

UNIVERSIDADE FEDERAL DE SÃO CARLOS
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOTECNOLOGIA

Técnicas Computacionais Inteligentes para a Inferência de
Estado e a Otimização de Cultivos de *Streptococcus pneumoniae*

ANTONIO CARLOS LUPERNI HORTA

SÃO CARLOS
2008

UNIVERSIDADE FEDERAL DE SÃO CARLOS
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOTECNOLOGIA

Técnicas Computacionais Inteligentes para a Inferência de
Estado e a Otimização de Cultivos de *Streptococcus pneumoniae*

ANTONIO CARLOS LUPERNI HORTA

Dissertação Apresentada ao Programa
de Pós-Graduação em Biotecnologia da
Universidade Federal de São Carlos,
como parte dos requisitos para
Obtenção do Título de Mestre em
Biotecnologia.

Orientadoras:

Profa. Dra. Teresa Cristina Zangirolami

Profa. Dra. Maria do Carmo Nicoletti

SÃO CARLOS
2008

**Ficha catalográfica elaborada pelo DePT da
Biblioteca Comunitária da UFSCar**

H821tc

Horta, Antonio Carlos Luperni.

Técnicas computacionais inteligentes para a inferência de estado e a otimização de cultivos de *Streptococcus pneumoniae* / Antonio Carlos Luperni Horta. -- São Carlos : UFSCar, 2008.

101 f.

Dissertação (Mestrado) -- Universidade Federal de São Carlos, 2008.

1. Comitê de redes neurais. 2. *Streptococcus pneumoniae*. 3. Busca heurística. 4. Fases de crescimento. 5. Estimativa de parâmetro. I. Título.

CDD: 660.6 (20^a)

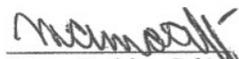


**ATA DO EXAME DE DISSERTAÇÃO DE MESTRADO DO CANDIDATO:
Antônio Carlos Luperni Horta**

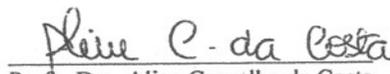
Aos vinte e sete dias do mês de março do ano de dois mil e oito, às quatorze horas, no departamento de Engenharia Química da UFSCar, reuniu-se a Comissão Examinadora nas formas e termos do Artigo 28º do Regimento Interno do Programa de Pós-Graduação em Biotecnologia da UFSCar, composta pelos membros: Profa. Dra. Teresa Cristina Zangirolami (DEQ-UFSCar, Orientadora), Profa. Dra. Maria do Carmo Nicoletti (DC-UFSCar, Orientadora), Profa. Dra. Heloísa de Arruda Camargo (DC-UFSCar) e Profa. Dra. Aline Carvalho da Costa (FEQ/UNICAMP), para Exame de Dissertação de Mestrado do candidato Antônio Carlos Luperni Horta, sob o título “Técnicas Computacionais Inteligentes para a Inferência de Estado e a Otimização de Cultivos de Streptococcus pneumoniae”. A sessão foi aberta pela Profa. Dra. Teresa Cristina Zangirolami (Presidente), iniciando-se pela apresentação da dissertação. Em seguida, foi feita a arguição do candidato pelos membros da Comissão Examinadora. A Comissão Examinadora considerou o tema relevante para a Biotecnologia e julgou a exposição feita pelo candidato clara e objetiva. O candidato respondeu satisfatoriamente as questões formuladas. Pelo apresentado acima, a comissão atribuiu os seguintes níveis: Profa. Dra. Teresa Cristina Zangirolami, nível A; Profa. Dra. Maria do Carmo Nicoletti, nível A; Profa. Dra. Heloísa de Arruda Camargo, nível A; e Profa. Dra. Aline Carvalho da Costa, nível A. De acordo com os parágrafos 1º e 2º do Artigo 33º, o candidato foi considerado aprovado com nível A. Encerrada a sessão secreta, o Presidente informou o resultado da defesa. Nada mais havendo a tratar, eu, Mauro Rodrigo Calér, Estagiário deste Programa, lavei a presente ata, que assino juntamente com os membros da Banca Examinadora.


Profa. Dra. Teresa Cristina Zangirolami


Profa. Dra. Maria do Carmo Nicoletti


Mauro Rodrigo Calér


Profa. Dra. Heloísa de Arruda Camargo


Profa. Dra. Aline Carvalho da Costa



1º TABELÃO DE NOTAS E DE PROTESTOS DE LETRAS E TÍTULOS
RUA D. ALEXANDRINA, 950 - FONE/FAX: (033) 3373-9000 - SÃO CARLOS - SP

AUTENTICAÇÃO

Autentico, a presente cópia reprográfica
extraída nestas notas a qual confere com
o original do que dou fé. São Carlos,
18/11/2008
JOSE GUILHERME RIBEIRO PORTO
FERREIRA - ESCRIVENTE
Valor recebido R\$ 1,65



AGRADECIMENTOS

À professora Doutora Teresa Cristina Zangirolami e professora Doutora Maria do Carmo Nicoletti pelo apoio, dedicação e, sobretudo por tudo o que me ensinaram.

Aos meus pais por terem possibilitado meus estudos e meu desenvolvimento científico.

À minha esposa e meus filhos por terem me suportado nos momentos estressantes desta jornada.

À minha avó e aos meus irmãos que me apoiaram e aconselharam nos momentos de dúvida.

Aos meus amigos, colegas e professores que me ajudaram a propor e a programar os modelos abordados no trabalho; em especial à Luciana, ao Pedro, ao Charles, ao Gilson, ao Thiago, ao Daniel, à Adriana, ao Roberto e ao Antônio.

À equipe do Butantan que colaborou para a realização deste trabalho fornecendo os dados experimentais.

À CAPES pelo financiamento.

RESUMO

Streptococcus pneumoniae (pneumococo) é uma bactéria patogênica causadora de várias infecções que são agravadas pelo aumento de cepas com resistência aos antibióticos. O desenvolvimento de uma vacina efetiva contra este patógeno é crucial para a prevenção das doenças pneumocócicas.

Vacinas conjugadas, constituídas pelo polissacarídeo capsular ligado a uma proteína carregadora, são mais eficientes no estímulo da memória imunológica. O polissacarídeo capsular (PS) está presente na cápsula que envolve a célula e, desta forma, a elaboração de vacinas conjugadas envolve o cultivo da bactéria para a produção do mesmo. Como o microrganismo é cultivado na ausência de oxigênio, a produção de lactato é inevitavelmente elevada e o seu acúmulo no meio provoca a inibição do crescimento.

Para minimizar os efeitos inibitórios da acumulação de lactato e aumentar a produção de PS é necessário monitorar o processo e controlar adequadamente a adição de meio suplementar e a retirada de meio saturado em cultivos operados em batelada alimentada, utilizando biorreatores acoplados a sistema de perfusão. O sucesso no monitoramento, no controle e na otimização deste bioprocessos depende da utilização de recursos de modelagem e de simulação que sejam eficientes.

Este trabalho de pesquisa propõe o uso de técnicas computacionais inteligentes, especificamente a técnica de busca heurística chamada de *simulated annealing* (SA) aliada a redes neurais, para a inferência de estado e a otimização de cultivos de *S. pneumoniae*. A proposta foi concretizada via desenvolvimento de um sistema computacional que:

- a) faz uso do SA para a identificação do conjunto de valores de parâmetros associados a modelos não estruturados e;
- b) usa redes neurais (individualmente e em regime de comitê) para a inferência de estado de um cultivo.

O trabalho apresenta e discute os resultados do sistema em conjuntos de dados obtidos experimentalmente e evidencia a importância da proposta para uma maior eficiência no controle de processos de cultivo.

Palavras-chave: Comitê de Redes Neurais, *Streptococcus pneumoniae*, Simulated Annealing, identificação das fases de crescimento, modelagem não-estruturada de crescimento microbiano, estimativa de parâmetros biocinéticos.

ABSTRACT

Streptococcus pneumoniae (pneumococo) is a pathogenic bacterium that causes several infections which are aggravated by the increase of serotypes with antibiotics resistance. The development of an effective vaccine against this pathogen is crucial for the prevention of the pneumococcal illnesses.

Conjugated vaccines, consisting of the capsular polysaccharide joined to a carrier protein, are more efficient in the stimulation of the immunologic memory. The capsular polysaccharide (PS) is present in the capsule that involves the cell. Thus, the conjugated vaccine elaboration involves bacterial cells cultivation for its production. As the organism is cultivated in the oxygen absence, the lactate production is inevitably high, leading to growth inhibition due to lactate accumulation in the medium.

To minimize the inhibitory effects of the lactate accumulation and to increase the PS production it is necessary to monitor the process and adequately control the addition of supplementary medium along with the withdrawal of saturated medium. This kind of operation can be performed by carrying out a fed-bath cultivation in a bioreactor connected to a perfusion system. The success on the monitoring, control and optimization of this bioprocess depends on the efficiency of the modeling and simulation resources employed.

This research work considers the uses intelligent computational techniques, specifically the technique of heuristical search called simulated annealing (SA) combined with neural networks for the state inference and the optimization of *S. pneumoniae* cultivations. The proposal was implemented as a computational system that:

- a) uses the SA for the identification of the values for a set of parameters associated to unstructured models and;
- b) uses neural networks (individually and grouped as a committee) for the state inference of a culture.

The work presents and discusses the results of the system for data sets experimentally obtained and highlights the importance of the proposal for achieving a higher efficiency in the culture control processes.

Keywords: Neural Networks Committee, *Streptococcus pneumoniae*, Simulated Annealing, growth phase identification, microbial growth unstructured modeling, biokinetic parameters estimation.

LISTA DE ABREVIÇÕES E SÍMBOLOS

C_X : Concentração celular
 C_L : Concentração de lactato
 C_S : Concentração de substrato
 C_{PS} : Concentração de polissacarídeo capsular
 CO_2 : Gás carbônico
DIP : Doenças invasivas por pneumococo
DO : Densidade óptica
EDO: Equação diferencial ordinária
F: Vazão de alimentação de meio suplementar
F': Vazão de retirada de meio (perfusão)
FBR: Rede de função base radial
FO: Função objetivo
IA: Inteligência Artificial
K: Parâmetro do modelo de Tessier análogo ao K_S
 K_D : Constante de morte
 K_S : Constante de saturação
m: Coeficiente de manutenção celular
MNE: modelagem não estruturada
n: Parâmetro do modelo de Moser
NaOH: Massa de solução de hidróxido de sódio
OG: Otimização global
 O_2 : Oxigênio
PS: Polissacarídeo capsular
PMC: Perceptron multi-camada
RN: Rede neural
SRES: "Evolution Strategy using Stochastic Ranking"
t: tempo
T: temperatura
uES: "unconstrained Evolution Strategy"
V: volume do biorreator
 Y_{LS} : Fator de formação de lactato a partir do substrato
 Y_{XS} : Fator de conversão de substrato a células
 α_L : Coeficiente pseudo-estequiométrico para formação de lactato associada ao crescimento
 α_{PS} : Coeficiente pseudo-estequiométrico para formação de PS associada ao crescimento
 β_L : Coeficiente cinético para formação de lactato não associada ao crescimento
 β_{PS} : Coeficiente cinético para formação de PS não associada ao crescimento
 μ : Velocidade específica de crescimento
 μ_M : Velocidade máxima específica de crescimento

LISTA DE FIGURAS

Figura 2.1. Estrutura da parede celular de bactérias gram-positivas (Koneman <i>et al.</i> , 2001).....	17
Figura 2.2. Esquema de reator tanque agitado. F corresponde à vazão de alimentação de meio suplementar, F' à vazão de retirada de meio de cultivo e V o volume. Operação em Batelada → F=F'=0; Operação em Batelada Alimentada → F≠0 e F'=0; Operação em modo contínuo → F≠0 e F'≠0 sem reciclo de células (ausência de filtro); Operação em Batelada alimentada com perfusão → F≠0 e F'≠0 com reciclo de células.	21
Figura 2.3. Curva típica de crescimento de uma população de bactérias, em um cultivo em batelada em meio líquido, na qual são evidenciadas fases: 1-lag; 2-aceleração; 3-crescimento exponencial; 4-desaceleração; 5-estacionária e 6-morte.	22
Figura 2.4. Perspectivas para a construção de modelos cinéticos do cultivo de microrganismos (Bailey e Ollis, 1986).....	25
Figura 2.5. Modelo geral de neurônio. x_i são as entradas do neurônio j , w_{ij} são os pesos associados a cada entrada, d_j é a saída da função soma (f_1), y_j é a saída da função de ativação (f_2) e S_j a saída da função saída (f_3).	33
Figura 2.6. Diferentes funções ativação: a) linear; b) <i>threshold</i> linear; c) degrau; d) degrau reverso; e) sigmóide; f) gaussiana.	34
Figura 3.1. Sistema de microfiltração acoplado ao biorreator. 1-Solução 5M de NaOH 5M (para controle de pH); 2- Sistema de membranas Durapore (filtra o meio retirando os metabólitos); 3- Meio para alimentação (supre o biorreator com glicose e nutrientes); 4- Balança e recipiente com permeado; 5- Fermentador (biorreator).	41
Figura 4.1. Dados e alisamento do Exp ₂ . C_X corresponde à concentração celular, C_S à concentração de glicose, C_L à concentração de lactato e C_{PS} à concentração de PS. Os pontos correspondem aos dados experimentais, enquanto que as linhas ao alisamento e interpolação. A descontinuidade de C_S em 7 horas corresponde ao pulso de glicose.....	47
Figura 4.2. Esquema geral da escolha do melhor modelo não estruturado para representar os cultivos em batelada do <i>S. pneumoniae</i>	52
Figura 4.3. Ajuste do modelo de Tessier ao cultivo Exp ₂ ^{est} . Os pontos/símbolos representam os valores experimentais, enquanto que as linhas, os valores preditos pelo modelo. C_X representa a concentração celular, C_S a concentração de glicose, C_L a concentração de lactato e C_{PS} a concentração de PS. (a) representa toda simulação até a sétima hora; (b) representa apenas C_X , e C_{PS}	54
Figura 4.4. Convergência da função objetivo; Modelo de Tessier calibrado pelo SA utilizando os dados Exp ₂ ^{est} . O valor da função objetivo está em escala logarítmica. O eixo do Número de iterações mostra os valores até 10.000 pois a partir disso não houve modificação no valor da função objetivo.	57
Figura 4.5. Validação do modelo proposto para o cultivo de <i>S. pneumoniae</i> em batelada, juntamente com o modelo de Tessier e os parâmetros calibrados a partir dos dados do cultivo Exp ₂ . (a) Exp ₃ (b) Exp ₁ . Os símbolos representam os valores experimentais medidos, enquanto que as linhas, os valores preditos pelo modelo. C_X representa a concentração celular, C_S representa a concentração de glicose, C_L representa a concentração de lactato e C_{PS} a concentração de PS.	58
Figura 4.6. Validação do modelo proposto para o cultivo de <i>S. pneumoniae</i> em batelada, juntamente com o modelo de Tessier e os parâmetros calibrados a partir dos dados do cultivo Exp ₂ . (a) Exp ₄ (b) Exp ₅ . Os símbolos representam os valores experimentais	

- medidos, enquanto que as linhas, os valores preditos pelo modelo. C_X representa a concentração celular, C_S representa a concentração de glicose, C_L representa a concentração de lactato e C_{PS} a concentração de PS. 59
- Figura 4.7. Evolução das concentrações de células, C_X (+), de substrato, C_S (●), de lactato, C_L (◆) e de polissacarídeo capsular, C_{PS} (□) ao longo do tempo para o cultivo de *S. pneumoniae* em biorreator acoplado a sistema de perfusão. Símbolos: dados experimentais; Linhas Contínuas: dados gerados por alisamento e interpolação. 62
- Figura 4.8. Comparação entre os resultados simulados e os experimentais obtidos no cultivo empregando biorreator acoplado às membranas de perfusão para retenção de células e drenagem de meio. Linhas tracejadas: dados simulados; Linhas contínuas: dados interpolados e alisados. C_X representa a concentração celular, C_S representa a concentração de glicose, C_L representa a concentração de lactato e C_{PS} a concentração de PS. 64
- Figura 4.9. Simulação do experimento. Evolução das concentrações de células, C_X (linha verde), de substrato, C_S (linha preta), de lactato, C_L (linha vermelha) e de polissacarídeo capsular, C_{PS} (linha azul) ao longo do tempo para o cultivo de *S. pneumoniae* em biorreator acoplado a sistema de perfusão. Dados gerados por simulação (reator de 5L). 67
- Figura 4.10. Interpolação dos dados do Exp₂. DO corresponde à densidade óptica, C_X à concentração celular, C_S à concentração de glicose, C_L à concentração de lactato, C_{PS} à concentração de PS e NaOH à massa de solução de NaOH (g). Os pontos correspondem aos dados experimentais, enquanto que as linhas ao alisamento e interpolação. 69
- Figura 4.11. Divisão das fases de crescimento em função do perfil da dDO/dt (aproximação da $\Delta DO/\Delta t$). As fases indicadas correspondem à: I (fase lag e aceleração); II (fase exponencial); III (fase de desaceleração) e IV (fase estacionária e morte). DO é a densidade óptica do meio, e t o tempo. Dados estendidos do cultivo Exp₂. 70
- Figura 4.12. Esquema da preparação dos dados de cada experimento para serem utilizados no treinamento de cada rede neural individual, RN_i , ($i=1, 2, 3, 4$). Na figura IT é referente ao processo de interpolação; EH à classificação das instâncias de dados em sua respectiva fase de crescimento por um especialista humano; SB à divisão dos dados em conjunto de treinamento (TR_Exp_i) e conjunto de validação (VAL_Exp); BP refere se ao treinamento backpropagation resultando na rede neural i (RN_i). 71
- Figura 4.13. Esquema do treinamento das redes neurais com os dados estendidos e classificados dos respectivos experimentos. 73
- Figura 4.14. Comitê de redes neurais classificatório. O procedimento contador é encarregado de decidir a classe final (i.e., a fase de crescimento), baseado na saída dada pelas quatro redes neurais. 74
- Figura 4.15. Arquiteturas das redes neurais mais promissoras. O número de neurônios testados na camada intermediária foram 2, 3, 4, 5, 6, 7, 9, 10 e 15; e com duas camadas as combinações 4-5, 5-4, 5-5, 5-7, 7-5 e 7-7, respectivamente. 74
- Figura 4.16. Desempenho de classificação do Comitê e da configuração de Rede única (RN) para diferentes variáveis de entrada. Treinamento feito com 200 épocas e uma camada intermediária. O desempenho corresponde à média das configurações com 5, 7, 9 e 10 neurônios na camada intermediária. 75
- Figura 4.17. Desempenho do comitê classificatório treinado com as variáveis de entrada {t, DO, NaOH} e com redes com uma camada intermediária com 5, 7, 9, 10 e 12 neurônios. 76

Figura 4.18. Desempenho do comitê classificatório com variáveis de entrada {t, DO, NaOH} e com duas camadas intermediárias com 4-5, 5-4, 5-5, 5-7, 7-5 e 7-7 neurônios nas respectivas camadas.	77
Figura 4.19. Desempenho do comitê classificatório com variáveis de entrada {DO, NaOH} e com duas camadas intermediárias com 4-5, 5-4, 5-5, 5-7, 7-5 e 7-7 neurônios nas respectivas camadas.	77
Figura 4.20. Resultados do ajuste do modelo de Tessier para a concentração celular (C_X), concentração de substrato (C_S), concentração de lactato (C_L) e concentração de polissacarídeo (C_{PS}). As linhas grossas correspondem aos dados simulados pelo modelo segregado por fases; as linhas finas correspondem aos dados simulados pelo modelo geral (Seção 4.1); os símbolos correspondem aos dados experimentais brutos do cultivo Exp ₂ . ..	79
Figura 4.21. Simulação do (a) Exp ₂ ^{est} e (b) Exp ₃ ^{est} com os parâmetros do modelo de Tessier ajustado em cada respectiva fase de crescimento do Exp ₂ ^{est}	82
Figura 4.22. Simulação do (a) Exp ₄ ^{est} e (b) Exp ₅ ^{est} com os parâmetros do modelo de Tessier ajustado em cada respectiva fase de crescimento do Exp ₂ ^{est} ; f é o valor da função objetivo. 83	
Figura 4.23. Esquema geral da ferramenta de inferência <i>on-line</i> . DO se refere à densidade ótica do meio, enquanto que NaOH é referente à massa de solução alcalina adicionada no meio de cultura para controlar o pH. ESPM é o algoritmo de escolha de parâmetros; t o tempo, C_X a concentração celular, C_S a concentração de glicose, C_L a concentração de lactato e C_{PS} a concentração de polissacarídeo capsular.	85
Figura 4.24. Simulação do Exp ₂ (a) e Exp ₃ (b), com a ferramenta de inferência tendo como entrada os atributos {t, DO, NaOH}.....	87
Figura 4.25. Simulação do Exp ₄ (a) e Exp ₅ (b), com a ferramenta de inferência tendo como entrada os atributos {t, DO, NaOH}.....	88

LISTA DE QUADROS

Quadro 1.1. Doenças causadas por pneumococo (Bricks e Berezin, 2006).	12
Quadro 2.1. Grupos de risco para doenças pneumocócicas invasivas (Bricks e Berezin, 2006)..	18
Quadro 2.2. Parâmetros das expressões cinéticas de Monod, Tessier, Contois e Moser.	27
Quadro 4.1. Variáveis e parâmetros dos modelos.	48
Quadro 4.2. Pseudocódigo do SA. ΔE corresponde à diferença entre o valor da função objetivo na solução corrente e o valor da mesma função na solução candidata.....	49

LISTA DE TABELAS

Tabela 3.1. Meio de cultivo das 5 fermentações usadas como dados.	39
Tabela 3.2. Condições experimentais utilizadas nas fermentações.	39
Tabela 3.3. Condições experimentais utilizadas no Exp ₆	42
Tabela 4.1. Restrições dos valores dos parâmetros.	51
Tabela 4.2. Valores estimados pelo SA dos parâmetros dos modelos de Monod, Tessier e Moser.	52
Tabela 4.3. Valores das funções objetivo dos modelos testados.	53
Tabela 4.4. Intervalo de confiança dos parâmetros ajustados ao Exp ₂	55
Tabela 4.5. Matriz de correlação dos parâmetros do modelo de Tessier ajustado aos dados do Exp ₂	56
Tabela 4.6. Valores da função objetivo para o ajuste do modelo aos dados experimentais brutos e estendidos dos cultivos Exp ₁ , Exp ₂ , Exp ₃ , Exp ₄ e Exp ₅	60
Tabela 4.7. Valores do parâmetro λ utilizados na suavização.	61
Tabela 4.8. Parâmetros ajustados via SA-LM na operação em batelada e em perfusão.	63
Tabela 4.9. Condições iniciais da simulação para a batelada e para o período de operação do sistema integrado.	65
Tabela 4.10. Intervalo das fases de crescimento quanto à variação da DO com o tempo das fermentações Exp ₁ ^{est} , Exp ₂ ^{est} , Exp ₃ ^{est} , Exp ₄ ^{est} e Exp ₅ ^{est}	70
Tabela 4.11. Esquema adotado para representar instâncias de treinamento que caracterizam quatro fases.	71
Tabela 4.12. Valores da função objetivo para cada respectiva fase, Lag e Aceleração, Exponencial, Desaceleração e Estacionária e Morte.	80
Tabela 4.13. Valores dos parâmetros ajustados a cada fase de crescimento do cultivo Exp ₂ - dados estendidos.	80
Tabela 4.14. Valor da função objetivo para as simulações do Exp ₂ ^{est} , Exp ₃ ^{est} , Exp ₄ ^{est} e Exp ₅ ^{est} , comparando os resultados do modelo ajustado a tudo batelada com o segregado por fases de crescimento.	84
Tabela 4.15. Valores das respectivas funções objetivo das simulações dos Exp ₂ ^{est} , Exp ₃ ^{est} , Exp ₄ ^{est} e Exp ₅ ^{est} com a ferramenta de inferência. DO corresponde à densidade óptica, NaOH à massa de solução da base para controlar o pH e t corresponde ao tempo.	88

SUMÁRIO

Capítulo 1 – INTRODUÇÃO	12
Capítulo 2 – FUNDAMENTOS TEÓRICOS	17
2.1 Aspectos Gerais sobre <i>S. pneumoniae</i>, Produção de Vacinas, Cultivo de Microrganismos e Modelagem de Bioprocessos	17
2.1.1 Microrganismo e Vacina	17
2.1.2 Biorreator e Modos de Operação	19
2.1.3. Crescimento Celular em Biorreator Batelada.....	21
2.1.4 Modelagem Matemática do Bioprocesso	24
2.2 Técnicas de Inteligência Computacional Relevantes ao Problema	29
2.2.1 Busca Heurística do Melhor Conjunto de Valores de Parâmetros que Otimizam um Modelo Determinístico Via <i>Simulated Annealing</i>	29
2.2.2 Principais Conceitos Relativos a Redes Neurais Artificiais e Sua Aplicação	32
2.2.2.1. Conceitos básicos (Palma Neto e Nicoletti, 2005)	32
2.2.2.2. Aplicações em bioprocessos.....	35
Capítulo 3 – OBTENÇÃO, DESCRIÇÃO E TRATAMENTO DOS DADOS EXPERIMENTAIS	38
3.1. Obtenção dos Dados Experimentais	38
3.1.1. Cultivos de <i>S. pneumoniae</i> em Batelada	38
3.1.2. Cultivo de <i>S. pneumoniae</i> em Biorreator Acoplado a Sistema de Retenção de Células e Retirada de Meio	40
3.2. Descrição dos Dados	43
3.3 Tratamento dos Dados para Utilização Pelas Técnicas de Inteligência Computacional – Interpolação e Alisamento	43
Capítulo 4 – MODELAGEM DE CULTIVO DE <i>S. PNEUMONIAE</i> UTILIZANDO ALGORITMO DE BUSCA HEURÍSTICA E COMITÊ DE REDES NEURAI	45
4.1 Modelagem não Estruturada e Validação de Cultivos de <i>Streptococcus pneumoniae</i> em Batelada	46
4.2 Modelagem não Estruturada de Cultivo em Biorreator Operado em Batelada Alimentada Acoplado a Sistema de Perfusão	61
4.3 O Uso de um Comitê de Redes Neurais Para o Reconhecimento Das Fases de Crescimento...	68
4.4 Modelagem não Estruturada de Cultivo em Batelada Segregada por Fases de Crescimento..	78
4.5 Ferramenta de Inferência baseada na modelagem segregada por fases de crescimento e no Comitê de Redes Neurais Classificatório	84
Capítulo 5 – CONSIDERAÇÕES FINAIS	91
SUGESTÕES PARA TRABALHOS FUTUROS	96
REFERÊNCIAS	97

Capítulo 1 – INTRODUÇÃO

Streptococcus pneumoniae (pneumococo) é o principal agente causador de doenças invasivas no ser humano. O Quadro 1.1 lista algumas doenças causadas por esta bactéria. De acordo com Bricks e Berezin (2006), estima-se que o pneumococo tenha causado mais de um milhão de óbitos por ano de crianças menores de cinco anos (em países em desenvolvimento). No Brasil, o número de casos de meningites pneumocócicas foi de 29.600 com 29 % de óbitos, entre os anos de 1983 a 2003. Nos Estados Unidos, o pneumococo liderou os casos de mortalidade e morbidade em crianças e idosos, com mais de 7 milhões de casos por ano de pneumonia, bronquite crônica, sinusite e otite média (Bhavnani *et al.*, 2005).

Quadro 1.1. Doenças causadas por pneumococo (Bricks e Berezin, 2006).

Doenças não invasivas	Doenças invasivas*
Otite média aguda	Bacteremia**
Sinusite	Pneumonia bacterêmica / empiema
Conjuntivite	Meningite
Bronquite	Sepse***
Pneumonia	Peritonite****
	Artrite/ osteomielite
*Doenças invasivas: isolamento do pneumococo de locais habitualmente estéreis (sangue, líquido cefalorraquidiano); ** Bacteremia: presença de bactérias no sangue; ***Sepse ou septicemia: infecção geral grave; ****Peritonite: Infecção do peritônio, membrana que reveste a parede do abdome e protege os órgãos.	

De acordo com Bricks e Berezin (2006), no Brasil, em 2004, os gastos públicos com crianças menores de 5 anos hospitalizadas por pneumonia foram de aproximadamente 180 milhões de dólares. Supondo que 33% destas pneumonias tenham sido causadas pelo pneumococo, então 90 milhões de dólares foram gastos apenas no tratamento de crianças hospitalizadas por pneumonia pneumocócica. Já os gastos públicos com meningite pneumocócica chegam a 11,8 milhões de dólares por ano. Além dos elevados gastos no tratamento das infecções, a possibilidade de evitar muitas mortes relacionadas às mesmas mais do que justifica o investimento na vacina conjugada: no Brasil, 30% das crianças menores de 5 anos com meningite por *S. pneumoniae* não sobrevivem (Bricks e Berezin, 2006).

O impacto das doenças pneumocócicas, tanto em países em desenvolvimento quanto em países desenvolvidos, aliada ao aumento da resistência bacteriana aos antibióticos, torna crucial o desenvolvimento de vacinas para prevenir estas doenças. O desenvolvimento da

vacina de polissacarídeo capsular (PS) é baseado no conhecimento de que anticorpos formados contra o PS são efetivos na imunização. Entretanto, os anticorpos formados são específicos a cada sorotipo, e existem mais de 90 sorotipos de *S. pneumoniae* (Brilles *et al.*, 1998; Gonçalves, 2002). Além disso, de acordo com Gonçalves (2002), o PS não possui a capacidade de ativar os linfócitos T, levando à incapacidade de gerar memória imunológica. A conjugação covalente de polissacarídeos com uma proteína aumenta a resposta do sistema imune, gerando memória imunológica (Gonçalves, 2002).

O sucesso obtido pela vacina conjugada na proteção contra infecções causadas por *Haemophilus influenzae* tipo b estimulou a pesquisa e o desenvolvimento de novas vacinas conjugadas, sendo que em 2000 foi licenciada a primeira vacina conjugada contra o *S. pneumoniae* nos Estados Unidos (Bricks e Berezin, 2006).

Desde 2001 está disponível no mercado brasileiro a vacina Pnc7 denominada Prevenar[®], produzida pelo Laboratório Wyeth, contendo os polissacarídeos dos sorotipos 4, 6B, 9V, 14, 18C, 19F e 23F do pneumococo, conjugados à proteína CRM197 (mutante da toxina diftérica). A vacina oferece proteção contra 80% dos pneumococos mais comuns nos Estados Unidos, inclusive os mais resistentes. Essa proteção cai para 60% nos países latino-americanos, principalmente devido à prevalência de diferentes sorotipos. Além da menor eficiência na imunização, uma dose de Prevenar[®] custa cerca de 50 dólares. Assim, uma campanha de vacinação em massa para imunizar 3,5 milhões de crianças, com a aplicação de três doses no primeiro ano de vida e uma dose no segundo ano, consumiria 700 milhões de dólares em recursos do Ministério da Saúde (Costa, 2007).

Visando introduzir no mercado um produto nacional de baixo custo e com alta capacidade de imunização, o Laboratório de Fermentação do Centro de Biotecnologia do Instituto Butantan (SP) vem, desde 1998, investindo no desenvolvimento de uma vacina conjugada para combater infecções de *S. pneumoniae*. As pesquisas envolvem desde os cultivos de diferentes sorotipos de *S. pneumoniae* para a produção de polissacarídeos e os cultivos de *E. coli* recombinante que expressa a proteína PspA utilizada na conjugação, até o estudo dos processos de separação requeridos para obtenção dos polissacarídeos e da proteína na forma purificada, requisito essencial para a última etapa, a reação de conjugação.

O LaDABio (Laboratório de Desenvolvimento e Automação de Bioprocessos) do Departamento de Engenharia Química da UFSCar vem interagindo com o Laboratório de

Fermentação do Butantan desde 2000, buscando estimular a incorporação de técnicas de modelagem, simulação e monitoramento de bioprocessos aos cultivos de *S. pneumoniae*, *E. coli* recombinante e *H. influenzae*. A incorporação destas técnicas é fundamental para que a produção em larga escala de vacinas antipneumocócicas conjugadas possa ter seu custo reduzido e um produto de alta qualidade possa ser obtido.

No caso específico dos cultivos de bactérias encapsuladas, como é o caso do pneumococo, o custo está diretamente relacionado à formação do polissacarídeo capsular, o principal antígeno vacinal, uma vez que perdas significativas são inevitáveis nas etapas de purificação e conjugação (Gonçalves, 2002). No entanto, a definição das condições de cultivo que levem à máxima produção do polissacarídeo não é trivial. Como comentado anteriormente as cepas de relevância clínica são homofermentativas (a fermentação da glicose origina apenas um tipo de produto), sendo o lactato o produto final da fermentação da glicose.

