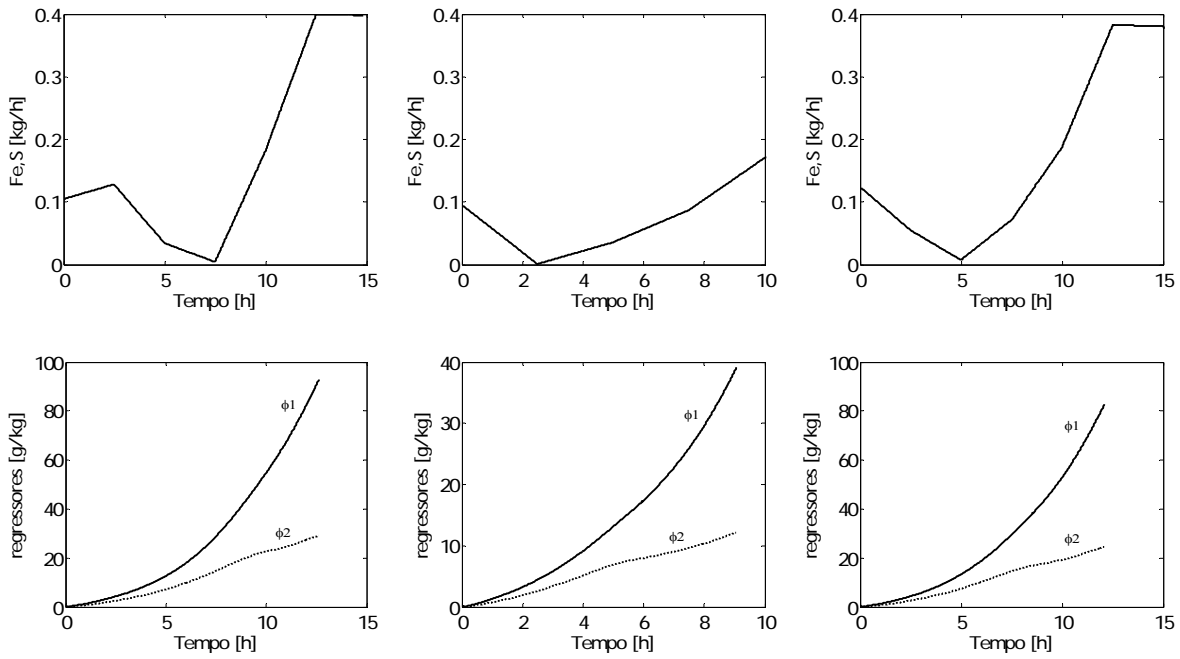


Figura 3.9 Perfis do caudal de alimentação e dos regressores obtidos para a partição SO_{RF} na otimização do \det_{MIF} (coluna da esquerda), do Λ_{MIF} (coluna do meio) e do $\det_{MIF} + \Lambda_{MIF}$ (coluna da direita).

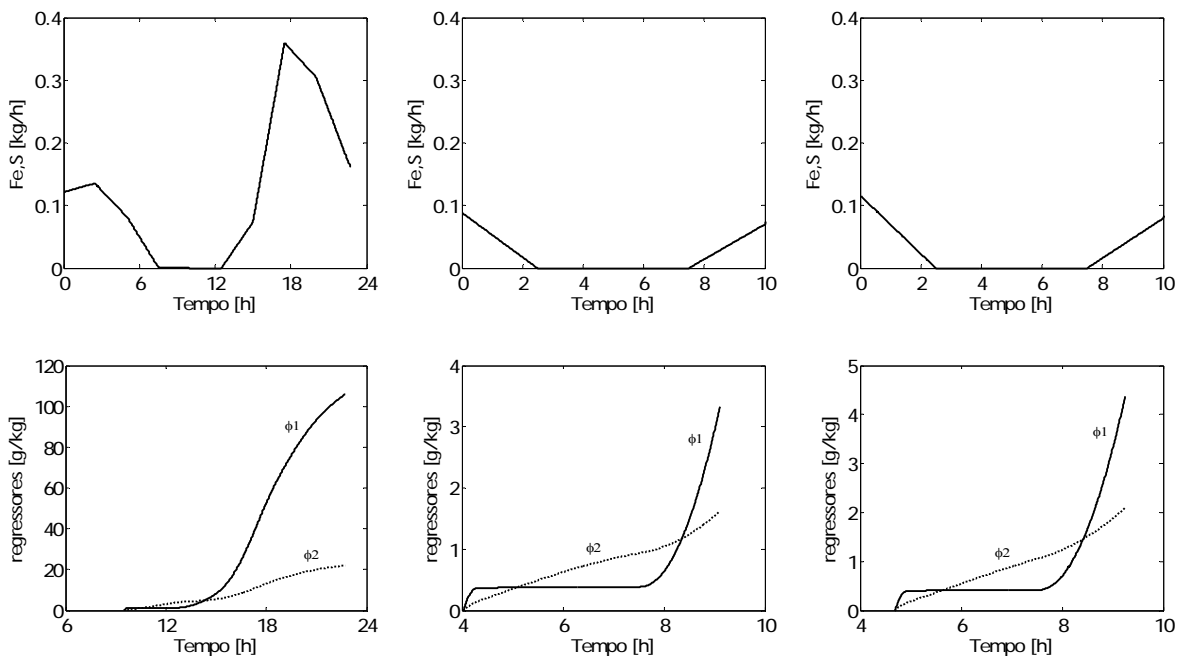


Figura 3.10 Perfis do caudal de alimentação e dos regressores obtidos para a partição SO_R na otimização do \det_{MIF} (coluna da esquerda), do Λ_{MIF} (coluna do meio) e do $\det_{MIF} + \Lambda_{MIF}$ (coluna da direita).

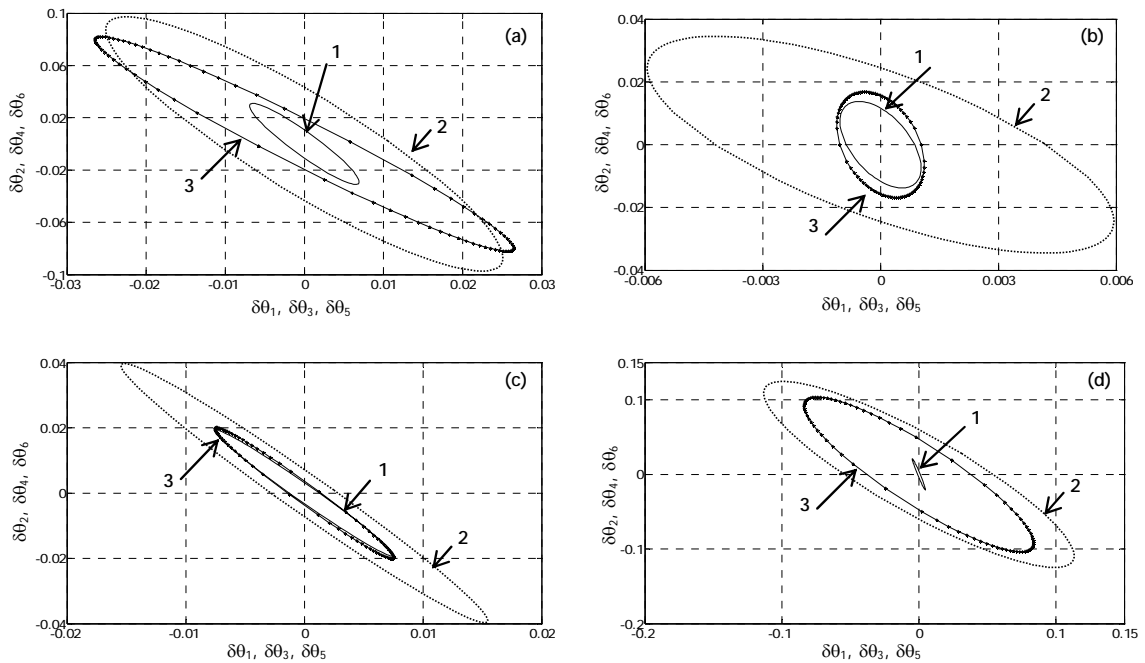


Figura 3.11 Elipses dos desvios dos parâmetros para XA_{RF} (a); XA_R (b); SO_{RF} (c) e SO_R (d):
1- optimização do \det_{MIF} ; 2- optimização do Λ_{MIF} e 3- optimização $\det_{MIF} + \Lambda_{MIF}$.

Convém referir que nas optimizações realizadas, independentemente do critério de optimalidade escolhido, verificaram-se problemas de convergência que poderão estar associados à não linearidade do problema sob estudo, às restrições impostas bem como à existência de óptimos locais, isto é, diferentes perfis de alimentação conduzem ao mesmo valor da função objectivo.

3.4.2.2.2. Diagnóstico da regressão linear múltipla do projecto de regressores

Tal como foi referido na secção anterior, os coeficientes de regressão θ_i da Equação 3.72 à Equação 3.74 são estimados pelo método dos mínimos quadrados. O Modelo de Regressão Linear Múltipla (*MRLM*) usado, não é mais do que uma técnica estatística, descritiva e inferencial que permite a análise da relação entre uma variável dependente, no caso do modelo univariado, e um conjunto de variáveis independentes. A aplicação do referido modelo conduz a equações ligeiramente diferentes das apresentadas (Equação 3.72 à Equação 3.74):

$$y_1(t) = \theta_1 \phi_1(t) + \theta_2 \phi_2(t) + \varepsilon_1 \quad \text{Equação 3.79}$$

$$y_2(t) = \theta_3 \phi_1(t) + \theta_4 \phi_2(t) + \varepsilon_2 \quad \text{Equação 3.80}$$

$$y_3(t) = \theta_5 \phi_1(t) + \theta_6 \phi_2(t) + \varepsilon_3$$

Equação 3.81

em que ε_i representa os erros ou resíduos aleatórios do modelo associados aos valores de y_i .

O modelo de regressão adoptado é do tipo I, pelo que se assume que apenas a variável dependente é afectada por erros de medição, enquanto que as variáveis independentes são medidas sem erros, ou quanto muito que o erro é desprezável face à amplitude das suas variações, ou seja, a variável dependente é aleatória e as variáveis independentes são fixas (Maroco, 2003). A utilização deste modelo carece da verificação das seguintes hipóteses (Maroco, 2003; Quinn e Keough, 2004; Pestana e Gageiro, 2005):

- linearidade do fenómeno em estudo;
- para cada valor fixo das variáveis independentes, a variável dependente tem uma distribuição normal com variância constante;
- as observações da variável independente são independentes umas das outras;
- as variáveis aleatórias residuais, que descrevem os efeitos na variável independentes não explicados pelas variáveis dependentes, têm uma distribuição normal, com média zero e variância constante;
- as variáveis aleatórias residuais referentes a duas observações diferentes não estão correlacionadas, sendo portanto independentes entre si, apresentando uma covariância nula;
- as variáveis independentes são ortogonais e portanto não há multicolinearidade, isto é, as variáveis explicativas devem ser independentes.

Assim sendo, após a estimação do *MRLM* convém verificar a qualidade do ajustamento na amostra, com base em medidas absolutas (desvios padrões dos parâmetros ou coeficientes da regressão) e em medidas relativas (R de *Pearson* ou coeficiente de correlação amostral, R^2 ou coeficiente de determinação e o R^2 ajustado). Mais ainda, é conveniente validar, isoladamente, os coeficientes da regressão e em termos globais o modelo, o que pode ser feito recorrendo aos testes *t* e *F*, este último sendo uma análise de variância (ANOVA).

Uma vez que o *MRLM* explica a relação linear entre as variáveis independentes e a variável dependente numa situação em que os seis pressupostos anteriormente enunciados se verificam, torna-se imperioso, após a estimação do modelo, proceder ao seu diagnóstico, analisando-se, separadamente as hipóteses da homocedasticidade, da independência e da normalidade dos resíduos, bem como a hipótese de multicolinearidade entre as variáveis independentes (Maroco, 2003; Pestana e Gageiro, 2005):

▪ **Homocedasticidade**

O estudo da hipótese da variância constante das variáveis aleatórias residuais pode ser feito por observação das relações, por um lado, entre os resíduos estudantizados e os resíduos estandardizados previstos, e por outro lado, entre os resíduos estandardizados e os valores estimados da variável dependente. Assume-se a homocedasticidade dos resíduos se ambos os gráficos indicam que os resíduos mantêm uma amplitude aproximadamente constante em relação ao eixo horizontal zero, não mostrando tendências crescentes ou decrescentes.

▪ **Independência dos resíduos (auto-correlação)**

Para testar este pressuposto, ou seja, que a magnitude de um resíduo não influencia a magnitude do resíduo seguinte, testa-se a existência ou a ausência da correlação serial entre resíduos consecutivos. A estatística proposta por *Durbin-Watson* (d_{DW}) pode ser utilizada para testar a presença ou ausência de auto-correlação (positiva ou negativa) entre os resíduos do *MRLM*. Se $d_{DW} \approx 2 (\pm 0,2)$ pode concluir-se que não existe auto-correlação entre os resíduos. Para valores muito inferiores ou superiores a 2 existe auto-correlação positiva ou negativa, respectivamente.

Uma forma mais exacta de usar este teste consiste em comparar o valor de d_{DW} com valores tabelados (Savin e White, 1977) para modelos de regressão com constante (limites inferior e superior: d_L e d_U) e sem constante (limites inferior e superior: d_M e d_U), para níveis de significância de 1 e 5%, para diferentes níveis de dados e variáveis independentes usados na estimação do *MRLM*. Convém referir ainda que, no caso de *MRLM* sem constante existem valores de d_M tabelados para as situações de auto-correlação positiva e negativa, d_{M+} e d_{M-} , respectivamente. A tabela seguinte indica as decisões a tomar em função dos valores de d_{DW} , de d_M e de d_U , para o *MRLM* sem constante, usado neste estudo.

Tabela 3.12 Intervalos de rejeição e de não rejeição de auto-correlação de dados.

Regiões de rejeição e de não rejeição					
Valor de d_{DW}	$[0, d_{M+}[$	$[d_{M+}, d_U[$	$[d_U, 4-d_U[$	$[4-d_U, 4-d_{M-}[$	$[4-d_{M-}, 4[$
Decisão	Auto-correlação positiva	Nada se pode concluir	Não existe auto-correlação	Nada se pode concluir	Auto-correlação negativa

▪ **Normalidade dos resíduos**

Tanto o teste anterior como o do modelo de regressão pressupõem a verificação da normalidade nas variáveis aleatórias residuais. A normalidade dos resíduos (estandardizados e não estandardizados) pode ser testada usando o teste de *Kolmogorov-Smirnov*. Se o valor de prova (valor p) for superior ao nível de significância do teste (α), fixado em 0,05 neste estudo, então concluir-se-á pela normalidade da distribuição. Os desvios à normalidade podem ainda ser observados graficamente representando os resíduos estandardizados ou os resíduos estudantizados em função dos valores previstos. No primeiro caso a distribuição será aproximadamente normal se a

maioria dos pontos estiver aproximadamente em cima da diagonal principal. No outro caso, convém que os pontos se distribuam aleatoriamente e o mais próximo possível da linha horizontal zero.

▪ **Multicolinearidade**

A intensidade da multicolinearidade pode ser estudada através da análise de: (i) Correlação entre as variáveis independentes consideradas no modelo: se os coeficientes de correlação entre as variáveis independentes são elevados (superiores em valor absoluto a 0,9) indiciam a possibilidade de existência de multicolinearidade; (ii) Tolerância, que mede a proporção da variação de uma variável independente que não é explicada pelas restantes, e o factor de inflação da variância (VIF), que é o inverso da tolerância. Se o valor da tolerância for superior a 0,1 ou o de VIF for inferior a 10 concluir-se-á pela inexistência de multicolinearidade; e (iii) Valores próprios (caso sejam próximos de zero, em geral inferiores a 0,5), índice de condicionamento (superior a 30) e proporção de variância (superior a 0,9) indicam possível multicolinearidade entre as variáveis independentes do modelo.

A análise estatística da regressão linear múltipla foi realizada recorrendo ao programa SPSS 11.5. Para o efeito usaram-se dados simulados para cada uma das duas partições estudadas (XA e SO) e para cada um dos dois modelos parciais considerados (RF e R).

O estudo estatístico realizado permitiu verificar a qualidade do ajuste com base nas medidas absolutas e relativas, assim como validar os coeficientes de regressão isoladamente e globalmente. Com efeito, para as duas partições consideradas e para os dois modelos parciais tidos em conta, os *MRLM* (que passam na origem) obtidos para as regressões de y_i (para $i = 1, 2, 3$) em função de ϕ_1 e ϕ_2 são altamente significativos (teste ANOVA: $p=0,000 < 0,05$), os coeficientes da regressão determinados são também significativos (teste t : $p=0,000 < 0,05$) e com desvios padrão associados reduzidos ($\leq 0,008$), sendo os coeficientes de correlação amostral, de determinação e o R^2 ajustado iguais a 1,000.

A homocedasticidade dos resíduos foi avaliada, graficamente, para os casos *XA_RF*, *XA_R*, *SO_RF* e *SO_R* (Figura 3.12 à Figura 3.15, respectivamente). Os gráficos obtidos mostram que, em geral, os resíduos mantêm uma amplitude aproximadamente constante em relação ao eixo horizontal zero, isto é, não mostram tendências crescentes ou decrescentes, pelo que não se rejeita a hipótese da homocedasticidade, indicando que os parâmetros estimados do *MRLM* são centrados e eficientes.

Apesar da qualidade dos resultados, procedeu-se ainda ao diagnóstico do *MRLM* para cada uma das partições estudadas e modelos parciais considerados. Da Tabela 3.13 à Tabela 3.16 são apresentados os valores obtidos para cada uma das três regressões realizadas em cada situação estudada.

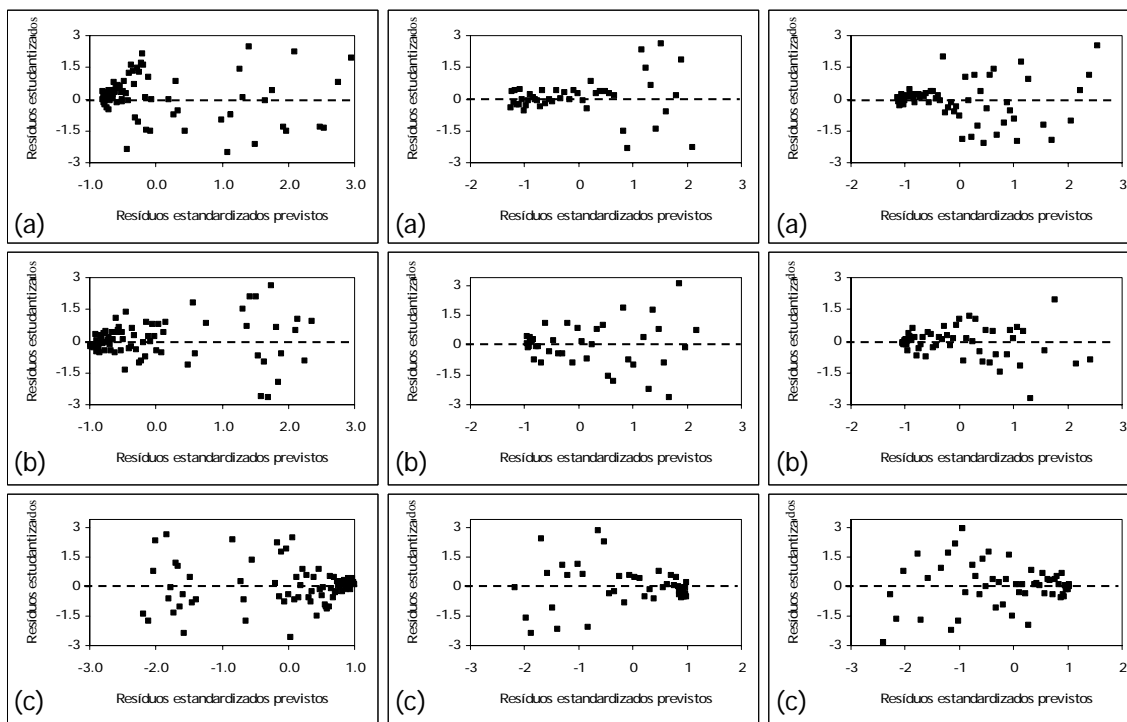


Figura 3.12 Estudo da homocedasticidade das variáveis aleatórias residuais (ε_i) para a partição XA , modelo parcial RF na otimização do \det_{MIF} (coluna da esquerda), do Λ_{MIF} (coluna do meio) e do $\det_{MIF} + \Lambda_{MIF}$ (coluna da direita): (a) Equação 3.79; (b) Equação 3.80; (c) Equação 3.81.

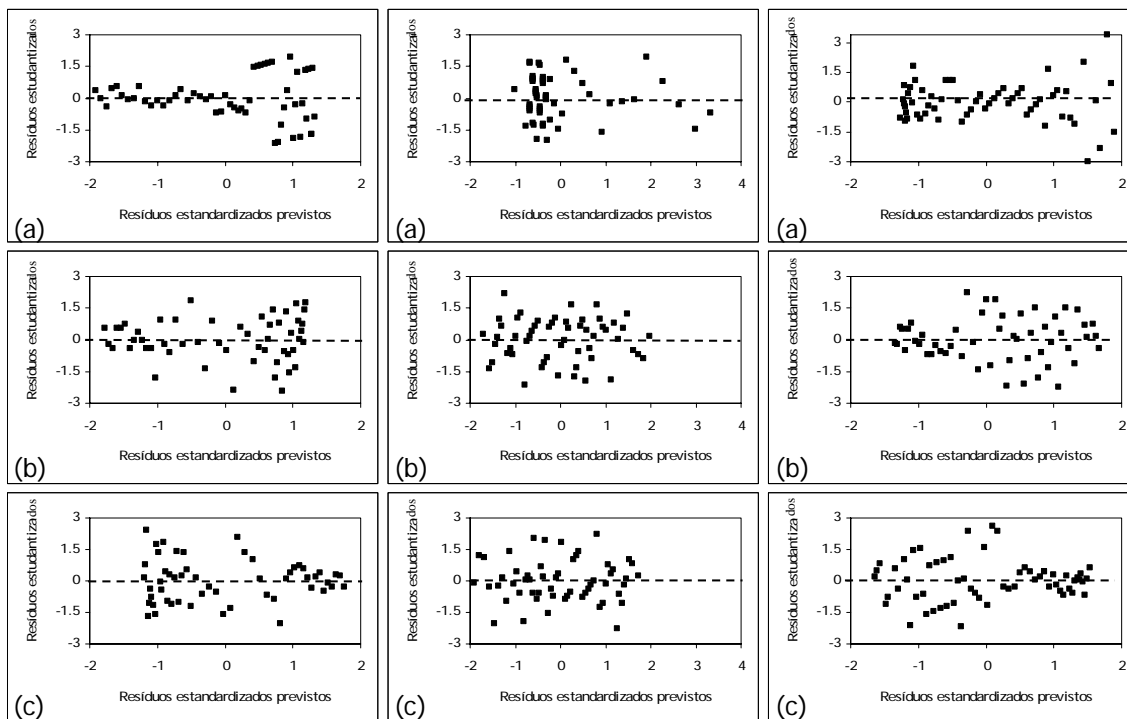


Figura 3.13 Estudo da homocedasticidade das variáveis aleatórias residuais (ε_i) para a partição XA , modelo parcial R na otimização do \det_{MIF} (coluna da esquerda), do Λ_{MIF} (coluna do meio) e do $\det_{MIF} + \Lambda_{MIF}$ (coluna da direita): (a) Equação 3.79; (b) Equação 3.80; (c) Equação 3.81.

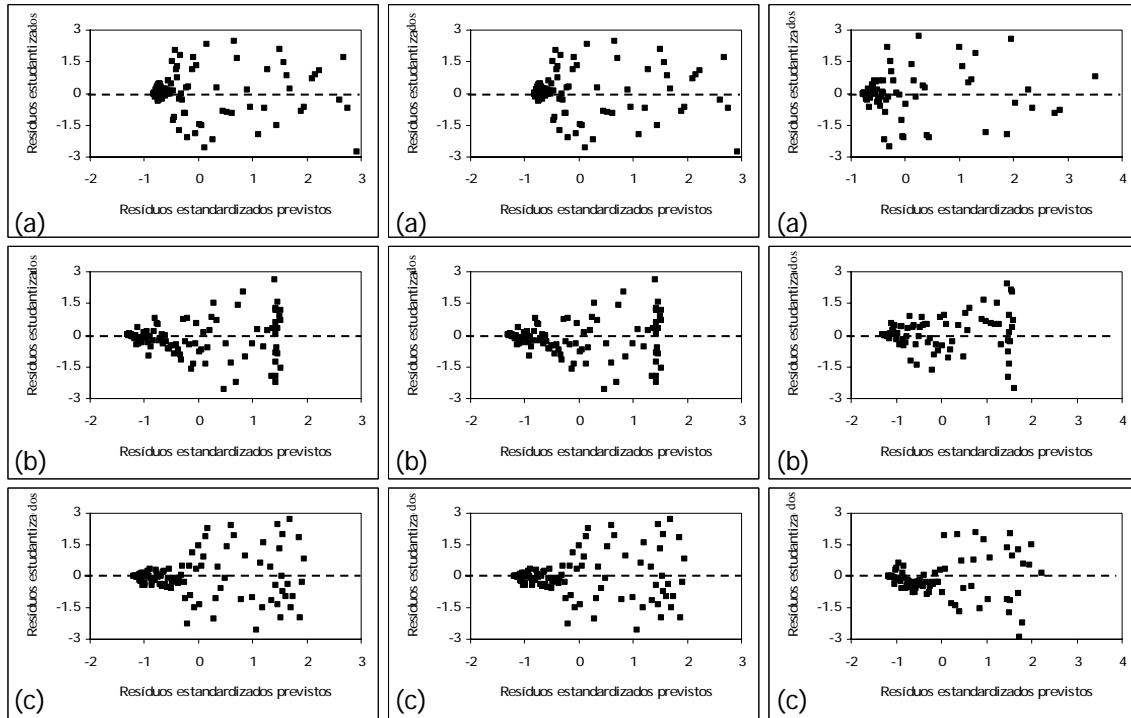


Figura 3.14 Estudo da homocedasticidade das variáveis aleatórias residuais (ε_i) para a partição SO , modelo parcial RF na otimização do \det_{MIF} (coluna da esquerda), do Λ_{MIF} (coluna do meio) e do $\det_{MIF} + \Lambda_{MIF}$ (coluna da direita): (a) Equação 3.79; (b) Equação 3.80; (c) Equação 3.81.

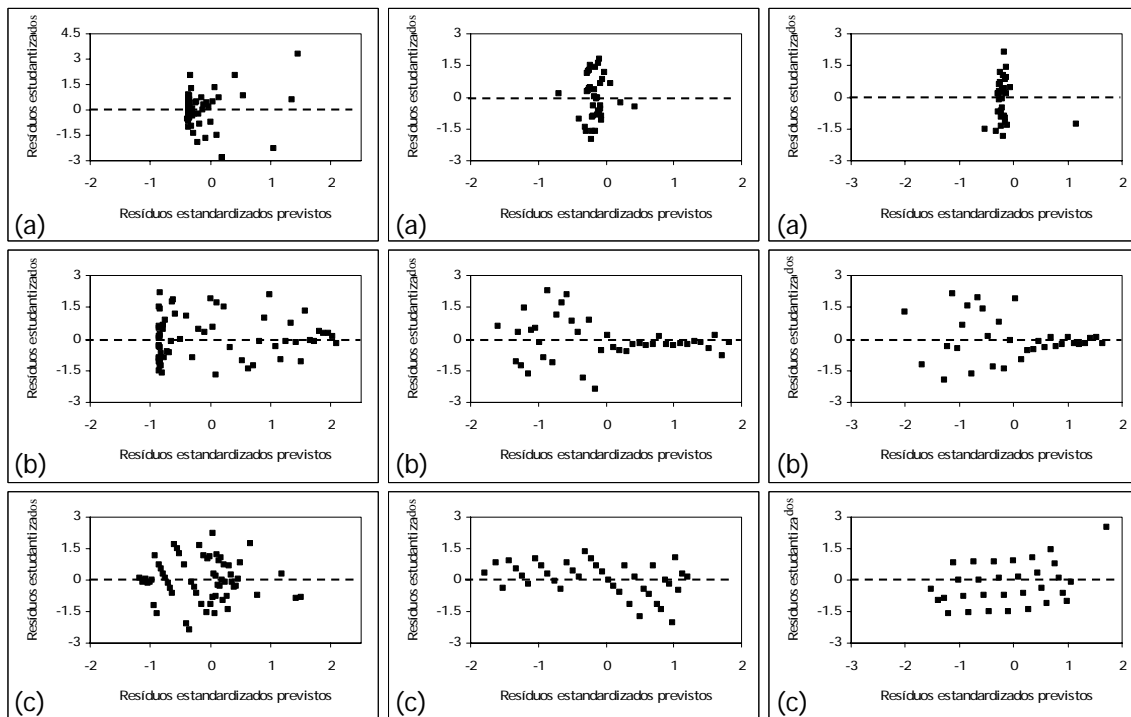


Figura 3.15 Estudo da homocedasticidade das variáveis aleatórias residuais (ε_i) para a partição SO , modelo parcial R na otimização do \det_{MIF} (coluna da esquerda), do Λ_{MIF} (coluna do meio) e do $\det_{MIF} + \Lambda_{MIF}$ (coluna da direita): (a) Equação 3.79; (b) Equação 3.80; (c) Equação 3.81.

Tabela 3.13 Diagnóstico do modelo *MRLM* para a partição *XA* e modelo parcial RF.

		Partição <i>XA</i> , modelo parcial RF								
Pressuposto	Parâmetro	\det_{MIF}			Λ_{MIF}			$\det_{MIF+\Lambda_{MIF}}$		
		y_1	y_2	y_3	y_1	y_2	y_3	y_1	y_2	y_3
Independência dos resíduos	d_{DW}	1,835	2,000	2,191	1,981	2,677	2,014	2,165	2,325	2,605
Normalidade dos resíduos	K-S (valor p)	0,081	0,120	0,120	0,052	0,502	0,320	0,054	0,183	0,135
	Coeficiente de correlação		0,940			0,879			0,955	
	Tolerância		0,116			0,227			0,088	
	VIF		8,628			4,413			11,337	
Multicolinearidade	Valor próprio (mínimo)		0,0597			0,121			0,0451	
	Índice de condicionamento (máximo)		5,699			3,948			6,582	
	Proporção de variância (máxima)		0,97			0,94			0,98	

Tabela 3.14 Diagnóstico do modelo *MRLM* para a partição *XA* e modelo parcial R.

		Partição <i>XA</i> , modelo parcial R								
Pressuposto	Parâmetro	\det_{MIF}			Λ_{MIF}			$\det_{MIF+\Lambda_{MIF}}$		
		y_1	y_2	y_3	y_1	y_2	y_3	y_1	y_2	y_3
Independência dos resíduos	d_{DW}	1,814	2,083	1,684	1,676	1,879	2,013	2,323	2,355	2,515
Normalidade dos resíduos	K-S (valor p)	0,541	0,680	0,931	0,775	0,722	0,891	0,882	0,882	0,980
	Coeficiente de correlação		0,532			0,705			0,399	
	Tolerância		0,717			0,503			0,840	
	VIF		1,394			1,989			1,190	
Multicolinearidade	Valor próprio (mínimo)		0,468			0,295			0,601	
	Índice de condicionamento (máximo)		1,809			2,405			1,527	
	Proporção de variância (máxima)		0,77			0,85			0,70	

Tabela 3.15 Diagnóstico do modelo *MRLM* para a partição *SO* e modelo parcial RF.

		Partição <i>SO</i> , modelo parcial RF								
Pressuposto	Parâmetro	\det_{MIF}			Λ_{MIF}			$\det_{MIF+\Lambda_{MIF}}$		
		y_1	y_2	y_3	y_1	y_2	y_3	y_1	y_2	y_3
Independência dos resíduos	d_{DW}	2,089	1,857	2,312	1,986	2,219	2,027	1,788	2,211	2,214
Normalidade dos resíduos	K-S (valor p)	0,061	0,187	0,067	0,053	0,369	0,880	0,062	0,091	0,052
	Coeficiente de correlação		0,984			0,986			0,980	
Multicolinearidade	Tolerância		0,031			0,028			0,040	
	VIF		31,942			35,253			25,039	
	Valor próprio (mínimo)		0,0158			0,0143			0,0202	
	Índice de condicionamento (máximo)		11,214			11,790			9,907	
	Proporção de variância (máxima)		0,99			0,99			0,99	

Tabela 3.16 Diagnóstico do modelo *MRLM* para a partição *SO* e modelo parcial R.

		Partição <i>SO</i> , modelo parcial R								
Pressuposto	Parâmetro	\det_{MIF}			Λ_{MIF}			$\det_{MIF+\Lambda_{MIF}}$		
		y_1	y_2	y_3	y_1	y_2	y_3	y_1	y_2	y_3
Independência dos resíduos	d_{DW}	2,241	1,702	2,197	1,871	2,343	2,079	2,047	2,279	1,920
Normalidade dos resíduos	K-S (valor p)	0,632	0,564	0,881	0,918	0,384	0,513	0,589	0,198	0,894
Multicolinearidade	Coeficiente de correlação		0,852			0,836			0,827	
	Tolerância		0,275			0,301			0,317	
	VIF		3,638			3,319			3,158	
	Valor próprio (mínimo)		0,148			0,164			0,173	
	Índice de condicionamento (máximo)		3,531			3,345			3,246	
Proporção de variância (máxima)		0,93			0,92			0,91		

A análise das tabelas anteriores mostra que as hipóteses da independência e da normalidade dos resíduos são verificadas para ambas as partições estudadas e modelos parciais analisados. É de referir que, embora em alguns casos se tenham obtido valores de d_{DW} diferentes de 2 ($\pm 0,2$), após efectuada a análise indicada na Tabela 3.12, pode concluir-se pela ausência de auto-correlação dos

resíduos. Relativamente à ausência de colinearidade entre as variáveis independentes (ϕ_1 e ϕ_2), o estudo realizado permitiu verificar que a partição *SO_RF* apresentava fortes indícios de multicolinearidade. Assim sendo, a Equação 3.72 à Equação 3.74 (ou Equação 3.79 à Equação 3.81) não são válidas para a partição *SO_RF*, podendo ser aplicadas apenas para a partição *XA_RF*, *XA_R* e *SO_R*.

Com base nos resultados anteriores e, embora teoricamente a identificação dos parâmetros devesse ser independente da partição escolhida (Chen, 1992), no estudo efectuado tal não ocorreu, pelo que é aconselhável verificar a aplicabilidade do *MRLM*.

3.4.2.2.3. Estabilidade matemática do problema de recuperação dos *k*'s através dos θ 's

Tal como anteriormente referido, as partições *XA* e *SO* foram escolhidas com o objectivo de avaliar a sua influência na propagação de erros bem como na estabilidade do método de recuperação dos valores de *k* através dos valores de θ estimados. O problema em análise, para cada partição e regime estudados, pode ser expresso pela Equação 3.30, que após manipulação algébrica adequada pode ser traduzida por um sistema de equações lineares (SEL) da forma genérica $Ax=b$. Assim sendo, o estudo pretendido pode ser realizado pela análise dos efeitos que pequenas perturbações introduzidas nos dados do problema, podem originar nos resultados *x*. Podem ser considerados três casos: as perturbações são introduzidas no vector *b* (termo independente); as perturbações são introduzidas apenas nos elementos da matriz *A* e por fim alterações são introduzidas quer em *b* quer em *A* (Fernandes, 1998).

Deste modo, a sensibilidade da solução do SEL, relativa às variações introduzidas nos dados do problema (*b*, *A*, ou *b* e *A*), mede a estabilidade matemática do problema, pelo que, se os resultados não forem afectados de grandes variações, o SEL diz-se bem condicionado. Nas três situações possíveis verifica-se que o erro relativo nos resultados é limitado pelos erros nos dados multiplicados pelo produto das normas subordinadas⁷ da matriz *A* e da sua matriz inversa A^{-1} , o qual pode ser definido como sendo o número de condição da matriz *A* (Fernandes, 1998):

$$n_{cond}(A) = \|A\| \|A^{-1}\| \geq 1 \quad \text{Equação 3.82}$$

O número de condição de *A* mede a sensibilidade da solução *x*, do SEL, em relação a pequenas variações introduzidas nos dados (vector *b* e matriz *A*). Daí designar-se o sistema bem condicionado se $n_{cond}(A)$ for pequeno e de, provavelmente, mal condicionado se $n_{cond}(A)$ for elevado.

⁷ Dada a norma de um vector $\|\cdot\|$ e uma matriz *A*, defina-se $\|Ax\|$ para todos os vectores que verificam $\|x\| = 1$. A norma subordinada (ou induzida) à norma de vectores é dada por $\|A\| = \text{máximo}_{\|x\|=1} \{\|Ax\|\}$.

Analogamente, se o determinante da matriz dos coeficientes do sistema na forma escalonada⁸ for próximo de zero, em valor absoluto, o sistema diz-se mal condicionado (Fernandes, 1998).

Na Tabela 3.17 apresentam-se os números de condição e os valores dos determinantes da matriz escalonada para a partição *XA* e *SO* e regimes RF e R.

Tabela 3.17 Valores do número de condição e do determinante da forma escalonada para as duas partições estudadas.

	Partição <i>XA</i>		Partição <i>SO</i>	
	Regime RF	Regime R	Regime RF	Regime R
Número de condição	9,82	30,8	11,4	94,4
Determinante da forma escalonada	-0,203	1,00	0,559	0,538

Os resultados anteriores mostram que, à excepção da partição *SO* e regime R, pode-se admitir que os restantes casos analisados são bem condicionados (baixos valores do número de condição e valores dos determinantes da forma escalonada suficientemente afastados de zero, em valor absoluto), apresentando por isso uma estabilidade matemática satisfatória. De facto, quando comparados os números de condição com os valores dos elementos envolvidos em cada sistema, verifica-se que são cerca de 3, 30, 8 e 50 vezes superiores para *XA_RF*, *XA_R*, *SO_RF* e *SO_R*, respectivamente. De um modo geral, e de um ponto de vista de estabilidade matemática, pode afirmar-se que, globalmente, a partição *XA* é aquela que evidencia melhor desempenho, o que é reforçado tendo em conta as conclusões obtidas pelo diagnóstico do *MRLM*, efectuado na secção precedente, segundo o qual para caso *SO_RF* os regressores apresentavam fortes indícios de multicolinearidade.

3.4.2.2.4. Validação da identificação

A validação da metodologia de identificação foi realizada por comparação das estimativas da concentração de glucose, oxigénio e dióxido de carbono dissolvidos, calculadas através da Equação 3.75, da Equação 3.76 e da Equação 3.77, respectivamente, com os valores “*pseudo* experimentais” dessas variáveis. Estes últimos foram determinados recorrendo ao modelo dinâmico do processo, implementado em MATLAB 7.1, para todas as variáveis medidas (*X*, *S*, *A*, *O*, *C*, *F_{e,s}*, *W*, *OTR*, *CTR*). Exemplificando para o caso da biomassa, a *pseudo* medida, *X_m*, é obtida pela equação:

$$X_m(t) \equiv X(t) + e(t) = X(t)[1 + w(t)] \quad \text{Equação 3.83}$$

⁸ A matriz na forma escalonada é obtida dividindo cada linha pelo elemento de maior módulo dessa mesma linha, em valor absoluto.

em que $e(t)$ representa o erro da medição calculado através de ruído branco, $w(t)$, de distribuição normal, não correlacionado de média nula e variância unitária. Considerou-se para todas as variáveis “medidas” um desvio padrão igual ao obtido em experiências laboratoriais (0,0350 g/kg, 0,0250 g/kg, 0,0750 g/kg, 0,0316 g/kg, 0,00298 g/kg, 0,000545 kg, 0,0995 g/kg.h e 0,0125 g/kg.h para as variáveis X , S , A , O , C , W , OTR , CTR , respectivamente).

O processo de validação da identificação foi realizado para a partição XA . Esta partição foi escolhida, em detrimento da partição SO , uma vez que é a que apresenta globalmente melhores resultados quer no estudo da regressão linear múltipla, quer na avaliação da estabilidade matemática do problema da recuperação dos k 's através dos valores dos θ 's estimados. Com efeito, quando estes últimos são calculados com incertezas, a partição seleccionada deve ser aquela que permite obter funções não lineares mais simples (Tabela 3.9) para o cálculo dos k 's de modo a limitar a propagação de erros.

O estudo apresentado de seguida refere-se a simulações efectuadas com o caudal de alimentação óptimo obtido na optimização segundo o critério D , para os regimes RF e R. Foram analisados dois casos, um em que todas as variáveis medidas foram afectadas de ruído (Equação 3.83) e outro onde as variáveis medidas, excepto X e A , foram também afectadas de ruído. Esta última situação foi motivada pelo facto do $MRLM$ utilizado ser do tipo I, no qual se considera que as variáveis independentes (X e A) são medidas isentas de erro, associado ainda ao facto de experimentalmente, essas mesmas variáveis serem medidas em diferido, com tempos de amostragem de hora em hora. Na Tabela 3.18 e na Tabela 3.19 apresentam-se os valores dos parâmetros k 's calculados para os dois casos estudados, bem como os respectivos erros relativos percentuais. Como se pode inferir da análise dessas tabelas, apesar do ruído nas variáveis medidas obtiveram-se erros relativos baixos ($\leq 9\%$), os quais são significativamente melhorados quando os regressores (X e A) não são afectados de ruído. Este facto permite admitir que a identificação numérica dos coeficientes de rendimento através de dados experimentais será possível.

Tabela 3.18 Valores calculados dos coeficientes de rendimento para a partição XA , regime RF, critério D .

Parâmetro	valor “real”	ruído em todas as variáveis		sem ruído nos regressores	
		valor estimado	erro relativo (%)	valor estimado	erro relativo (%)
k_1	1,649	1,658	0,55	1,649	0,00
k_2	29,8	29,8	0,0	29,7	0,34
k_3	17,27	17,76	2,8	17,24	0,17
k_5	0,3036	0,3172	4,5	0,3042	0,20
k_8	0,2761	0,2989	8,3	0,2758	0,11
k_9	52,48	52,22	0,50	52,42	0,11

Tabela 3.19 Valores calculados dos coeficientes de rendimento para a partição XA, regime R, critério D.

Parâmetro	valor "real"	ruído em todas as variáveis		sem ruído nos regressores	
		valor estimado	erro relativo (%)	valor estimado	erro relativo (%)
k_1	1,649	1,649	0,0	1,649	0,0
k_4	31,91	29,07	8,9	32,24	1,0
k_5	0,3036	0,3004	1,1	0,3038	0,066
k_7	17,11	15,87	7,2	17,45	2,0
k_8	0,2761	0,2733	1,0	0,2761	0,0
k_{10}	17,24	16,04	7,0	17,40	0,93

Na Figura 3.16 ilustra-se a validação da identificação para o caudal de alimentação ótimo obtido para a partição XA e regime R, com o critério D, onde se comparam os valores "pseudo experimentais" com os valores estimados para a glucose, o oxigénio dissolvido e o dióxido de carbono dissolvido.

