

UNIVERSIDADE ESTADUAL DE CAMPINAS

FACULDADE DE ENGENHARIA QUÍMICA

ÁREA DE CONCENTRAÇÃO

DESENVOLVIMENTO DE PROCESSOS QUÍMICOS

Desenvolvimento de Modelos Híbrido-Neurais para Fermentação

Alcoólica e Estudo de Técnicas de Otimização do Processo

Autor: **Elver Radke** (*Engenheiro Químico, Universidade Estadual de Maringá, 1999*)

Orientador: **Prof. Dr. Rubens Maciel Filho**

Co-Orientador: **Dra. Aline Carvalho da Costa**

Dissertação de Mestrado
apresentada à Faculdade de Engenharia
Química como parte dos requisitos
exigidos para a obtenção do título de
Mestre em Engenharia Química.

Campinas – São Paulo

Novembro – 2002

UNICAMP
BIBLIOTECA CENTRAL
SEÇÃO CIRCULANTE

200336534

UNIDADE	BC
Nº CHAMADA	T/UNICAMP
	R119d
V	EX
TOMBO BCI	56590
PROC.	16-124/03
C	<input type="checkbox"/>
D	<input checked="" type="checkbox"/>
PREÇO	R\$ 11,00
DATA	
Nº CPD	

ii

CM00192860-9

Bib id 307087

FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DA ÁREA DE ENGENHARIA - BAE - UNICAMP

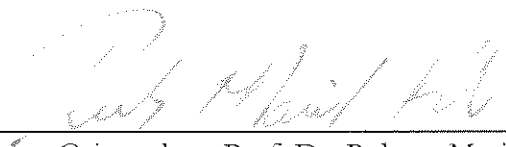
R119d Radke, Elver
Desenvolvimento de modelos híbrido-neuronais para fermentação alcoólica e estudo de técnicas de otimização do processo / Elver Radke.--Campinas, SP: [s.n.], 2002.

Orientadores: Rubens Maciel Filho; Aline Carvalho da Costa

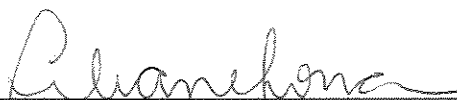
Dissertação (mestrado) - Universidade Estadual de Campinas, Faculdade de Engenharia Química

1. Fermentação. 2. Redes neurais (Computação). 3. Programação quadrática. 4. Simulação por computador. 5. Modelagem de dados. I. Maciel Filho, Rubens. II. Costa Aline Carvalho. III. Universidade Estadual de Campinas. Faculdade de Engenharia Química. IV. Título.

Dissertação de Mestrado defendida por Elver Radke e aprovada em 29-11-2002 pela banca examinadora constituída pelos doutores:



Orientador - Prof. Dr. Rubens Maciel Filho

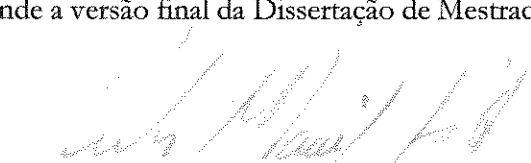


Profª. Dra. Liliane Maria Ferrareso Lona



Prof. Dr. Francisco Maugeri Filho

Este exemplar corresponde a versão final da Dissertação de Mestrado em Engenharia Química.



Orientador – Prof. Dr. Rubens Maciel Filho

Dedico,

À minha querida esposa Vanessa,

Meus pais Ervino e Leonil,

Minha irmã Ruby.

AGRADECIMENTOS

- ✓ Agradeço eternamente a Deus por sua infinita graça e misericórdia revelada a nós através de Seu filho Cristo Jesus, sem o qual não teríamos a esperança da vida eterna;

Porque dEle, e por Ele, e para Ele, são todas as coisas; glória, pois, a Ele eternamente. Amém.

(Romanos 11:36)

- ✓ A minha amada Vanessa pelo apoio, muitas vezes com lágrimas de saudades, pelo amor, carinho e paciência;
- ✓ Aos meus pais por serem pais, educadores, companheiros e estarem disponíveis tantas vezes quanto eu precisei;
- ✓ A minha irmã pelo incentivo e cartas carinhosas;
- ✓ Aos meus orientadores professor Dr. Rubens Maciel Filho e Dra. Aline Carvalho da Costa pela compreensão, disponibilidade, paciência, auxílio e conhecimento;
- ✓ Aos meus sogros Luiz e Vera pelo apoio constante;
- ✓ Ao pessoal da república “Route 66”, Denis, Gilzão, Kleber e Kalil;
- ✓ A família Berbert e todo o pessoal da 2ª IPI de Campinas que me acolheram tão carinhosamente quanto os irmãos da 3ª IPB de Maringá que ficaram torcendo por mim;
- ✓ Aos amigos do mestrado: Vanessa Ito, Wesley, Edvaldo, Leonardo, Maria, Edson Fugita e Rose (in memoriam);
- ✓ A todo o pessoal do Laboratório de Otimização, Projeto e Controle Avançado, LOPCA pela amizade, convívio e dicas;
- ✓ A Fundação de Amparo à Pesquisa do Estado de São Paulo, FAPESP, pelo suporte financeiro.

EPÍGRAFE

“O temor do Senhor é o princípio sabedoria; e o
conhecimento do Santo é o entendimento”.

(Provérbios 9:10)

“Tudo deveria se tornar o mais simples possível,
mas não simplificado”.

(Albert Einstein)

“A capacidade de *aprender* através de exemplos e de
generalizar a informação aprendida é, sem dúvida, o
atrativo principal da solução de problemas através de
Redes Neurais Artificiais”.

(BRAGA et al., 2000)

SUMÁRIO

Sumário viii

Lista de Figuras xi

Lista de Tabelas xiv

Nomenclatura xv

Resumo xvi

1	INTRODUÇÃO	1
1.1	OBJETIVOS DO TRABALHO.....	3
2	REVISÃO BIBLIOGRÁFICA.....	5
2.1	PANORAMA DO ÁLCOOL COMBUSTÍVEL NO BRASIL.....	5
2.2	PRODUÇÃO DE ETANOL.....	7
2.2.1	Microorganismos Produtores de Etanol	8
2.2.2	Processos Fermentativos para a Produção de Etanol	10
2.3	MODELAGEM MATEMÁTICA.....	11
2.4	REDES NEURAIS ARTIFICIAIS	17
2.4.1	O Sistema nervoso e o Neurônio.....	17
2.4.2	Histórico	19
2.4.3	Algoritmos de Treinamento	22
2.4.4	O Algoritmo Back-Propagation.....	26
2.4.5	Tipos e Topologias de Redes.....	29
2.4.6	Minimização de Erros	34
2.5	MODELAGEM HÍBRIDA COM REDES NEURAIS ARTIFICIAIS	36
2.6	OTIMIZAÇÃO.....	39
2.6.1	Planejamento Fatorial Completo 2^k	40
2.6.2	Método das Superfícies de Resposta (RSM).....	41
2.6.3	Programação Quadrática Sucessiva	43

3	O PROCESSO FERMENTATIVO E A MODELAGEM MATEMÁTICA....	46
3.1	TRABALHO EXPERIMENTAL (COSTA , 2000).....	46
3.1.1	Material.....	46
3.2	MODELAGEM MATEMÁTICA E PARÂMETROS CINÉTICOS	47
4	REDES NEURAIS ARTIFICIAIS, DESENVOLVIMENTO E APLICAÇÃO	50
4.1	ORGANIZAÇÃO E ANÁLISE DOS DADOS EXPERIMENTAIS.....	50
4.2	MODELAGEM COM REDES NEURAIS ARTIFICIAIS.....	64
4.2.1	Normalização, Homogeneização e Conjunto de Validação.....	64
4.2.2	Algoritmos de Treinamento e Arquitetura da Rede.....	69
4.2.3	Modelo Híbrido com Variáveis Primárias	73
4.2.4	Modelo Híbrido com Variáveis Secundárias (Software Sensor)	88
5	ESTUDO DE TÉCNICAS DE OTIMIZAÇÃO USANDO MODELOS GERADOS POR PLANEJAMENTO FATORIAL.....	105
5.1	ESTUDO PARA IDENTIFICAÇÃO DAS VARIÁVEIS RELEVANTES.....	105
5.2	DETERMINAÇÃO DE UM MODELO EMPÍRICO USANDO PLANEJAMENTO FATORIAL.....	109
5.3	METODOLOGIA DAS SUPERFÍCIES DE RESPOSTA.....	112
5.4	PROGRAMAÇÃO QUADRÁTICA SUCESSIVA.....	116
6	CONCLUSÕES.....	119
6.1	AJUSTE DOS DADOS EXPERIMENTAIS E TREINAMENTO DAS RNAs ...	119
6.2	MODELO HÍBRIDO COM VARIÁVEIS PRIMÁRIAS.....	120
6.3	MODELO HÍBRIDO COM VARIÁVEIS SECUNDÁRIAS	120
6.4	OTIMIZAÇÃO.....	121
6.5	SUGESTÕES	121
7	BIBLIOGRAFIA	122
8	APÊNDICES E ANEXOS	129
8.1	MÉTODOS ANALÍTICOS.....	129
8.1.1	Concentração de células totais e viáveis:	129
8.1.2	Viabilidade	129

8.1.3	Concentração de açúcares redutores totais (ART) e etanol	130
8.2	TABELAS	130
8.3	ALGORITMO BACKPROPAGATION	132
8.4	ESTATÍSTICA.....	133

LISTA DE FIGURAS

Figura 2.1 – Ascensão e queda do álcool. A produção nacional de veículos a álcool em 06/1999 representava 0,5%.....	5
Figura 2.2 - Constituintes da célula neural – Esquema	18
Figura 2.3 - Soma de sinal no neurônio artificial, modelo tipo monocamada Perceptron.....	19
Figura 2.4 - Aprendizado Supervisionado	23
Figura 2.5 - Aprendizado não-supervisionado	24
Figura 2.6 - Aprendizado por Reforço.....	25
Figura 2.7 - Ilustração da função erro $E(w)$ vista como uma superfície sobre a área dos pesos....	27
Figura 2.8 - Exemplos de Arquiteturas de RNAs.....	30
Figura 2.9 - Rede monocamada tendo d entradas x_1, \dots, x_d e m saídas z_1, \dots, z_m	32
Figura 2.10 - Rede MLP tendo duas camadas de pesos.....	33
Figura 2.11 - Algumas Funções de Ativação.....	34
Figura 2.12 - a) Planejamento Fatorial de dois níveis com ponto central. b) Planejamento em Estrela. Cada ponto representa um ensaio.....	41
Figura 4.1 - Comparação entre os diversos ajustes, para as concentrações de a) Biomassa; b) Substrato e c) Produto. Temperatura de 28°C	54
Figura 4.2 – Comparação entre os diversos ajustes, para as concentrações de a) Biomassa; b) Substrato e c) Produto. Temperatura de 31°C	55
Figura 4.3 – Comparação entre os diversos ajustes, para as concentrações de a) Biomassa; b) Substrato e c) Produto. Temperatura de 34°C	56
Figura 4.4 - Comparação entre os diversos ajustes, para as concentrações de a) Biomassa; b) Substrato e c) Produto. Temperatura de 37°C	57
Figura 4.5 - Comparação entre os diversos ajustes, para as concentrações de a) Biomassa; b) Substrato e c) Produto. Temperatura de 40°C	58
Figura 4.6 – Comparação entre os modelos ajustados para as taxas cinéticas. a) Crescimento de biomassa; b) Consumo de substrato; c) Formação de produto. Temperatura 28° C ...	59
Figura 4.7 – Comparação entre os modelos ajustados para as taxas cinéticas. a) Crescimento de biomassa; b) Consumo de substrato; c) Formação de produto. Temperatura 31° C ...	60
Figura 4.8 – Comparação entre os modelos ajustados para as taxas cinéticas. a) Crescimento de biomassa; b) Consumo de substrato; c) Formação de produto. Temperatura 34° C ...	61
Figura 4.9 – Comparação entre os modelos ajustados para as taxas cinéticas. a) Crescimento de biomassa; b) Consumo de substrato; c) Formação de produto. Temperatura 37° C ...	62
Figura 4.10 – Comparação entre os modelos ajustados para as taxas cinéticas. a) Crescimento de biomassa; b) Consumo de substrato; c) Formação de produto. Temperatura 40° C ...	63