Considerando que o crescimento de *S. pneumoniae* ocorre em condições anaeróbicas, o acúmulo de lactato no meio de cultivo causa inibição do crescimento, tornando-se o principal obstáculo para a obtenção de altas densidades celulares e, conseqüentemente, elevadas concentrações de polissacarídeo. Além disso, a glicose deve ser estrategicamente fornecida ao longo do cultivo, para evitar a ocorrência de inibição do crescimento pelo substrato. Sendo assim, a realização de cultivos em biorreatores operados em batelada alimentada, com monitoramento da produção de lactato e adoção de estratégias de controle da vazão de alimentação de meio suplementar, em associação com sistemas que minimizem o acúmulo de lactato, se torna alternativa promissora para obtenção de altas produtividades em polissacarídeos capsulares. Porém, para implementação efetiva de uma estratégia de controle, é fundamental que o bioprocessos seja formalmente modelado.

Modelos utilizados para descrever bioprocessos podem ser caracterizados como estruturados ou não-estruturados. Modelos não-estruturados são modelos simples nos quais os componentes celulares são agrupados em apenas uma variável, C_X (concentração celular). Já os modelos estruturados são mais complexos, incluindo até variáveis intracelulares tal como a concentração de enzimas controladoras do metabolismo (Nielsen e Villadsen, 1994). Devido à complexidade do metabolismo dos microrganismos, a modelagem não-estruturada convencional nem sempre consegue representar a evolução de um cultivo. Já os modelos metabolicamente estruturados requerem grande quantidade de informação, muitas vezes de componentes

intracelulares, para que possam ser utilizados. Os modelos não-estruturados segregados nas diferentes fases de um cultivo, entretanto, podem ser mais simples que os modelos não estruturados clássicos. Modelos clássicos não se adequam facilmente à representação de todas as fases metabólicas pelas quais o microrganismo passa durante um cultivo (Honda e Kobayashi, 2000; Shuler e Kargi, 2001; Nielsen e Villadsen, 1994).

A utilização da modelagem segregada por fases requer o conhecimento de cada respectiva fase do cultivo. A detecção *on-line* das fases de crescimento tem sido implementada com base em diferentes técnicas, como por meio de redes neurais (Simon e Karim, 2001; Silva, 2003) e lógica fuzzy, bem como por meio da associação de modelagem fuzzy com redes neurais (Honda e Kobayashi, 2000).

A associação da modelagem segregada por fases a um software de detecção *on-line* da fase de crescimento celular, pode auxiliar no controle e otimização dos bioprocessos, inclusive para a produção de polissacarídeo capsular. Neste contexto de redução de custos, o aumento de quantidade e melhoria na qualidade da vacina conjugada contra o pneumococo, o uso de redes neurais artificiais (RNs) e modelagem não estruturada (MNE) no auxílio ao controle e otimização das condições de cultivo do microrganismo, se mostram técnicas promissoras.

RNs são um método geral e prático para o aprendizado de aproximação de funções a partir de exemplos e têm sido utilizadas com sucesso em muitos problemas, tais como interpretação visual de cenas, reconhecimento de escrita, controle de processos, etc. (Palma Neto e Nicoletti, 2005). Modelos baseados em redes neurais são inerentemente paralelos e muito promissores devido principalmente à habilidade desses modelos no aprendizado de relações não lineares. Almeida (2002) destaca o aumento de popularidade das RNs em aplicações cujas descrições mecanicistas da dependência entre variáveis dependentes e independentes são desconhecidas ou até mesmo muito complexas.

Para a finalidade deste trabalho, um comitê de RNs foi utilizado para identificar e classificar as fases de crescimento microbiano dos cultivos do pneumococo. Além disso, sua associação com modelos cinéticos de crescimento possibilitou o desenvolvimento de uma ferramenta de inferência das concentrações das variáveis do processo, ou seja, uma ferramenta que fornece os valores das variáveis de interesse *on-line*.

Desta forma, os objetivos principais deste trabalho foram:

- a) gerar uma ferramenta baseada em um comitê de redes neurais para a classificação *on-line* das fases do cultivo;
- b) desenvolver um modelo não estruturado segregado por fases para descrever com precisão o cultivo do pneumococo; ao mesmo tempo desenvolver um software de calibração que usa o Simulated Annealing (SA) para encontrar os parâmetros do modelo;
- c) criar usando um comitê de redes neurais e o modelo segregado por fases, uma ferramenta de inferência de estado;
- d) utilizar uma técnica de busca heurística, no caso SA, para encontrar as condições que levam à maior produtividade de PS a partir das simulações realizadas com o modelo do processo com perfusão.

Capítulo 2 – FUNDAMENTOS TEÓRICOS

2.1 Aspectos Gerais sobre *S. pneumoniae*, Produção de Vacinas, Cultivo de Microrganismos e Modelagem de Bioprocessos

Esta seção apresenta as principais informações sobre o microrganismo *S. pneumoniae*, as infecções por ele causadas e as vacinas antipneumocócicas. Nas demais seções são apresentados conceitos fundamentais sobre os biorreatores e sua operação; o crescimento microbiano e a modelagem de bioprocessos.

2.1.1 Microrganismo e Vacina

Streptococcus pneumoniae é uma bactéria Gram-positiva, aerotolerante, ou seja, que tolera oxigênio em baixas concentrações (Chapuy-Regaud *et al.*, 2001), a qual apresenta-se na forma de cocos, diplococos lanceolados ou cadeias (Klein e Ellis, 1997). Possui uma parede celular espessa com quase 80 nm contendo polissacarídeos, proteínas e moléculas específicas, denominadas ácidos teicóicos, responsáveis pela virulência da bactéria (Madigan *et al.*, 2004; Koneman *et al.*, 2001). A estrutura celular Gram-positiva é representada na Figura 2.1.

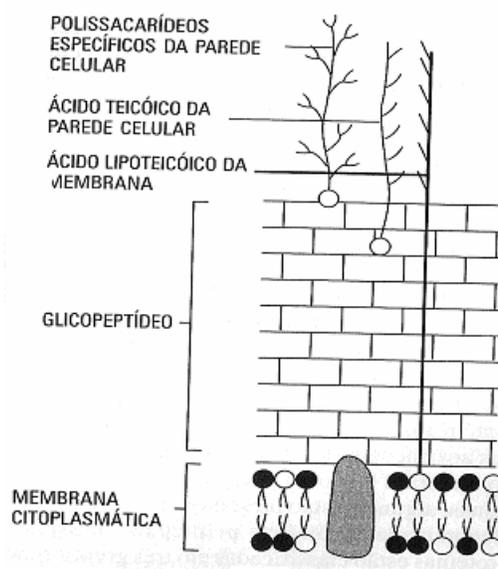


Figura 2.1. Estrutura da parede celular de bactérias gram-positivas (Koneman *et al.*, 2001).

Esta bactéria pode ser identificada por teste de detecção sorológica das formas capsulares. A cápsula bacteriana protege a célula, promove a concentração de nutrientes na

superfície celular, além de possuir a função de adesão da bactéria às células e superfícies de mucosas. A cápsula oferece ainda resistência à ação bactericida dos anticorpos séricos (Koneman *et al.*, 2001), dificultando a fagocitose destes microrganismos pelas células de defesa do hospedeiro (Gonçalves, 2002). A bactéria cresce tipicamente de modo difuso em caldo de cultivo e requer meios ricos em nutrientes para seu pleno desenvolvimento. Quanto aos requerimentos de oxigênio, são microrganismos anaeróbios facultativos (Kim e Weiser, 1998). Caracterizam-se ainda por serem catalase e citocromo-oxidase negativos, de forma que, com o envelhecimento da cultura, ocorre rápida perda da viabilidade celular devido ao acúmulo de peróxido de hidrogênio (Brandileone *et al.*, 1994).

Além de ser líder em infecções do trato respiratório humano, esta bactéria tem a habilidade de se disseminar no sistema sanguíneo, causando meningite e bacteremia (Gonçalves *et al.*, 2006). Primeiramente coloniza a nasofaringe daí pode se espalhar, causando infecções nas vias aéreas. Regulando a síntese de estruturas-chaves de superfície, pode entrar na corrente sanguínea, espalhando-se pelo corpo do hospedeiro (Gonçalves, 2002).

Esta habilidade de se adaptar a condições tão diferentes se deve à síntese de várias estruturas de superfície como polissacarídeo capsular (PS) e várias proteínas de membranas, que dão capacidade ao microrganismo de interagir com o hospedeiro (Gonçalves *et al.*, 2006).

Em 2006, Bricks e Berezin apontaram o pneumococo como o principal agente causador de pneumonias e doenças de mucosas menos graves, como sinusites e otites. O uso indiscriminado de medicamentos no tratamento destas doenças, por sua vez, vem sendo associado ao surgimento de cepas resistentes aos antibióticos. Os grupos de risco a doenças pneumocócicas invasivas estão listados no Quadro 2.1.

Quadro 2.1. Grupos de risco para doenças pneumocócicas invasivas (Bricks e Berezin, 2006).

Crianças menores de 5 anos
Idosos, maiores de 65 anos
Portadores de HIV ou em tratamento imunossupressor
Portadores de asplenia anatômica ou funcional (anemias hemolíticas)
Portadores de doenças crônicas como diabetes, cirrose, asma, etc.
Indivíduos submetidos à cirurgia de crânio
Pessoas que vivem em ambientes aglomerados como creches e escolas
Pessoas expostas à fumaça de cigarro
Após infecções virais, particularmente pós-influenza e infecção pelo vírus sincicial respiratório
Grupos étnicos como índios Navajos, Apaches e raça negra.

O corpo humano possui um sistema de defesa que o protege de microrganismos invasores. De acordo com Lima *et al.* (2001), o processo de imunização ocorre de duas formas distintas, ativa e passiva. A imunização ativa é resultante da atividade individual das células do próprio corpo; pode resultar de infecções conhecidas ou não, contraídas naturalmente ou resultante da administração de vacinas preparadas com o patógeno causador da infecção ou com alguma substância produzida por ele. Já a imunização passiva corresponde à injeção de soro imune ao patógeno. A grande vantagem da imunização ativa obtida com vacinas é que sua duração é medida em meses e anos, enquanto que a imunização passiva dura apenas algumas semanas.

Bricks e Berezin (2006) apontam ainda outro fator muito importante que deve ser considerado: bactérias que colonizam as vias respiratórias, como o pneumococo, sofrem freqüentes alterações genéticas, podendo adquirir genes de resistência de outros sorotipos. Como os sorotipos mais isolados de crianças com doenças invasivas e não invasivas são os que apresentam maior variação clonal, é essencial desenvolver novas vacinas contra o *Streptococcus pneumoniae*. Além disso, as vacinas devem ser desenvolvidas voltadas ao país ou região onde serão utilizadas, pois as cepas prevalentes variam de região para região (Gonçalves, 2002). De fato, de acordo com um estudo feito no Hospital da Faculdade de Medicina de São Paulo, foram identificadas as cepas 18C, 23F, 6B, 19F, 10A, 23B, 19A, 6A, 33 colonizando a nasofaringe dos pacientes estudados (Fonseca *et al.*, 2005). A vacina comercial Prevenar[®] contém os polissacarídeos correspondentes a apenas 4 destes sorotipos e este fato poderia ser uma explicação para a menor imunização conferida a crianças brasileiras.

Para produzir o componente principal da vacina, o polissacarídeo capsular, deve-se estabelecer condições de cultivo em biorreator que maximizem o crescimento de *S. pneumoniae* e a produção do PS.

2.1.2 Biorreator e Modos de Operação

A produção do polissacarídeo capsular constituinte da vacina conjugada é feita em cultivos do pneumococo em biorreator operado de forma descontínua. O reator descontínuo, também conhecido como reator batelada, é simplesmente um tanque com agitação que contém em um momento inicial todos os reagentes, dos quais são feitas medidas de tempos em tempos

para se aferir a extensão da reação. É o reator mais utilizado em laboratórios experimentais, pela simplicidade e adequação (Levenspiel, 1974).

Biorreatores tanque agitados são tipicamente cilíndricos, com fundo liso ou ligeiramente curvo. A razão entre a altura e o diâmetro do tanque pode variar, mas geralmente obedece a escala de 1:1, 2:1 ou 3:1 (Nielsen e Villadsen, 1994). Em tanques pequenos, a temperatura é controlada por meio de uma camisa na qual circulam fluidos de aquecimento e refrigeração. Já para tanques grandes são necessários trocadores de calor internos e externos. Em cultivos aeróbios, o gás (ar ou oxigênio) é normalmente suprido por um tubo posicionado no fundo do tanque próximo ao último agitador. A saída do gás geralmente tem um formato de anel perfurado. Muitos tanques agitados são equipados com defletores para impedir a formação de vórtice devido às forças de rotação introduzidas pela ação dos agitadores (Nielsen e Villadsen, 1994).

A Figura 2.2 esquematiza um reator tanque agitado, que pode ser operado de diferentes formas. Quando $F=F'=0$, a operação é em batelada, e não entram nutrientes nem saem produtos durante o cultivo. Em algumas situações, como em cultivos em que altas concentrações de substrato inibem o crescimento celular, é vantajoso iniciar o experimento com concentrações de substrato moderado e acrescentar meio nutriente no decorrer do cultivo, o que caracteriza uma batelada alimentada ($F \neq 0$ e $F'=0$).

Existem também, casos em que um produto do cultivo, quando em altas concentrações, inibe o crescimento do microrganismo. Nestes casos é interessante retirar o meio de cultivo saturado com estes metabólitos inibitórios. Este modo de operação com F e F' não nulos caracteriza o modo contínuo. No “quimiostato”, $F=F' \neq 0$ e o efluente do reator é o próprio meio de cultivo, incluindo as células. Alternativamente, conectando-se uma membrana de ultrafiltração ao biorreator e operando-se o conjunto com $F' \neq F \neq 0$ ou $F'=F \neq 0$, promove-se a retirada do meio de cultivo saturado, isento de células, que são recirculadas para o biorreator. Este modo de operação pode ser considerado uma variação da operação batelada. Devido à associação com a membrana de ultrafiltração, recebe a denominação de operação batelada alimentada em perfusão ou operação contínua com reciclo ($F' \neq F \neq 0$) (Shuler e Kargi, 2001).

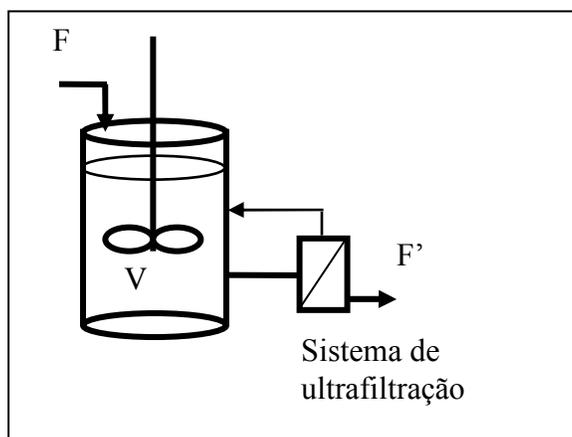


Figura 2.2. Esquema de reator tanque agitado. F corresponde à vazão de alimentação de meio suplementar, F' à vazão de retirada de meio de cultivo e V o volume. Operação em Batelada → F=F'=0; Operação em Batelada Alimentada → F≠0 e F'=0; Operação em modo contínuo → F≠0 e F'≠0 sem reciclo de células (ausência de filtro); Operação em Batelada alimentada com perfusão → F≠0 e F'≠0 com reciclo de células.

2.1.3. Crescimento Celular em Biorreator Batelada

O crescimento microbiano é resultado tanto da replicação celular quanto da mudança no tamanho celular. Os microrganismos podem crescer sob a influência de uma variedade de condições físicas, químicas e nutricionais. Em um meio com nutrientes apropriados, organismos convertem os nutrientes em compostos biológicos. Parte dos nutrientes é transformada em energia e parte é usada em biossíntese e formação de produto. Como resultado da utilização de nutrientes, a massa microbiana cresce com o tempo (Shuler e Kargi, 2001).

A velocidade de crescimento microbiano é diretamente proporcional à concentração celular, e é caracterizada pela velocidade de crescimento específica líquida, definida pelas equações (2.1) e (2.2), para cultivos em batelada.

$$\mu_{\text{liq}} = \frac{1}{C_X} \frac{dC_X}{dt} \quad (2.1)$$

$$\mu_{\text{liq}} = \mu - k_d \quad (2.2)$$

Nas quais C_X é a concentração mássica de células (g/L), t é o tempo (h) e μ_{liq} é a velocidade específica de crescimento líquida (por hora, h^{-1}). A velocidade específica de crescimento líquida é a diferença entre a velocidade de crescimento bruto (μ) e a velocidade de perda de massa celular devido à morte celular (k_D) (Shuler e Kargi, 2001).

Durante o crescimento em um reator operado em batelada simples contendo meio líquido, os organismos absorvem seletivamente os nutrientes do meio e os convertem em

biomassa levando a uma curva de crescimento típica. A curva de crescimento para cultivo batelada é caracterizada pelas seguintes fases em ordem cronológica: (1) fase lag, (2) fase de aceleração, (3) crescimento exponencial, (4) fase de desaceleração, (5) estacionária, e (6) fase de morte. A Figura 2.3 descreve um ciclo de crescimento em reator descontínuo, representando a cinética deste processo de crescimento.

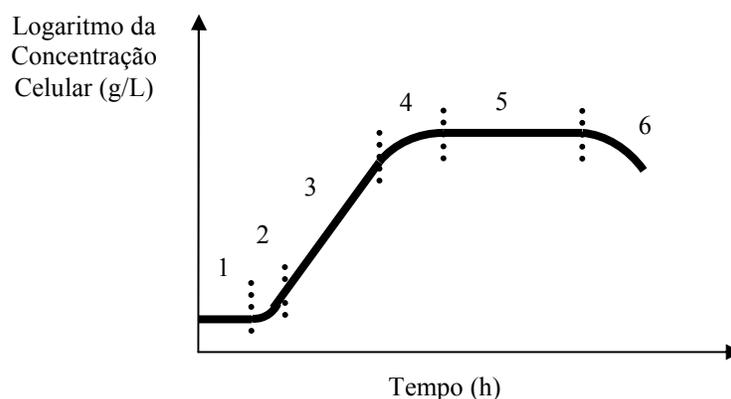


Figura 2.3. Curva típica de crescimento de uma população de bactérias, em um cultivo em batelada em meio líquido, na qual são evidenciadas fases: 1-lag; 2-aceleração; 3-crescimento exponencial; 4-desaceleração; 5-estacionária e 6-morte.

A fase lag ocorre imediatamente após a inoculação (que marca o ponto inicial do processo) e é um período de adaptação das células ao novo meio ambiente. Nesta fase, os microrganismos reorganizam seus constituintes moleculares. Dependendo da composição de nutrientes, novas enzimas são sintetizadas enquanto outras são inibidas e a célula se adapta às novas condições. A baixa concentração de alguns nutrientes e fatores de crescimento podem causar uma fase lag mais longa, o que é industrialmente indesejado. Quanto menor a fase lag mais produtivo se torna o processo (Shuler e Kargi, 2001).

Após a fase lag, a célula começa a crescer intensamente e sua velocidade de crescimento aumenta gradativamente. Esta é a fase de aceleração, que precede a fase de crescimento exponencial, também conhecida como fase logarítmica. Nesta fase, as células já estão adaptadas ao meio. As células podem se multiplicar rapidamente, e a massa e o número de células cresce exponencialmente com o tempo. Desde que a concentração de nutrientes não seja limitante, a taxa de crescimento pode ser expressa pela equação (2.3).

$$\frac{dC_X}{dt} = \mu_{liq} \cdot C_X \quad (2.3)$$

na qual $C_X = C_{X_{in}}$ em $t = t_{in}$, cuja integração, descrito pela equação (2.4), tem como resultado a equação (2.5) ou equação (2.6).

$$\int_{C_{X_{in}}}^{C_X} \frac{dC_X}{C_X} = \int_{t_{in}}^t \mu_{liq} \cdot dt \quad (2.4)$$

$$\ln \frac{C_X}{C_{X_{in}}} = \mu_{liq} \cdot (t - t_{in}) \quad (2.5)$$

$$C_X = C_{X_{in}} \cdot \exp(\mu_{liq} t - t_{in}) \quad (2.6)$$

nas quais C_X e $C_{X_{in}}$ são as concentrações celulares em t e em t_{in} , tempo de início da fase exponencial, respectivamente. A equação (2.5) define o perfil linear da fase exponencial de crescimento na Figura 2.3. Esta é uma forma de se calcular a velocidade máxima de crescimento, que é correspondente à maior inclinação da curva representada na Figura 2.3.

A fase de desaceleração segue a fase exponencial. Nesta fase, o crescimento desacelera devido a dois fatores: diminuição de nutrientes essenciais e acúmulo de toxinas ou produtos do crescimento.

A fase estacionária começa no fim da fase de desaceleração, quando a velocidade de crescimento é zero (não há divisão celular) ou a velocidade de morte é equivalente à velocidade de crescimento. Mesmo nesta fase, as células continuam ativas. De fato alguns produtos de interesse são produzidos na fase estacionária (como antibióticos e alguns hormônios), devido à regulação metabólica. Durante a fase estacionária a célula gasta energia para manter suas funções metabólicas essenciais. Esta energia é chamada de energia de manutenção (Shuler e Kargi, 2001).

A fase de morte segue a fase estacionária. Entretanto, a morte celular pode ter se iniciado durante a fase estacionária, e uma distinção clara entre esta duas fases não é sempre possível.

Como descrito, durante o cultivo as células crescem, consomem substrato e formam produtos. O produto pode se formar em momentos distintos durante o crescimento celular ou após o crescimento celular. O produto formado durante o crescimento é associado ao

crescimento, enquanto que o formado durante a fase estacionária não é associado ao crescimento. A equação de Luedeking-Piret, (2.7), fornece a velocidade específica de formação mista de produto (q_p).

$$q_p = \alpha\mu + \beta \quad (2.7)$$

na qual α corresponde ao termo associado ao crescimento enquanto β corresponde ao termo não associado, e μ à velocidade de crescimento bruto (Shuler e Kargi, 2001).

2.1.4 Modelagem Matemática do Bioprocesso

Para se alcançar bons resultados em um cultivo microbiano, algumas estratégias de otimização do processo devem ser consideradas. Muitas das estratégias disponíveis são baseadas no uso de modelos matemáticos do crescimento celular. Os modelos quantificam a relação entre a velocidade específica de crescimento do microrganismo e a concentração de substrato limitante, além de relacionarem o consumo de substrato ao coeficiente de rendimento do produto (Yurt *et al.*, 2003).

O estudo da influência das condições de cultivo na velocidade de crescimento se mostrou de fundamental importância, pois é a partir da expressão desta velocidade que se pode simular a produtividade em biorreatores. Como comentado anteriormente, existem diferentes tipos de modelos, dos mais complexos, representados por modelos estruturados que descrevem as principais reações dentro das células, aos mais simples, os não-estruturados, que procuram uma equação geral para descrever o crescimento. Os modelos podem ser ainda, segregados (distinguem os diferentes tipos de células presentes no reator) ou não segregados (consideram uniforme a população de um cultivo). Os modelos estruturados segregados são os mais realísticos, porém são também muito complexos computacionalmente (Shuler e Kargi, 2001). A Figura 2.4 apresenta a ordem de complexidade dos modelos cinéticos.

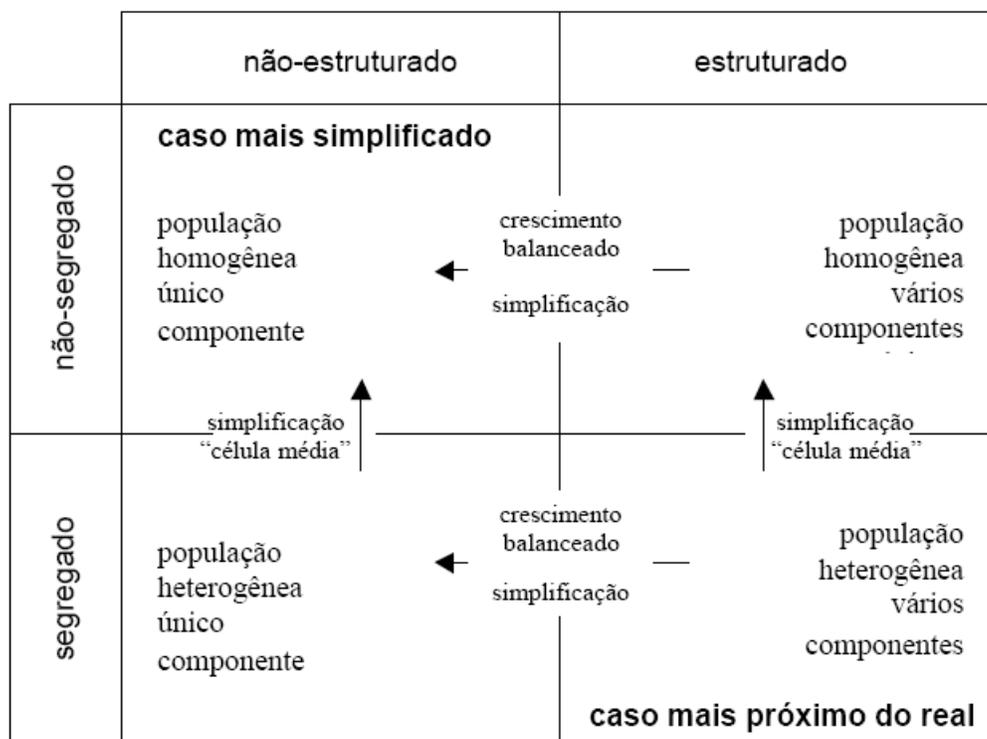


Figura 2.4. Perspectivas para a construção de modelos cinéticos do cultivo de microrganismos (Bailey e Ollis, 1986).

De acordo com Nielsen e Villadsen (1994) em modelos não-estruturados, a cinética de crescimento é modelada normalmente por apenas uma reação. Já os modelos estruturados são de certa forma melhorias nos modelos não-estruturados, na medida em que alguns mecanismos básicos do comportamento celular são incorporados ao menos qualitativamente. Em modelos estruturados, todos os componentes da biomassa são agrupados indiscriminadamente em algumas variáveis-chaves, que são supostamente representativas do comportamento celular. Desta forma a atividade microbiana se torna uma função não apenas das variáveis abióticas, que podem mudar em pequenos e constantes intervalos de tempo, mas também da composição celular, e desta maneira depende das condições do meio às quais as células foram submetidas. Existem muitas aproximações que visam estruturar a biomassa e que se subdividem em: Modelos estruturados simples, nos quais apenas alguns poucos componentes celulares são considerados; e Modelos altamente estruturados nos quais são considerados mais de 20 componentes intracelulares. A complexidade das reações metabólicas é tamanha que mesmo em modelos altamente estruturados os componentes celulares inclusos no modelo representam *pools* de enzimas diferentes, metabólitos, ou outros componentes celulares. As reações celulares

consideradas em modelos estruturados são conseqüentemente, empíricas desde que não podem representar a conversão entre componentes verdadeiros. Devido a essa natureza empírica, os modelos estruturados são baseados normalmente em alguns mecanismos celulares bem conhecidos, e assim, estes modelos têm a habilidade de simular muito bem algumas características dos experimentos (Nielsen e Villadsen, 1994).

Um aspecto essencial de modelagem estruturada é a verificação experimental, que requer medidas de componentes intracelulares (como DNA, RNA e proteínas) em diferentes condições de operação. Apesar dos avanços tecnológicos, estas análises de componentes intracelulares são trabalhosas, requerem equipamentos sofisticados e ficam restritas a estudos conduzidos em laboratório. Porém, sem o conhecimento das reações metabólicas torna-se impossível determinar os inúmeros parâmetros cinéticos tipicamente presentes em modelos metabolicamente estruturados (Moles *et al.*, 2003).

De acordo com Shuler e Kargi (2001) e Nelles, (2001), o modelador deve sempre escolher o modelo mais simples que descreve adequadamente o processo em questão, seguindo o princípio da parcimônia. Como os modelos estruturados são muito complexos, os modelos não-estruturados se apresentam como a primeira opção para a descrição de um processo de cultivo. O modelo não-estruturado mais comumente utilizado para descrever a cinética de crescimento de microrganismos é o modelo de Monod. Porém existem diversos modelos propostos na literatura e dentre eles, modificações do próprio modelo de Monod.

Considerando que o crescimento seja limitado apenas pela quantidade de nutrientes disponível no meio, e que exista apenas um nutriente limitante, a cinética de crescimento pode ser expressa pela equação de Monod, descrita na expressão (2.8).

$$\mu_{\text{Monod}} = \frac{\mu_M \cdot C_S}{K_S + C_S} \quad (2.8)$$

na qual μ_M é a velocidade máxima específica de crescimento. A constante K_S é a constante de saturação. A equação de Monod é semi-empírica, pois considera que a velocidade de reação seja limitada por uma reação enzimática suficientemente lenta. Gonçalves (2002) utiliza cinética de Monod com inibição pelo lactato para descrever o crescimento de pneumococo sorotipo 23F. Outras expressões empíricas, não estruturadas, comumente encontradas na literatura para

descrever o crescimento limitado pelo substrato são os modelos de Tessier, de Moser e de Contois, descritos nas equações (2.9), (2.10) e (2.11), respectivamente.

$$\mu_{\text{Tessier}} = \mu_M \cdot \left(1 - e^{-K \cdot C_S}\right) \quad (2.9)$$

$$\mu_{\text{Moser}} = \frac{\mu_M \cdot C_S^n}{K_S + C_S^n} \quad (2.10)$$

$$\mu_{\text{Contois}} = \frac{\mu_M \cdot C_S}{K \cdot C_X + C_S} \quad (2.11)$$

Nas quais C_X é a concentração celular e C_S a concentração de substrato. A equação de Moser é equivalente à de Monod para $n=1$ (Shuler e Kargi, 2001). A descrição dos parâmetros cinéticos presentes nas equações (2.8) a (2.11) se encontra no Quadro 2.2.

O modelo de Contois é mais apropriado para culturas com alta densidade celular, quando a velocidade específica de crescimento diminui com o aumento da concentração celular.

Quadro 2.2. Parâmetros das expressões cinéticas de Monod, Tessier, Contois e Moser.

μ_M (h^{-1})	Velocidade máxima específica de crescimento.
K_S (g/L)	Constante de saturação (o valor $1/K_S$ é interpretado como a afinidade da célula pelo substrato (Kovárová-Kovar e Egli, 1998)).
K	Parâmetro da equação de Tessier (L/g_S) e de Contois (g_S/g_X) análogo ao K_S de Monod
n	Parâmetro do modelo de Moser, expoente da concentração de substrato
Unidades: g → grama; mg → miligrama; L → litro; h → hora; g_i → grama de i, com $i=X, S, L$ e P (célula, substrato, Lactato e Produto, respectivamente)	
O parâmetro μ_M é comum a todos os 4 modelos, K_S pertence aos modelos de Monod e Moser, n apenas ao modelo de Moser enquanto que K ao modelo de Tessier.	

Beyenal *et al.* (2003) avaliaram os modelos de Monod, Moser, Tessier e de Contois, para múltiplos substratos e dentre estes modelos, o que melhor descreveu a cinética de crescimento de *Pseudomonas aeruginosa* foi o modelo de Tessier.

Além de descrever o crescimento em função de um substrato limitante, muitas vezes é necessário incluir inibições no modelo. O microrganismo pode apresentar seu crescimento inibido pelo substrato, por algum produto ou, ainda, por alguma toxina no meio. Várias expressões estão disponíveis na literatura para descrever cada situação relatada (Shuler e Kargi, 2001; Levenspiel, 1974; Aiba, 1973).

As concentrações das principais variáveis de um processo podem ser descritas associando-se uma expressão para a velocidade de reação (equações (2.8) a (2.11)) e equações de balanço material. Neste trabalho, consideramos modelos biológicos dinâmicos determinísticos não-lineares, descritos por meio de equações diferenciais ordinárias (EDOs). Neste caso a equação (2.12) compõe a formulação do espaço de estados:

$$\frac{dx}{dt} = f(\mathbf{x}, \mathbf{x}_a, \mathbf{w}, \mathbf{u}, t) \quad (2.12)$$

sendo:

x: vetor de estado do sistema (ou seja, concentrações de biomassa, substrato, produto, metabólitos, concentração de oxigênio dissolvido, fração molar de gás carbônico e oxigênio na corrente gasosa efluente);

x_a: vetor de variáveis representando a fase abiótica no biorreator (e.g, temperatura, pH);

w: vetor de parâmetros do modelo;

u: vetor de variáveis que podem ser manipuladas (velocidade de agitação, vazão de alimentação);

t: tempo.