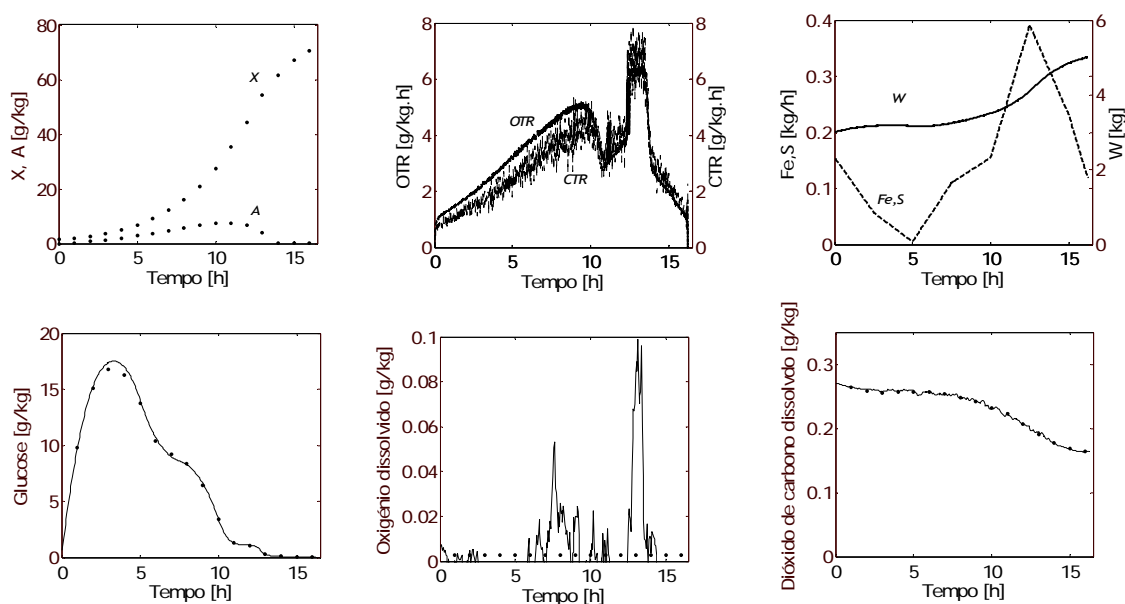


Figura 3.16 Experiência de identificação com caudal de alimentação obtido, para a partição XA_R e optimização do critério D (pontos: valores "pseudo experimentais", linha: valores estimados).

Na Figura 3.17 representam-se ainda as funções de sensibilidade dos y 's (Equação 3.72 à Equação 3.74) em relação aos k 's, para a mesma simulação da figura anterior. É de referir que foram utilizadas funções de sensibilidade relativas:

$$\partial_{y_k}^r = \frac{k}{y} \frac{\partial y}{\partial k}$$

Equação 3.84

Estas funções permitem determinar, teoricamente, se os parâmetros podem ou não ser estimados através dos dados experimentais. Se a sensibilidade de y em relação a k é muito pequena ou nula, então o período de amostragem pode não ter sido bem escolhido ou então a variável de controlo pode não excitar suficientemente a sensibilidade dos parâmetros, reduzindo a precisão dos parâmetros estimados.

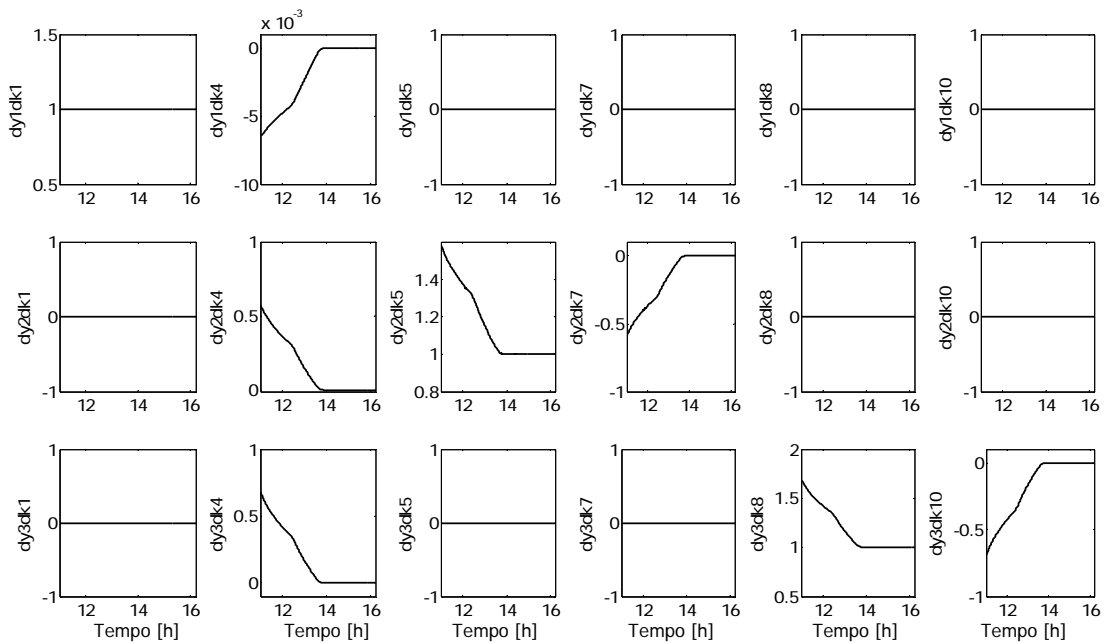


Figura 3.17 Funções de sensibilidade relativas dos y 's em relação aos k 's para XA_R .

Como se pode observar os parâmetros k_4 , k_7 e k_{10} apresentam funções de sensibilidade de amplitude reduzida, sendo nula a partir das 14 horas, isto é, em cerca de metade da duração desta experiência. Este facto vem realçar os resultados anteriormente obtidos para os erros relativos destes parâmetros, que são os mais elevados. Com base nestes resultados e para este regime seria desejável aumentar a frequência de amostragem até às 14 horas, de modo a otimizar a precisão da estimativa destes parâmetros.

3.4.3. Identificação dos parâmetros do modelo: implementação prática

Na secção anterior procedeu-se ao estudo, em termos de simulações numéricas, da planificação experimental para a determinação dos coeficientes de rendimento do sistema biológico em estudo. Nesta secção apresentam-se os resultados da validação experimental da metodologia do desenho óptimo. Os resultados apresentados são referentes a duas fermentações, Exp_2 e Exp_3, sendo que

a primeira consistiu na implementação de um caudal de alimentação não otimizado e para a qual só se obteve regime respirativo. Na Exp_3 observou-se a existência dos dois regimes, respirativo e respiro-fermentativo e, corresponde à implementação prática do caudal de alimentação otimizado de acordo com o critério D , para o regime respirativo, como apresentado na secção anterior (Figura 3.8, coluna da esquerda, primeiro gráfico). O procedimento experimental encontra-se descrito em pormenor no Capítulo 2.

As duas fermentações realçam a importância do desenho óptimo de experiências como ferramenta útil para maximizar o conteúdo de informação dos dados experimentais e permitiram o cálculo dos parâmetros θ , parâmetros calculados por regressão, de acordo com a metodologia anteriormente apresentada. A precisão associada ao cálculo dos θ 's foi avaliada através de índices relativos ao modelo de regressão linear múltipla e da matriz de Fisher e pelo desenho das elipses dos desvios. Com os valores de θ foi possível determinar, posteriormente, o valor dos coeficientes de rendimento, como era objectivo deste trabalho.

Apresentam-se de seguida os resultados obtidos e faz-se uma discussão dos mesmos.

A Exp_2 permitiu calcular apenas os coeficientes de rendimento associados ao regime respirativo (k_1, k_4, k_5, k_7, k_8 e k_{10}), enquanto que a Exp_3 permitiu calcular para além desses coeficientes os associados ao regime respiro-fermentativo (k_2, k_3 e k_9), embora para esta última experiência as condições não tenham sido optimizadas para este regime. A análise dos pressupostos do *MRLM* permitiu concluir que para ambas as experiências efectuadas e para o regime respirativo esses mesmos pressupostos eram, de um modo geral, verificados, sendo os resultados para a Exp_3 de melhor qualidade que os da Exp_2. Quanto ao regime respiro-fermentativo, constatou-se que, os referidos pressupostos não eram verificados.

Na Tabela 3.20 apresentam-se os valores de θ calculados com base nos dados das experiências Exp_2 e Exp_3, bem como as incertezas que lhes estão associadas. Os resultados obtidos permitem verificar que, para o regime respirativo os valores de θ calculados a partir dos dados da Exp_2 apresentam uma incerteza superior aos calculados com a Exp_3. Relativamente à Exp_3 os valores de θ calculados para o regime respiro-fermentativo apresentam uma maior incerteza do que os obtidos para o regime respirativo. Nessa mesma tabela indicam-se ainda os valores dos coeficientes de rendimento e das incertezas que lhes estão associadas. Estas últimas foram calculadas recorrendo à fórmula de propagação de erros. Mais uma vez verifica-se que para o regime respirativo estas incertezas são maiores para a Exp_2 do que para a Exp_3. Para esta última constata-se também que as incertezas associadas aos k 's do regime respiro-fermentativo são superiores às obtidas para os k 's do regime respirativo.

A qualidade dos valores dos θ 's e, indirectamente dos valores dos k 's, foi ainda avaliada recorrendo a parâmetros associados à Matriz de Fisher bem como à construção das elipses dos desvios, sendo os resultados apresentados na Tabela 3.21 e na Figura 3.18, respectivamente.

Tabela 3.20 Valores de θ e de k e, respectivas incertezas, calculados por *MRLM* para as experiências Exp_2 e Exp_3.

Exp_2				Exp_3							
regime respirativo				regime respirativo				regime respiro-fermentativo			
$\theta \pm \delta\theta^{(i)}$		$k \pm \delta k^{(ii)}$		$\theta \pm \delta\theta^{(i)}$		$k \pm \delta k^{(ii)}$		$\theta \pm \delta\theta^{(i)}$		$k \pm \delta k^{(ii)}$	
θ_1	3,127±0,013	k_1	3,13±0,013	θ_1	3,184±0,007	k_1	3,18±0,007	θ_1	1,669±0,21	k_2	30,4±6,6
θ_2	0,9169±0,32	k_4	3,41±1,2	θ_2	0,5843±0,021	k_4	5,44±0,20	θ_2	0,9271±0,28	k_3	31,0±4,8
θ_3	2,356±0,018	k_5	2,36±0,018	θ_3	1,211±0,011	k_5	1,21±0,011	θ_3	1,007±0,13	k_9	44,3±6,9
θ_4	-1,561±0,44	k_7	7,68±2,4	θ_4	-1,754±0,034	k_7	10,8±0,39	θ_4	1,145±0,18		
θ_5	-2,752±0,022	k_8	2,75±0,022	θ_5	-2,593±0,008	k_8	2,59±0,008	θ_5	-0,1651±0,045		
θ_6	2,153±0,52	k_{10}	10,1±3,1	θ_6	0,4081±0,026	k_{10}	4,82±0,16	θ_6	-1,426±0,06		

(i) Incerteza associada ao valor de θ obtida pela regressão: $\delta\theta = t_{GL}\left(\frac{\alpha}{2}\right) \times s_{\theta}$, sendo s_{θ} o desvio padrão.

(ii) Incerteza associada ao valor de k obtida pela equação de propagação de erros: $\delta k = \sqrt{\left(\frac{\partial k}{\partial \theta}\right)^2} \delta\theta^2$

Tabela 3.21 Parâmetros associados à Matriz de Fisher usados no estudo da qualidade dos valores de θ .

	Exp_2		Exp_3	
	regime respirativo		regime respirativo	regime respiro-fermentativo
\det_{MIF}	4,322		3083	13,50
Λ_{MIF}	658,8		20,62	218,7
$s_{\text{aprox}}(\theta_i)$	0,148		0,0927	1,21
$s_{\text{aprox}}(\theta_j)$	3,51		0,278	1,61
c_{ij}	-0,37		-0,71	-0,99
IC _{aprox} (θ_1)	[2,84; 3,42]		[3,00; 3,37]	[-0,734; 4,07]
IC _{aprox} (θ_2)	[-5,97; 7,80]		[0,0323; 1,14]	[-2,27; 4,12]
IC _{aprox} (θ_3)	[2,07; 2,65]		[1,03; 1,39]	[-1,40; 3,41]
IC _{aprox} (θ_4)	[-8,45; 5,33]		[-2,31; -1,20]	[-2,05; 4,34]
IC _{aprox} (θ_5)	[2,46; 3,04]		[-2,78; -2,41]	[-2,57; 2,24]
IC _{aprox} (θ_6)	[-4,73; 9,04]		[-0,144; 0,960]	[-4,62; 1,77]

$\theta_j = \theta_1 = \theta_3 = \theta_5$; $\theta_j = \theta_2 = \theta_4 = \theta_6$; s_{aprox} - desvio padrão aproximado; c_{ij} - coeficiente de correlação aproximado; IC - intervalo de confiança aproximado.

A Matriz de Informação de Fisher é o elemento central da teoria do desenho ótimo de experiências. De facto, já foi referido neste capítulo que a *MIF* define a forma das elipses dos desvios na vizinhança dos valores verdadeiros dos parâmetros do processo e que as regiões de

confiança diferem das elipses dos desvios apenas numa constante, tendo as mesmas propriedades geométricas. Deste modo, pode-se utilizar uma função escalar da *MIF*, como o critério *D* e *E* modificado, para minimizar as incertezas associadas aos parâmetros estimados.

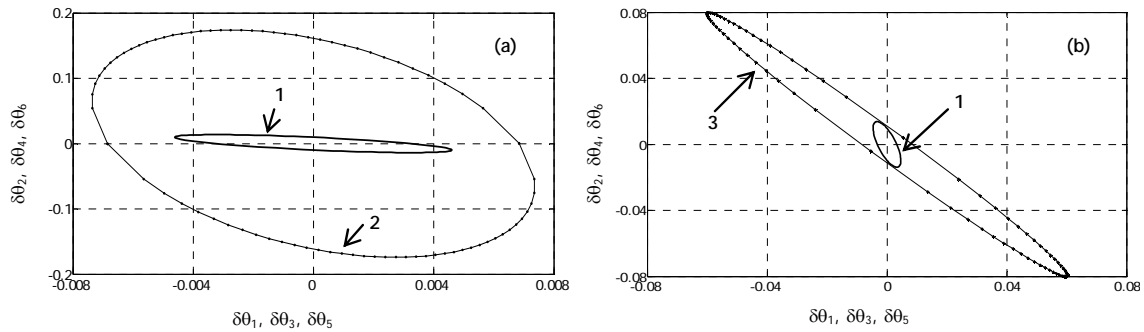


Figura 3.18 Elipses dos desvios para as experiências Exp_2 e Exp_3: 1- regime respirativo da Exp_3; 2- regime respirativo da Exp_2 e 3- regime respiro-fermentativo da Exp_3.

Os resultados apresentados na Tabela 3.21 e na Figura 3.18 (a) mostram que a Exp_3, sendo a implementação prática de um perfil de alimentação otimizado para o regime respirativo, apresenta uma melhoria face aos obtidos com a Exp_2. Com efeito, o desvio padrão associado aos valores de θ calculados a partir da Exp_3 são bastante inferiores aos obtidos para a Exp_2, observando-se ainda que, para a Exp_3, a elipse dos desvios apresenta uma área bastante inferior à da Exp_2, o que está de acordo com os valores obtidos para o determinante e o número de condicionamento da *MIF*, traduzindo uma maior precisão dos parâmetros calculados com base na Exp_3. Contudo, embora os coeficientes de correlação aproximados obtidos para o regime respirativo, para ambas as experiências, sejam pequenos, a região de confiança conjunta será menos sobreestimada no caso da Exp_2, que apresenta um valor de c_{ij} inferior em valor absoluto.

Como se referiu anteriormente a Exp_3 foi utilizada para calcular os parâmetros para o regime respirativo e para o regime respiro-fermentativo, embora não tenha sido otimizada para este último. Como se pode constatar a partir da análise da tabela anterior os parâmetros obtidos para este último regime apresentam uma menor precisão que é traduzida pelos elevados valores do número de condicionamento da *MIF*, do desvio padrão aproximado e do coeficiente de correlação, e pelo pequeno valor do determinante da *MIF*. Este facto é ainda realçado pela Figura 3.18 (b), onde se observa uma elipse mais alongada e com uma área superior à obtida para o regime respirativo.

O valor obtido para Λ_{MIF} para o regime respirativo da Exp_3 ($\Lambda_{MIF}=20,62$) é bastante satisfatório. Já foi referido noutros trabalhos (Versyck, 2000) a obtenção do valor óptimo para este critério ($\Lambda_{MIF}=1$), no entanto, deve-se referir que este valor apenas foi conseguido em simulações numéricas e utilizando um modelo mais simples constituído apenas por duas equações e envolvendo dois

parâmetros estimados. Na respectiva implementação prática a autora apresenta também outros valores de Λ_{MF} da mesma ordem de grandeza do obtido neste trabalho.

Deve realçar-se ainda que para o regime respirativo se observou uma melhoria significativa da qualidade dos valores estimados da Exp_2 (não otimizada) para a Exp_3 (otimizada), demonstrando assim as potencialidades da metodologia do desenho óptimo de experiências apresentada.

Deste modo, pode concluir-se que os coeficientes de rendimento k_1 , k_5 e k_8 são aqueles que apresentam uma maior precisão, enquanto que as estimativas dos coeficientes k_2 , k_3 , k_4 , k_7 , k_9 e k_{10} têm associadas uma incerteza maior. Deve ainda referir-se que no estudo efectuado na secção 3.4.2.1 concluiu-se que dos coeficientes de rendimento, o k_1 , deveria ser o estimado com maior precisão, o que realmente foi conseguido.

Estes resultados podem ser ainda sustentados por uma análise de sensibilidade. Na Figura 3.19 encontram-se representadas as funções de sensibilidade relativas dos y 's (Equação 3.72 à Equação 3.74) em relação aos k 's, para os dados da Exp_3. Apresentam-se os resultados desta experiência por ser a otimizada. As funções de sensibilidade foram calculadas de acordo com a Equação 3.84.

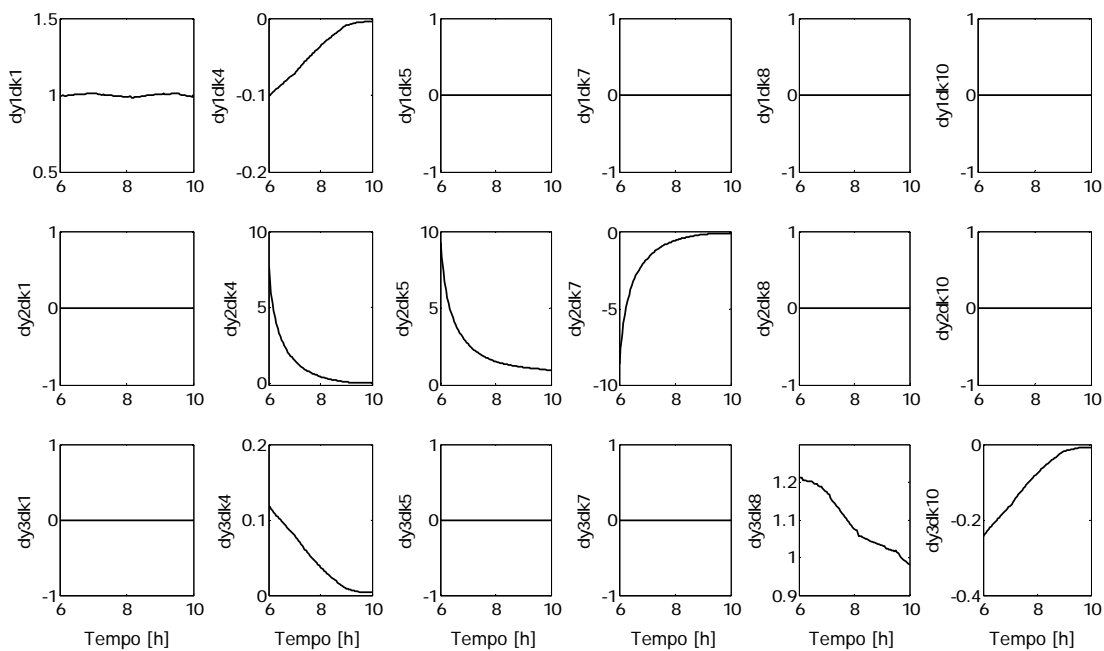


Figura 3.19 Funções de sensibilidade relativas dos y 's em relação aos k 's, para a Exp_3, regime respirativo.

A análise de sensibilidade é também uma ferramenta útil, permitindo avaliar se os parâmetros teoricamente identificáveis podem ser ou não correctamente calculados com base em dados experimentais. Convém referir que, para um parâmetro ser correctamente identificável os dados experimentais devem ser altamente sensíveis a variações desse parâmetro. Uma reduzida

sensibilidade indica que será difícil obter um valor único para esse parâmetro. Mais ainda, a semelhança entre os perfis de sensibilidade das variáveis medidas relativamente aos diferentes parâmetros estudados indica que as suas estimativas podem estar correlacionadas (Pratt *et al.*, 2004).

Da análise da figura anterior constata-se que em algumas situações as funções de sensibilidade são nulas. Esta situação deve-se ao modo como os k 's são calculados (equações da Tabela 3.9), uma vez que no cálculo de cada um desses parâmetros não são utilizados todos os θ 's. Verifica-se ainda que as sensibilidades dos sinais medidos relativamente a variações dos valores de k_1 , k_4 , k_5 , k_7 e k_8 são em geral elevadas (superiores a 1), indicando que estes parâmetros poderão ser estimados com uma pequena incerteza. No entanto, uma vez que as formas dos perfis das sensibilidades obtidos para os valores de k_4 e k_7 (para y_2) e k_4 e k_{10} (para y_3) têm magnitudes semelhantes, sendo proporcionais, sugere que estes parâmetros podem estar correlacionados, pelo que a sua identificação prática, usando aquelas variáveis medidas, pode ser difícil. Por outro lado, a sensibilidade para k_{10} é baixa (inferior a 1) indicando que será difícil obter uma estimativa correcta para este coeficiente. Assim, a análise de sensibilidade realizada permite verificar que a Exp_3 permite determinar com maior precisão os coeficientes k_1 , k_5 e k_8 , sendo que os coeficientes k_4 , k_7 e k_{10} têm associada alguma incerteza. Para estes últimos coeficientes será de esperar que uma experiência em que ocorra apenas consumo de acetato melhore a estimativa destes parâmetros por permitir o desacoplamento do seu cálculo da determinação de k_1 , k_5 e k_8 , ou seja da reacção relacionada com μ_1 e μ_3 .

Na Figura 3.20 apresenta-se a validação da identificação para a Exp_3 onde se comparam os valores experimentais com os valores estimados para a glucose, o oxigénio dissolvido e o dióxido de carbono dissolvido (Equação 3.75 à Equação 3.77). Apesar dos resultados obtidos serem satisfatórios, é evidente que neste caso de estudo é necessário uma nova iteração da metodologia do desenho óptimo de experiências, uma vez que os erros associados a alguns dos valores de θ são elevados. Será também necessário a realização de mais experiências de modo a conseguir-se uma convergência para os verdadeiros valores dos parâmetros do processo, em especial para o regime respiro-fermentativo, para o qual não foi implementada nenhuma experiência otimizada. Assim, o primeiro passo seria efectuar uma nova optimização em que os valores dos coeficientes de rendimento estimados pela Exp_3 serão os novos valores nominais.

A convergência deste ciclo iterativo do desenho óptimo de experiências para os verdadeiros valores dos parâmetros não pode ser, de um modo geral, comprovada. No entanto, pode ser visualizada de iteração para iteração, uma melhoria das regiões de confiança em termos de área e forma, indicando um progresso na direcção correcta, como se observou com os resultados obtidos da Exp_2 e Exp_3.

Assim sendo, a contribuição deste trabalho consistiu numa maior consolidação da metodologia proposta por Chen (1992) e sua aplicação à fermentação semi-contínua de *E. coli*. Foi possível demonstrar que a planificação experimental é uma ferramenta adequada, partindo inicialmente de

simulações numéricas, para se obterem experiências suficientemente informativas que traduzem na prática o cálculo de valores de parâmetros mais precisos. Por outro lado, o teste de identificabilidade teórica utilizado por esta metodologia é muito simples, não sendo necessário recorrer a métodos mais complexos como a aproximação por séries de *Taylor*, o que é uma vantagem principalmente, para a identificação de modelos, como o do caso em estudo, constituídos por várias equações diferenciais e parâmetros.

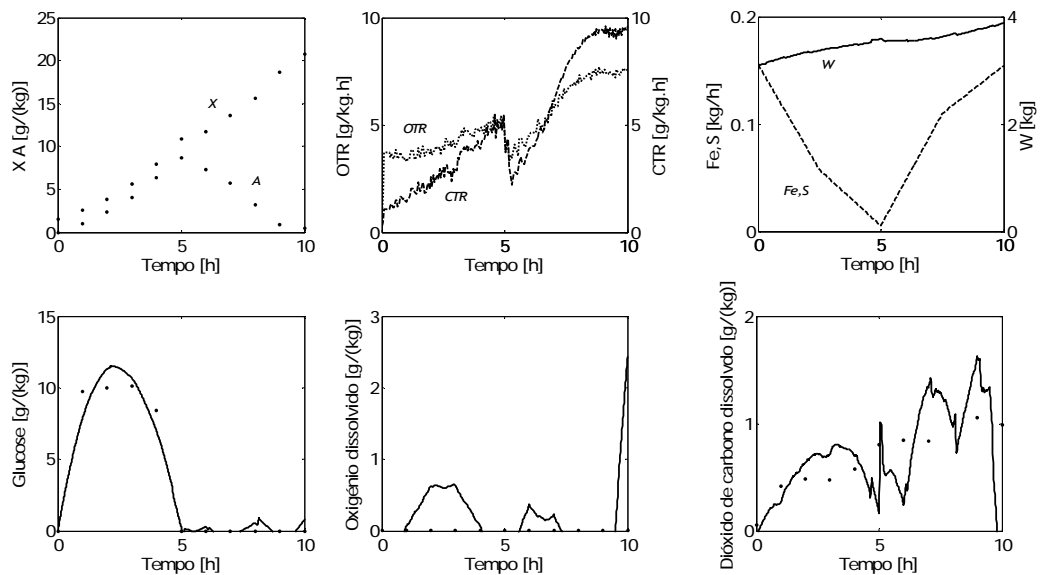


Figura 3.20 Validação experimental para a Exp_3 (pontos: valores experimentais, linha: valores estimados).

A metodologia apresentada permitiu a determinação dos coeficientes de rendimento sem o conhecimento preciso do modelo cinético. Seria interessante prosseguir com este estudo e alargar o desenho óptimo de experiências para otimizar a estrutura cinética e determinar os respectivos parâmetros.

A aplicação do *MRLM* no cálculo dos valores de θ e, indirectamente dos valores de k , afigura-se como uma metodologia simples, cuja análise é facilmente verificada por depender do grau de independência dos regressores, embora a sua aplicação se revista de alguns pressupostos que deverão ser respeitados, como se demonstrou neste estudo. Verificou-se que apesar de teoricamente todas as partições permitirem a identificação dos k , na prática isso pode não ser verdadeiro.

Com a aplicação desta metodologia na identificação dos coeficientes de rendimento reduziu-se o número de parâmetros identificados, sendo de dois parâmetros por cada equação de regressão. Este facto é relevante pois simplifica o processo de optimização.

Alternativamente, poderá ser usado um modelo de regressão não linear que permita calcular de forma directa os valores dos k 's. Neste caso, as funções de sensibilidade utilizadas no cálculo da *MIF*

estariam directamente relacionados com os k 's, o que poderá traduzir-se em experiências com maior sensibilidade relativamente a esses parâmetros. Contudo, este método apresenta-se mais difícil de analisar envolvendo directamente os coeficientes de rendimento que são os parâmetros a estimar. Por outro lado, o número de parâmetros directamente identificados seria maior, de seis para o regime respirativo ou de sete para o regime respiro-fermentativo, em oposição aos dois parâmetros por equação de regressão no *MRLM*. Este facto traduzir-se-ia numa *MIF* de dimensão 6×6 ou 7×7 , para o regime respirativo ou respiro-fermentativo, respectivamente, enquanto que a *MIF* determinada no caso do *MRLM* é de dimensão 2×2 . Quando o número de parâmetros estimados é maior que dois, a dificuldade associada à optimização de uma função escalar da *MIF* aumenta significativamente. Assim, seria necessário ponderar se o aumento da sensibilidade aos k 's compensaria o aumento da complexidade do problema de optimização que poderia levar a maiores dificuldades de convergência.

3.5. SÍNTESE

A teoria do desenho óptimo de experiências permite a optimização de variáveis de entrada do modelo de modo a garantir um nível de precisão elevado para todos os parâmetros estimados. Esta metodologia tem sido aplicada com sucesso na identificação dos parâmetros da cinética de *Monod* (Baltes *et al.*, 1994; Versyck *et al.*, 1997; Versyck *et al.*, 1998) e em várias áreas como a microbiologia alimentar (Grijpspeerdt e Vanrolleghem, 1999; Versyck *et al.*, 1999; Bernaerts *et al.*, 2000), a biologia de sistemas (Kutalik *et al.*, 2004), a farmacologia (Jacquez, 1998) e em processos de tratamentos de águas residuais (Noykova *et al.*, 2002).

Neste capítulo foram apresentados algoritmos para a identificação de coeficientes de rendimento em bioprocessos através de medições das variáveis de estado. Foram propostas estratégias de planificação experimental que visam a optimização da riqueza informativa da experiência, quantificada por índices relativos à matriz de informação de Fisher. O objectivo do desenho de experiências foi endereçado em termos da programação do perfil do caudal de alimentação.

Embora o modelo utilizado para descrever a fermentação semi-contínua de *E. coli* não esteja ainda optimizado em termos cinéticos e de algumas dificuldades encontradas na implementação prática dos resultados obtidos por simulação para o desenho óptimo de experiências, a qualidade da estimativa dos parâmetros, especialmente os do regime respirativo, é promissora. A incerteza das estimativas foi verificada através de índices relacionados com o modelo de regressão linear múltipla, índices relativos à matriz de Fisher e pelo desenho das correspondentes elipses dos desvios.

Os valores dos coeficientes de rendimento determinados através da Exp_3 são diferentes dos valores nominais utilizados durante a planificação experimental. Os desvios associados a cada coeficiente mostram que ainda não foram encontrados os melhores valores. Assim, é esperado que se consiga diminuir as incertezas das estimativas se a metodologia adoptada continuar a ser utilizada, num processo iterativo. Sendo assim, pode-se concluir que são necessárias mais experiências de modo a convergir para os valores verdadeiros dos coeficientes de rendimento do sistema biológico em estudo. Estes coeficientes constituem uma peça fundamental para a implementação de estratégias de monitorização avançada como se apresentará no próximo capítulo, no qual, estes parâmetros são utilizados para o desenho dos sensores por programação.

3.6. BIBLIOGRAFIA

Aiba, S. e Furuse, H., (1990). Some comments on respiratory quotient (RQ) determination from the analysis of exit gas from a fermenter. *Biotechnology and Bioengineering*, 36, 534-538.

Akesson, M., Karlsson, E.N., Hagander, P., Axelsson, J.P. e Tocaj, A., (1999). On-line detection of acetate formation in *Escherichia coli* cultures using dissolved oxygen responses to feed transients. *Biotechnology and Bioengineering*, 64 (5), 590-598.

Alagappan, G. e Cowan, R.M., (2001). Biokinetic models for representing the complete inhibition of microbial activity at high substrate concentrations. *Biotechnology and Bioengineering*, 75 (4), 393-405.

Arrua, L.A., McCoy, B.J. e Smith, J.M., (1990). Gas-liquid mass transfer in stirred tanks. *AIChE Journal*, 36 (11), 1768-1772.

Asprey, S.P. e Macchietto, S., (2000). Statistical tools for optimal dynamic model building. *Computers & Chemical Engineering*, 24 (2-7), 1261-1267.

Asprey, S.P. e Macchietto, S., (2002). Designing robust optimal dynamic experiments. *Journal of Process Control*, 12 (4), 545-556.

Bailey, J.E., (1998). Mathematical modeling and analysis in biochemical engineering: past accomplishments and future opportunities. *Biotechnology Progress*, 14 (1), 8-20.

Bailey, J.E. e Ollis, D.F., (1986). *Biochemical Engineering Fundamentals*, McGraw-Hill Book Co., Singapore.

Baltes, M., Schneider, R., Sturm, C. e Reuss, M., (1994). Optimal experimental design for parameter estimation in unstructured growth models. *Biotechnology Progress*, 10 (5), 480-488.

Banga, J.R., Versyck, K.J. e Van Impe, J.F., (2002). Computation of optimal identification experiments for nonlinear dynamic process models: a stochastic global optimization approach. *Ind.Eng.Chem.Res.*, 41, 2425-2430.

Bastin, G. e Dochain, D., (1990). *On-line estimation and adaptive control of bioreactors*, Elsevier Science Publishers, Amsterdam.

Berkholz, R. e Guthke, R., (2001). Model Based Sequential Experimental Design for Bioprocess Optimisation - An Overview. Em: Hofman, M. e Thonart, P. (eds), *Fokus Biotechnology. Engineering and Manufacturing for Biotechnology*. pp. 129-141, Kluwer Academic Publishers. Netherland.

Berkholz, R., Röhligh, D. e Guthke, R., (2000). Data and knowledge based experimental design for fermentation process optimization. *Enzyme and Microbial Technology*, 27, 784-788.

Bernaerts, K., Servaes, R.D., Kooyman, S., Versyck, K.J. e Van Impe, J.F., (2002). Optimal temperature input design for estimation of the Square Root model parameters: parameter accuracy and model validity restrictions. *International Journal of Food Microbiology*, 73 (2-3), 145-157.

Bernaerts, K., Versyck, K.J. e Van Impe, J.F., (2000). On the design of optimal dynamic experiments for parameter estimation of a Ratkowsky-type growth kinetics at suboptimal temperatures. *International Journal of Food Microbiology*, 54 (1-2), 27-38.

Bernard, O., Hadj-Sadok, Z., Dochain, D., Genovesi, A. e Steyer, J.P., (2001). Dynamical model development and parameter identification for an anaerobic wastewater treatment process. *Biotechnology and Bioengineering*, 75 (4), 424-438.

Bogaerts, P., (1999). *Contribution à la modélisation mathématique pour la simulation et l'observation d'états des bioprocédés*. Ph.D., Université Libre de Bruxelles (Belgium).

Bogaerts, P., Delcoux, J.L. e Hanus, R., (2003). Maximum likelihood estimation of pseudo-stoichiometry in macroscopic biological reaction schemes. *Chemical Engineering Science*, 58 (8), 1545-1563.

Chen, L., (1992). *Modelling, identifiability and control of complex biotechnological systems*. Tese de Doutorado, Universidade Católica de Louvain, Bélgica.

Chen, L. e Bastin, G., (1996). Structural identifiability of the yield coefficients in bioprocess models when the reaction rates are unknown. *Mathematical Biosciences*, 132 (1), 35-67.

Chouakri, N., Fonteix, C., Corriou, J.P. e Marc, I., (1994a). Parameter estimation of a Monod-type model. Part II: use of a neural network with associative memories. *Biotechnology Techniques*, 8 (10), 689-694.

Chouakri, N., Fonteix, C., Marc, I. e Corriou, J.P., (1994b). Parameter estimation of a Monod-type model. Part I: theoretical identifiability and sensitivity analysis. *Biotechnology Techniques*, 8 (10), 683-688.

Claes, J.E. e Van Impe, J.F., (2000). Combining yield coefficients and exit-gas analysis for monitoring of the baker's yeast fed-batch fermentation. *Bioprocess Engineering*, 22 (3), 195-200.

Claes, J.E. e Van Impe, J.F., (1999). On-line estimation of the specific growth rate based on viable biomass measurements: experimental validation. *Bioprocess Engineering*, 21 (5), 389-395.

Cockshott, A.R. e Bogle, I.D.L., (1999). Modelling the effects of glucose feeding on a recombinant *E. coli* fermentation. *Bioprocess Engineering*, 20, 83-90.

Degenring, D., Froemel, C., Dikta, G. e Takors, R., (2004). Sensitivity analysis for the reduction of complex metabolism models. *Journal of Process Control*, 14 (7), 729-745.

Ejiofor, A.O., Posten, C., Solomon, B.O. e Deckwer, W.-D., (1994a). A robust fed-batch feeding strategy for optimal parameter estimation for baker's yeast production. *Bioprocess Engineering*, 11, 135-144.

- Ejiofor, A.O., Solomon, B.O. e Posten, C., (1994b). Analysis of the respiro-fermentative growth of *Saccharomyces cerevisiae* on glucose in a fed-batch fermentation strategy for accurate parameter estimation. *Applied Microbiology and Biotechnology*, 41, 664-669.
- Espie, D. e Macchietto, S., (1989). The optimal design of dynamic experiments. *AIChE Journal*, 35, 223-229.
- Fernandes, E.M.G.P., (1998). *Computação Numérica*, Publicações da Universidade do Minho, Braga.
- Ferreira, E.C., (1995). *Identificação e controlo adaptativo de processos biotecnológicos*. Tese de Doutoramento, Universidade do Porto, Portugal.
- Galvanauskas, V., Simutis, R., Volk, N. e Lübert, A., (1998). Model based design of a biochemical cultivation process. *Bioprocess Engineering*, 18, 227-234.
- Godfrey, K.R. e DiStefano III, J.J., (1985). Identifiability of Model Parameters. *Identification and System Parameter Estimation 7th IFAC/IFORS Symposium*, 89-114.
- Goodwin, G.C., (1987). Identification: Experiment design. *Systems & Control Encyclopedia*, Pergamon Press, 4, 2257-2264.
- Grijspeerdt, K. e Vanrolleghem, P., (1999). Estimating the parameters of the Baranyi model for bacterial growth. *Food Microbiology*, 16 (6), 593-605.
- Haag, J.E., Vande Wouwer, A. e Remy, M., (2005). A general model of reaction kinetics in biological systems. *Bioprocess and Biosystems Engineering*, 27, 303-309.
- Heinzle, E. e Dunn, I.J., (1990). Methods and instruments in fermentation gas analysis. Em: Karl Schügerl (ed), *Biotechnology - Measuring, Modelling and Control*. pp. 27-72, VCH, Weinheim.
- Hidalgo, M.E. e Ayesa, E., (2001). Numerical and graphical description of the information matrix in calibration experiments for state-space models. *Water Research*, 35 (13), 3206-3214.
- Hodgson, B.J., Taylor, C.N., Ushio, M., Leigh, J.R., Kalganova, T. e Baganz, F., (2004). Intelligent modelling of bioprocesses: a comparison of structured and unstructured approaches. *Bioprocess and Biosystems Engineering*, 26 (6), 353-359.
- Holmberg, A., (1982). On the Practical Identifiability of Microbial-Growth Models Incorporating Michaelis-Menten Type Nonlinearities. *Mathematical Biosciences*, 62 (1), 23-43.
- Jacquez, J.A., (1998). Design of experiments. *Journal of the Franklin Institute-Engineering and Applied Mathematics*, 335B (2), 259-279.
- Keesman, K.J. e Stigter, J.D., (2002). Optimal parametric sensitivity control for the estimation of kinetic parameters in bioreactors. *Mathematical Biosciences*, 179, 95-111.
- Kutalik, Z., Cho, K.H. e Wolkenhauer, O., (2004). Optimal sampling time selection for parameter estimation in dynamic pathway modeling. *Biosystems*, 75 (1-3), 43-55.
- Lee, J. e Ramirez, W.F., (1992). Mathematical modelling of induced foreign production by recombinant bacteria. *Biotechnology and Bioengineering*, 39, 635-646.
- Lei, F. e Jorgensen, S.B., (2001). Estimation of kinetic parameters in a structured yeast model using regularisation. *Journal of Biotechnology*, 88 (3), 223-237.

Lin, H.Y., Mathisizik, B., Xu, B., Enfors, S.O. e Neubauer, P., (2001). Determination of the maximum specific uptake capacities for glucose and oxygen in glucose-limited fed-batch cultivations of *Escherichia coli*. *Biotechnology and Bioengineering*, 73 (5), 347-357.

Maroco, J., (2003). *Análise Estatística com Utilização do SPSS*, Edições Sílabo, Lda., Lisboa.

Mehra, R.K., (1974). Optimal input signal for parameter estimation in dynamic systems - survey and new results. *IEEE Transactions on Automatic Control*, 19 (6), 753-768.

Merkel, W., Schwarz, A., Fritz, S., Reuss, M. e Krauth, K., (1996). New strategies for estimating kinetic parameters in anaerobic wastewater treatment plants. *Water Science and Technology*, 34 (5-6), 393-401.

Mhaskar, P., Hjortso, M.A. e Henson, M.A., (2002). Cell population modeling and parameter estimation for continuous cultures of *Saccharomyces cerevisiae*. *Biotechnology Progress*, 18 (5), 1010-1026.

Monod, J., (1942). *Recherches sur la Croissance des Cultures Bactériennes*, Hermann, Paris.

Muller, T.G., Noykova, N., Gyllenberg, M. e Timmer, J., (2002). Parameter identification in dynamical models of anaerobic waste water treatment. *Mathematical Biosciences*, 177, 147-160.

Munack, A., (1989). Optimal feeding strategy for identification of Monod-type models by fed-batch experiments. Em: Fish, N.M., Fox, R.I. e Thornhill, N.F. (eds), *Computer Applications in Fermentation Technology: Modelling and Control of Biotechnological Processes*. pp. 195-204, Elsevier Applied Science, Londres.

Munack, A., (1984). Optimal feeding strategy for identification of Monod-type models by fed-batch experiments. *4th.Int.Congr.on Computer Application on Fermentation Techenology*, 195-204.

Munack, A., (1991). Optimization and sampling. Em: Rehm, H.J., Reed, G., Pühler, A. e Stadler, P. (eds), *Measurements in Bioreactor Systems*. pp. 252-264, VCH, Weinheim.

Na, J.-G., Chang, Y.K., Chung, B.H. e Lim, H.C., (2002). Adaptive optimization of fed-batch culture of yeast by using genetic algorithms. *Bioprocess and Biosystems Engineering*, 24, 299-308.

Nougués, J.M., Grau, M.D. e Puigjaner, L., (2002). Parameter estimation with genetic algorithm in control of fed-batch reactors. *Chemical Engineering and Processing*, 41, 303-309.

Noykova, N., Müller, T.G., Gyllenberg, M. e Timmer, J., (2002). Quantitative analyses of anaerobic wastewater treatment processes: identifiability and parameter estimation. *Biotechnology and Bioengineering*, 78 (1), 91-103.

Onken, U. e Liefke, E., (1989). Effect of total and partial pressure (oxygen and carbon dioxide) on aerobic microbial processes. *Advances in Biochemical Engineering / Biotechnology*, 40, 135-169.

Patnaik, P.R., (2002). Comparative evaluation of batch and fed-batch bioreactors for GAPDH production by recombinant *Escherichia coli* with distributed plasmid copy number. *Chemical Engineering Journal*, 87 (3), 357-366.

Pauw, D.J.W. e Vanrolleghem, P.A., (2006). Designing and performing experiments for model calibration using an automated iterative procedure. *Water Science and Technology*, 53 (1), 117-127.

Pestana, M.H. e Gageiro, J.N., (2005). *Análise de Dados para Ciências Sociais. A Complementaridade do SPSS*, Edições Sílabo, Lda., Lisboa.