Figura 4.11 - Funções de ativação	65
Figura 4.12 - Comparação entre as funções de ativação na camada intermediária. À saída a função purelin foi empregada. Algoritmo de aprendizado: Levenberg-Marquardt. Arquitetura Feedforward	66
Figura 4.13 - Comparação entre as funções de ativação na camada intermediária. Arquitetura: Feedforward. Algoritmo: Levenberg-Marquardt. 5 nodos na camada intermediária e 1 na de saída.....	67
Figura 4.14 - Relação entre o erro quadrático com o número de nodos na camada de saída. Arquitetura: Feedforward. Algoritmo: Levenberg-Marquardt.....	68
Figura 4.15 - Exemplo de uma rede Feedforward, completamente conectada, com duas camadas intermediárias	70
Figura 4.16 - Exemplo de superfície de erro e passos do gradiente descendente.....	71
Figura 4.17 - Fluxograma do modelo híbrido	74
Figura 4.18 - Resultados do treinamento da rede para a Taxa de Crescimento de Biomassa e aplicação do Modelo Híbrido para Concentração de Biomassa. Temperatura de: 34°C. Conjunto de Treinamento TIPO I	76
Figura 4.19 - Resultados do treinamento da rede para a Taxa de Consumo de Substrato e aplicação do Modelo Híbrido para Concentração de Substrato. Temperatura de: 34°C. Conjunto de Treinamento TIPO I	77
Figura 4.20 - Resultados do treinamento da rede para a Taxa de Formação de Produto e aplicação do Modelo Híbrido para Concentração de Produto. Temperatura de: 34°C. Conjunto de Treinamento TIPO I	78
Figura 4.21 - Resultados do treinamento da rede para a Taxa de Crescimento de Biomassa e aplicação do Modelo Híbrido para Concentração de Biomassa. Temperatura de: 34°C. Conjunto de Treinamento TIPO II.....	79
Figura 4.22 - Resultados do treinamento da rede para a Taxa de Consumo de Substrato e aplicação do Modelo Híbrido para Concentração de Substrato. Temperatura de: 34°C. Conjunto de Treinamento TIPO II.....	80
Figura 4.23 - Resultados do treinamento da rede para a Taxa de Formação de Produto e aplicação do Modelo Híbrido para Concentração de Produto. Temperatura de: 34°C. Conjunto de Treinamento TIPO II.....	81
Figura 4.24 – Comparação entre os modelos Híbrido, determinístico (COSTA, 2000) e os dados experimentais reais. Temperatura 28°C.....	83
Figura 4.25 – Comparação entre os modelos Híbrido, determinístico (COSTA, 2000) e os dados experimentais reais. Temperatura 31°C.....	84
Figura 4.26 – Comparação entre os modelos Híbrido, determinístico (COSTA, 2000) e os dados experimentais reais. Temperatura 34°C.....	85
Figura 4.27 – Comparação entre os modelos Híbrido, determinístico (COSTA, 2000) e os dados experimentais reais. Temperatura 37°C.....	86

Figura 4.28 – Comparação entre os modelos Híbrido, determinístico (COSTA, 2000) e os dados experimentais reais. Temperatura 40°C.....	87
Figura 4.29 - Fluxograma do modelo híbrido com variáveis secundárias na entrada. (tz : turbidez; Bx : Brix; pH : potencial de Hidrogênio).	88
Figura 4.30 - Ajustes para as variáveis secundárias de a) pH; b) Turbidez e c) Brix. Temperatura de 28°C.....	90
Figura 4.31 - Ajustes para as variáveis secundárias de a) pH; b) Turbidez e c) Brix. Temperatura de 31°C.....	91
Figura 4.32 - Ajustes para as variáveis secundárias de a) pH; b) Turbidez e c) Brix. Temperatura de 34°C.....	92
Figura 4.33 - Ajustes, para as variáveis secundárias de a) pH; b) Turbidez e c) Brix. Temperatura de 37°C.....	93
Figura 4.34 - Ajustes, para as variáveis secundárias de a) pH; b) Turbidez e c) Brix. Temperatura de 40°C.....	94
Figura 4.35 - Resultados do treinamento da rede para a Taxa de Crescimento de Biomassa e aplicação do Modelo Híbrido com variáveis secundárias para Concentração de Biomassa. Temperatura de: 34°C.	96
Figura 4.36 - Resultados do treinamento da rede para a Taxa de Consumo de Substrato e aplicação do Modelo Híbrido com variáveis secundárias para Concentração de Substrato. Temperatura de: 34°C.	97
Figura 4.37 - Resultados do treinamento da rede para a Taxa de Formação de Produto e aplicação do Modelo Híbrido com variáveis secundárias para Concentração de Produto. Temperatura de: 34°C.	98
Figura 4.38 – Comparação entre os modelos Híbrido com variáveis secundárias, ajuste aos dados experimentais com a Equação Sigmoidal-Boltzman, Determinístico COSTA (2000) e dados experimentais. Temperatura de 28°C.....	100
Figura 4.39 – Comparação entre os modelos Híbrido com variáveis secundárias, ajuste aos dados experimentais com a Equação Sigmoidal-Boltzman, Determinístico COSTA (2000) e dados experimentais. Temperatura de 31°C.....	101
Figura 4.40 – Comparação entre os modelos Híbrido com variáveis secundárias, ajuste aos dados experimentais com a Equação Sigmoidal-Boltzman, Determinístico COSTA (2000) e dados experimentais. Temperatura de 34°C.....	102
Figura 4.41 – Comparação entre os modelos Híbrido com variáveis secundárias, ajuste aos dados experimentais com a Equação Sigmoidal-Boltzman, Determinístico COSTA (2000) e dados experimentais. Temperatura de 37°C.....	103
Figura 4.42 – Comparação entre os modelos Híbrido com variáveis secundárias, ajuste aos dados experimentais com a Equação Sigmoidal-Boltzman, Determinístico COSTA (2000) e dados experimentais. Temperatura de 40°C.....	104
Figura 5.1 - Histograma dos efeitos principais das variáveis sobre a concentração de produto final.....	108

Figura 5.2 – Superfície de resposta e diagrama de contorno para a concentração de produto final (PROD, g/l) em função da concentração inicial de substrato (SUBS, g/l) e temperatura (TEMP, °C).	112
Figura 5.3 – Superfície de resposta e diagrama de contorno para o tempo de fermentação (TEMPO, horas) em função da concentração inicial de substrato (SUBS, g/l) e temperatura (TEMP, °C).	114
Figura 5.4 – Superposição dos diagramas de contorno das Figuras 5.2 e 5.3	115

LISTA DE TABELAS

Tabela 2.1- Rumo a auto-suficiência, capacidade de produção nacional de petróleo em relação ao seu consumo.....	6
Tabela 2.2 Quantidade de etanol utilizada na formulação de alguns combustíveis	7
Tabela 2.3- Alguns microrganismos que produzem etanol a partir de carboidratos.....	8
Tabela 3.1 - Parâmetros em função da Temperatura em °C.....	48
Tabela 4.1 - Polinômios ajustados para os dados de cada temperatura e suas derivadas.....	51
Tabela 4.2 - Parâmetros ajustados para as concentrações monitoradas experimentalmente, em cada temperatura, para a equação Sigmoidal-Boltzman	53
Tabela 4.3 – Parâmetros ajustados para as variáveis secundárias monitoradas experimentalmente, em cada temperatura, para a equação Sigmoidal-Boltzman.....	89
Tabela 5.1 – Variáveis e Níveis para o Planejamento Fatorial.	106
Tabela 5.2 – Primeiro Planejamento Fatorial	107
Tabela 5.3 – Efeitos principais para o Primeiro planejamento Fatorial	107
Tabela 5.4 – Variáveis e níveis para o planejamento fatorial completo	110
Tabela 5.5 – Planejamento Experimental, fatorial completo. Segundo planejamento	110
Tabela 5.6 – Análise da Variância (ANOVA) para o ajuste de um modelo quadrático de duas variáveis, sem repetição no ponto central	111
Tabela 8.1 - Dados experimentais do processo batelada, variáveis de estado e variáveis secundárias para a temperatura de 28°C	130
Tabela 8.2 - Dados experimentais do processo batelada, variáveis de estado e variáveis secundárias para a temperatura de 31°C	130
Tabela 8.3 - Dados experimentais do processo batelada, variáveis de estado e variáveis secundárias para a temperatura de 34°C	131
Tabela 8.4 - Dados experimentais do processo batelada, variáveis de estado e variáveis secundárias para a temperatura de 37°C	131
Tabela 8.5 - Dados experimentais do processo batelada, variáveis de estado e variáveis secundárias para a temperatura de 40°C	131

NOMENCLATURA

μ_{\max}	Taxa máxima de crescimento (h^{-1})
d_i	Saída desejada para o treinamento da rede neural
EQ	Erro quadrático
K_i	Coefficiente de inibição do Substrato (l/g)
K_s	Constante de saturação do Substrato (g/l)
m	Constante na equação 3.1
m_p	Produção de etanol associado ao crescimento (g/[g.h])
m_x	Coefficiente de manutenção (g/[g.h])
n	Constante na equação 3.1 ou quantidade de amostras
P	Concentração de Produto (g/l)
P_{\max}	Concentração de Produto quando o crescimento de célula cessa (g/l)
r_p	Taxa cinética de formação de Produto (g/l.h)
r_s	Taxa cinética de consumo de Substrato (g/l.h)
r_x	Taxa cinética de crescimento de Biomassa (g/l.h)
S	Concentração de Substrato (g/l)
X	Concentração de Biomassa (g/l)
X_{\max}	Concentração de Biomassa quando o crescimento de célula cessa (g/l)
y_i	Saída obtida no treinamento da rede neural
Y_{px}	Rendimento de produto baseado no crescimento celular (g/g)
Y_x	Rendimento Celular limite (g/g)

RESUMO

Neste trabalho, pretende-se estudar a modelagem híbrido neural de um processo de fermentação alcoólica. O objetivo é desenvolver de forma simples e rápida modelos que sejam capazes de descrever as características mais representativas do processo, podendo ser usados para otimização, controle ou como *software sensors*. Para isso, combinam-se as equações de balanço de massa do processo com redes neurais, que descrevem a cinética desconhecida. São estudadas medidas secundárias, como turbidez, pH e Brix como entradas para a rede neural artificial. São usados dados experimentais por COSTA (2000) e ATALA (2000). Os resultados referentes à organização e análise dos dados experimentais e estudo das rede neurais do tipo *feedforward*, bem como o desenvolvimento do modelo híbrido usando variáveis primárias e secundárias, são apresentados.

Neste trabalho, são estudadas técnicas de otimização do processo de fermentação alcoólica. Com este objetivo os conceitos de planejamento fatorial são usados aplicados à simulação para determinar faixas operacionais adequadas e um modelo do processo. Este modelo, juntamente com o modelo determinístico detalhado, são usados para otimização empregando-se a metodologia das superfícies de resposta e programação quadrática sucessiva. Os resultados são comparados.

Palavras-chave: Redes Neurais Artificiais, Fermentação Alcoólica, Modelagem Híbrida, Planejamento Fatorial.