Destaca-se ainda que, no vetor de estado do sistema, as variáveis primárias são aquelas adquiridas em tempo real, enquanto as demais são medidas geralmente por amostragem, mas podem ser estimadas a partir das variáveis primárias e dos algoritmos de inferência de estado.

A construção de um modelo começa pela definição do objetivo e algum conhecimento a priori (i.e. dados preliminares, análises básicas e hipóteses iniciais). A partir dos dados disponíveis, é feita a estimativa dos parâmetros gerando o primeiro modelo. Este modelo inicial deve ser validado com dados de outros experimentos similares, que em muitos casos revela algumas deficiências. Neste caso a estrutura do modelo deve ser refinada e reavaliada até que um modelo satisfatório seja encontrado (Rodriguez-Fernandez *et al.*, 2006).

Além da modelagem convencional, podem-se utilizar modelos híbridos. As estruturas híbridas combinam os princípios básicos de modelagem com técnicas quimiométricas para extrair informações dos dados do processo. A estrutura híbrida mais adotada de sistemas de biorreatores combina equações macroscópicas de balanço material/energia com redes neurais artificiais tal como a PMC (perceptron multi-camada) e a FBR (função de base radial). A função

das redes neurais nestas estruturas é da modelagem não paramétrica de reações cinéticas desconhecidas (Peres *et al.*, 2007).

Após a determinação dos valores dos parâmetros do modelo e validação, é possível utilizar o modelo para simular diferentes condições de operação e identificar aquelas que levam à maior produtividade. Porém, a determinação dos valores dos parâmetros é um problema matematicamente complexo e a utilização de técnicas computacionais de busca heurística ou determinística ou, ainda, uma associação de ambas, é recomendável e até necessária no caso de sistemas nos quais o número de parâmetros é relativamente alto.

2.2 Técnicas de Inteligência Computacional Relevantes ao Problema

Esta seção apresenta os principais conceitos relacionados a duas subáreas da inteligência computacional relevantes ao trabalho sendo desenvolvido: a relacionada a métodos de busca heurística, com foco no *simulated annealing* (SA) e redes neurais. Neste trabalho a busca heurística realizada pelo SA será utilizada para a identificação do melhor conjunto de valores associados a parâmetros que governam três importantes modelos matemáticos que representam o crescimento do *S. pneumoniae* em uma cultura.

Redes neurais, particularmente um conjunto delas organizado em um comitê, serão utilizadas para a identificação da fase do processo de crescimento de *Streptococcus pneumoniae* em um cultivo. Com base na fase identificada, um modelo matemático será alocado para a inferência de estado e desta forma, auxílio no controle do experimento.

2.2.1 Busca Heurística do Melhor Conjunto de Valores de Parâmetros que Otimizam um Modelo Determinístico Via *Simulated Annealing*

A simulação de processos biotecnológicos requer o conhecimento preciso dos parâmetros cinéticos que, em geral, são obtidos pelo ajuste de um conjunto de dados experimentais a uma equação de crescimento. Este ajuste dos dados ao modelo tem sido feito por métodos não lineares de estimativa de parâmetros. Dentre estes métodos, o algoritmo de Levenberg-Marquardt (LM) tem sido o mais utilizado (Eftaxias *et al.*, 2002).

Este processo de ajuste é feito comparando os valores dos dados medidos experimentalmente aos valores calculados pelo modelo através de uma função objetivo (FO). A

estimativa de parâmetros também é conhecida como calibração do modelo (Rodriguez-Fernandes *et al.*, 2006).

Encontrar o conjunto ótimo de valores de parâmetros de um modelo significa encontrar o mínimo de uma função objetivo que contabiliza, por exemplo, o erro cometido ao se estimar o dado experimental usando o modelo. Moles *et al.* (2003) e Hodgson (2004) utilizam como função objetivo uma medida ponderada dos desvios entre os dados calculados pelo modelo e os dados medidos experimentalmente. No presente trabalho, a FO mede a diferença entre os dados experimentais e os dados calculados pelo modelo. É mais comumente expressa pela diferença dos mínimos quadrados. A função objetivo pode ser expressa pela equação (2.13), na qual C_i^{mod} refere-se à variável i calculada pelo modelo enquanto que C_i^{exp} à variável i medida experimentalmente, com $i \in \{X, S, L \text{ e } PS\}$, t corresponde ao tempo e $w_{C_i}(t)$ corresponde aos pesos associados a cada variável, isto é, à normalização.

$$f(C^{\text{mod}}, C^{\text{exp}}) = \sum_i \left\{ \sum_{t_0}^{t_f} \left[\left(\frac{C_i^{\text{mod}}(t) - C_i^{\text{exp}}(t)}{w_{C_i}(t)} \right)^2 \right] \right\} \quad (2.13)$$

Moles *et al.* (2003) define a função peso de acordo com a equação (2.14).

$$w_{C_i}(t) = \text{máximo}(C_i^{\text{exp}}(t)) \quad (2.14)$$

Outra forma de definir a função peso seria pelo maior valor das respectivas variáveis, semelhante à equação (2.14), porém sem a dependência do tempo (Hodgson *et al.*, 2004). No presente trabalho, optou-se por definir a função w de acordo com equação (2.15), a qual tem como base a ponderação pelo desvio calculado a partir da equação (2.16) (Horta *et al.*, 2007).

$$w_{C_i}(C^{\text{exp}}) = \sqrt{\frac{\sum_{t_0}^{t_f} (C_i^{\text{exp}}(t) - M)^2}{n^{\circ} \text{ Instâncias} - 1}} \quad (2.15)$$

$$M = \frac{\sum_{t_0}^{t_f} C_i^{\text{exp}}}{n^{\circ} \text{ Instâncias}} \quad (2.16)$$

onde o **n.º Instâncias** é o número de instâncias de dados disponíveis para o processo de ajuste do modelo. Instâncias são definidas pelos valores das variáveis que descrevem o processo a cada intervalo de tempo.

O algoritmo estocástico *Simulated Annealing* (SA) é um procedimento computacional de grandes potencialidades que garante a convergência global da função objetivo desde que um determinado número de iterações seja realizado. Em contrapartida, algoritmos determinísticos, tal como o Levenberg-Marquardt (LM), garantem a convergência da FO, porém podem se prender a mínimos locais (Moles *et al.*, 2003; Mendes e Kell, 1998). Uma alternativa para evitar esse problema, é iniciar o algoritmo inúmeras vezes, de forma aleatória, que é um procedimento muito adotado, porém questionável, uma vez que existe a probabilidade dos mesmos mínimos locais poderem ser encontrados várias vezes (Moles *et al.*, 2003).

O SA é iniciado com a escolha, pelo usuário, dos dados experimentais, dos parâmetros do modelo a serem ajustados e com a escolha da função objetivo. A partir disso, o algoritmo faz uma busca sobre o espaço de todos os possíveis valores de parâmetros do modelo, com o objetivo de encontrar o conjunto ótimo de valores dos parâmetros. A cada passo a função objetivo é recalculada. Assim uma nova solução durante a busca é aceita ou não, de acordo com o critério Metropolis, representado na equação (2.17) (Chao *et al.*, 2007).

$$P_{AB} = \min \left\{ 1, e^{\left[\frac{-(S_A - S_B)}{T} \right]} \right\} \quad (2.17)$$

onde P_{AB} é a probabilidade de aceitação da solução B em detrimento da solução A, e T é um valor que controla o mínimo de iterações realizadas.

A diminuição controlada do valor da variável T tem a finalidade de garantir que o sistema não fique preso em mínimos locais (estado meta-estável), de forma que o estado estável (mínimo global) seja alcançado invariavelmente (Chao *et al.*, 2007).

Na literatura podem ser encontrados trabalhos descrevendo a utilização de diferentes métodos de estimativa de parâmetros em modelos cinéticos, tais como: simplex, utilizado por Henriques *et al.* (2005) para ajuste do modelo matemático de crescimento de *Neisseria meningitidis* na produção de polissacarídeo capsular; “Quasi-Newton” ou Marquardt,

utilizado por Baptista Neto (2004), para a determinação dos parâmetros do modelo de crescimento de *Streptomyces clavuligerus* na produção de ácido clavulânico; Marquardt, utilizado por Silva (2003) para estimar os parâmetros do modelo cinético do crescimento de *Bacillus megaterium* na produção de penicilina G acilase; Algoritmos Genéticos, utilizados por Park e Froment (1998) para estimativa de parâmetros de modelo cinético da desidrogenação catalítica de etanol; Simulated annealing e Levenberg-Marquardt para a estimativa dos parâmetros de modelos com até 38 parâmetros, da oxidação catalítica de fenol (Eftaxias *et al.*, 2002).

Se o modelo matemático que descreve o processo for apropriado e o método de ajuste de parâmetros for bem sucedido, o resultado é um modelo cuja implementação computacional resulta em uma ferramenta capaz de simular novos experimentos. Uma outra forma de utilização da ferramenta é como parte de um sistema supervisor, no qual a trajetória de uma ou mais variáveis é monitorada em tempo real durante os experimentos, permitindo a inferência dos valores das variáveis mais importantes, os quais são comparados com os previstos, viabilizando o diagnóstico do desempenho do processo.

2.2.2 Principais Conceitos Relativos a Redes Neurais Artificiais e Sua Aplicação

Como comentado anteriormente, devido à variação no metabolismo celular durante um cultivo, é essencial o reconhecimento das fases de crescimento para o processo de otimização; a inferência da fase de um cultivo permite que se possa estender ou retardar a fase de interesse, fornecendo controladamente o substrato, de acordo com a cinética de formação do produto de desejado.

2.2.2.1. Conceitos básicos (Palma Neto e Nicoletti, 2005)

Muitos livros sobre Redes Neurais (RNs) não definem claramente o que é uma RN. A maioria dos pesquisadores da área, entretanto concordaria que, definida de uma maneira informal uma RN é uma rede com muitos processadores simples (cada um deles tendo, possivelmente, uma pequena quantidade de memória local), conectados por meio de canais de comunicação (conexões), aos quais, usualmente estão associados valores (pesos) numéricos. Uma rede neural pode então ser caracterizada:

- a) por seus processadores básicos, os neurônios;
- b) pela função de ativação, que representa o estado de um neurônio;
- c) pelo padrão de conexão existente entre neurônios;
- d) por seu algoritmo de treinamento (chamado também de algoritmo de aprendizado).

A função de ativação de um neurônio estabelece seu estado, chamado de ativação. O valor da função de ativação depende das entradas que o neurônio recebe em um determinado momento. Essas entradas podem ter origem externa ou ser provenientes de outros neurônios. Neurônios geram um valor de saída que é dependente de seu valor de ativação e interagem enviando suas saídas por meio de conexões. Como essas conexões geralmente são ponderadas, o valor de saída do neurônio é modificado pelo peso associado à conexão. Em uma situação de treinamento ou de uso de uma RN, um neurônio envia sua saída aos vários outros neurônios aos quais está conectado; estes, por sua vez, recebem aquela saída modificada pelo peso da respectiva conexão como entrada e procedem de maneira análoga.

O comportamento de um neurônio é determinado pelas funções a ele associadas e pelas conexões por meio das quais recebe entradas e transmite sua saída. Um modelo geral de neurônio como descrito em Tschichold-Gürman (1995), pode ser visualizado na Figura 2.5. Nesse modelo geral, um neurônio é caracterizado por meio da função soma, da função ativação e da função saída.

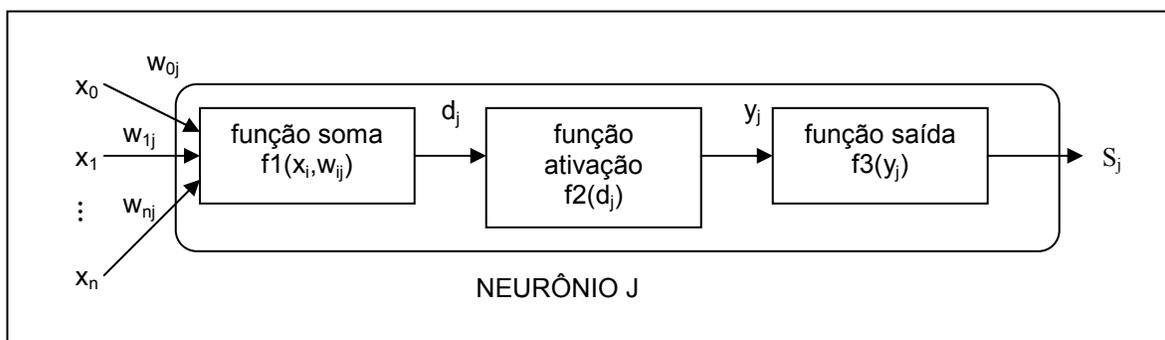


Figura 2.5. Modelo geral de neurônio. x_i são as entradas do neurônio j , w_{ij} são os pesos associados a cada entrada, d_j é a saída da função soma ($f1$), y_j é a saída da função de ativação ($f2$) e S_j a saída da função saída ($f3$).

Primeiro, a função soma combina as entradas que chegam ao neurônio em um valor. Como pode ser visto na figura, essas entradas dependem dos valores transmitidos e do peso associado às conexões por meio das quais os valores são transmitidos. Uma função soma frequentemente usada é o produto interno definido pela equação (2.18).

$$d_j = \sum_{i=0}^n w_{ij} \cdot x_i \quad (2.18)$$

O nível de atividade ou estado de um neurônio é calculado pela função ativação. Há uma grande variedade de funções que são usadas para calcular a ativação de um neurônio. A Figura 2.6 mostra o gráfico de algumas das mais comuns.

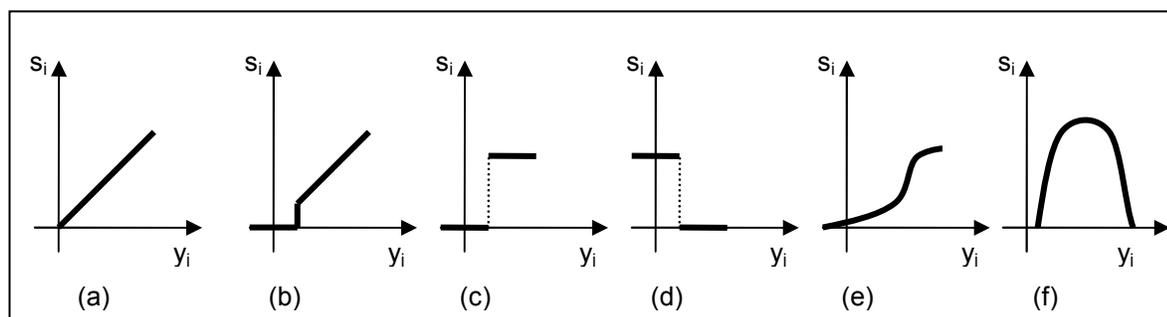


Figura 2.6. Diferentes funções ativação: a) linear; b) *threshold* linear; c) degrau; d) degrau reverso; e) sigmóide; f) gaussiana.

A função saída associa a ativação de um neurônio a um valor de saída. Usualmente, a função saída é a função identidade $S_j = f_3(y_j) = I(y_j) = y_j$.

É conveniente abordar os neurônios de uma RN como organizados em camadas, de maneira a caracterizar o padrão de conexão existente entre eles. Geralmente os neurônios que pertencem a uma mesma camada têm as mesmas funções (soma, ativação e saída) e o mesmo padrão de conexão a outros neurônios.

Em muitas redes neurais, os neurônios que pertencem a uma mesma camada ou estão completamente interconectados ou, então, não estão conectados de maneira alguma. A organização dos neurônios em camadas, bem como os padrões de conexão dentro e entre camadas, caracteriza a arquitetura (também chamada de topologia) de rede neural.

Geralmente, em uma RN, a camada de neurônios identificada como camada de entrada é formada por neurônios que servem apenas como *buffers* para os valores de entrada para a rede. Cada uma das funções (soma, ativação e saída) associada a cada um dos neurônios dessa camada replica o valor que recebe.

A maneira de contar o número de camadas existentes em uma RN não está padronizada. Algumas abordagens contam o número de camadas de neurônios; ainda assim, há divergência entre elas, pois algumas incluem na contagem a camada de entrada e outras não. Outras abordagens contam as camadas de pesos em vez de contar as de neurônios. Para evitar

ambigüidade, os assuntos tratados neste trabalho farão referência explícita ao número de camadas intermediárias de neurônios (isto é, aquelas camadas que estão entre a camada de entrada e a saída de uma rede).

O algoritmo de treinamento (ou aprendizado) de uma rede é um conjunto de regras por meio das quais os pesos das conexões são ajustados, usando, para o ajuste, um conjunto de exemplos que representa a função a ser aprendida. Esse conjunto de exemplos (também chamadas instâncias) é geralmente chamado de conjunto de treinamento. Uma RN geralmente aprende a partir de um conjunto de treinamento; o aprendizado se reflete na capacidade de generalização que a rede exibe, quando usada em novas situações.

Assim como acontece com qualquer método de aprendizado indutivo de máquina, o conjunto de treinamento é de fundamental importância para o aprendizado da rede. Não existem métodos de treinamentos de redes que criem informações que não estejam já presentes no conjunto de treinamento. Uma rede é sempre um reflexo do conjunto de treinamento usado para treiná-la.

2.2.2.2. Aplicações em bioprocessos

Há inúmeros trabalhos na literatura que ilustram diferentes aplicações de redes neurais em bioprocessos, envolvendo classificação, filtração de ruídos e estimativa de velocidades de reação, dentre outros.

De acordo com Simon *et al.* (1998) uma das maiores aplicações de redes neurais em processos de fermentação é em classificação de fases. Os autores dividem as fases de crescimento do *B. subtilis* nas quatro principais fases, lag, exponencial, estacionária e morte, a partir de medidas de densidade óptica obtidas de experimentos em frascos agitados com retirada de amostras a cada 2 horas. Os autores utilizaram dados interpolados para treinamento de uma rede neural *feed-forward* com duas camadas intermediárias, com trinta neurônios na primeira e 10 na segunda camada e quatro neurônios tanto na camada de entrada quanto na camada de saída. As fases de crescimento foram subdivididas de acordo com a inclinação da curva do logaritmo na base 2 da densidade óptica. Uma inclinação de 0,09 demarcou a fase lag e a estacionária. Uma inclinação superior a 0,09 foi definida como fase exponencial e a inclinação negativa como fase de morte. Os vetores de entrada das redes neurais foram formados por 4 variáveis, tempo, densidade óptica (DO), primeira derivada da DO no tempo (dDO^1) e segunda derivada da DO no

tempo (dDO^2). Simon *et al.* (1998) obtiveram um desempenho de 98% em classificações corretas com as quatro variáveis de entrada (t , DO , dDO^1 , dDO^2). Este desempenho diminuiu para 67% ao utilizar como entradas apenas o tempo e a DO .

Um trabalho mais recente sobre classificação de fases de crescimento com redes neurais publicado por Simon e Karim (2001), subdivide as fases de crescimento do *Bacillus subtilis* em três fases (lag, exponencial e estacionária). Neste caso, os autores utilizaram como entrada da rede neural as variáveis tempo e a terceira derivada da DO no tempo (dDO^3). As fases foram divididas com base na dDO^3 . Redes neurais probabilísticas foram utilizadas. Os autores levantam várias vantagens destas redes em relação às redes “backpropagation” tais como: menos tempo de treinamento e a não necessidade de otimizar o número de neurônios na camada intermediária (o número de neurônios na camada intermediária é controlado pelo número de entradas). Porém a grande desvantagem das redes neurais probabilísticas, apontada pelos autores, é a grande dimensão da rede. Com a configuração de rede neural experimentada os autores obtiveram acurácia de 95% no teste e 99% no treinamento. Além disso, levantam a possibilidade do erro de classificação estar relacionado à terceira derivada, ressaltando a importância da escolha das variáveis de entrada que maximizam as saídas do processo.

Bacauskiene e Verikas (2004) relatam algumas técnicas de composição de comitês mais utilizadas: voto majoritário, média, média ponderada, aproximação Bayesiana, teoria Dempster-Shafer, contagem de Borda, agregação probabilística, agregação por redes neurais entre outras.

Silva (2003) propõe o uso de redes neurais artificiais em bioprocessos com quatro finalidades: (1) como filtro de ruídos dos dados experimentais com redes recorrentes; (2) na identificação das três principais fases de crescimento (fase lag, exponencial e estacionária) do *Bacillus megaterium*; (3) na inferência de C_X no processo de produção de PGA (penicilina G acilase) por *Bacillus megaterium* e (4) na estimativa da concentração de produto através de enfoque híbrido, usando a velocidade de crescimento inferida por rede “feedforward”, acoplada ao fator de rendimento célula-produto estimado no ajuste de modelos não-estruturados. Silva (2003) utiliza uma rede neural PMC “feedforward” para o reconhecimento das fases de crescimento. A separação das fases para o treinamento da rede foi feita visualmente (usando a curva experimental de C_X alisada pelo método B-spline do MatLab 5.2). A entrada da RN foi o tempo (valor atual) e a variável Y_{CO_2} (tempo atual e atrasado 9 amostragens), totalizando um

vetor de 11 dimensões. A rede apresentou apenas uma saída entre 0 e 2 (0 representando fase lag, 1 fase exponencial e 2 a fase estacionária) e 10 neurônios na camada intermediária. Os autores apresentam resultados bem sucedidos na classificação das fases pela rede neural.

Patnaik (2002) ressalta a grande dificuldade em fazer o *scale-up* de bancada para um de biorreator em escala piloto (industrial) devido à grande ocorrência de ruídos e à heterogeneidade (mistura incompleta) do meio de cultivo. Patnaik (2002) apresenta alguns trabalhos de otimização de cultivos com o uso de redes neurais. O autor defende que o grau de mistura, a extensão da filtragem dos ruídos e a distribuição do número de cópias plasmidiais (em uma fermentação recombinante) pode ser controlada por redes neurais de forma efetiva *on-line*. Para a produção de estreptoquinase por *Streptococcus equisimilis* a otimização da mistura do meio com rede neural, aumentou a atividade máxima de estreptoquinase em 34%, e uma rede neural otimizada para filtrar os ruídos forneceu mais 22% de aumento (Patnaik, 1999-b). A junção da rede neural filtro e a rede neural de controle aumentou a atividade máxima de estreptoquinase em 42% (Patnaik, 2001-c). Patnaik (2002) comenta outros resultados similares obtidos com a utilização de redes neurais em cultivos de *Escherichia coli* para obtenção de desidrogenase gliceraldeído-3-fosfato (Patnaik, 2001-b) e β -galactosidase (Patnaik, 2001-a, e Patnaik, 1999-a).

Linko *et al.* (1999) citam um exemplo de utilização de redes neurais como *software sensor*. No exemplo os autores avaliam várias configurações de rede neural para estimativa e predição de atividade da glucoamilase e da concentração de biomassa no processo de produção de glucoamilase a partir de cultivos de *Aspergillus niger*. 2000 ciclos de iteração e taxa de aprendizado de 0,001 foram utilizados para o treinamento das redes PMC “feedforward”. Com uma arquitetura de rede com 3 neurônios na camada de entrada, 7 na única camada intermediária e 2 na camada de saída, e mantendo como entrada as variáveis velocidade de liberação de dióxido de carbono, produção total de dióxido de carbono e consumo de amônia, a estimativa da atividade de glucoamilase e a concentração de biomassa apresentaram valores para o coeficiente de determinação (R^2) iguais a 0,990 e 0,964, respectivamente (Linko *et al.*, 1999).

Linko *et al.* (1999) utilizam também um exemplo de aplicação de redes neurais na estimativa da biomassa e da atividade de lipase em cultivo de *Candida rugosa*. Neste caso os autores também obtêm R^2 acima de 0,9 para a configuração de redes com uma camada intermediária com 3, 7 e 9 neurônios; tendo como variáveis de entrada CO_2 total produzido, concentração de O_2 dissolvido, atividade de lipase e concentração de biomassa.

Capítulo 3 – OBTENÇÃO, DESCRIÇÃO E TRATAMENTO DOS DADOS EXPERIMENTAIS

3.1. Obtenção dos Dados Experimentais

Os dados experimentais foram obtidos no período de agosto de 2005 a setembro de 2007, com cepas liofilizadas ST 433/03 I e II de *S. pneumoniae* sorotipo 6B, fornecidas pelo Serviço de Bacteriologia do Instituto Adolfo Lutz. Os cultivos foram realizados no Laboratório de Fermentação do Centro de Biotecnologia do Instituto Butantan pela pesquisadora Viviane M. Gonçalves e pela aluna de doutorado Talita S. Carmo, como parte das atividades desenvolvidas dentro do Projeto Temático “Escalonamento de bioprocessos para a produção de vacinas”, PRONEX, CNPq, FAPESP, Proc. 03/10125-3. Os dados utilizados neste trabalho foram coletados em dois tipos de experimentos: cultivos em batelada com suplementação de glicose por meio de pulsos e cultivo em batelada alimentada em biorreator acoplado a sistema de retenção de células. Os experimentos em batelada 1, 2, 3 e o de perfusão foram realizados com criotubos preparados a partir da cepa liofilizada I. Já os experimentos em batelada 4 e 5 foram realizados com criotubos preparados a partir da cepa liofilizada II. A cepa II foi requisitada ao Instituto Adolfo Lutz devido à perda por descongelamento da cepa I durante falta de energia elétrica. A seguir é apresentada uma breve descrição das condições de cultivo empregadas, da metodologia analítica e dos dados experimentais coletados.

3.1.1. Cultivos de *S. pneumoniae* em Batelada

Foram disponibilizados dados de 5 experimentos, dos quais, dois foram acompanhados pelo autor deste trabalho. Os cinco experimentos (nomeados Exp₁, Exp₂, Exp₃, Exp₄ e Exp₅) foram realizados sob condições similares, em batelada com um pulso de meio suplementar contendo glicose entre a quinta e a oitava hora do cultivo. Os experimentos 4 e 5 foram realizados com um lote diferente da mesma cepa da bactéria. O meio de cultura utilizado nas fermentações foi o meio de Hoeprich modificado (Hoeprich, 1955; Gonçalves *et al.*, 2002), cujas principais características estão descritas na Tabela 3.1.

Tabela 3.1. Meio de cultivo das 5 fermentações usadas como dados.

Substância	Concentração	Substância	Concentração
Hidrolisado ácido caseína	37,5 g/L	L-Glutamina	0,78 g/L
Glicose monohidratada	25 g/L	Asparagina	0,125 g/L
Dialisado de extrato de levedura	25 g/L	Solução de ácido tioglicólico 10%	1 mL/L
Fosfato bipotássico	6,25 g/L	solução de sais	2 mL/L
Bicarbonato de sódio	1 g/L	Colina	0,012 g/L
Solução de sais			
Sulfato de magnésio		250 g/L	
Sulfato ferroso		2,50 g/L	
Sulfato de zinco		0,40 g/L	
Sulfato de manganês		0,18 g/L	
Ácido clorídrico		10 mL/L	

Os cultivos em batelada foram conduzidos em biorreatores BioFlo 2000 (New Brunswick Scientific Inc.), sendo que, com exceção do Exp₁, todos foram monitorados por um programa desenvolvido em ambiente de programação LabView 7.1 (National Instruments) (Jesus *et al.*, (2007); Cruz *et al.*, (2006)). Além disso, o experimento Exp₄ foi o realizado com aquisição automática (*on-line*) da massa de solução de NaOH. O meio de cultivo foi esterilizado por filtração em membrana de 0,22 µm. Utilizou-se estoque celular congelado a -70°C para inocular o mesmo meio. Polipropilenoglicol foi adicionado quando necessário para controlar a formação de espuma, enquanto que o pH foi controlado automaticamente com adição de solução de NaOH a 5M. Demais informações estão descritas na Tabela 3.2. O pulso de glicose foi de 100 mL, contendo glicose na concentração de 50% (m/v), adicionado quando a glicose no meio estava próxima da exaustão, com o objetivo de prolongar um pouco mais o experimento.

Tabela 3.2. Condições experimentais utilizadas nas fermentações.

	Exp ₁	Exp ₂	Exp ₃	Exp ₄	Exp ₅
Vazão de N ₂ (L/min)	1,0	1,0	1,5	1,0	1,0
pH	7,0-7,2	7,0-7,3	7,2-7,4	7,0-7,4	7,0-7,4
Agitação mecânica (rpm)	200	200	250	150	150
Temperatura (°C)	37	37	36	37	37
Volume (L)	5,0	8,8	8,8	4,6	4,6
Concentração inicial de glicose (g/L)	29,3	30,9	27	26,7	28,5
Inóculo (L)	0,5	0,8	0,8	0,4	0,4
Tempo de incubação (h)	12	12	13	13	13
Momento de aplicação do pulso de glicose (h)	5	7	6	6	6

Os experimentos Exp₁, Exp₂, Exp₃ e Exp₅ foram feitos com retiradas de amostras de hora em hora para a determinação das concentrações de biomassa, glicose, polissacarídeo

capsular, ácido láctico e outros metabólitos. O volume da solução de NaOH adicionado para controlar o pH também foi registrado a cada hora, manualmente. Já no experimento Exp₄, além da amostragem a cada hora, os dados da massa de solução de NaOH foram adquiridos em tempo real conectando-se a balança eletrônica (APX-6001, Denver Instruments Company) pela porta serial (module Breakout Box RS 232) ao computador com programa desenvolvido em LabView 7.1, empregado para monitoramento e aquisição de dados.

Das amostras retiradas de hora em hora uma alíquota foi convenientemente diluída e utilizada imediatamente para acompanhamento da concentração celular por meio de determinação da densidade ótica (DO) em espectrofotômetro (Hitachi U-1800 Spectrophotometer) no comprimento de onda de 600 nm. O restante da amostra foi centrifugado a 3.220 g (g - aceleração da gravidade) a 4°C por 30 minutos. O sobrenadante foi utilizado na determinação das concentrações de glicose (C_S), ácido láctico (C_L) e polissacarídeo (C_{PS}). Já os sedimentos foram utilizados para medidas da concentração de biomassa (C_X) pelo método gravimétrico após re-suspensão em solução salina (0,9%), re-centrifugação e secagem em estufa a 60°C. A concentração de glicose e de ácido láctico foram medidas pelos métodos: colorimétrico de glicose oxidase (Enz-Color Biodiagnóstica, Brasil) e por HPLC, respectivamente. As concentrações de ácidos orgânicos e açúcares foram determinadas com coluna de HPLC Aminex HPX87H (Biorad), 0,005M ácido sulfúrico como fase móvel, fluxo de 0,6mL/min, temperatura 60° C. O volume de injeção na coluna foi de 20µL. A concentração de PS foi determinada após diálise em água destilada, por quantificação de ramnose (Dische e Shettles, 1948).

Os dados dos Exp₁, Exp₂, Exp₃, Exp₄ e Exp₅ foram utilizados no treinamento das redes neurais com o objetivo de aprendizado da fase do processo tratado na Seção 4.3; no ajuste dos modelos não estruturados clássicos, tratado na Seção 4.1, e segregados em fases, na Seção 4.4 e para a construção e avaliação da ferramenta de inferência, Seção 4.5.

3.1.2. Cultivo de *S. pneumoniae* em Biorreator Acoplado a Sistema de Retenção de Células e Retirada de Meio

Um sexto experimento Exp₆ foi conduzido em condições diferentes dos descritos anteriormente. Este experimento foi iniciado como uma batelada simples, sem adição de pulso. Quando a concentração de glicose estava próxima da exaustão, iniciou-se a alimentação com

meio suplementar, caracterizando a operação como batelada alimentada. Além disso, um sistema de microfiltração foi acoplado ao biorreator permitindo a retirada do filtrado constituído pelo meio de cultivo saturado em metabólitos e isento de células, que permaneciam retidas no biorreator. A Figura 3.1 mostra o conjunto experimental empregado para a realização deste experimento. A complexidade de operação é compensada pelas perspectivas de obtenção de altas densidades celulares em consequência da retenção de células, da minimização do acúmulo de metabólitos inibidores no meio de cultivo e da suplementação com nutrientes.

Para a realização do Exp₆, foi acoplado ao biorreator um sistema constituído por 3 membranas Durapore de poro 0,22 μ m, 50cm² (Millipore[®] / PelliconXL, lote no P2JN6608-27, cat. no PXGVPPC50). Foi empregado o biorreator BioFlo 3000 (New Brunswick) com monitoramento e aquisição de dados em LabView 7.1, já mencionado anteriormente. A operação em perfusão foi iniciada após 6 horas de cultivo, coletando-se o permeado produzido em um frasco posicionado sobre uma balança. A medida da massa de permeado produzida foi utilizada para estabelecer o valor da vazão de alimentação de meio suplementar, definindo a operação do reator a volume constante. Devido a problemas de entupimento da membrana de filtração, houve muita variação na vazão de filtrado e dificuldade de adequar a vazão do meio suplementar e, conseqüentemente, de manter a condição de operação a volume constante.