- Pohlheim, H., (2004). Documentation 3.5a for Genetic and Evolutionary Algorithm Toolbox for use with Matlab: GEATbx version 3.5
- Posten, C. e Munack, A., (1990). On-line application of parameter estimation accuracy to biotechnical processes. *ACC*, 2181-2186.
- Pratt, S., Yuan, Z.G. e Keller, J., (2004). Modeling aerobic carbon oxidation and storage by integrating respirometric, titrimetric, and off-gas CO₂ measurements. *Biotechnology and Bioengineering*, 88 (2), 135-147.
- Quinn, G.P. e Keough, M.J., (2004). *Experimental Design and Data Analysis for Biologists*, Cambridge University Press, Cambridge.
- Rao, M.S. e Iyengar, S.S., (1984). Application of Statistical Techniques in Modelling of Complex Systems. Em: Iyengar, S.S. (ed), *Computer Modeling of complex Biological Systems*. pp. 29-53, CRC Press, Boca Raton, Florida (USA).
- Renard, F., Wouwer, A.V., Hulhoven, X. e Bogaerts, Ph., (2004). Modelling and Optimal Experiment Design for Cultures of *S. cerevisiae* (CAB9 - 9th International Symposium Computer Applications in Biotechnology (Nancy, France))
- Renko, E.K. e Sirvio, H., (1997). Optimal sampling design for a sludge blanket interface settling model. *Water Research*, 31 (5), 1148-1154.
- Rocha, C. e Ferreira, E.C., (1996). Design of optimal experiments for identification of yield coefficients in a Baker's Yeast Model (1st European Symposium on Biochemical Engineering Science (Dublin, Irlanda)), 99-100.
- Rocha, I., (2003). *Model-based strategies for computer-aided operation of recombinant E. coli fermentation*. Tese de Doutoramento, Universidade do Minho, Portugal.
- Rocha, I. e Ferreira, E.C., (2004). Yield and kinetic parameters estimation for model reduction in a recombinant *E. coli* fermentation. *ESCAPE-14: European Symposium on Computer Aided Process Engineering (accepted)*
- Sarkar, D. e Modak, J.M., (2005). Pareto-optimal solutions for multi-objective optimization of fed-batch bioreactors using nondominated sorting genetic algorithm. *Chemical Engineering Science*, 60 (2), 481-492.
- Savin, N.E. e White, K.J., (1977). The Durbin-Watson test for serial correlation with extreme sample sizes or array regressors. *Econometrica*, 15, 1989-1996.
- Schneider, R. e Munack, A., (1995). Improvements in the on-line parameter identification of bioprocesses. *6th Int.Conf.Computer Applications on Biotechnology, Garmisch-Patenkirchen, Germany*, 177-182.
- Schumpe, A. e Quicker, G., (1982). Gas solubilities in microbial culture media. *Advances in Biochemical Engineering / Biotechnology*, 24, 1-37.
- Smets, I., Bernaerts, K., Sun, J., Marchal, K., Vanderleyden, J. e Van Impe, J.F., (2002a). Sensitivity function-based model reduction. A bacterial gene expression case study. *Biotechnology and Bioengineering*, 80 (2), 195-200.
- Smets, I.Y., Versyck, K.J. e Van Impe, J.F., (2002b). Optimal Control Theory: A Generic Tool for Identification and Control of (Bio-)Chemical Reactors. *Annual Reviews in Control*, 26, 57-73.

Stigter, J.D. e Keesman, K.J., (2004). Optimal parametric sensitivity control of a fed-batch reactor. *Automatica*, 40 (8), 1459-1464.

Van Impe, J.F., (1993). *Modeling and optimal adaptive control of biotechnological processes*. Tese de Doutorado, Universidade Católica de Leuven, Bélgica.

Vanrolleghem, P.A. e Dochain, D., (1998). Bioprocess model identification. Em: Van Impe, J.F.M., Vanrolleghem, P.A. e Iserentant, D.M. (eds), *Advanced Instrumentation Data Interpretation, and Control of Biotechnological Processes*. pp. 251-318, Kluwer Academic Publishers, Holanda.

Veloso, A.C.A., Rocha, I. e Ferreira, E.C., (2004). Identification of yield coefficients in an *E. coli* model - an optimal experimental design using Genetic Algorithms. *9th Symposium on Computer Application in Biotechnology, Nancy, France*

Versyck, K.J., (2000). *Dynamic Input Design for Optimal Estimation of Kinetic Parameters in Bioprocess Models*. Tese de Doutorado, Universidade Católica de Leuven, Bélgica.

Versyck, K.J., Bernaerts, K., Geeraerd, A.H. e Van Impe, J.F., (1999). Introducing optimal experimental design in predictive modeling: A motivating example. *International Journal of Food Microbiology*, 51 (1), 39-51.

Versyck, K.J., Claes, J.E. e Van Impe, J.F., (1998). Optimal experimental design for practical identification of unstructured growth models. *Mathematics and Computers in Simulation*, 46 (5-6), 621-629.

Versyck, K.J., Claes, J.E. e Van Impe, J.F., (1997). Practical identification of unstructured growth kinetics by application of optimal experimental design. *Biotechnology Progress*, 13 (5), 524-531.

Versyck, K.J. e Van Impe, J.F., (1999). Feed rate optimization for fed-batch bioreactors: from optimal process performance to optimal parameter estimation. *Chemical Engineering Communications*, 172, 107-124.

Walter, E. e Pronzato, L., (1997). *Identification of parametric models from experimental data*, Springer, Reino Unido.

Walter, E. e Pronzato, L., (1990). Qualitative and quantitative experiment design of phenomenological models - a survey. *Automatica*, 26 (2), 195-213.

Wang, F.S. e Cheng, W.M., (1999). Simultaneous optimization of feeding rate and operation parameters for fed-batch fermentation processes. *Biotechnology Progress*, 15 (5), 949-952.

Wang, F.S., Su, T.-L. e Jang, H.-J., (2001). Hybrid differential evolution for problems of kinetic parameter estimation and dynamic optimization of an ethanol fermentation process. *Industrial and Engineering Chemistry Research*, 40 (13), 2876-2885.

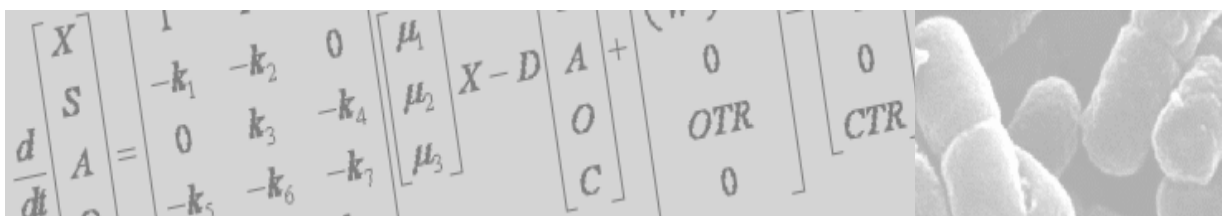
Wirtz, K.W., (2002). A generic model for changes in microbial kinetic coefficients. *Journal of Biotechnology*, 97 (2), 147-162.

Xu, B., Jahic, M. e Enfors, S.O., (1999). Modeling of overflow metabolism in batch and fed-batch cultures of *Escherichia coli*. *Biotechnology Progress*, 15 (1), 81-90.

Zhou, Y.H. e Titchener-Hooker, N.J., (2003). The application of a Pareto optimisation method in the design of an integrated bioprocess. *Bioprocess and Biosystems Engineering*, 25 (6), 349-355.

CAPÍTULO 4

MONITORIZAÇÃO DO PROCESSO



4. MONITORIZAÇÃO DO PROCESSO

Durante as duas últimas décadas, tem-se assistido a um crescimento e grande desenvolvimento de processos biotecnológicos à escala industrial, quer a nível da produção quer da biotransformação.

De um modo geral, os problemas encontrados na implementação destes processos são semelhantes aos verificados em processos industriais clássicos sendo necessário a sua monitorização e controlo com vista ao melhoramento do seu desempenho, à obtenção de um produto de melhor qualidade assim como, na detecção de perturbações no processo. Contudo, a supervisão de bioprocessos tem sido dificultada devido ao facto de, hoje em dia, ser ainda difícil encontrar no mercado sensores baratos e robustos capazes de permitirem a monitorização em tempo real de importantes variáveis do processo necessárias à implementação de estratégias eficientes que influenciam o seu desempenho. Por outro lado, as deficiências dos sensores existentes é notória e estão associadas ao facto destes terem de ser esterilizáveis e das técnicas de amostragem provocarem, muitas vezes, a degradação das propriedades hidrodinâmicas dos reactores. Assim, a medição de variáveis críticas do processo como a concentração da biomassa, do substrato, produtos intermediários e finais é feita normalmente, de modo discreto no tempo e muitas vezes com um tempo longo de análise. Este facto tem impedido a implementação de técnicas de controlo automáticas que facilitem a reprodutibilidade dos processos biológicos e assim permitam a sua afirmação à escala industrial.

Estes problemas têm encorajado o desenvolvimento de algoritmos designados por *software sensors* (Dochain, 2003b) ou sensores por programação (Ferreira, 1995) os quais permitem estimativas em linha, contínuas no tempo, de variáveis de estado que não são medidas em tempo real, baseadas, por um lado, nos sensores físicos disponíveis e, por outro lado, no modelo matemático do processo.

É possível encontrar na literatura várias aplicações de sensores por programação em bioprocessos (Farza *et al.*, 1999; Bogaerts, 1999; Zhang e Guay, 2002; Dochain, 2003a; Rapaport e Dochain, 2005; Raissi *et al.*, 2005; Lemesle e Gouzé, 2005; Chachuat e Bernard, 2006; Nadri *et al.*, 2006), embora a sua aplicação a processos biológicos complexos seja ainda muito reduzida.

Neste capítulo será discutida a aplicação dos observadores de Kalman, assintóticos e por intervalo, ao processo de crescimento de *E. coli* por operação semi-contínua. Será estabelecida uma comparação do desempenho desses observadores, sendo a sua validação feita através de resultados de simulação e dados experimentais.

4.1. SENSORES POR PROGRAMAÇÃO

Monitorizar e operar um processo de fermentação de modo efectivo requer informação contínua sobre as suas principais variáveis de estado. Contudo, como já referido, essas variáveis críticas para o desempenho do processo, em muitas situações, não podem ser medidas ou quando o são é a partir de amostragens pouco frequentes, em intervalos de tempo irregulares. Este problema resulta da inadequação dos sensores disponíveis e também de limitações operacionais como o aumento da possibilidade de contaminação do meio de cultura quanto maior for o número de amostras retirado para análise em diferido. Assim, os sensores por programação são de facto, uma alternativa interessante para ultrapassar estas dificuldades.

O modelo matemático do processo poderá ser utilizado para a obtenção, em linha, de informação não directamente mensurável através de medições mais frequentes de outras variáveis como a temperatura, o pH, a pressão, as concentrações de oxigénio e de dióxido de carbono dissolvidos, a análise dos gases de saída do reaktor, entre outros.

Podem-se considerar dois grupos de variáveis de processo: (i) as variáveis de estado e (ii) os parâmetros do processo. O primeiro grupo refere-se às variáveis que caracterizam unicamente o estado do processo, enquanto que, o segundo grupo é constituído pelas variáveis que afectam o seu estado, como os coeficientes de rendimento e as taxas específicas de crescimento. Os parâmetros do processo são também parâmetros do modelo matemático que podem ser constantes ou variarem no tempo. Como normalmente não são variáveis directamente mensuráveis, o seu valor ou é conhecido *a priori* ou tem que ser estimado a partir de medições de variáveis de estado disponíveis. Assim, os algoritmos utilizados para estimar as variáveis do processo podem ser designados por observadores de estado quando relacionados com a estimação de variáveis de estado ou estimadores se forem algoritmos de estimação de parâmetros (Bastin e Dochain, 1990; Ferreira, 1995).

Um sensor por programação pode ser descrito como a associação entre o sensor (*hardware*) e o estimador (*software*) como representado na Figura 4.1.

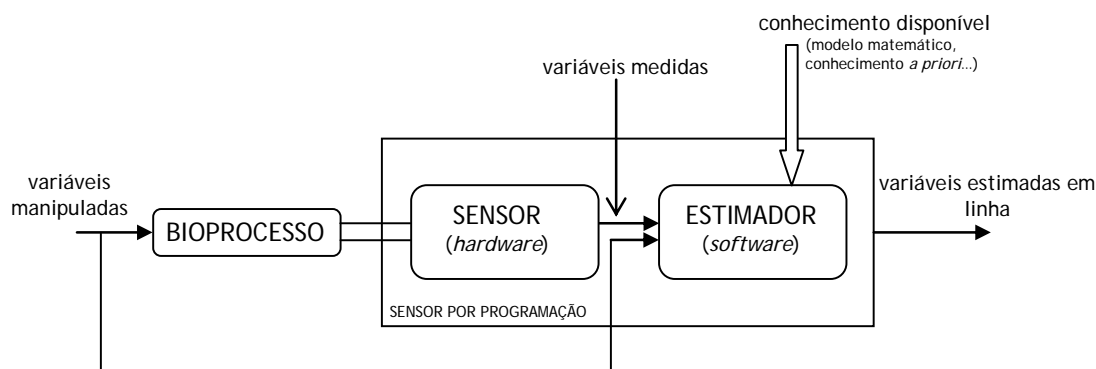


Figura 4.1 Sensores por programação (adaptado de Farza *et al.* (1998)).

O observador/estimador pode ser definido como o algoritmo que fornece a estimativa em linha das variáveis de estado/parâmetros que não são normalmente medidos a partir da medição de variáveis de saída utilizando sensores (Cheruy, 1997).

Actualmente, tem-se assistido a um crescente interesse no desenvolvimento de sensores por programação para processos bioquímicos e químicos (Soroush, 1997; Lombardi *et al.*, 1999; Zhang e Guay, 2002; Aguilar-Lopez e Maya-Yescas, 2005; Nadri *et al.*, 2006). Contudo, ao projecto e à aplicação desses sensores a bioprocessos está associada uma dificuldade importante que reside na incerteza de alguns dos termos dos modelos usados para a descrição da sua dinâmica.

Assim, na literatura podem encontrar-se duas classes de observadores para bioprocessos. A primeira classe inclui os observadores clássicos do tipo Luenberger e Kalman e observadores não lineares, que se baseiam no conhecimento completo da estrutura do modelo. Têm sido publicados vários trabalhos com a aplicação destes observadores a processos biológicos, particularmente do observador estendido de Kalman (Dubach e Märkl, 1992; Bogaerts, 1999; Neeleman *et al.*, 2000; Neeleman, 2001; Arndt e Hitzmann, 2004; Arndt *et al.*, 2005). No entanto, e apesar dos bons resultados, estes observadores apresentam um problema, uma vez que a sua teoria é desenvolvida usando o conhecimento completo dos parâmetros do sistema, em particular da sua cinética. Em alguns casos, estão associadas grandes incertezas aos parâmetros e à cinética do processo, o que pode originar grandes desvios nas estimativas das variáveis não medidas, não se conseguindo a convergência nem a estabilidade destes algoritmos (Dochain, 2003b; Chachuat e Bernard, 2006).

A segunda classe de observadores, designados por Observadores Assimptóticos, proposta por Bastin e Dochain (1990), é baseada na hipótese de que a incerteza dos modelos dos bioprocessos se concentra nos modelos cinéticos dos processos. O projecto destes observadores é baseado em balanços de massa, não requerendo o conhecimento da cinética dos processos. Estes algoritmos apresentam o inconveniente da taxa de convergência da estimação ser completamente dependente das condições de operação (Bastin e Dochain, 1990; Bogaerts, 1999; Dochain, 2001).

Mais recentemente, foram propostos os Observadores por Intervalo e por Probabilidade (Gouzé *et al.*, 2000; Rapaport e Gouzé, 2003; Chachuat e Bernard, 2004; Rapaport e Dochain, 2005; Chachuat e Bernard, 2006). A sua aplicação prende-se com a estimação de variáveis não medidas ou parâmetros para uma classe de sistemas dinâmicos com incertezas, isto é, quando existe uma parte do seu modelo que não é conhecida com precisão. Estes algoritmos têm como objectivo estimar um intervalo que contém o verdadeiro valor da variável ou a sua densidade de probabilidade, respectivamente, atendendo às incertezas do modelo ou de medições (por exemplo, parâmetros do modelo, composição da alimentação). O projecto destes observadores é baseado na definição de sistemas cooperativos (Rapaport e Dochain, 2005). Têm sido descritas na literatura, algumas aplicações desta abordagem a sistemas biológicos (Hadj-Sadok e Gouzé, 2001; Alcaraz-Gonzalez *et al.*, 2002).

Durante a década passada surgiram ainda algumas técnicas alternativas que podem ser aplicadas ao problema da estimação de estados e de parâmetros, como redes neuronais e outras ferramentas

do campo da inteligência artificial (de Assis e Maciel, 2000; James *et al.*, 2000; James *et al.*, 2002; Komives e Parker, 2003; Chen *et al.*, 2004; Vanek *et al.*, 2004; Jenzsch *et al.*, 2006), análise multivariada e quimiometria (Franz *et al.*, 2005; Jenzsch *et al.*, 2006).

4.1.1. Observadores de estado

Genericamente, não existem sensores robustos para medir cada variável de um sistema biológico. Nesse contexto, o projecto de observadores é muito interessante em virtude de permitir estimar as concentrações dos principais compostos químicos e biológicos no biorreactor.

Considerando o modelo dinâmico não linear:

$$\begin{aligned} \frac{dx}{dt} &= f(x, u) \\ y &= h(x) \end{aligned} \quad \text{Equação 4.1}$$

em que y representa as variáveis medidas, pode-se definir um observador para o sistema anterior como:

$$\begin{aligned} \frac{d\hat{x}}{dt} &= f(\hat{x}, u) + G(\hat{x})(y - \hat{y}) \\ \hat{y} &= h(\hat{x}) \end{aligned} \quad \text{Equação 4.2}$$

onde \hat{x} e \hat{y} são as estimativas em linha de x e y obtidas através do observador e G é o ganho do observador.

Originalmente, o observador da Equação 4.2 foi desenvolvido para problemas lineares. Contudo, devido às características não lineares da dinâmica dos bioprocessos houve interesse em estender estes conceitos e explorar estruturas particulares para a sua aplicação a sistemas biológicos.

Definindo o erro de observação como se segue:

$$e = x - \hat{x} \quad \text{Equação 4.3}$$

a dinâmica do erro de observação é dada por:

$$\frac{de}{dt} = f(\hat{x} + e, u) - f(\hat{x}, u) - G(\hat{x})(h(\hat{x} + e) - h(\hat{x})) \quad \text{Equação 4.4}$$

O projecto do observador pode ser formulado como a escolha da matriz $G(\hat{x})$ tal que a dinâmica do erro (Equação 4.4) tenha determinadas propriedades. Deste conceito resultou o projecto de diferentes tipos de observadores.

Como definido por Bastin e Dochain (1990), um bioprocessos pode ser descrito por um conjunto de m reacções que envolvem n compostos. Assim, no presente trabalho são apresentados algoritmos de estimação de estados baseados no denominado modelo dinâmico geral de reactores biológicos, já apresentado no capítulo anterior:

$$\frac{d\xi}{dt} = Kr(\xi, t) - D\xi + F - Q \quad \text{Equação 4.5}$$

Supondo conhecidos os coeficientes de rendimento da matriz K , as expressões para as taxas de reacção $r(\xi, t)$, a taxa de diluição D , os vários caudais dos vectores F e Q e um subconjunto do vector de estado, ξ_1 , por medição em linha, é possível projectar um observador de estado para reconstrução das variáveis de estado não medidas.

Bastin e Dochain (1990) propõem a seguinte equação genérica de um observador de estado para sistemas não lineares do tipo descrito pela equação anterior:

$$\frac{d\hat{\xi}}{dt} = Kr(\hat{\xi}, t) - D\hat{\xi} + F - Q + \Omega(\hat{\xi})[\xi_1 - \hat{\xi}_1] \quad \text{Equação 4.6}$$

em que $\hat{\xi}$ e $\hat{\xi}_1$ representam a estimativa em linha de ξ e ξ_1 , respectivamente. $\Omega(\hat{\xi})$ é uma matriz de ganhos de dimensão $n \times q$, função de $\hat{\xi}$ e q o número de variáveis medidas em linha. Esta equação pode ser vista como sendo uma cópia do modelo dado pela Equação 4.5 acrescentada de um termo matriz do erro de observação na parte medida do estado.

O projecto do observador consiste em escolher a matriz ganho, $\Omega(\hat{\xi})$, mais adequada.

4.1.1.1. Observador estendido de Kalman

O observador de Kalman é o mais adequado para processos lineares se o modelo do sistema e algumas propriedades das variáveis medidas estiverem disponíveis. Quando são incluídas não linearidades no modelo matemático do processo, deverá ser utilizada a versão estendida do observador de Kalman (Biagiola e Figueroa, 2004).

Devem ser assumidas as seguintes considerações de modo a obter-se o observador estendido de Kalman (OEK):

- (i) conhecimento completo do modelo, isto é, a estrutura das reacções cinéticas $r(\xi, t)$ é totalmente conhecida;

- (ii) os valores numéricos de todos os parâmetros do modelo (coeficientes de rendimento e cinéticos) são também conhecidos; e
- (iii) D , F e Q são medidos em linha, juntamente com um subconjunto q de variáveis de estado.

O vector das variáveis de estado medidas é designado por ξ_1 e está relacionado com o estado do sistema como se segue:

$$\xi_1 = L\xi \tag{Equação 4.7}$$

onde a matriz L de dimensão $(q \times n)$ é uma matriz elementar que selecciona os componentes medidos de ξ . Por outro lado, o vector das variáveis de estado não medidas é representado por ξ_2 . (ξ_1, ξ_2) constituem partições de estado de ξ .

Como referido anteriormente, a Equação 4.6 representa o modelo geral dos observadores para sistemas não lineares. O problema da formulação do observador de estado reduz-se à escolha adequada da matriz ganho $\Omega(\xi, t)$. Para resolver este problema, é introduzido o erro de observação $e = \xi - \hat{\xi}$, e a sua dinâmica deduzida (Bastin e Dochain, 1990).

Considerando a aproximação da tangente linearizada da dinâmica do modelo do erro de observação na vizinhança de $e=0$, vem:

$$\frac{de}{dt} = [M(\hat{\xi}) - \Omega(\hat{\xi})L]e \tag{Equação 4.8}$$

com:

$$M(\hat{\xi}) \equiv K \left[\frac{\partial r(\xi, t)}{\partial \xi} \right]_{\xi=\hat{\xi}} - D I_N \tag{Equação 4.9}$$

onde I_N é a matriz identidade de dimensão $n \times n$.

Considerando que o modelo dado pela Equação 4.5 é exponencialmente observável, sendo a condição de observabilidade dada por:

$$\text{característica} \begin{bmatrix} L \\ LM(\hat{\xi}) \\ LM(\hat{\xi})^2 \\ \vdots \\ LM(\hat{\xi})^{N-1} \end{bmatrix} = N \tag{Equação 4.10}$$

o desenho do OEK reduz-se a um problema de optimização quadrática para determinar a matriz $\Omega(\hat{\xi}, t)$ que minimize o erro de observação médio quadrático, tendo em conta a restrição do modelo do erro da tangente linear (Equação 4.8 e Equação 4.9). A solução deste problema de optimização é dada por:

$$\Omega(\hat{\xi}, t) = R(\hat{\xi}, t)L^T \quad \text{Equação 4.11}$$

onde a matriz $R(\hat{\xi}, t)$, simétrica e quadrada de ordem n , é gerada pela equação de *Riccati*:

$$\frac{dR}{dt} = -RL^T LR + RM^T(\hat{\xi}, t) + M(\hat{\xi}, t)R \quad \text{Equação 4.12}$$

A maior desvantagem destes observadores advém do facto da qualidade da estimativa $\hat{\xi}$ ser fortemente dependente da precisão do modelo (Equação 4.5) e especialmente do modelo cinético, $r(\xi, t)$.

Um observador estendido de Kalman foi usado por Arndt e Hitzmann (2004) para estimar a concentração de biomassa, a taxa específica de crescimento e o volume do meio de cultura durante uma fermentação de *Saccharomyces cerevisiae*. A combinação da monitorização em linha da concentração de glucose com este observador permitiu controlar a glucose em níveis de 0,08 e 0,05 g/L.

4.1.1.2. Observadores Assintóticos

A ideia subjacente aos observadores assintóticos (OA) consiste em tirar partido da estrutura do modelo dinâmico dos processos biológicos de modo a obter uma parte do modelo numa forma independente da cinética (Bastin e Dochain, 1990). Estes observadores permitem então, a estimação das variáveis de estado não medidas quando o processo não é exponencialmente observável e a estrutura das reacções cinéticas é desconhecida. Assim, estes observadores são propostos considerando as seguintes condições:

- (i) os coeficientes de rendimento (matriz K) são conhecidos; e
- (ii) o número de q variáveis de estado medidas é igual ou superior à característica da matriz K : $q \geq p = \text{rank}(K)$ (Bastin e Dochain, 1990).

Definindo uma partição do vector de estado como:

$$\xi = \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} \quad \text{Equação 4.13}$$

em que ξ_1 e ξ_2 representam as variáveis de estado medidas e não medidas, respectivamente, o modelo dinâmico geral pode ser então reescrito na forma:

$$\frac{d\xi_1}{dt} = K_1 r(\xi, t) - D\xi_1 + F_1 - Q_1 \quad \text{Equação 4.14}$$

$$\frac{d\xi_2}{dt} = K_2 r(\xi, t) - D\xi_2 + F_2 - Q_2 \quad \text{Equação 4.15}$$

Introduzindo a seguinte transformação de estado, que corresponde a uma mudança linear de coordenadas:

$$Z \equiv \xi_2 - K_2 K_1^{-1} \xi_1 \quad \text{Equação 4.16}$$

onde K_1^{-1} representa a pseudo inversa da matriz K_1 , considerando que a matriz K_1 apresenta característica completa. As matrizes K_1 e K_2 são obtidas através da matriz K por aplicação da partição escolhida.

A dinâmica de Z é independente das velocidades de reacção $r(\xi, t)$ e é dada pela seguinte equação:

$$\frac{dZ}{dt} = -DZ - K_2 K_1^{-1} (F_1 - Q_1) + (F_2 - Q_2) \quad \text{Equação 4.17}$$

O problema da estimação de estados consistirá no projecto de um algoritmo para reconstruir as variáveis de estado não medidas em linha a partir das variáveis medidas, considerando-se conhecidos (por imposição ou medição em linha) a taxa de diluição D , os caudais de entrada F , os caudais gasosos Q , e um subconjunto do vector de variáveis de estado medidas em linha, designados por ξ_1 , em que q é o número de variáveis de estado medidas em linha. Representando por ξ_2 o vector das variáveis de estado não medidas, pode derivar-se a equação do observador assintótico com base nas duas equações anteriores:

$$\frac{d\hat{Z}}{dt} = -D\hat{Z} - K_2 K_1^{-1} (F_1 - Q_1) + (F_2 - Q_2) \quad \text{Equação 4.18}$$

$$\hat{\xi}_2 = \hat{Z} + K_2 K_1^{-1} \xi_1 \quad \text{Equação 4.19}$$

em que \hat{Z} e $\hat{\xi}_2$ representam as estimativas em linha de Z e ξ_2 , respectivamente.

Este observador é classificado de assintótico pelo facto da velocidade de convergência ser determinada pelas condições experimentais (pela taxa de diluição). A análise de erro deste observador $(\tilde{\xi}_2 = \xi_2 - \hat{\xi}_2)$ conduz ao ponto de equilíbrio $(\tilde{\xi}_2 = 0)$ de estabilidade assintótica. A convergência deste algoritmo foi demonstrada pelos autores (Bastin e Dochain, 1990).

A principal vantagem destes observadores é a sua completa independência da cinética do processo. Contudo, ao contrário do OEK, a sua convergência de estimação é completamente determinada pelas condições experimentais através do valor da taxa de diluição, implicando assim que $D(t)$ não possa permanecer igual a zero por um período de tempo excessivamente longo (Bastin e Dochain, 1990).

4.1.1.3. Observadores por Intervalo

Os observadores por intervalo (OI) têm sido aplicados em problemas de estimação de estado em bioprocessos quando o valor dos parâmetros do processo e/ou o valor das variáveis de entrada do processo não são conhecidos com precisão. O seu objectivo é então estimar um intervalo que contenha o verdadeiro valor da variável de estado, o qual está relacionado com as incertezas associadas ao modelo e/ou às medições.

O projecto destes observadores é baseado nas propriedades de sistemas cooperativos, isto é, sistemas dinâmicos para os quais os elementos da não diagonal da matriz Jacobiano são positivos ou nulos (Smith, 1996; Rapaport e Dochain, 2005). Deste modo, se para todo (x, t) , o sistema representado pela Equação 4.1 é cooperativo, então verifica-se a condição:

$$\frac{\partial f_i}{\partial x_j}(x, t) \geq 0, \quad \text{para } i \neq j \quad \text{Equação 4.20}$$

Sendo assim, considerando o sistema cooperativo da Equação 4.1 e sabendo que f^- e f^+ são dois vectores tais que:

$$f^-(x) \leq f(x, t) \leq f^+(x), \quad \forall (x, t) \quad \text{Equação 4.21}$$

e as condições iniciais x_0^-, x_0, x_0^+ tais que $x_0^- \leq x_0 \leq x_0^+$, a solução do sistema dinâmico:

$$\frac{d}{dt} \begin{pmatrix} x^- \\ x \\ x^+ \end{pmatrix} = \begin{pmatrix} f^-(x^-) \\ f(x, t) \\ f^+(x^+) \end{pmatrix}, \quad \text{com} \quad \begin{pmatrix} x^-(0) \\ x(0) \\ x^+(0) \end{pmatrix} = \begin{pmatrix} x_0^- \\ x_0 \\ x_0^+ \end{pmatrix} \quad \text{Equação 4.22}$$

admite a seguinte propriedade:

$$x^-(t) \leq x(t) \leq x^+(t), \quad t \geq 0$$

Equação 4.23

O pressuposto da estimação usando observadores por intervalo pode ser formulado do seguinte modo: sendo conhecidos o limite superior e inferior das incertezas do modelo (por exemplo, da concentração do substrato na alimentação ou do valor dos coeficientes de rendimento) podem-se estimar também os limites superior e inferior das variáveis estimadas. Sendo assim, com estes observadores pretendem-se obter duas estimativas, limite superior e limite inferior, que contenham as variáveis não medidas. Por outro lado, pretende-se que o intervalo $[x^-(t), x^+(t)]$ seja óptimo, isto é, o mais pequeno possível e que idealmente tenda para o valor de $x(t)$ quando t aumenta.

Os observadores por intervalo têm sido particularmente aplicados com sucesso em sistemas de tratamento de águas residuais onde a composição da alimentação é normalmente pouco conhecida (Hadj-Sadok e Gouzé, 2001; Alcaraz-Gonzalez *et al.*, 2002). No trabalho de Alcaraz-Gonzalez *et al.* (2002) é feita a validação de observadores por intervalo usando dados experimentais de um reactor de leito fixo de 1 m³ utilizado no tratamento de águas residuais provenientes de destilarias industriais. O seu desempenho foi satisfatório, apresentando boas propriedades de convergência, estimando convenientemente o intervalo dinâmico das variáveis não medidas (carência química de oxigénio (CQO), concentração de biomassa e alcalinidade) a partir das medições em linha da taxa de diluição, da concentração de ácidos gordos voláteis, de carbono inorgânico total e do caudal molar de dióxido de carbono gasoso.

Bernard e Gouzé (2004) apresentam um estudo em que utilizam não apenas um observador por intervalo, mas um conjunto de observadores inicializados simultaneamente, permitindo que a intercepção dos intervalos obtidos para cada um dos observadores melhorasse simultaneamente a velocidade de convergência e os intervalos assintóticos das variáveis estimadas.

Rapaport e Dochain (2005) propõem o projecto de um observador por intervalo para o crescimento de um microrganismo num reactor perfeitamente agitado usando como modelo cinético a equação de *Monod* e *Haldane* e conhecendo o limite superior e inferior da taxa específica de crescimento e da concentração de substrato na alimentação. Nesse trabalho é estudada a estabilidade e convergência do observador com e sem incertezas nas medidas. Estes autores concluem que no caso particular do processo onde as variáveis de entrada são conhecidas, o erro de observação das variáveis não medidas é inversamente proporcional a um dos ganhos do observador. O desempenho dos observadores por intervalo foi estudado a partir de simulações numéricas.

4.1.2. Estimadores de cinética

Como já referido, é frequente, em processos biológicos, encontrar situações em que alguns dos parâmetros do processo são desconhecidos ou conhecidos com reduzida precisão o que dificulta o controlo do processo. Sendo assim, os sensores por programação têm que ser usados para estimar não só as variáveis de estado não medidas mas também os parâmetros cinéticos associados

(Lubenova, 1996; Tatiraju *et al.*, 1999; Keesman, 2002; Lubenova *et al.*, 2003; Bogaerts e Vande Wouwer, 2004).

A aplicação de sensores por programação para estimar as velocidades de reacção é muito importante, uma vez que são funções extremamente complexas das condições de operação e das variáveis de estado do processo. A modelação destas funções é difícil e continua a ser motivo de contínua e intensa investigação.

Na literatura podem encontrar-se trabalhos que apresentam sensores por programação adequados para estimar simultaneamente as variáveis de estado e os parâmetros do processo quando a estrutura do modelo cinético é conhecida, mas os seus parâmetros não (Farza *et al.*, 1998; Dochain, 2003a). Uma aproximação apresentada nesses trabalhos baseia-se no projecto dos observadores estendidos de Luenberger e Kalman nos quais os parâmetros cinéticos desconhecidos são usados como parâmetros extra do observador de modo a garantir que a dinâmica do erro de estimação das variáveis do processo estimadas seja estacionária (Dochain, 2003a).

Uma outra aproximação consiste num observador adaptativo onde os parâmetros cinéticos são também estimados em linha (Bastin e Dochain, 1990; Dochain, 2003a).

Estes dois tipos de estimadores combinam a estimação das variáveis de estado e dos parâmetros no mesmo algoritmo, sendo esperado que se consiga um melhor desempenho, uma maior estabilidade e convergência do estimador quando a estimação destas variáveis do processo esteja desacoplada. Nesse sentido, Bastin e Dochain (1990) propuseram um estimador baseado num observador que consiste num algoritmo que permite o desacoplamento da estimação de estados da estimação dos parâmetros cinéticos. A maior parte dos trabalhos propostos sobre a estimação das velocidades de reacção em bioprocessos é baseada na teoria desenvolvida por estes autores (Biagiola *et al.*, 2000; November e Van Impe, 2002). Farza e co-autores (Farza *et al.*, 1997a; Farza *et al.*, 1997b; Farza *et al.*, 1999) propuseram um estimador baseado num observador fundamentado pela teoria dos sistemas não lineares. Nestes trabalhos as velocidades de reacção são estimadas através das variáveis de estado. Outros autores utilizaram a velocidade de transferência de oxigénio (*OTR*) e de dióxido de carbono (*CTR*) como medidas para estimar as taxas de crescimento (Lubenova, 1999). Existem ainda trabalhos que se centram no estudo da sintonização e calibração destes algoritmos, sendo propostos estimadores de dinâmica de segunda ordem (Oliveira *et al.*, 1996; Perrier *et al.*, 2000; Oliveira *et al.*, 2002).

No presente trabalho, a velocidade específica de crescimento (μ) é estimada recorrendo ao desenho de um estimador baseado num observador, como proposto por Bastin e Dochain (1990) e reformulado por Pomerleau e Perrier (1990).

A equação geral do estimador é dada por:

$$\frac{d\hat{\psi}}{dt} = \hat{\mu}\hat{X} - D\hat{\psi} + K_a^{-1}(F_a - Q_a) - \Omega_1(\psi - \hat{\psi}) \quad \text{Equação 4.24}$$

$$\frac{d\hat{\mu}}{dt} = \Omega_2(\psi - \hat{\psi}) \quad \text{Equação 4.25}$$

em que ψ é dado pela transformação, $\psi = K_a^{-1} \xi_a$, onde ξ_a é a partição do vector ξ , constituído pelas variáveis de estado medidas e K_a , F_a , Q_a as partições correspondentes de K , F e Q . Esta transformação é introduzida para desacoplamento do modelo dinâmico geral relativamente às taxas de reacção (Pomerleau e Perrier, 1990). A escolha das matrizes dos ganhos é feita de modo a que se obtenha uma dinâmica de segunda ordem como apresentado no trabalho de Oliveira *et al.* (1996). A determinação dos ganhos ω_{ij} para cada instante t é feita de acordo com as equações:

$$\omega_{ij,t} = \frac{2\zeta_j}{\tau_j} - \frac{X_t - X_{t-1}}{TX_t}, \quad \text{com } i=1 \text{ e } j=1, \dots, m \quad \text{Equação 4.26}$$

$$\omega_{ij,t} = \frac{1}{\tau_j^2 X_t}, \quad \text{com } i=2 \text{ e } j=1, \dots, m \quad \text{Equação 4.27}$$

onde m é o número de velocidades específicas de crescimento e T é o passo de integração. A implementação destes algoritmos requer a sintonização, por tentativa erro, dos parâmetros τ_j e ζ_j .

4.2. RESULTADOS

A monitorização de bioprocessos utilizando sensores que permitem medições em linha é essencial para garantir o sucesso das estratégias de controlo, uma vez que estas dependem de medições regulares e precisas. Contudo, como já referido, é ainda difícil encontrar no mercado sensores robustos e precisos que permitam a medição em linha de todas as variáveis de estado fundamentais para o controlo do processo.

O processo em estudo, fermentação semi-contínua de elevada densidade celular de *E. coli*, é de grande importância na indústria farmacêutica, visto que a *E. coli* é o organismo mais utilizado na produção de várias proteínas recombinadas. Existem ainda dificuldades, difíceis de ultrapassar, na implementação de algoritmos de controlo deste processo, pois algumas das variáveis de estado não são facilmente medidas em linha. De entre as variáveis de estado, a concentração de biomassa adquire um papel de destaque no controlo, na estimação das velocidades específicas de crescimento e na optimização da produção de proteínas recombinadas, tendo em conta quer a produtividade, quer o tempo de indução. Assim, a monitorização da concentração da biomassa, para além de ser essencial para as estratégias de controlo, permite verificar a sua consistência e

consequentemente garantir a qualidade do produto final. No entanto, hoje em dia, é ainda difícil medir em linha a concentração de biomassa, embora estejam a ser feitos esforços significativos para o desenvolvimento de sensores que permitam a sua monitorização (Sonnleitner, 2003; Vojinovic *et al.*, 2006). Por outro lado, a disponibilidade e precisão dos instrumentos já desenvolvidos é limitada, principalmente para processos industriais. É então urgente obter informação acerca dessa variável através de outras que sejam mais facilmente medidas, utilizando para isso, técnicas de estimação como os sensores por programação.

A importância da concentração de biomassa no desempenho do processo motiva o estudo realizado neste capítulo. A seguir serão apresentados e discutidos os resultados obtidos no estudo da aplicabilidade de observadores de estado na estimação da concentração da biomassa. Este estudo consistiu na comparação do desempenho dos observadores estendido de Kalman, assintóticos e por intervalo, sendo a sua validação feita através de resultados de simulação e dados experimentais. Os resultados de simulação do modelo foram obtidos por integração numérica da Equação 3.55 (Capítulo 3), usando a subrotina ODE23s do MATLAB. A implementação dos observadores, quer para dados simulados quer para dados experimentais, foi efectuada através do método de integração de Euler. O estudo da observabilidade do modelo, assim como a maior parte das operações matemáticas necessárias para o desenho dos observadores foram efectuadas através da rotina *Symbolic Math Toolbox* que corre em MATLAB.

4.2.1. Observador estendido de Kalman

No caso da fermentação semi-contínua de *E. coli* considerada neste trabalho, a condição de observabilidade exponencial (Bastin e Dochain, 1990) foi estudada para nove combinações diferentes de variáveis medidas e estimadas com o intuito de verificar a aplicabilidade do OEK para este processo particular. Cada situação foi classificada como observável ou não observável. Nos casos observáveis distinguiram-se ainda as situações em que é usado o modelo global - *MG* (Equação 3.55) e quando para ser observável são utilizados os modelos parciais - *MP* (Equação 3.59 e Equação 3.60). Esses resultados são apresentados na Tabela 4.1.

O estudo realizado permitiu concluir que o OEK pode ser aplicado a fermentações semi-contínuas de *E. coli* num número limitado de situações. No entanto, é de realçar que com os sensores comercialmente disponíveis para medir a concentração de oxigénio e de dióxido de carbono dissolvidos, é possível estimar a concentração em linha de biomassa e de outras variáveis de estado se as células não crescerem apenas em regime oxidativo em acetato ou oxidativo em glucose.

Para o caso 4, referido na Tabela 4.1, em que as variáveis medidas em linha são o O e C , a seguinte partição de estado é escolhida: $\xi_1^T = [O \ C]$ e $\xi_2^T = [X \ S \ A]$. Deste modo, a matriz L toma o seguinte valor:

$$L = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Equação 4.28

Tabela 4.1 Estudo da observabilidade do modelo para diferentes combinações de variáveis medidas e estimadas usando o OEK⁽ⁱ⁾.

Casos estudados	Variáveis medidas	Variáveis estimadas	Observabilidade do modelo
1	<i>A, O, C</i>	<i>X, S</i>	Observável - <i>MG</i>
2	<i>S, A, O</i>	<i>X, C</i>	Não observável
3	<i>S, O, C</i>	<i>X, A</i>	Observável - <i>MG</i>
4	<i>O, C</i>	<i>X, S, A</i>	Observável - <i>MP</i>
5	<i>A, O</i>	<i>X, S, C</i>	Não observável
6	<i>S, O</i>	<i>X, A, C</i>	Não observável
7	<i>S, C</i>	<i>X, A, O</i>	Não observável
8	<i>S, A</i>	<i>X, O, C</i>	Não observável
9	<i>A, C</i>	<i>X, S, O</i>	Não observável

(i) *OTR, CTR, W* e *F* são medidos em linha em todos os casos.

A matriz $M(\hat{\xi}) = M(\hat{X}, \hat{S}, \hat{A}, \hat{O}, \hat{C})$ é dada pela Equação 4.9, onde:

$$\left[\frac{\partial r(\xi, t)}{\partial \xi} \right]_{\xi=\hat{\xi}} = \begin{bmatrix} \frac{\partial(\mu_1 \hat{X})}{\partial \hat{X}} & \frac{\partial(\mu_1 \hat{X})}{\partial \hat{S}} & \frac{\partial(\mu_1 \hat{X})}{\partial \hat{A}} & \frac{\partial(\mu_1 \hat{X})}{\partial \hat{O}} & \frac{\partial(\mu_1 \hat{X})}{\partial \hat{C}} \\ \frac{\partial(\mu_2 \hat{X})}{\partial \hat{X}} & \frac{\partial(\mu_2 \hat{X})}{\partial \hat{S}} & \frac{\partial(\mu_2 \hat{X})}{\partial \hat{A}} & \frac{\partial(\mu_2 \hat{X})}{\partial \hat{O}} & \frac{\partial(\mu_2 \hat{X})}{\partial \hat{C}} \\ \frac{\partial(\mu_3 \hat{X})}{\partial \hat{X}} & \frac{\partial(\mu_3 \hat{X})}{\partial \hat{S}} & \frac{\partial(\mu_3 \hat{X})}{\partial \hat{A}} & \frac{\partial(\mu_3 \hat{X})}{\partial \hat{O}} & \frac{\partial(\mu_3 \hat{X})}{\partial \hat{C}} \end{bmatrix}$$

Equação 4.29

O observador é então escrito de acordo com a Equação 4.5 e a Equação 4.6, sendo o último termo da Equação 4.6 definido como:

$$\Omega(\hat{\xi}, t) \begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \\ \xi_4 \\ \xi_5 \\ \xi_6 \\ \xi_7 \\ \xi_8 \\ \xi_9 \\ \xi_{10} \end{bmatrix} = \begin{bmatrix} \Omega_1 & \Omega_2 \\ \Omega_3 & \Omega_4 \\ \Omega_5 & \Omega_6 \\ \Omega_7 & \Omega_8 \\ \Omega_9 & \Omega_{10} \end{bmatrix} \begin{bmatrix} O - \hat{O} \\ C - \hat{C} \end{bmatrix}$$

Equação 4.30

A matriz ganho $\Omega(\hat{X}, \hat{S}, \hat{A}, \hat{O}, \hat{C})$ é determinada pela Equação 4.11 como se segue:

$$\Omega(\hat{\xi}) = \begin{bmatrix} \Omega_1 & \Omega_2 \\ \Omega_3 & \Omega_4 \\ \Omega_5 & \Omega_6 \\ \Omega_7 & \Omega_8 \\ \Omega_9 & \Omega_{10} \end{bmatrix} = \begin{bmatrix} R_8 & R_9 \\ R_{11} & R_{12} \\ R_{13} & R_{14} \\ R_4 & R_{15} \\ R_{15} & R_5 \end{bmatrix} \quad \text{Equação 4.31}$$

com a matriz R , quadrada e simétrica, definida como:

$$R = \begin{bmatrix} R_1 & R_6 & R_7 & R_8 & R_9 \\ R_6 & R_2 & R_{10} & R_{11} & R_{12} \\ R_7 & R_{10} & R_3 & R_{13} & R_{14} \\ R_8 & R_{11} & R_{13} & R_4 & R_{15} \\ R_9 & R_{12} & R_{14} & R_{15} & R_5 \end{bmatrix} \quad \text{Equação 4.32}$$

Os únicos parâmetros de sintonização são os valores iniciais dos elementos desta matriz, necessários para a resolução numérica da Equação 4.12.