ABSTRACT

In this work the neural hybrid modeling of an alcoholic fermentation process is studied. The objective is the simple and fast development of models capable to describe the most representative features of the process, being able to be used for optimization, control or as software sensors. For this, the mass balance equations of the process are combined with neural networks, which describe the complex kinetic. Secondary measures as turbidity, pH and Brix are also studied as input data for the artificial neural networks. From an experimental study, COSTA (2000) and ATALA (2000), the necessary data were obtained. The referring results to the organization and analysis of the experimental data and feedforward neural network study, as well as the hybrid model development using primary and secondary variables, are presented.

Techniques of alcoholic fermentation process optimization are studied. With this objective the concepts of factorial design are applied to the simulation to determine adequate operating bands and a process model. This model, with the detailed deterministic model, is used for optimization using the response surface methodology and successive quadratic programming. The results are compared.

Keywords: Artificial Neural Networks, Alcoholic Fermentation, Hybrid Models, Factorial Design.

1 INTRODUÇÃO

A partir de 1975, devido a um aumento do preço internacional do barril de petróleo, o governo brasileiro iniciou o programa Pró-álcool, dando incentivos ao desenvolvimento tecnológico para viabilizar a produção de etanol como combustível, tentando desta forma tornar o país independente do petróleo.

Entretanto, como o país durante a década de 80 passou a extrair uma quantidade maior de petróleo, e também como no mercado internacional o preço do barril teve uma queda significativa, o governo brasileiro ao final daquela mesma década passou a retirar os incentivos. Considerando que na época a produção de veículos a álcool somava praticamente 90% da produção total de veículos no país, isto proporcionou uma crise neste setor, levando o mesmo a uma falta de credibilidade pelos consumidores. Este foi o panorama do álcool no Brasil durante praticamente toda a década de 90.

Em 1999 uma nova alta do preço internacional do barril de petróleo, saltando de US\$ 10,00 para US\$ 25,00 juntamente com o ajuste cambial e os ganhos de produtividade, tornaram o custo de produção do álcool menor do que o preço da gasolina (FURTADO, 1999), o que renovou as perspectivas para este setor. Para que o etanol seja um produto economicamente competitivo como a gasolina, no entanto, é necessário que se continue estudando maneiras de minimizar o seu custo de produção.

Entre as várias formas para se atingir este objetivo, existe a possibilidade da implementação de sistemas de controle eficientes nas usinas. No entanto, um dos desafios no controle e otimização de bioprocessos é a falta de sensores capazes de medir as variáveis importantes do processo de maneira confiável, em tempo real e com baixos custos. Uma solução para este problema é o uso de *software sensors*, que são algoritmos desenvolvidos para estimar as variáveis de estado.

Para que seja possível estimar tais variáveis, no entanto, é necessária a determinação de modelos realmente capazes de descrever o processo com alto grau de confiabilidade. Porém, a modelagem de bioprocessos é difícil devido à natureza das reações bioquímicas envolvidas, dados apresentando ruídos ou até mesmo não disponíveis em quantidade suficiente, além de existir a dificuldade em se obter os dados do processo em tempo real (LINKO & ZHU, 1991).

Como uma alternativa para solucionar esta dificuldade na modelagem, a literatura cita o uso de redes neurais artificiais (RNA) (COSTA et al, 1999). As RNAs geram, a partir de dados de entrada e saída, relações que serão responsáveis pelo modelo interno da rede. Uma vez gerado este modelo, treinada a rede, a mesma pode ser utilizada para prever as saídas correspondentes a novos dados de entrada. O modelo interno desenvolvido pela rede, entretanto, não é baseado em nenhum mecanismo do processo. Além disso, as redes neurais podem:

- Ser retreinadas facilmente por ocasião da mudança das condições de operação do processo;
- Lidar melhor com ruídos normalmente presentes nos dados.

Uma RNA também pode ser adaptada para trabalhar em um modelo híbrido, principalmente em situações onde são desconhecidos alguns fatores essenciais, como por exemplo, em bioprocessos, a cinética. Desta forma, em modelos híbrido neurais os aspectos do problema cujo comportamento quantitativo é bem conhecido são descritos por equações matemáticas determinísticas (equações de balanço), enquanto as RNAs representam a cinética. As vantagens deste sistema híbrido, quando comparados com as redes neurais puras, são:

- São interpretados e analisados mais facilmente;
- Necessitam menos dados para o treinamento das redes;
- São desenvolvidos mais rapidamente;
- Interpolam e extrapolam melhor;
- A incerteza é embutida somente em uma parte do modelo;
- Facilmente usados de forma adaptativa, já que o esforço para reestimação da rede é menor do que quando modelos determinísticos são utilizados.

As redes neurais artificiais têm se mostrado muito versáteis e eficientes em várias aplicações. De acordo com um exemplo de uma refinaria no Reino Unido, um sistema utilizando RNA chegou a economizar tanto quanto outro sistema com controle preditivo (*Predictive Integral* (PI)), no entanto, enquanto a implementação do sistema PI levou dois anos, a rede neural foi desenvolvida e implementada em duas semanas (PARKINSON & FOUHY, 1995).

1.1 Objetivos do Trabalho

Este trabalho tem como objetivo, de acordo com o que foi apresentado, o desenvolvimento de modelos híbridos neurais para o processo de fermentação alcoólica. Os modelos desenvolvidos podem ser usados como *software sensors* ou para descrever o comportamento dinâmico do processo.

Os dados necessários a este desenvolvimento foram obtidos através de experimentos realizados por COSTA (2000) e ATALA (2000). Tais experimentos avaliaram a influência da temperatura na cinética do processo de fermentação alcoólica. As variáveis medidas foram: concentrações de Biomassa, Substrato e Produto; além das variáveis secundárias como pH, turbidez e Brix. Estuda-se, portanto, a possibilidade da utilização destas medidas no desenvolvimento de um modelo híbrido.

O estudo da otimização do processo a partir de modelos gerados pelo planejamento fatorial também é proposto. Os resultados da otimização são comparados usando as metodologias das superfícies de resposta e a programação quadrática sucessiva (EDGAR & HIMMELBLAU, 1989).

O trabalho foi dividido nas seguintes etapas:

- Organização e análise dos dados disponíveis, para determinar quais informações devem ser usadas no desenvolvimento do modelo híbrido;
- Estudo das redes *feedforward*, que são usadas junto com as equações determinísticas para compor o modelo híbrido neural;
- Considerando todos os dados disponíveis para o treinamento, estudo da melhor estrutura para o modelo híbrido neural;
- Determinação do modelo híbrido neural para a fermentação alcoólica e estudo do seu desempenho para descrever a dinâmica do processo. O seu desempenho é comparado ao do modelo determinístico de COSTA (2000) e ATALA (2000);
- Determinação de um modelo quadrático da concentração de produto final da batelada em função da concentração inicial de substrato e da temperatura usando-se os conceitos de planejamento fatorial. A concentração final de

produto é otimizada usando-se a metodologia das superfícies de resposta e a programação quadrática sucessiva.

2 REVISÃO BIBLIOGRÁFICA

2.1 Panorama do Álcool Combustível no Brasil

Desde a sua descoberta, o petróleo cada vez mais é utilizado como matéria prima para muitos produtos utilizados no cotidiano. Entretanto, o seu emprego em nenhum caso é tão representativo quanto como fonte de energia.

Desde o aumento do preço internacional do petróleo no final da década de 70, vários países têm procurado buscar novas fontes de energia, ou até mesmo extrair seu próprio petróleo, entre estes países o Brasil também está incluso.

Desta forma, o governo brasileiro passou a incentivar, naquela época, a busca de petróleo em nosso território, além de fontes novas e alternativas de energia.

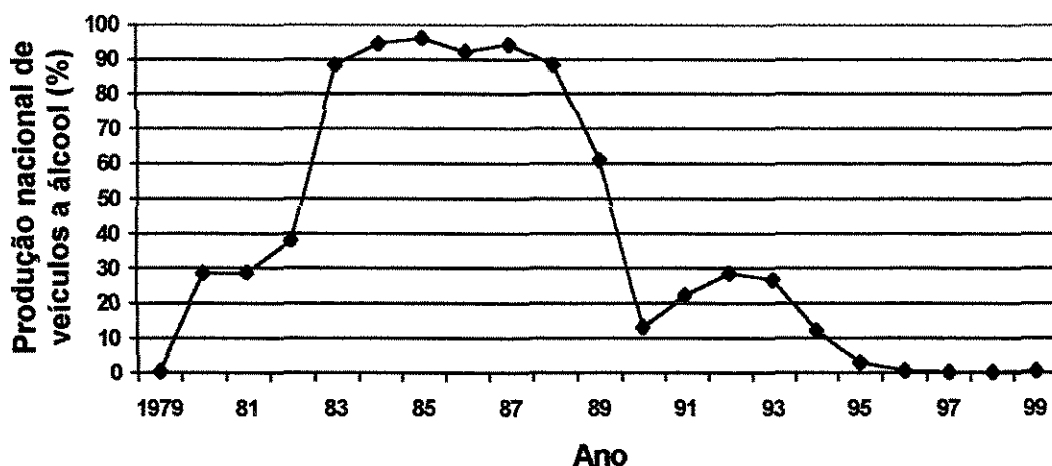


Figura 2.1 - Ascensão e queda do álcool. A produção nacional de veículos a álcool em 06/1999 representava 0,5%.

(Fonte: Revista Quatro Rodas 06/1999)

Assim, a partir de 1975 o Brasil iniciou o programa Pró-álcool, visando a substituição da gasolina pelo álcool combustível. O programa funcionou, pois além do governo conceder incentivos financeiros para os veículos a álcool, o combustível é renovável, tendo como matéria-prima a cana-de-açúcar. O álcool é também menos poluente, produzindo menos monóxido de carbono (CO), óxidos de nitrogênio (NO_x) e substâncias fotoquímicas em comparação com a gasolina (WHEELS, 1999). Logo ganhou a adesão da maioria dos brasileiros. Assim, a produção de veículos nacionais foi praticamente toda voltada para carros a álcool, como pode ser visto na Figura 2.1. Esta iniciativa proporcionou um grande desenvolvimento tecnológico, com o

aprimoramento das variedades de canas cultivadas, técnicas de plantio e colheita, criação de novos processos de extração do álcool, diferentes e melhores microrganismos, além da geração de empregos (GOUVÊIA, 2000). Desta forma, segundo BERG (1999), o país tornou-se o maior produtor mundial de álcool, chegando a produzir 10,768 bilhões de litros deste produto hidratado entre 1991/92.

A partir da segunda metade da década de 80, devido à estabilização do preço internacional do petróleo e ao aumento em direção a auto-suficiência do Brasil com relação a esta matéria-prima, conforme pode ser observado na Tabela 2.1, o governo passou a retirar os incentivos outrora dados ao programa Pró-álcool, tanto em pesquisa quanto sobre o seu valor de comércio.

Tabela 2.1 - Rumo a auto-suficiência, capacidade de produção nacional de petróleo em relação ao seu consumo.

ANO	CAPACIDADE DE PRODUÇÃO NACIONAL
1973	22 %
1983	35 %
1991	59 %
1999	71 %

(Fonte: Revista Veja 23/02/2000)

Durante quase toda a década de 90 os produtores de álcool permaneceram diante da mesma perspectiva. Entretanto, no ano de 1999 um novo aumento no preço internacional do barril de petróleo, a pressão pela retomada do combustível exercida pelo setor, o ajuste cambial e o custo de produção do álcool menor do que o preço da gasolina, além da situação financeira da população favorável, tornaram este combustível novamente uma atrativa fonte de investimentos.

Para isso, tanto o governo quanto o próprio setor alcooleiro adotaram algumas medidas, como por exemplo, a criação de uma Bolsa Brasileira de Álcool (BBA), a qual possui a finalidade de centralizar as vendas de álcool feitas para as distribuidoras de combustível, inclusive trabalhando com o sistema de vendas futuras, ou seja, vender o álcool antes para produzir depois (FURTADO,1999).