Figura 3.1. Sistema de microfiltração acoplado ao biorreator. 1-Solução 5M de NaOH 5M (para controle de pH); 2-Sistema de membranas Durapore (filtra o meio retirando os metabólitos); 3- Meio para alimentação (supre o biorreator com glicose e nutrientes); 4- Balança e recipiente com permeado; 5- Fermentador (biorreator).

As condições de cultivo adotadas para o experimento Exp₆ estão descritas na Tabela 3.3. A composição do meio de cultivo é a mesma descrita na Tabela 3.1, com exceção da concentração de glicose monohidratada que foi de 15 g/L. Foram empregados os mesmos métodos analíticos descritos na Subseção 3.1.1.

Tabela 3.3. Condições experimentais utilizadas no Exp₆.

Vazão de N ₂ (L/min)	0,5
pH	7,6-7,1
Agitação mecânica (rpm)	100
Temperatura (°C)	37
Volume (L)	2,0
Concentração inicial de glicose (g/L)	15
Inóculo (L)	0,5
Tempo de incubação (h)	13
Período de início de alimentação e retirada de meio (h)	6

A já mencionada complexidade operacional de um cultivo com retenção de células requer a utilização de recursos de modelagem, simulação, monitoramento, controle, enfim, a presença de um sistema supervisor, para que o potencial do sistema biorreator+microfiltração seja explorado plenamente. Dentre os vários aspectos que influenciam a operação do sistema, e que devem ser controlados e otimizados, destacam-se:

- a) concentração de glicose e demais nutrientes no meio suplementar;
- b) vazão de alimentação do meio suplementar;
- c) vazão de perfusão (retirada de meio de cultivo saturado e retenção de células por ultrafiltração);
- d) modo de operação do sistema de perfusão: contínuo/intermitente.

A definição destas condições de operação por experimentos exclusivamente “in vitro” é inviável em termos de tempo consumido e recursos gastos. No Capítulo 4, Seção 4.2 são apresentados e discutidos os resultados da utilização de modelagem matemática e simulação para otimização deste sistema.

3.2. Descrição dos Dados

Foram medidos 11 variáveis durante os cultivos, dentre os quais 7 foram *on-line* ou *quase on-line* (“quase on-line” refere-se às medidas de DO que podem ser realizadas em até 5 minutos) e 4 *off-line*. As variáveis medidas foram: tempo (h), agitação (rpm), temperatura (°C), pH, pressão (bar), massa de solução de NaOH, densidade óptica (DO), concentração celular (C_X), concentração de substrato (C_S), concentração de lactato (C_L) e concentração de polissacarídeo (C_{PS}). As variáveis agitação, temperatura, pH e pressão são condições de operação e foram mantidos constantes. C_X , C_S , C_L e C_{PS} são variáveis cujos valores podem demorar dias para serem determinados. Já a massa de solução de NaOH é uma medida que pode ser obtida *on-line*. A variável, DO, também pode ser considerada como atributo de medida *on-line*, visto que o seu valor pode ser determinado em até 3 minutos após a coleta de amostra. As demais variáveis (agitação, temperatura, pH e pressão), apesar de serem adquiridos em tempo real, não foram utilizados por não conterem informações relevantes para o presente trabalho.

3.3 Tratamento dos Dados para Utilização Pelas Técnicas de Inteligência Computacional – Interpolação e Alisamento

Todos os dados destinados à calibração dos modelos matemáticos e treinamento das redes neurais foram alisados de forma a amenizar os ruídos das medidas experimentais, bem como passaram por um processo de interpolação com o objetivo de aumentar a sua quantidade. O algoritmo utilizado para estes fins foi o Alisador Perfeito (“Perfect Smoother”) de Eilers (2003). É importante ressaltar que o alisamento foi bem suave de forma a preservar a maioria dos dados medidos.

A interpolação foi necessária, pois os algoritmos de ajuste requerem uma quantidade de dados relativamente grande em comparação ao número de parâmetros a serem ajustados. Não obstante, as redes neurais também requerem maiores quantidades de dados, visto que estes dados são subdivididos em conjuntos de treinamento, conjunto de teste e de validação. Como os dados fornecidos experimentalmente correspondem a cultivos de 11 horas com medidas de hora em hora (i.e. 12 instâncias), optou-se por gerar um número maior de instâncias por interpolação.

Em 1964 Savitzky e Golay introduziram a metodologia moderna de regressão, quando publicaram um algoritmo para alisamento e interpolação de dados com ruídos. O alisador

de Savitzky e Golay (SGS) ajusta um polinômio de baixa ordem em um intervalo de dados e, deixando o dado ajustado no meio do intervalo, desloca o intervalo (uma distância amostral) para a direita e repete este procedimento até que todo o intervalo desejado seja alisado. Porém, os valores ausentes e as extremidades dos domínios dos dados (o primeiro e último dado) tornam os programas SGS complicados. Em 1923, Whittaker publicou um algoritmo de alisamento denominado de “graduação das tabelas da vida”, que Eilers mesclou com o SGS, criando o “Alisador Perfeito” (Eilers, 2003).

O Alisador Perfeito funciona da seguinte forma: dado um conjunto de dados ruidosos y de comprimento m , aos quais se quer substituir por dados alisados z , balanceando dois objetivos conflitantes: 1- manter a fidelidade aos dados e 2-reduzir a rugosidade (ruídos) de z . Quanto maior for o alisamento de z , maior será a distância entre z e y , por outro lado, se o alisamento for muito pequeno, z coincidirá com y . A rugosidade (R) de z pode ser expressa por diferenças: $R = \sum_i (\Delta z_i)^2$. A falta de ajuste aos dados pode ser medida pelo usual mínimos quadrados: $S = \sum_i (y_i - z_i)^2$. A combinação balanceada dos dois objetivos é a soma $Q = S + \lambda R$, onde λ (o parâmetro de alisamento) é um número escolhido pelo usuário. A idéia dos mínimos quadrados penalizados é encontrar a série z que minimiza Q . Quanto maior for o valor de λ , maior será a influência de R em Q e z será menos rugoso (enquanto que o ajuste se distancia dos “ruídos”).

Em notação matricial $Q = |y - z|^2 + \lambda |D \cdot z|^2$, onde D é uma matriz tal que $D \cdot z = \Delta z$. Usando os resultados dos cálculos matriciais, encontramos para o vetor das derivadas parciais: $\partial Q / \partial z = -2(y - z) + 2\lambda D' D z$, igualando isso a zero, chega-se a um sistema linear de equações: $(I + \lambda \cdot D' D)z = y$, onde I é a matriz identidade.

Para simplificar a apresentação, o algoritmo foi introduzido com penalidade por diferenças de primeira ordem, o caso mais simples. Whittaker usou diferenças de terceira ordem, e em muitos casos, diferenças de segunda ordem funcionam bem: $\Delta^2 z_i = (z_i - 2z_{i-1} + z_{i-2})$ e $\Delta^3 z_i = z_i - 3z_{i-1} + 3z_{i-2} - z_{i-3}$. Para mais detalhes sobre o algoritmo, ver em Eilers (2003).

Em suma, o algoritmo requer a escolha de três parâmetros de entrada, λ , o parâmetro de alisamento, INT, o intervalo de tempo em minutos no qual os dados são interpolados e a ordem das diferenças. Neste trabalho, os parâmetros utilizados no Alisador Perfeito variaram de acordo com as características e necessidades de cada experimento (Exp₁ a Exp₆), além de variarem também para cada respectivo atributo, com exceção da ordem que foi mantida constante e igual a 2 para todas as interpolações.

Capítulo 4 – MODELAGEM DE CULTIVO DE *S. PNEUMONIAE* UTILIZANDO ALGORITMO DE BUSCA HEURÍSTICA E COMITÊ DE REDES NEURAIAS

Este capítulo apresenta três sistemas de modelagem de cultivo de *Streptococcus pneumoniae* que fazem uso do algoritmo SA para a determinação do melhor conjunto de valores de parâmetros. Com este objetivo o capítulo está organizado em cinco seções.

A Seção 4.1 descreve a modelagem não estruturada de cultivos em batelada. Apresenta os resultados obtidos tanto na escolha do melhor modelo cinético, quanto da configuração utilizada no algoritmo de ajuste. Estes resultados são importantes na medida em que permitem um melhor entendimento sobre o comportamento da bactéria, além de fornecer dados suficientes para que, a partir dos valores da função objetivo empregada pelo algoritmo de estimativa de parâmetros (neste caso o SA), o modelo cinético mais adequado dentre os testados possa ser selecionado.

A Seção 4.2 descreve a utilização da modelagem não estruturada do cultivo em biorreator operado em batelada alimentada acoplado a sistema de perfusão. Apresenta os resultados da modelagem bem como das simulações em diferentes condições de operação do sistema, visando à identificação daquelas que maximizam a produtividade de polissacarídeo capsular.

A Seção 4.3 descreve uma proposta de utilização de um comitê de redes neurais para a identificação das fases de crescimento, que é usado com a modelagem descrita na Seção 4.4. Apresenta os resultados das configurações mais promissoras encontradas para o treinamento das redes neurais e constituição do comitê.

A Seção 4.4 utiliza os resultados obtidos na Seção 4.1, emprega novamente o algoritmo SA e o modelo selecionado, para a estimativa dos parâmetros e a simulação do modelo não estruturado segregado por fases de crescimento da bactéria. Apresenta os resultados obtidos nesta modelagem segregada por fases de crescimento.

A Seção 4.5 utiliza o comitê de redes neurais treinadas na Seção 4.3 e os modelos ajustados a cada fase de crescimento da Seção 4.4 para construir uma ferramenta de inferência de estado (C_X , C_S , C_L e C_{PS}). Apresenta e discute os resultados da ferramenta de inferência.

4.1 Modelagem não Estruturada e Validação de Cultivos de *Streptococcus pneumoniae* em Batelada

Três modelos cinéticos não estruturados convencionais foram empregados para a descrição dos cultivos em batelada: o de Monod, o de Tessier e o de Moser. O conjunto de dados do Exp₂ foi utilizado para estimativa dos valores dos parâmetros por meio de ajuste de cada um dos modelos aos dados experimentais obtidos neste cultivo, após alisamento e interpolação. Os dados do Exp₁, Exp₃, Exp₄ e Exp₅ foram utilizados para validação, bem como suas versões estendidas, para posterior comparação com os resultados da Seção 4.4 e 4.5 deste trabalho.

As instâncias de dados utilizadas relativas aos experimentos descrevem valores do tempo de cultivo e das quatro variáveis: concentração celular (C_X); concentração de substrato (C_S); concentração de PS (C_{PS}) e concentração de lactato (C_L).

Os dados do Exp₁, Exp₂, Exp₃, Exp₄ e Exp₅ foram utilizados como entrada para o processo de interpolação que forneceu como saída, uma versão alisada e estendida dos conjuntos Exp₁^{est}, Exp₂^{est}, Exp₃^{est}, Exp₄^{est} e Exp₅^{est}, com 841 instâncias originadas a partir das 8 instâncias até o pulso de alimentação do Exp₂, 721 instâncias para o Exp₃, Exp₄ e Exp₅ que continham 7 instâncias até o pulso de alimentação e 602 instâncias para o Exp₁. Vale ressaltar que no Exp₁ o pulso de glicose foi feito na quinta hora, no Exp₂ na sétima hora e no Exp₃, Exp₄ e Exp₅ na sexta hora, pela razão exposta na seção 3.1.1. Como o modelo foi construído visando a descrição apenas de um cultivo convencional em batelada foram utilizados, nesta seção, apenas os dados que antecedem o pulso de glicose. O parâmetro de interpolação INT, usado pelo alisador foi ajustado de forma a interpolar 119 instâncias entre duas pré-existentes, com INT = 0,5 minuto. O valor do parâmetro de alisamento, λ , foi escolhido depois de várias tentativas. O conjunto ($\lambda_{CX} = 10^4$, $\lambda_{CL} = 10^4$, $\lambda_{CS} = 10^6$ e $\lambda_{CPS} = 10^5$) apresentou a melhor descrição dos dados experimentais pelos interpolados (pouco alisamento). O alisamento aplicado aos dados do Exp₂ pode ser visualizado na Figura 4.1.

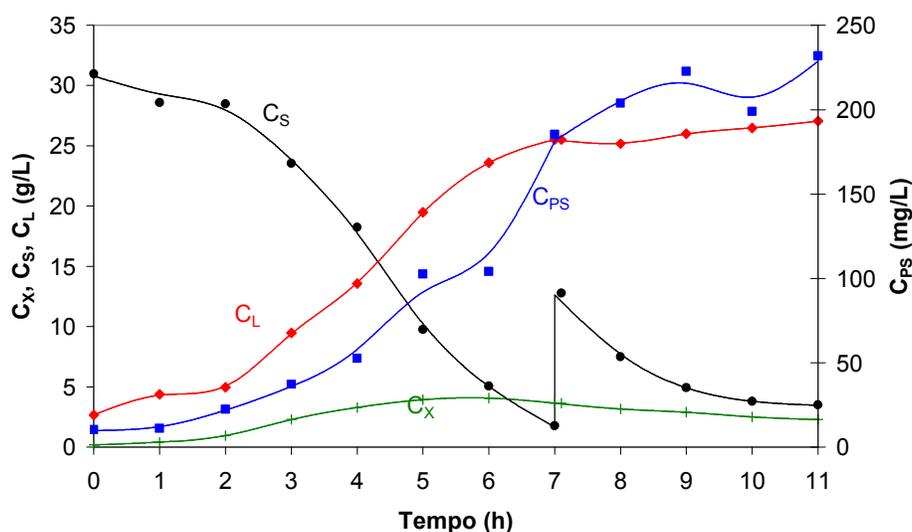


Figura 4.1. Dados e alisamento do Exp₂. C_X corresponde à concentração celular, C_S à concentração de glicose, C_L à concentração de lactato e C_{PS} à concentração de PS. Os pontos correspondem aos dados experimentais, enquanto que as linhas ao alisamento e interpolação. A descontinuidade de C_S em 7 horas corresponde ao pulso de glicose.

Após o tratamento, os dados estendidos foram então utilizados em dois procedimentos: (1) busca pelo melhor conjunto de valores de parâmetros associados a cada um dos modelos, Monod, Tessier e Moser, usando apenas o Exp₂^{est} para o cálculo dos valores dos parâmetros; (2) seleção do melhor modelo cinético dentre os três investigados por meio de simulações dos experimentos Exp₁, Exp₃, Exp₄ e Exp₅. As expressões cinéticas investigadas assim como as equações diferenciais resultantes do balanço material para as principais variáveis são representadas nas equações (4.1) a (4.9). O Quadro 4.1 contém a definição de variáveis e parâmetros das equações.

As equações de cada modelo cinético investigado para representar o crescimento celular são representadas nas equações (4.1) a (4.3). O modelo matemático que descreve o cultivo em batelada é composto pelas quatro equações de balanço (equações (4.6) a (4.9)) e por uma das equações de crescimento (equações (4.1) a (4.3)).

Quadro 4.1. Variáveis e parâmetros dos modelos.

Variáveis			
C_S (g/L)	C_X (g/L)	C_L (g/L)	C_{PS} (mg _{PS} /L)
Concentração de glicose	Concentração celular	Concentração de lactato	Concentração de polissacarídeo
Parâmetros			
Y_{XS} (g _X /g _S)	Fator de conversão de substrato a células		
Y_{LS} (g _L /g _S)	Fator de formação de lactato a partir do substrato		
α_{PS} (mg _{PS} /g _X)	Coeficiente pseudo-estequiométrico para formação de PS associada ao crescimento		
α_L (g _L /g _X)	Coeficiente pseudo-estequiométrico para formação de lactato associada ao crescimento		
β_{PS} (mg _{PS} /g _X .h)	Coeficiente para formação de PS não associada ao crescimento		
β_L (g _L /g _X .h)	Coeficiente para formação de lactato não associada ao crescimento		
m (g _S /g _X .h)	Coeficiente de manutenção celular		
K_D (1/h)	Constante de morte celular		
μ_M (1/h)	Velocidade específica máxima de crescimento celular		
K_S (g/L)	Constante de saturação		
K	Parâmetro do modelo de Tessier análogo ao K_S de Monod		
n	Parâmetro do modelo de Moser		

Monod:

$$\mu = \frac{\mu_M \cdot C_S}{K_S + C_S} \quad (4.1)$$

Tessier:

$$\mu = \mu_M \cdot \left(1 - e^{-K \cdot C_S}\right) \quad (4.2)$$

Moser:

$$\mu = \frac{\mu_M \cdot C_S^n}{K_S + C_S^n} \quad (4.3)$$

A expressão para a velocidade específica de formação de produtos é descrita pelas equações (4.4) e (4.5) baseadas no modelo de Luedeking e Piret, nos quais μ_L representa a velocidade específica de formação de lactato e μ_{PS} a velocidade específica de formação de PS (Nielsen e Villadsen, 1994).

$$\mu_L = \alpha_L \cdot \mu + \beta_L \quad (4.4)$$

$$\mu_{PS} = \alpha_{PS} \cdot \mu + \beta_{PS} \quad (4.5)$$

Equações de balanço:

$$\frac{dC_X}{dt} = (\mu - k_d) \cdot C_X \quad (4.6)$$

$$\frac{dC_S}{dt} = -\left(\frac{\mu}{Y_{XS}} + \frac{\mu_L}{Y_{LS}} + m\right) \cdot C_X \quad (4.7)$$

$$\frac{dC_{PL}}{dt} = \mu_L \cdot C_X \quad (4.8)$$

$$\frac{dC_{PS}}{dt} = \mu_{PS} \cdot C_X \quad (4.9)$$

O *Simulated Annealing*, cujo pseudocódigo canônico está descrito no Quadro 4.2, foi usado como algoritmo de busca heurística para encontrar os valores otimizados (ou quase-ótimos) dos parâmetros dos três modelos não estruturados de crescimento ajustados aos dados do $\text{Exp}_2^{\text{est}}$.

```

procedimento SA
início
  param_corrente = Escolhe_param(restrições_param)
  custo_param_corrente = custo(param_corrente)
  T = 10000
  fator_decrécimo = 0,999
Enquanto (T > 0,01) faça
  início
    param_novo = Escolhe_param_vizinho(param_corrente,restrições_param)
    custo_param_novo = custo(param_novo)
    Se (custo_param_novo < custo_param_corrente) então
      param_corrente = param_novo // muda o conjunto de valores dos parâmetros
      custo_param_corrente = custo_param_novo
    Senão
      num = random()
      Se (num <  $\exp\left(\frac{-\Delta E}{T}\right)$ ) então
        param_corrente = param_novo
        custo_param_corrente = custo_param_novo
    T = T × fator_decrécimo // decresce a temperatura
  fim
fim

```

Quadro 4.2. Pseudocódigo do SA. ΔE corresponde à diferença entre o valor da função objetivo na solução corrente e o valor da mesma função na solução candidata.

A função objetivo utilizada pelo SA - equações (4.10), (4.11) e (4.12) - representa a diferença entre os valores medidos experimentalmente e os valores calculados pelo modelo, sendo que C_i^{mod} refere-se à variável calculada pelo modelo enquanto que C_i^{exp} à variável medida

experimentalmente, com $i \in \{X, S, L \text{ e } PS\}$; t corresponde ao tempo e w_{Ci} corresponde à normalização.

$$f(C^{\text{mod}}, C^{\text{exp}}) = \sum_t \left[\left(\frac{C_X^{\text{mod}}(t) - C_X^{\text{exp}}(t)}{w_{C_X}} \right) + \left(\frac{C_S^{\text{mod}}(t) - C_S^{\text{exp}}(t)}{w_{C_S}} \right) + \left(\frac{C_L^{\text{mod}}(t) - C_L^{\text{exp}}(t)}{w_{C_L}} \right) + \left(\frac{C_{PS}^{\text{mod}}(t) - C_{PS}^{\text{exp}}(t)}{w_{C_{PS}}} \right) \right] \quad (4.10)$$

$$w_{Ci}(C^{\text{exp}}) = \sqrt{\frac{\sum_{t_0}^{t_f} (C_i^{\text{exp}}(t) - M)^2}{n^{\circ} \text{Instâncias} - 1}} \quad (4.11)$$

$$M = \frac{\sum_{t_0}^{t_f} C_i^{\text{exp}}}{n^{\circ} \text{Instâncias}} \quad (4.12)$$

O algoritmo começa pela escolha randômica de uma solução possível, isto é, um conjunto de valores para os parâmetros do modelo (p_1, p_2, \dots, p_9 como mostra a Tabela 4.1). Esta solução é nomeada *param_corrente*. Qualquer valor randômico é aceitável, desde que se enquadre no intervalo de restrição de cada parâmetro (limites inferiores e superiores apresentados na Tabela 4.1). Em seguida, o valor da função objetivo associado à solução *param_corrente* é calculado (representado no pseudocódigo pelo procedimento *custo(param-corrente)*). A cada passo do algoritmo, uma nova solução candidata, identificada como *param_novo* é escolhida randomicamente na vizinhança da solução corrente, respeitando os limites inferiores e superiores de cada parâmetro e, além disso, a seguinte condição dos valores dos parâmetros deve ser satisfeita, para todos os parâmetros: $0 \leq \text{param_novo}[p_i] \leq \text{param_corrente}[p_i] \times 2$. O custo da nova solução é então calculado e ambas as soluções têm seus custos comparados; a que tiver menor custo se torna a solução corrente (ΔE corresponde à diferença entre o valor função objetivo com a solução corrente e o valor com a solução candidata). Entretanto, existe uma probabilidade do *param_novo*, mesmo tendo um custo maior, se tornar a solução corrente; essa possibilidade depende tanto do valor da variável T quanto de um número randômico.

Tabela 4.1. Restrições dos valores dos parâmetros.

Limites	Parâmetros dos Modelos										
	p ₁ K/K _s	p ₂ m	p ₃ K _D	p ₄ Y _{XS}	p ₅ α _{PS}	p ₆ μ _M	p ₇ α _L	p ₈ Y _{LS}	p ₉ β _{PS}	p ₁₀ n	p ₁₁ β _L
Inferior	0,001	0,0001	0,0001	0,01	1	0,4	0,01	0.01	1	0,0001	0,001
Superior	10	0,5	0,5	0,8	50	0,9	9	2	50	10	10

A aceitação de uma solução pior que a atual, é uma tentativa do SA de não cair em soluções que são mínimos locais, aumentando a cobertura do espaço de busca. Como pode ser observado no pseudocódigo, à medida que a “temperatura” T diminui, a probabilidade do algoritmo aceitar um conjunto de valores não muito bom diminui. Este fato garante a sua convergência. O valor inicial das variáveis T e *fator_decrécimo* devem ser determinados empiricamente uma vez que essas variáveis controlam o número de iterações que o algoritmo faz e, conseqüentemente, desempenham um papel importante na convergência. Para os experimentos descritos nesta seção, as variáveis T e *fator_decrécimo* foram inicializadas com 10.000 e 0,999 respectivamente.

A Figura 4.2 apresenta um diagrama que exhibe o processo de busca do melhor modelo não-estruturado para representar um conjunto de dados experimentais. O diagrama supõe a existência de K possíveis modelos $\{M(1), M(2), \dots, M(K)\}$ e de um conjunto de dados experimentais (Exp). Para cada um dos modelos o algoritmo SA é utilizado para a determinação do conjunto de valores de parâmetros do modelo em questão que permitem o seu melhor ajuste aos dados experimentais.

O procedimento descrito pelo diagrama tem como saída o “melhor” conjunto de valores de parâmetros relativo a cada um dos K modelos bem como o modelo, dentre os K disponíveis (já com seu “melhor” conjunto de valores de parâmetros) que melhor representa os dados experimentais.

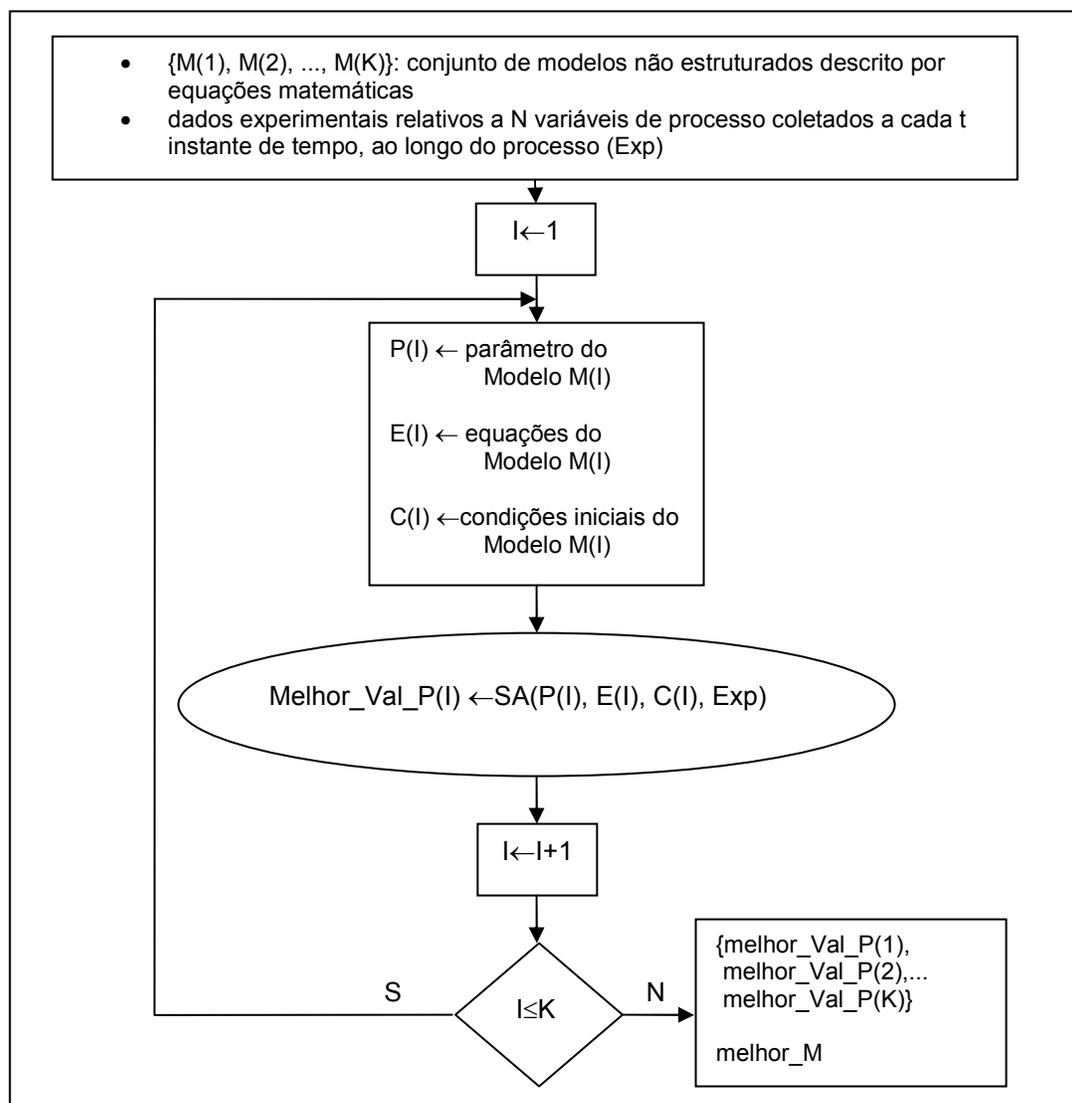


Figura 4.2. Esquema geral da escolha do melhor modelo não estruturado para representar os cultivos em batelada do *S. pneumoniae*.

Em seguida, são apresentados na Tabela 4.2, os resultados da calibração dos modelos de Monod, de Tessier e de Moser, aos dados do experimento $\text{Exp}_2^{\text{est}}$.

Tabela 4.2. Valores estimados pelo SA dos parâmetros dos modelos de Monod, Tessier e Moser.

	Parâmetros dos Modelos										
	p_1 K K_s	p_2 m	p_3 K_D	p_4 Y_{XS}	p_5 α_{PS}	p_6 μ_M	p_7 α_L	p_8 Y_{LS}	p_9 β_{PS}	p_{10} n	p_{11} β_L
Monod	0,2081	0,0002	0,0002	0,4263	16,9009	0,5997	5,3106	1,0533	3,8124	-	0,0761
Tessier	0,0551	0,0001	0,0008	0,2873	1,8722	0,8875	5,2018	1,5924	10,3542	-	0,0164
Moser	0,0536	0,0016	0,0201	0,0857	4,4053	0,5942	0,0109	0,0115	14,9034	1,6694	2,9624

Analisando-se os valores dos parâmetros apresentados na Tabela 4.2, especificamente os parâmetros β_{PS} e α_{PS} , observa-se que a formação de produto, de acordo com o modelo de Luedeking-Piret, pode ser equivocadamente interpretada como do tipo parcialmente associado ao crescimento para *S. pneumoniae*. No entanto, o alto valor de β_{PS} é, na realidade, decorrente da intensificação na liberação do polissacarídeo, que originalmente integra a parede celular, à medida que as células perdem a viabilidade ao final da fase exponencial.

De acordo com os valores das respectivas funções objetivo (Tabela 4.3), o modelo de Tessier se mostrou mais adequado para representar o perfil das variáveis, C_X , C_S , C_L e C_{PS} no cultivo Exp_2^{est} . A Figura 4.3 mostra os resultados obtidos na simulação do modelo de Tessier.

Tabela 4.3. Valores das funções objetivo dos modelos testados.

	Monod	Tessier	Moser
Valor da Função Objetivo	404,34	201,56	958,07

Na Figura 4.3.a., observa-se um bom ajuste do modelo aos dados experimentais representando a variação da concentração de polissacarídeo com o tempo ao longo de todo o cultivo. Por outro lado, o modelo representa bem apenas a variação das concentrações de glicose e lactato nas 3 primeiras horas de cultivo. A partir deste tempo, os valores simulados para a concentração de glicose são sempre inferiores aos valores experimentais desta variável. Da mesma forma, os valores simulados para a concentração de lactato são sempre superiores aos respectivos valores experimentais. Esta falta de ajuste pode ser explicada pelo fato do modelo não considerar outros nutrientes presentes no meio como fontes de carbono. Conforme descrito na Subseção 3.1.1, o meio de cultivo é composto por extrato de levedura, hidrolisado de caseína e aminoácidos que, com certeza, são também consumidos e utilizados pela célula para a formação de novas células. Como o modelo considera apenas a glicose como fonte de carbono e de energia, é natural que o consumo de glicose previsto pelo modelo seja maior do que o real, levando a menores valores simulados de concentração de glicose no meio. Como a produção de lactato está associada ao crescimento e, portanto, ao consumo de glicose que sustenta o mesmo, os valores simulados para a concentração de lactato se tornam maiores do que os valores reais. Os valores previstos pelo modelo para a concentração celular, representados na Figura 4.3.b, também não conseguem acompanhar a trajetória de aumento na concentração celular experimental na fase exponencial de crescimento. Da mesma forma, considerando apenas a glicose como fonte de carbono e energia, o modelo falha ao descrever a intensa formação de biomassa que ocorre na

fase exponencial e que, provavelmente, é resultado do consumo de outros nutrientes, além da glicose.