Deve ainda ser referido que o desempenho do OEK é altamente dependente da precisão do modelo do processo, exigindo grande esforço e tempo para a sua formulação. Mais ainda, poderão existir problemas numéricos e dificuldades de convergência devido a aproximações associadas com a linearização do modelo. Contudo, este é um algoritmo muito utilizado em aplicações práticas de problemas de estimação, como previamente referido.

O desempenho do observador foi avaliado pelo cálculo da diferença quadrática entre os dados experimentais e os estimados, de acordo com a seguinte equação:

$$dif_{\xi} = \sum_{j=1}^{np} \left(\frac{\xi_{exp,j} - \xi_{est,j}}{\xi_{exp,j}} \right)^2 \quad \text{Equação 4.33}$$

onde np é o número de pontos experimentais e ξ_{exp} e ξ_{est} representam os valores experimentais e estimados da variável de estado ξ , respectivamente.

O estudo da robustez do algoritmo desenvolvido foi efectuado recorrendo a simulações numéricas, tendo sido seleccionados os casos 1 e 4 constantes na Tabela 4.1. Valores "pseudo reais" das variáveis de estado foram obtidos por integração numérica da Equação 3.55 (Capítulo 3). Estes valores foram corrompidos com ruído branco, de acordo com o desvio padrão calculado para as diferentes variáveis medidas neste processo, originando dados "pseudo experimentais". De seguida, o observador foi aplicado de modo a obter as variáveis "estimadas" através dos dados "pseudo experimentais" correspondentes às variáveis medidas.

Os índices de desempenho do observador, calculados pela Equação 4.33, para os casos 1 e 4 são apresentados na Tabela 4.2. Embora o objectivo principal deste estudo seja a estimação da

concentração de biomassa, apresentam-se também os resultados obtidos na estimação das concentrações de glucose e de acetato. A Figura 4.2 apresenta os resultados da simulação para o caso 4, realçando o facto da concentração de biomassa ser razoavelmente estimada, mesmo após introdução de ruído, demonstrando assim a robustez do OEK.

Tabela 4.2 Índice de desempenho para o OEK para os casos 1 e 4 da Tabela 4.1.

Observador	Caso estudado	dif_x	dif_s	dif_A
OEK	1	$4,55 \times 10^{-3}$	0,0480	---
	4	$4,81 \times 10^{-3}$	0,101	0,165

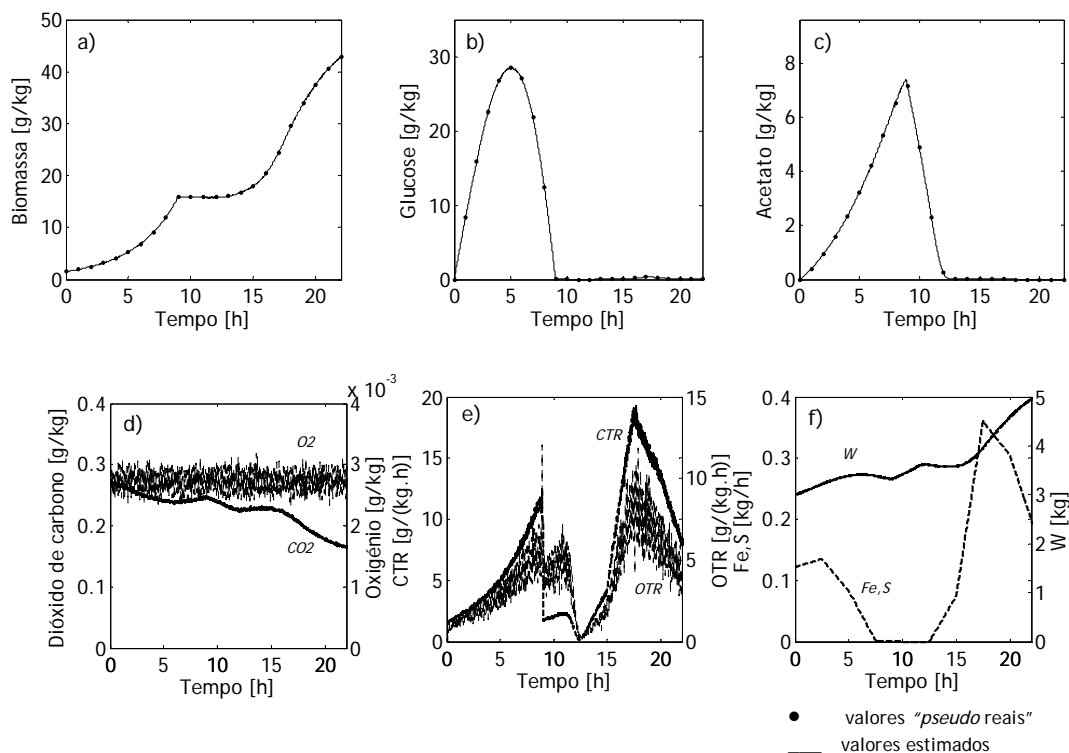


Figura 4.2 Desempenho do OEK na estimativa da concentração de biomassa, de glucose e de acetato para o caso 4 estudado por simulação. Os gráficos d), e) e f) representam as variáveis medidas em linha.

O OEK foi também validado para o caso 4, usando dados experimentais (Exp_4 e Exp_5). Na Tabela 4.3 são apresentados os índices de desempenho obtidos para as referidas experiências. Verificou-se que este observador exibiu um desempenho satisfatório para ambas as experiências, conseguindo estimar razoavelmente a concentração de biomassa. Contudo, como se pode observar a partir da Tabela 4.3, na validação experimental, o desempenho deste observador é pouco satisfatório para a estimação das concentrações de glucose e de acetato.

Tabela 4.3 Índice de desempenho⁽ⁱ⁾ para o OEK para as experiências Exp_4 e Exp_5.

Observador	Experiência	dif_X	dif_S	dif_A
OEK	Exp_4	8,34	442	369
	Exp_5	17,7	2069	738

(i) Índice de desempenho calculado a partir de 250 pontos.

Na Figura 4.3 estão representados os dados medidos em linha usados para estimar a concentração de biomassa referentes à Exp_4. As variáveis medidas em linha foram o O , C , CTR , OTR , $F_{e,s}$ e W .

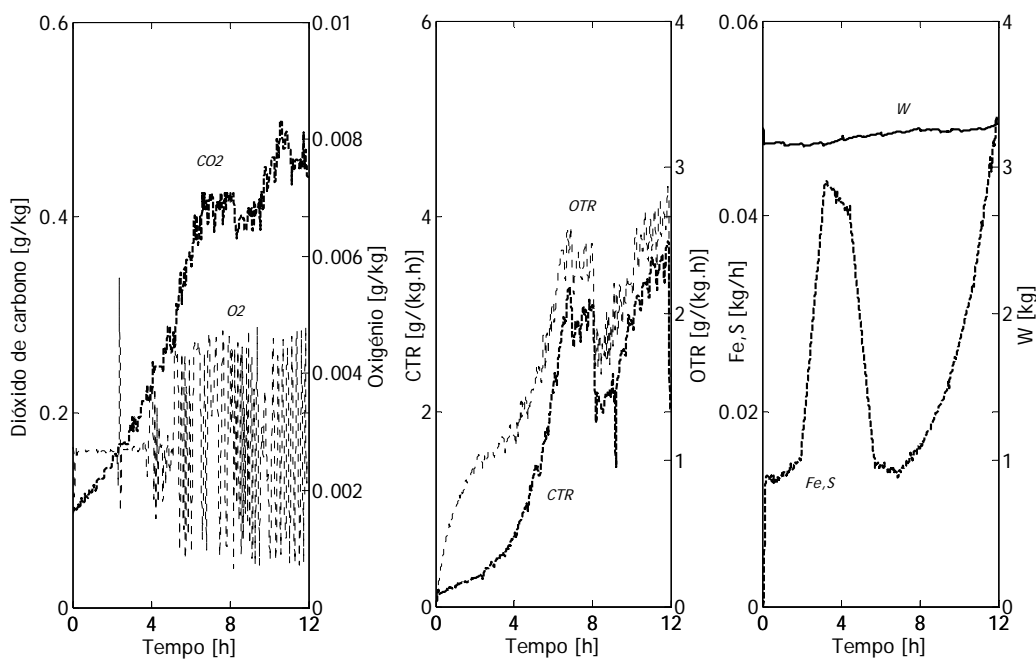


Figura 4.3 Observação de estado numa fermentação semi-contínua de *E. coli* (Exp_4): dados medidos em linha com um intervalo de amostragem de 18 segundos.

A convergência do algoritmo deste observador exigiu um intervalo de amostragem de pelo menos 18 segundos para as variáveis medidas, sendo que intervalos superiores não permitiram a sua convergência. Embora o tempo de amostragem usado seja compatível com o sistema de aquisição de dados existente no laboratório, esta frequência de amostragem não é adequada se outros sensores, como por exemplo o sistema FIA (Rocha e Ferreira, 2002), forem usados para a medição em linha de outras variáveis de estado, nomeadamente a concentração de acetato, que requer cerca de 3 minutos por análise. Contudo, estudos de simulação numérica realizados permitiram verificar que o tempo de amostragem necessário para garantir a convergência do algoritmo deste observador era afectado pelo modelo cinético utilizado. Com efeito, o recurso a um modelo cinético

mais pormenorizado (Rocha, 2003) do que o adoptado neste estudo, permite que, em situações idênticas, o algoritmo convirja com amostragens de 3 em 3 minutos. Este facto vem realçar a importância de um modelo cinético bem ajustado para o comportamento do observador estendido de Kalman, o que representa uma dificuldade acrescida atendendo à complexidade dos modelos cinéticos dos processos biológicos.

A Figura 4.4 apresenta a concordância entre os valores estimados e os valores medidos em diferido para a concentração de biomassa.

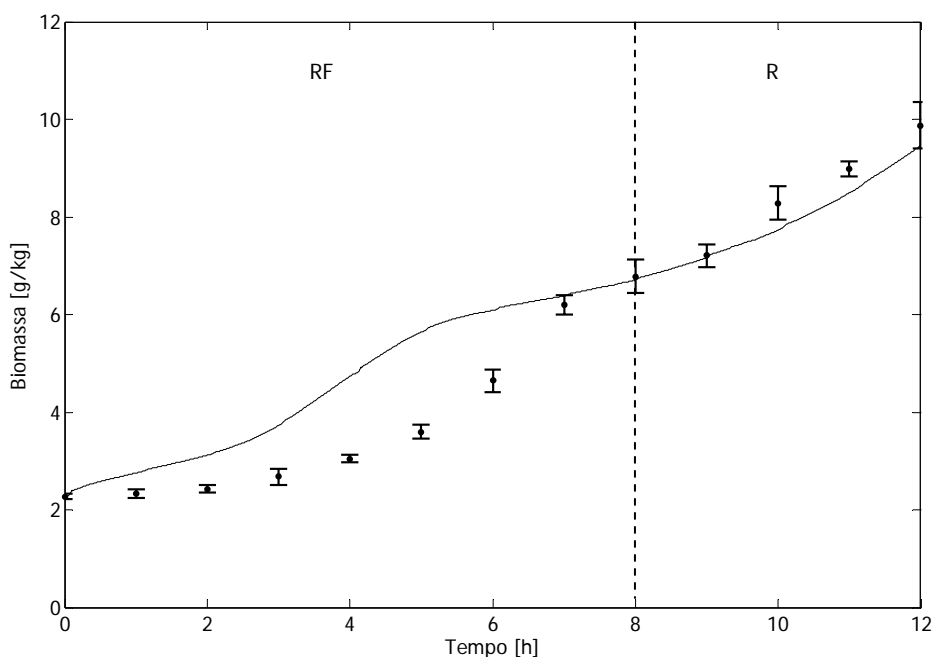


Figura 4.4 Observação de estado numa fermentação semi-contínua de *E. coli* (Exp_4): variável estimada (pontos: valores medidos em diferido; linha: valores estimados; RF: regime respiro-fermentativo; R: regime respirativo).

A análise da Figura 4.4 permite verificar que os valores estimados e medidos em diferido da concentração de biomassa apresentam uma evolução semelhante pelo que se poderá inferir que a estrutura do modelo cinético utilizado neste estudo, embora simplificado, descreve razoavelmente o crescimento da *E. coli* para esta fermentação, embora os valores dos parâmetros cinéticos não tenham sido otimizados.

Uma desvantagem adicional para este tipo de observador é a sintonização dos seus parâmetros (isto é, dos valores iniciais da equação de *Riccati*) que, para o caso em estudo, consistiu na calibração, por tentativa-erro, de 15 parâmetros, o que pode conduzir a uma etapa bastante demorada, requerendo um tempo considerável. Contudo, neste trabalho valores iniciais de 0,001 para todos os parâmetros de sintonização permitiram a convergência e estabilidade do OEK.

4.2.2. Observadores Assintóticos

Os observadores assintóticos (OA) permitem a estimação das variáveis de estado não medidas mesmo quando o processo não é exponencialmente observável e a estrutura das reacções cinéticas é desconhecida.

Para a estrutura do modelo da fermentação semi-contínua da *E. coli* considerada neste trabalho, se o modelo global é considerado, e devido às limitações impostas pela condição $q \geq p = \text{rank}(K)$, o número de variáveis medidas tem que ser igual ou superior a 3. Neste caso, ao contrário do que acontece para o OEK, todas as combinações de três variáveis medidas são teoricamente possíveis. O mesmo ocorre quando os modelos parciais são considerados para duas variáveis de estado medidas, como ilustrado na Tabela 4.4.

Tabela 4.4 Estudo da observabilidade do modelo para diferentes combinações de variáveis medidas e estimadas usando o OA⁽ⁱ⁾.

Casos estudados	Variáveis medidas	Variáveis estimadas	Observabilidade do modelo
1	<i>A, O, C</i>	<i>X, S</i>	Observável - <i>MG</i>
2	<i>S, A, O</i>	<i>X, C</i>	Observável - <i>MG</i>
3	<i>S, O, C</i>	<i>X, A</i>	Observável - <i>MG</i>
4	<i>O, C</i>	<i>X, S, A</i>	Observável - <i>MP</i>
5	<i>A, O</i>	<i>X, S, C</i>	Observável - <i>MP</i>
6	<i>S, O</i>	<i>X, A, C</i>	Observável - <i>MP</i>
7	<i>S, C</i>	<i>X, A, O</i>	Observável - <i>MP</i>
8	<i>S, A</i>	<i>X, O, C</i>	Observável - <i>MP</i>
9	<i>A, C</i>	<i>X, S, O</i>	Observável - <i>MP</i>

(i) *OTR, CTR, W* e *F* são medidos em linha em todos os casos.

Se as variáveis medidas forem, por exemplo, *O* e *C*, como na secção anterior, a matriz usada na transformação de estado da Equação 4.16, para o regime repero-fermentativo, será:

$$K_2 K_1^{-1} = \begin{bmatrix} \alpha_1 & \alpha_4 \\ \alpha_2 & \alpha_5 \\ \alpha_3 & \alpha_6 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ -k_1 & -k_2 \\ 0 & k_3 \end{bmatrix} \begin{bmatrix} -k_5 & -k_6 \\ k_8 & k_9 \end{bmatrix}^{-1} \quad \text{Equação 4.34}$$

e para o regime respirativo:

$$K_2 K_1^{-1} = \begin{bmatrix} \alpha_1 & \alpha_4 \\ \alpha_2 & \alpha_5 \\ \alpha_3 & \alpha_6 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ -k_1 & 0 \\ 0 & -k_4 \end{bmatrix} \begin{bmatrix} -k_5 & -k_7 \\ k_8 & k_{10} \end{bmatrix}^{-1} \quad \text{Equação 4.35}$$

O observador, neste caso, é dado pelas equações seguintes:

$$\frac{d}{dt} \begin{bmatrix} \hat{Z}_1 \\ \hat{Z}_2 \\ \hat{Z}_3 \end{bmatrix} = -D \begin{bmatrix} \hat{Z}_1 \\ \hat{Z}_2 \\ \hat{Z}_3 \end{bmatrix} - \begin{bmatrix} \alpha_1 & \alpha_4 \\ \alpha_2 & \alpha_5 \\ \alpha_3 & \alpha_6 \end{bmatrix} \begin{bmatrix} OTR \\ -CTR \end{bmatrix} + \begin{bmatrix} 0 \\ \left(\frac{F_{e,s}}{W}\right) S_e \\ 0 \end{bmatrix} \quad \text{Equação 4.36}$$

$$\begin{bmatrix} \hat{X} \\ \hat{S} \\ \hat{A} \end{bmatrix} = \begin{bmatrix} \hat{Z}_1 \\ \hat{Z}_2 \\ \hat{Z}_3 \end{bmatrix} + \begin{bmatrix} \alpha_1 & \alpha_4 \\ \alpha_2 & \alpha_5 \\ \alpha_3 & \alpha_6 \end{bmatrix} \begin{bmatrix} O \\ C \end{bmatrix} \quad \text{Equação 4.37}$$

Uma das principais vantagens deste tipo de observadores é, contrariamente ao OEK, a ausência de parâmetros de sintonização, uma vez que os valores iniciais das variáveis Z podem ser obtidos directamente da Equação 4.19 ou da Equação 4.37 substituindo os valores das variáveis estimadas pelos seus valores experimentais iniciais, normalmente conhecidos. Contudo, nos casos em que são utilizados modelos parciais, os dois regimes são descritos por expressões matemáticas diferentes, sendo necessário o recurso em alternância de dois algoritmos. Para tal, é utilizado um mecanismo de detecção do modelo a usar baseado na estimativa das velocidades específicas de crescimento utilizando para isso o estimador baseado no observador descrito na secção 4.1.2. Os valores estimados de μ_1 , μ_2 e μ_3 foram usados como critério para determinar o regime do processo: respiro-fermentativo se μ_1 e μ_2 forem diferentes de zero e μ_3 for nulo; e respirativo se μ_1 for diferente de zero e μ_2 e μ_3 forem nulos, ou se μ_1 e μ_2 forem nulos e μ_3 diferente de zero, ou ainda se μ_1 e μ_3 forem diferentes de zero e μ_2 for nulo. Deste modo torna-se necessário a sintonização de 6 parâmetros como se poderá inferir a partir do algoritmo que se apresenta de seguida.

Introduzindo a transformação $\psi = K_a^{-1} \xi_a$ e fazendo $\xi_a^T = \xi_1^T = [A \ O \ C]$ obtém-se:

$$K_a^{-1} = \frac{1}{\tilde{k}} \begin{bmatrix} (-k_6 k_{10} + k_7 k_9) & (-k_3 k_{10} - k_4 k_9) & (-k_3 k_7 - k_4 k_6) \\ (k_5 k_{10} - k_7 k_8) & (k_4 k_8) & (k_4 k_5) \\ (-k_5 k_9 + k_6 k_8) & (k_3 k_8) & (k_3 k_5) \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 \\ a_7 & a_8 & a_9 \end{bmatrix} \quad \text{Equação 4.38}$$

$$\tilde{k} = (k_5 k_3 k_{10} + k_5 k_4 k_9 - k_8 k_3 k_7 - k_8 k_4 k_6)$$

em que,

$$\begin{bmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 \\ a_7 & a_8 & a_9 \end{bmatrix} \begin{bmatrix} A \\ O \\ C \end{bmatrix} \quad \text{Equação 4.39}$$

Com estas transformações e usando a Equação 4.24 e a Equação 4.25 é possível escrever as seguintes equações para estimar as velocidades específicas de crescimento μ_1 , μ_2 e μ_3 :

$$\frac{d\hat{\psi}_1}{dt} = \hat{\mu}_1 \hat{X} - D\hat{\psi}_1 + a_2 OTR - a_3 CTR + \omega_{11}(\psi_1 - \hat{\psi}_1) \quad \text{Equação 4.40 (a)}$$

$$\frac{d\hat{\psi}_2}{dt} = \hat{\mu}_2 \hat{X} - D\hat{\psi}_2 + a_5 OTR - a_6 CTR + \omega_{12}(\psi_2 - \hat{\psi}_2) \quad \text{Equação 4.40 (b)}$$

$$\frac{d\hat{\psi}_3}{dt} = \hat{\mu}_3 \hat{X} - D\hat{\psi}_3 + a_8 OTR - a_9 CTR + \omega_{13}(\psi_3 - \hat{\psi}_3) \quad \text{Equação 4.40 (c)}$$

$$\frac{d\hat{\mu}_1}{dt} = \omega_{21}(\psi_1 - \hat{\psi}_1) \quad \text{Equação 4.41 (a)}$$

$$\frac{d\hat{\mu}_2}{dt} = \omega_{22}(\psi_2 - \hat{\psi}_2) \quad \text{Equação 4.41 (b)}$$

$$\frac{d\hat{\mu}_3}{dt} = \omega_{23}(\psi_3 - \hat{\psi}_3) \quad \text{Equação 4.41 (c)}$$

O cálculo dos ganhos ω_{ij} para cada instante t é efectuado pelas equações que se seguem de modo a obter-se uma dinâmica de segunda ordem:

$$\omega_{11,t} = \frac{2\zeta_1}{\tau_1} - \frac{\hat{X}_t - \hat{X}_{t-1}}{T\hat{X}_t} \quad \text{Equação 4.42 (a)}$$

$$\omega_{12,t} = \frac{2\zeta_2}{\tau_2} - \frac{\hat{X}_t - \hat{X}_{t-1}}{T\hat{X}_t} \quad \text{Equação 4.42 (b)}$$

$$\omega_{13,t} = \frac{2\zeta_3}{\tau_3} - \frac{\hat{X}_t - \hat{X}_{t-1}}{T\hat{X}_t} \quad \text{Equação 4.42 (c)}$$

$$\omega_{21,t} = \frac{1}{\tau_1^2 \hat{X}_t} \quad \text{Equação 4.43 (a)}$$

$$\omega_{22,t} = \frac{1}{\tau_2^2 \hat{X}_t} \quad \text{Equação 4.43 (b)}$$

$$\omega_{23,t} = \frac{1}{\tau_3^2 \hat{X}_t} \quad \text{Equação 4.43 (c)}$$

onde T é o passo de integração. Assim, como referido anteriormente, a implementação deste algoritmo requer a sintonização de 6 parâmetros (ζ_1 , ζ_2 , ζ_3 , τ_1 , τ_2 , e τ_3).

O estudo da robustez deste algoritmo foi realizado de modo semelhante ao realizado para o OEK, tendo sido efectuadas simulações numéricas para os casos 1 e 4 constantes na Tabela 4.4. No caso 4 foi necessário estimar as velocidades específicas de crescimento recorrendo ao algoritmo descrito

pela Equação 4.40 à Equação 4.43. Os valores utilizados para os parâmetros τ_1 , τ_2 , e τ_3 foram 0,15 h, 0,20 h e 0,10 h, respectivamente tendo sido os valores de ζ_1 , ζ_2 e ζ_3 iguais a um. Os índices de desempenho para o observador assintótico, calculados pela Equação 4.33, para os casos 1 e 4 e referentes às concentrações de biomassa, de glucose e de acetato, são apresentados na Tabela 4.5.

Tabela 4.5 Índices de desempenho para o OA para os casos 1 e 4 da Tabela 4.4.

Observador	Caso estudado	dif_X	dif_S	dif_A
OA	1	0,196	1,42	---
	4	$1,09 \times 10^{-2}$	0,311	0,389

A Figura 4.5 apresenta os resultados da simulação para o caso 4, onde se pode observar que a concentração de biomassa é bem estimada, mesmo depois da introdução do ruído nas variáveis medidas em linha. Contudo, para este caso foi necessário reduzir o ruído do OTR em 85%, uma vez que o algoritmo mostrou ser sensível ao ruído para esta variável, em especial na estimativa dos valores de μ .

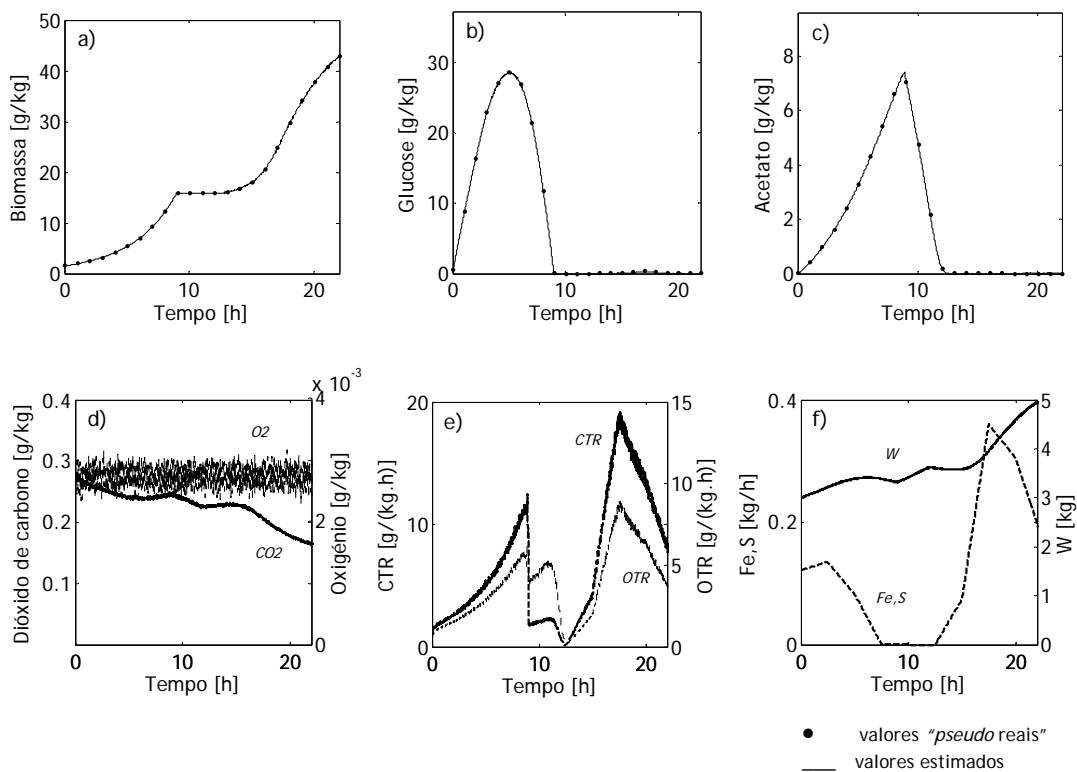


Figura 4.5 Desempenho do OA na estimação da concentração de biomassa, de glucose e de acetato para o caso 4 estudado por simulação. Os gráficos d), e) e f) representam as variáveis medidas em linha.

A validação experimental foi realizada também para o caso 4 utilizando as mesmas experiências como no OEK (Exp_4 e Exp_5). Na Tabela 4.6 são apresentados os índices de desempenho obtidos para as referidas experiências.

Tabela 4.6 Índices de desempenho⁽ⁱ⁾ para o OA para as experiências Exp_4 e Exp_5.

Observador	Experiência	dif_x	dif_s	dif_A
OA	Exp_4	8,02	656	370
	Exp_5	4,42	734	738

(i) Índice de desempenho calculado a partir de 250 pontos.

Na Figura 4.6 estão representados os dados medidos em linha usados para estimar a concentração de biomassa referentes à Exp_5.

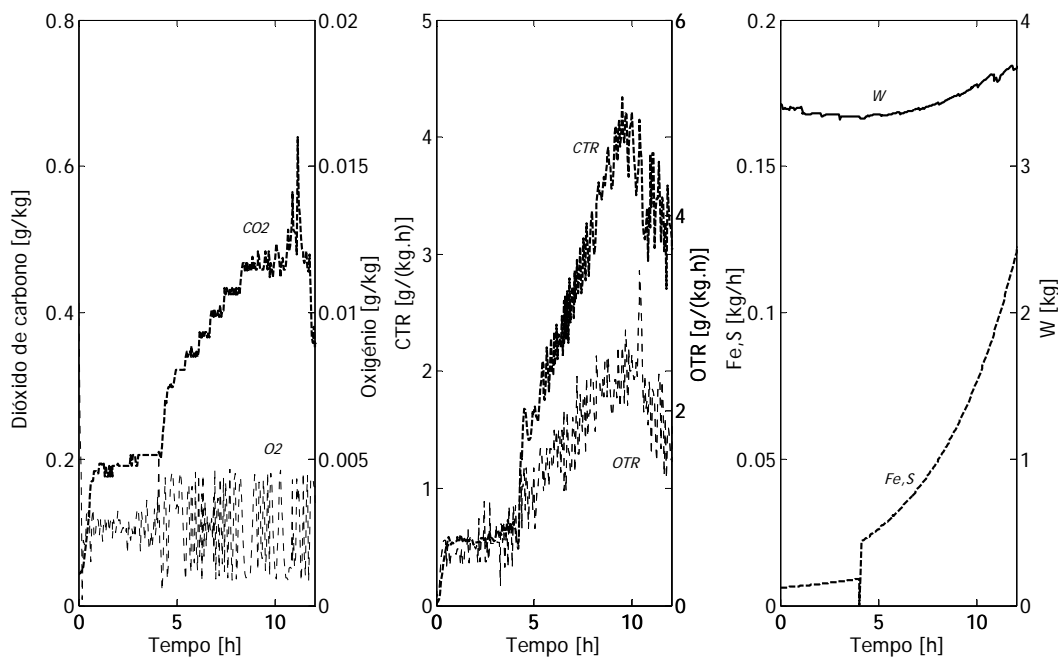


Figura 4.6 Observação de estado numa fermentação semi-contínua de *E. coli* (Exp_5): dados medidos em linha com um intervalo de amostragem de 3 minutos.

Na Figura 4.7 a) são comparados os valores estimados e os valores medidos em diferido para a concentração de biomassa, verificando-se uma concordância satisfatória, principalmente no início da fermentação. O desempenho menos satisfatório observado no final da fermentação, a que corresponde o período de regime respiro-fermentativo, poderá ser devido à maior incerteza associada aos valores dos coeficientes de rendimento desse regime, tal como discutido no Capítulo 3. Por outro lado, na validação experimental, o desempenho deste observador não é tão satisfatório

na estimação das concentrações de glucose e de acetato. Contudo, o intervalo de amostragem utilizado e que conduziu à convergência do observador foi de 3 minutos, o que permite, tal como já referido, a utilização de outros sensores como por exemplo o sistema FIA.

Na Figura 4.7 b) apresentam-se os valores experimentais do CTR obtidos por análise dos gases de saída do reactor por espectrometria de massa e os valores estimados através da expressão $\hat{CTR} = (k_8 \hat{\mu}_1 + k_9 \hat{\mu}_2 + k_{10} \hat{\mu}_3) \hat{X}$. A concordância entre esses valores permite inferir acerca da qualidade da estimativa dos valores de μ .

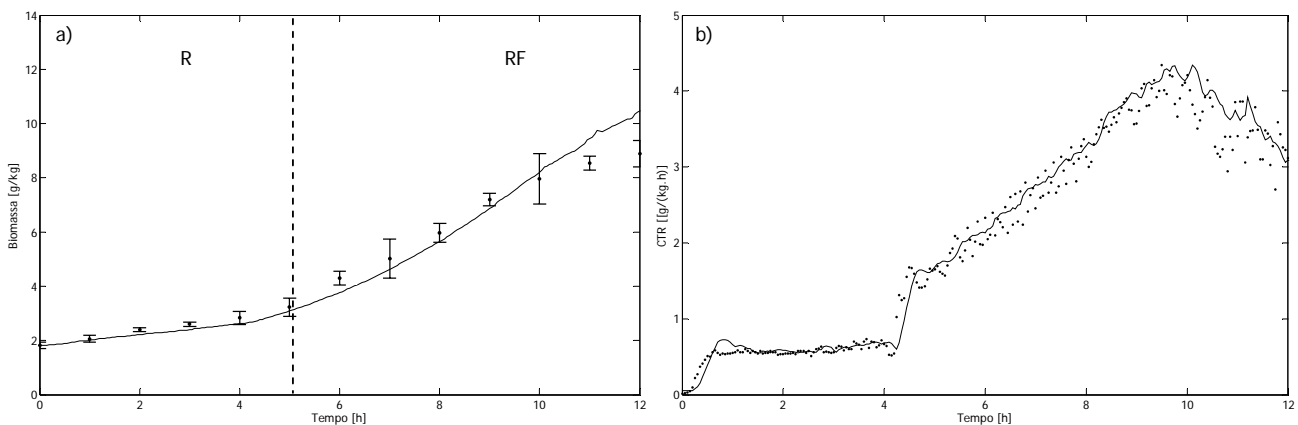


Figura 4.7 Observação de estado numa fermentação semi-contínua de *E. coli* (Exp_5): a) variável estimada: concentração de biomassa; b) variável estimada: CTR (pontos: valores medidos em diferido no caso da biomassa ou valores medidos em linha para o CTR ; linha: valores estimados; RF: regime respiro-fermentativo; R: regime respirativo).

Os resultados obtidos nesta secção e na anterior, permitem concluir que o OEK é menos flexível que o OA tendo em conta as diferentes combinações das variáveis medidas e estimadas como se pode inferir a partir da análise da Tabela 4.1 e da Tabela 4.4. No entanto, o estudo efectuado por simulação numérica mostrou que o desempenho do OEK é ligeiramente superior ao do OA para as mesmas condições (Tabela 4.2 e Tabela 4.5). Por outro lado, o OA mostrou-se menos robusto relativamente a erros experimentais sendo o seu desempenho melhorado quando os dados "pseudo experimentais" são afectados de menor ruído, como anteriormente referido. Esta limitação poderá ser ultrapassada usando filtros numéricos no tratamento dos dados experimentais.

Na validação experimental obteve-se uma concordância satisfatória entre os dados medidos experimentalmente e os estimados, para ambos os observadores embora, o desempenho do AO seja ligeiramente superior. Este facto pode estar relacionado com problemas numéricos e com as dificuldades de convergência do OEK devido à aproximação associada à linearização do modelo, assim como com a necessidade de conhecer com precisão a estrutura do modelo cinético. Verificou-se ainda que o OEK necessita de intervalos de amostragem inferiores ao OA (0,3 e 3 minutos,

respectivamente) o que poderá acarretar dificuldades na sua implementação prática. Mais ainda, o OEK apresenta uma dificuldade extra quando comparado com o OA, uma vez que necessita da sintonização por tentativa erro de 15 parâmetros, enquanto que os OA não requerem a sintonização de qualquer parâmetro, quando se utiliza o modelo global, ou apenas a sintonização de 6 parâmetros quando se utiliza o modelo parcial.

4.2.3. Observadores por Intervalo

Nesta secção pretende-se introduzir uma solução alternativa para a monitorização avançada de bioprocessos em casos onde a estrutura do modelo cinético é desconhecida e a concentração do substrato na alimentação e as concentrações iniciais de biomassa e de substrato, não são conhecidas com precisão, sendo uma situação que acontece frequentemente em várias aplicações práticas.

Assim sendo, a fermentação semi-contínua de alta densidade de *E. coli* é estudada em termos da aplicabilidade de observadores por intervalo na estimativa da concentração de biomassa, através das medições em linha das concentrações de acetato, oxigénio dissolvido e dióxido de carbono dissolvido. Admite-se ainda que D , W , OTR , CTR e $F_{e,s}$ são conhecidos em linha.

O desempenho do observador por intervalo foi avaliado por simulação, sendo o seu desenho baseado nas propriedades de cooperatividade da dinâmica do erro de observação (Equação 4.20). Por outro lado, considera-se que são conhecidos os limites inferior e superior da concentração de substrato na alimentação bem como das concentrações iniciais de biomassa e de substrato, tal que: $S_e^- \leq S_e(t) \leq S_e^+$; $X_0^- \leq X_0(t) \leq X_0^+$ e $S_0^- \leq S_0(t) \leq S_0^+$.

Convém referir que não é possível desenhar um observador por intervalo para \hat{X} directamente a partir de um observador do modelo dinâmico dado pela Equação 3.55. Com efeito, o observador (sendo g_1 a g_{15} os ganhos do observador) obtido directamente a partir do referido modelo dinâmico é dado por:

$$\frac{d\hat{X}}{dt} = (\mu_1 + \mu_2 + \mu_3)\hat{X} - D\hat{X} + g_1(A - \hat{A}) + g_2(O - \hat{O}) + g_3(C - \hat{C}) \quad \text{Equação 4.44 (a)}$$

$$\frac{d\hat{S}}{dt} = (-k_1\mu_1 - k_2\mu_2)\hat{X} - D\hat{S} + \frac{F_{e,s}}{W}S_e + g_4(A - \hat{A}) + g_5(O - \hat{O}) + g_6(C - \hat{C}) \quad \text{Equação 4.44 (b)}$$

$$\frac{d\hat{A}}{dt} = (k_3\mu_2 - k_4\mu_3)\hat{X} - D\hat{A} + g_7(A - \hat{A}) + g_8(O - \hat{O}) + g_9(C - \hat{C}) \quad \text{Equação 4.44 (c)}$$

$$\frac{d\hat{O}}{dt} = (-k_5\mu_1 - k_6\mu_2 - k_7\mu_3)\hat{X} - D\hat{O} + OTR + g_{10}(A - \hat{A}) + g_{11}(O - \hat{O}) + g_{12}(C - \hat{C}) \quad \text{Equação 4.44 (d)}$$

$$\frac{d\hat{C}}{dt} = (k_8\mu_1 + k_9\mu_2 + k_{10}\mu_3)\hat{X} - D\hat{C} - CTR + g_{13}(A - \hat{A}) + g_{14}(O - \hat{O}) + g_{15}(C - \hat{C}) \quad \text{Equação 4.44 (e)}$$

Este observador conduz a uma matriz Jacobiano (Equação 4.45) que não satisfaz a condição de cooperatividade dada pela Equação 4.20, uma vez que o elemento da não diagonal $(-k\mu)$ é negativo e os elementos $\left(\pm K \frac{\partial \mu}{\partial S} X\right)$ poderão ser, nalguns casos, também negativos:

$$J(\hat{X}, \hat{S}, \hat{A}, \hat{O}, \hat{C}) = \begin{bmatrix} (\mu - D) & \left(\frac{\partial \mu}{\partial S} X\right) & \left(\frac{\partial \mu}{\partial A} X + g_1\right) & g_2 & g_3 \\ (-K\mu) & \left(-K \frac{\partial \mu}{\partial S} X - D\right) & \left(-K \frac{\partial \mu}{\partial A} X + g_4\right) & g_5 & g_6 \\ (K\mu) & \left(K \frac{\partial \mu}{\partial S} X\right) & \left(K \frac{\partial \mu}{\partial A} X - D + g_7\right) & g_8 & g_9 \\ (-K\mu) & \left(-K \frac{\partial \mu}{\partial S} X\right) & \left(-K \frac{\partial \mu}{\partial A} X + g_{10}\right) & -D + g_{11} & g_{12} \\ (K\mu) & \left(K \frac{\partial \mu}{\partial S} X\right) & \left(K \frac{\partial \mu}{\partial A} X + g_{13}\right) & g_{14} & -D + g_{15} \end{bmatrix} \quad \text{Equação 4.45}$$

Contudo, é de realçar que a noção de cooperatividade é dependente de coordenadas, pelo que pode ser conseguida pela introdução da mesma transformação de estado descrita na secção 4.1.1.2, para os observadores assintóticos (Equação 4.16). Recorrendo à mesma abordagem realizada nessa secção e utilizando a Equação 4.14, que reescreve o modelo dinâmico geral, procedeu-se ao projecto de três observadores por intervalo que a seguir se apresentam. Nesta abordagem admite-se que os coeficientes de rendimento são conhecidos e que o modelo cinético é desconhecido.

▪ **Observador por intervalo simples**

Como referido anteriormente, pretende-se obter um observador por intervalo capaz de estimar a concentração de biomassa, \hat{X} , através da medição em linha das concentrações de acetato, oxigénio e dióxido de carbono dissolvidos.

A matriz usada na transformação de estado da Equação 4.16, que permite a mudança linear de coordenadas é, neste caso particular:

$$K_2 K_1^{-1} = \begin{bmatrix} 1 & 1 & 1 \\ -k_1 & -k_2 & 0 \end{bmatrix} \begin{bmatrix} 0 & k_3 & -k_4 \\ -k_5 & -k_6 & -k_7 \\ k_8 & k_9 & k_{10} \end{bmatrix}^{-1} = \begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 \\ \alpha_4 & \alpha_5 & \alpha_6 \end{bmatrix} \quad \text{Equação 4.46}$$

O observador correspondente é dado pelas seguintes equações:

$$\frac{d}{dt} \begin{bmatrix} \hat{Z}_1 \\ \hat{Z}_2 \end{bmatrix} = -D \begin{bmatrix} \hat{Z}_1 \\ \hat{Z}_2 \end{bmatrix} - \begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 \\ \alpha_4 & \alpha_5 & \alpha_6 \end{bmatrix} \begin{bmatrix} 0 \\ OTR \\ -CTR \end{bmatrix} + \begin{bmatrix} 0 \\ F_{e,s} \\ W \end{bmatrix} S_e \quad \text{Equação 4.47 (a)}$$

$$\begin{bmatrix} \hat{X} \\ \hat{S} \end{bmatrix} = \begin{bmatrix} \hat{Z}_1 \\ \hat{Z}_2 \end{bmatrix} + \begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 \\ \alpha_4 & \alpha_5 & \alpha_6 \end{bmatrix} \begin{bmatrix} A \\ O \\ C \end{bmatrix} \quad \text{Equação 4.47 (b)}$$

Para verificar se o observador satisfaz as propriedades de cooperatividade, definiram-se inicialmente os erros do observador como: $e_{z_1} = \hat{Z}_1 - Z_1$, $e_{z_2} = \hat{Z}_2 - Z_2$, $e_x = \hat{X} - X$ e $e_s = \hat{S} - S$, cuja dinâmica é dada pela seguinte equação:

$$\frac{d}{dt} \begin{bmatrix} e_{z_1} \\ e_{z_2} \\ e_x \\ e_s \end{bmatrix} = -D \begin{bmatrix} e_{z_1} \\ e_{z_2} \\ e_x \\ e_s \end{bmatrix} \quad \text{Equação 4.48}$$

à qual corresponde a seguinte matriz Jacobiano:

$$J(e_{z_1}, e_{z_2}, e_x, e_s) = \begin{bmatrix} -D & 0 & 0 & 0 \\ 0 & -D & 0 & 0 \\ 0 & 0 & -D & 0 \\ 0 & 0 & 0 & -D \end{bmatrix} \quad \text{Equação 4.49}$$

A observação da matriz anterior permite verificar facilmente que o sistema é cooperativo sendo possível desenhar um observador por intervalo. Considerando os limites inferior e superior dos valores iniciais das concentrações de biomassa e substrato, pode-se definir o seguinte conjunto de equações para o observador por intervalo:

$$\frac{d}{dt} \begin{bmatrix} \hat{Z}_1^+ \\ \hat{Z}_2^+ \end{bmatrix} = -D \begin{bmatrix} \hat{Z}_1^+ \\ \hat{Z}_2^+ \end{bmatrix} - \begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 \\ \alpha_4 & \alpha_5 & \alpha_6 \end{bmatrix} \begin{bmatrix} 0 \\ OTR \\ -CTR \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{F_{e,s}}{W} S_e^+ \end{bmatrix} \quad \text{Equação 4.50 (a)}$$

$$\begin{bmatrix} \hat{X}^+ \\ \hat{S}^+ \end{bmatrix} = \begin{bmatrix} \hat{Z}_1^+ \\ \hat{Z}_2^+ \end{bmatrix} + \begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 \\ \alpha_4 & \alpha_5 & \alpha_6 \end{bmatrix} \begin{bmatrix} A \\ O \\ C \end{bmatrix} \quad \text{Equação 4.50 (b)}$$

$$\frac{d}{dt} \begin{bmatrix} \hat{Z}_1^- \\ \hat{Z}_2^- \end{bmatrix} = -D \begin{bmatrix} \hat{Z}_1^- \\ \hat{Z}_2^- \end{bmatrix} - \begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 \\ \alpha_4 & \alpha_5 & \alpha_6 \end{bmatrix} \begin{bmatrix} 0 \\ OTR \\ -CTR \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{F_{e,s}}{W} S_e^- \end{bmatrix} \quad \text{Equação 4.50 (c)}$$

$$\begin{bmatrix} \hat{X}^- \\ \hat{S}^- \end{bmatrix} = \begin{bmatrix} \hat{Z}_1^- \\ \hat{Z}_2^- \end{bmatrix} + \begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 \\ \alpha_4 & \alpha_5 & \alpha_6 \end{bmatrix} \begin{bmatrix} A \\ O \\ C \end{bmatrix} \quad \text{Equação 4.50 (d)}$$

▪ **Observador por intervalo com ganhos**

Neste ponto apresenta-se um observador por intervalo semelhante ao caso anterior, mas no qual é introduzido um termo correctivo com ganho, com o intuito de facilitar a sua convergência.