O governo também com interesse na retomada do combustível “verde”, estabeleceu o programa de renovação da frota no segundo semestre de 1999. Pelo programa quem trocasse um carro com mais de 15 anos por um novo ganharia um bônus de aproximadamente R\$

2.200,00, mais 1.000 litros de combustível gratuitos, além de isenção de IPVA por dois anos, o que ao final somaria até R\$ 3.200,00 economizados com a promoção, naquela época (KUTNEY, 1999).

Existe a necessidade da diminuição de custos na produção do álcool para que o etanol seja um produto economicamente competitivo com a gasolina em longo prazo. Portanto, é necessário que se continue estudando maneiras de minimizar o seu custo de produção (WHEALS et al., 1999).

De acordo com o que pôde ser visto, a ênfase inicial era tornar o mercado de combustíveis independente do petróleo. Entretanto, a nova ênfase, ou tendência, é a redução da poluição, assim como o consumo de produtos ecologicamente corretos. A Tabela 2.2 mostra a quantidade de etanol utilizada na formulação de alguns combustíveis. Segundo WHEALS et al. (1999), a união européia recomenda que até 2010 toda a gasolina tenha um aditivo de pelo menos 5% em etanol.

Tabela 2.2 - Quantidade de etanol utilizada na formulação de alguns combustíveis.

COMBUSTIVEL	ETANOL (% v/v)
Etanol hidratado (Brasil)	95,5
E85 (América do Norte)	85
Gasolina (Brasil)	24
E10 (gasohol – América do Norte)	10
Combustível Oxigenado (EUA)	7,6
Gasolina Reformulada (EUA)	5,7
Biodiesel (Suécia)	15

(Fonte: ATALA, 2000)

2.2 Produção de Etanol

A utilização de microrganismos em processos para converter uma substância em outra é uma ciência muito estudada. Assim, a fermentação de frutos e cereais até o álcool é conhecida desde culturas primitivas, muito embora somente no século XX tornou-se reconhecida a vasta aplicação deste procedimento, bem como o seu estudo mais aprofundado (SHEREVE & BRINK, 1997).

Um processo biotecnológico ganha um grande destaque por sua versatilidade, alta taxa de produção em produtos específicos e com alto valor agregado. Uma característica comum a estes processos consiste no comportamento complexo, onde vários mecanismos estão envolvidos para a obtenção dos produtos. Tais bioprocessos, devido à sua complexidade,

apresentam não-linearidades como consequência de alguns fatores como: cinética celular de crescimento e formação de produtos, limitações termodinâmicas do processo, efeitos de transferência de massa e calor, agitação do fluido no reator (GOUVÊIA, 2000).

Conforme previamente exposto, nos países pobres em petróleo, é econômico misturar o álcool com a gasolina ou até mesmo utilizá-lo diretamente como combustível (ver Tabela 2.2). No Brasil, a produção de álcool etílico, também conhecido como álcool combustível, é feita a partir de materiais sacarificados (melaço de cana-de-açúcar). Os melaços contêm de 50 a 55% de açúcar fermentável, consistindo principalmente de 70% de sacarose e 30 % de açúcar invertido¹ (SHREVE & BRINK, 1997).

2.2.1 Microrganismos Produtores de Etanol

Muitos microrganismos apresentam etanol como um dos produtos de seu metabolismo. Os principais microrganismos ou mais utilizados são as leveduras dos gêneros *Saccharomyces* e *Schizosaccharomyces*, de acordo com SIQUEIRA (1997), citado por ATALA (2000).

O emprego de bactérias para a produção do álcool pode também ser observado pela utilização da bactéria gram-negativa *Zimomonas mobilis*, capaz de produzir quase a quantidade teórica de etanol a partir da via glicolítica Entner-Doudoroff sobre condições anaeróbicas (LEE & HUANG, 2000). Outros microrganismos são ainda empregados para a fermentação alcoólica, como pode ser observado na Tabela 2.3.

Tabela 2.3 - Alguns microrganismos que produzem etanol a partir de carboidratos.

MICROORGANISMO	SUBSTRATO	PRINCIPAIS METABÓLITOS
<i>Saccharomyces cerevisiae</i>	Hexoses	Etanol, CO ₂
<i>Zimomonas mobilis</i>	Hexoses	Etanol, CO ₂
<i>Ruminococcus albus</i>	Celulose e hemicelulose	Etanol, CO ₂ , ácido acético
<i>Clostridium thermosaccharolyticum</i>	Pentoses	Etanol, CO ₂ , ácidos Acético e Lático
<i>Bacillus macerans</i>	Amido, hexoses	Etanol, CO ₂ , acetonas
<i>Erwinia amylovora</i>	Hexoses	Etanol, CO ₂ , ácido lático
<i>Leuconostoc mesenteroides</i>	Pentoses	Etanol, CO ₂ , ácido lático

(Fonte: ATALA, 2000)

¹ Açúcar invertido: mistura de glicose e frutose.

A planta de etanol produz uma grande quantidade de levedura ou fermento, o qual é reutilizado como inóculo. O excedente é adicionado, depois de seco, à ração animal ou é usado como matéria-prima para o extrato de levedura. Em ambos os casos trata-se de uma rica fonte de proteínas. ECHEGARAY et al. (2000) destacam a utilização das células de levedura como fonte de vários outros produtos de interesse comercial. A *S. cerevisiae* é uma ótima fonte de enzimas como hexokinase, glicose-6-fosfato desidrogenase e invertase, todos com grande interesse comercial.

O processo fermentativo para a produção do etanol pode ser compreendido em três etapas (GOUVÊIA, 2000):

- Preparo da Matéria Prima: são as etapas preliminares que incluem a moagem da cana, peneiragem, decantação (a quente) e resfriamento. Alguns resíduos gerados nestas etapas são o bagaço e a torta, que podem posteriormente ser utilizados como combustível para a caldeira, ou até mesmo como adubo;
- Transformação: trata-se da etapa que envolve os microrganismos, onde estes, agindo como catalisadores biológicos, transformam o açúcar em álcool;
- Separação: é a última etapa e tem a finalidade de separar o álcool do vinhoto; envolve operações como a centrifugação, destilação e no caso do álcool anidro, ainda a anidrização.

De acordo com SILCOX & LEE (1948), citado por SHREVE & BRINK (1997), existem cinco pré-requisitos básicos para um bom processo fermentativo:

- Um microrganismo que leva ao produto final desejado. Este organismo deve se propagar facilmente e ser capaz de manter a uniformidade biológica, levando, por isso, a rendimentos previsíveis;
- Matérias-primas baratas para o substrato; por exemplo, amido ou um entre os vários açúcares;
- Rendimentos aceitáveis;
- Fermentação rápida;
- Produto recuperado e purificado com facilidade.

Alguns fatores que devem ser observados durante a fermentação, são, segundo SHREVE & BRINK (1997), o próprio microrganismo, o equipamento e a fermentação, onde estão inclusos fatores críticos como o pH, a temperatura, a aeração-agitação, fermentação única (cultura pura), e a uniformidade de rendimento.

2.2.2 *Processos Fermentativos para a Produção de Etanol*

A fermentação para o etanol pode ser desenvolvida em batelada, batelada-alimentada ou de modo contínuo. Ou seja, de uma forma geral, a fermentação alcoólica ocorre por dois processos: descontínuos (bateladas), ou contínuos.

- Processos descontínuos: são os processos mais empregados nas destilarias brasileiras. Segundo WHEALS et al. (1999), cerca de 70%; com uma capacidade de fermentação de 1,5 milhões (l etanol)/dia. Controlar a taxa de alimentação do meio em uma cultura em batelada-alimentada é muito vantajoso, pois evita a inibição causada pelo excesso de substrato ou produto sobre o metabolismo da célula (ECHEGARAY et al., 2000; I et al., 1996), sendo esta uma das vantagens deste processo em relação ao processo em batelada. ATALA (2000) apresenta uma série de outras vantagens para este tipo de processo, e entre elas cita o grande potencial para automação e redução de custos operacionais em relação às operações em batelada;
- Processos contínuos: este tipo de processo começou a ser utilizado industrialmente no Brasil ao final da década de 70, mas somente a partir da metade da década seguinte é que o mesmo se consolidou como uma alternativa a produção do álcool. ANDRIETTA (1994) e SIQUEIRA (1997), ambos citados por ATALA (2000), informam as vantagens e desvantagens deste processo, que envolve:
 - Maior produtividade;
 - Maior uniformidade do produto;
 - Maior adaptabilidade ao controle automático;
 - Dificuldade de manutenção de um meio estéril;
 - Ocorrência de mutantes não interessantes ao processo de fermentação.

Segundo ANDRIETTA (1998), citado por GOUVÊIA (2000), estão sendo estudados processos de fermentação contínua que dispensam o uso de centrífugas para a separação da levedura, uma vez que esta operação corresponde a 25% do custo total de uma destilaria.

Além disto, como o processo convencional de fermentação apresenta características inibitórias pelo produto, algumas técnicas vêm sendo estudadas para contornar este problema. HARADA (2001) cita várias destas técnicas, entre elas:

- Fermentação sob vácuo: como o etanol é mais volátil que a água, a sua separação é feita com pressão e temperaturas reduzidas durante o próprio processo de fermentação;
- Pervaporação: trata-se da difusão das substâncias voláteis através de uma membrana hidrofóbica. No caso, o mosto líquido é difundido na superfície da membrana e o álcool separado é então evaporado, sendo condensado posteriormente;
- Fermentação acoplada ao sistema de absorção (stripping) em coluna de recheio: o mosto é enviado para a coluna empacotada, onde o etanol é removido. Segundo o autor a técnica de absorção estudada utilizou o CO_2 formado na própria fermentação, como gás absorvedor.

Da forma convencional, tanto nos processos contínuos quanto descontínuos, as leveduras são separadas do meio fermentado pela centrifugação e reutilizadas em fermentações posteriores após a descontaminação pela lavagem com ácido sulfúrico diluído. O licor fermentado é destilado produzindo uma mistura azeotrópica de álcool hidratado em 95,5% v/v (WHEALS et al., 1999). Para produzir o álcool anidro outras etapas subseqüentes são envolvidas, como a absorção ou a extração líquido-líquido usando benzeno como solvente.

2.3 Modelagem Matemática

A modelagem constitui uma tarefa essencial para o controle de processos, além de ser relevante com relação à otimização. Segundo COSTA (1998), a importância da otimização de processos descontínuos vem crescendo com o desenvolvimento da indústria biotecnológica, pois devido ao alto valor agregado aos produtos oriundos desta indústria, mesmo pequenas melhorias

visando aumentar a eficiência do processo ou diminuir seus custos podem levar a um aumento significativo dos lucros.

Em comum consenso na literatura, a modelagem é tida como fundamental para a engenharia de bioprocessos.

Os modelos desenvolvidos devem possuir a habilidade de representar satisfatoriamente determinado processo, denotando suas características com uma grande capacidade de aproximação. Além disso, devem também, em alguns casos, ser simples (MONTAGUE & MORRIS, 1994).

Tanto a modelagem quanto a otimização e controle de bioprocessos são difíceis, devido ao mecanismo complexo das reações bioquímicas envolvidas, dados apresentando ruídos ou até mesmo não disponíveis em quantidade satisfatória. Aliado a isto ainda, muitas vezes, está a dificuldade em obter dados do processo em tempo real. Ao contrário, são usadas medidas que necessitam testes em laboratório sujeitos a várias fontes de erros (LINKO & ZHU, 1991).

Segundo COSTA (1998), olhando pela perspectiva do modelo determinístico, outro aspecto que contribui para a dificuldade na modelagem é a existência de muitos parâmetros, fenômenos e interações que determinam a cinética do processo fermentativo. SYU & TSAO (1993) declaram que a aproximação clássica é regida pelos princípios da cinética, conservação de massa e energia. Devido à quantidade de informações que devem ser empregadas no desenvolvimento do modelo, torna-se praticamente infactível considerar a todas.