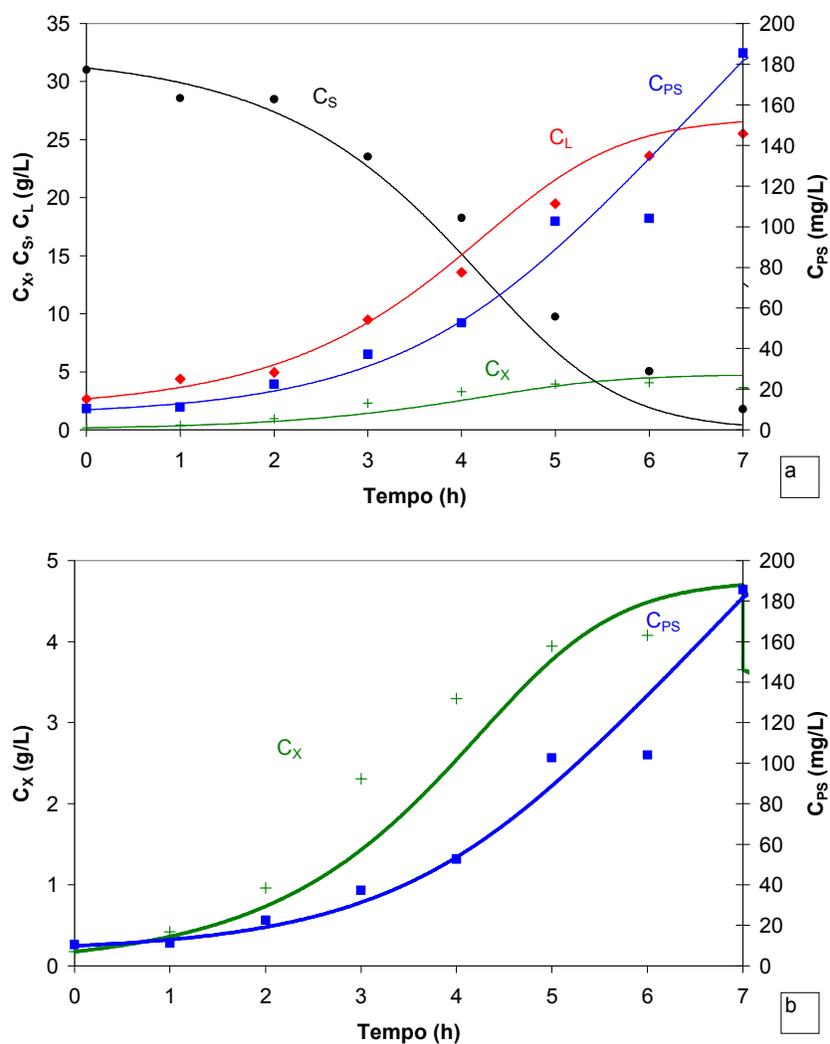


Figura 4.3. Ajuste do modelo de Tessier ao cultivo Exp_2^{est} . Os pontos/símbolos representam os valores experimentais, enquanto que as linhas, os valores preditos pelo modelo. C_X representa a concentração celular, C_S a concentração de glicose, C_L a concentração de lactato e C_{PS} a concentração de PS. (a) representa toda simulação até a sétima hora; (b) representa apenas C_X e C_{PS} .

Para efeito de determinação do nível de significância associado aos valores estimados para os parâmetros assim como da matriz de correlação entre eles, realizou-se novo ajuste do modelo de Tessier aos dados do cultivo Exp_2 . Porém, este ajuste foi realizado sobre os dados experimentais reais e com o SA tendo como valor inicial do conjunto de parâmetros os valores previamente encontrados no ajuste aos dados estendidos. O cálculo do intervalo de

confiança e da matriz de correlação foram feitos em notação matricial, de acordo com as equações (4.13) a (4.17).

$$IC(i) = \text{tst} \cdot \sqrt{\text{covar}(i,i)} \quad (4.13)$$

$$\text{covariância} = \frac{\text{hessiana}^{-1} \cdot \text{FO}}{n - m} \quad (4.14)$$

$$\text{hessiana} = J' \cdot J \quad (4.15)$$

$$J(i,j) = \frac{x(p(i,j) + 10^{-8}) - x(p(i,j) - 10^{-8})}{2 \cdot 10^{-8}} \quad (4.16)$$

$$\text{correlação}(i,j) = \frac{\text{covariância}(i,j)}{\sqrt{\text{covariância}(j,j) \cdot \text{covariância}(i,i)}} \quad (4.17)$$

onde i vai de 1 até m (m representa o número de parâmetros), j vai de 1 até n (n representa o número de dados experimentais), FO representa a função objetivo, IC o intervalo de confiança, x as variáveis calculadas pelo modelo (C_X , C_S , C_L e C_{PS}) e tst representa os valores fornecidos pela distribuição de probabilidade t de Student, que é função do número de graus de liberdade ($n-m$), por isso a necessidade de se fazer o ajuste aos dados experimentais sem interpolação. Se n for muito grande em relação a m , então a tst se aproxima da distribuição normal. Mais detalhes são fornecidos em Nelles (2001) e Rodriguez-Fernandez *et al.* (2006).

Os valores do intervalo de confiança dos respectivos parâmetros do modelo de Tessier ajustado aos dados do Exp_2 estão na Tabela 4.4. A Tabela 4.5 apresenta a matriz de correlação dos respectivos parâmetros, sendo que valores próximos a um indicam parâmetros altamente correlacionados e valores negativos indicam que os parâmetros são inversamente correlacionados.

Tabela 4.4. Intervalo de confiança dos parâmetros ajustados ao Exp_2 .

	Valor do Parâmetro	Intervalo de confiança
K	0,0551	$\pm 0,0176$
m	0,0001	$\pm 6067,5312$
K_D	0,0008	$\pm 0,0450$
Y_{XS}	0,2873	$\pm 158971,45$
α_{PS}	1,8722	$\pm 1,4922$
μ_M	0,8875	$\pm 0,1562$
α_L	5,2018	$\pm 0,6982$
β_{PS}	10,3542	$\pm 0,8443$
β_L	0,0164	$\pm 0,1870$
Y_{LS}	1,5924	$\pm 938585,66$

Tabela 4.5. Matriz de correlação dos parâmetros do modelo de Tessier ajustado aos dados do Exp₂.

	K	m	K _D	Y _{XS}	α _{PS}	μ _M	α _L	β _{PS}	β _L	Y _{LS}
K	1,00	-0,03	0,05	0,03	0,55	-0,90	0,50	0,02	-0,60	-0,03
m	-0,03	1,00	0,41	-1,00	-0,22	0,23	-0,41	-0,07	-0,13	1,00
K _D	0,05	0,41	1,00	-0,41	-0,68	0,37	-0,56	0,43	-0,21	0,41
Y _{XS}	0,03	-1,00	-0,41	1,00	0,22	-0,23	0,41	0,07	0,13	-1,00
α _{PS}	0,55	-0,22	-0,68	0,22	1,00	-0,77	0,65	-0,61	-0,37	-0,22
μ _M	-0,90	0,23	0,37	-0,23	-0,77	1,00	-0,74	0,07	0,45	0,23
α _L	0,50	-0,41	-0,56	0,41	0,65	-0,74	1,00	0,05	-0,50	-0,41
β _{PS}	0,02	-0,07	0,43	0,07	-0,61	0,07	0,05	1,00	0,16	-0,07
β _L	-0,60	-0,13	-0,21	0,13	-0,37	0,45	-0,50	0,16	1,00	-0,13
Y _{LS}	-0,03	1,00	0,41	-1,00	-0,22	0,23	-0,41	-0,07	-0,13	1,00

Pela análise da Tabela 4.4 pode-se inferir que os parâmetros “K”, “μ_M”, “α_L” e “β_{PS}” são significativos uma vez que apresentam um intervalo de confiança menor do que o valor do parâmetro. Já os demais parâmetros apresentam baixo nível de significância. Os parâmetros “m” e “K_D” praticamente assumiram o valor do limite inferior estipulado (Tabela 4.1). No caso destes parâmetros, é importante ressaltar que a quantidade de dados disponível na região de crescimento aonde estes parâmetros são significativos era muito restrita, o que justifica a incerteza na estimativa dos mesmos. O parâmetro “β_L” é pouco significativo, apresentando valor próximo ao limite inferior estipulado (Tabela 4.1). Um valor baixo para este parâmetro já era esperado, visto que para cultivos anaeróbios, a formação de lactato é necessariamente associada ao crescimento, por ser a única alternativa da célula para geração de energia.

O baixo nível de confiança associado aos parâmetros “Y_{XS}” e o “Y_{LS}” pode ser explicado pelo fato dos mesmos serem altamente correlacionados, como mostra a Tabela 4.5. Ainda de acordo com a Tabela 4.5, também são altamente correlacionados os parâmetros “m” e “Y_{XS}”, “K_D” e “α_{PS}”, “β_L” e “K”, “μ_M” e “K” e “β_{PS}” e “α_{PS}”.

O algoritmo SA implementado se mostrou bem eficiente. Foram realizadas 13.810 iterações em aproximadamente 15 minutos (em um computador com processador Pentium 4 com 1GB de memória RAM). Tal número de iterações foi suficiente para garantir a convergência do SA, conforme ilustrado na Figura 4.4.

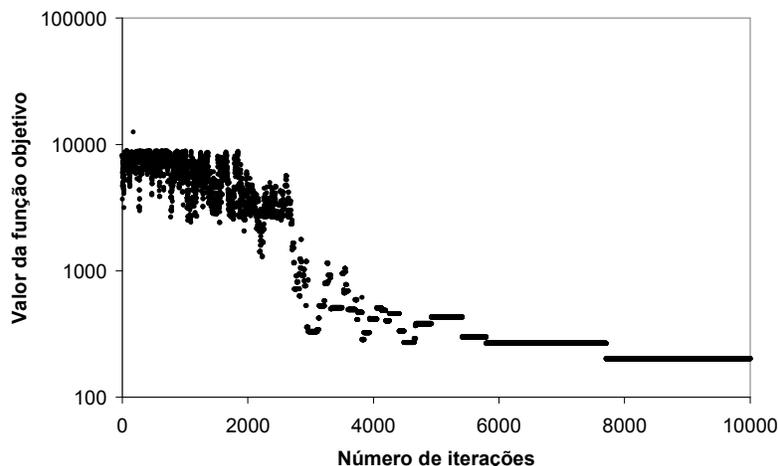


Figura 4.4. Convergência da função objetivo; Modelo de Tessier calibrado pelo SA utilizando os dados $\text{Exp}_2^{\text{est}}$. O valor da função objetivo está em escala logarítmica. O eixo do Número de iterações mostra os valores até 10.000 pois a partir disso não houve modificação no valor da função objetivo.

O modelo matemático proposto para descrever o crescimento de *S. pneumoniae* em batelada, equações (4.4) a (4.9), juntamente com a expressão de Tessier (4.2) para a cinética de crescimento e os valores dos parâmetros estimados a partir dos dados do cultivo Exp_2 , foi validado utilizando-se os dados experimentais brutos e estendidos dos cultivos Exp_1 (até a quinta hora), Exp_3 , Exp_4 e Exp_5 (até a sexta hora).

A Figura 4.5 e a Figura 4.6 apresentam os resultados da validação do modelo proposto, permitindo a comparação dos valores simulados com os experimentais de C_X , C_S e C_{PS} para o experimento Exp_1 e C_X , C_S , C_L e C_{PS} para os experimentos Exp_3 , Exp_4 e Exp_5 . Observa-se na Figura 4.5 (a) que os valores calculados pelo modelo, C_{PS}^{mod} , C_X^{mod} , C_S^{mod} e C_L^{mod} acompanham muito bem os dados experimentais, C_{PS}^{exp} , C_X^{exp} , C_S^{exp} e C_L^{exp} . Observa-se ainda na Figura 4.5 (a e b) e na Figura 4.6 (a e b) que foi necessário alterar as condições iniciais da simulação para obter um bom ajuste. É importante notar que a determinação de valores confiáveis para a concentração celular e para a concentração de PS no início do cultivo pela metodologia descrita na Subseção 3.1.1 é particularmente difícil, já que a concentração de células e de PS fica próxima do limite de sensibilidade do método analítico. Por isso, a atribuição de valores diferentes dos valores experimentais às condições iniciais é justificável. A dificuldade de ajuste aos dados experimentais dos cultivos 4 e 5 também está relacionada à mudança da cepa. O experimento 2, utilizado para calibração, foi realizado com a cepa liofilizada I, enquanto os experimentos 4 e 5 foram realizados com a cepa liofilizada II.

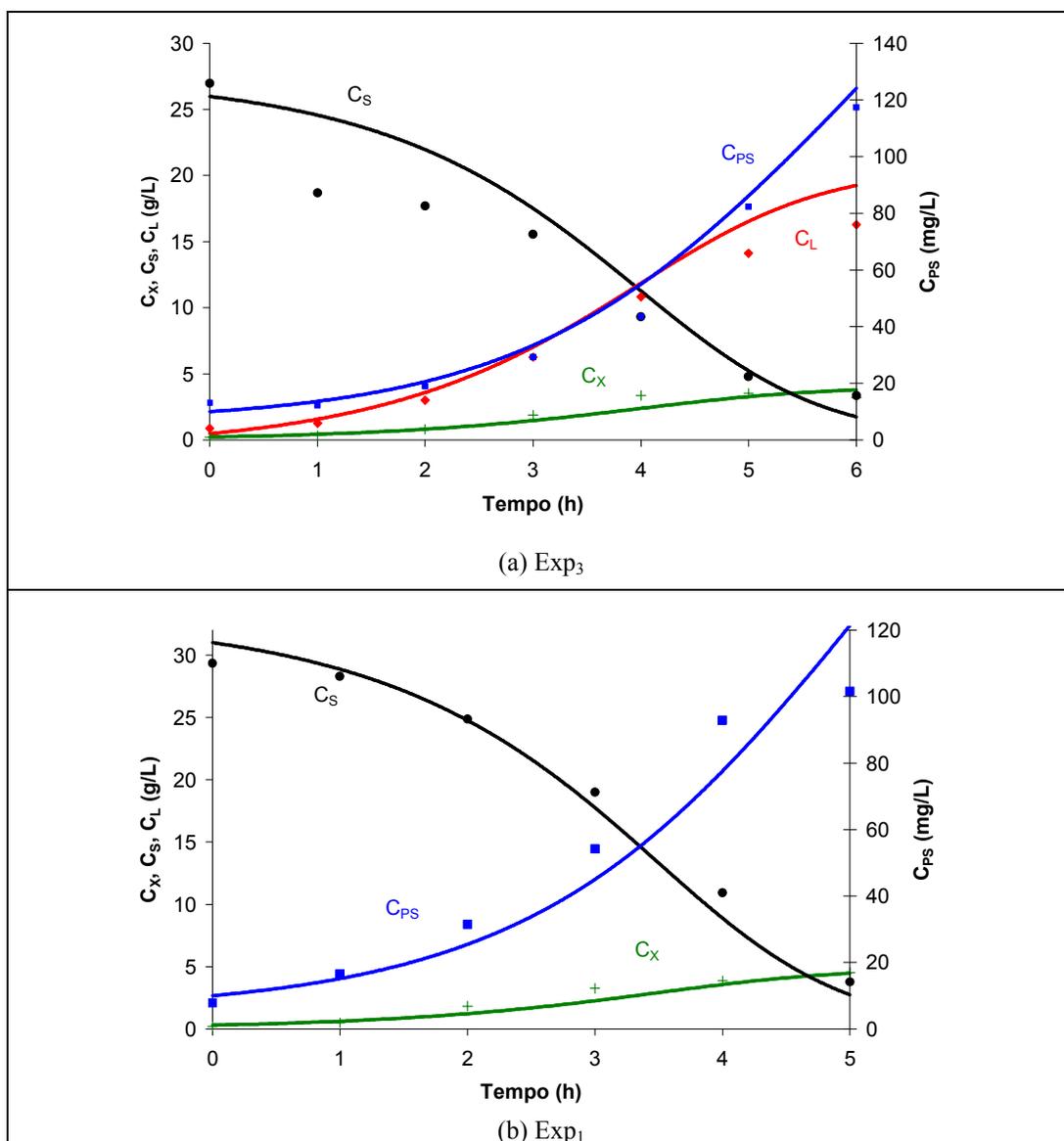


Figura 4.5. Validação do modelo proposto para o cultivo de *S. pneumoniae* em batelada, juntamente com o modelo de Tessier e os parâmetros calibrados a partir dos dados do cultivo Exp₂. (a) Exp₃ (b) Exp₁. Os símbolos representam os valores experimentais medidos, enquanto que as linhas, os valores preditos pelo modelo. C_X representa a concentração celular, C_S representa a concentração de glicose, C_L representa a concentração de lactato e C_{PS} a concentração de PS.

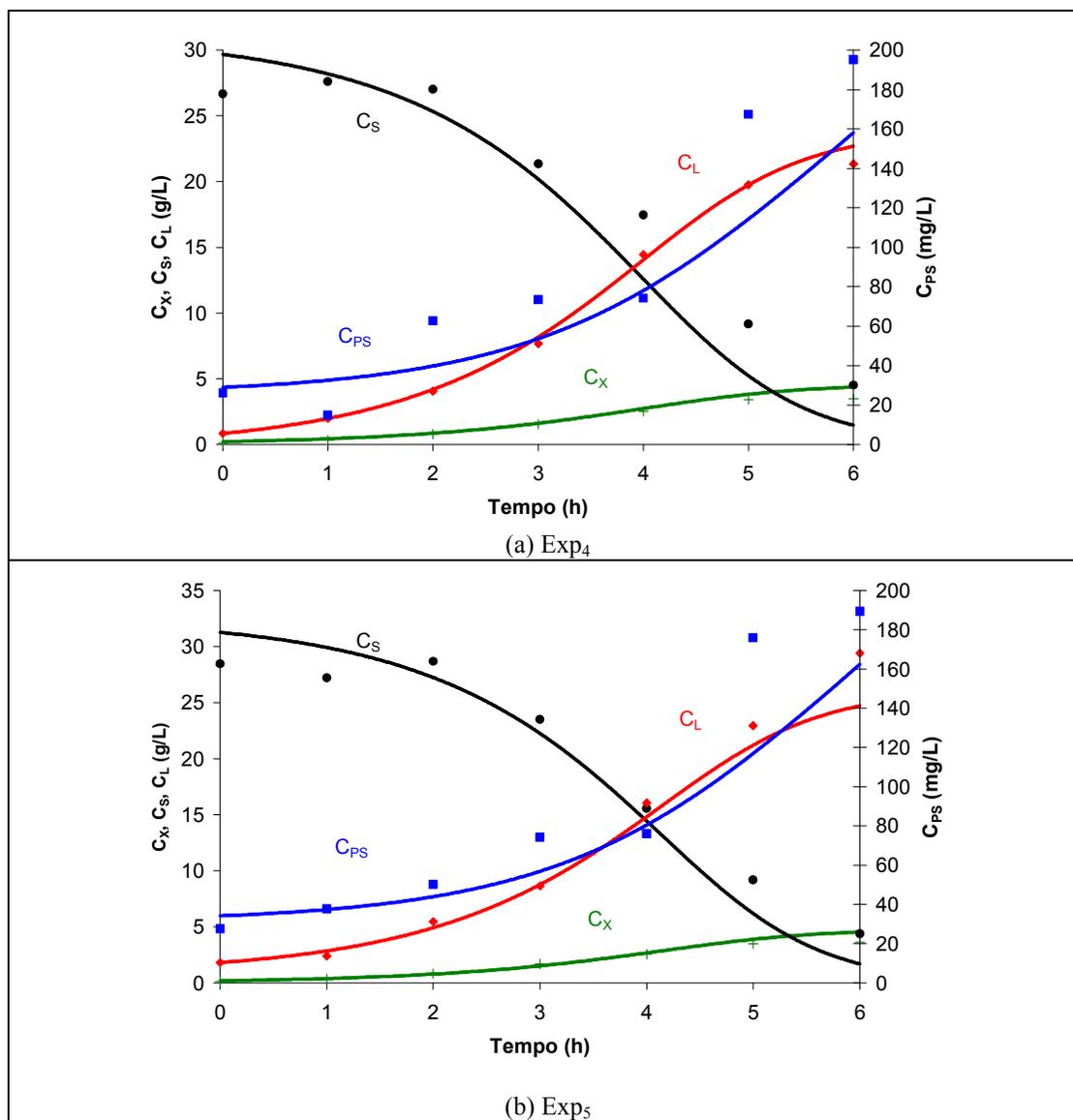


Figura 4.6. Validação do modelo proposto para o cultivo de *S. pneumoniae* em batelada, juntamente com o modelo de Tessier e os parâmetros calibrados a partir dos dados do cultivo Exp₂. (a) Exp₄ (b) Exp₅. Os símbolos representam os valores experimentais medidos, enquanto que as linhas, os valores preditos pelo modelo. C_X representa a concentração celular, C_S representa a concentração de glicose, C_L representa a concentração de lactato e C_{PS} a concentração de PS.

A comparação dos valores da função objetivo obtidos no ajuste do modelo aos conjuntos de dados experimentais de diferentes cultivos é útil na avaliação da capacidade descritiva do modelo. Os valores da função objetivo para os cultivos Exp₁, Exp₃, Exp₄ e Exp₅ se encontram na Tabela 4.6. Para comparação, a Tabela 4.6 também inclui o valor da função objetivo para o cultivo Exp₂ (até a sexta hora), utilizado para a calibração do modelo.

Tabela 4.6. Valores da função objetivo para o ajuste do modelo aos dados experimentais brutos e estendidos dos cultivos Exp₁, Exp₂, Exp₃, Exp₄ e Exp₅.

Experimento	Função Objetivo	
	Dados Brutos	Dados Estendidos
Exp ₁ (sem lactato)	1,25	145,26
Exp ₂	1,99	186,39
Exp ₃	2,55	296,34
Exp ₄	2,52	279,37
Exp ₅	2,46	226,19

Observa-se que os valores da função objetivo (dados brutos) para os cultivos Exp₃, Exp₄ e Exp₅ reproduzidos na Tabela 4.6 são cerca de 25% maiores que o valor da função objetivo do cultivo Exp₂. Já a diferença entre os valores da função objetivo (dados estendidos) de Exp₂ e dos cultivos Exp₃, Exp₄ e Exp₅ chega a 50% ou mais. Deve-se notar, no entanto, que a qualidade dos dados experimentais (brutos ou estendidos) tem grande influência no valor da função objetivo. Assim, analisando os dados apresentados nas Figuras 4.7(a,b), observa-se que a qualidade das medidas experimentais da concentração de PS para os cultivos Exp₄ e Exp₅ deixa muito a desejar quando comparada à das medidas da mesma variável no cultivo Exp₂ (Figura 4.4). Com certeza, o erro experimental nas medidas da concentração de PS contribuiu para os valores mais altos da função objetivo no ajuste aos dados dos cultivos Exp₄ e Exp₅. Da mesma forma, o elevado erro experimental na determinação das concentrações de glicose no início do cultivo Exp₃ (Figura 4.6 a) foi determinante para que o valor da respectiva função objetivo fosse superior aos demais. Quanto ao menor valor da função objetivo para o ajuste aos dados do cultivo Exp₁, o mesmo se deve ao cálculo da FO sem medidas da concentração de lactato, as quais não estavam disponíveis para este experimento.

Com base nos resultados apresentados na Tabela 4.6, na Figura 4.5 (a e b) e na Figura 4.6 (a e b), pode-se afirmar que o modelo foi capaz de descrever experimentos similares com boa precisão e que, portanto, a validação do mesmo foi bem sucedida. Na continuidade do trabalho, será explorada a abordagem de modelagem por fases de crescimento, sendo que o mesmo algoritmo aqui descrito para ajuste do modelo a cada fase de crescimento do microrganismo será utilizado.

4.2 Modelagem não Estruturada de Cultivo em Biorreator Operado em Batelada Alimentada Acoplado a Sistema de Perfusão

O trabalho de modelagem e estimativa de parâmetros teve início com o tratamento dos dados experimentais. O conjunto de dados obtidos em cultivo realizado em perfusão, Exp₆, com 12 instâncias foi dividido em três partes: fase lag (0-2 h com 3 instâncias), período de batelada (2-6 h com 5 instâncias) e período de perfusão (6-11 h com 6 instâncias). Os dados da fase lag foram descartados, pois não é explicada pelo modelo, enquanto que os dados do período de batelada e de perfusão foram tratados pela metodologia de alisamento e a interpolação de dados, com a finalidade principal de aumentar o número de dados para a estimativa dos parâmetros ajustáveis do modelo. Para tanto foi empregado o Alisador Perfeito (Eilers, 2003).

As 10 instâncias de dados experimentais correspondentes às fases selecionadas foram suavizadas e interpoladas, a uma frequência de inserção de 120 novas instâncias gerando o conjunto renomeado de Exp₆^{est}, com 1081 instâncias de dados. A escolha dos valores do parâmetro de suavização λ foi feita de acordo com as características dos respectivos dados (mais ou menos rugosidade) de forma a manter o perfil das medidas experimentais. Os valores escolhidos para λ são fornecidos na Tabela 4.7.

Tabela 4.7. Valores do parâmetro λ utilizados na suavização.

Variável	λ
C _x	20000
C _s	30000
C _L	15000
C _{ps}	20000

Os resultados do alisamento e interpolação podem ser observados na Figura 4.7.

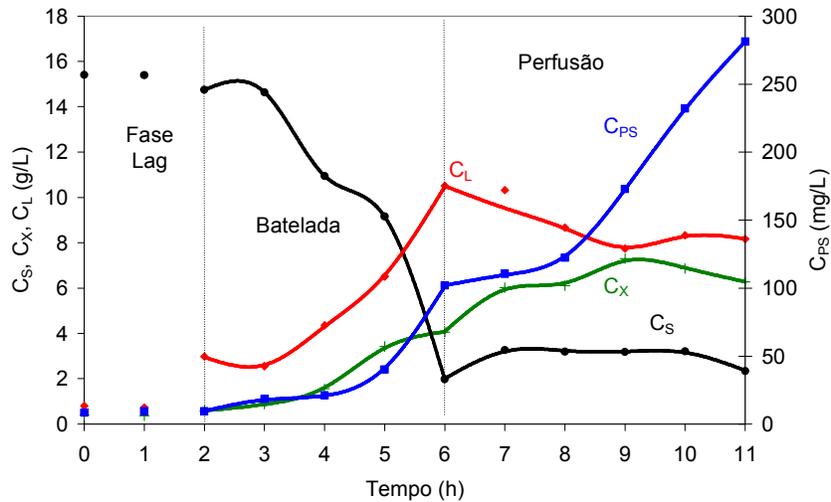


Figura 4.7. Evolução das concentrações de células, C_X (+), de substrato, C_S (●), de lactato, C_L (♦) e de polissacarídeo capsular, C_{PS} (□) ao longo do tempo para o cultivo de *S. pneumoniae* em biorreator acoplado a sistema de perfusão. Símbolos: dados experimentais; Linhas Contínuas: dados gerados por alisamento e interpolação.

Para a seleção do modelo não estruturado mais adequado para a descrição do processo, foram avaliados diferentes modelos representando a cinética do crescimento celular, tais como o modelo de Hoppe e Hansford, com inibição não competitiva pelo produto; o modelo de Levenspiel, com inibição por produto; e ainda com o modelo híbrido de Andrews e Levenspiel, com inibição por substrato e por produto (Shuler e Kargi, 1992). O modelo cinético utilizado na simulação foi o de Hoppe e Hansford (equação (4.18)), o qual apresentou melhor ajuste aos dados experimentais. Como os dados disponíveis foram obtidos em cultivo conduzido em baixa concentração de glicose, a inibição do crescimento pelo substrato não foi incluída no modelo proposto. A expressão cinética escolhida para representar o crescimento celular e as equações diferenciais presentes no modelo são descritas nas equações (4.18) a (4.24) baseadas no trabalho de Gonçalves (2002).

$$\mu = \frac{\mu_M}{\left(1 + \frac{K_S}{C_S}\right) \left(1 + \frac{C_L}{C_{lactar}}\right)} \quad (4.18)$$

$$\frac{dC_X}{dt} = (\mu - kd) \cdot C_X \quad (4.19)$$

$$\frac{dC_S}{dt} = (S_F - C_S) \cdot \frac{F}{V} - \left(\frac{\mu}{Y_{XS}} + \frac{\mu_L}{Y_{LS}} + m \right) \cdot C_X \quad (4.20)$$

$$\mu_L = \alpha_L \cdot \mu + \beta_L \quad (4.21)$$

$$\frac{dC_L}{dt} = \mu_L \cdot C_X - C_L \cdot \frac{F}{V} \quad (4.22)$$

$$\mu_{PS} = \alpha_{PS} \cdot \mu + \beta_{PS} \quad (4.23)$$

$$\frac{dC_{PS}}{dt} = \mu_{PS} \cdot C_X - C_{PS} \cdot \frac{F}{V} \quad (4.24)$$

nas quais F (L/h) representa a vazão de alimentação e de retirada de meio, S_F (g/L) a concentração de glicose no meio de alimentação, V (L) o volume do biorreator e C_{lactcr} (g/L) corresponde à constante de inibição pelo lactato. Os demais parâmetros já foram descritos anteriormente na Seção 4.1.

A estimativa de todos os parâmetros foi realizada simultaneamente ao processo de escolha do melhor modelo, também em duas etapas, operação em batelada e em perfusão, após exclusão dos dados pertencentes à fase de adaptação (fase lag). Foram estimados os onze parâmetros presentes no modelo, de acordo com a estratégia descrita a seguir.

Neste trabalho o modelo teve seus parâmetros ajustados adotando-se as técnicas de otimização supervisionada não linear pelo algoritmo de busca global *Simulated Annealing* (SA).

A solução ótima do modelo, ou seja, os valores para os parâmetros que levam ao melhor ajuste, é alcançada quando a função objetivo atinge seu valor mínimo. Novamente, a função objetivo utilizada foi o somatório dos mínimos quadrados da diferença entre os valores experimentais alisados e interpolados e os valores estimados pelo modelo, ponderados, conforme descrito anteriormente.

Os valores estimados pelo SA para os parâmetros nas duas etapas de operação estão representados na Tabela 4.8.

Tabela 4.8. Parâmetros ajustados via SA-LM na operação em batelada e em perfusão.

Etapa	K_S	m	K_D	Y_{XS}	α_{PS}	μ_M	α_L	C_{lactcr}	β_{PS}	β_L	Y_{LS}
Batelada	0,4	0,5	0	0,4	2	0,5	2,5	44,4	9,1	0	1,5
Perfusão	4,5	0,4	0	0,4	2	0,8	2,5	47,6	11,5	0	1,6

Os resultados do ajuste do modelo aos dados experimentais podem ser visualizados na Figura 4.8. Em geral o modelo descreve bem as 4 variáveis de estado, porém existem diferenças entre os valores simulados e os dados gerados por alisamento e interpolação. Esta falta de ajuste do modelo aos dados gerados, principalmente aos da glicose na fase de perfusão, está relacionada à dificuldade de manter a vazão de retirada do meio de cultivo na faixa de operação desejada devido ao gradativo entupimento das membranas de microfiltração. Com isso, os ajustes da vazão de alimentação do meio fresco, os quais eram feitos visualmente

procurando-se manter o nível de meio constante no biorreator, foram dificultados e o volume de meio no reator foi aumentando. Além disso, por uma falha no sistema de aquisição *on-line*, apenas um valor médio da vazão de retirada do meio pôde ser estipulado para a operação em perfusão, o qual foi empregado para simular o experimento. Como a dificuldade de manter o fluxo de permeado através da membrana foi se acentuando com o tempo de cultivo, não foi possível simular o experimento após nove horas. Apesar das dificuldades experimentais e da falta de informações mais precisas sobre fatores importantes, como a vazão de alimentação exata a cada instante, a capacidade descritiva do modelo para a produção de polissacarídeo, de lactato e de crescimento celular foi preservada.

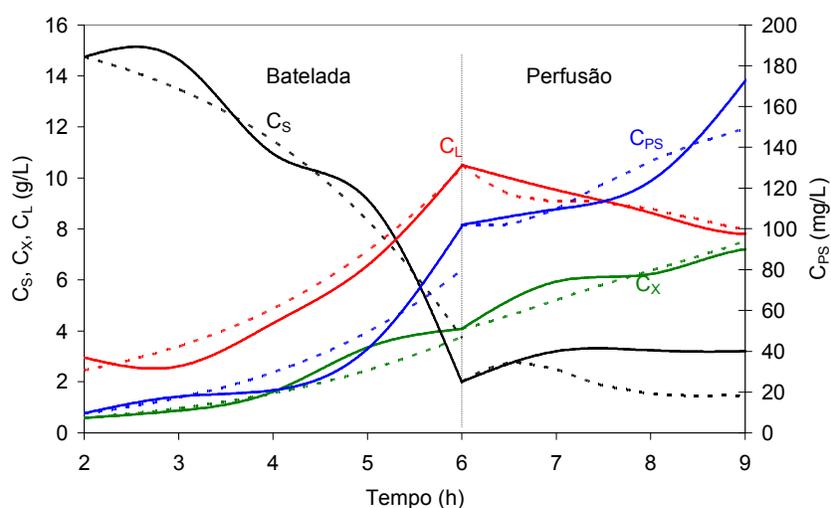


Figura 4.8. Comparação entre os resultados simulados e os experimentais obtidos no cultivo empregando biorreator acoplado às membranas de perfusão para retenção de células e drenagem de meio. Linhas tracejadas: dados simulados; Linhas contínuas: dados interpolados e alisados. C_X representa a concentração celular, C_S representa a concentração de glicose, C_L representa a concentração de lactato e C_{PS} a concentração de PS.

É interessante ainda comentar que o perfil simulado para a concentração de glicose na fase de perfusão pode ser atribuído à presença de vários nutrientes no meio de cultivo, que podem ter sido utilizados pela bactéria durante o crescimento. Esta possível utilização de outros nutrientes manteve os valores experimentais da concentração de glicose mais elevados do que os simulados, já que no modelo, a glicose é considerada a única fonte de carbono, e tem que ser consumida mais intensamente para manutenção do crescimento. Se o comportamento descrito for observado em experimentos similares a serem realizados futuramente, o modelo deverá ser modificado para representar de forma mais precisa o consumo de glicose.