Considerando a Equação 4.16 sabe-se que:

$$\xi_1 = (K_2 K_1^{-1})^{-1} (\xi_2 - Z) \quad \text{Equação 4.51}$$

Neste caso particular:

$$\begin{bmatrix} A \\ O \\ C \end{bmatrix} = \begin{bmatrix} \bar{\alpha}_1 & \bar{\alpha}_4 \\ \bar{\alpha}_2 & \bar{\alpha}_5 \\ \bar{\alpha}_3 & \bar{\alpha}_6 \end{bmatrix} \begin{bmatrix} X - Z_1 \\ S - Z_2 \end{bmatrix} \text{ em que } (K_2 K_1^{-1})^{-1} = \begin{bmatrix} \bar{\alpha}_1 & \bar{\alpha}_4 \\ \bar{\alpha}_2 & \bar{\alpha}_5 \\ \bar{\alpha}_3 & \bar{\alpha}_6 \end{bmatrix} \quad \text{Equação 4.52}$$

Assim, o observador por intervalo com ganhos é dado por:

$$\frac{d}{dt} \begin{bmatrix} \hat{Z}_1 \\ \hat{Z}_2 \end{bmatrix} = -D \begin{bmatrix} \hat{Z}_1 \\ \hat{Z}_2 \end{bmatrix} - \begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 \\ \alpha_4 & \alpha_5 & \alpha_6 \end{bmatrix} \begin{bmatrix} 0 \\ OTR \\ -CTR \end{bmatrix} + \begin{bmatrix} 0 \\ F_{e,s} \\ W \end{bmatrix} S_e + \begin{bmatrix} g_1 & g_2 & g_3 \\ g_4 & g_5 & g_6 \end{bmatrix} \begin{bmatrix} A - \hat{A} \\ O - \hat{O} \\ C - \hat{C} \end{bmatrix} \quad \text{Equação 4.53 (a)}$$

$$\text{com } \begin{bmatrix} \hat{A} \\ \hat{O} \\ \hat{C} \end{bmatrix} = \begin{bmatrix} \bar{\alpha}_1 & \bar{\alpha}_4 \\ \bar{\alpha}_2 & \bar{\alpha}_5 \\ \bar{\alpha}_3 & \bar{\alpha}_6 \end{bmatrix} \begin{bmatrix} \hat{X} - \hat{Z}_1 \\ \hat{S} - \hat{Z}_2 \end{bmatrix}$$

$$\begin{bmatrix} \hat{X} \\ \hat{S} \end{bmatrix} = \begin{bmatrix} \hat{Z}_1 \\ \hat{Z}_2 \end{bmatrix} + \begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 \\ \alpha_4 & \alpha_5 & \alpha_6 \end{bmatrix} \begin{bmatrix} A \\ O \\ C \end{bmatrix} \quad \text{Equação 4.53 (b)}$$

A dinâmica dos erros do observador é agora dada por:

$$\frac{d}{dt} \begin{bmatrix} e_{z_1} \\ e_{z_2} \\ e_x \\ e_s \end{bmatrix} = -D \begin{bmatrix} e_{z_1} \\ e_{z_2} \\ e_x \\ e_s \end{bmatrix} + \begin{bmatrix} g_1 & g_2 & g_3 \\ g_4 & g_5 & g_6 \end{bmatrix} \begin{bmatrix} \bar{\alpha}_1 (e_{z_1} - e_x) + \bar{\alpha}_4 (e_{z_2} - e_s) \\ \bar{\alpha}_2 (e_{z_1} - e_x) + \bar{\alpha}_5 (e_{z_2} - e_s) \\ \bar{\alpha}_3 (e_{z_1} - e_x) + \bar{\alpha}_6 (e_{z_2} - e_s) \end{bmatrix} \quad \text{Equação 4.54}$$

à qual corresponde a seguinte matriz Jacobiano:

$$J(e_{z_1}, e_{z_2}, e_x, e_s) = \begin{bmatrix} (-D + g_1\bar{\alpha}_1 + g_2\bar{\alpha}_2 + g_3\bar{\alpha}_3) & (g_1\bar{\alpha}_4 + g_2\bar{\alpha}_5 + g_3\bar{\alpha}_6) & (-g_1\bar{\alpha}_1 - g_2\bar{\alpha}_2 - g_3\bar{\alpha}_3) & (-g_1\bar{\alpha}_4 - g_2\bar{\alpha}_5 - g_3\bar{\alpha}_6) \\ (g_4\bar{\alpha}_1 + g_5\bar{\alpha}_2 + g_6\bar{\alpha}_3) & (-D + g_4\bar{\alpha}_4 + g_5\bar{\alpha}_5 + g_6\bar{\alpha}_6) & (-g_4\bar{\alpha}_1 - g_5\bar{\alpha}_2 - g_6\bar{\alpha}_3) & (-g_4\bar{\alpha}_4 - g_5\bar{\alpha}_5 - g_6\bar{\alpha}_6) \\ 0 & 0 & -D & 0 \\ 0 & 0 & 0 & -D \end{bmatrix}$$

Equação 4.55

Atendendo à matriz obtida, verifica-se que o sistema só é cooperativo se:

$$\begin{cases} g_1\bar{\alpha}_4 + g_2\bar{\alpha}_5 + g_3\bar{\alpha}_6 = 0 \\ g_4\bar{\alpha}_1 + g_5\bar{\alpha}_2 + g_6\bar{\alpha}_3 = 0 \\ -g_1\bar{\alpha}_1 - g_2\bar{\alpha}_2 - g_3\bar{\alpha}_3 \geq 0 \\ -g_4\bar{\alpha}_4 - g_5\bar{\alpha}_5 - g_6\bar{\alpha}_6 \geq 0 \end{cases}$$

Deste modo, os valores de g_1 , g_2 , g_3 , g_4 , g_5 e g_6 deverão ser tais que permitam satisfazer simultaneamente aquelas quatro condições.

Sendo assim, pode-se definir o seguinte conjunto de equações para o referido observador por intervalo:

$$\frac{d}{dt} \begin{bmatrix} \hat{z}_1^+ \\ \hat{z}_2^+ \end{bmatrix} = -D \begin{bmatrix} \hat{z}_1^+ \\ \hat{z}_2^+ \end{bmatrix} - \begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 \\ \alpha_4 & \alpha_5 & \alpha_6 \end{bmatrix} \begin{bmatrix} 0 \\ OTR \\ -CTR \end{bmatrix} + \begin{bmatrix} 0 \\ F_{e,s} \\ W \end{bmatrix} S_e^+ + \begin{bmatrix} g_1 & g_2 & g_3 \\ g_4 & g_5 & g_6 \end{bmatrix} \begin{bmatrix} A - \hat{A}^+ \\ O - \hat{O}^+ \\ C - \hat{C}^+ \end{bmatrix} \quad \text{Equação 4.56 (a)}$$

$$\text{com } \begin{bmatrix} \hat{A}^+ \\ \hat{O}^+ \\ \hat{C}^+ \end{bmatrix} = \begin{bmatrix} \bar{\alpha}_1 & \bar{\alpha}_4 \\ \bar{\alpha}_2 & \bar{\alpha}_5 \\ \bar{\alpha}_3 & \bar{\alpha}_6 \end{bmatrix} \begin{bmatrix} \hat{X}^+ - \hat{z}_1^+ \\ \hat{S}^+ - \hat{z}_2^+ \end{bmatrix}$$

$$\begin{bmatrix} \hat{X}^+ \\ \hat{S}^+ \end{bmatrix} = \begin{bmatrix} \hat{z}_1^+ \\ \hat{z}_2^+ \end{bmatrix} + \begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 \\ \alpha_4 & \alpha_5 & \alpha_6 \end{bmatrix} \begin{bmatrix} A \\ O \\ C \end{bmatrix} \quad \text{Equação 4.56 (b)}$$

$$\frac{d}{dt} \begin{bmatrix} \hat{z}_1^- \\ \hat{z}_2^- \end{bmatrix} = -D \begin{bmatrix} \hat{z}_1^- \\ \hat{z}_2^- \end{bmatrix} - \begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 \\ \alpha_4 & \alpha_5 & \alpha_6 \end{bmatrix} \begin{bmatrix} 0 \\ OTR \\ -CTR \end{bmatrix} + \begin{bmatrix} 0 \\ F_{e,s} \\ W \end{bmatrix} S_e^- + \begin{bmatrix} g_1 & g_2 & g_3 \\ g_4 & g_5 & g_6 \end{bmatrix} \begin{bmatrix} A - \hat{A}^- \\ O - \hat{O}^- \\ C - \hat{C}^- \end{bmatrix} \quad \text{Equação 4.56 (c)}$$

$$\text{com } \begin{bmatrix} \hat{A}^- \\ \hat{O}^- \\ \hat{C}^- \end{bmatrix} = \begin{bmatrix} \bar{\alpha}_1 & \bar{\alpha}_4 \\ \bar{\alpha}_2 & \bar{\alpha}_5 \\ \bar{\alpha}_3 & \bar{\alpha}_6 \end{bmatrix} \begin{bmatrix} \hat{X}^- - \hat{z}_1^- \\ \hat{S}^- - \hat{z}_2^- \end{bmatrix}$$

$$\begin{bmatrix} \hat{X}^- \\ \hat{S}^- \end{bmatrix} = \begin{bmatrix} \hat{Z}_1^- \\ \hat{Z}_2^- \end{bmatrix} + \begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 \\ \alpha_4 & \alpha_5 & \alpha_6 \end{bmatrix} \begin{bmatrix} A \\ O \\ C \end{bmatrix} \quad \text{Equação 4.56 (d)}$$

▪ **Observador por intervalo com ganhos através do OTR**

Alternativamente à introdução de ganhos através das variáveis de estado A , O e C projectou-se um novo observador por intervalo com ganhos, sendo estes introduzidos através da estimativa do OTR , obtendo-se o seguinte observador:

$$\frac{d}{dt} \begin{bmatrix} \hat{Z}_1 \\ \hat{Z}_2 \end{bmatrix} = -D \begin{bmatrix} \hat{Z}_1 \\ \hat{Z}_2 \end{bmatrix} - \begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 \\ \alpha_4 & \alpha_5 & \alpha_6 \end{bmatrix} \begin{bmatrix} 0 \\ OTR \\ -CTR \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{F_{e,s}}{W} S_{e,s} \end{bmatrix} + \begin{bmatrix} g_1 \\ g_2 \end{bmatrix} [OTR - \hat{OTR}] \quad \text{Equação 4.57 (a)}$$

$$\text{com } \hat{OTR} = (k_5 \hat{\mu}_1 + k_6 \hat{\mu}_2 + k_7 \hat{\mu}_3) \hat{X}$$

$$\begin{bmatrix} \hat{X} \\ \hat{S} \end{bmatrix} = \begin{bmatrix} \hat{Z}_1 \\ \hat{Z}_2 \end{bmatrix} + \begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 \\ \alpha_4 & \alpha_5 & \alpha_6 \end{bmatrix} \begin{bmatrix} A \\ O \\ C \end{bmatrix} \quad \text{Equação 4.57 (b)}$$

Neste caso a dinâmica dos erros pode ser determinada por:

$$\frac{d}{dt} \begin{bmatrix} e_{z_1} \\ e_{z_2} \\ e_x \\ e_s \end{bmatrix} = -D \begin{bmatrix} e_{z_1} \\ e_{z_2} \\ e_x \\ e_s \end{bmatrix} - \begin{bmatrix} g_1(k_5 \hat{\mu}_1 + k_6 \hat{\mu}_2 + k_7 \hat{\mu}_3) \\ g_2(k_5 \hat{\mu}_1 + k_6 \hat{\mu}_2 + k_7 \hat{\mu}_3) \\ 0 \\ 0 \end{bmatrix} e_x \quad \text{Equação 4.58}$$

com a seguinte matriz Jacobiano:

$$J(e_{z_1}, e_{z_2}, e_x, e_s) = \begin{bmatrix} -D & 0 & -g_1(k_5 \hat{\mu}_1 + k_6 \hat{\mu}_2 + k_7 \hat{\mu}_3) & 0 \\ 0 & -D & -g_2(k_5 \hat{\mu}_1 + k_6 \hat{\mu}_2 + k_7 \hat{\mu}_3) & 0 \\ 0 & 0 & -D & 0 \\ 0 & 0 & 0 & -D \end{bmatrix} \quad \text{Equação 4.59}$$

Verifica-se que o sistema só é cooperativo se g_1 e g_2 forem negativos ou nulos. No caso de serem nulos o observador obtido resume-se ao primeiro observador por intervalo apresentado, isto é, ao observador simples (sem ganhos). Quando os valores dos ganhos são negativos obtém-se o seguinte observador:

$$\frac{d}{dt} \begin{bmatrix} \hat{Z}_1^+ \\ \hat{Z}_2^+ \end{bmatrix} = -D \begin{bmatrix} \hat{Z}_1^+ \\ \hat{Z}_2^+ \end{bmatrix} - \begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 \\ \alpha_4 & \alpha_5 & \alpha_6 \end{bmatrix} \begin{bmatrix} 0 \\ OTR \\ -CTR \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{F_{e,s}}{W} S_{e,s}^+ \end{bmatrix} + \begin{bmatrix} g_1 \\ g_2 \end{bmatrix} [OTR - (k_5 \hat{\mu}_1^+ + k_6 \hat{\mu}_2^+ + k_7 \hat{\mu}_3^+) \hat{X}^+]$$

Equação 4.60 (a)

$$\begin{bmatrix} \hat{X}^+ \\ \hat{S}^+ \end{bmatrix} = \begin{bmatrix} \hat{Z}_1^+ \\ \hat{Z}_2^+ \end{bmatrix} + \begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 \\ \alpha_4 & \alpha_5 & \alpha_6 \end{bmatrix} \begin{bmatrix} A \\ O \\ C \end{bmatrix}$$

Equação 4.60 (b)

$$\frac{d}{dt} \begin{bmatrix} \hat{Z}_1^- \\ \hat{Z}_2^- \end{bmatrix} = -D \begin{bmatrix} \hat{Z}_1^- \\ \hat{Z}_2^- \end{bmatrix} - \begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 \\ \alpha_4 & \alpha_5 & \alpha_6 \end{bmatrix} \begin{bmatrix} 0 \\ OTR \\ -CTR \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{F_{e,s}}{W} S_{e,s}^- \end{bmatrix} + \begin{bmatrix} g_1 \\ g_2 \end{bmatrix} [OTR - (k_5 \hat{\mu}_1^- + k_6 \hat{\mu}_2^- + k_7 \hat{\mu}_3^-) \hat{X}^-]$$

Equação 4.60 (c)

$$\begin{bmatrix} \hat{X}^- \\ \hat{S}^- \end{bmatrix} = \begin{bmatrix} \hat{Z}_1^- \\ \hat{Z}_2^- \end{bmatrix} + \begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 \\ \alpha_4 & \alpha_5 & \alpha_6 \end{bmatrix} \begin{bmatrix} A \\ O \\ C \end{bmatrix}$$

Equação 4.60 (d)

Para obter a estimativa das velocidades específicas de crescimento μ_1 , μ_2 e μ_3 utilizou-se o estimador baseado no observador descrito na secção 4.2.2. Contudo, sendo um observador por intervalo, é necessário ter em conta os limites inferior e superior no seu cálculo. Assim sendo, a Equação 4.40 à Equação 4.43 dão origem às seguintes equações:

$$\frac{d\hat{\psi}_1^\pm}{dt} = \hat{\mu}_1^\pm \hat{X}^\pm - D\hat{\psi}_1^\pm + a_2 OTR - a_3 CTR + \omega_{11}^\pm (\psi_1 - \hat{\psi}_1^\pm)$$

Equação 4.61 (a)

$$\frac{d\hat{\psi}_2^\pm}{dt} = \hat{\mu}_2^\pm \hat{X}^\pm - D\hat{\psi}_2^\pm + a_5 OTR - a_6 CTR + \omega_{12}^\pm (\psi_2 - \hat{\psi}_2^\pm)$$

Equação 4.61 (b)

$$\frac{d\hat{\psi}_3^\pm}{dt} = \hat{\mu}_3^\pm \hat{X}^\pm - D\hat{\psi}_3^\pm + a_8 OTR - a_9 CTR + \omega_{13}^\pm (\psi_3 - \hat{\psi}_3^\pm)$$

Equação 4.61 (c)

$$\frac{d\hat{\mu}_1^\pm}{dt} = \omega_{21}^\pm (\psi_1 - \hat{\psi}_1^\pm)$$

Equação 4.62 (a)

$$\frac{d\hat{\mu}_2^\pm}{dt} = \omega_{22}^\pm (\psi_2 - \hat{\psi}_2^\pm)$$

Equação 4.62 (b)

$$\frac{d\hat{\mu}_3^\pm}{dt} = \omega_{23}^\pm (\psi_3 - \hat{\psi}_3^\pm)$$

Equação 4.62 (c)

$$\omega_{11,t}^\pm = \frac{2\zeta_1}{\tau_1} - \frac{\hat{X}_t^\pm - \hat{X}_{t-1}^\pm}{T\hat{X}_t^\pm}$$

Equação 4.63 (a)

$$\omega_{12,t}^{\pm} = \frac{2\zeta_2}{\tau_2} - \frac{\hat{X}_t^{\pm} - \hat{X}_{t-1}^{\pm}}{T\hat{X}_t^{\pm}} \quad \text{Equação 4.63 (b)}$$

$$\omega_{13,t}^{\pm} = \frac{2\zeta_3}{\tau_3} - \frac{\hat{X}_t^{\pm} - \hat{X}_{t-1}^{\pm}}{T\hat{X}_t^{\pm}} \quad \text{Equação 4.63 (c)}$$

$$\omega_{21,t}^{\pm} = \frac{1}{\tau_1^2 \hat{X}_t^{\pm}} \quad \text{Equação 4.64 (a)}$$

$$\omega_{22,t}^{\pm} = \frac{1}{\tau_2^2 \hat{X}_t^{\pm}} \quad \text{Equação 4.64 (b)}$$

$$\omega_{23,t}^{\pm} = \frac{1}{\tau_3^2 \hat{X}_t^{\pm}} \quad \text{Equação 4.64 (c)}$$

Qualquer um dos três observadores por intervalo propostos permite estimar os limites inferiores e superiores das concentrações de biomassa e de substrato, X e S , respectivamente: X^* , X^* e S^* , S^* .

Um ponto-chave associado ao projecto dos observadores por intervalo diz respeito às propriedades de cooperatividade da dinâmica do erro do observador. Sendo assim, e tal como já referido, neste caso foi necessário proceder a uma transformação de estado apropriada capaz de garantir as condições para que o sistema fosse cooperativo.

Neste estudo considerou-se que a concentração da glucose na alimentação S_e é desconhecida, embora se assuma que é conhecida a variação no tempo dos seus limites inferior e superior, bem como o limite inferior e superior da concentração inicial de biomassa. O desempenho dos 3 tipos de observadores por intervalo propostos foi deste modo testado considerando incertezas de $\pm 50\%$ na concentração de glucose na alimentação, relativamente ao seu valor nominal ($S_e = 250$ g/kg). Para a concentração inicial de biomassa considerou-se também incertezas de $\pm 50\%$ face ao seu valor nominal ($X_0 = 1,5$ g/kg).

No estudo realizado foram ainda fixados os seguintes valores de ganhos, para os dois últimos observadores propostos:

- Observador por intervalo com ganhos: $g_1 = g_3 = g_4 = g_6 = 0,01$; $g_2 = -0,1$; $g_5 = 0,5$.
- Observador por intervalo com ganhos através do *OTR*: $g_1 = g_2 = -0,01$.

Os resultados obtidos são apresentados na Figura 4.8. Como se pode inferir por observação desta figura, os resultados são satisfatórios sendo de realçar que o melhor desempenho foi o do observador por intervalo com ganhos através do *OTR*, em que os valores "pseudo reais" permanecem sempre entre os valores do limite inferior e superior estimados.

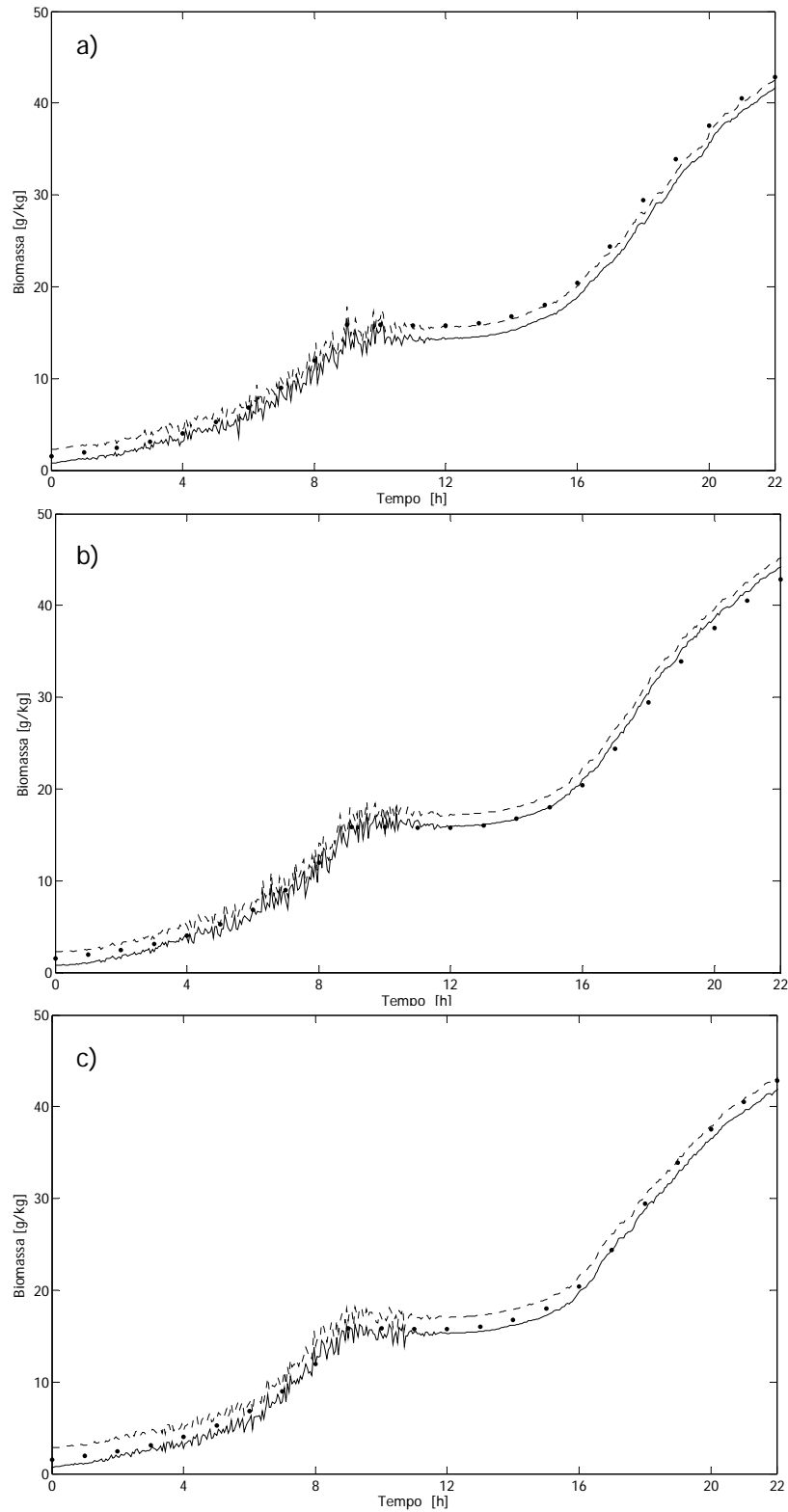


Figura 4.8 Observador por intervalo simulado com incerteza associada à concentração de glucose no caudal de alimentação. (a) observador por intervalo simples; (b) observador por intervalo com ganhos; (c) observador por intervalo com ganhos através do *OTR*.

Mais ainda, embora não seja muito evidente, a simulação obtida com este observador parece diminuir ligeiramente o ruído comparativamente aos outros dois observadores por intervalo tidos em conta. Este facto foi também observado por Rapaport e Dochain (2005). Relativamente aos dois observadores por intervalo com ganhos aqui analisados, é clara a vantagem de usar o último, atendendo não só aos resultados alcançados mas também pelo facto de envolver um menor número de ganhos a fixar.

Os observadores por intervalo propostos apresentam, tal como os observadores assintóticos, a vantagem de não exigirem o conhecimento do modelo cinético do processo em estudo, e relativamente àqueles apresentarem ainda uma vantagem adicional, uma vez que conseguem estimar o intervalo de variação relativamente pequeno mesmo considerando incertezas significativas em algumas variáveis, para além do ruído nas variáveis de estado medidas em linha, utilizando um tempo de amostragem semelhante (3 minutos). Este facto reveste-se de grande importância em situações práticas para as quais a composição do caudal de alimentação pode variar consideravelmente com o tempo, como por exemplo no caso de fermentações onde se utiliza melaço como meio de cultura. Mais ainda, atendendo a que nas fermentações efectuadas neste trabalho se usou como alimentação uma solução concentrada de glucose e que estas tinham uma duração superior a 10 horas, a glucose adicionada poderá variar com o tempo. Por outro lado, durante a esterilização da solução poderá haver evaporação concentrando-a ainda mais. Os valores iniciais das concentrações de biomassa e substrato não são conhecidos com precisão, visto que durante as análises em diferido, que permitem conhecer os seus valores, continua a haver crescimento e consumo de substrato.

Os princípios dos observadores por intervalo podem também ser aplicados de modo a ultrapassar e considerar as incertezas nos valores dos coeficientes de rendimento, o que poderá constituir uma mais valia, permitindo melhorar a robustez do algoritmo quando existem incertezas nos valores desses coeficientes.

4.3. SÍNTESE

Hoje em dia os bioprocessos adquiriram um papel preponderante na produção de substâncias de elevado valor acrescentado na indústria farmacêutica (proteínas recombinadas, vacinas, etc.), na indústria agro-alimentar (leveduras, cerveja, vinho, etc.) e no tratamento de sólidos orgânicos assim como no tratamento de águas residuais quer urbanas, quer industriais.

A monitorização desses processos é fundamental para o seu desempenho, sendo por isso importante conseguir medir a concentração dos principais compostos envolvidos como a concentração da biomassa, do substrato e de produtos finais. Uma alternativa interessante à medição de todas essas variáveis é aquela que explora o uso de um modelo matemático em

conjunto com um número mais reduzido de variáveis medidas, ou seja a utilização de sensores por programação. Assim sendo, pode ser interessante usar estes algoritmos para estimar em linha e em contínuo a concentração quer das variáveis não medidas, quer das variáveis medidas mas de uma forma discreta no tempo. Com efeito, no caso em estudo, fermentação semi-contínua de *E.coli*, variáveis como a concentração de biomassa, são em geral, obtidas por análises laboratoriais em diferido, tornando-as de uso limitado no que diz respeito a estratégias de controlo. No entanto, esta variável pode ser obtida em linha aplicando sensores por programação.

Neste capítulo foram apresentados sensores por programação para a estimação de variáveis de estado e de parâmetros (cinéticas de reacção) num processo biotecnológico. Estes algoritmos baseiam-se na estrutura do modelo geral de reactores biológicos (Bastin e Dochain, 1990).

Neste trabalho foram aplicados três observadores - observador estendido de Kalman, observador assintótico e observador por intervalo - para estimar a concentração de biomassa, tendo sido avaliado o seu desempenho bem como a sua flexibilidade. De um modo geral os algoritmos apresentados apenas requerem a medição em linha das concentrações de oxigénio e de dióxido de carbono dissolvidos juntamente com as taxas de transferência gasosa, que representam medições típicas quer à escala laboratorial quer à escala industrial. Mais ainda, o intervalo de amostragem requerido é também compatível com a generalidade dos sistemas de aquisição de dados disponíveis.

Os sensores por programação foram validados por simulação e com dados experimentais para o processo em estudo.

Tal como discutido anteriormente, os observadores assintóticos apresentam, em geral, um melhor desempenho que os observadores estendidos de Kalman. Por outro lado, os observadores por intervalo apresentam vantagens em termos de implementação prática, apresentando-se bastante promissores neste tipo de estudo, embora a sua validação experimental seja necessária.

O estudo aqui apresentado e que conduziu à estimativa da concentração da biomassa e das taxas específicas de crescimento pode constituir uma peça fundamental para a implementação de estratégias de controlo, como se discutirá no próximo capítulo.

4.4. BIBLIOGRAFIA

Aguilar-Lopez, R. e Maya-Yescas, R., (2005). State estimation for nonlinear systems under model uncertainties: a class of sliding-mode observers. *Journal of Process Control*, 15 (3), 363-370.

Alcaraz-Gonzalez, V., Harmand, J., Rapaport, A., Steyer, J.P., Gonzalez-Alvarez, V. e Pelayo-Ortiz, C., (2002). Software sensors for highly uncertain WWTPs: a new approach based on interval observers. *Water Research*, 36 (10), 2515-2524.

Arndt, M. e Hitzmann, B., (2004). Kalman filter based glucose control at small set points during fed-batch cultivation of *Saccharomyces cerevisiae*. *Biotechnology Progress*, 20 (1), 377-383.

Arndt, M., Kleist, S., Miksch, G., Friehs, K., Flaschel, E., Trierweiler, J. e Hitzmann, B., (2005). A feedforward-feedback substrate controller based on a Kalman filter for a fed-batch cultivation of *Escherichia coli* producing phytase. *Computers & Chemical Engineering*, 29 (5), 1113-1120.

Bastin, G. e Dochain, D., (1990). *On-line estimation and adaptive control of bioreactors*, Elsevier Science Publishers, Amsterdam.

Bernard, O. e Gouzé, J.L., (2004). Closed loop multi-observers for uncertain biotechnological models (CAB9 - 9th International Symposium Computer Applications in Biotechnology (Nancy, France))

Biagiola, S., Solsona, J. e Milocco, R., (2000). Estimation of kinetic rates in batch *Thiobacillus ferrooxidans* cultures. *Journal of Biotechnology*, 84 (1), 13-25.

Biagiola, S.I. e Figueroa, J.L., (2004). Application of state estimation based NMPC to an unstable nonlinear process. *Chemical Engineering Science*, 59 (21), 4601-4612.

Bogaerts, P., (1999). A hybrid asymptotic-Kalman observer for bioprocesses. *Bioprocess Engineering*, 20 (3), 249-255.

Bogaerts, P. e Vande Wouwer, A., (2004). Parameter identification for state estimation - application to bioprocess software sensors. *Chemical Engineering Science*, 59 (12), 2465-2476.

Chachuat, B. e Bernard, O., (2004). Design of Probabilistic Observers for Mass-Balance Based Bioprocess Models (CAB9 - 9th International Symposium Computer Applications in Biotechnology (Nancy, France))

Chachuat, B. e Bernard, O., (2006). Probabilistic observers for a class of uncertain biological processes. *International Journal of Robust and Nonlinear Control*, 16, 157-171.

Chen, L.Z., Nguang, S.K., Li, X.M. e Chen, X.D., (2004). Soft sensors for on-line biomass measurements. *Bioprocess and Biosystems Engineering*, 26 (3), 191-195.

Cheruy, A., (1997). Software sensors in bioprocess engineering. *Journal of Biotechnology*, 52 (3), 193-199.

de Assis, A.J. e Maciel, R., (2000). Soft sensors development for on-line bioreactor state estimation. *Computers & Chemical Engineering*, 24 (2-7), 1099-1103.

Dochain, D., (2001). State observation and adaptive linearizing control for distributed parameter (bio)chemical reactors. *International Journal of Adaptive Control and Signal Processing*, 15 (6), 633-653.

Dochain, D., (2003a). State observers for processes with uncertain kinetics. *International Journal of Control*, 76 (15), 1483-1492.

Dochain, D., (2003b). State and parameter estimation in chemical and biochemical processes: a tutorial. *Journal of Process Control*, 13 (8), 801-818.

Dubach, A. e Märkl, H., (1992). Application of an extended kalman filter method for monitoring high density cultivation of *Escherichia coli*. *Journal of Fermentation and Bioengineering*, 73 (5), 396-402.

Farza, M., Busawon, K. e Hammouri, H., (1998). Simple nonlinear observers for on-line estimation of kinetic rates in bioreactors. *Automatica*, 34 (3), 301-318.

Farza, M., Hammouri, H., Jallut, C. e Lieto, J., (1999). State observation of a nonlinear system: Application to (bio)chemical processes. *AIChE Journal*, 93-106.

- Farza, M., Hammouri, H., Othman, S. e Busawon, K., (1997a). Nonlinear observers for parameter estimation in bioprocesses. *Chemical Engineering Science*, 52 (23), 4251-4267.
- Farza, M., Othman, S., Hammouri, H. e Biston, J., (1997b). A nonlinear approach for the on-line estimation of the kinetic rates in bioreactors - Application to a lactic acid production process. *Bioprocess Engineering*, 17 (3), 143-150.
- Ferreira, E.C., (1995). *Identificação e controlo adaptativo de processos biotecnológicos*. Tese de Doutoramento, Universidade do Porto, Portugal.
- Franz, C., Jürgen, K., Florentina, P. e Karl, B., (2005). Sensor combination and chemometric modelling for improved process monitoring in recombinant *E. coli* fed-batch cultivations. *Journal of Biotechnology*, 120, 183-196.
- Gouzé, J.L., Rapaport, A. e Hadj-Sadok, M.Z., (2000). Interval observers for uncertain biological systems. *Ecological Modelling*, 133 (1-2), 45-56.
- Hadj-Sadok, M.Z. e Gouzé, J.L., (2001). Estimation of uncertain models of activated sludge processes with interval observers. *Journal of Process Control*, 11 (3), 299-310.
- James, S., Legge, R. e Budman, H., (2002). Comparative study of black-box and hybrid estimation methods in fed-batch fermentation. *Journal of Process Control*, 12 (1), 113-121.
- James, S.C., Legge, R.L. e Budman, H., (2000). On-line estimation in bioreactors: A review. *Reviews in Chemical Engineering*, 16, 311-340.
- Jenzsch, M., Simutis, R., Eisbrenner, G., Stückrath, I. e Lübbert, A., (2006). Estimation of biomass concentrations in fermentation processes for recombinant protein production. *Bioprocess and Biosystems Engineering*, 29 (1), 19-27.
- Keesman, K.J., (2002). State and parameter estimation in biotechnical batch reactors. *Control Engineering Practice*, 10 (2), 219-225.
- Komives, C. e Parker, R.S., (2003). Bioreactor state estimation and control. *Current Opinion in Biotechnology*, 14 (5), 468-474.
- Lemesle, V. e Gouzé, J.L., (2005). Hybrid bounded error observers for uncertain bioreactor models. *Bioprocess and Biosystems Engineering*, 27, 311-318.
- Lombardi, M., Fiaty, K. e Laurent, P., (1999). Implementation of observer for on-line estimation of concentration in continuous-stirred membrane bioreactor: Application to the fermentation of lactose. *Chemical Engineering Science*, 54 (13-14), 2689-2696.
- Lubenova, V., (1996). On-line estimation of biomass concentration and non stationary parameters for aerobic bioprocesses. *Journal of Biotechnology*, 46 (3), 197-207.
- Lubenova, V., Rocha, I. e Ferreira, E.C., (2003). Estimation of multiple biomass growth rates and biomass concentration in a class of bioprocesses. *Bioprocess and Biosystems Engineering*, 25 (6), 395-406.
- Lubenova, V.N., (1999). Stable adaptive algorithm for simultaneous estimation of time-varying parameters and state variables in aerobic bioprocesses. *Bioprocess Engineering*, 21, 219-226.
- Nadri, M., Trezzani, I., Hammouri, H., Dhurjati, P., Longin, R. e Lieto, J., (2006). Modeling and observer design for recombinant *Escherichia coli* strain. *Bioprocess and Biosystems Engineering*, 28, 217-225.

Neeleman, R., (2001). Respiratory quotient: estimation during batch cultivation in bicarbonate buffered media. *Engineering and Manufacturing for Biotechnology*, 203-216.

Neeleman, R., van den End, E.J. e Van Boxtel, A.J.B., (2000). Estimation of the respiratory quotient in a bicarbonate buffered batch cell cultivation. *Journal of Biotechnology*, 80 (1), 85-95.

November, E.J. e Van Impe, J.F., (2002). The tuning of a model-based estimator for the specific growth rate of *Candida utilis*. *Bioprocess and Biosystems Engineering*, 25 (1), 1-12.

Oliveira, R., Ferreira, E.C. e de Azevedo, S.F., (2002). Stability, dynamics of convergence and tuning of observer-based kinetics estimators. *Journal of Process Control*, 12 (2), 311-323.

Oliveira, R., Ferreira, E.C., Oliveira, F. e Fayo de Azevedo, S., (1996). A study on the convergence of observer-based kinetic estimators in fed-batch fermentations. *Journal of Process Control*, 6 (6), 367-371.

Perrier, M., Fayo de Azevedo, S., Ferreira, E.C. e Dochain, D., (2000). Tuning of observer-based estimators: theory and application to the on-line estimation of kinetic parameters. *Control Engineering Practice*, 8, 377-388.

Pomerleau, Y. e Perrier, M., (1990). Estimation of multiple specific growth rates in bioprocesses. *AIChE Journal*, 36 (2), 207-215.

Raissi, T., Ramdani, N. e Candau, Y., (2005). Bounded error moving horizon state estimator for non-linear continuous- time systems: application to a bioprocess system. *Journal of Process Control*, 15 (5), 537-545.

Rapaport, A. e Dochain, D., (2005). Interval observers for biochemical processes with uncertain kinetics and inputs. *Mathematical Biosciences*, 193 (2), 235-253.

Rapaport, A. e Gouzé, J.L., (2003). Parallelotopic and practical observers for non-linear uncertain systems. *International Journal of Control*, 76 (3), 237-251.

Rocha, I., (2003). *Model-based strategies for computer-aided operation of recombinant E. coli fermentation*. Tese de Doutorado, Universidade do Minho, Portugal.

Rocha, I. e Ferreira, E.C., (2002). On-line simultaneous monitoring of glucose and acetate with FIA during high cell density fermentation of recombinant *E. coli*. *Analytica Chimica Acta*, 462 (2), 293-304.

Smith, H.L., (1996). Monotone dynamical systems: An introduction to the theory of competitive and cooperative systems. *Bulletin (New Series) of the American Mathematical Society*, 33 (2), 203-209.

Sonnleitner, B., (2003). Monitoring of bioprocesses. Em: Berovic, M. e Kieran, P. (eds), *Bioprocess Engineering: doctoral/post-doctoral level course*. pp. 299-323, Faculdade de Química e Tecnologia Química, Ljubljana.

Soroush, M., (1997). Nonlinear state-observer design with application to reactors. *Chemical Engineering Science*, 52 (3), 387-404.

Tatiraju, S., Soroush, M. e Mutharasan, R., (1999). Multi-rate nonlinear state and parameter estimation in a bioreactor. *Biotechnology and Bioengineering*, 63 (1), 22-32.

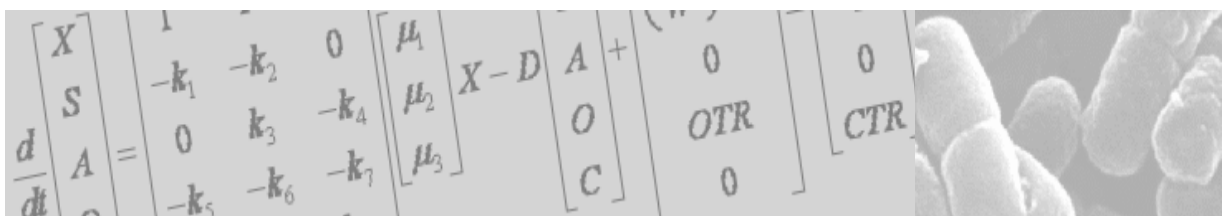
Vanek, M., Hrnčirik, P., Vovsik, J. e Nahlik, J., (2004). On-line estimation of biomass concentration using a neural network and information about metabolic state. *Bioprocess and Biosystems Engineering*, 27 (1), 9-15.

Vojinovic, V., Cabral, J.M.S. e Fonseca, L.P., (2006). Real-time bioprocess monitoring. Part I: In situ sensors. *Sensors and Actuators B-Chemical*, 114, 1083-1091.

Zhang, T. e Guay, M., (2002). Adaptive nonlinear observers of microbial growth processes. *Journal of Process Control*, 12 (5), 633-643.

CAPÍTULO 5

CONTROLO DO PROCESSO



5. CONTROLO DO PROCESSO

O controlo de bioprocessos tem-se revelado uma área de investigação de grande interesse, tendo-se assistido por isso, e em especial nos últimos anos, a um crescente número de trabalhos de investigação nesta área (Ferreira e Foyo de Azevedo, 1996; Costa *et al.*, 1998; Gordillo *et al.*, 1998; Siegwart *et al.*, 1999; Rani e Rao, 1999; Levisauskas, 2001; Aguilar *et al.*, 2001; Akesson *et al.*, 2001a; Frahm *et al.*, 2002; Szederkenyi *et al.*, 2002; Rocha e Ferreira, 2002a; Zhang *et al.*, 2003; Jobe *et al.*, 2003; Jenzsch *et al.*, 2004; Whiffin *et al.*, 2004; Mailleret *et al.*, 2004; Chung *et al.*, 2006).

A capacidade de controlar os processos fermentativos na proximidade do seu desempenho óptimo, de um modo preciso e automático, é de grande importância e interesse para a maior parte das indústrias biotecnológicas. De facto, o controlo apropriado de um processo permite reduzir não só os custos de produção, mas também, aumentar a sua produtividade assegurando e mantendo a qualidade do produto final.

Num passado ainda recente, o controlo na indústria biotecnológica tinha apenas como objectivo manter, a um nível desejado, variáveis como a temperatura de operação, o pH e a concentração de oxigénio dissolvido. Contudo, é hoje reconhecida a necessidade de controlar outras variáveis como por exemplo, a concentração de substrato, a concentração de produtos intermédios, a velocidade de transferência de oxigénio e a taxa específica de crescimento, entre outras, para as quais é imprescindível desenvolver algoritmos simples e estáveis, adequados às necessidades das aplicações industriais.

Se, por um lado, é verdade que desde há várias décadas os métodos de controlo avançado têm sido aplicados com sucesso na indústria química, por outro lado, só mais recentemente essas estratégias de controlo, baseadas em modelos matemáticos, têm sido implementadas em processos biológicos (Lübert e Jorgensen, 2001; Komives e Parker, 2003; Alford, 2006). Este facto poder-se-á dever ao desenvolvimento de áreas relacionadas com o controlo, que têm permitido ultrapassar as dificuldades no controlo de biorreactores. De entre essas áreas podem destacar-se os avanços verificados no melhoramento da monitorização desses processos, nomeadamente por intermédio da disponibilidade, em tempo real, de medições quantitativas das suas principais variáveis, permitindo incrementar a qualidade do controlo efectuado. Esses avanços têm-se verificado quer no âmbito do desenvolvimento de novos sensores físicos quer, principalmente, ao nível do projecto de sensores por programação, como abordado no capítulo anterior.