ATALA (2000) citando BAYLEY & OLLIS (1986), descreve a classificação dos modelos cinéticos empregados para fermentação, os quais ficam divididos da seguinte forma:

- Não estruturados e não segregados: a população celular é considerada como soluto monocomponente;
- Estruturados e não segregados: as células são consideradas multicomponentes com composição média semelhante;
- Não estruturados e segregados: as células são consideradas como indivíduos heterogêneos, porém descritos por um único componente;
- Estruturados e segregados: as células são elementos distintos e consideradas como multicomponentes.

De acordo com o trabalho de SAXÉN & SAXÉN (1996), muito embora os modelos estruturados sejam bioquimicamente significantes, eles apresentam certa desvantagem devido a grande quantidade de parâmetros, como já foi comentado. Assim, tanto GOUVÊIA (2000) quanto outros autores consideram o modelo não segregado e não estruturado como uma maneira de contornar este problema, pois este último modelo reduz em muito o número de parâmetros no ajuste. Para o desenvolvimento do modelo são feitas ainda algumas considerações, como por exemplo: o crescimento é balanceado, o microrganismo já está adaptado ao meio de cultura, todos os nutrientes com exceção de um encontram-se em excesso no meio de cultura, o biorreator é perfeitamente agitado (COSTA, 2000).

Entretanto, mesmo quando são escolhidos estes modelos simplificados para se trabalhar, a modelagem convencional de processos biotecnológicos é normalmente difícil e conseqüentemente cara, além de demorada (LANOUILLE et al., 1999).

Com o advento do computador como um equipamento acessível e sua evolução nas últimas décadas como ferramenta de cálculo, modelos alternativos como os baseados em redes neurais artificiais (RNAs) podem ser desenvolvidos.

As RNAs são capazes de, a partir de um conjunto de dados, ou relações de entrada e saída, gerarem um modelo interno apto para descrever o processo em questão. Uma vez gerado este modelo interno, ao serem alimentados novos dados de entrada, a rede pode prever os valores das saídas correspondentes. Alguns aspectos podem ser levantados entre as vantagens das redes neurais artificiais: havendo necessidade, as redes podem ser nova e facilmente treinadas (gerando novo modelo interno) de maneira que se adaptem a alterações no processo; o modelo interno desenvolvido não é baseado em nenhuma especificação dos mecanismos envolvidos no processo a ser modelado (COSTA, 1998), as RNAs apresentam uma boa habilidade para ajustar relações não lineares (SAXÉN & SAXÉN, 1996).

Como desvantagem das RNAs, SCHENKER & AGARWAL (1996) comentam que a modelagem envolve duas etapas, a primeira é a criação do modelo a partir dos dados mensurados; a segunda etapa é a avaliação do melhor modelo e previsão da sua performance. Segundo os autores, o uso de RNAs simplifica a geração do modelo, mas agrava a etapa de seleção devido ao número razoável de modelos a serem considerados, bem como os parâmetros que envolvem cada um.

A estrutura de uma rede pode ser brevemente descrita como um número de nodos ou neurônios interconectados; onde cada neurônio realiza uma transformação não-linear sobre a soma das suas entradas. A maneira pela qual estes nodos são conectados e a natureza das conexões determinam a estrutura da rede. Um algoritmo chamado “algoritmo de aprendizado” é responsável pela maneira como os pesos entre as conexões são ajustados ou treinados para fornecerem o comportamento genérico da rede. Assim, as RNAs são classificadas de acordo com as suas estruturas e algoritmos de treinamento (PHAM, 1995). Ainda de um modo grosseiro, as redes podem ser classificadas como *cíclicas* ou *acíclicas*. Como as redes são divididas em camadas, as primeiras, também conhecidas como *recorrentes*, recebem a saída de algum neurônio na i -ésima camada da rede em níveis inferiores. Já para as redes acíclicas ou *feedforward*, as saídas de neurônios na i -ésima camada não podem ser utilizadas como entrada de nodos em camadas de índice menor ou igual a i (BRAGA et al., 2000).

Considerando o estudo e a aplicação de redes neurais muito promissores, vários trabalhos têm sido publicados nos últimos anos.

LINKO & ZHU (1991) obtiveram um bom resultado para a modelagem de um bioprocessos aplicando uma RNA *feedforward* com três camadas. O algoritmo de treinamento e outros detalhes são descritos no artigo. A rede mostrou-se capaz de prever as variáveis de estado num determinado tempo dadas as variáveis no tempo de amostragem anterior. A RNA bem treinada estimou com precisão e rapidamente as variáveis de estado com ou sem ruídos, mesmo com mudanças na dinâmica do processo, que foram simuladas alterando-se um dos parâmetros do modelo.

YOU & NIKOLAOU (1993) estudaram a capacidade de redes neurais recorrentes para aprender tanto as relações estacionárias quanto a dinâmica de processos contínuos ou em batelada, estudando diferentes formas de apresentar o conjunto de dados à rede. Em seu trabalho os autores concluíram o uso da arquitetura recorrente como uma boa alternativa às redes *feedforward*.

GOH (1995) demonstra o uso de redes neurais com o algoritmo *backpropagation* para a modelagem de sistemas complexos. Tais sistemas envolvem relações empíricas freqüentemente empregadas na estimativa de parâmetros de projeto e propriedades de engenharia. Geralmente os valores experimentais, usados no treinamento da rede, estão acompanhados por ruídos. Os

autores apresentam dois exemplos que demonstram a capacidade da rede em lidar com sistemas não-lineares.

Segundo I et al (1996), para o monitoramento de uma fermentação em batelada alimentada algumas variáveis de estado como substrato, massa de células e concentração de produtos devem estar disponíveis. Infelizmente a maioria destas variáveis não podem ser medidas diretamente. Para contornar isto, vários métodos podem ser aplicados para a estimativa destas variáveis *on-line*, entre eles a modelagem com redes neurais, pois o desenvolvimento destes modelos é mais barato que a aplicação de sensores para a medida direta, além de serem aplicáveis praticamente a todos os processos fermentativos. Assim os autores implementaram satisfatoriamente um controle para a produção de L-lysina usando as RNAs como modelo, que combinam três redes neurais do tipo *feedforward* como estimadores.

TEISSIER et al. (1996) investigaram vários modelos construídos com configurações estáticas e dinâmicas de redes neurais. O objetivo principal era obter uma estimativa e predição em tempo real da concentração do fermento durante o crescimento na segunda etapa da fermentação na produção de Champagne. Segundo os autores, a rede, que foi modelada com arquitetura recorrente, tinha a temperatura e o pH como entradas do modelo e conseguiu estimar a concentração desejada com um erro menor que 3%. Assim, a partir de dados iniciais da fermentação foi possível prever a quantidade final de biomassa ao final de 21 horas de crescimento.

LINKO et al. (1997) denotam que o controle do processo de produção de enzimas é difícil devido a falta de sensores e às incertezas próprias das reações biológicas. Entretanto, empregando uma RNA do tipo *feedforward* com o algoritmo *backpropagation* foram conseguidos bons resultados para o ajuste da atividade de lipase estimada e concentração de biomassa.

CRUZ et al. (1997) simularam a produção de penicilina G por *Penicillium chrysogenum* IFO 8644, empregando uma rede neural *feedforward* com três camadas. Além do *backpropagation* foram também empregados outros algoritmos de treinamento. Segundo os autores os resultados foram promissores e a rede mostrou-se capaz de descrever o comportamento não-linear do processo. Além disso, os mesmos recomendam, baseados nos resultados obtidos, a aplicação desta técnica em algoritmos de controle de processos.

SYU & CHANG (1997) propuseram o uso de uma rede recorrente com *backpropagation* para o controle adaptativo do pH na fermentação da penicilina acylase com *Arthrobacter viscosus*. A

justificativa do uso do modelo escolhido para o controlador deve-se ao fato do mesmo possuir facilidade quanto a identificação. A saída da rede do controlador é a vazão de substrato predita para o próximo intervalo de controle. Como conclusão o pH foi eficientemente controlado e em consequência uma densidade máxima ótima foi obtida.

KASPARIAN & BATUR (1998) apresentam em seu artigo um modelo de referência estruturado em redes neurais que pode ser usado no controle adaptativo de processos lineares e não lineares. O modelo é testado em diversos sistemas não-lineares simulados. Os autores testaram também a eficiência do algoritmo de treinamento de Davidon, o qual é similar aos métodos quasi-Newton de otimização, onde os pesos do índice de performance são atualizados recursivamente em proporção à aproximação quadrática a este índice. Este algoritmo apresentou um resultado melhor do que o algoritmo *backpropagation*, convergindo mais rapidamente.

COTTIS et al. (1999) empregam a técnica de redes neurais artificiais para a modelagem do comportamento da corrosão em aço inoxidável como função da composição da solução e temperatura. As predições da rede exibem correlações razoáveis para dados de soluções com um ou dois componentes, contudo, segundo os autores, o problema de testar a capacidade de generalização da rede ainda procede. O artigo denota também dificuldades como a variabilidade do comportamento da corrosão e o comportamento imprevisível da rede que pode ocorrer em regiões do domínio onde não existem dados suficientemente disponíveis. Para tanto sugerem o uso de dados gerados em simulação para o treinamento da rede, a fim de assegurar a aplicabilidade e a boa performance do método.

De acordo com THOLUDUR et al. (2000), a modelagem com redes neurais tem se mostrado muito eficaz quando aplicada a dinâmicas complexas como crescimento celular e produção de proteínas em processos biotecnológicos. O desenvolvimento de modelos com redes neurais requer, geralmente, quantidades grandes de dados obtidos experimentalmente. O trabalho em questão examina a habilidade do uso de funções interpoladas em conjunto com RNAs para reduzir a quantidade de dados experimentais requeridos ao desenvolvimento do modelo para sistemas dinâmicos. Os autores apresentam uma redução em 40% na quantidade de experimentos realizados para a modelagem, o que confirma a eficiência do método proposto.

2.4 Redes Neurais Artificiais

A definição de Redes Neurais Artificiais segundo BRAGA et al. (2000), consegue sintetizar o conceito e a estrutura do sistema, assim citando o referido autor:

“RNAs são sistemas paralelos distribuídos compostos por unidades de processamento simples (nodos) que calculam determinadas funções matemáticas (normalmente não lineares). Tais unidades são dispostas em uma ou mais camadas e interligadas por um grande número de conexões, geralmente unidirecionais. Na maioria dos modelos estas conexões estão associadas a pesos, os quais armazenam o conhecimento representado no modelo e servem para ponderar a entrada recebida por cada neurônio da rede. O funcionamento destas redes é inspirado em uma estrutura física concebida pela natureza: o cérebro humano”.

2.4.1 O Sistema nervoso e o Neurônio

O sistema nervoso trata-se de um complexo conjunto de células chamados neurônios. Tais células se interconectam de forma específica e precisa, formando os chamados circuitos neurais.

Em consequência deste avançado sistema nervoso, todo ser vivo que disponha do mesmo é capaz de modificar ações ou seu comportamento em função de experiências passadas. Essa modificação comportamental é chamada de aprendizado.

O Neurônio é o principal componente do sistema nervoso e estima-se que no cérebro humano existam cerca de 15 bilhões destas células, que são responsáveis por todas as funções do sistema. Os neurônios são formados pelos dendritos, que são um conjunto de terminais de entrada, pelo corpo central, e pelos axônios, que são longos terminais de saída.