Ainda analisando a Figura 4.8, as descontinuidades observadas nas transições entre a batelada e a perfusão estão relacionadas às condições iniciais adotadas para iniciar a simulação da fase de perfusão, as quais correspondem aos valores dos dados gerados por interpolação e alisamento ao final da batelada.

Com os valores dos parâmetros estimados e confirmada a capacidade descritiva do modelo proposto, foi iniciada a etapa de simulação das diferentes condições de operação visando identificar aquelas que maximizavam a produtividade em PS. O modelo ajustado foi simulado primeiro para o período de batelada, adotando-se uma concentração inicial de glicose de 10 g/L, menor que a concentração utilizada no experimento. O critério escolhido para término da batelada e início da operação contínua, com adição de meio fresco, retirada de meio saturado e retenção celular, foi o momento no qual a concentração residual de glicose no meio atingiu 2 g/L. Os valores finais da batelada para as variáveis, concentração de glicose, de biomassa, de polissacarídeo e de lactato foram fornecidos como condições iniciais à simulação do modelo aplicado à fase de perfusão. As condições iniciais para a simulação da fase batelada e da fase perfusão estão relacionadas na Tabela 4.9.

Tabela 4.9. Condições iniciais da simulação para a batelada e para o período de operação do sistema integrado.

Condições iniciais	Tempo (h)	C_X (g/L)	C_S (g/L)	C_L (g/L)	C_{PS} (mg/L)
Batelada	0	0,5	10	2,5	9
Perfusão	3,72	2,8	1,98	8,3	60

Para o estudo da influência da vazão de alimentação sobre a produção de polissacarídeo em condições de crescimento não limitado, escolheu-se o perfil exponencial de alimentação representado pela equação (4.25), em função dos parâmetros C_1 e C_2 , o qual foi implementado no algoritmo de simulação juntamente com o modelo descrito anteriormente. O perfil exponencial foi escolhido para prolongar o crescimento celular na fase exponencial.

$$F = C_2 \cdot e^{C_1 \cdot t} \quad (4.25)$$

A complexidade do modelo proposto inviabiliza a obtenção de uma solução analítica para a vazão de alimentação. Por isso, adotou-se novamente o SA com a finalidade de

buscar um perfil de alimentação exponencial ótimo e uma concentração de substrato ótima no meio de alimentação.

Na simulação do cultivo em biorreator acoplado ao sistema de microfiltração, foram consideradas as seguintes restrições operacionais:

- a) se a vazão de alimentação for muito elevada, o tempo de vida da membrana de microfiltração seria muito curto devido ao entupimento e ao desgaste. Desta forma, o valor máximo para F foi estabelecido em 0,8 L/h. Já o mínimo para F , o qual é definido pela velocidade de rotação mínima da bomba peristáltica, foi estabelecido em 0,01 L/h;
- b) pelo conhecimento do tempo de vida útil da membrana, o tempo de simulação (correspondente ao tempo de operação em perfusão) foi limitado a 12 horas;
- c) por razões relacionadas à esterilização do meio por microfiltração, a concentração máxima de glicose no meio de alimentação foi limitada a 100 g/L. Já a concentração mínima foi definida em 20 g/L;
- d) como, durante o experimento de perfusão, a vazão de meio de alimentação foi, a princípio, manipulada de forma que fosse igual à de retirada do meio de cultivo, a condição de operação a volume constante foi adotada tanto na modelagem quanto na simulação do processo.

Conforme descrito anteriormente, a busca pelas melhores condições para operação do processo utilizando o sistema de microfiltração para retirada de meio de cultivo foi realizada por meio do algoritmo SA já implementado. Desta forma, os parâmetros ajustados pelo SA foram C_1 , C_2 e S_F (C_1 e C_2 são constantes do perfil de alimentação exponencial e S_F corresponde à concentração de glicose na alimentação). Neste caso a função objetivo (FO) a ser minimizada passou a ser o inverso da soma da concentração máxima de polissacarídeo no meio com o inverso da concentração máxima de glicose acumulada no biorreator, equação (4.26). A produtividade foi obtida dividindo-se a concentração máxima de polissacarídeo capsular (PS_{MAX}) pelo tempo de cultivo total (t_{TOTAL_PSmax}) necessário para atingi-la.

$$FO = \frac{1}{\max(C_{PS}) + \frac{1}{\max(C_S)}} \quad (4.26)$$

O algoritmo de ajuste (SA) iniciou com T igual a 5 até atingir $T = 0,0000001$ para convergir a FO. Estes valores foram encontrados empiricamente. O resultado da otimização pode ser visualizado na Figura 4.9 com $C_1 = 0,003 \text{ h}^{-1}$; $C_2 = 0,8 \text{ L/h}$ e $S_F = 99,5 \text{ g/L}$, para as simulações de um reator com 5L de volume.

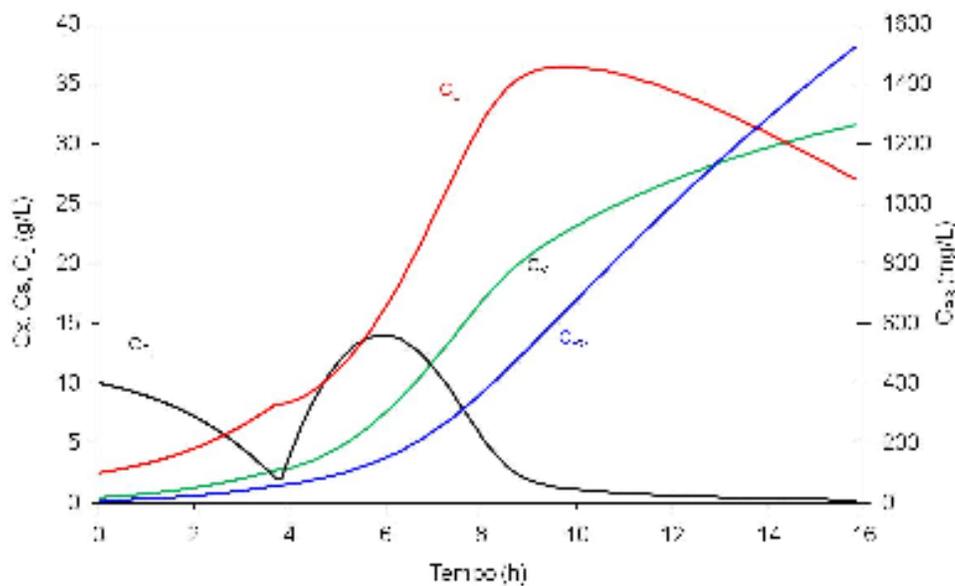


Figura 4.9. Simulação do experimento. Evolução das concentrações de células, C_x (linha verde), de substrato, C_s (linha preta), de lactato, C_L (linha vermelha) e de polissacarídeo capsular, C_{PS} (linha azul) ao longo do tempo para o cultivo de *S. pneumoniae* em biorreator acoplado a sistema de perfusão. Dados gerados por simulação (reator de 5L).

Ao contrário do esperado, a concentração de glicose residual não atingiu valores mais elevados devido ao aumento da concentração de glicose no meio de alimentação. Além disso, como o valor estimado para o parâmetro concentração crítica de lactato foi relativamente alto ($\sim 47 \text{ g/L}$), a modulação do crescimento celular pelo acúmulo de lactato só ocorre quando a concentração do metabólito se aproxima de 30 g/L . Para melhorar o modelo, é necessário realização de cultivo em condições de fornecimento de glicose correspondentes às empregadas nas simulações, de forma que valores reais para as concentrações críticas de lactato e de glicose possam ser determinados.

Os resultados da simulação do perfil exponencial em reator de 5L representados na Figura 4.9, podem ser comparados com os valores dos dados experimentais obtidos no cultivo em biorreator acoplado à membrana de microfiltração ($V=2\text{L}$) Figura 4.7. Nota-se que, mesmo que não seja possível manter a operação de perfusão por 12 h, caso haja entupimento da membrana, a

concentração de polissacarídeo simulada após 5 h de perfusão, nas condições de operação definidas para a simulação, ultrapassa 550 mg/L, ou seja, duas vezes superior à obtida experimentalmente (~ 250 mg/L).

Os resultados se mostram promissores, dobrando a concentração de polissacarídeo capsular, porém são necessários experimentos para validar o modelo proposto, e principalmente, para aprimorá-lo de forma que a inibição pelo substrato seja incorporada e a inibição pelo lactato realmente estabelecida.

A incerteza envolvida em processos biológicos demanda que, além da otimização, ferramentas de monitoramento *on-line* dos cultivos estejam disponíveis. Na próxima seção uma ferramenta para reconhecimento de fases de cultivo é apresentada e discutida.

4.3 O Uso de um Comitê de Redes Neurais Para o Reconhecimento Das Fases de Crescimento

Esta seção inicialmente descreve como as instâncias de dados coletados em experimentos foram tratadas para viabilizar o aprendizado de redes neurais para a identificação da fase de crescimento durante o cultivo de *Streptococcus pneumoniae*. Em seguida, descreve o processo de treinamento e validação de um comitê de redes neurais encarregado do processo de identificação da fase de crescimento.

Para o treinamento e teste das redes neurais foram utilizados os dados alisados e interpolados do Exp₁, Exp₂, Exp₃, Exp₄ e Exp₅ descritos por seis variáveis: densidade ótica (DO), concentração celular (C_X), concentração de lactato (C_L), concentração de substrato (C_S), concentração de PS (C_{PS}) e massa de solução de NaOH (NaOH).

Via processo de interpolação, as 13 instâncias (12 referentes às medidas de hora em hora e 1 referente à medida feita logo após o pulso de alimentação) originais de cada experimento resultaram em 1441 instâncias para cada conjunto, renomeados de Exp₁^{est}, Exp₂^{est}, Exp₃^{est}, Exp₄^{est} e Exp₅^{est}. O parâmetro de interpolação INT, usado pelo alisador foi ajustado de forma a interpolar 119 instâncias entre duas pré-existentes, com INT=0,5 minuto. O valor do parâmetro de alisamento, λ , foi escolhido depois de várias tentativas, e o que melhor ajustou o alisamento aos dados foi: ($\lambda_{DO} = 10^5$, $\lambda_{CX} = 10^5$, $\lambda_{CL} = 10^4$, $\lambda_{CS} = 10^5$, $\lambda_{CPS} = 10^5$ e $\lambda_{NaOH} = 10^5$), passando os dados interpolados sobre a maioria dos pontos experimentais (pouco alisamento). O alisamento aos dados do Exp₂ pode ser visualizado na Figura 4.10.

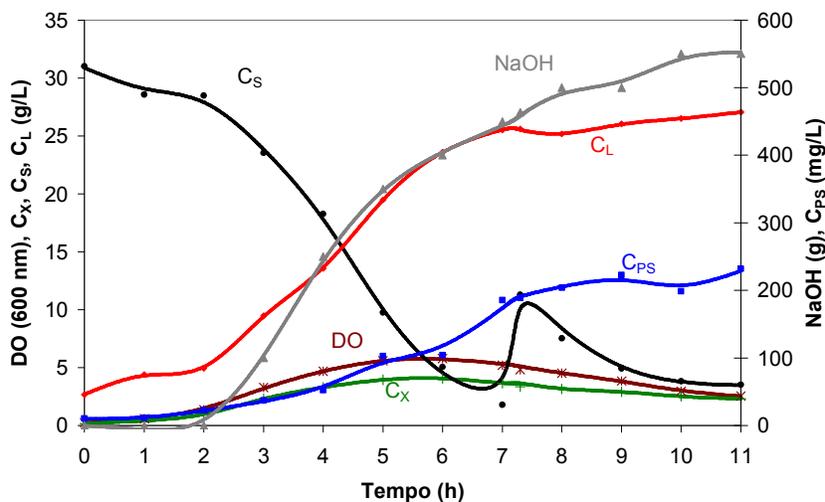


Figura 4.10. Interpolação dos dados do Exp_2 . DO corresponde à densidade óptica, C_x à concentração celular, C_s à concentração de glicose, C_L à concentração de lactato, C_{ps} à concentração de PS e NaOH à massa de solução de NaOH (g). Os pontos correspondem aos dados experimentais, enquanto que as linhas ao alisamento e interpolação.

Em seguida, cada uma das instâncias de dados do $\text{Exp}_1^{\text{est}}$, $\text{Exp}_2^{\text{est}}$, $\text{Exp}_3^{\text{est}}$, $\text{Exp}_4^{\text{est}}$ e $\text{Exp}_5^{\text{est}}$ foi classificada por um especialista humano da área, em uma das fases de crescimento. A classificação foi baseada na variação da densidade óptica no tempo ($d\text{DO}/dt$) (a densidade óptica é proporcional à concentração celular). As mudanças de fases foram determinadas pelos pontos de inflexão da curva $d\text{DO}/dt$ versus t , representado na Figura 4.11 (para o $\text{Exp}_2^{\text{est}}$). Neste caso, a $d\text{DO}/dt$ foi aproximada da variação $\Delta\text{DO}/\Delta t$; esta aproximação se justifica com o grande número de dados interpolados.

Na Figura 4.11 os algarismos romanos identificam as quatro fases como I, II, III e IV, que correspondem às fases, lag e aceleração, exponencial, desaceleração, estacionária e morte, respectivamente. Como algumas fases são de difícil distinção, optou-se por mesclar as fases lag e aceleração como fase I, e as fases estacionária e morte, como fase IV. A identificação da fase do cultivo associada a cada uma das instâncias de dados de $\text{Exp}_1^{\text{est}}$, $\text{Exp}_2^{\text{est}}$, $\text{Exp}_3^{\text{est}}$, $\text{Exp}_4^{\text{est}}$ e $\text{Exp}_5^{\text{est}}$ foi realizada com base no respectivo valor da $d\text{DO}/dt$ da instância avaliada por um especialista humano por critério visual (Tabela 4.10).

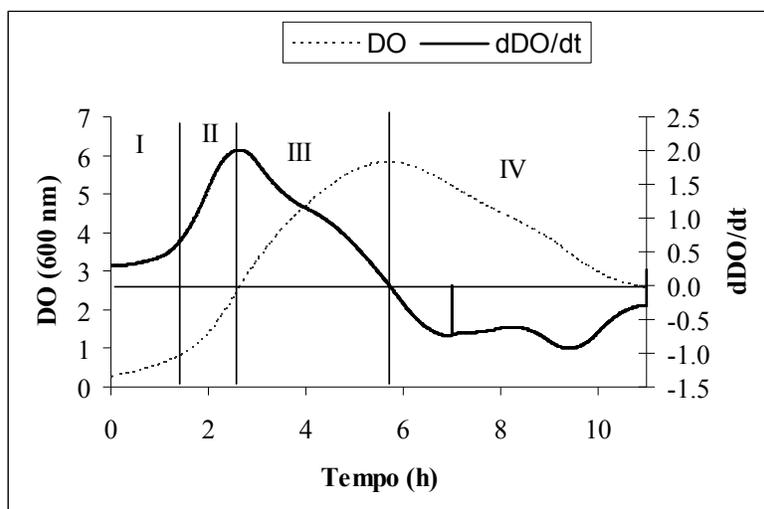


Figura 4.11. Divisão das fases de crescimento em função do perfil da dDO/dt (aproximação da $\Delta DO/\Delta t$). As fases indicadas correspondem à: I (fase lag e aceleração); II (fase exponencial); III (fase de desaceleração) e IV (fase estacionária e morte). DO é a densidade óptica do meio, e t o tempo. Dados estendidos do cultivo Exp_2 .

Tabela 4.10. Intervalo das fases de crescimento quanto à variação da DO com o tempo das fermentações Exp_1^{est} , Exp_2^{est} , Exp_3^{est} , Exp_4^{est} e Exp_5^{est} .

Fase	Exp_1^{est} t (h)	Exp_2^{est} t (h)	Exp_3^{est} t (h)	Exp_4^{est} t (h)	Exp_5^{est} t (h)
Lag e Aceleração	0 a 0,65	0 a 1,32	0 a 0,85	0 a 0,73	0 a 1,31
Exponencial	0,65 a 3,34	1,32 a 2,64	0,85 a 3,43	0,73 a 4	1,31 a 4,31
Desaceleração	3,34 a 4,3	2,64 a 5,72	3,43 a 5,75	4 a 5,73	4,31 a 5,87
Estacionária e Morte	4,3 a 11	5,72 a 11	5,75 a 11	5,73 a 11	5,87 a 11

Para uma representação de dados adequada com vistas ao treinamento de uma rede neural usando o backpropagation, as quatro possíveis fases de crescimento foram codificadas usando a codificação 1-de-4. Com a adoção dessa codificação, quatro variáveis binárias (0/1) de saída, correspondentes às 4 fases de crescimento foram codificadas ao final da representação de cada instância de dado, em um processo chamado de introdução de variáveis “dummy” (auxiliares). A cada variável “dummy” foi então atribuído o valor zero, exceto àquela correspondente à fase correta, à qual foi dado o valor 1, como mostram os exemplos listados na Tabela 4.11.

A última modificação nos conjuntos de dados antes do treinamento das redes neurais foi dividir cada um deles em dois subconjuntos: um subconjunto para a realização do processo de treinamento com validação cruzada e outro para uma validação independente. A validação com conjunto de dados independente é importante para verificar se houve treinamento

em excesso (over fitting). O conjunto estendido Exp_5^{est} com 1441 instâncias, foi escolhido para compor o conjunto para validação independente, VAL_Exp. Os demais dados (1441 dados para cada Exp_i^{est} , com $i=1, 2, 3$ e 4) foram utilizados para o aprendizado das redes neurais com validação cruzada. A

Figura 4.12 resume as etapas de tratamento e preparação dos dados para o aprendizado de redes neurais.

Tabela 4.11. Esquema adotado para representar instâncias de treinamento que caracterizam quatro fases.

Tempo (h)	DO (600 nm)	NaOH (g)	Fase		Tempo	DO	NaOH	Classe			
0,2	0,16	0	Lag e Aceleração	⇒	0,2	0,16	0	1	0	0	0
1,8	0,66	10,57	Exponencial		1,8	0,66	10,57	0	1	0	0
3,8	3,61	133,76	Desaceleração		3,8	3,61	133,76	0	0	1	0
6,0	5,46	261,18	Estacionária e Morte		6,0	5,46	261,18	0	0	0	1

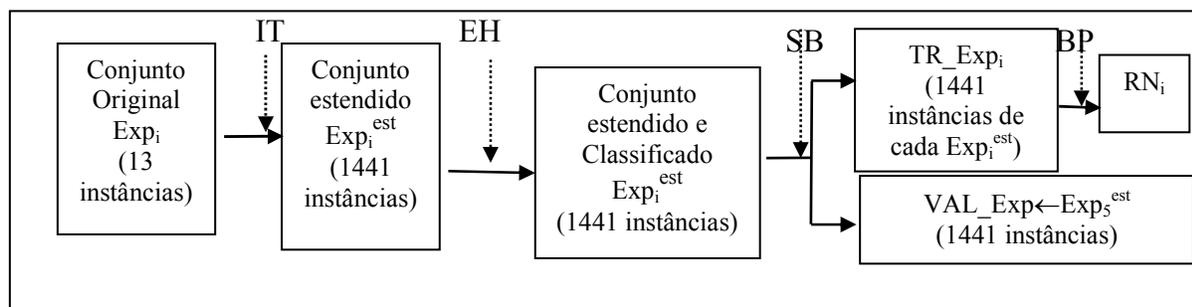


Figura 4.12. Esquema da preparação dos dados de cada experimento para serem utilizados no treinamento de cada rede neural individual, RN_i , ($i=1, 2, 3, 4$). Na figura IT é referente ao processo de interpolação; EH à classificação das instâncias de dados em sua respectiva fase de crescimento por um especialista humano; SB à divisão dos dados em conjunto de treinamento (TR_Exp_i) e conjunto de validação (VAL_Exp); BP refere-se ao treinamento backpropagation resultando na rede neural i (RN_i).

O estabelecimento de uma rede neural com a finalidade de reconhecer as fases de crescimento de um microrganismo não é uma tarefa trivial. Os algoritmos tradicionais, como o backpropagation, requerem a definição prévia da arquitetura da rede antes do treinamento. Em geral estes métodos só apresentam bons resultados se a arquitetura da rede neural for apropriada. Como não existe um método que defina a arquitetura da rede neural para cada aplicação, várias devem ser avaliadas na busca da mais conveniente (Giordano *et al.*, 2007; Nicoletti e Bertini, 2007; Palma Neto e Nicoletti, 2005; Linko *et al.*, 1999). O primeiro passo foi definir as diferentes arquiteturas a serem avaliadas:

- a) uma, duas e três camadas intermediárias;
- b) diferentes números de neurônios da camada intermediária (2, 3, 4, 5, 6, 7, 9, 10, 12 e 15);
- c) diferentes funções de ativação, tanto na camada intermediária quanto na de saída. As diferentes combinações testadas foram: (a) tan-sigmoide e linear, (b) log-sigmoide e linear e (c) log-sigmoide e log-sigmoide, respectivamente;
- d) diferentes épocas de treinamento; foram utilizadas 10, 25, 50, 100, 200 e 2000 épocas;
- e) diferentes variáveis de entrada. Cinco conjuntos de variáveis (variáveis input) foram avaliados: {DO}, {DO, NaOH}, {t, DO}, {t, NaOH}, {t, DO, NaOH} e todos {t, DO, C_X, C_S, C_L, C_{PS}, NaOH}.

Além disso, duas abordagens diferentes usando redes neurais foram analisadas: i) um comitê de redes neurais, constituído por quatro redes neurais treinadas cada uma com um respectivo conjunto de treinamento, Exp_1^{est} , Exp_2^{est} , Exp_3^{est} ou Exp_4^{est} e ii) uma rede neural clássica, treinada com os dados dos quatro experimentos. A Figura 4.13 mostra com detalhe o esquema de treinamento das redes neurais. O backpropagation¹ foi implementado usando o algoritmo de treinamento Levenberg-Marquardt.

¹ Todos os programas foram implementados no programa MatLab[®] 6.5 (Mathworks).

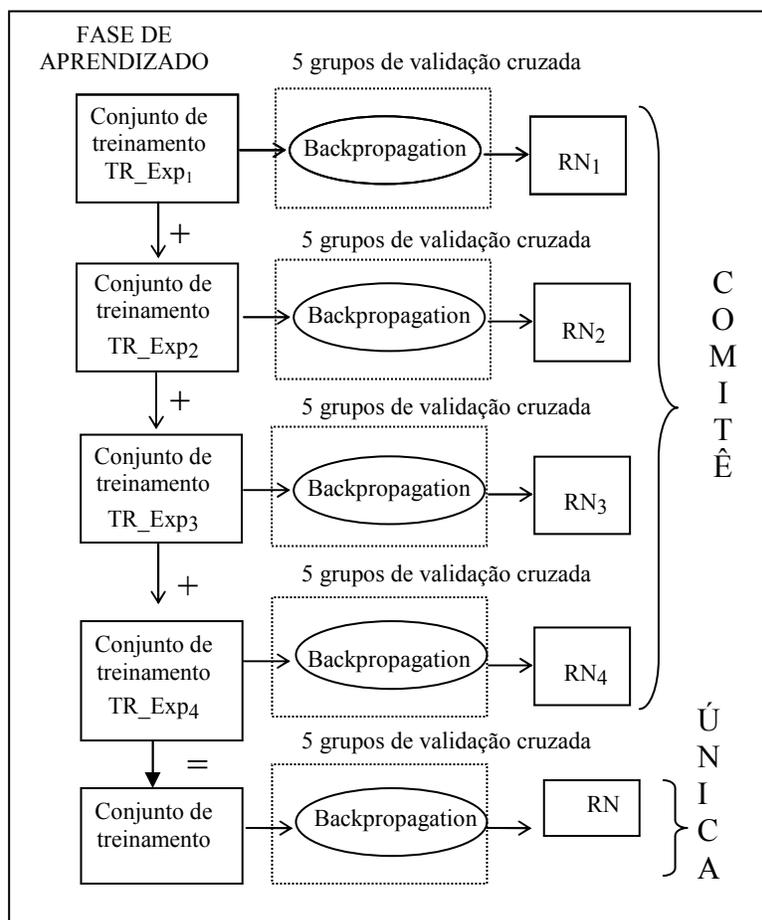


Figura 4.13. Esquema do treinamento das redes neurais com os dados estendidos e classificados dos respectivos experimentos.

O processo de avaliação chamado de 5-validação cruzada divide os dados de treinamento em 5 grupos, treinando a rede neural com os dados de 4 grupos e validando o desempenho de classificação da rede treinada com os dados do quinto grupo. O processo é repetido 5 vezes; a cada vez que é repetido um grupo diferente é considerado para validação.

Existem várias formas de integrar redes neurais em um comitê, conforme detalha Haykin (2001) e Bacauskiene e Verikas (2004). A forma escolhida de se computar a saída do comitê foi usando frequência, assim, a fase que apresentou maior frequência nas saídas das redes neurais foi tomada como resultado da classificação do comitê. Na Figura 4.14 cada uma das quatro redes neurais dá como saída uma fase de crescimento; o módulo de contagem verifica qual, dentre as quatro saídas é a mais frequente e esta é a saída do comitê.

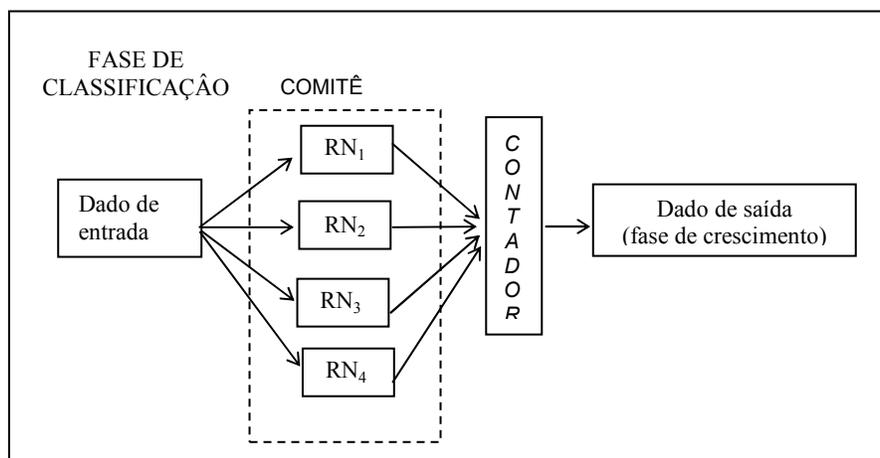


Figura 4.14. Comitê de redes neurais classificatórias. O procedimento contador é encarregado de decidir a classe final (i.e., a fase de crescimento), baseado na saída dada pelas quatro redes neurais.

Dentre as várias experimentadas, três arquiteturas tiveram bom desempenho: a com dois neurônios na camada de entrada, a com três neurônios na camada de entrada e a com 7 neurônios na camada de entrada; mantendo a configuração com 4 neurônios na camada de saída e variando o número de neurônios nas camadas intermediárias. As configurações mais promissoras estão descritas na Figura 4.15.

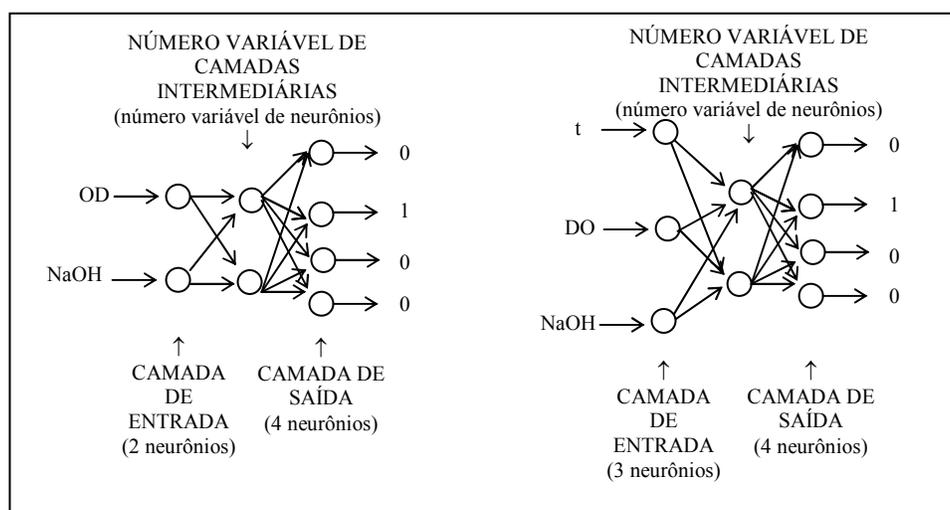


Figura 4.15. Arquiteturas das redes neurais mais promissoras. O número de neurônios testados na camada intermediária foram 2, 3, 4, 5, 6, 7, 9, 10 e 15; e com duas camadas as combinações 4-5, 5-4, 5-5, 5-7, 7-5 e 7-7, respectivamente.

Alguns resultados preliminares indicaram que os treinamentos feitos com 200 épocas apresentam bom resultado. Além disso, apesar de apresentarem resultados estatisticamente semelhantes, a utilização da função de ativação tan-sigmoide não é tão eficiente

quanto a log-sigmoide, visto que há como saída, valores negativos, o que demanda um esforço maior no tratamento destas saídas. Outra informação obtida com os resultados preliminares indica que o atributo de entrada DO não é um dado preciso para as redes classificatórias, quando utilizado sozinho. Percebe-se que em fases diferentes este atributo pode apresentar os mesmos valores.

Como os testes preliminares indicaram que a função log-sigmoide é a mais adequada para o problema, esta será a única utilizada para o treinamento das redes neurais.

Para se comparar a melhor configuração de entrada das redes neurais construiu-se a Figura 4.16 que representa o desempenho médio de classificação das redes neurais treinadas com as seguintes configurações:

- uma camada intermediária, com 5, 7, 9 e 10 neurônios;
- diferentes variáveis de entrada, $\{t, DO\}$, $\{DO, NaOH\}$, $\{t, NaOH\}$, $\{t, DO, NaOH\}$ e $\{Todos\}$;
- configuração de comitê e de rede única.

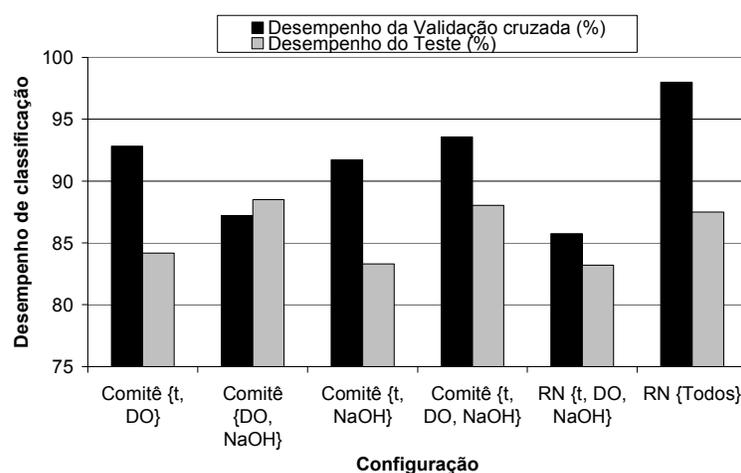


Figura 4.16. Desempenho de classificação do Comitê e da configuração de Rede única (RN) para diferentes variáveis de entrada. Treinamento feito com 200 épocas e uma camada intermediária. O desempenho corresponde à média das configurações com 5, 7, 9 e 10 neurônios na camada intermediária.

Infere-se da Figura 4.16 que as variáveis de entrada DO e massa de solução de NaOH conferem um poder de discernimento elevado às redes classificatórias, similar ao obtido na utilização de todas as variáveis e com as variáveis $\{t, DO, NaOH\}$. Além disso, as redes neurais únicas apresentaram ótimo desempenho de validação cruzada, porém não foram superiores no teste de classificação. Estes resultados devem ser analisados, tomando em

consideração, dois fatores: a rede única foi treinada com todos os dados de treinamento (isto é, os dados dos quatro experimentos), e assim, generaliza sobre os quatro experimentos. Em segundo lugar, cada rede neural, parte do comitê, representa a generalização de um único experimento.

Um aspecto bem interessante do comitê, que é consequência direta de sua inerente natureza incremental, é que novas redes neurais podem ser introduzidas ao comitê, à medida que novos experimentos forem sendo feitos, sem nenhum esforço adicional que não o de treinar uma rede neural com os dados do novo experimento disponível. Este procedimento já não se aplica ao caso da rede única uma vez que a obtenção de novos dados vai implicar o treinamento da rede usando todo o conjunto de dados (antigo & novo).