O desenvolvimento de estratégias de controlo baseadas em modelos matemáticos adquire assim, uma relevância primordial. Com efeito, essas metodologias têm provado ser capazes de melhorar, na maior parte dos casos, o desempenho do processo, reduzindo os seus custos de produção, em especial quando implementadas em linha, uma vez que permitem a eliminação de perturbações do

processo. Por outro lado, o recurso a estes métodos afigura-se vantajoso no que se refere à sintonização dos parâmetros do controlador, permitindo-lhe obter uma dinâmica mais estável.

No que diz respeito ao crescimento e produção de proteínas recombinadas em *E. coli* é sabido que existem vários factores que influenciam a obtenção de culturas de alta densidade celular e elevada produtividade. De entre esses factores a velocidade específica de crescimento adquire um papel de destaque quer na monitorização quer no controlo e optimização do bioprocessos. De facto, através do controlo deste parâmetro poder-se-á reduzir ou mesmo eliminar a produção de acetato que, como referido em capítulos anteriores, quando presente no meio em concentração elevada, inibe o crescimento da *E. coli* diminuindo a produção de proteínas recombinadas. Mais ainda, a possibilidade de controlar a taxa específica de crescimento em valores diferentes antes e depois da indução da produção da proteína poderá traduzir-se no aumento da produtividade (Curless *et al.*, 1990; Yee e Blanch, 1993; Lim e Jung, 1998; Akesson *et al.*, 2001b; Sanden *et al.*, 2003; Jana e Deb, 2005).

Neste capítulo será apresentada uma lei de controlo adaptativo desenvolvida para o controlo da taxa específica de crescimento num processo semi-contínuo de crescimento de *E. coli*. O estudo foi realizado a partir de simulações matemáticas desse processo, avaliando-se a robustez do algoritmo através da variação dos parâmetros do modelo, da alteração do valor de referência da taxa específica de crescimento e pela introdução de perturbações externas através da variação da concentração da glucose na alimentação, S_e . Foram ainda avaliados quatro mecanismos de sintonização dos parâmetros do controlador, sendo os seus desempenhos comparados entre si e com algoritmos de controlo frequentemente referidos na literatura.

5.1. ESTRATÉGIAS DE CONTROLO

A aplicação de metodologias de controlo a sistemas biológicos tem como principais objectivos melhorar a estabilidade de operação, aumentando a eficiência da produção e o lucro, bem como lidar com variações dinâmicas que ocorrem durante as várias fases das fermentações (Lim e Lee, 1990; Lee *et al.*, 1999).

Contudo, os bioprocessos são de um modo geral muito difíceis de controlar. Os principais factores que contribuem para essa dificuldade são: (i) o facto da dinâmica do seu comportamento ser não linear, (ii) o desconhecimento, em geral, de modelos precisos desses processos em virtude da complexidade inerente aos processos bioquímicos, (iii) a variação, quase sempre imprevisível, dos parâmetros do modelo, e (iv) a não disponibilidade no mercado de sensores robustos e baratos que permitam medir variáveis relacionadas com a actividade celular, fazendo com que o estado do processo seja difícil de caracterizar. Deste modo, verifica-se que o desenho de algoritmos de controlo baseado em modelos matemáticos para estes processos é dificultado principalmente por

dois problemas, os quais necessitam de soluções de engenharia adequadas. O primeiro relaciona-se com os pontos *i)*, *ii)* e *iii)* mencionados anteriormente, podendo-se realçar o facto das suas cinéticas, na maioria das vezes, serem funções não lineares e pouco compreendidas, às quais correspondem parâmetros geralmente variáveis no tempo. O segundo problema, relacionado com o ponto *iv)*, reside no facto de até à presente data, existirem poucos sensores disponíveis capazes de monitorizar em tempo real as variáveis do processo que são necessárias para os algoritmos de controlo avançado (Lee *et al.*, 1999; Rani e Rao, 1999; Komives e Parker, 2003). Assim, e apesar de reconhecida a importância dos modelos matemáticos para controlar os bioprocessos, estes são ainda pouco utilizados a nível industrial para esse fim. Sendo assim, as trajectórias resultantes de experiências com sucesso são a base do controlo ainda usado em muitos processos de fermentação utilizados na indústria.

O número de maneiras de controlar um processo biotecnológico assim como a literatura existente acerca deste controlo é vasto, pelo que aqui, apenas se fará uma breve descrição de alguns conceitos bem conhecidos do controlo em geral (Stephanopoulos, 1984; Chen *et al.*, 1995; Jadot *et al.*, 1998; Lee *et al.*, 1999; Rani e Rao, 1999; Dochain, 2001).

5.1.1. Controlo por retroacção (*feedback*)

O controlo por retroacção é um modelo clássico de controlo automático de um sistema, podendo ser representado esquematicamente pela Figura 5.1.

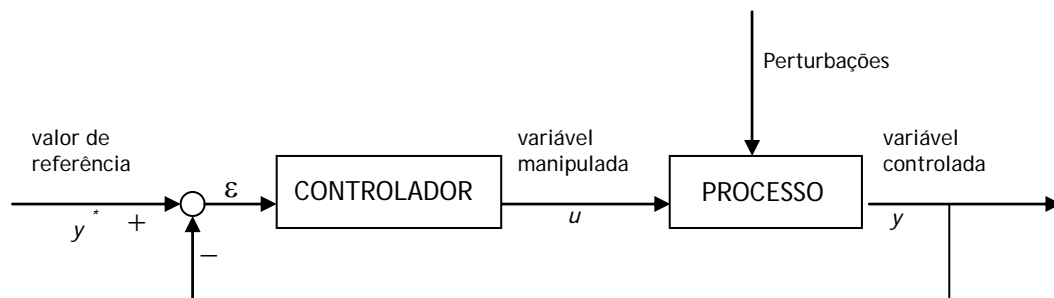


Figura 5.1 Esquema de controlo por retroacção.

O ciclo de controlo começa pela medição da variável de saída y . É gerado um sinal erro, ε , subtraindo o valor medido da variável de saída, y , a um valor de referência desejado, y^* , sendo calculado o sinal de controlo por aplicação de um dado algoritmo que decide qual a variação da variável manipulada, u , de modo a compensar esse desvio e assim reduzir a diferença entre o valor da variável y e o seu valor de referência.

Esta estratégia de controlo não requer um modelo matemático preciso sendo insensível quer a erros do modelo quer a variações nos seus parâmetros. No entanto, a sua acção apenas ocorre após

uma dada perturbação ter ocorrido, não sendo adequado no controlo de sistemas cuja dinâmica é lenta podendo criar instabilidade.

5.1.2. Controlo por antecipação (*feedforward*)

Embora, em várias situações o controlo por retroacção seja adequado, nalguns casos é necessário recorrer a outro tipo de controlo como por exemplo, o controlo por antecipação. Neste caso, contrariamente ao controlo por retroacção normal, não se espera até que a perturbação afecte a saída. Pelo contrário, a perturbação é medida e um controlo correctivo é aplicado por antecipação ao efeito esperado. Esta estratégia pode apresentar um melhor desempenho, quando permite a eliminação do efeito da perturbação, sendo adequado para sistemas cuja dinâmica é lenta, não introduzindo instabilidade no processo. A Figura 5.2 apresenta o esquema do controlo por antecipação.

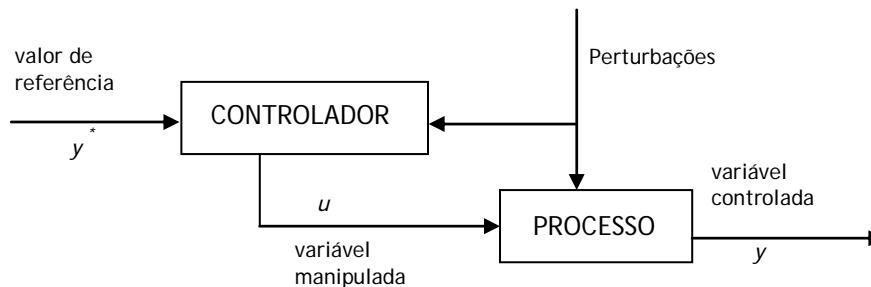


Figura 5.2 Esquema de controlo por antecipação.

Como se pode inferir pela observação da figura anterior, este tipo de controlo possui algumas desvantagens nomeadamente, a necessidade de dispor de um bom modelo do processo, sensibilidade a variações dos parâmetros do modelo, a necessidade de conhecer a perturbação *a priori* e a sua incapacidade para lidar com perturbações não medidas. Nesta estratégia o modelo matemático do processo tem um papel preponderante pelo facto de ser através deste que se consegue prever os efeitos da perturbação antes que estes sejam efectivamente sentidos pelo sistema.

5.1.3. Controlo por antecipação/retroacção (*feedforward/feedback*)

O controlo por antecipação pode não ser eficiente pelas razões anteriormente apresentadas, contudo, esses problemas podem ser ultrapassados combinando este tipo de controlo com o controlo por retroacção de modo a corrigir erros provocados pelo conhecimento deficiente na previsão do efeito de perturbações nas variáveis de saída. A combinação entre modos de controlo por

retroacção e por antecipação permite que a maioria do efeito provocado pela perturbação possa ser corrigida pelo controlador por antecipação deixando o controlo fino para o ciclo de retroacção. Na Figura 5.3 representa-se esquematicamente o controlo por antecipação/retroacção de um processo.

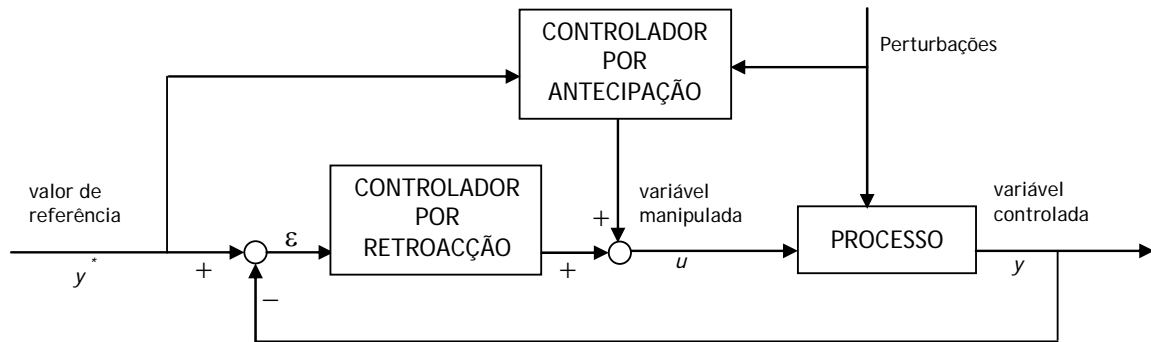


Figura 5.3 Esquema de controlo por antecipação/retroacção.

5.1.4. Tipos de controladores

A escolha do tipo de controlador que se utilizará para assegurar a acção de controlo desejada é uma das decisões mais importantes para garantir o seu correcto funcionamento. Na literatura estão descritos vários tipos de controladores utilizados no controlo de bioprocessos destacando-se de entre eles os controladores "on-off", os controladores PID (controlador proporcional, integral, derivativo) e os controladores adaptativos.

O controlador "on-off" é o mais simples sendo utilizado quando a dinâmica da variável a controlar não requer um controlo muito complexo, sendo a sua acção independente da magnitude do erro entre o valor de referência pretendido e o valor actual da variável. Este controlador é o tipicamente utilizado no controlo do pH. Em muitos casos o controlador "on-off" não é suficientemente preciso para controlar variáveis que se pretendem manter estáveis dentro de uma gama restrita de valores, conseguindo-se um melhor desempenho quando a acção do controlador é função da magnitude do erro, o que é conseguido com um controlador PID. Neste tipo de controlo podem ser diferenciados três possibilidades dependendo do tipo de acção envolvida: o controlo proporcional (P), o proporcional integral (PI) e o proporcional integral derivativo (PID). Nalguns casos é ainda possível utilizar o controlo proporcional derivativo (PD). O controlo PID é o mais utilizado devido à sua simplicidade, robustez e versatilidade. Contudo, este tipo de controlo não é aplicável num reduzido número de situações, para as quais as condições e a resposta do sistema são variáveis. Nesses casos podem utilizar-se controladores alternativos como os controladores adaptativos (Lim e Lee, 1990).

Uma panorâmica geral do controlo adaptativo pode ser encontrada em diferentes referências bibliográficas (Aström e Wittenmark, 1989; Sastry e Bodson, 1989; Narendra e Annaswamy, 1989;

Bastin e Dochain, 1990; Åström *et al.*, 1995). Podem, assim, ser encontradas e distinguidas duas abordagens sobre o controlo adaptativo.

No primeiro grupo a adaptação do controlador está relacionada com o seu desempenho onde o que conduz a adaptação é o erro entre o comportamento desejado e o comportamento actual do sistema. Este erro pode ser devido a incertezas associadas aos parâmetros cinéticos ou a todo o modelo do processo. Este método de controlo tem sido aplicado em várias situações como o controlo do oxigénio dissolvido, da concentração de substrato e da temperatura de operação (Frahm *et al.*, 2002; Akay *et al.*, 2002; Babuska *et al.*, 2003). Nesta abordagem, para se conseguir uma adaptação adequada é necessário que sejam introduzidos erros excitando sistematicamente o sistema (Soons *et al.*, 2006).

No segundo grupo de controladores adaptativos as expressões que permitem a adaptação dos parâmetros do controlador estão relacionados com as variáveis de estado medidas ou estimadas do processo (Bastin e Dochain, 1990; Van Impe e Bastin, 1995; Smets *et al.*, 2002b; 2004). Esta abordagem tem sido adoptada por vários autores (Levisauskas *et al.*, 1996; Rocha e Ferreira, 2002b; Soons *et al.*, 2006).

No trabalho de Chang (2003) a sintonização dos parâmetros de um controlador PI²D é baseada nas variáveis de estado medidas. Levisauskas *et al.* (1996) aplicam a mesma abordagem usando um controlador cujos parâmetros se adaptam à alteração das variáveis de estado. No trabalho de Alvarez e Simutis (2004) é apresentada e estudada por simulação numérica uma lei de controlo da velocidade específica de crescimento, onde esse parâmetro e a biomassa são estimados através de um observador de Kalman.

No trabalho desenvolvido a lei de controlo apresentada baseia-se na abordagem de adaptação do segundo grupo referido como será apresentado das secções que se seguem.

5.2. MODELO DO PROCESSO

O modelo usado para o desenho do algoritmo de controlo é baseado no modelo dinâmico geral para reactores biológicos desenvolvido por Bastin e Dochain (1990) e já apresentado no Capítulo 3.

$$\frac{d\xi}{dt} = Kr(\xi, t) - D\xi + F - Q \quad \text{Equação 5.1}$$

Quando aplicado à fermentação semi-contínua de *E. coli* origina o modelo seguinte, como descrito nos capítulos anteriores:

$$\frac{d}{dt} \begin{bmatrix} X \\ S \\ A \\ O \\ C_{tot} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ -k_1 & -k_2 & 0 \\ 0 & k_3 & -k_4 \\ -k_5 & -k_6 & -k_7 \\ k_8 & k_9 & k_{10} \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} X - D \begin{bmatrix} X \\ S \\ A \\ O \\ C_{tot} \end{bmatrix} + \begin{bmatrix} 0 \\ \left(\frac{F_{e,s}}{W}\right) S_e \\ 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ OTR \\ -CTR \end{bmatrix} \quad \text{Equação 5.2}$$

O modelo cinético, assim como os valores dos parâmetros cinéticos e coeficientes de rendimento são apresentados no Capítulo 3.

5.3. CONTROLO DA TAXA ESPECÍFICA DE CRESCIMENTO

Tradicionalmente, o controlo de bioprocessos é feito em ciclo aberto, ou seja, o perfil do caudal de alimentação $F_{e,s}$ é determinado previamente e aplicado durante o processo de produção sem o recurso a qualquer controlo por retroacção (Gregory e Turner, 1993; Korz *et al.*, 1995; Shin *et al.*, 1997; Martinez *et al.*, 1998; Suárez e Kilikian, 2000; Cheng *et al.*, 2002; Smets *et al.*, 2002b). A aplicação de um caudal exponencial é a estratégia mais comum. Contudo, em geral, a fermentação está sujeita a perturbações que podem ser o resultado de alterações das condições iniciais (por exemplo as concentrações iniciais de substrato ou de biomassa) relativamente às que foram assumidas na determinação do perfil de alimentação ou devido a perturbações que ocorram no decurso da fermentação (por exemplo variações do caudal de alimentação ou problemas relacionados com a transferência de massa do oxigénio). Deste modo, o efeito da ocorrência aleatória de perturbações só pode ser tido em conta recorrendo ao controlo por retroacção, o qual é dificultado, como já referido, quer pela falta de sensores comercialmente disponíveis para medir o valor actual da variável controlada e/ou do estado do processo, quer pela dinâmica complexa e não linear dos bioprocessos. Por estes motivos, os algoritmos de controlo mais simples, como os controladores P, PI e PID, não funcionam adequadamente o que motiva o desenvolvimento de algoritmos de controlo adaptativo baseados em modelos matemáticos (Delisa *et al.*, 2001; Johnston *et al.*, 2002; Jenzsch *et al.*, 2006).

Com vista a um melhor controlo e optimização de bioprocessos é essencial medir em linha parâmetros fisiológicos chave. No entanto, esse tipo de medidas nem sempre se encontra disponível, embora alguns esforços tenham sido realizados para desenvolver sensores que permitam fornecer essa informação. Por tal motivo, existe uma necessidade urgente de aceder a este tipo de informação estimando os parâmetros chave a partir de outros que sejam mais simples e fáceis de medir. Por exemplo, no que diz respeito à indústria farmacêutica, a consistência na produção é um dos principais, senão, o mais importante objectivo. Para alcançar tal objectivo durante o processo de cultivo, o estado metabólico dos microrganismos no biorreactor deve ser mantido constante,

podendo-se garantir essa estabilidade através do controlo da taxa específica de crescimento, μ . De facto, esta taxa é um dos parâmetros mais importantes do processo caracterizando o estado metabólico dos microrganismos durante a fermentação, sobretudo devido ao facto da biossíntese de muitos produtos de interesse estar frequentemente associada ao valor desse parâmetro (Levisauskas, 2001). Assim sendo, o controlo do valor de μ é uma tarefa primordial, o qual pode ser conseguido manipulando convenientemente o caudal de alimentação de substrato. Deste modo, o recurso a fermentações de operação semi-contínua ou contínua afigura-se uma alternativa vantajosa à utilização de fermentações de operação descontínua, permitindo controlar a taxa específica de crescimento num dado valor que possibilite a manutenção de um estado metabólico adequado, maximizando a produção e simultaneamente aumentando a consistência do processo.

No caso particular da fermentação semi-contínua de *E. coli* para produção de proteínas recombinadas é importante manter a taxa específica de crescimento inferior a um dado valor, de modo a evitar a acumulação de acetato durante a fermentação e, adicionalmente, controlar cuidadosamente as velocidades específicas de crescimento pré- e pós-indução com o intuito de obter uma produtividade máxima da proteína pretendida (Lee e Bailey, 2002; Pinsach *et al.*, 2006).

Deste modo, o desenvolvimento de algoritmos para o controlo automático da taxa específica de crescimento em processos fermentativos com operação semi-contínua reveste-se de enorme relevância. Contudo, o desempenho destes algoritmos depende da capacidade em determinar eficientemente a taxa específica de crescimento, que não pode ser obtida de forma directa, como referido anteriormente.

Na literatura encontram-se descritos poucos exemplos de aplicação de estratégias de controlo em ciclo fechado da taxa específica de crescimento, a fermentações de *E. coli* ou a outros bioprocessos complexos (Jenzsch *et al.*, 2006; Soons *et al.*, 2006). Nos trabalhos de Levisauskas *et al.* (1996) e Levisauskas (2001) é descrita uma estratégia de controlo automático da taxa específica de crescimento para uma fermentação semi-contínua de *E. coli* que requer a monitorização em linha de OUR, CER, caudal de alimentação e volume do meio de alimentação. No trabalho de Jenzsch *et al.* (2006) é também apresentado um algoritmo de controlo por antecipação/retroacção para a taxa específica de crescimento numa fermentação semi-contínua de *E. coli*. De modo a implementar este algoritmo de controlo é necessário estimar ou medir em linha as variáveis concentração de biomassa, concentração de substrato, o valor da taxa específica de crescimento e o peso do meio de fermentação. Os referidos autores utilizaram um observador estendido de Kalman para estimar os valores das variáveis que não conseguiam medir em linha, X , S e μ . Para determinar os parâmetros de sintonização do controlador utilizaram uma técnica de optimização de procura aleatória. O desempenho deste algoritmo é comparado com o de um controlador PI clássico, tendo verificado que o primeiro apresentou um melhor desempenho. O estudo foi realizado através de simulações numéricas tendo sido validado numa situação real. Soons *et al.* (2006) apresentam um algoritmo de controlo adaptativo com modelo de referência para a taxa específica de crescimento para uma fermentação de *Bordetella pertussis*, tendo recorrido igualmente a um observador

estendido de Kalman para estimar a concentração de biomassa e a taxa específica de crescimento. A sintonização do controlador foi também efectuada por optimização dos seus parâmetros. O estudo foi realizado recorrendo a simulações numéricas tendo sido ainda validado experimentalmente.

O principal objectivo deste capítulo é assim, o desenvolvimento de uma lei de controlo para a taxa específica de crescimento da *E. coli* numa fermentação semi-contínua, tendo-se adoptado como estratégia o controlo adaptativo com modelo de referência adaptado de Soons *et al.* (2006). Para esse propósito utilizou-se o modelo matemático do processo (Equação 5.2) em conjunto com medições em linha da concentração de oxigénio dissolvido e da concentração de oxigénio na corrente gasosa de saída do reactor juntamente, com um observador assintótico que permite a estimação em linha da concentração da biomassa e de um estimador baseado num observador que permite a estimação da taxa específica de crescimento, os quais foram apresentados no Capítulo 4.

5.3.1. Lei de controlo

A metodologia adoptada na derivação da lei de controlo baseia-se no modelo dinâmico geral dado pela Equação 5.1, e para este caso particular, no modelo descrito pela Equação 5.2.

O objectivo do controlador que se pretende desenvolver é manter a taxa específica de crescimento próxima do valor de referência pré-estabelecido mesmo na presença de várias incertezas, incluindo perturbações externas e parâmetros do modelo variantes no tempo ou não conhecidos com precisão, manipulando o caudal de alimentação do substrato, $F_{e,s}$.

O algoritmo de controlo adaptativo foi derivado combinando a dinâmica de μ , determinada através do modelo matemático do processo e da sua dinâmica obtida pelo modelo de referência que permite a resposta ideal do controlador (Bastin e Dochain, 1990), como será apresentado de seguida.

A lei de controlo foi derivada para a fermentação semi-contínua de *E. coli*, na qual se pretende que ocorra apenas o regime respirativo com glucose ($\mu_1 \neq 0$ e $\mu_2 = \mu_3 = 0$) de modo a não se verificar produção de acetato.

A primeira etapa consistiu em obter uma expressão da derivada da taxa específica de crescimento através do modelo matemático do processo. Sendo assim, sabe-se, pelo modelo cinético (Capítulo 3) que:

$$\mu_1 = \frac{q_s}{k_1} \tag{Equação 5.3}$$

$$\text{com } q_s = q_{s,\text{máx}} \frac{S}{K_s + S} \frac{K_{i,s}}{K_{i,s} + A}$$

no entanto, é considerado que o crescimento da *E. coli* é oxidativo em glucose, não havendo, por isso, acetato no meio ($A=0$). Sendo assim, vem:

$$q_s = q_{s,max} \frac{S}{K_s + S} \quad \text{Equação 5.4}$$

e, conseqüentemente pode-se então definir que a taxa específica de crescimento ($\mu=\mu_1$) é dada pela equação que se segue,

$$\mu = \mu_1 = \frac{q_{s,max}}{k_1} \frac{S}{K_s + S} \quad \text{Equação 5.5}$$

Derivando a Equação 5.5 em ordem ao tempo obtém-se a expressão da derivada de μ . Atendendo a que $q_{s,max}$, k_1 e K_s são parâmetros do modelo cujo valor é constante, a taxa específica de crescimento é função apenas da concentração de substrato:

$$\frac{d\mu}{dt} = \frac{\partial\mu}{\partial S} \frac{dS}{dt} \quad \text{Equação 5.6}$$

Para o caso em estudo pode-se então determinar $\frac{\partial\mu}{\partial S}$ que é dado por:

$$\frac{\partial\mu}{\partial S} = \frac{q_{s,max}}{k_1} \frac{K_s}{(K_s + S)^2} \quad \text{Equação 5.7}$$

Considerando o modelo do processo e considerando que na derivação da lei de controlo o caudal de alimentação de substrato é igual ao caudal total de alimentação ($F \approx F_{e,s}$) obtém-se o seguinte balanço ao substrato que é dado por:

$$\frac{dS}{dt} = -k_1\mu X + \frac{F_{e,s}}{W}(S_e - S) \quad \text{Equação 5.8}$$

sendo assim,

$$\frac{d\mu}{dt} = \frac{q_{s,max}}{k_1} \frac{K_s}{(K_s + S)^2} \left(-k_1\mu X + \frac{F_{e,s}}{W}(S_e - S) \right) \quad \text{Equação 5.9}$$

Esta equação representa a derivada da taxa específica de crescimento obtida através do modelo do processo. Pretendeu-se, numa segunda etapa, determinar essa mesma derivada recorrendo ao

modelo de referência, o qual define o modo ideal como a taxa específica de crescimento deverá responder à acção de controlo, isto é, a forma como o erro de convergência do controlador ($\varepsilon = \mu^* - \mu$) será orientado.

O objectivo do controlo consistirá em impor uma dinâmica estável em ciclo fechado dada por um modelo de referência de segunda ordem que, como referido na literatura (Jenzsch *et al.*, 2006; Soons *et al.*, 2006), permite maior estabilidade:

$$\frac{d^2(\mu^* - \mu)}{dt^2} + \gamma_1 \frac{d(\mu^* - \mu)}{dt} + \gamma_2(\mu^* - \mu) = 0, \quad \gamma_1, \gamma_2 > 0 \quad \text{Equação 5.10}$$

em que μ^* é o valor de referência pré-estabelecido da taxa específica de crescimento e γ_1 e γ_2 são os parâmetros de sintonização do controlador que determinam a sua velocidade de convergência. Num problema de regulação μ^* é um valor constante, ou seja a sua derivada é nula. Sendo assim,

$$\frac{d^2\mu}{dt^2} = \gamma_2(\mu^* - \mu) - \gamma_1 \frac{d\mu}{dt} \quad \text{Equação 5.11}$$

O modelo de referência para a taxa específica de crescimento é obtido integrando a equação anterior entre zero e t e considerando que para $t=0$, $\mu=\mu^*$, sendo dado por:

$$\frac{d\mu}{dt} = \gamma_1(\mu^* - \mu) + \int_0^t \gamma_2(\mu^* - \mu) dt \quad \text{Equação 5.12}$$

Embora a estabilidade do processo não possa ser garantida *a priori*, uma vez que a concentração de biomassa segue uma trajectória de crescimento exponencial, o que é característico de um comportamento instável (Bastin e Dochain, 1990), a estabilidade do modelo de referência pode ser garantida. O modelo de referência é estável se γ_1 e γ_2 forem estritamente positivos. Sendo a dinâmica do processo forçada a seguir a dinâmica do modelo de referência por acção do controlador, a convergência deste também é garantida.

Igualando a Equação 5.9 à Equação 5.12, eliminando $\frac{d\mu}{dt}$ e resolvendo em ordem $F_{e,s}$, obtém-se a seguinte lei de controlo adaptativo:

$$F_{e,s} = \frac{k_1 \mu X W}{(S_e - S)} + \left[\frac{\gamma_1}{\frac{q_{s,max} K_s}{(K_s + S)^2 k_1}} (\mu^* - \mu) + \frac{1}{\frac{q_{s,max} K_s}{(K_s + S)^2 k_1}} \int_0^t \gamma_2 (\mu^* - \mu) dt \right] \frac{W}{(S_e - S)} \quad \text{Equação 5.13}$$

Verificou-se que a qualidade do controlo melhorava se no primeiro termo da equação anterior se substituísse μ por μ^* , o que foi também constatado por Jenzsch *et al.* (2006). Assim sendo, a lei de controlo é definida pela expressão:

$$F_{e,s} = \frac{k_1 \mu^* XW}{(S_e - S)} + \left[\frac{\gamma_1}{\frac{q_{s,max} K_s}{(K_s + S)^2 k_1}} (\mu^* - \mu) + \frac{1}{\frac{q_{s,max} K_s}{(K_s + S)^2 k_1}} \int_0^t \gamma_2 (\mu^* - \mu) dt \right] \frac{W}{(S_e - S)} \quad \text{Equação 5.14}$$

Da Equação 5.14 verifica-se que será necessário conhecer os parâmetros do modelo $q_{s,max}$, k_1 e K_s , os parâmetros do controlador γ_1 e γ_2 bem como os valores em linha da taxa específica de crescimento e das concentrações de biomassa e substrato. No Capítulo 4 foram descritos sensores por programação que permitiram estimar de modo satisfatório o valor de μ e da concentração de biomassa, embora não se tenham obtido resultados tão satisfatórios na estimação da concentração de substrato. Sendo assim, estudou-se a possibilidade de eliminar essa variável da lei de controlo.

Resolvendo a Equação 5.5 em ordem a S , obtém-se:

$$S = \frac{\mu K_s}{\left(\frac{q_{s,max}}{k_1} - \mu \right)} \quad \text{Equação 5.15}$$

Atendendo a que a concentração de substrato na alimentação é muito superior à referida concentração no interior do reactor ($S_e - S \approx S_e$), e substituindo S na Equação 5.14 pela equação anterior, vem:

$$F_{e,s} = \frac{k_1 \mu^* \hat{X}W}{S_e} + \left[\frac{\gamma_1}{\frac{(1 - m\mu^*)^2}{mK_s}} (\mu^* - \hat{\mu}) + \frac{1}{\frac{(1 - m\mu^*)^2}{mK_s}} \int_0^t \gamma_2 (\mu^* - \hat{\mu}) dt \right] \frac{W}{S_e} \quad \text{Equação 5.16}$$

$$\text{em que } m = \frac{k_1}{q_{s,max}}$$

A Equação 5.16 traduz, como se pretendia, a lei de controlo adaptativo para a qual só é necessário estimar a concentração de biomassa e a taxa específica de crescimento. Esta expressão pode ser facilmente interpretada como sendo um controlador por antecipação/retroacção. O primeiro termo representa o caudal de alimentação derivado do modelo, assumindo um processo não perturbado, constituindo assim o controlo por antecipação. O segundo e o terceiro termos podem ser interpretados como um controlador por retroacção do tipo PI, corrigindo para desvios em

μ devidos a perturbações do processo e/ou a pequenas incorrecções do modelo/processo. Como se pode inferir a acção PI do controlador é também influenciada e de certa maneira adaptada pela variação de W , que vai sendo medida em linha através de uma balança, como referido no Capítulo 2.

Como referido anteriormente, a implementação da lei de controlo apresentada só é possível quando os valores da concentração da biomassa, do peso do meio de cultura (W), assim como da taxa específica de crescimento estão disponíveis em linha durante a fermentação. Contudo, nem a concentração de biomassa nem a taxa específica de crescimento podem ser medidas directamente, pelo que deverão ser estimadas através de outras variáveis de estado cuja medição em linha esteja disponível. Para tal recorreu-se a um observador assintótico e a um estimador baseado num observador para a determinação de X e de μ , respectivamente, utilizando a medição em linha das concentrações de oxigénio dissolvido e de oxigénio na corrente gasosa de saída do reactor. A Equação 5.17 e a Equação 5.18 correspondem ao observador assintótico e ao estimador baseado no observador, respectivamente, como anteriormente apresentadas no Capítulo 4.

$$\frac{d\hat{Z}}{dt} = -D\hat{Z} + \frac{1}{k_5} OTR \quad \text{Equação 5.17 (a)}$$

$$\hat{X} = \hat{Z} - \frac{1}{k_5} O \quad \text{Equação 5.17 (b)}$$

$$\frac{d\hat{\psi}}{dt} = \hat{\mu}\hat{X} - D\hat{\psi} - \frac{1}{k_5} OTR + \omega_{11}(\psi - \hat{\psi}) \quad \text{Equação 5.18 (a)}$$

$$\frac{d\hat{\mu}}{dt} = \omega_{21}(\psi - \hat{\psi}) \quad \text{Equação 5.18 (b)}$$

5.3.2. Sintonização do controlador

O desempenho e a estabilidade dos métodos de controlo de processos semi-contínuos são afectados pelo comportamento dinâmico destes e pela existência de ruído. Nas metodologias de controlo baseadas em modelos matemáticos, a discrepância entre o modelo e o comportamento real do processo conduz a desvios da acção de controlo pretendida, pelo que o desempenho do controlador poderá não ser satisfatório se os seus parâmetros não forem reajustados. A redução das incertezas e imprecisões aliada às variações temporais do processo incentivam o recurso a controladores inteligentes, ou seja, que permitam a adaptação automática dos seus parâmetros e assim conduzam a um comportamento mais robusto.

O objectivo de um processo fermentativo em operação semi-contínua não é alcançar a estabilidade do processo em si, mas optimizá-lo, mesmo que o seu comportamento se mantenha instável, embora sob controlo. Os parâmetros do controlador, γ_1 e γ_2 , determinam a sua velocidade

de convergência. O controlador será estável se γ_1 e γ_2 forem dois números estritamente positivos, atendendo à Equação 5.12. Duas situações podem ser tidas em conta: (i) sintonização dos parâmetros γ_1 e γ_2 admitindo que os seus valores são constantes durante todo o processo, por técnicas de optimização (Jenzsch *et al.*, 2006; Soons *et al.*, 2006) ou ajuste por tentativa erro, e (ii) recorrendo a uma sintonização automática dos referidos parâmetros, sendo estes variáveis com o tempo e ajustados com base no erro do controlador. A adaptação dos parâmetros de controlo ocorre devido à persistência de desvios em relação ao valor de referência para a taxa específica de crescimento. Na ausência de desvios, o desempenho do controlador é satisfatório estando-se na presença dos valores óptimos dos parâmetros do controlador.

Neste trabalho ambas as situações foram estudadas, sendo que na primeira optou-se pela abordagem de ajuste por tentativa erro, uma vez que Soons *et al.* (2006) verificaram que a escolha de parâmetros de controlo diferentes dos valores óptimos não conduz necessariamente a uma diminuição significativa do desempenho do controlador. No decurso do presente estudo, verificou-se que parâmetros de controlo inferiores aos utilizados e que permitiram a convergência, não implicaram uma diminuição acentuada do desempenho do controlador, embora valores elevados destes parâmetros tenham provocado oscilações importantes, em especial no início da fermentação.

No caso da sintonização automática foram utilizados três estratégias: uma regra de adaptação em função do erro relativo do controlador, a regra de MIT com recurso à sensibilidade do erro e a regra de adaptação em função do modelo de referência.

▪ **Regra de adaptação em função do erro relativo do controlador**

Um método possível para a sintonização automática dos parâmetros do controlador é admitir que a velocidade de adaptação dos mesmos é uma função do erro relativo, ε^{rel} , ou seja, do erro dividido pelo valor de referência:

$$\frac{d\gamma}{dt} = f(\varepsilon^{rel}) \quad \text{Equação 5.19}$$

Esta equação pode ser aproximada por uma série de Taylor na vizinhança de um dado valor pré-definido do erro relativo, ε_0^{rel} :

$$\frac{d\gamma}{dt} = f(\varepsilon_0^{rel}) + \left. \frac{df}{d\varepsilon^{rel}} \right|_{\varepsilon_0^{rel}} (\varepsilon^{rel} - \varepsilon_0^{rel}) + \left. \frac{d^2 f}{d(\varepsilon^{rel})^2} \right|_{\varepsilon_0^{rel}} \frac{(\varepsilon^{rel} - \varepsilon_0^{rel})^2}{2!} + \left. \frac{d^3 f}{d(\varepsilon^{rel})^3} \right|_{\varepsilon_0^{rel}} \frac{(\varepsilon^{rel} - \varepsilon_0^{rel})^3}{3!} + \dots \quad \text{Equação 5.20}$$

Quando ε_0^{rel} é igual a zero a série de Taylor resume-se à série de MacLaurin, conduzindo à seguinte expressão simplificada:

$$\frac{d\gamma}{dt} = \beta_0 + \beta_1 \varepsilon^{rel} + \beta_2 (\varepsilon^{rel})^2 + \beta_3 (\varepsilon^{rel})^3 + \dots \quad \text{Equação 5.21}$$

em que os valores de β_i ($i=0,1,2,\dots$) são constantes.

Se o desempenho do controlador for satisfatório e o erro for nulo então a adaptação não é necessária pelo que β_0 será também igual a zero. Atendendo a este facto e substituindo na equação anterior o ε^{rel} por $\frac{\varepsilon}{\mu^*}$ obtêm-se as seguintes expressões para a velocidade de adaptação dos parâmetros γ_1 e γ_2 , admitindo que os coeficientes β são os mesmos para os dois parâmetros e ignorando os termos de ordem 3 e superiores:

$$\frac{d\gamma_1}{dt} = \beta_1 \frac{\varepsilon}{\mu^*} + \beta_2 \left(\frac{\varepsilon}{\mu^*} \right)^2 = \beta_1 \frac{\mu^* - \mu}{\mu^*} + \beta_2 \left(\frac{\mu^* - \mu}{\mu^*} \right)^2 \quad \text{Equação 5.22 (a)}$$

$$\frac{d\gamma_2}{dt} = \beta_1 \frac{\varepsilon}{\mu^*} + \beta_2 \left(\frac{\varepsilon}{\mu^*} \right)^2 = \beta_1 \frac{\mu^* - \mu}{\mu^*} + \beta_2 \left(\frac{\mu^* - \mu}{\mu^*} \right)^2 \quad \text{Equação 5.22 (b)}$$

Deste modo, a sintonização automática do controlador adapta os seus parâmetros de um modo proporcional simultaneamente ao erro e ao erro quadrático.

- **Regra de MIT recorrendo à sensibilidade do erro**

A regra de MIT é um mecanismo de adaptação dos parâmetros do controlador utilizada no modelo de referência original de sistemas adaptativos (Aström e Wittenmark, 1989). Assumindo que se pretende que os parâmetros variem de modo que o erro entre as saídas do processo e do modelo de referência tenda para zero, pode escolher-se arbitrariamente uma função perda, como por exemplo,

$$J(\gamma) = \frac{1}{2} \varepsilon^2 \quad \text{Equação 5.23}$$

em que ε é o erro e γ os parâmetros do controlador.

Uma abordagem possível para minimizar a função $J(\gamma)$ é variar os parâmetros γ na direcção do gradiente negativo de J , ou seja,

$$\frac{d\gamma}{dt} = -\beta \frac{dJ}{d\gamma} = -\beta \varepsilon \frac{d\varepsilon}{d\gamma} \quad \text{Equação 5.24}$$

em que β é um parâmetro da sintonização ou do mecanismo de adaptação.

Assume-se que os parâmetros variam mais lentamente que as outras variáveis do sistema, permitindo que a derivada $\frac{d\varepsilon}{d\gamma}$ possa ser calculada admitindo que γ é constante. Esta derivada é designada por *derivada de sensibilidade* do erro do sistema relativamente ao parâmetro γ , ou sensibilidade do erro. A lei de sintonização dada pela Equação 5.24, onde $\frac{d\varepsilon}{d\gamma}$ é sensibilidade do erro, é em geral referida como a regra de MIT.

A regra de MIT pode ser aplicada em situações com vários parâmetros. No caso do presente trabalho, para dois parâmetros do controlador, γ_1 e γ_2 , a referida regra poderá ser escrita do seguinte modo:

$$\frac{d\gamma_1}{dt} = -\beta\varepsilon \frac{\partial\varepsilon}{\partial\gamma_1} \quad \text{Equação 5.25 (a)}$$

$$\frac{d\gamma_2}{dt} = -\beta\varepsilon \frac{\partial\varepsilon}{\partial\gamma_2} \quad \text{Equação 5.25 (b)}$$

Como se pode inferir da equação anterior a regra de MIT requer o conhecimento da sensibilidade do erro, o que em geral é difícil de obter. Por tal motivo, na literatura são propostas algumas aproximações para a derivada da sensibilidade do erro (Aström e Wittenmark, 1989).

Neste trabalho utilizou-se a regra de MIT recorrendo à sensibilidade do erro. Admita-se que a função erro é uma função dos parâmetros γ e do tempo t . Sabe-se que, se esta função é contínua em todo o seu domínio de definição e se as derivadas parciais de primeira e de segunda ordens são também contínuas nesse domínio, então as derivadas parciais cruzadas de segunda ordem são iguais (Piskounov, 1993). Assim sendo, pode assumir-se que a evolução temporal da derivada parcial do erro relativamente a γ iguala a derivada em relação a γ da derivada do erro relativamente ao tempo, para pequenas variações do erro (Smets *et al.*, 2002a):

$$\frac{\partial}{\partial t} \left(\frac{\partial\varepsilon}{\partial\gamma} \right) = \frac{\partial}{\partial\gamma} \left(\frac{\partial\varepsilon}{\partial t} \right) \quad \text{Equação 5.26}$$

A dinâmica do erro, $\varepsilon = \mu^* - \mu$, é obtida a partir do modelo de referência dado pela Equação 5.12:

$$\frac{d\varepsilon}{dt} = - \left(\gamma_1 \varepsilon + \int_0^t \gamma_2 \varepsilon dt \right) \quad \text{Equação 5.27}$$

A derivação parcial da equação anterior relativamente a γ_1 e γ_2 e subsequente integração em relação ao tempo (de acordo com a Equação 5.26) permite obter as expressões das sensibilidades do erro em relação a cada um daqueles dois parâmetros:

$$\frac{\partial \varepsilon}{\partial \gamma_1} = - \int_0^t \varepsilon dt \quad \text{Equação 5.28 (a)}$$

$$\frac{\partial \varepsilon}{\partial \gamma_2} = - \int_0^t \int_0^t \varepsilon dt dt \quad \text{Equação 5.28 (b)}$$

Combinando a Equação 5.28 (a) e (b) com a regra de MIT (Equação 5.25 (a) e (b)) obtém-se o seguinte mecanismo de adaptação:

$$\frac{d\gamma_1}{dt} = \beta \varepsilon \int_0^t \varepsilon dt \quad \text{Equação 5.29 (a)}$$

$$\frac{d\gamma_2}{dt} = \beta \varepsilon \int_0^t \int_0^t \varepsilon dt dt \quad \text{Equação 5.29 (b)}$$

O mecanismo de adaptação descrito pela equação anterior permite concluir que se um erro persistir durante um longo período de tempo a sensibilidade do erro aumentará (e portanto a adaptação do parâmetro de sintonização) até que o erro seja cancelado.