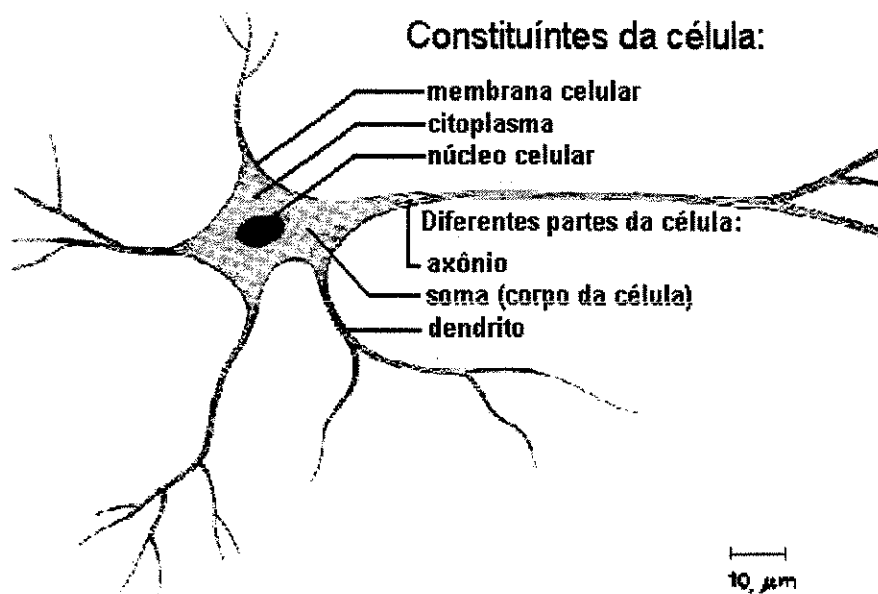


Figura 2.2 - Constituintes da célula neural - Esquema.

Embora existam várias formas de células neurais, com diferentes funções dependendo da sua localização e estruturas morfológicas, em geral possuem a mesma constituição conforme descrito na Figura 2.2.

A junção interneural através da qual os impulsos nervosos são transmitidos, denominada sinapse, é a forma de comunicação entre uma célula nervosa e outra. Esta junção pode ser compreendida da seguinte forma: quando um potencial de ação chega a um terminal pré-sináptico, ou seja, um determinado neurônio recebe os impulsos, em um certo momento são processados, e atingindo o limiar de ação, a despolarização da membrana pode fazer com que um pequeno número de vesículas sinápticas secrete um composto químico denominado neurotransmissor na fenda sináptica, axônio, que pode estar conectado a um dendrito de um outro neurônio. O neurotransmissor irá atuar, por sua vez, nas proteínas receptoras do neurônio seguinte produzindo alterações na permeabilidade da membrana neural pós-sináptica, podendo excitá-lo, inibi-lo, ou modificar de alguma forma sua intensidade. Este processo depende de vários fatores, como a geometria da sinapse e o tipo de neurotransmissor (BASHEER & HAJMEER, 2000).

A aprendizagem pode levar a alterações estruturais no cérebro. Esta característica conhecida como plasticidade neural, é unicamente do sistema nervoso. Trata-se, portanto, de uma propriedade que permite o desenvolvimento de alterações estruturais em resposta à

experiência, e como adaptação a condições mutantes e a estímulos repetidos. Assim, a cada nova experiência do indivíduo, redes de neurônios são rearranjadas, outras tantas sinapses são reforçadas e múltiplas possibilidades de respostas ao ambiente tornam-se possíveis. Assim, o mapa cortical de um adulto está sujeito a constantes modificações com base no uso ou atividade de seus caminhos sensoriais periféricos.

Segundo BASHEER & HAJMEER (2000), a crua analogia entre os neurônios artificiais e biológicos é que as conexões entre os nodos representam os axônios e os dendritos, os pesos as sinapses enquanto o *threshold* aproxima-se da atividade na soma. Ambos, a RNA e a rede biológica, aprendem pelo ajuste incremental dos pesos ou da força das sinapses. A Figura 2.3 ilustra um nodo ou neurônio artificial recebendo vários sinais de intensidade x e força sináptica (pesos) w , com um *threshold* ou ativação b .

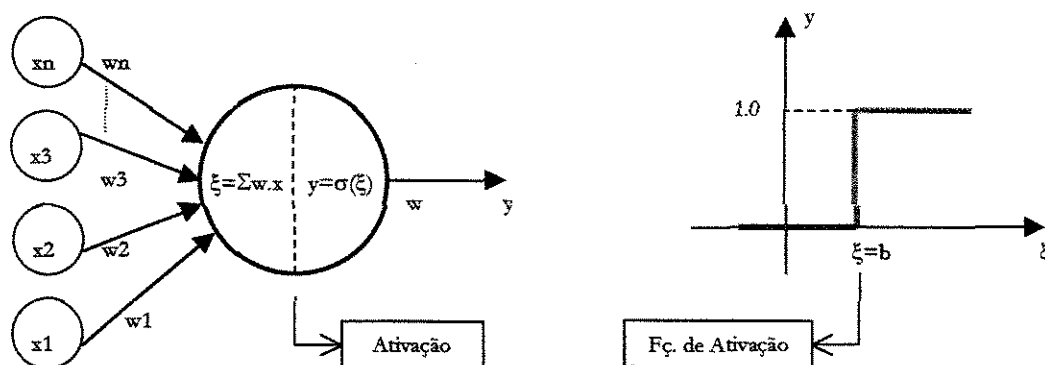


Figura 2.3 - Soma de sinal no neurônio artificial, modelo tipo monocamada Perceptron.

2.4.2 Histórico

O conceito de redes neurais foi primeiramente apresentada com o artigo de McCulloch e Pitts em 1943. Estes pesquisadores mostraram que as redes com modelos neurais são capazes de simular máquinas computacionais com aplicação em qualquer finalidade (BISHOP, 1994). O modelo destes pesquisadores propõe elementos computacionais, abstraídos das propriedades fisiológicas dos neurônios biológicos e de suas conexões, introduzindo assim a referência número um para a teoria de RNA's (OLIVEIRA; 2000).

Em consequência à teoria apresentada no início da década de 40, no ano de 1949 foi publicado o livro *"The Organization of Behaviour"* escrito por Hebb. Neste livro o autor propôs um mecanismo específico para o aprendizado em redes neurais biológicas, sugerindo que o

aprendizado ocorra através de modificações na força ou peso nas interconexões sinápticas entre os neurônios, ou seja, uma entrada sináptica, que é ativada quando a saída do neurônio² também é ativada, deverá ter seu peso modificado visando o fortalecimento da interconexão.

O terceiro passo antológico sucedido em RNA foi no ano de 1958 com a criação do modelo *Perceptron* por Rosenblatt. Baseado no modelo de McCulloch e Pitts, Rosenblatt desenvolveu o primeiro aparelho, *hardware*, de redes neurais, juntamente com o algoritmo de aprendizado da rede. Ele acreditava na idéia em que o cérebro humano trabalha como um associador de padrões adaptável, e não como um circuito lógico determinístico como era a idéia anterior. O Perceptron trata-se de uma rede de múltiplos neurônios do tipo discriminadores lineares que já implementava o princípio hebbiano de treinamento, como brevemente descrito no parágrafo anterior. Redes similares foram estudadas por Widrow que desenvolveu a rede chamada ADALINE (*ADaptative LINear Element*) e uma regra de treinamento correspondente conhecida como Regra de Widrow-Hoff de treinamento. Mais tarde a rede ADALINE recebeu sua versão multicamadas, daí então foi renomeada como MADALINE, acrescentado a palavra *Multiple* ao nome já conhecido. A contribuição principal desta rede desenvolvida por Widrow foi um princípio muito eficaz de treinamento, que passou a ser conhecido como a Regra Delta, sendo mais tarde também generalizada para modelos mais complexos.

Apesar do “sucesso” de até então das redes neurais, o interesse na área sofreu uma diminuição gradativa, porém acentuada, a partir do final da década de 60. Isto devido ao grande número de problemas que surgiram e que, entretanto, não podiam ser resolvidos com os algoritmos disponíveis na época. Desta forma, baseando-se na classe de problemas *linearmente não-separáveis*, os quais não podiam ser resolvidos pelo Perceptron ou ADALINE, várias críticas foram feitas ao sistema de RNA's, desacreditando até o final da década de 70 o sistema.

Pelo motivo apresentado vários autores não apresentam em sua literatura uma revisão histórica sobre o que foi produzido durante este período. No entanto, alguns modelos importantes foram desenvolvidos, entre eles vale salientar o sistema de redes neurais sem peso. Esse sistema proporcionava modelos de rápida aprendizagem e fácil implementação. O princípio de funcionamento de uma Rede Neural sem Peso (RNSP) consiste em armazenar a função do nodo em uma memória RAM, ao invés de ajustar pesos entre os nodos.

² neurônio ou nodo: desempenha uma função de soma e threshold ou outras, onde os pesos nas conexões entre os nodos correspondem às sinapses inibidoras e/ou excitadoras de um neurônio real.

A partir do início da década de 80, no entanto, marca-se o ressurgimento das redes neurais com a publicação dos trabalhos de Hopfield, em 1982, relatando a utilização de redes simétricas para otimização e de Rumelhart, Hinton e Williams que introduziram o eficaz método de retropropagação do erro, conhecido como *Backpropagation* (PSICHOGIOS & UNGAR; 1992). Além destes, alguns outros aspectos contribuíram para o novo fôlego, entre eles: 1) a disponibilidade de microcomputadores mais fortes e baratos; 2) o desenvolvimento de modelos de redes neurais não lineares; 3) a redescoberta de técnicas para treinamento das camadas intermediárias dos neurônios ou nodos.

A partir de então muitos pesquisadores desenvolveram o interesse pelo sistema de RNA, principalmente com a publicação do livro "*Parallel Distributed Processing*" escrito por Rumelhart et al. (BISHOP, 1994).

Segundo BRAGA et al. (2000), hoje determinados tipos de redes podem resolver qualquer problema. E ainda segundo BISHOP (1994), atualmente aplicações de redes neurais podem ser encontradas em produtos eletrônicos de consumo, desde câmeras de vídeo até máquinas de lavar "inteligentes".

2.4.2.1 Vantagens e Desvantagens das RNAs

Algumas das razões para a utilização de redes neurais segundo BLAESI (1992), são:

- Uma rede neural pode ser desenvolvida com menos tempo de programação e menos tempo de verificação do que um modelo teórico convencional;
- Uma rede neural aprende com o processo corrente, portanto, ajustes devidos à diferença entre a teoria e os valores práticos não são necessários;
- Caso as condições de operação sejam trocadas a rede neural pode novamente ser treinada para responder corretamente a estas modificações, sem, contudo necessitar especialistas tanto no processo quanto para o modelo.
- BASHEER & HAJMEER (2000) ainda acrescentam: não-linearidade, alto paralelismo, robustez, tolerância a falhas ou faltas, habilidade em lidar com informações imprecisas e capacidade de generalização.

Ainda segundo BLAESI (1992), a desvantagem no uso de RNAs está na grande quantidade de dados representativos do processo necessários para gerar um bom modelo neural.

Isto, contudo, varia de acordo com a qualidade destes dados do processo. Pode, ainda, ser destacado que um bom conjunto de treinamento contém desde dados obtidos quando o processo possui um ótimo funcionamento, até situações não usuais de operação ou sua paradas, se uma modelagem tão completa for desejável, o que aumenta em muito a complexidade do sistema.

Um dos aspectos que ajudam na velocidade de implementação do modelo neural está no fato de que alguns parâmetros que possuam pouco ou nenhum efeito sobre os resultados podem ser descartados; isto, no entanto, deve ser feito levando em consideração a técnica adequada.

O uso de RNAs não é recomendado quando dados suficientes para a criação do modelo não são disponíveis, existe um modelo disponível que representa adequadamente o processo ou um pequeno modelo pode ser facilmente gerado a partir de técnicas de regressão estatísticas.