A configuração de entrada {t, DO, NaOH} com apenas uma camada intermediária apresentou o desempenho de classificação ilustrado na Figura 4.17, para números diferentes de neurônios na camada intermediária.

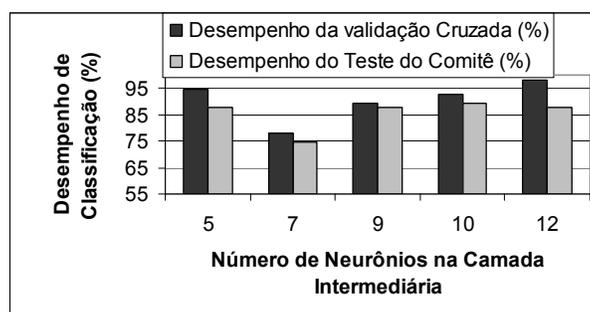


Figura 4.17. Desempenho do comitê classificatório treinado com as variáveis de entrada {t, DO, NaOH} e com redes com uma camada intermediária com 5, 7, 9, 10 e 12 neurônios.

De acordo com a Figura 4.17, apesar do desempenho de validação cruzada aumentar com o aumento no número de neurônios na camada intermediária, o desempenho do teste diminui, apresentando um máximo com 10 neurônios.

Em seguida testaram-se diferentes números de camadas intermediárias, para a configuração de Comitê com as variáveis de entrada {t, DO, NaOH}. Os resultados estão na Figura 4.18.

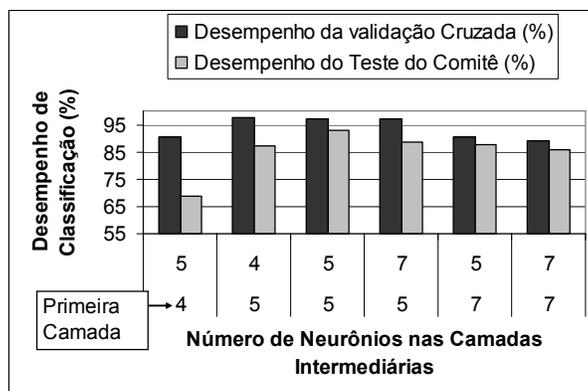


Figura 4.18. Desempenho do comitê classificatório com variáveis de entrada {t, DO, NaOH} e com duas camadas intermediárias com 4-5, 5-4, 5-5, 5-7, 7-5 e 7-7 neurônios nas respectivas camadas.

De acordo com a Figura 4.18 a melhor configuração encontrada foi a com 5 neurônios em ambas as camadas intermediárias, com 92% de classificações corretas no teste. Porém a configuração de entrada com {DO, NaOH} com duas camadas intermediárias com 4 e cinco neurônios respectivamente também apresentou classificação de 92%, como pode ser observado na Figura 4.19. A dificuldade de classificação é aumentada com a utilização de cepas diferentes, para calibração foram empregados dados de 3 experimentos da cepa I e 1 experimento da cepa II, enquanto que os dados de validação provieram da cepa II.

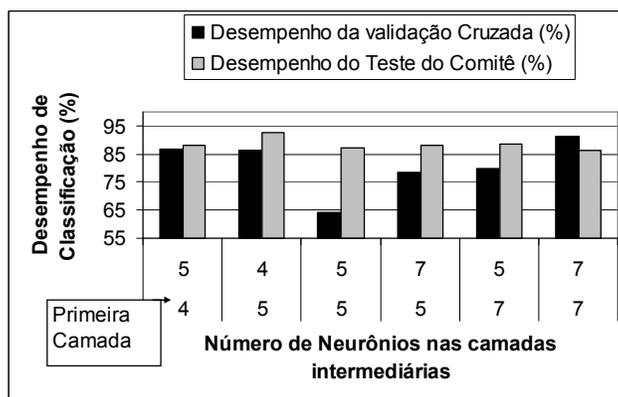


Figura 4.19. Desempenho do comitê classificatório com variáveis de entrada {DO, NaOH} e com duas camadas intermediárias com 4-5, 5-4, 5-5, 5-7, 7-5 e 7-7 neurônios nas respectivas camadas.

Em geral, todas as variáveis juntas conseguem dar boa precisão para as redes neurais no processo de classificação das fases de crescimento, porém as variáveis DO e NaOH com o tempo ou não, também fornecem dados satisfatórios para as classificações. Este resultado é de extrema importância visto que estas variáveis podem ser medidas *on-line*, ou quase *on-line*,

o que indica que o comitê de redes neurais pode ser usado como parte de uma ferramenta de inferência e controle *on-line*.

Dos resultados obtidos, as configurações mais promissoras foram as redes neurais compostas por duas camadas intermediárias, com duas possíveis configurações de entrada: {DO, NaOH} com 5 neurônios na primeira camada intermediária e 4 na segunda camada intermediária; e {t, DO, NaOH} com 5 neurônios em ambas as camadas intermediárias.

Depois de treinadas as redes e composto o comitê, a próxima etapa antes da construção da ferramenta híbrida de inferência é o estudo da modelagem segregada por fases.

4.4 Modelagem não Estruturada de Cultivo em Batelada Segregada por Fases de Crescimento

A modelagem não estruturada clássica para cultivos de *S. pneumoniae* foi apresentada e discutida nas Seções 4.1 e 4.2. Na modelagem segregada por fases de crescimento, foram empregadas as mesmas equações de balanço material (4.6) a (4.9), a cinética de crescimento de Tessier (4.2) e o modelo de Luedking-Piret [(4.4) e (4.5)] para a cinética de formação dos produtos ácido lático e polissacarídeo capsular já descritas na Seção 4.1. A mesma seqüência de etapas presente na Seção 4.1 para o estabelecimento do modelo e verificação da sua capacidade descritiva também foi adotada.

A estimativa dos parâmetros do modelo não estruturado segregado por fases foi feita com os dados experimentais do cultivo Exp₂, após serem tratados de acordo com a metodologia de alisamento e interpolação, que produziu a versão estendida daqueles dados, referenciada como Exp₂^{est}. Os parâmetros de alisamento e interpolação utilizados no algoritmo Alisador Perfeito foram os mesmos da Seção 4.1. A Figura 4.1 da Seção 4.1 ilustra o resultado do alisamento juntamente com os dados do Exp₂. Porém, para a estimativa dos parâmetros na modelagem segregada por fases, foi necessário primeiramente subdividir os dados experimentais estendidos em conjuntos correspondentes às diferentes fases de crescimento para então proceder à calibração do modelo. O mesmo critério descrito na Seção 4.3, ou seja, a identificação das fases de crescimento por meio de análise da curva dDO/dt em função do tempo (Figura 4.11) por um especialista humano, foi aplicado. A classificação resultante da aplicação do critério é mostrada na Tabela 4.10. Finalmente, procedeu-se à estimativa dos parâmetros para cada fase de

crescimento, usando o algoritmo de busca heurística *Simulated Annealing*, o qual foi implementado com a mesma configuração definida na modelagem convencional descrita na Seção 4.1: variável T inicializada com 10^4 e fator de decréscimo de 0,999; função objetivo ponderada; limites superiores e inferiores estipulados para cada parâmetro do modelo descritos na Tabela 4.1.

Os resultados da simulação do modelo segregado por fases podem ser visualizados na Figura 4.20, juntamente com os valores experimentais das variáveis concentração de células, glicose, PS e lactato obtidos no cultivo Exp₂. Para comparação, os resultados da simulação do modelo não-estruturado convencional descrito na Seção 4.1 também foram incluídos na Figura 4.20.

Pela análise da Figura 4.20, observa-se que os resultados do modelo ajustado por fases se aproximam mais dos dados experimentais que o modelo convencional, que descreve todo o cultivo, destacando-se principalmente o ajuste aos dados da concentração glicose. Além disso, a soma dos valores da função objetivo do ajuste por fases foi melhor que o ajuste a toda batelada, com 195,08 para a modelagem por fases e 201,56 para a modelagem clássica. Os valores da função objetivo correspondentes a cada fase podem ser observados na Tabela 4.12 e os valores dos parâmetros ajustados na Tabela 4.13.

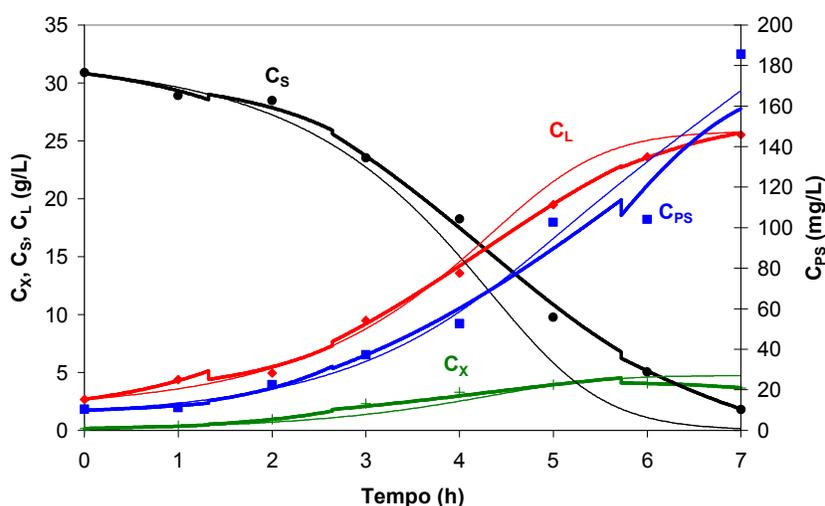


Figura 4.20. Resultados do ajuste do modelo de Tessier para a concentração celular (C_X), concentração de substrato (C_S), concentração de lactato (C_L) e concentração de polissacarídeo (C_{PS}). As linhas grossas correspondem aos dados simulados pelo modelo segregado por fases; as linhas finas correspondem aos dados simulados pelo modelo geral (Seção 4.1); os símbolos correspondem aos dados experimentais brutos do cultivo Exp₂.

Tabela 4.12. Valores da função objetivo para cada respectiva fase, Lag e Aceleração, Exponencial, Desaceleração e Estacionária e Morte.

Valor da Função objetivo	Fase				TOTAL
	Lag e Aceleração	Exponencial	Desaceleração	Estacionária e Morte	
	90,82	26,03	54,69	23,54	195,08

Tabela 4.13. Valores dos parâmetros ajustados a cada fase de crescimento do cultivo **Exp₂** - dados estendidos.

Fase de crescimento	K	m	K _D	Y _{XS}	α _{PS}	μ _m	α _L	β _{PS}	β _L	Y _{LS}
Lag e Aceleração	0,1616	0,0001	0,0001	0,3836	3,5419	0,8259	6,7923	6,3041	0,0041	1,7399
Exponencial	0,2302	0,0032	0,0002	0,5529	4,6598	0,8935	2,2238	9,1294	0,0033	1,9078
Desaceleração	0,0299	0,0002	0,0006	0,2105	1,5536	0,8185	5,6002	8,1075	0,0036	1,9588
Estacionária e Morte	0,0631	0,0001	0,1384	0,2624	1,4759	0,3282	8,6645	12,9851	0,0011	0,9056

A Tabela 4.13 mostra que os valores dos parâmetros efetivamente mudam de acordo com a fase de crescimento. A mudança mais significativa é observada no parâmetro K_D , o qual se torna expressivo na fase de morte, como esperado. Os valores estimados para os parâmetros K e Y_{XS} tendem a diminuir do início para o fim do cultivo, possivelmente refletindo a adaptação do metabolismo celular à gradual diminuição na concentração de substrato. A variação do valor do parâmetro μ_m também é coerente com as condições de crescimento em cada fase e apenas na fase estacionária e de morte, assume um valor significativamente reduzido, refletindo o efeito de inibição provocado pelo acúmulo de ácido lático. Em relação à cinética de formação de PS, o termo de produção não associado ao crescimento (β_{PS}) se torna maior em relação ao termo associado (α_{PS}) para acompanhar a intensificação da liberação de PS da superfície das células para o meio à medida que o cultivo avança. Já para a cinética de produção de lactato, apenas o termo associado ao crescimento (α_L) é significativo e aumenta da fase exponencial para a fase estacionária/morte, provavelmente como consequência do aumento gradativo do consumo de glicose para suprir os requerimentos energéticos voltados à manutenção celular. O parâmetro m apresenta variação desprezível, como já observado na modelagem convencional (Seção 4.1). Porém, este parâmetro geralmente se torna mais relevante quando os cultivos atingem altas concentrações celulares (10 g/L ou mais), o que não acontece nos experimentos utilizados no

presente trabalho.

Os valores assumidos pelos parâmetros na modelagem por fases (Tabela 4.13) podem ainda ser comparados com os valores estimados na modelagem convencional (Tabela 4.2). Restringindo-se a comparação aos parâmetros estatisticamente significativos, observa-se que os valores estimados para μ_m nas 3 fases iniciais do crescimento são similares ao valor de μ_m estimado no ajuste do modelo não estruturado convencional ($0,887 \text{ h}^{-1}$). Já os valores de α_L e de β_{PS} estimados na modelagem convencional (5,2 e 10,3, respectivamente) se aproximam da média dos valores apresentados na Tabela 4.13 para estes parâmetros. O valor do parâmetro K (0,055, na modelagem convencional) é semelhante aos valores menores de K, estimados para as duas últimas fases de crescimento, que coincidem com o período do cultivo no qual a concentração de glicose é mais baixa e a contribuição da constante K na expressão cinética de Tessier é determinante para o ajuste.

A análise do desempenho do modelo segregado por fases de crescimento em relação à sua capacidade de descrever outros cultivos conduzidos em condições semelhantes às do cultivo Exp_2 é discutida a seguir. Para a validação, foram selecionados os cultivos Exp_3 , Exp_4 e Exp_5 . No entanto, para gerar uma quantidade de dados suficiente para representar cada fase do crescimento assim como a transição entre elas, os dados experimentais brutos dos cultivos selecionados foram submetidos ao tratamento de interpolação e alisamento, seguindo os mesmos critérios adotados para o Exp_2 . Foram assim gerados os conjuntos de dados experimentais estendidos $\text{Exp}_3^{\text{est}}$, $\text{Exp}_4^{\text{est}}$ e $\text{Exp}_5^{\text{est}}$. As variáveis dos dados utilizados foram: tempo (t), concentração celular (C_X), concentração de substrato, (C_S), concentração de lactato (C_L) e concentração de polissacarídeo (C_{PS}), como no caso de modelagem convencional. Os dados experimentais obtidos no cultivo Exp_1 não foram utilizados na validação por estarem incompletos (as medidas experimentais de concentração de lactato não foram realizadas).

Os resultados da simulação do modelo para os cultivos Exp_3 , Exp_4 e Exp_5 são apresentados na Figura 4.21 e na Figura 4.22, juntamente com os dados experimentais brutos obtidos nos respectivos cultivos. Como esperado, a capacidade descritiva do modelo segregado por fases é superior a do modelo não segregado convencional. Esta superioridade fica evidente quando observamos o ajuste dos dados simulados aos pontos experimentais da concentração de polissacarídeo capsular dos cultivos Exp_4 e Exp_5 na Figura 4.22 e comparamos com os resultados exibidos na Figura 4.6, para a modelagem convencional. O alto valor estimado para o parâmetro

β_{PS} na última fase de crescimento (Tabela 4.13) efetivamente contribuiu para melhorar a capacidade descritiva do modelo.

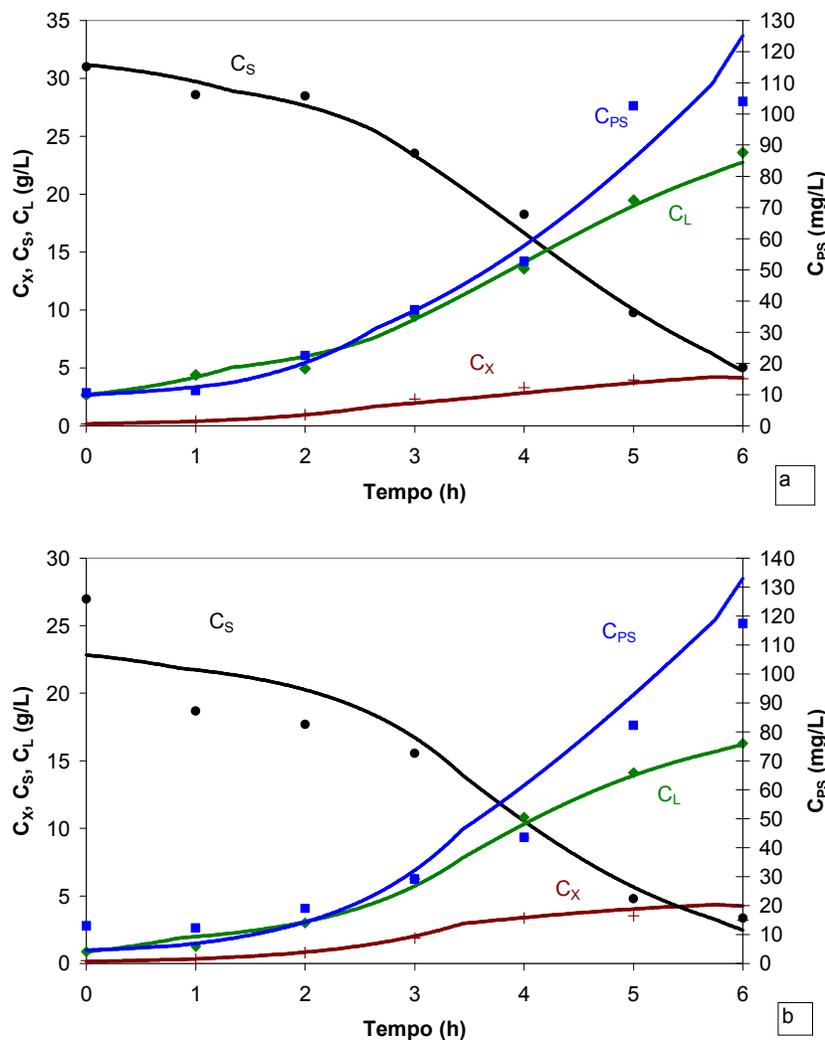


Figura 4.21. Simulação do (a) Exp_2^{est} e (b) Exp_3^{est} com os parâmetros do modelo de Tessier ajustado em cada respectiva fase de crescimento do Exp_2^{est} .

Os valores da função objetivo expressando a diferença entre os dados simulados e os dados experimentais estendidos para a modelagem por fases estão listados na Tabela 4.14. A simulação do cultivo Exp_2^{est} foi interrompida na sexta hora, para permitir a comparação entre os valores da função objetivo de diferentes cultivos. Pode-se constatar, ao analisar os valores apresentados na Tabela 4.14, que a modelagem por fases consegue representar com mais precisão a maioria dos experimentos, com exceção apenas do Exp_5^{est} . Em geral o modelo clássico ajustado

a toda batelada apresentou função objetivo 40% superior ao modelo segregado por fases. Os altos valores da função objetivo para o ajuste por fases aos dados Exp_3^{est} , Exp_4^{est} e Exp_5^{est} é principalmente devido à baixa qualidade das medidas experimentais da concentração de PS ou da glicose nestes cultivos, bem como da utilização de cepas diferentes, conforme já discutido na Seção 4.1.

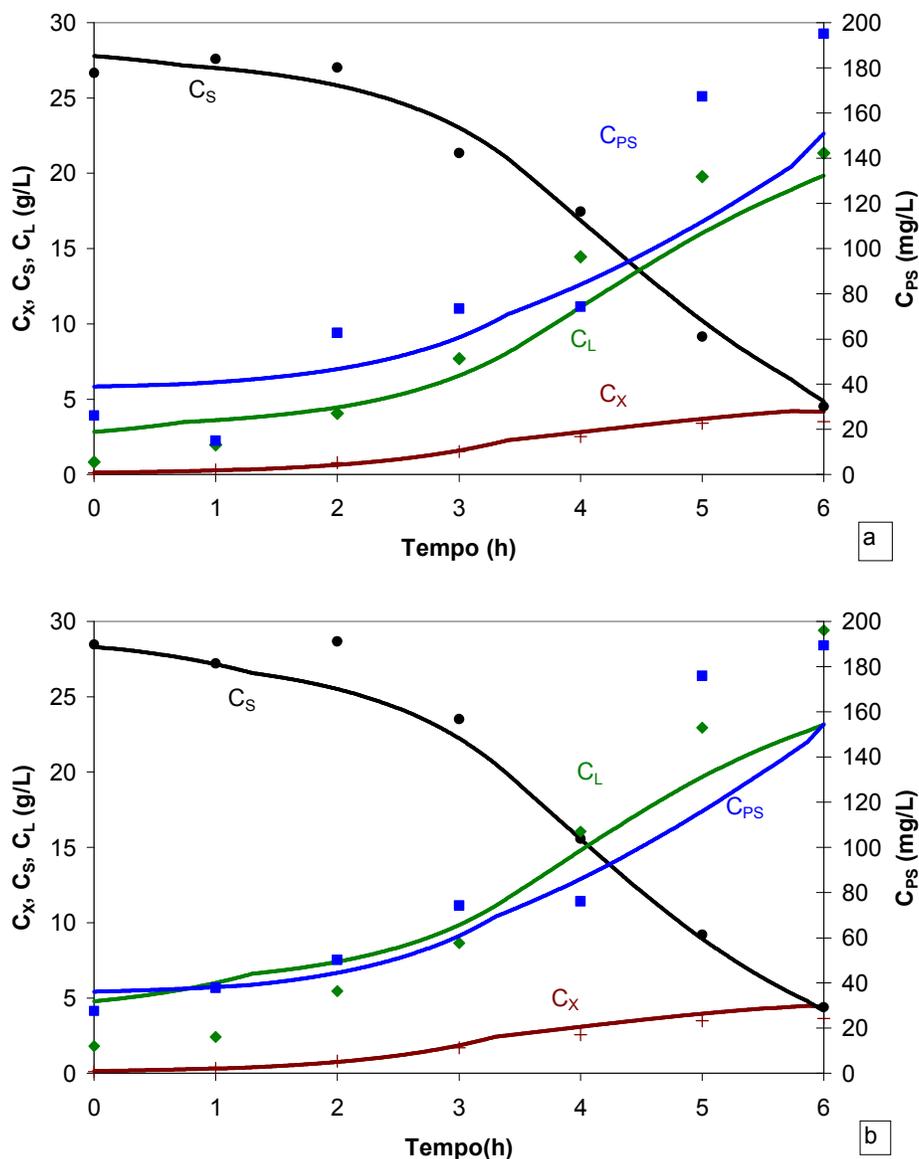


Figura 4.22. Simulação do (a) Exp_4^{est} e (b) Exp_5^{est} com os parâmetros do modelo de Tessier ajustado em cada respectiva fase de crescimento do Exp_2^{est} ; f é o valor da função objetivo.

Tabela 4.14. Valor da função objetivo para as simulações do Exp_2^{est} , Exp_3^{est} , Exp_4^{est} e Exp_5^{est} , comparando os resultados do modelo ajustado a tudo batelada com o segregado por fases de crescimento.

Experimento	Valor da Função Objetivo	
	Modelo segregado por fases	Modelo ajustado a toda batelada
Exp_2^{est}	33,40	186,39
Exp_3^{est}	140,67	296,34
Exp_4^{est}	250,15	279,37
Exp_5^{est}	277,71	226,19

É importante ressaltar que a modelagem por fases provavelmente mostrará maior superioridade em relação à abordagem clássica, quando for aplicada a cultivos mais longos, tais como batelada alimentada ou batelada alimentada com sistema de retenção de células. Nestes tipos de cultivo, as diferentes fases de crescimento se prolongarão, e fases de crescimento intermediárias entre a fase exponencial e a estacionária surgirão. Na modelagem clássica, essas mudanças são artificialmente incorporadas na cinética de crescimento por meio de expressões que levam em conta a inibição pelo substrato e pelo produto. Na modelagem por fases, o modelo permanece o mesmo, porém os valores dos parâmetros mudam para acompanhar as alterações na cinética de crescimento e formação de produtos, conforme comentado anteriormente, levando a uma descrição mais fiel do que acontece a nível celular.

De posse do modelo segregado em fases e do comitê de redes neurais a construção da ferramenta de inferência se torna um trabalho de mesclar as duas ferramentas em uma única.

4.5 Ferramenta de Inferência baseada na modelagem segregada por fases de crescimento e no Comitê de Redes Neurais Classificatório

A ferramenta de inferência corresponde à associação entre o comitê de redes neurais que identifica as fases de crescimento (Seção 4.3) com a modelagem por fases, que por sua vez, simula as concentrações das principais variáveis de estado do cultivo, calculando os valores desconhecidos de C_X , C_S , C_L e C_{PS} (Seção 4.4). A Figura 4.23 ilustra o esquema geral do funcionamento da ferramenta de inferência. Inicialmente, os dados de densidade ótica (DO) e de massa da solução de NaOH para controle do pH (NaOH) são alimentados ao comitê classificatório. O algoritmo ESPM (algoritmo de escolha de parâmetros), recebe a fase prevista

pelo comitê de redes neurais, escolhe um vetor de parâmetros previamente ajustado à respectiva fase e envia este vetor para o algoritmo de simulação do modelo. O algoritmo de simulação recebe o vetor de parâmetros, as condições iniciais do cultivo (que devem ser digitadas pelo operador), o intervalo de tempo de integração do modelo e estima os valores das variáveis C_X , C_S , C_L , C_{PS} no intervalo de tempo definido.

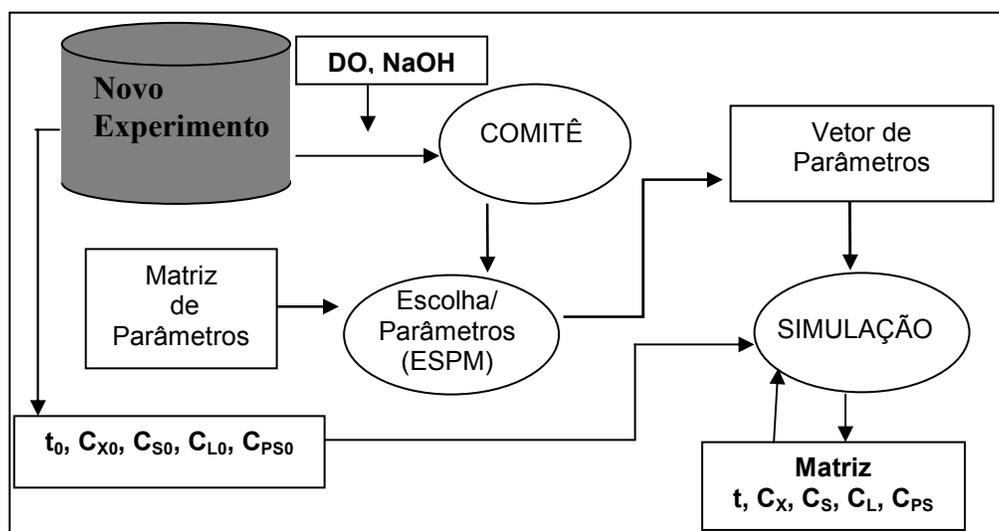


Figura 4.23. Esquema geral da ferramenta de inferência *on-line*. DO se refere à densidade ótica do meio, enquanto que NaOH é referente à massa de solução alcalina adicionada no meio de cultura para controlar o pH. ESPM é o algoritmo de escolha de parâmetros; t o tempo, C_X a concentração celular, C_S a concentração de glicose, C_L a concentração de lactato e C_{PS} a concentração de polissacarídeo capsular.

A precisão da ferramenta de inferência depende da atualização constante do vetor de parâmetros. Por isso, a concepção da ferramenta aqui descrita foi baseada em uma entrada constituída pelas variáveis $\{DO \text{ e } NaOH\}$ ou $\{t, DO \text{ e } NaOH\}$, adquiridos automaticamente com uma frequência de 1 minuto ou menos e fornecidos em tempo real ao comitê de redes neurais. Nesta situação, a ferramenta pode ser utilizada tanto para a previsão da trajetória das diferentes variáveis do cultivo, como para a detecção precoce de cultivos mal sucedidos, ou seja, que apresentem desvios significativos em relação à trajetória simulada. As limitações práticas para a implementação de um algoritmo de inferência alimentado por dados de DO *on-line* serão analisadas no final desta seção.

A partir dos dados experimentais de DO e NaOH, dados fictícios de DO e NaOH foram gerados em intervalos de tempo de 30 em 30 segundos por interpolação. Estes dados foram utilizados para simular cultivo em tempo real. Os dados parâmetros utilizados no procedimento

de interpolação e alisamento foram os mesmos utilizados nas Seções 4.1, 4.2 e 4.3. O comitê de redes neurais foi composto por cinco redes neurais, cada uma treinada com os dados de um experimento Exp_1^{est} , Exp_2^{est} , Exp_3^{est} , Exp_4^{est} e Exp_5^{est} .

A capacidade descritiva da ferramenta de inferência foi avaliada comparando os dados simulados com os dados estendidos Exp_2^{est} , Exp_3^{est} , Exp_4^{est} e Exp_5^{est} . O valor da função objetivo foi calculado em relação aos pontos experimentais interpolados até a sexta hora, para os quatro experimentos, visando viabilizar a comparação entre o desempenho da ferramenta.

Os resultados da ferramenta de inferência podem ser observados na Figura 4.24 e Figura 4.25, nas quais as linhas correspondem aos resultados simulados dos respectivos experimentos, enquanto os símbolos correspondem aos pontos medidos experimentalmente. Pela análise dos resultados apresentados nas figuras, nota-se que o desempenho da ferramenta de inferência não foi satisfatório, mostrando-se inferior ao desempenho obtido na modelagem segregada por fases e até na modelagem convencional. Esta constatação é reforçada pela análise dos valores da função objetivo para a ferramenta de inferência, relacionados na Tabela 4.15.

Os valores da função objetivo presentes na Tabela 4.15 são visivelmente superiores aos valores relacionados na Tabela 4.6 (Seção 4.1), para a modelagem convencional, e na Tabela 4.14 (Seção 4.4), para a modelagem por fases. Em porcentagem, o valor médio da função objetivo para a ferramenta foi 56% maior que o valor médio da FO para o modelo convencional descrevendo a batelada integral e 100% maior que o valor da FO para a modelagem segregada por fases.

Analisando ainda os resultados apresentados na Tabela 4.15, observa-se que as diferentes configurações de entrada $\{DO, NaOH\}$ e $\{t, DO, NaOH\}$ investigadas obtiveram praticamente o mesmo desempenho, apresentando praticamente o mesmo valor médio para a função objetivo.

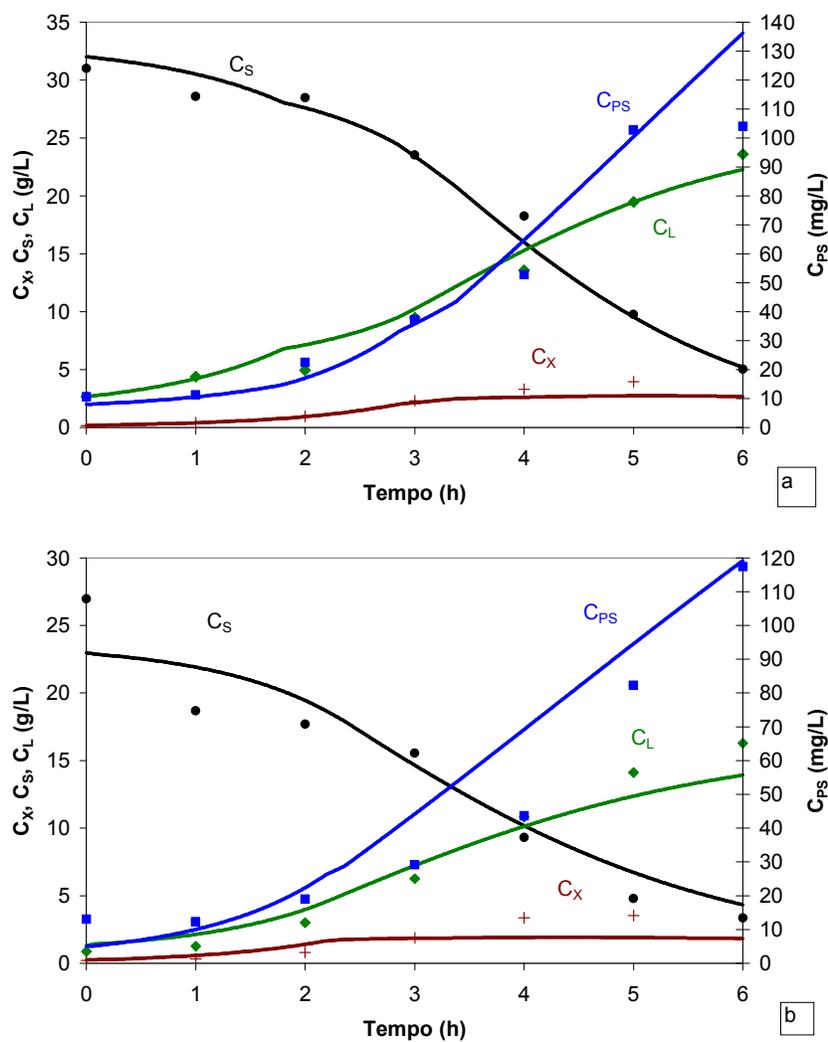


Figura 4.24. Simulação do Exp₂ (a) e Exp₃ (b), com a ferramenta de inferência tendo como entrada os atributos $\{t, DO, NaOH\}$.