▪ **Regra de adaptação em função do modelo de referência**

Considerou-se ainda um outro mecanismo alternativo de sintonização. Supondo que a sensibilidade do erro é proporcional à sua evolução temporal desejada, então a Equação 5.28 (a) e (b) reduz-se a:

$$\frac{\partial \varepsilon}{\partial \gamma_1} = \frac{d\varepsilon}{dt} \quad \text{Equação 5.30 (a)}$$

$$\frac{\partial \varepsilon}{\partial \gamma_2} = \frac{d\varepsilon}{dt} \quad \text{Equação 5.30 (b)}$$

Atendendo ao modelo de referência e à Equação 5.27 derivada a partir daquele, obtém-se a seguinte expressão alternativa de sintonização para os parâmetros do controlador:

$$\frac{d\gamma_1}{dt} = \beta \varepsilon \left(\gamma_1 \varepsilon + \int_0^t \gamma_2 \varepsilon dt \right) \quad \text{Equação 5.31 (a)}$$

$$\frac{d\gamma_2}{dt} = \beta \varepsilon \left(\gamma_1 \varepsilon + \int_0^t \gamma_2 \varepsilon dt \right) \quad \text{Equação 5.31 (b)}$$

Esta equação permite inferir que a sintonização automática adapta os parâmetros de controlo γ_1 e γ_2 de modo proporcional ao erro ε e à evolução do erro com o tempo, que aumentará se o erro for grande e persistente.

5.4. RESULTADOS

O estudo apresentado nesta secção foi realizado a partir de simulações matemáticas. Foi desenvolvida uma rotina em MATLAB (versão 7.1) que permite a integração das equações do modelo do processo, a implementação do observador e do estimador que permitem estimar em linha a concentração da biomassa e da taxa específica de crescimento, respectivamente, assim como a geração do caudal de alimentação determinado pela lei de controlo. O método de Euler com um passo fixo de 0,01 horas foi o método de integração utilizado. Os valores iniciais das variáveis de estado encontram-se na Tabela 5.1. O tempo final considerado para todas as simulações foi de 25 horas. Os valores de *OTR* e *CTR* foram determinados considerando estado estacionário quer para o oxigénio dissolvido quer para o dióxido de carbono.

Tabela 5.1 Valores iniciais das variáveis de estado.

Variável de estado	Valor inicial
$X(0)$	1,5 g/kg
$S(0)$	0 g/kg
$A(0)$	0 g/kg
$C(0)$	0,272 g/kg
$O(0)$	0,0027 g/kg
$W(0)$	3 kg
S_e	250 g/kg

A validação dos algoritmos propostos foi efectuada recorrendo à integração numérica da Equação 5.2. Os valores “*pseudo* reais” das variáveis que podem ser medidas em linha, ou seja, de *O*, *OTR* e *W*, foram obtidos por simulação e depois corrompidos com ruído branco, de acordo com o desvio padrão calculado para as diferentes variáveis medidas neste processo, originando dados “*pseudo* experimentais”. De seguida, os algoritmos do observador e do estimador foram aplicados de modo a

obter as variáveis “estimadas” através dos dados “pseudo reais” correspondentes às variáveis medidas.

De modo a seleccionar a melhor alternativa para o projecto do controlador foram estudadas as diferentes estratégias de sintonização dos seus parâmetros apresentadas na secção anterior: mecanismos de sintonização com parâmetros constantes e com parâmetros variáveis. Neste último caso, consideraram-se três alternativas: regra de adaptação dos parâmetros γ_1 e γ_2 em função do erro relativo do controlador (Equação 5.22 (a) e (b), respectivamente), regra de MIT recorrendo à sensibilidade do erro para γ_1 e γ_2 (Equação 5.29 (a) e (b), respectivamente) e, por fim, a regra de adaptação em função do modelo de referência para γ_1 e γ_2 (Equação 5.31 (a) e (b), respectivamente). Na Tabela 5.2 encontram-se os valores iniciais dos parâmetros γ bem como os valores dos parâmetros β , tendo sido os utilizados em todas as simulações efectuadas.

Tabela 5.2 Valores iniciais dos parâmetros γ e β utilizados nas diferentes estratégias de sintonização estudadas.

Estratégia de sintonização	Valores iniciais					
	Parâmetros γ		Parâmetros β			
	γ_1	γ_2	β	β_1	β_2	
Constante	3	1	---	---	---	
Variável ou automática	$f(\varepsilon^{rel})$	3	1	---	10	20
	MIT com sensibilidade do erro	3	1	1	---	---
	adaptação em função do modelo de referência	3	1	1	---	---

Duas situações foram consideradas atendendo à Equação 5.16. Numa delas admitiu-se que, como é o caso deste trabalho, os valores de k_1 , $q_{s,máx}$, μ^* e K_S são conhecidos, embora lhes estejam associadas incertezas, pelo que a referida equação foi utilizada sem qualquer alteração. Uma outra situação foi tida em conta, onde se assumiu que esses valores eram desconhecidos e por isso foram incorporados nos parâmetros do controlador γ_1 e γ_2 , sendo a Equação 5.16 reescrita do seguinte modo:

$$F_{e,s} = \frac{k_1 \mu^* \hat{X}W}{S_e} + \left[\gamma_1 (\mu^* - \hat{\mu}) + \int_0^t \gamma_2 (\mu^* - \hat{\mu}) dt \right] \frac{W}{S_e} \quad \text{Equação 5.32}$$

O estudo realizado permitiu verificar que o conhecimento daqueles valores não demonstrou ser uma mais valia, pelo que, como se poderá constatar a seguir, foi adoptada a segunda situação, traduzida pela Equação 5.32.

O valor de referência escolhido para a taxa específica de crescimento foi de $0,1 \text{ h}^{-1}$ para a qual corresponde um crescimento da *E. coli* em regime oxidativo apenas com glucose e sem formação de acetato. Da Figura 5.4 à Figura 5.7 apresentam-se os resultados obtidos para as quatro estratégias de sintonização estudadas, que correspondem a simulações onde são gerados dados “pseudo reais” pela introdução de ruído branco nas variáveis medidas em linha (O , OTR e W) como já referido. Em cada uma dessas figuras indica-se a evolução com o tempo t , do erro médio, $E(t)$, bem como do erro de oscilação, $O(t)$, para os últimos n pontos (gráficos (d) e (f), respectivamente). Estes erros podem ser vistos como medidas do desempenho dos controladores, sendo dados pelas seguintes equações (Babuska *et al.*, 2003):

$$E(t) = \frac{1}{n} \sum_{t-n}^t |\varepsilon(t)| \quad \text{Equação 5.33}$$

$$O(t) = \frac{1}{n} \left[\sum_{t-n}^t |\varepsilon(t)| - \sum_{t-n}^t \varepsilon(t) \right] \quad \text{Equação 5.34}$$

A análise da Figura 5.4 à Figura 5.7 permite concluir que as quatro estratégias de sintonização estudadas apresentam um comportamento idêntico verificando-se que os erros médio e de oscilação diminuem com o tempo de modo semelhante. Em todos os casos constata-se uma boa convergência da variável controlada μ para o valor de referência (μ^*).

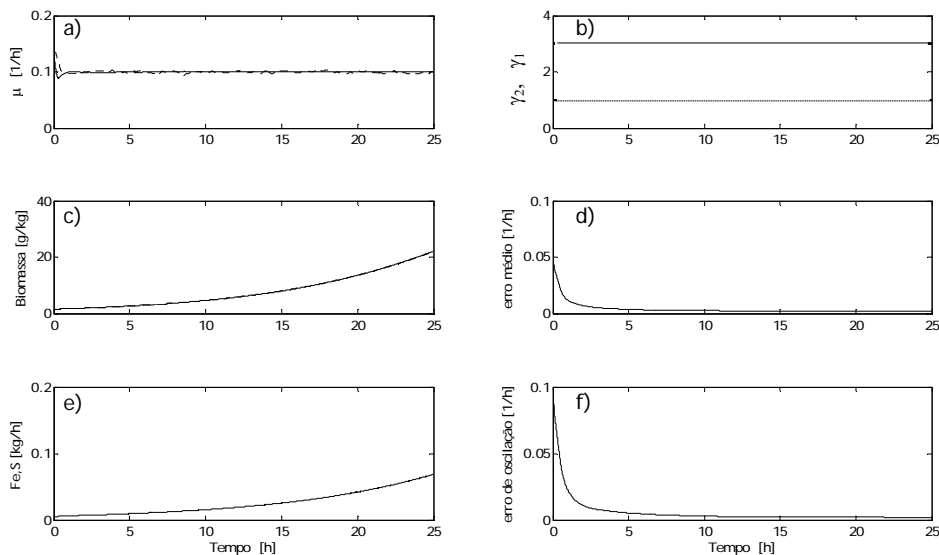


Figura 5.4 Resultados de simulação do controlo da taxa específica de crescimento com parâmetros de sintonização constantes. (a) Variável controlada: taxa específica de crescimento; (b) Parâmetros de sintonização; (c) Concentração de biomassa; (d) Erro médio do controlador; (e) Caudal de alimentação, $F_{e,s}$; (f) Erro de oscilação do controlador. Linha a tracejado corresponde aos valores estimados e linha a cheio aos valores simulados.

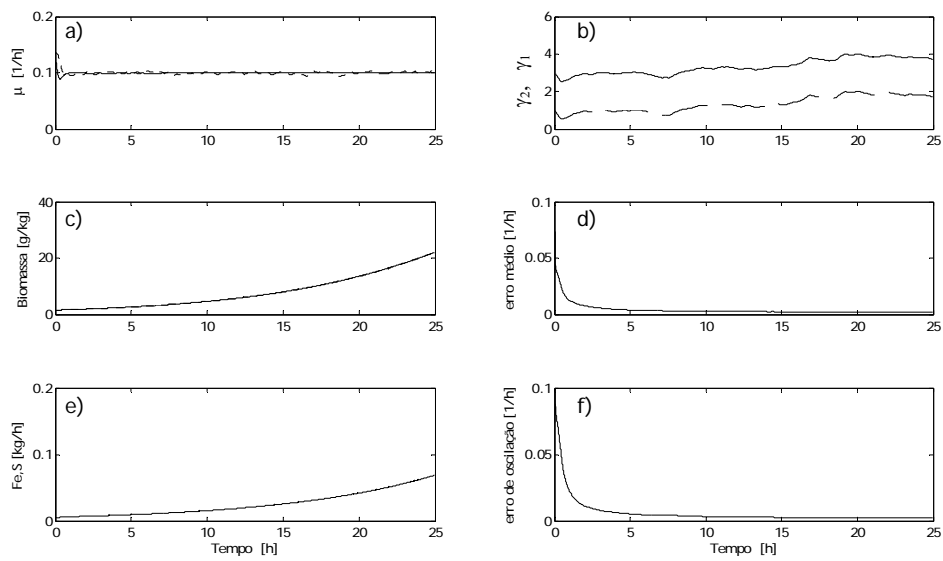


Figura 5.5 Resultados de simulação do controlo da taxa específica de crescimento com parâmetros de sintonização variáveis: regra de adaptação em função do erro relativo do controlador. (a) Variável controlada: taxa específica de crescimento; (b) Parâmetros de sintonização; (c) Concentração de biomassa; (d) Erro médio do controlador; (e) Caudal de alimentação, $F_{e,s}$; (f) Erro de oscilação do controlador. Linha a tracejado corresponde aos valores estimados e linha a cheio aos valores simulados.

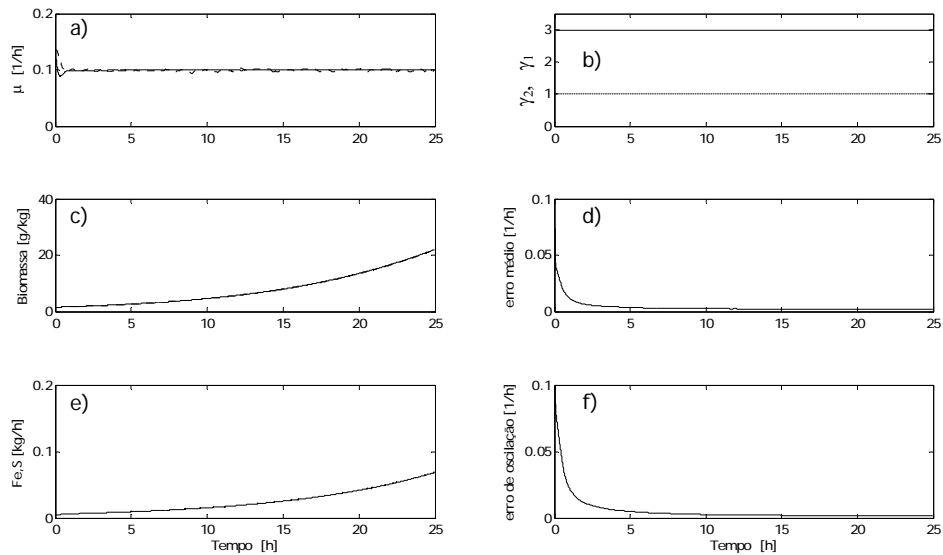


Figura 5.6 Resultados de simulação do controlo da taxa específica de crescimento com parâmetros de sintonização variáveis: regra de MIT recorrendo à sensibilidade do erro. (a) Variável controlada: taxa específica de crescimento; (b) Parâmetros de sintonização; (c) Concentração de biomassa; (d) Erro médio do controlador; (e) Caudal de alimentação, $F_{e,s}$; (f) Erro de oscilação do controlador. Linha a tracejado corresponde aos valores estimados e linha a cheio aos valores simulados.

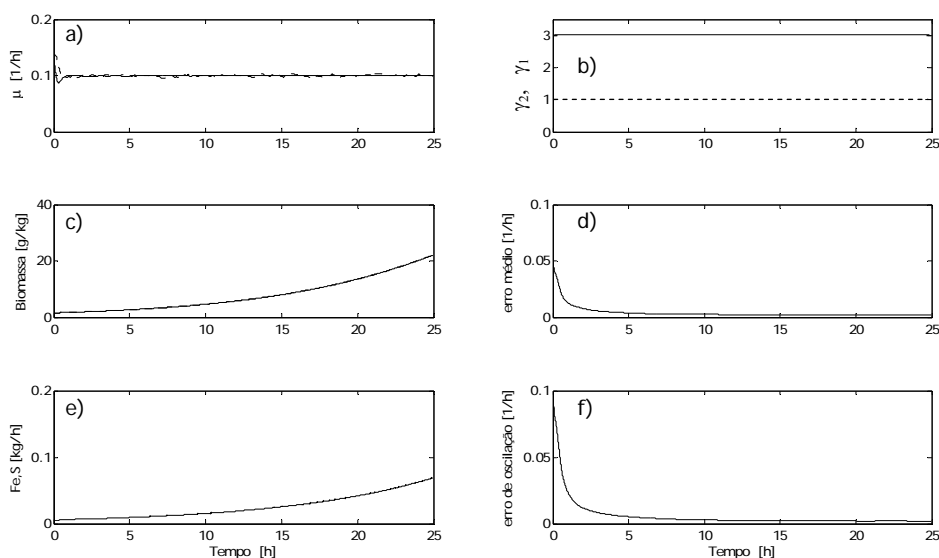


Figura 5.7 Resultados de simulação do controlo da taxa específica de crescimento com parâmetros de sintonização variáveis: regra de adaptação em função do modelo de referência. (a) Variável controlada: taxa específica de crescimento; (b) Parâmetros de sintonização; (c) Concentração de biomassa; (d) Erro médio do controlador; (e) Caudal de alimentação, $F_{e,S}$; (f) Erro de oscilação do controlador. Linha a tracejado corresponde aos valores estimados e linha a cheio aos valores simulados.

De modo a avaliar a robustez do controlador realizaram-se outras simulações nas quais o valor de referência da taxa específica de crescimento é alterado de $0,1 \text{ h}^{-1}$ para $0,05 \text{ h}^{-1}$ às 10 horas de fermentação, como se mostra na Figura 5.8. Esta situação poderá corresponder ao tempo de indução num processo de produção de proteínas recombinadas em que normalmente valores de μ mais baixos permitem alcançar uma maior produtividade. Convém referir que os valores iniciais dos parâmetros γ e β são os mesmos utilizados nas simulações anteriores e que se encontram na Tabela 5.2.

A observação da Figura 5.8 permite concluir que a sintonização automática recorrendo à regra de adaptação dos parâmetros γ_1 e γ_2 em função do erro relativo do controlador é a única estratégia que permite, num intervalo de tempo útil (inferior a 5 horas), atingir o novo valor de referência da taxa específica de crescimento e, a partir desse momento, controlar μ no valor pretendido (Figura 5.8 (b)). Por outro lado, as restantes estratégias estudadas, embora tendam para o novo valor de referência, não foram capazes de o alcançar ao fim de 15 horas, apresentando um desvio de aproximadamente 20% (Figura 5.8 (a), (c) e (d)).

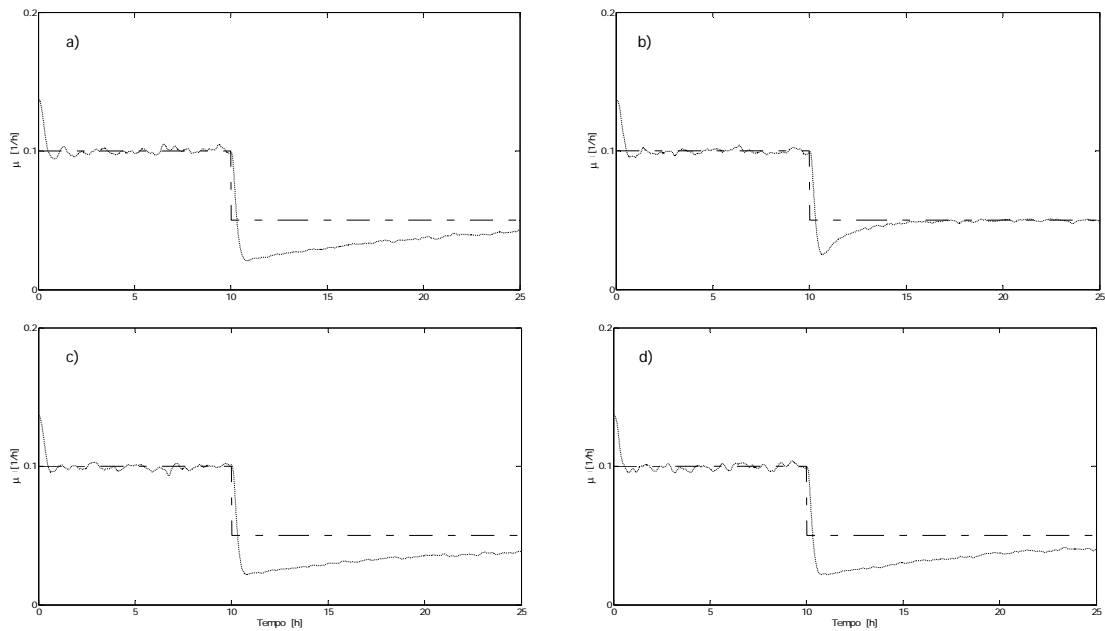


Figura 5.8 Resultados de simulação do controlo da taxa específica de crescimento quando alterado o valor de referência de 0,1 para 0,05 h⁻¹. (a) Parâmetros de sintonização constantes; (b) Parâmetros de sintonização variáveis de acordo com a regra de adaptação em função do erro relativo; (c) Parâmetros de sintonização variáveis de acordo com a regra de MIT recorrendo à sensibilidade do erro; (d) Parâmetros de sintonização variáveis de acordo com a regra de adaptação em função do modelo de referência.

Foi ainda avaliado o desempenho do controlador, tendo em conta as quatro estratégias de sintonização propostas, face a uma alteração de menos 20% na concentração de glucose no caudal de alimentação, introduzida após 5 horas de fermentação. Esta perturbação poderá ocorrer na prática devido à ocorrência de alterações ao nível do funcionamento da bomba de alimentação ou em virtude da composição da solução de alimentação poder sofrer degradação. Os resultados obtidos são apresentados na Figura 5.9. A análise desta figura permite inferir que, mais uma vez, apenas a sintonização automática recorrendo à regra de adaptação dos γ 's em função do erro relativo é a única estratégia que permite reagir satisfatoriamente à perturbação imposta, permitindo atingir de novo o valor de referência de μ pretendido.

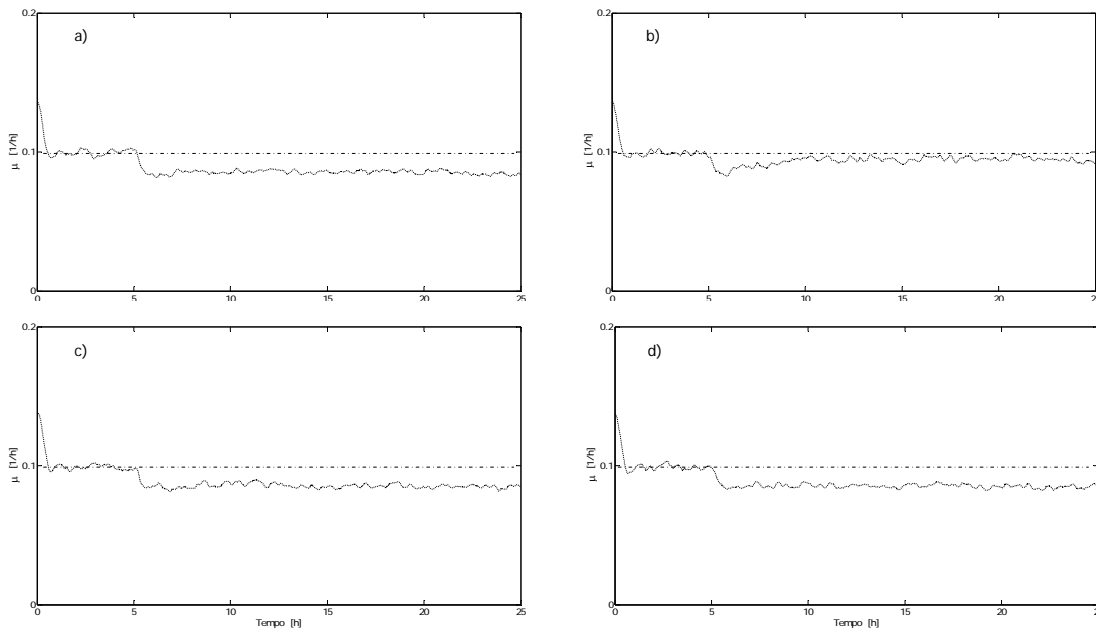


Figura 5.9 Resultados de simulação do controlo da taxa específica de crescimento quando o valor da concentração de glucose no caudal de alimentação diminui 20%, após 5 horas de fermentação.

(a) Parâmetros de sintonização constantes; Parâmetros de sintonização variáveis: (b) regra de adaptação em função do erro relativo; (c) regra de MIT recorrendo à sensibilidade do erro; (d) regra de adaptação em função do modelo de referência.

A robustez do controlador foi ainda estudada em relação à variação dos parâmetros do modelo, ou seja dos coeficientes de rendimento e dos parâmetros cinéticos. Duas situações foram analisadas: (i) os parâmetros do modelo mantiveram-se constantes com o tempo embora os seus valores nominais (Capítulo 3) tenham sido afectados de uma variação aleatória de $\pm 20\%$ e, (ii) os seus valores variaram com o tempo de acordo com um ruído imposto de $\pm 10\%$, face aos referidos valores nominais. Esta última situação pode ser visualizada na Figura 5.10 e na Figura 5.11, para a os coeficientes de rendimento e os parâmetros cinéticos, respectivamente.

É de referir que, os parâmetros do modelo afectados das variações anteriormente referidas, apenas foram utilizados nas equações do modelo. Por outro lado, no que diz respeito ao observador, estimador bem como ao controlador, utilizaram-se os valores constantes e iguais aos nominais para os parâmetros do modelo.

As simulações efectuadas tentam descrever duas situações que poderão ocorrer na prática em que, por um lado, na situação i) anteriormente referida, pretende-se simular o caso em que o algoritmo de controlo aqui proposto é aplicado ao estudo de uma fermentação na qual se utiliza uma estirpe de *E. coli* diferente da utilizada neste trabalho, pelo que os valores dos parâmetros do modelo serão diferentes dos valores nominais que aqui foram utilizados. Por outro lado, com a situação ii) pretende-se simular o que ocorre na prática, isto é, o facto de normalmente os valores

dos coeficientes de rendimento bem como os valores dos parâmetros cinéticos variarem com tempo. Os resultados obtidos podem ser observados na Figura 5.12 e na Figura 5.13, para a situação *i)*, e na Figura 5.14 e na Figura 5.15, para a situação *l)*.

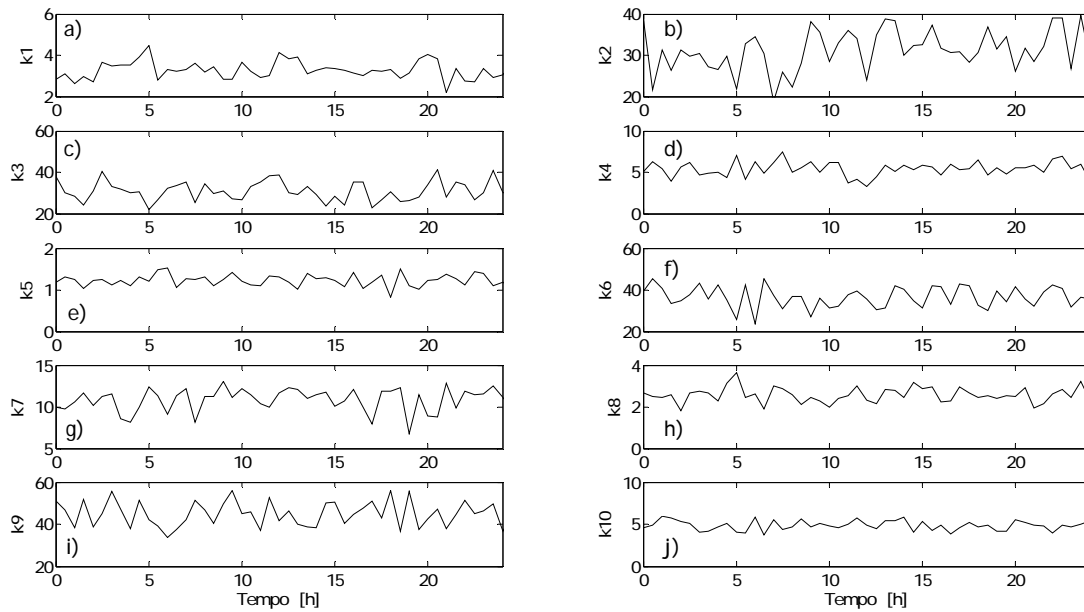


Figura 5.10 Variação com o tempo, por aplicação de ruído $\pm 10\%$, imposta aos coeficientes de rendimento durante a simulação para avaliar a robustez do algoritmo.

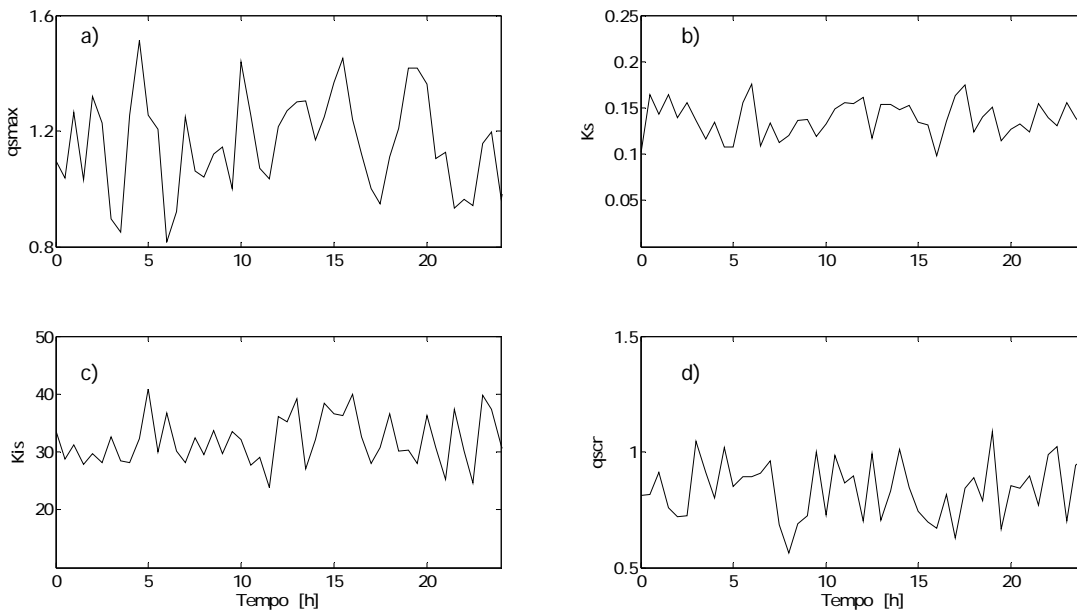


Figura 5.11 Variação com o tempo, por aplicação de ruído $\pm 10\%$, imposta aos coeficientes de rendimento durante a simulação para avaliar a robustez do algoritmo.

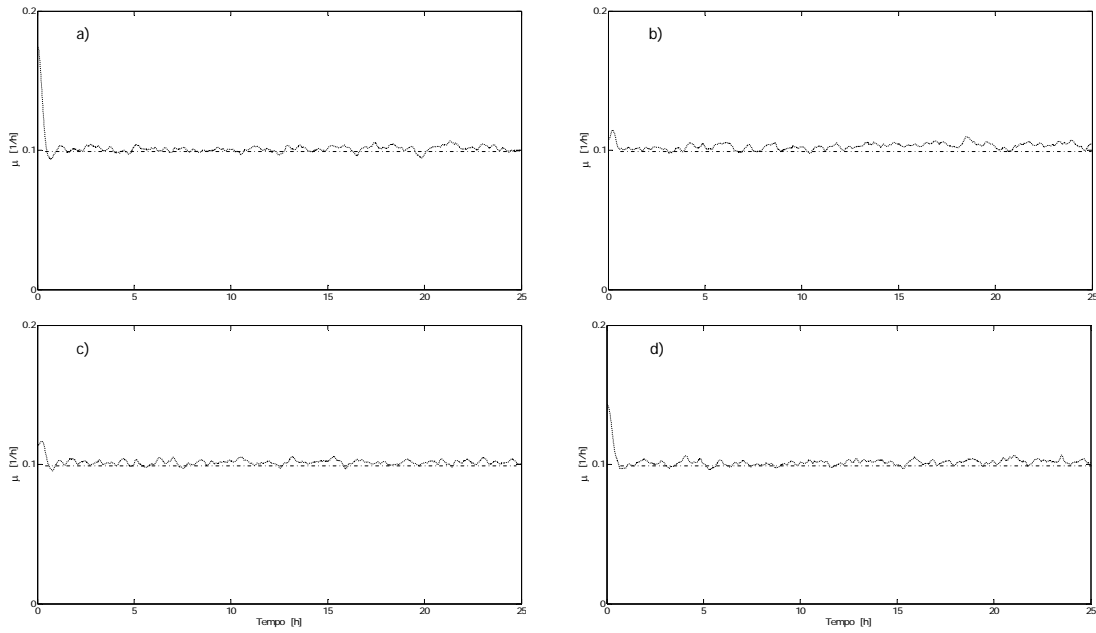


Figura 5.12 Resultados de simulação do controlo da taxa específica de crescimento quando os coeficientes de rendimento variam com o tempo com um ruído imposto de $\pm 10\%$. (a) Parâmetros de sintonização constantes; Parâmetros de sintonização variáveis: (b) regra de adaptação em função do erro relativo; (c) regra de MIT recorrendo à sensibilidade do erro; (d) regra de adaptação em função do modelo de referência.

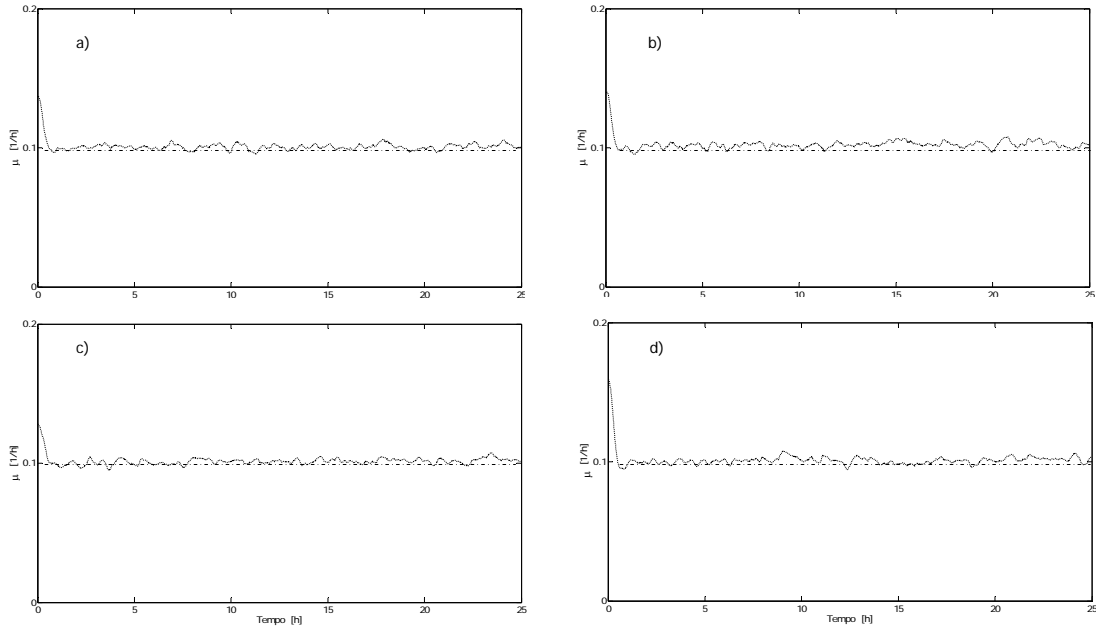


Figura 5.13 Resultados de simulação do controlo da taxa específica de crescimento quando os coeficientes de rendimento e parâmetros cinéticos variam com o tempo com um ruído imposto de $\pm 10\%$. (a) Parâmetros de sintonização constantes; Parâmetros de sintonização variáveis (b) regra de adaptação em função do erro relativo; (c) regra de MIT recorrendo à sensibilidade do erro; (d) regra de adaptação em função do modelo de referência.

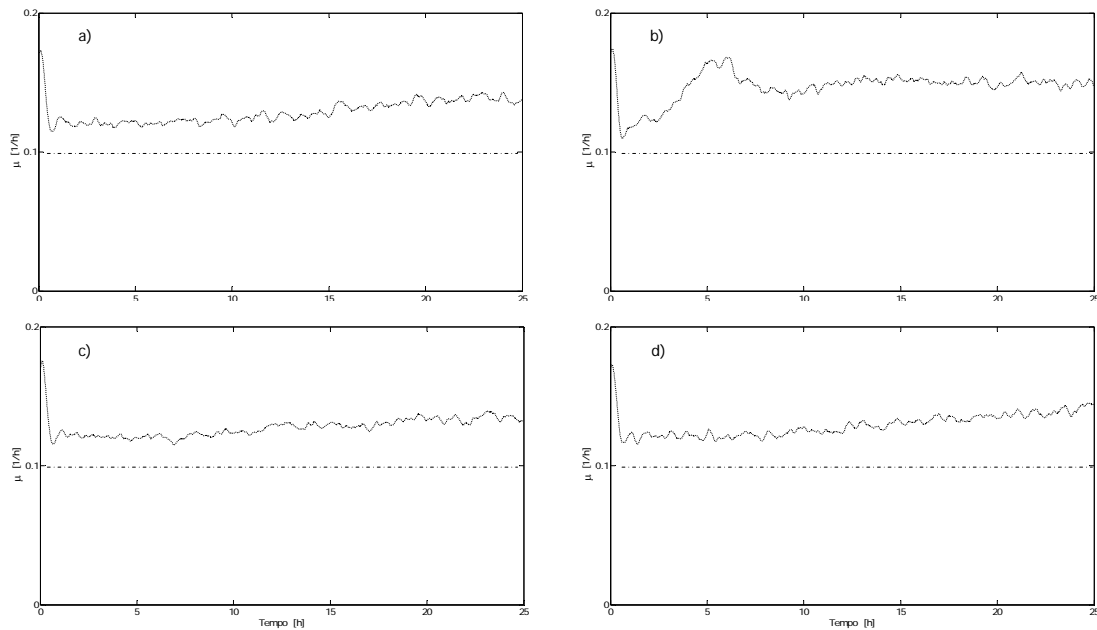


Figura 5.14 Resultados de simulação do controlo da taxa específica de crescimento quando os coeficientes de rendimento sofrem uma variação de $\pm 20\%$ dos seus valores nominais. (a) Parâmetros de sintonização constantes; Parâmetros de sintonização variáveis: (b) regra de adaptação em função do erro relativo; (c) regra de MIT recorrendo à sensibilidade do erro; (d) regra de adaptação em função do modelo de referência.

Como se pode inferir pela análise da Figura 5.12 e da Figura 5.13 verifica-se que, embora os valores dos parâmetros do modelo sejam variáveis no tempo, o algoritmo de controlo, para as quatro estratégias de sintonização adoptadas, apresenta um comportamento idêntico, permitindo controlar o valor da taxa específica de crescimento na vizinhança do valor de referência pretendido. No entanto, no caso em que uma variação constante face aos valores nominais dos parâmetros do modelo é imposta, isto é, para a situação λ , verifica-se que o algoritmo de controlo não é tão robusto, obtendo-se um desvio entre 30 e 40%. Contudo, apesar do desvio observado face ao valor de referência desejado, o valor de μ é mantido abaixo do valor para o qual, em geral, há formação de acetato (inferior a $0,25 \text{ h}^{-1}$ para a estirpe utilizada, embora este valor seja dependente da estirpe considerada e composição do meio utilizado).

Na Figura 5.15 são apresentados os resultados das simulações efectuadas onde para além de se ter imposto uma variação $\pm 20\%$ nos valores nominais dos parâmetros do modelo, foi ainda alterado o valor de referência de μ , após 10 horas de fermentação. Foi considerada esta situação, visto englobar os casos mais desfavoráveis apresentados anteriormente de modo a possibilitar uma avaliação mais efectiva do algoritmo de controlo. Nesta figura, para além do algoritmo de controlo proposto com as diferentes estratégias de sintonização adoptadas, são ainda estudados outros três

algoritmos de controlo: controlo adaptativo com modelo de referência de primeira ordem, controlo por antecipação e controlo PI clássico (Figura 5.15 (f), (g) e (h), respectivamente), cujas equações são as seguintes:

Para o controlo com modelo de referência de primeira ordem vem:

$$F_{e,s} = \frac{k_1 \mu^* \hat{X}W}{S_e} + \gamma_1 (\mu^* - \hat{\mu}) \frac{W}{S_e} \quad \text{Equação 5.35}$$

O controlo por antecipação é dado por:

$$F_{e,s} = \frac{k_1 \mu^* \hat{X}W}{S_e} \quad \text{Equação 5.36}$$

O controlo PI clássico é dado pela equação que se segue, em que $\bar{F}_{e,s}$ é o valor nominal de $F_{e,s}$:

$$F_{e,s} = \bar{F}_{e,s} + \gamma_1 (\mu^* - \hat{\mu}) + \int_0^t \gamma_2 (\mu^* - \hat{\mu}) dt \quad \text{Equação 5.37}$$

A Figura 5.15 (a) e a Figura 5.15 (b) mostram o desempenho do algoritmo desenvolvido neste trabalho considerando que os parâmetros γ_1 e γ_2 são constantes, embora a primeira corresponda à situação relativa à Equação 5.32 e a segunda à situação tendo em conta a Equação 5.16. Como se pode observar pela análise das referidas situações, não ocorreu uma melhoria significativa quando os parâmetros do modelo são utilizados. Uma vez que estes últimos parâmetros nem sempre são conhecidos e quando o são apresentam por vezes incertezas nos seus valores, a primeira abordagem considerada (Equação 5.32) afigura-se vantajosa, tendo sido esta a abordagem adoptada no desenho da lei de controlo, como já referido. Este facto foi ainda reforçado quando se consideraram estratégias de sintonização automática.

Na Figura 5.15 (c), (d) e (e) são apresentados os desempenhos do controlador considerado com sintonização automática dos seus parâmetros pela regra de adaptação em função do erro relativo, pela regra de MIT recorrendo à sensibilidade do erro e regra de adaptação em função do modelo de referência, respectivamente. Mais uma vez, o comportamento do controlador com sintonização em função do erro relativo demonstrou ser mais estável à alteração do valor de referência da variável controlada apesar do ruído das variáveis medidas e da alteração em $\pm 20\%$ do valor nominal dos parâmetros do modelo (coeficientes de rendimento e coeficientes cinéticos).

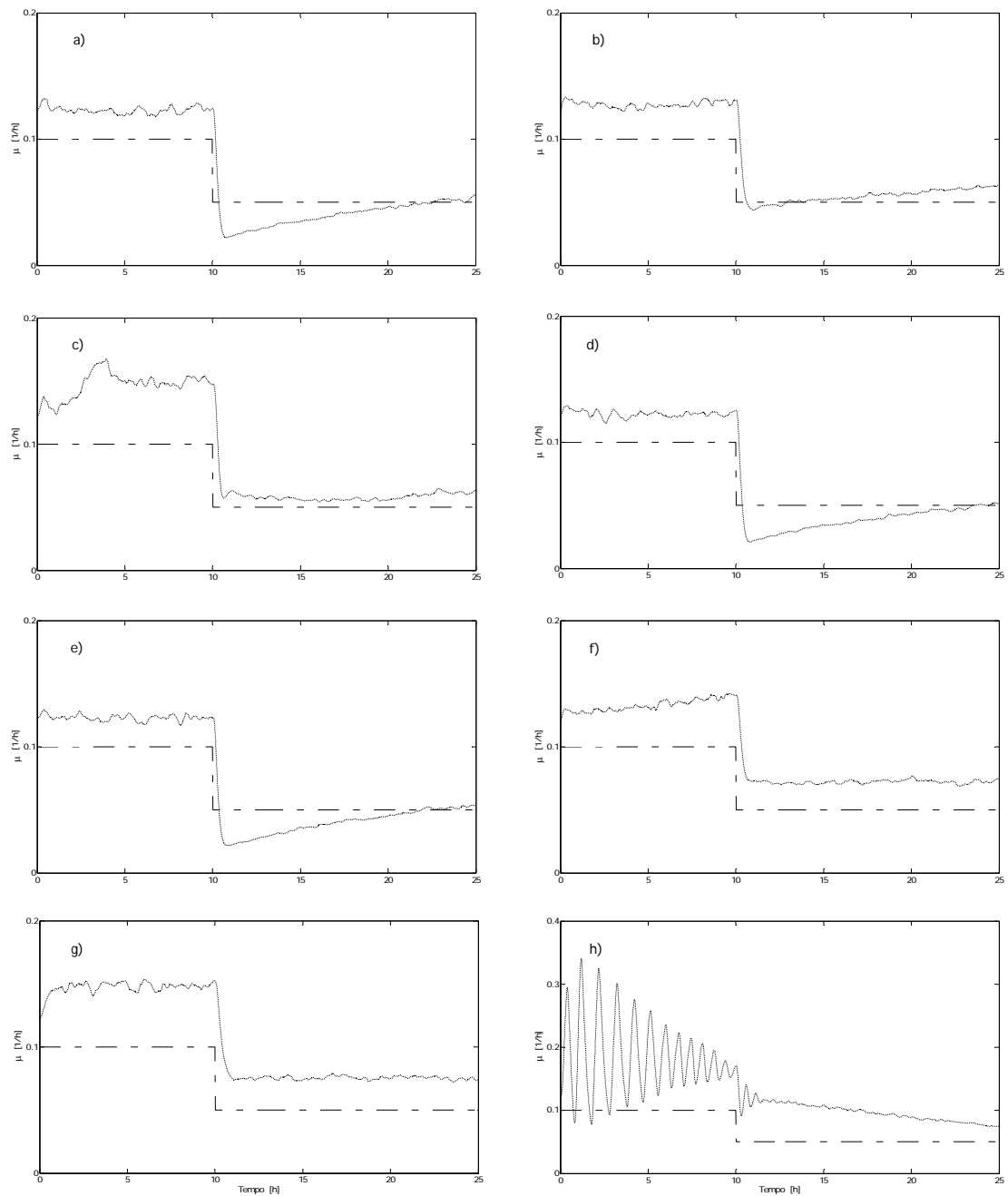


Figura 5.15 Comparação do desempenho de vários algoritmos de controlo da taxa específica de crescimento. Parâmetros de sintonização constantes: (a) algoritmo de controlo dado pela Equação 5.32; (b) algoritmo de controlo dado pela Equação 5.16. Parâmetros de sintonização variáveis: (c) regra de adaptação em função do erro relativo; (d) regra de MIT recorrendo à sensibilidade do erro; (e) regra de adaptação em função do modelo de referência. (f) controlo adaptativo com modelo de referência de primeira ordem; (g) controlo por antecipação; (h) controlo PI clássico.