2.4.3 Algoritmos de Treinamento

Existem vários tipos de algoritmos de treinamento. Cada algoritmo ou método apresenta suas vantagens e desvantagens, diferindo entre si, basicamente, pela maneira pela qual o ajuste de pesos é feito. Portanto a utilização de um método ou outro depende exclusivamente do caso em estudo.

Citando a definição geral do que vem a ser a aprendizagem, segundo BRAGA et al. (2000):

“Aprendizagem é o processo pelo qual os parâmetros de uma rede neural são ajustados através de uma forma continuada de estímulo pelo ambiente no qual a rede está operando, sendo o tipo específico de aprendizagem realizado definido pela maneira particular como ocorrem os ajustes realizados nos parâmetros”.

As várias metodologias de treinamento de RNAs podem ser compreendidas em duas categorias, conhecidas como paradigmas: aprendizado supervisionado e aprendizado não-supervisionado. Alguns autores citam outros paradigmas, entretanto são apenas casos mais específicos destes apresentados.

O método de aprendizado supervisionado é o mais utilizado para o treinamento de redes neurais. Este método pode ser implementado de formas *on-line* e *off-line*. No primeiro a rede deve estar em processo contínuo de treinamento, pois os dados mudam continuamente. No sistema *off-line* o conjunto de treinamento é fixo, e uma vez obtida a solução para a rede esta deve permanecer fixa (BRAGA et al., 2000).

No aprendizado supervisionado ocorre a interação de um supervisor externo, daí o nome do método. Os valores de entrada e saída são fornecidos por este supervisor, conforme ilustra a Figura 2.4. A idéia é encontrar uma ligação entre os pares de entrada e saída, e para tanto é feita a comparação entre a resposta calculada pela rede e a desejada (solução ótima). A informação sobre o erro da resposta atual é complementada pelo supervisor, indicando um comportamento bom ou ruim para a rede. A partir disto os pesos das conexões são ajustados para minimizar, incrementalmente, o erro. A soma dos erros quadráticos de todas as saídas é geralmente utilizada tanto como função a ser minimizada quanto como medida de desempenho da rede.

Alguns dos exemplos mais conhecidos de aprendizado supervisionado são a regra delta e a sua generalização, mais conhecida como algoritmo *Backpropagation*, para redes de múltiplas camadas. A desvantagem deste tipo de aprendizado está na impossibilidade da rede aprender novas estratégias para situações não apresentadas dentro dos exemplos de treinamento.

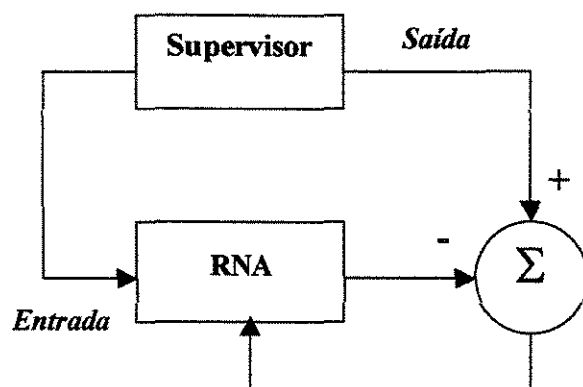


Figura 2.4 - Aprendizado Supervisionado.

O aprendizado não-supervisionado, assim como o próprio nome o diz, não possui um supervisor para acompanhar o processo de treinamento. Este tipo de aprendizado, de uma forma mais geral, necessita somente padrões ou dados de entrada, conforme ilustrado na Figura 2.5.

Segundo BRAGA et al. (2000), a partir do momento em que a rede estabelece uma harmonia com as regularidades estatísticas da entrada de dados, desenvolve-se nela uma habilidade de formar representações internas para codificar características da entrada e criar novas classes ou grupos automaticamente. Deve-se observar, no entanto, que os dados de entrada precisam apresentar redundância para que os padrões ou características dos dados de entrada possam ser encontrados, e conseqüentemente se chegar às respostas.

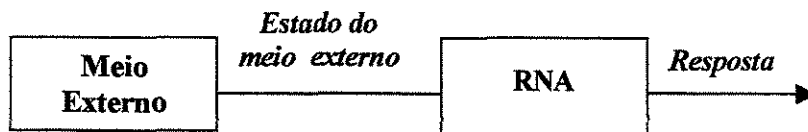


Figura 2.5 - Aprendizado não-supervisionado.

A metodologia do aprendizado não-supervisionado pode adquirir várias formas diferentes, variando os tipos de conexões, o número de camadas ou a maneira como são modificados os pesos sinápticos entre as conexões. Entre as várias formas para implementação desta metodologia, as principais são as seguintes: Aprendizado hebbiano; Modelo de Linsker; Regra de Oja; Regra de Yuille; Aprendizado por competição.

A descrição detalhada de cada método para implementação do sistema de aprendizado foge ao escopo deste trabalho. Serão discutidos, entretanto, os métodos utilizados, e serão citados aqueles secundários.

Existe ainda um terceiro sistema de aprendizado, aprendizado por reforço, que assim como dito anteriormente trata-se apenas de um caso particular dos sistemas já apresentados, no caso o aprendizado supervisionado. O que diferencia esse método daqueles discutidos é a medida de desempenho utilizado em cada um. Enquanto no primeiro esta medida é baseada em um conjunto de respostas desejadas usando um critério de erro conhecido, no aprendizado por reforço é realimentado à rede somente um simples sinal de reforço (sim/não), sem qualquer informação instrutiva que melhore o desempenho da rede. De uma maneira geral, se uma ação tomada pelo sistema de aprendizagem não é seguida de estados satisfatórios, a tendência do sistema de produzir esta ação é enfraquecida (BRAGA et al., 2000).

Como pode ser acompanhado da Figura 2.6, no aprendizado por reforço um crítico externo avalia a resposta fornecida pela rede.

O aprendizado por reforço ainda pode ser “associativo” ou “não-associativo”. No primeiro, um mapeamento, na forma estímulo/ação, deve ser aprendido com as informações adicionais fornecidas pelo meio. Já no segundo o sistema não recebe informações adicionais, funcionando apenas com a avaliação do desempenho da rede (sim/não).

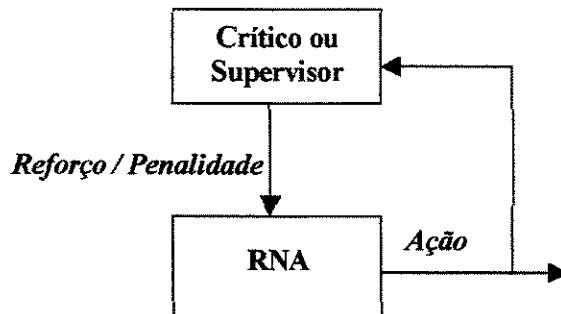


Figura 2.6 - Aprendizado por Reforço.

A maneira como são feitas as correções dos pesos em uma rede também pode ser dividida em duas formas, “Modo Padrão” e “Modo Batelada”. No modo padrão os pesos são corrigidos a cada apresentação para a rede de um par do conjunto de treinamento. No modo batelada, como o nome sugere, é feita apenas uma correção por ciclo, ou seja, depois que todos os pares de exemplos foram apresentados à rede, seu erro médio é calculado e a partir deste erro são feitas as correções dos pesos. Lembrando que um conjunto de treinamento possui N pares de exemplos (entrada e saída), e que a apresentação destes pares de exemplos constitui um ciclo na rede.

Existem algumas vantagens e desvantagens em se utilizar o Modo Padrão ou o Modo Batelada para a apresentação dos pares de exemplos do conjunto de treinamento. No primeiro, modo padrão, quando são empregadas taxas de aprendizado elevadas a rede geralmente torna-se instável, entretanto, ocupa menos memória computacional. No segundo modo, a abordagem apresenta uma estimativa mais precisa do vetor gradiente, sendo geralmente mais estável, porém necessitando mais memória.

2.4.4 O Algoritmo Back-Propagation

Muitos algoritmos de aprendizado supervisionado foram desenvolvidos para treinar redes do tipo MLP (*Multilayer Perceptron*), o mais conhecido entre todos é o algoritmo *backpropagation*, pois a maioria dos métodos de aprendizado usa variações deste.

Historicamente derivado da regra delta padrão proposta por Widrow e Hoff, e conhecido como regra delta generalizada, este algoritmo foi um dos principais responsáveis pelo fôlego novo dado as RNAs durante a década de 80. Isto porque combinado com uma topologia adequada de rede, fornece bons resultados para qualquer problema implementado.

O treinamento com o *backpropagation*, utilizando pares de exemplos (entrada, saída desejada), ocorre em duas fases. Na primeira, conhecida como fase *forward*, os padrões de entrada são apresentados à rede, ou melhor, são apresentados a camada de entrada, fluindo pelas camadas intermediárias, até que uma resposta seja produzida pela camada de saída. Na segunda, fase *backward*, a resposta obtida da rede é comparada com a saída desejada. Caso não esteja correta, o erro é calculado e retropropagado da camada de saída à entrada, ajustando assim os pesos das conexões nas unidades das camadas internas. Este procedimento tem continuidade até que a rede atinja um critério de parada pré-definido. O algoritmo do *backpropagation* é apresentado em seguida.

2.4.4.1 Treinamento da Rede

O treinamento da rede é feito procurando-se por valores de pesos que minimizem a função erro, que é geralmente a soma dos erros quadráticos:

$$E = \frac{1}{2} \sum_p \sum_{i=1}^k (d_i^p - y_i^p)^2 \quad (2.1)$$

Na equação acima, E é a medida de erro total, k corresponde ao número de unidades de saída, p é o número de padrões, y é a i -ésima saída gerada, d a i -ésima saída desejada. Geometricamente, a função erro corresponde à superfície de erro sobre as coordenadas de peso (ver Figura 2.7). O treinamento da rede consiste em se procurar um ponto mínimo na superfície do erro. Existem, contudo, pontos de mínimo que não correspondem ao mínimo global, são pontos conhecidos como mínimos locais, e podem em alguns casos fornecer a resposta de treinamento para a rede. Entretanto, o objetivo principal do treinamento é a procura pelo ponto de mínimo global.

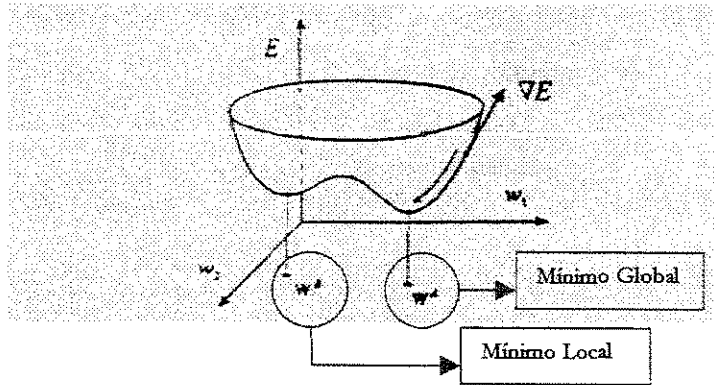


Figura 2.7 - Ilustração da função erro $E(w)$ vista como uma superfície sobre a área dos pesos.

Ao contrário de redes com apenas uma camada, nas redes multicamadas a função erro é uma função altamente não-linear dos pesos. A procura pelo ponto de mínimo é feita iterativamente, tendo como partida algum peso aleatório. Para redes monocamada com funções de ativação lineares não existem pontos de mínimo, em seu treinamento é utilizada a regra delta padrão.

O algoritmo *backpropagation* é um procedimento muito eficiente para realizar a distribuição dos pesos entre as conexões das funções não lineares utilizadas na rede MLP. Este método é obtido a partir da função soma dos erros quadráticos, para tanto considere a Equação 2.1.

A medida total de erro (eq. 2.1) pode ser dividida em duas equações:

$$E = \sum_p E^p, \quad E^p = \frac{1}{2} \sum_{j=1}^k (d_j^p - y_j^p)^2 \quad (2.2)$$

(a)

(b)

Para simplificar a notação será omitido o índice p das variáveis da rede.