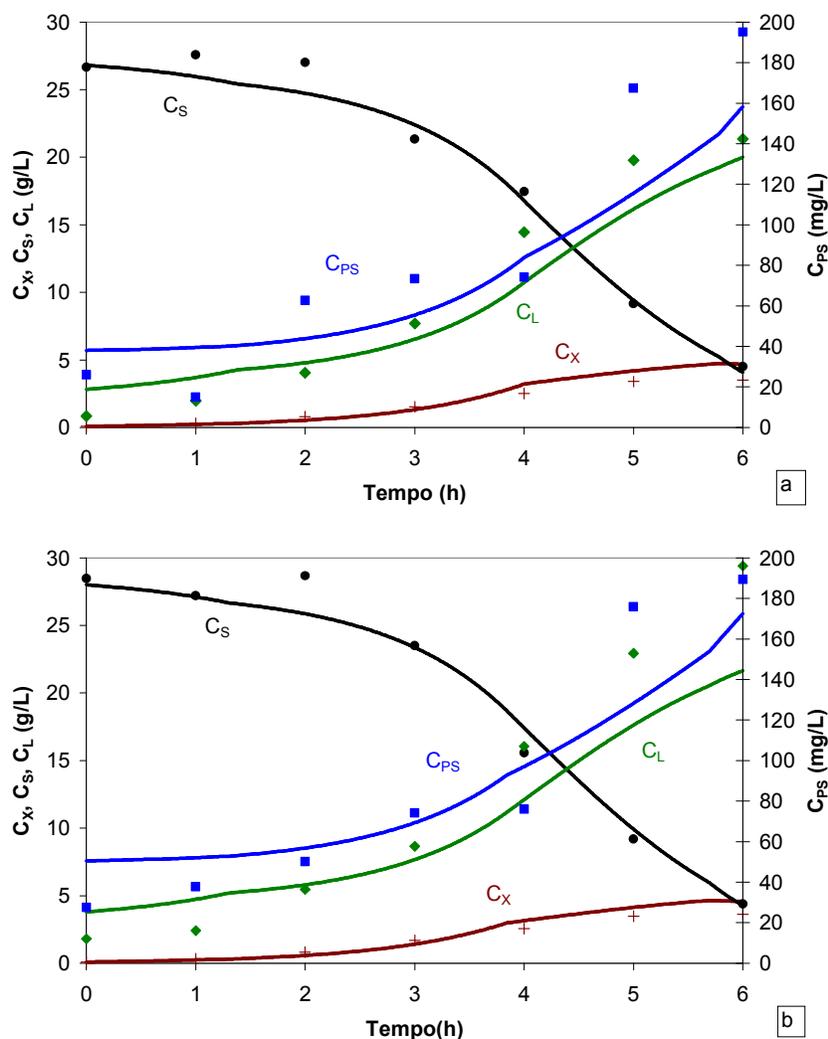


Figura 4.25. Simulação do Exp₄ (a) e Exp₅ (b), com a ferramenta de inferência tendo como entrada os atributos {t, DO, NaOH}.

Tabela 4.15. Valores das respectivas funções objetivo das simulações dos Exp₂^{est}, Exp₃^{est}, Exp₄^{est} e Exp₅^{est} com a ferramenta de inferência. DO corresponde à densidade óptica, NaOH à massa de solução da base para controlar o pH e t corresponde ao tempo.

Atributos de entrada	Valor da Função Objetivo				Média
	Exp ₂ ^{est}	Exp ₃ ^{est}	Exp ₄ ^{est}	Exp ₅ ^{est}	
t, DO, NaOH	237,95	597,14	363,71	314,30	378,27
DO, NaOH	231,65	573,03	370,96	337,30	378,23

Uma série de fatores contribuiu para que a capacidade descritiva da ferramenta de inferência deixasse a desejar. O primeiro fator a ser considerado refere-se ao treinamento das

redes neurais. Na Seção 4.3, o comitê foi constituído por redes neurais treinadas com dados estendidos correspondentes a 11 h de cultivo. Porém, na ferramenta, este mesmo comitê foi utilizado para classificar fases em cultivos de 6 horas. O principal problema, neste caso, é que os cultivos com 11 horas de duração foram executados mediante suplementação por pulsos de glicose, os quais não são descritos pelo modelo. Isso explica por que o mesmo comitê que acertou na classificação de dados semelhantes aos do treinamento (Seção 4.3), falhou na classificação das fases quando inserido na ferramenta de inferência. Deve-se ainda levar em conta que a quantidade de experimentos (apenas 5 cultivos) disponíveis para treinamento das redes neurais é pequeno e o erro de inferência tende a diminuir à medida que dados de novos experimentos sejam acrescentados ao comitê de redes neurais. Outra consideração importante refere-se ao critério de divisão dos cultivos em fases de crescimento por meio da identificação visual dos pontos de inflexão na curva de dDO/dt contra o tempo por especialista humano. Ele é subjetivo, de difícil padronização e pode estar levando a erros que prejudicam o desempenho da ferramenta de inferência. Além disso, a utilização de cepas diferentes prejudica o desempenho da ferramenta, como comentado anteriormente nas seções 4.1, 4.3 e 4.4.

Finalizando os comentários sobre a ferramenta de inferência, deve ser retomada a questão mencionada no início desta seção sobre a importância da atualização constante do vetor de parâmetros, o que requer que as variáveis de entrada do comitê de redes neurais sejam adquiridas em intervalos de tempo curtos e fornecidas em tempo real. A aquisição *on-line* dos dados da variável de entrada “massa de solução de NaOH consumida” já está implementada no programa LabView e portanto preenche os requisitos. Porém, para a variável “densidade ótica” a situação é muito mais complexa. Embora existam dispositivos experimentais como o sistema FIA (Flow Injection Analysis) que permitem a quantificação em tempo real da DO (Nielsen e Villadsen, 1994), estes dispositivos são de custo elevado, exigem manutenção constante e principalmente, não são indicados para cultivos de microrganismos patogênicos, como é o caso da bactéria *S. pneumoniae*.

Para contornar esta limitação, uma alternativa seria introduzir mais um comitê de redes neurais na ferramenta. Este novo comitê auxiliar teria como entrada os valores de C_X , C_S , C_L , e C_{PS} calculados pelo simulador e como saída, a fase de cultivo obtida a partir dos dados simulados. Esta fase seria a entrada para o ESPM que atualizaria a cada iteração o vetor de parâmetros até que um novo dado experimental de densidade ótica esteja disponível. Neste

momento, o comitê principal assumiria a identificação da fase do cultivo, que seria enviada para o ESPM para a identificação do vetor de parâmetros, seguido pela transferência da tarefa de identificação das fases do cultivo para o comitê auxiliar até a obtenção de um novo valor experimental de DO.

Capítulo 5 – CONSIDERAÇÕES FINAIS

Várias conclusões importantes podem ser extraídas dos resultados apresentados nas diferentes seções.

Na Seção 4.1, verificou-se que o algoritmo SA foi adequado ao problema proposto: apresentou convergência da função objetivo de forma eficiente; permitiu a escolha de um modelo não-estruturado, o modelo de Tessier, dentre os três testados, que por sua vez representou satisfatoriamente o perfil das principais variáveis do processo de produção de polissacarídeo em cultivos em batelada de *Streptococcus pneumoniae*. O mesmo algoritmo de busca heurística foi ainda utilizado para a estimativa de parâmetros cinéticos na modelagem não estruturada do cultivo em batelada alimentada com reciclo celular (Seção 4.2) e na modelagem por fases de crescimento (Seção 4.4). Em outra aplicação, permitiu a identificação do perfil ótimo de alimentação e retirada do meio para o cultivo em batelada alimentada com retenção de células. Em todas as aplicações, o algoritmo SA mostrou-se versátil, robusto e confiável, o que o torna altamente indicado para a estimativa de parâmetros na modelagem de bioprocessos e na otimização dos mesmos.

Na Seção 4.3 um comitê de redes neurais foi construído e validado, com a finalidade de reconhecer a fase de crescimento da bactéria a partir de dados obtidos *on-line*. A precisão de classificação do comitê formado foi de 92% de classificações corretas de um experimento não utilizado no treinamento das redes. Um desafio enfrentado na implementação do comitê foi definir as variáveis de entrada que seriam empregadas no treinamento das redes neurais. Em cultivos anaeróbios como o abordado no presente estudo, a disponibilidade de dados *on-line* é bastante restrita. Dados experimentais tipicamente adquiridos em tempo real como a concentração de oxigênio dissolvido e a fração molar de gás carbônico produzido na corrente de gases que sai do biorreator não estão disponíveis, já que não há consumo de oxigênio nem expressiva produção de gás carbônico em cultivos de *S. pneumoniae*. Assim, foram escolhidos como dados de entrada do comitê de redes neurais as variáveis massa de solução alcalina consumida e densidade ótica. A diminuição do pH em cultivos de *S. pneumoniae* se deve em grande parte à produção de ácido lático, o principal produto metabólico deste microrganismo. Assim, o consumo de solução de hidróxido de sódio é praticamente proporcional à produção de ácido lático, constituindo-se em valiosa informação sobre o desenvolvimento do cultivo. Porém,

é importante ressaltar que outros fatores também contribuem para a variação do pH em culturas de *S. pneumoniae*. Além do ácido láctico, outros ácidos orgânicos, que também causam diminuição do pH, são produzidos em proporção muito menor. Por outro lado, a degradação, pela bactéria, de fontes de nitrogênio complexas presentes no extrato de levedura e em outros nutrientes que constituem o meio de cultivo, pode levar à formação de amônia como subproduto e assim provocar uma elevação no valor do pH. A interferência destes fatores é uma possível explicação para o desempenho inferior no reconhecimento das fases de crescimento para o comitê treinado apenas com a massa consumida de solução de NaOH. Já a densidade ótica está diretamente relacionada à concentração celular e desde que possa ser fornecida ao sistema supervisor pelo operador assim que for medida, poderá ser considerada uma informação quase em tempo real. Porém, dificilmente poderia ser determinada experimentalmente em intervalos de tempo menores que 30 minutos, uma vez que a sua determinação requer retirada de amostras do meio de cultivo e manipulação das mesmas. De qualquer forma, é uma medida que pode ser realizada em poucos minutos, aproximadamente 5 minutos; o que justifica a utilização o termo “quase em tempo real”.

Como esperado, na Seção 4.4 mostrou-se que a modelagem não-estruturada pode simular, com mais precisão, as fases de crescimento do microrganismo quando separadas. Além disso, foi possível associar a variação dos valores estimados para os parâmetros nas diferentes fases de crescimento com as mudanças metabólicas correspondentes. Isso permite a construção de um modelo “enxuto”, com reduzido número de parâmetros, que não necessita da incorporação de termos adicionais para representar a inibição do crescimento por acúmulo de metabólitos ou por repressão catabólica. A simplicidade do modelo juntamente com a sua capacidade preditiva reforça a relevância da modelagem por fases para a implementação de uma ferramenta de inferência *on-line*. Ao analisarmos os resultados da modelagem e da ferramenta de classificação de fases, não podemos deixar de lado a informação de que dois dos cultivos realizados (Exp₄ e Exp₅) foram realizados partindo da mesma cepa de *S. pneumoniae*, a qual porém pertencia a um lote diferente da cepa inicialmente utilizada. Esta informação é de extrema importância, pois ressalta a robustez do modelo e do comitê de redes neurais.

Assim, confirmada a viabilidade do reconhecimento das fases de crescimento pelo comitê de redes neurais bem como da modelagem por fases de crescimento (Seções 4.3 e 4.4), a ferramenta de inferência de C_X , C_S , C_L e C_{PS} foi construída e avaliada na Seção 4.5. O comitê de

redes neurais foi formado por 5 redes neurais, cada uma treinada com os dados de um experimento específico (Exp_1^{est} , Exp_2^{est} , Exp_3^{est} , Exp_4^{est} e Exp_5^{est}). O desempenho da ferramenta de inferência não se mostrou superior ao desempenho da simulação do modelo ajustado a toda batelada, pelas razões já expostas ao final da Seção 4.5. No entanto, este fato não deve, de forma alguma, ser interpretado como uma indicação de fracasso da metodologia proposta no presente trabalho para o desenvolvimento de uma ferramenta de inferência. De fato, para cultivos simples como os conduzidos em batelada, o esforço adicional de modelagem por fases de crescimento não se justifica, já que os modelos não estruturados convencionais são razoavelmente capazes de simular as principais variáveis de estado, nestes casos. Até a divisão do cultivo em diferentes fases de crescimento é difícil, pois as bateladas são curtas (terminam após 6 a 10 h de cultivo, dependendo da concentração inicial de substrato), contêm fases de crescimento com duração inferior a uma hora, sendo possível retirar apenas uma ou duas amostras neste período, o que restringe muito a quantidade de informação disponível.

Por outro lado, no modo de operação em batelada alimentada, com ou sem reciclo de células, o cultivo geralmente se prolonga por 24 horas ou mais e o crescimento pode ser efetivamente inibido pelo acúmulo de ácidos orgânicos excretados pelas células ou pelo acúmulo de glicose caso a vazão de alimentação de meio suplementar não seja controlada adequadamente. Como resultado da influência de todos estes fatores, as diferentes fases de crescimento se tornariam mais longas, conteriam maior número de dados experimentais e a definição do momento de transição entre elas seria mais fácil. Conclui-se, portanto, que o maior potencial de aplicação da modelagem segregada por fases é justamente na descrição de cultivos em batelada alimentada. Infelizmente, devido a acontecimentos imprevistos, a equipe de pesquisadores do Butantan realizou, até março de 2008, apenas um cultivo em batelada alimentada, no qual um sistema de perfusão para retirada de meio de cultivo foi acoplado ao biorreator, conforme detalhado na Seção 4.2. O referido experimento, neste trabalho denominado cultivo Exp_6 , foi na realidade realizado em março de 2006, quando o programa LabView ainda estava em implementação. Por isso, não houve aquisição de dados *on-line* e até os dados experimentais *off-line* apresentaram muitas perturbações devido às dificuldades encontradas na operação manual do sistema.

Assim, a potencialidade do sistema biorreator acoplado a membrana de perfusão ainda não foi plenamente explorada, mesmo por que as condições simuladas no presente trabalho

ainda não foram verificadas experimentalmente. A associação entre reator em batelada alimentada e membrana de microfiltração se mostra como uma alternativa promissora para contornar os efeitos inibitórios de metabólitos acumulados no meio de cultivo, que constituem o maior obstáculo para a realização de cultivos de alta densidade celular e, conseqüentemente, para obtenção de concentrações elevadas do produto desejado no menor tempo possível. No entanto, devido à complexidade e custo de operação deste sistema, os valores da vazão de alimentação, de retirada de meio e da concentração de substrato no meio suplementar devem ser cuidadosamente escolhidos. A ferramenta de inferência pode ser ainda utilizada para simular outras formas de operação do sistema que prolonguem o tempo de vida útil da membrana, como operação em batelada alimentada e acionamento intermitente da bomba de retirada de permeado, apenas para manter o acúmulo de lactato abaixo da concentração crítica. Porém, tais simulações de novas formas de operação dependem da realização de vários experimentos, empregando diferentes condições de operação simuladas, de forma que o treinamento das redes neurais que constituem o comitê de classificação das fases de crescimento possa ser realizado a partir de dados obtidos na operação em batelada alimentada com reciclo celular. Da mesma forma, estes dados são necessários para a estimativa dos vetores de parâmetros na modelagem segregada por fases. Estes dois elementos essenciais da ferramenta de inferência dependem apenas da disponibilidade de dados experimentais, uma vez que tanto o comitê de classificação como a modelagem por fases já foram testados com dados de cultivos em batelada e apresentaram excelentes resultados.

Já o terceiro elemento essencial para o funcionamento da ferramenta, ou seja, os dados experimentais adquiridos em tempo real que constituirão as variáveis de entrada para o comitê classificatório, permanece a ser resolvido. Diante da impossibilidade de obtenção de dados experimentais da densidade ótica com a frequência necessária, conforme discutido na Seção 4.5, pode-se investigar a alternativa de incorporação de um comitê classificatório auxiliar à ferramenta, o qual seria alimentado pelos dados simulados, conforme detalhado também na Seção 4.5. Outra alternativa seria investir na aquisição de outra variável de entrada que seja representativa do estado metabólico das células e de possível obtenção em tempo real. Neste sentido, o sensor de capacitância que é capaz de detectar mudanças nas propriedades dielétricas da membrana celular apresenta-se como uma excelente escolha.

É importante ressaltar ainda que todo o esforço investido na concepção da ferramenta de inferência se justifica não só pela possibilidade de maximizar a produção de

polissacarídeo capsular em cultivos de *S. pneumoniae*, mas também pela possibilidade de extrapolar a mesma concepção para a otimização de cultivos de quaisquer bactérias que sejam sujeitas à inibição do crescimento pelo acúmulo de metabólitos.

SUGESTÕES PARA TRABALHOS FUTUROS

1. Melhorias na simulação/validação a serem testadas:
 - Investir na modelagem por fases;
 - Fazer mais cultivos diferenciados para obter maior quantidade de dados para treinamento de redes neurais bem como calibração de modelos;
 - Fazer a calibração de outros modelos cinéticos, utilizando experimentos diferentes;
 - Utilizar outras funções objetivo para otimizar a produtividade de PS e minimizar a glicose residual;

2. Aplicação de algoritmo de controle ótimo:
 - Pesquisar a utilização de novas medidas “on-line” para inferir as variáveis de estado. Um exemplo é o sensor de capacitância, que mede de forma “on-line” a concentração de células viáveis;
 - Construir uma ferramenta de simulação para ajudar no planejamento de novos cultivos.
 - Construir uma base de dados de parâmetros cinéticos obtidos a partir das diversas variações dos cultivos; associar esta base de dados a um algoritmo (exemplo: fuzzy) que compare as semelhanças dos cultivos e defina um conjunto de parâmetros da base de dados para simular o cultivo vigente;

REFERÊNCIAS

- Almeida, J.S. (2002) Predictive non-linear modeling of complex data by artificial neural networks. *Current Opinion in Biotechnology*, 13, 72-76.
- Aiba, S.; Humprey, A.E.; Millis, N.F. (1973) *Biochemical engineering*. 2.ed. Tóquio: University of Tokyo Press.
- Bacauskiene, M., Verikas, A. (2004) Selecting salient features for classification based on neural network committees. *Pattern Recognition Letters*, 25, 1879–1891.
- Bailey, J. E.; Ollis, D. F. (1986) *Biochemical Engineering Fundamentals*. 2 ed. McGraw-Hill do Brasil.
- Baptista Neto, A. (2004) Estudo do processo de produção de ácido clavulânico por *Streptomyces clavuligerus* em cultivos com alta concentração celular. Tese de Doutorado apresentada ao programa de Pós-graduação em Engenharia Química da Universidade federal de São Carlos.
- Beyenal, H.; Chen, A.N.; Lewandowski, Z. (2003) The double substrate growth kinetics of *Pseudomonas aeruginosa*. *Enzyme and Microbial Technology*, vol. 32, 92-98.
- Bhavnani, S. M.; Hammel, J. P.; Jones, R. N.; Ambrose, P. G. (2005) Relationship between increased levofloxacin use and decreased susceptibility of *Streptococcus pneumoniae* in the United States. *Diagnostic Microbiology and Infectious Disease*, vol. 51, 31-37.
- Brandileone, M.C.C.; Vieira, V.S.D.; Almeida, R.G.; Landgraf, I.; Buschinelli, S.; Casagrande, S.T.; Zanella, R.C. (1994) *Bacteriologia de Streptococcus pneumoniae*: manual técnico. São Paulo: Instituto Adolfo Lutz, 62.
- Bricks, L.F.; Berezin, E. (2006) Impact of pneumococcal conjugate vaccine on the prevention of invasive pneumococcal diseases. *J Pediatr (Rio J)*; 82 (3 SUPPL): S67-74 (Em Português).
- Brilles D.E.; Tart, R.C.; Swiatlo, E.; Dillard, J.P.; Smith, P.; Benton, K.A.; Ralph, B.A.; Brooks-Walter, A.; Crain, M.J.; Hollingshead, S.K.; McDaniel, L.S. (1998) Pneumococcal diversity: considerations for new vaccine strategies with emphasis on pneumococcal surface protein A (PspA), *Clinical Microbiology Reviews*, 11(4), 645-657.
- Chao, B.; Chae, S.H.; Zhang, X.; Lu, K.H., Im, J.; Ho, P.S. (2007), Investigation of diffusion and electromigration parameters for Cu–Sn intermetallic compounds in Pb-free solders using simulated annealing. *Acta Materialia*, 55 (8), 2805-2814.
- Chapuy-Regaud, S.; Duthoit, F.; Malfroy-Mastrorillo, L.; Gourdon, P.; Lindley, N.D.; Trombe, M.-C. (2001) Competence regulation by oxygen availability and Nox is not related to specific adjustment of central metabolism in *Streptococcus pneumoniae*. *J. Bacteriol.*, 138, 2957-2962.

- Costa, O. (2007) Guerra das Vacinas: Encomendas do ministério da saúde deflagram batalha entre grandes laboratórios farmacêuticos. *Isto é Dinheiro*, 522, 26/09/2007.
- Cruz, A. J. G.; Lourenço, A. A. B.; Jesus, C. D. F.; Ferreira, D. S.; Sturlini, D.; Giordano, R. C. (2006) WebLabs em Engenharia Química: Desenvolvimento, Implementação e Operação Remota de Experimentos de Transferência de Massa Via Internet. In: Congresso Brasileiro de Ensino de Engenharia, Passo Fundo, RS. XXXIV Congresso Brasileiro de Ensino de Engenharia, 1, 1-14.
- Dische, Z.; Shettles, B. (1948) A specific color reaction of methylpentoses and a spectrophotometric micromethod for their determination, *J. Biol. Chem.*, 175, 595-603.
- Eftaxias, A.; Font, J.; Fortuny, A.; Fabregat, A.; Stüber, F. (2002) Nonlinear kinetic parameter estimation using simulated annealing. *Computers and Chemical Engineering* 26, 1725-1733.
- Eilers, P.H.C. (2003) A Perfect Smoother, *Analytical Chemistry*, 75 (14), July 15, 3631-3636.
- Fonseca, P.B.B.; Braga, J.A.p.; Machado, A.M.O.; Brandileone, M.C.C.; Farhat, C.K. (2005) Colonização nasofaríngea pelo *Streptococcus pneumoniae* em crianças com doença falciforme usando penicilina profilática. *Jornal de Pediatria (Rio J.)*; 81(2), 149-154.
- Giordano, R. C.; Bertini, J. R.; Nicoletti, M. C.; Giordano, R. L. C. (2007) Online filtering of CO₂ signals from a bioreactor gas outflow using a committee of constructive neural networks, *Bioprocess and Biosystems Engineering* (accepted for publication).
- Gonçalves, V.M. (2002) Cultivo de *Streptococcus pneumoniae* sorotipo 23F: produção e purificação de seu polissacarídeo capsular para formulação de vacinas. Tese de doutorado. Universidade de São Paulo, Faculdade de Ciências Farmacêuticas, Programa de Pós-Graduação em Tecnologia Bioquímico-Farmacêutica.
- Gonçalves, V.M.; Takagi, M.; Carneiro, S.M.; Giordano, R.C.; Tanizaki, M.M. (2006) Introduction of air in the anaerobic culture of *Streptococcus pneumoniae* serotype 23F induces the release of capsular polysaccharide from bacterial surface into the cultivation medium. *Journal of Applied Microbiology*, 101, 1009–1014.
- Gonçalves, V.M.; Zangirolami, T.C.; Giordano, R.L.C.; Raw, I.; Tanizaki, M.M.; Giordano, R.C. (2002) Optimization of medium and cultivation conditions for capsular polysaccharide production by *Streptococcus pneumoniae* serotype 23F, *Appl. Microbiol. Biotechnol.* 59, 713-717.
- Haykin, S. (2001) *Neural Networks: A comprehensive foundation*; 2nd ed. India, Pearson Education.
- Henriques, A.W.S.; Jessouroun, E.; Lima, E.L.; Alves, T.L.M. (2005) Mathematical Modeling of Capsular Polysaccharide Production by *Neisseria meningitidis* serogroup C in Bioreactors. *Brazil Journal of Chemical Engineering*, 22 (04), 585-592.

- Hodgson, B. J.; Taylor, C. N.; Ushio, M.; Leigh, J.R.; Kalganova, T.; Baganz, F. (2004) Intelligent modelling of bioprocesses: a comparison of structured and unstructured approaches. *Bioprocess Biosyst Eng.* 26, 353-359.
- Hoepflich, P.D. (1955) Carbon-14 labeling of *Diplococcus pneumoniae*. *J. Bacteriol.*, 69, 682-688.
- Honda, H.; Kobayashi, T. (2000) Fuzzy Control of Bioprocess. *Journal of Bioscience and Bioengineering*, 89 (5), 401-408.
- Horta, A.C.L.; Carmo, T.S.; Zangirolami, T.C.; Gonçalves, V.M.; Giordano, R.C.; Tanikazi, M.M.; Nicoletti, M. do C. (2007) Modelagem e Simulação do Cultivo de *S. pneumoniae* em Biorreator Acoplado a Sistema de Perfusão: Identificação das Condições Ótimas de Operação. *Anais do SINAIFERM 2007*.
- Jesus, C. D. F.; Giordano, R. C.; Cruz, A. J. G.; Alliet-Gaubert, M.; Joulia, X.; Rouzineau, D.; Albet, J.; Coufort, C; Roux, G. A. C. L. (2007) Weblab in Chemical Engineering between France and Brazil: validation of the methodology. In: *International Conference on Engineering Education, Coimbra, Portugal. International Conference on Engineering Education ICEE 2007*, 1, 1-6.
- Kim, Jo, Weiser, Jn. (1998) Association of intrastain phase variation in quantity of capsular polysaccharide and teichoic acid with the virulence of *Streptococcus pneumoniae*. *J. Infect. Dis.*, 177, 368-377.
- Klein, D.L.; Ellis, R.W. (1997) Conjugate vaccines against *Streptococcus pneumoniae*. In - *New generation vaccines 2nded*, ed Levine, M.M., Woodrow, G.C., Kaper, J.B, Cobon, G.S. Marcel Dekker, New York, 503-525.
- Koneman, E. W.; Allen, S. D.; Janda, W. M.; Schreckenberger, P. C.; Winn, W. C., Jr. (2001) *Diagnóstico Microbiológico - Texto e Atlas Colorido – 5ª edição, Tradução e Supervisão Arlete Emily Cury, Guanabara Koogan S.A., Rio de Janeiro, RJ.*
- Kovárová-Kovar, K.; Egli, T. (1998) Growth kinetics of suspended microbial cells: from single-substrate-controlled growth to mixed-substrate kinetics, *Microbiol. Mol. Biol.*, 62, 646-666.
- Levenspiel, O. (1974) *Engenharia das reações químicas. Volume 1. Tradução: Sérgio Fucks Calil e Pedro Maurício Büchler. São Paulo, Edgard Blücher.*
- Lima, U.A.; Aquarone, E.; Borzani, W.; Schmidell, W. (2001) *Biotechnology Industrial: Processos Fermentativos e Enzimáticos, Volume 3. Edgard Blücher.*
- Linko, S., Zhu, Y.H., Linko, P. (1999) Applying neural networks as software sensors for enzyme engineering. *Elsevier Science*, 17, 155-162.

- Ma, C.; Jiang, L. (2007) Some research on Levenberg–Marquardt method for the nonlinear equations. *Applied Mathematics and Computation*, 184, 1032–1040.
- Madigan, M. T.; Martinko, J. M.; Parker, J. (2004) *Microbiologia de Brock*. Tradução de revisão técnica Cynthia Maria Kiaw, São Paulo, Prentice Hall.
- Mendes, P.; Kell, D.B. (1998) Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation. *Bioinformatics*, 14 (10), 869-883.
- Mitchell, T. M. (1997) *Machine Learning*, McGraw-Hill Science.
- Moles, C.G.; Mendes, P.; Banga, J. R. (2003) Parameter Estimation Pathways: A Comparison of Global Optimization Methods. *Genome Research*, 13, 2467-2474.
- Nelles, O. (2001) *Nonlinear system identification: from classical approaches to neural networks and fuzzy models*. Germany, Springer.
- Nicoletti, M. C.; Bertini, J. R. (2007) An empirical evaluation of constructive neural network algorithms in classification tasks, *International Journal of Innovative Computing and Applications*, 1 (1), 2-13.
- Nielsen J., Villadsen, J. (1994) *Bioreaction Engineering Principles*. Plenum Press, New York.
- Palma Neto, L. G.; Nicoletti, M. C. (2005) *Introdução às Redes Neurais Construtivas*. São Carlos, EdUFSCar.
- Park, T.Y.; Froment, G.F. (1998) A Hybrid Genetic Algorithm for the Estimation of Parameters in Detailed Kinetic Models. *Computers chem*, 22, S103-S110.
- Patnaik, P.R. (2002) Can imperfections help to improve bioreactor performance? *TRENDS in Biotechnology*, 20 (4), April, 135-137.
- Patnaik, P.R. (2001-a) A simulation study of dynamic neural filtering and control of a fed-batch bioreactor under nonideal conditions. *Chem. Eng. J.* 84, 533-541.
- Patnaik, P.R. (2001-b) Enhancement of protein activity in a recombinant fermentation by optimising fluid dispersion and initial plasmid copy number dispersion. *Biochem. Eng. J.* 9, 111-118.
- Patnaik, P.R. (2001-c) Further enhancement of fed-batch streptokinase yield in the presence of inflow noise by coupled neural networks. *Process Biochem.* 37, 145-151.
- Patnaik, P.R. (1999-a) Coupling of a neural filter and a neural controller for improvement of fermentation performance. *Biotechnol. Techniques* 13, 735-738.
- Patnaik, P.R. (1999-b) Improvement of the microbial production of streptokinase by controlled filtering of process noise. *Process Biochem.* 35, 309-315.

- Peres, J., Oliveira, R., Feyer de Azevedo, S. (2007) Bioprocess hybrid parametric/nonparametric modeling based on the concept of mixture of experts. *Biochemical Engineering Journal*, 39, 190-206.
- Rodriguez-Fernandez, M., Mendes, P., Banga, J.R. (2006) A hybrid approach for efficient and robust parameter estimation in biochemical pathways. *BioSystems*, 83, 248-265.
- Shuler, M.L.; Kargi, F. (2001) *Bioprocess Engineering: Basic Concepts*. Prentice Hall PTR, 2nd ed.
- Silva, R.G. (2003) Inferência de Variáveis do Processo de Produção de Penicilina G Acilase por *Bacillus megaterium* ATCC 14945. Tese de doutorado apresentada ao programa de pós-graduação em engenharia química da Universidade Federal de São Carlos.
- Simon, L., Karim, M.N., Schreiweis, A. (1998) Prediction and classification of different phases in a fermentation using neural networks. *Biotechnology Techniques*, 12 (4), 301–304.
- Simon, L.; Karim, M.N. (2001) Probabilistic neural networks using Bayesian decision strategies and modified Gomperts model for growth phase classification in the batch culture of *Bacillus subtilis*. *Biochemical Engineering Journal* 7, 41-48.
- Takagi, M.; Cabrera-Crespo, J.; Baruque-Ramos, J.; Zangirolami, T.C.; Raw, I.; Tanikazi, M.M. (2003) Characterization of polysaccharide production of *Haemophilus influenzae* type b and relationship to bacterial cell growth, *Applied Biochemistry and Biotechnology* (110), 91-100.
- Tschichold-Gürman, N.N. (1995) A knowledge-based neural network model with application examples in mobile robotics. Ph.D. (Thesis) – ETH No. 11356, Swiss Federal Institute of Technology, Zürich.
- Yurt, N.; Beyenal, H.; Sears, J.; Lewandowski, Z. (2003) Quantifying selected growth parameters of *Leptothrix discophora* SP-6 in biofilms from oxygen concentration profiles. *Chemical Engineering Science* 58, 4557-4566.