O algoritmo de controlo adaptativo com modelo de referência proposto traduz uma lei de controlo por antecipação/retroacção cuja acção de retroacção é do tipo PI. O desempenho deste controlador (Figura 5.15 (c), (d) ou (e) consoante o mecanismo de sintonização adoptado) foi comparado com três algoritmos de controlo diferentes dados pelas equações 5.35 à 5.37 (Figura 5.15 (f), (g) e (h), respectivamente). O algoritmo de controlo com modelo de referência de primeira ordem dado pela Equação 5.35, embora semelhante ao desenvolvido no presente trabalho, apenas apresenta uma acção por retroacção do tipo P. Este caso conduz a um pior desempenho do controlador não havendo convergência para o novo valor de referência, como se pode inferir pela observação da Figura 5.15 (f). Este comportamento era esperado uma vez que a acção P não é capaz de lidar com o desvio final, contrariamente à acção PI. Assim sendo, a acção integral parece ser de grande relevância quando há perturbações externas e incertezas no modelo matemático. Deste modo, pode concluir-se que a escolha das características do modelo de referência (primeira ou segunda ordem) é um factor importante para o desempenho do controlador da taxa específica de crescimento. Na Figura 5.15 (g) é apresentado o desempenho do controlador por antecipação. A aplicação deste algoritmo (Equação 5.36) implica em geral o conhecimento da concentração de biomassa e do peso do meio de fermentação. A estratégia mais comum para obter esses valores é estimá-los baseando-se nos valores iniciais do caudal de alimentação e assumindo um crescimento puramente exponencial. Contudo, perturbações externas juntamente com variações naturais do peso do meio causadas por exemplo devido à evaporação são uma fonte de erro importante. Por outro lado, esta metodologia pode não controlar adequadamente o processo devido às incertezas nos parâmetros do modelo, como por exemplo os coeficientes de rendimento, que têm de ser conhecidos de modo a calcular o caudal de alimentação e cujos valores poderão ainda variar ao longo do processo. Assim, a estratégia de alimentação exponencial é por vezes compensada incorporando uma acção apropriada de controlo por retroacção. O algoritmo de controlo dado pela Equação 5.36 tem em conta essa acção de retroacção devido à incorporação da estimação em linha da concentração de biomassa e da medida em linha do peso do meio de fermentação. Verificou-se que o comportamento não foi muito diferente do obtido com o controlador adaptativo com modelo de referência de primeira ordem (Figura 5.15 (g) e (f), respectivamente) o que vem realçar o facto de na presença de perturbações e incertezas no modelo a acção proporcional não é suficiente para alcançar a acção de controlo desejada. Foram ainda comparados os desempenhos do controlador proposto neste trabalho (Figura 5.15 (c), (d) ou (e)) e do controlador PI clássico de parâmetros constantes (Figura 5.15 (h)). Como se pode inferir facilmente, este último controlador não consegue uma qualidade de controlo satisfatória apresentando um comportamento muito oscilante, em especial no início da fermentação, situação que se encontra de acordo com o descrito por outros autores (Jenzsch *et al.*, 2006; Soons *et al.*, 2006). Este facto, vem realçar a importância da inclusão do termo nominal adaptado pela introdução do valor estimado da concentração de biomassa e do valor medido em linha do peso do meio de cultura.

5.5. SÍNTESE

O principal objectivo deste capítulo consistiu em desenvolver um controlador adaptativo com modelo de referência traduzindo-se num controlo por antecipação/retroacção cuja acção de retroacção é do tipo PI, para controlar a taxa específica de crescimento da *E. coli* numa fermentação semi-contínua. O controlador utiliza os valores estimados da concentração de biomassa obtidos através de um observador assintótico, os valores estimados da taxa específica de crescimento obtidos por um estimador baseado num observador e o valor, medido em linha, do peso do meio de fermentação. Para estimar a concentração de biomassa e da taxa específica de crescimento foi apenas necessário ter acesso em linha aos valores da concentração de oxigénio dissolvido, do *OTR* e do peso do meio para além de conhecer com relativa precisão os coeficientes de rendimento do processo.

A robustez do algoritmo de controlo para a taxa específica de crescimento foi estudada por simulação numérica gerando dados "pseudo reais", por aplicação de um ruído branco às variáveis medidas em linha, por alteração do valor de referência, por alteração do valor da concentração da glucose no caudal de alimentação e variando os valores nominais dos parâmetros do modelo. O estudo realizado permite concluir que a resposta do controlador é, em geral, satisfatória, sendo capaz de manter o valor da taxa específica de crescimento na vizinhança do valor de referência pretendido e inferior a um valor que conduz à formação de acetato, revestindo-se este facto de grande importância numa situação real, em especial, numa fermentação cujo objectivo seja a produção, nomeadamente de proteínas recombinadas.

Diferentes métodos de sintonização dos parâmetros do controlador foram analisados, podendo concluir-se que, em geral, o método de sintonização automática com recurso à regra de adaptação dos parâmetros γ_1 e γ_2 em função do erro relativo do controlador foi o que apresentou um melhor desempenho global. Este mecanismo de sintonização automática demonstrou capacidade para melhorar o desempenho do controlador ajustando continuamente os seus parâmetros. Demonstrou ainda uma maior robustez aos valores iniciais dos parâmetros de sintonização.

Contudo, novos estudos de simulação deverão ser realizados com o intuito de melhorar o procedimento de sintonização uma vez que se verificou que o desempenho do algoritmo de controlo é sensível aos valores dos seus parâmetros. Assim sendo, poder-se-á recorrer a um processo de optimização com vista à determinação dos melhores valores iniciais dos parâmetros do controlador, γ_1 e γ_2 , bem como dos valores de β , que são os valores das expressões das regras de sintonização utilizadas.

O algoritmo de controlo aqui apresentado, embora semelhante a outros descritos na literatura (Jenzsch *et al.*, 2006; Soons *et al.*, 2006) tem a vantagem não só de não requerer o conhecimento do modelo cinético e respectivos parâmetros como também de necessitar de um menor número de variáveis medidas em linha.

Apesar da qualidade dos resultados obtidos no estudo por simulação numérica aqui apresentado, de futuro, será pertinente implementar na prática o algoritmo de controlo proposto de modo a avaliar o seu desempenho numa situação real, quer numa situação onde haja apenas crescimento da *E. coli*, quer numa situação em que haja produção, por exemplo, de uma proteína recombinada.

5.6. BIBLIOGRAFIA

Aguilar, R., González, J., Barrón, M., Martínez-Guerra, R. e Maya-Yescas, R., (2001). Robust PI² controller for continuous bioreactors. *Process Biochemistry*, 36, 1007-1013.

Akay, B., Ertunç, C., Kahvecioglu, A., Hapoglu, H. e Alpbaz, M., (2002). Adaptive control of *S. cerevisiae* production. *Food and Bioprocess Processing*, 80, 28-38.

Akesson, M., Hagander, P. e Axelsson, J.P., (2001a). Avoiding acetate accumulation in *Escherichia coli* cultures using feedback control of glucose feeding. *Biotechnology and Bioengineering*, 73 (3), 223-230.

Akesson, M., Hagander, P. e Axelsson, J.P., (2001b). Probing control of fed-batch cultivations: analysis and tuning. *Control Engineering Practice*, 9 (7), 709-723.

Alford, J.S., (2006). Bioprocess control: Advances and challenges. *Computers and Chemical Engineering*, 30, 1464-1475.

Alvarez, A. e Simutis, R., (2004). Application of Kalman filter algorithm in GMC control strategy for fed-batch cultivation process. *Information Technology and Control*, 1 (30), 7-12.

Åström, K.J., Goodwin, G.C. e Kumar, P.R., (1995). *Adaptive control, filtering, and signal processing*, Springer-Verlag, Nova Iorque.

Aström, K.J. e Wittenmark, B., (1989). *Adaptive control*, Addison-Wesley, Reading.

Babuska, R., Damen, M.R., Hellinga, C. e Maarleveld, H., (2003). Intelligent adaptive control of bioreactors. *Journal of Intelligent Manufacturing*, 14, 255-265.

Bastin, G. e Dochain, D., (1990). *On-line estimation and adaptive control of bioreactors*, Elsevier Science Publishers, Amsterdam.

Chang, D., (2003). The snowball effect in fed-batch bioreactors. *Biotechnology Progress*, 19, 1064-1070.

Chen, L., Bastin, G. e Van Breusegem, V., (1995). A case study of adaptive nonlinear regulation of fed-batch biological reactors. *Automatica*, 31 (1), 55-65.

Cheng, L.-C., Wu, J.-Y. e Chen, T.-L., (2002). A pseudo-exponential feeding method for control of specific growth rate in fed-batch cultures. *Biochemical Engineering Journal*, 10 (227), 232.

Chung, Y.-C., Chien, I.-L. e Chang, D.-M., (2006). Multiple-model control strategy for fed-batch high cell-density culture processing. *Journal of Process Control*, 16, 9-26.

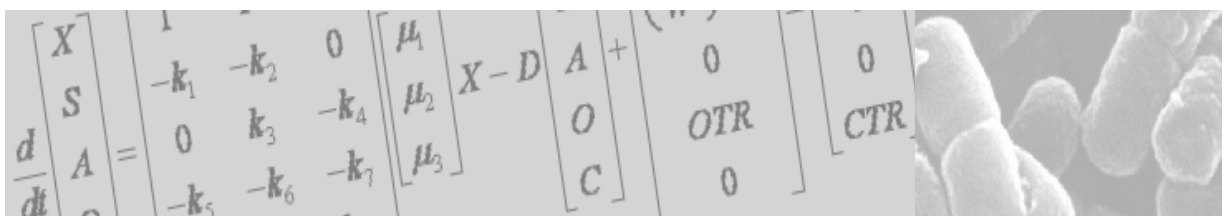
- Costa, A.C., Alves, T.L.M., Henriques, A.W.S., Maciel, R. e Lima, E.L., (1998). An adaptive optimal control scheme based on hybrid neural modelling. *Computers & Chemical Engineering*, 22, S859-S862.
- Curless, C., Pope, J. e Tsai, L., (1990). Effect of preinduction specific growth rate on recombinant alpha consensus interferon synthesis in *Escherichia coli*. *Biotechnology Progress*, 6, 149-152.
- Delisa, M.P., Chae, H.J., Weigand, W.A., Valdes, J.J., Rao, G. e Bentley, W.E., (2001). Generic model control of induced protein expression in high cell density cultivation of *Escherichia coli* using on-line GFP-fusion monitoring. *Bioprocess and Biosystems Engineering*, 24 (2), 83-91.
- Dochain, D., (2001). State observation and adaptive linearizing control for distributed parameter (bio)chemical reactors. *International Journal of Adaptive Control and Signal Processing*, 15 (6), 633-653.
- Ferreira, E.C. e Feyeo de Azevedo, S., (1996). A new adaptive scheme for the adaptive linearizing control of bioprocesses. *2nd portuguese conference on automatic control*, 153-157.
- Frahm, B., Lane, P., Atzert, H., Munack, A., Hoffmann, M., Hass, V.C. e Portner, R., (2002). Adaptive, model-based control by the Open-Loop-Feedback-Optimal (OLFO) controller for the effective fed-batch cultivation of hybridoma cells. *Biotechnology Progress*, 18 (5), 1095-1103.
- Gordillo, M.A., Sanz, A., Sanchez, A., Valero, F., Montesinos, J.L., Lafuente, J. e Sola, C., (1998). Enhancement of *Candida rugosa* lipase production by using different control fed-batch operational strategies. *Biotechnology and Bioengineering*, 60 (2), 156-168.
- Gregory, M.E. e Turner, C., (1993). Open-loop control of specific growth rate in fed-batch cultures of recombinant *E. coli*. *Biotechnology Techniques*, 7 (12), 889-894.
- Jadot, F., Bastin, G. e Van Impe, J.F., (1998). Optimal adaptive control of a bioprocess with yield-productivity conflict. *Journal of Biotechnology*, 65 (1), 61-68.
- Jana, S. e Deb, J.K., (2005). Strategies for efficient production of heterologous proteins in *Escherichia coli*. *Applied Microbiology and Biotechnology*, 67, 289-298.
- Jenzsch, M., Simutis, R. e Lübbert, A., (2004). Application of model predictive control to cultivation processes for protein production with genetically modified bacteria (CAB9 - 9th International Symposium Computer Applications in Biotechnology (Nancy, France))
- Jenzsch, M., Simutis, R. e Luebbert, A., (2006). Generic model control of the specific growth rate in recombinant *Escherichia coli* cultivations. *Journal of Biotechnology*, 12, 483-493.
- Jobe, A.M., Herwig, C., Surzyn, M., Walker, B., Marison, I. e von Stockar, U., (2003). Generally applicable fed-batch culture concept based on the detection of metabolic state by on-line balancing. *Biotechnology and Bioengineering*, 82 (6), 627-639.
- Johnston, W., Cord-Ruwisch, R. e Cooney, M.J., (2002). Industrial control of recombinant *E. coli* fed-batch culture: new perspectives on traditional controlled variables. *Bioprocess and Biosystems Engineering*, 25 (2), 111-120.
- Komives, C. e Parker, R.S., (2003). Bioreactor state estimation and control. *Current Opinion in Biotechnology*, 14 (5), 468-474.
- Korz, D.J., Rinas, U., Hellmuth, K., Sanders, E.A. e Deckwer, W.-D., (1995). Simple fed-batch technique for high cell density cultivation of *Escherichia coli*. *Journal of Biotechnology*, 39 (1), 59-65.

- Lee, J., Lee, S.Y., Park, S. e Middelberg, A., (1999). Control of fed-batch fermentations. *Biotechnology Advances*, 17 (1), 29-48.
- Lee, S.B. e Bailey, J.E., (2002). Analysis of growth rate effects on productivity of recombinant *Escherichia coli* populations using molecular mechanism models. *Biotechnology and Bioengineering*, 79 (5), 550-557.
- Levisauskas, D., (2001). Inferential control of the specific growth rate in fed-batch cultivation processes. *Biotechnology Letters*, 23, 1189-1195.
- Levisauskas, D., Simutis, R., Borvitz, D. e Lübert, A., (1996). Automatic control of the specific growth rate in fed-batch cultivation processes based on an exhaust gas analysis. *Bioprocess Engineering*, 15, 145-150.
- Lim, H.C. e Lee, K.-S., (1990). Control of Bioreactor Systems. Em: Karl Schügerl (ed), *Biotechnology - Measuring, Modelling and Control*. pp. 511-560, VCH, Weinheim.
- Lim, H.-K. e Jung, K.-H., (1998). Improvement of heterologous protein productivity by controlling postinduction specific growth rate in recombinant *Escherichia coli* under control of the P_L promoter. *Biotechnology Progress*, 14, 548-553.
- Lübert, A. e Jorgensen, S.B., (2001). Bioreactor performance: a more scientific approach for practice. *Journal of Biotechnology*, 85, 187-212.
- Mailleret, L., Bernard, O. e Steyer, J.-P., (2004). Nonlinear adaptive control for bioreactors with unknown kinetics. *Automatica*, 40, 1379-1385.
- Martinez, A., Ramirez, O. e Valle, F., (1998). Effect of growth rate on the production of β -galactosidase from *Escherichia coli* in *Bacillus subtilis* using glucose-limited exponentially fedbatch cultures. *Enzyme and Microbial Technology*, 22, 520-526.
- Narendra, K.S. e Annaswamy, A.M., (1989). *Stable adaptive systems*, Prentice-Hall, Englewood Cliffs.
- Pinsach, J., de Mas, C. e López-Santín, J., (2006). A simple feedback control of *Escherichia coli* growth for recombinant aldolase production in fed-batch mode. *Biochemical Engineering Journal*, 29, 235-242.
- Piskounov, N., (1993). *Cálculo Diferencial e Integral*, Edições Lopes da Silva, Porto.
- Rani, K.Y. e Rao, V.S.R., (1999). Control of fermenters - a review. *Bioprocess Engineering*, 21 (1), 77-88.
- Rocha, I. e Ferreira, E.C., (2002b). An integrated system for advanced monitoring and control of fed-batch fermentations of recombinant *E. coli*. *Computer Applications in Biotechnology 2001*, 349-354.
- Rocha, I. e Ferreira, E.C., (2002a). Model-based adaptive control of acetate concentration during the production of recombinant proteins with *E. coli* (IFAC - 15th Triennial World Congress (Barcelona, Spain))
- Sanden, A.M., Prytz, I., Tubulekas, I., Forberg, C., Le, H., Hektor, A., Neubauer, P., Pragai, Z., Harwood, C., Ward, A., Picon, A., de Mattos, J.T., Postma, P., Farewell, A., Nystrom, T., Reeh, S., Pedersen, S. e Larsson, G., (2003). Limiting factors in *Escherichia coli* fed-batch production of recombinant proteins. *Biotechnology and Bioengineering*, 81 (2), 158-166.

- Sastry, S. e Bodson, M., (1989). *Adaptive control. Stability, convergence, and robustness*, Prentice-Hall, Inc, Englewood Cliffs.
- Shin, C.S., Honh, M.S., Bae, C.S. e Lee, J., (1997). Enhanced production of human mini-proinsulin in fed-batch cultures at high cell density of *Escherichia coli* BL21 (DE3) [pET-3aT2M2]. *Biotechnology Progress*, 13, 249-257.
- Siegwart, P., Cote, J., Male, K., Luong, J.H.T., Perrier, M. e Kamen, A., (1999). Adaptive control at low glucose concentration of Hek-293 cell serum-free cultures. *Biotechnology Progress*, 15 (4), 608-616.
- Smets, I., Bernaerts, K., Sun, J., Marchal, K., Vanderleyden, J. e Van Impe, J.F., (2002a). Sensitivity function-based model reduction. A bacterial gene expression case study. *Biotechnology and Bioengineering*, 80 (2), 195-200.
- Smets, I.Y., Bastin, G.P. e Van Impe, J.F., (2002b). Feedback stabilization of fed-batch bioreactors: Non-monotonic growth kinetics. *Biotechnology Progress*, 18 (5), 1116-1125.
- Smets, I.Y., Claes, J.E., November, E.J., Bastin, G.P. e Van Impe, J.F., (2004). Optimal adaptive control of (bio)chemical reactors: past, present and future. *Journal of Process Control*, 14 (7), 795-805.
- Soons, Z.I.T.A., Voogt, J.A., Van Straten, G. e Van Boxtel, A.J.B., (2006). Constant specific growth rate in fed-batch cultivation of *Bordetella pertussis* using adaptive control. *Journal of Biotechnology*, 125 (2), 252-268.
- Stephanopoulos, G., (1984). *Chemical Process Control: An Introduction to Theory and Practice*, Prentice-Hall International (USA), New Jersey.
- Suárez, D.C. e Kilikian, B.V., (2000). Acetic acid accumulation in aerobic growth of recombinant *Escherichia coli*. *Process Biochemistry*, 35 (9), 1051-1055.
- Szederkenyi, G., Kristensen, N.R., Hangos, K.M. e Jorgensen, S.B., (2002). Nonlinear analysis and control of a continuous fermentation process. *Computers & Chemical Engineering*, 26 (4-5), 659-670.
- Van Impe, J.F. e Bastin, G., (1995). Optimal Adaptive-Control of Fed-Batch Fermentation Processes. *Control Engineering Practice*, 3 (7), 939-954.
- Whiffin, V.S., Cooney, M.J. e Cord-Ruwisch, R., (2004). Online detection of feed demand in high cell density cultures of *Escherichia coli* by measurement of changes in dissolved oxygen transients in complex media. *Biotechnology and Bioengineering*, 85 (4), 422-433.
- Yee, L. e Blanch, H.W., (1993). Recombinant trypsin production in high cell density fed-batch cultures in *Escherichia coli*. *Biotechnology and Bioengineering*, 41 (8), 781-790.
- Zhang, T., Guay, M. e Dochain, D., (2003). Adaptive extremum seeking control of continuous stirred tank bioreactors. *AIChE Journal*.

CAPÍTULO 6

CONCLUSÕES GERAIS E SUGESTÕES PARA TRABALHO FUTURO



6. CONCLUSÕES GERAIS E SUGESTÕES PARA TRABALHO FUTURO

O estudo realizado e apresentado nesta dissertação pretende ser um contributo no desenvolvimento de algoritmos matemáticos que permitam simplificar e melhorar o processo de identificação, a monitorização em tempo real e o controlo de bioprocessos, nomeadamente de uma fermentação semi-contínua de alta densidade celular de *Escherichia coli*.

Foi considerada uma abordagem matemática baseada em modelos determinísticos com uma estrutura global para reactores biológicos, o que permite generalizar os algoritmos propostos a outros processos.

O desenho óptimo de experiências baseado na Matriz de Informação de Fisher em conjugação com a metodologia estabelecida para a identificação dos coeficientes de rendimento do modelo do processo em estudo, permitiram atingir o objectivo central deste trabalho, identificando esses parâmetros e incrementando a sua precisão, pela aplicação de estratégias mais simples que as abordadas normalmente na literatura. É de destacar ainda, a tentativa de se obter algoritmos, em especial para monitorização e controlo, independentes do modelo cinético, mas sustentados no conhecimento dos coeficientes de rendimento. Este facto, reveste-se de grande relevância sendo uma vantagem inegável em virtude da incerteza dos modelos utilizados na descrição dos bioprocessos, residir essencialmente nos seus modelos cinéticos.

Neste capítulo são apresentadas as principais conclusões do trabalho realizado e ainda propostas algumas sugestões de trabalho futuro.

6.1. CONCLUSÕES GERAIS

Os principais objectivos deste trabalho, ou seja, o desenvolvimento de estratégias de alimentação de substrato para a identificação dos coeficientes de rendimento e de algoritmos que permitam a monitorização e controlo do processo de crescimento de *E. coli* em operação semi-contínua, foram alcançados com sucesso. De entre os resultados obtidos convém destacar as seguintes conclusões:

O desenho óptimo de experiências, aliado à metodologia de identificação dos coeficientes de rendimento (k), mostrou-se adequado, permitindo determinar o valor desses parâmetros sem o conhecimento preciso do modelo cinético do processo em estudo.

É de salientar a maior simplicidade da metodologia adoptada no cálculo dos valores de k a partir dos valores de θ , obtidos por aplicação do *MRLM* (Modelo de Regressão linear Múltipla), face a outras alternativas descritas na literatura, permitindo verificar a identificabilidade teórica e numérica do modelo de um modo simples.

Demonstrou-se, que apesar de teoricamente todas as partições permitirem a identificação dos coeficientes de rendimento, na prática tal nem sempre é possível, visto que a aplicação do *MRLM* exige a verificação de pressupostos que devem ser respeitados.

Foi ainda demonstrada a importância do processo iterativo no desenho óptimo de experiências. A realização de uma iteração deste processo permitiu realçar a sua importância como ferramenta útil para maximizar o conteúdo de informação dos dados experimentais. Para o regime respirativo, observou-se uma melhoria significativa da qualidade dos valores estimados dos coeficientes de rendimento por comparação dos resultados obtidos para uma experiência não otimizada, com os obtidos a partir de uma experiência otimizada.

A experiência otimizada para o regime respirativo permitiu ainda determinar os valores dos coeficientes de rendimento associados ao regime respiro-fermentativo, embora com maior incerteza.

Não obstante as incertezas associadas aos coeficientes de rendimento determinados para o regime respirativo e em especial para o regime respiro-fermentativo, estes foram uma peça fundamental para a implementação de estratégias de monitorização avançada e controlo.

Desenvolveram-se sensores por programação, para a estimação de variáveis de estado e de parâmetros (cinéticas de reacção), baseados na estrutura do modelo geral de reactores biológicos.

Os observadores estendidos de Kalman, assintóticos e por intervalo permitiram estimar satisfatoriamente a concentração de biomassa. Os algoritmos propostos apenas requerem o acesso a medições em linha típicas (concentrações de oxigénio e de dióxido de carbono dissolvidos juntamente com as taxas de transferência gasosa), sendo o intervalo de amostragem necessário compatível com a generalidade dos sistemas de aquisição de dados disponíveis.

De um modo geral, na validação com dados experimentais, os observadores assintóticos apresentaram um melhor desempenho que os observadores estendidos de Kalman. Os observadores por intervalo, tal como os observadores assintóticos, apresentam a vantagem de não exigirem o conhecimento do modelo cinético. Embora, a sua validação experimental seja crucial, estudos de simulação realizados permitiram concluir que os observadores por intervalo apresentam vantagens em termos de implementação prática.

Foi desenvolvido um algoritmo de controlo para a taxa específica de crescimento da *E. coli* numa fermentação semi-contínua. A lei de controlo foi obtida adoptando uma abordagem baseada na Teoria de Controlo, originando um controlador adaptativo com modelo de referência de segunda ordem, que corresponde a um controlo por antecipação/retroacção cuja acção de retroacção é do tipo PI.

O estudo realizado por simulação numérica permitiu verificar a robustez satisfatória do algoritmo de controlo mesmo na presença de ruído nas variáveis medidas em linha, de perturbações externas e alteração dos valores nominais dos parâmetros do modelo.

Foram analisados diferentes mecanismos de sintonização dos parâmetros do controlador tendo-se concluído que, geralmente, o método de sintonização automática com recurso à regra de adaptação desses parâmetros em função do erro relativo do controlador foi o que apresentou um melhor desempenho global.

O algoritmo de controlo aqui apresentado, embora semelhante a outros descritos na literatura, apresenta duas vantagens inequívocas: não requer o conhecimento do modelo cinético e dos respectivos parâmetros e necessita de um menor número de variáveis medidas em linha. Mais ainda, apresentou um desempenho muito superior ao conseguido por outros algoritmos normalmente utilizados no controlo da taxa específica de crescimento.

6.2. SUGESTÕES PARA TRABALHO FUTURO

Os resultados obtidos neste trabalho deixam em aberto linhas para prosseguir a investigação e aplicação das estratégias apresentadas. Assim, de seguida, apresentam-se algumas sugestões de trabalho futuro que, se por um lado poderão complementar o trabalho realizado, por outro poderão ser indicações para novas linhas de investigação que possibilitem o desenvolvimento de estratégias de fermentação, sustentadas numa abordagem integrada do conhecimento fisiológico e de análises à escala genómica, que conduzam a uma melhoria do desempenho de produção em processos fermentativos.

A metodologia apresentada para a identificação dos coeficientes de rendimento permitiu a sua determinação sem o conhecimento preciso do modelo cinético. Contudo, seria interessante prosseguir com este estudo e alargar o desenho óptimo de experiências para otimizar a estrutura cinética e determinar os respectivos parâmetros, quer de modelos determinísticos quer de modelos celulares à escala genómica.

Os resultados obtidos, por simulações numéricas, na aplicação dos observadores por intervalo revelaram-se muito interessantes, apresentando, estes observadores, vantagens em termos de implementação prática devido ao seu comportamento robusto em relação ao ruído nas variáveis medidas em linha e incertezas em algumas variáveis como a concentração de substrato na alimentação. Contudo, seria importante proceder à sua validação experimental.

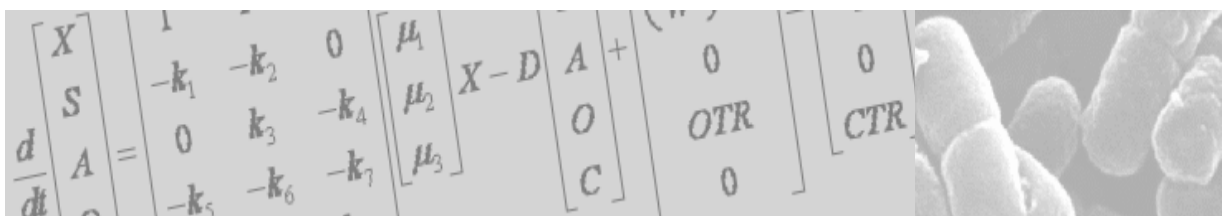
Os observadores por intervalo foram utilizados numa situação onde é considerado que a concentração de substrato na alimentação não é conhecida com precisão. O estudo destes observadores poderá ser alargado para situações onde são consideradas incertezas nos valores dos

coeficientes de rendimento, o que poderá constituir uma mais valia, permitindo melhorar a robustez do algoritmo quando existem incertezas nos valores desses coeficientes.

Seria, ainda conveniente proceder à verificação experimental do algoritmo de controlo da taxa específica de crescimento, de modo a avaliar o seu desempenho numa situação real de crescimento e produção, num processo fermentativo de *E. coli*, a nível laboratorial. Esta validação experimental reveste-se de grande importância atendendo a que várias análises à escala genómica, que permitem a caracterização do processo, como a análise do fluxoma e do proteoma, necessitem que as amostragens das culturas microbianas sejam feitas num estado *pseudo*-estacionário que pode ser conseguido impondo uma taxa específica de crescimento constante.

Por fim, seria pertinente realizar novos estudos de simulação de modo a melhorar o procedimento de sintonização dos parâmetros do controlador. Poder-se-á recorrer a um método de optimização com vista à determinação dos seus valores óptimos iniciais bem como dos valores dos parâmetros das expressões das regras de sintonização utilizadas.

ANEXO



A. CURVAS DE CALIBRAÇÃO

Neste anexo são apresentados exemplos de curvas de calibração referentes aos diferentes métodos analíticos descritos no Capítulo 2.

A.1. ESPECTRÓMETRO DE MASSA

As misturas gasosas de composição conhecida foram obtidas através de compostos puros impondo diferentes caudais. Para a calibração efectuou-se a análise das misturas gasosas e determinou-se a média dos valores da pressão obtidos para cada massa. As massas 28, 32 e 44 foram usadas para quantificar o azoto, oxigénio e dióxido de carbono, respectivamente. Para cada massa efectuou-se uma regressão linear entre a média da pressão obtida e a respectiva composição na mistura gasosa, obtendo-se assim uma correlação entre pressão e composição para cada gás analisado.

De seguida será apresentado um exemplo da curva de calibração obtida para o azoto.

Tabela A.1 Valores utilizados na curva de calibração do Azoto.

% molar	Pressão relativa (m/z=28)
69,39	0,6950
76,43	0,7817
78,53	0,7996

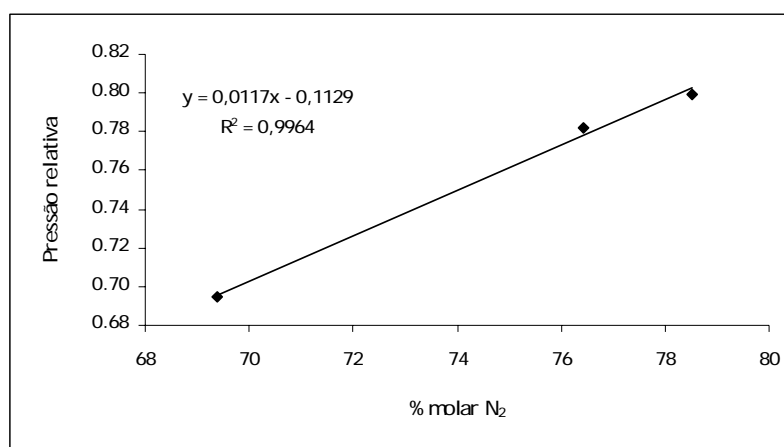


Figura A.1 Curva de calibração para o Azoto.

Para a curva de calibração do oxigénio utilizaram-se os dados apresentados na tabela que se segue.

Tabela A.2 Valores utilizados na curva de calibração do Oxigénio.

% molar	Pressão relativa (m/z=32)
10,20	0,1028
18,40	0,1822
20,53	0,2046
22,00	0,2188
25,51	0,2554

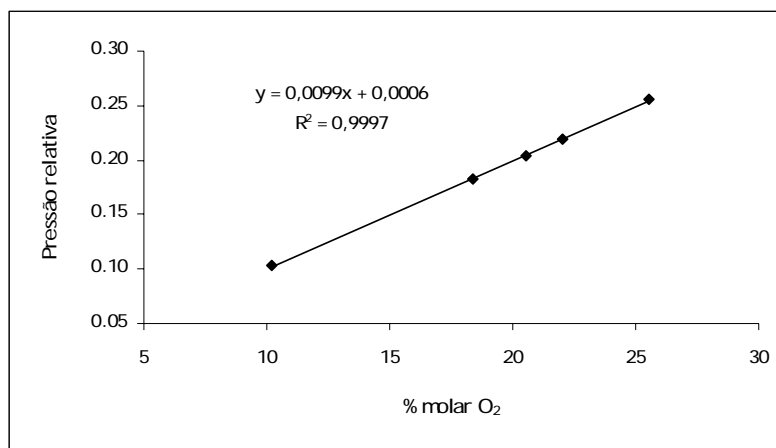


Figura A.2 Curva de calibração para o Oxigénio.

A curva de calibração do dióxido de carbono foi obtida com base nos dados apresentados na tabela seguinte:

Tabela A.3 Valores utilizados na curva de calibração do Dióxido de Carbono.

% molar	Pressão relativa (m/z=44)
1,55	0,0004240
5,10	0,04681
10,20	0,1138

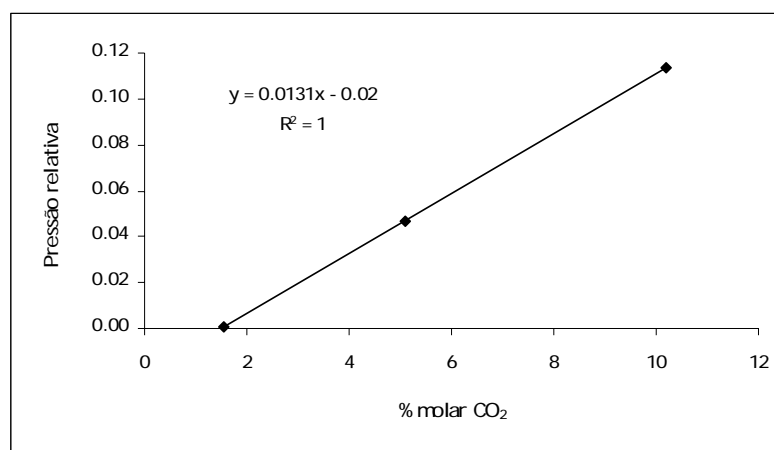


Figura A.3 Curva de calibração para o Dióxido de Carbono.

A.2. BIOMASSA

A curva de calibração da densidade óptica *versus* peso seco para a biomassa foi determinada por leitura da densidade óptica de suspensões, em duplicado, obtidas por diluição de uma suspensão de células de densidade óptica e de concentração em peso seco conhecidas. A densidade óptica destas suspensões está compreendida entre 0,1 e 0,6 unidades de absorvância. A determinação da concentração em peso seco da suspensão celular inicial é efectuada gravimetricamente como descrito no Capítulo 2. Nas tabelas seguintes são apresentados os valores utilizados no cálculo da concentração em peso seco das suspensões celulares.

Tabela A.4 Valores utilizados no cálculo da concentração em peso seco da suspensão celular inicial.

filtro nº	Peso (g)			suspensão total filtrada	Concentração de biomassa (g/kg)
	filtro	filtro+biomassa	biomassa		
1	0,0715	0,1912	0,1197	5,1509	24,24
2	0,0726	0,1980	0,1254	5,1248	24,47
3	0,0716	0,1986	0,1270	5,0802	25,00
4	0,0718	0,1938	0,1220	5,0158	24,32
		<i>média</i>	0,1235		<i>24,26</i>

Tabela A.5 Valores da concentração em peso seco das outras suspensões celulares utilizadas para a curva de calibração da biomassa.

balão nº	Volume (mL)		Concentração de biomassa (g/kg)	Densidade óptica		
	adicionado	balão		ensaio 1	ensaio 2	média
1	1	1000	0,02426	0,09090	0,0880	0,08940
2	1	500	0,04852	0,1617	0,1636	0,1626
3	1	250	0,09703	0,3040	0,3154	0,3097
4	0,5	100	0,1213	0,3811	0,3829	0,3820
5	0,67	100	0,1625	0,4948	0,5025	0,4986

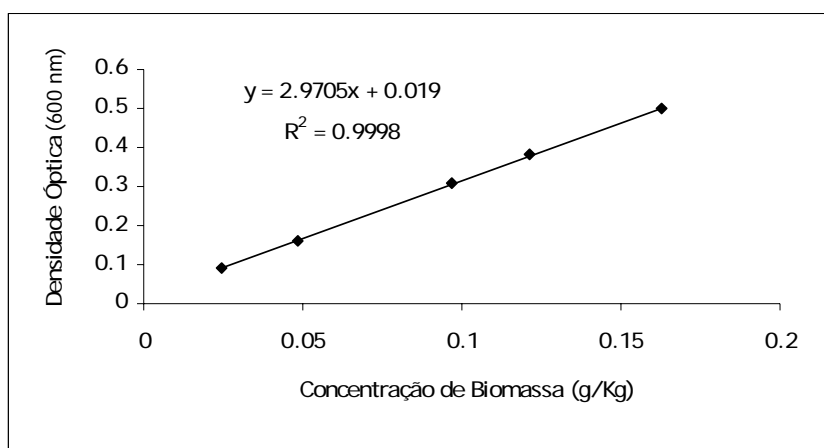


Figura A.4 Curva de calibração da Biomassa.

A.3. ACETATO

A concentração de acetato foi determinada por Cromatografia Líquida de Alta Resolução (HPLC). A curva de calibração para a determinação da concentração de acetato foi obtida a partir 6 soluções padrão cujas concentrações variavam entre 0,5 e 10 gramas de acetato por quilograma de solução. Cada uma das soluções padrão foi injectada em triplicado.

Na tabela que se segue são apresentados os valores utilizados na construção da curva de calibração para o acetato.

Tabela A.6 Valores utilizados na curva de calibração do Acetato.

Solução padrão	Concentração de acetato (g/kg)	área 1	área 2	área 3
1	0,5	443484	438147	440815
2	1,25	1049024	1070879	1068468
3	2,5	1877259	1813585	1878720
4	7,5	6706446	6625616	6464744
5	10	8370508	8416052	8437144

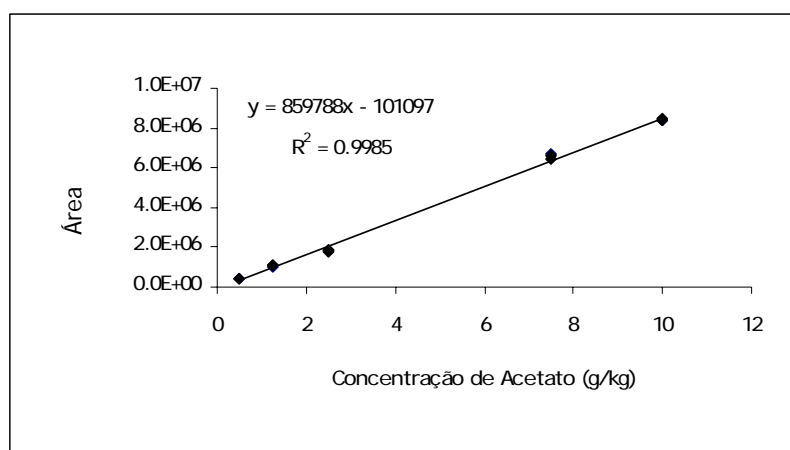


Figura A.5 Curva de calibração do Acetato.

A.4. GLUCOSE

A determinação da concentração de glucose foi realizada por HPLC e pelo *kit* enzimático Glucose Trinder GOD-POD.

A.4.1. Concentração de glucose determinada por HPLC

A curva de calibração para a determinação da concentração de glucose foi obtida a partir de 6 soluções padrão cujas concentrações variavam entre 0,5 e 10 gramas de glucose por quilograma de solução. Cada uma das soluções padrão foi injectada em triplicado.

Na tabela que se segue são apresentados os valores utilizados na construção da curva de calibração para a glucose.

Tabela A.7 Valores utilizados na curva de calibração da Glucose por HPLC.

Solução padrão	Concentração de glucose (g/kg)	área 1	área 2	área 3
1	0,5	155204	155434	---
2	1,25	379191	377845	372993
3	2,5	654375	642077	651907
4	7,5	2281010	2284154	2374358
5	10	3038533	3026125	3003294

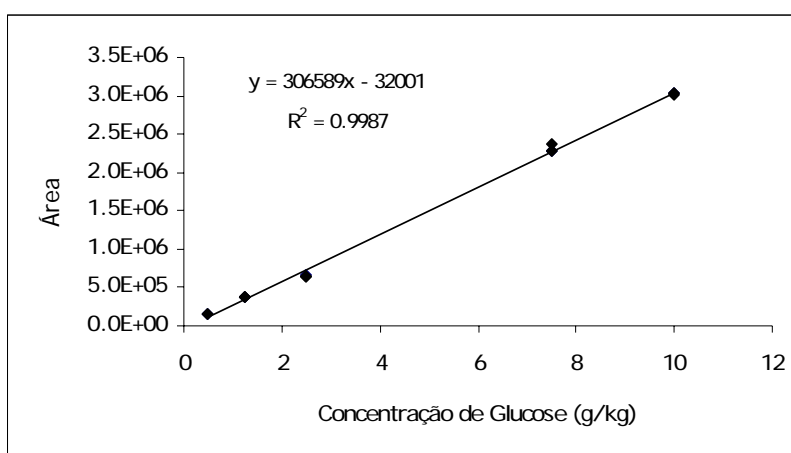


Figura A.6 Curva de calibração da Glucose por HPLC.

A.4.2. Concentração de glucose determinada por *kit* enzimático

Para a determinação da curva de calibração foram preparadas soluções padrão com concentração compreendida entre 0,1 e 5 g/kg. Posteriormente, e após agitação da microplaca contendo as soluções a analisar, procedeu-se à leitura da densidade óptica das soluções padrão (em triplicado) num leitor de microplacas ELISA, a um comprimento de onda de 505 nm.

Na tabela que se segue são apresentados os valores utilizados na construção da curva de calibração para a glucose.

Tabela A.8 Valores utilizados na curva de calibração da Glucose por *kit* enzimático.

Soluções padrão	Concentração de glucose (g/kg)	DO (505 nm)		
		ensaio 1	ensaio 2	ensaio 3
branco	---	0,0430	0,0410	0,0400
1	0,10	0,0700	0,0700	0,0710
2	0,25	0,110	0,109	0,110
3	0,50	0,179	0,180	0,180
4	0,75	0,250	0,243	0,249
5	1,00	0,320	0,319	0,319
6	2,00	0,577	0,611	0,598
7	3,00	0,868	0,873	0,866
8	4,00	1,15	1,16	1,14
9	5,00	1,43	1,43	1,45

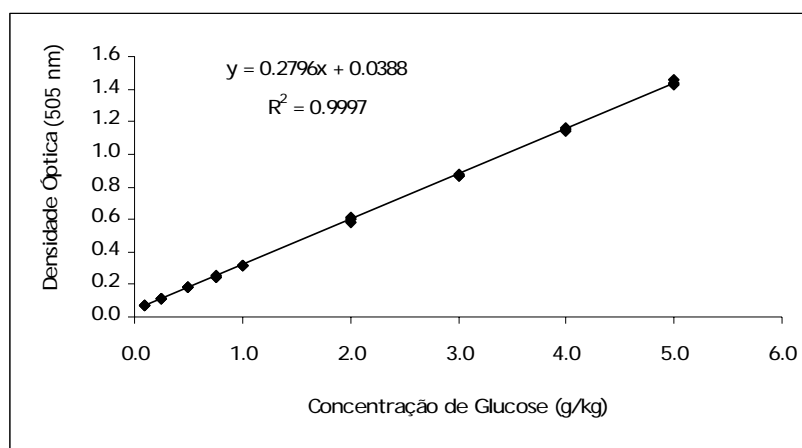


Figura A.7 Curva de calibração da Glucose por *kit* enzimático.