É sugerido pela regra delta que a variação dos pesos seja definida de acordo com o gradiente descendente, ou seja, o treinamento da rede começa com um peso aleatório, como já foi dito, e procede com pequenas variações no vetor peso com o objetivo de mover-se na direção do gradiente negativo da função erro, até que o vetor peso atinja um mínimo local ou global. Desta forma,

$$\Delta w_{ji} \propto - \frac{\partial E}{\partial w_{ji}} \quad (2.3)$$

Para definir como cada um dos pesos de cada nodo da rede deve ser ajustado de forma a diminuir o erro total gerado pela rede, a derivada encontrada na última equação deve ser dividida como o produto de duas partes, a primeira refletindo as mudanças no erro como uma função de mudanças em ξ , definido como $net_j = \xi = \sum_{i=1}^n x_i \cdot w_{ji}$, e a segunda parte representando os efeitos da mudança de um peso particular em ξ . Utilizando a regra da cadeia,

$$\frac{\partial E}{\partial w_{ji}} = \frac{\partial E}{\partial net_j} \cdot \frac{\partial net_j}{\partial w_{ji}} \quad (2.4)$$

A segunda parte da equação 2.4 pode ser calculada facilmente:

$$\frac{\partial net_j}{\partial w_{ji}} = \frac{\partial \sum_{i=1}^n x_i \cdot w_{ji}}{\partial w_{ji}} = x_i \quad (2.5)$$

A primeira parte do lado direito da equação 2.4 mede o erro do nodo j , e geralmente é abreviada para δ_j .

$$\delta_j = \frac{\partial E}{\partial net_j} \quad (2.6)$$

Novamente utilizando a regra da cadeia;

$$\delta_j = \frac{\partial E}{\partial net_j} = \frac{\partial E}{\partial y_j} \cdot \frac{\partial y_j}{\partial net_j} \quad (2.7)$$

Onde,

$$\frac{\partial y_j}{\partial net_j} = \frac{\partial f(net_j)}{\partial net_j} = f'(net_j) \quad (2.8)$$

O cálculo da 1ª derivada no lado direito da equação 2.7 que utiliza o erro, depende da camada onde o nodo j se encontra. Quando o nodo estiver na última camada, o erro pode ser definido utilizando a equação 2.2-b.

$$\frac{\partial E}{\partial y_j} = \frac{\partial \left(1/2 \sum_{i=1}^k (d_i - y_i)^2 \right)}{\partial y_j} = (d_i - y_i) \quad (2.9)$$

A equação 2.9 é a mesma fórmula da regra delta original. Agora, substituindo os dois termos do lado direito da equação 2.7:

$$\delta_j = (d_j - y_j)f'(\text{net}_j) \quad (2.10)$$

Quando o nodo não for de saída, a regra da cadeia é utilizada para escrever,

$$\frac{\partial E}{\partial y_j} = \sum_{i=1}^M \frac{\partial E}{\partial \text{net}_i} \frac{\partial \text{net}_j}{\partial y_j} = \sum_{i=1}^M \frac{\partial E}{\partial \text{net}_i} \frac{\partial \sum_{l=1}^n w_{il} y_l}{\partial y_j} = \sum_{i=1}^M \frac{\partial E}{\partial \text{net}_i} w_{ji} \quad (2.11)$$

Onde,

$$\sum_{i=1}^M \frac{\partial E}{\partial \text{net}_i} w_{ji} = \sum_{i=1}^M \delta_i w_{ji} \quad (2.12)$$

Substituindo novamente os dois termos do lado direito da Equação 2.7, é obtido o erro definido para os nodos das camadas intermediárias.

$$\delta_j = f'(\text{net}_j) \sum_i \delta_i w_{ji} \quad (2.13)$$

Assim, generalizando a fórmula de ajuste de pesos proposta na Equação 2.3,

$$\Delta w_{ji} = \eta \delta_j x_i \quad (a)$$

(2.14)

$$w_{ji}(t+1) = w_{ji}(t) + \eta \delta_j(t) x_i(t) \quad (b)$$

Portanto se o nodo for de saída, o erro será definido pela Equação 2.10, caso contrário, δ_j será dado pela Equação 2.13.

2.4.5 Tipos e Topologias de Redes

O tipo de uma determinada RNA é um importante aspecto, uma vez que pode definir quais as classes de problemas que poderão ser abordados pela rede. Um exemplo disto é a rede MCP (*McCulloch Pitts*) com uma única camada de nodos, que são hábeis para resolver problemas linearmente separáveis apenas. Como uma rede neural é composta por nodos ou neurônios, a maneira pela qual eles estão interconectados, bem como a quantidade dos mesmos, definem a estrutura da rede. O modo como são ajustados os pesos entre as conexões é definido pelo algoritmo de treinamento, conforme já discutido. Assim sendo, os parâmetros importantes que participam da definição de uma estrutura são, segundo BRAGA et al. (2000), os seguintes:

- número de camadas da rede;
- número de nodos em cada camada;
- tipo de conexão entre os nodos e topologia da rede.

Alguns tipos de arquiteturas de RNAs são apresentados a seguir na Figura 2.8:

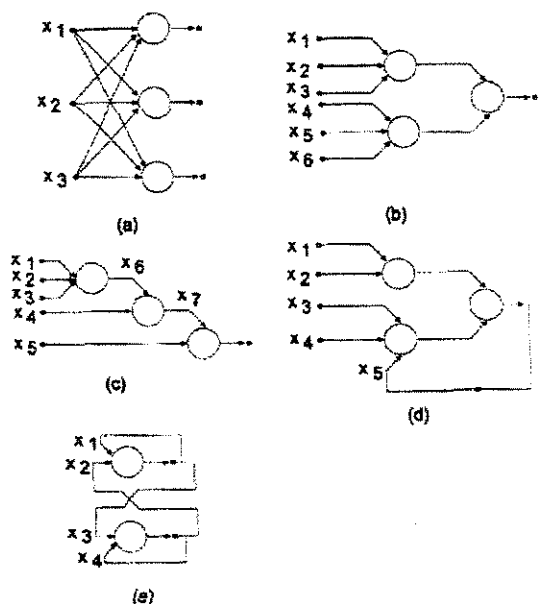


Figura 2.8 - Exemplos de Arquiteturas de RNAs

De acordo com o que pode ser observado da Figura 2.8, as redes podem ser de uma única camada, ou seja, apenas com um nó entre qualquer entrada e saída rede (Figura 2.8 a, e), ou ainda, de múltiplas camadas, quando existem mais de um neurônio entre alguma entrada e saída da rede (Figura 2.8 b,c,d).

Estruturalmente as redes podem ser divididas em Acíclicas, *feedforward*, ou Cíclicas, *feedback*. Na primeira, em nenhum momento a saída de um neurônio na i -ésima camada é utilizada para realimentar nodos em camadas de índice menor ou igual a i , como pode ser observado da Figura 2.8 a,b,c. Já quando a rede é cíclica a saída de algum neurônio da i -ésima camada da rede é retornada como entrada para os nodos das camadas inferiores (Figura 2.8 d,e).

Ainda, as RNAs podem ser classificadas como redes fracamente conectadas, Figura 2.8 b,c,d, ou redes completamente conectadas, Figura 2.8 a,e.

Dentro do que foi apresentado, como tipos de redes neurais podem ser destacadas as principais. Entre elas, como citado anteriormente, as redes Perceptron de uma única camada e as redes lineares do tipo ADALINE, ocupam um lugar de destaque historicamente. Ambas utilizam o nodo MCP (McCulloch Pitts) como unidade básica, porém se diferenciam quando o modelo Adaline calcula o erro de saída em relação à sua saída linear, e não em relação à saída não-linear, como na maioria dos algoritmos de treinamento de RNAs e o Perceptron.

A grande dificuldade destas e de outras redes de uma só camada é que resolvem apenas problemas linearmente separáveis. Para que problemas não separáveis linearmente pudessem ser resolvidos foram criadas as redes com uma ou mais camadas internas ou escondidas. Segundo BRAGA et al. (2000), uma rede com duas camadas intermediárias permite a aproximação de qualquer função, enquanto a rede contendo apenas uma camada escondida permite a aproximação de qualquer função contínua.

Alguns dos principais exemplos de redes neurais, segundo PHAM (1995), são:

- *Multi-layer Perceptron* (MLP);
- *Learning Vector Quantization* (LVQ);
- *Group Method of Data Handling* (GMDH);
- *Hopfield network*;
- *Elman and Jordan networks*;
- *Kohonen network*;

Cada um destes exemplos de redes apresenta características que os diferem entre si, entretanto são ordenados conforme a classificação já descrita. A rede LVQ, por exemplo, é uma rede *feedforward* que apresenta o aprendizado por reforço não-associativo, e é parcialmente conectada.

2.4.5.1 Multi-Layer Perceptron

Como citado por vários autores, o melhor tipo de rede é a MLP. O algoritmo *backpropagation* foi inclusive desenvolvido para esta rede.

A rede MLP possui três camadas, uma de entrada, uma intermediária e outra de saída, Figura 2.10. Os neurônios na camada de entrada atuam apenas como *buffers* para distribuir os

sinais que chegam (x_i) para os nodos das camadas internas. De acordo com PHAM (1995), cada nodo j na camada intermediária soma os sinais de entrada (x_i) depois de pesá-los com as forças das respectivas conexões w_{ji} da camada de entrada e computa suas saídas z_j como uma função g da soma. Para uma rede neural monocamada, Figura 2.9, a saída contendo a função soma, também conhecida como função de ativação, é dada pela Equação 2.15, a seguir.

$$z_j = g \left(\sum_{i=0}^d w_{ji} x_i \right) \quad (2.15)$$

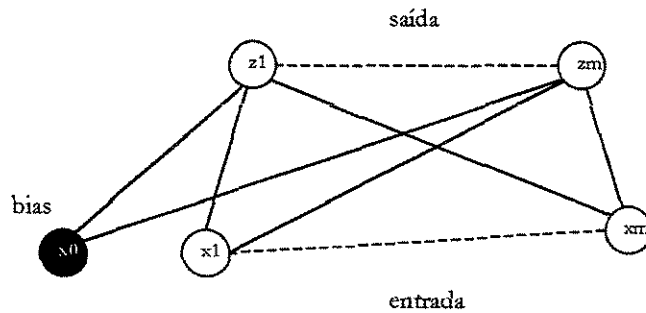


Figura 2.9 - Rede monocamada tendo d entradas x_1, \dots, x_d e m saídas z_1, \dots, z_m .

As saídas de uma rede multicamadas são conseguidas ativando as respostas z_j , conforme a Equação 2.15, com uma segunda transformação, correspondendo a segunda camada de unidades ou nodos da rede. A equação completa para a transformação representada pela rede multicamada é da forma a seguir.

$$y_k = \bar{g} \left(\sum_{j=0}^m \bar{w}_{kj} z_j \right)$$

$$y_k = \bar{g} \left(\sum_{j=0}^m \bar{w}_{kj} g \left(\sum_{i=0}^d w_{ji} x_i \right) \right) \quad (2.16)$$

Na expressão acima, as variáveis \bar{g} , \bar{w}_{kj} e y_k são, respectivamente, a função soma ou de ativação, os pesos e a saída obtida na segunda camada, conectando o nodo escondido j à saída k .

De acordo com BRAGA et al. (2000), a função de ativação sigmoideal é a mais utilizada, pois o treinamento de redes com mais de uma camada baseia-se no método do gradiente descendente, e, a fim de que este método funcione, a função de ativação precisa ser contínua, diferenciável e não-decrescente, e ao mesmo tempo informar os erros cometidos pela rede para as camadas anteriores com a maior precisão possível, características desta função.

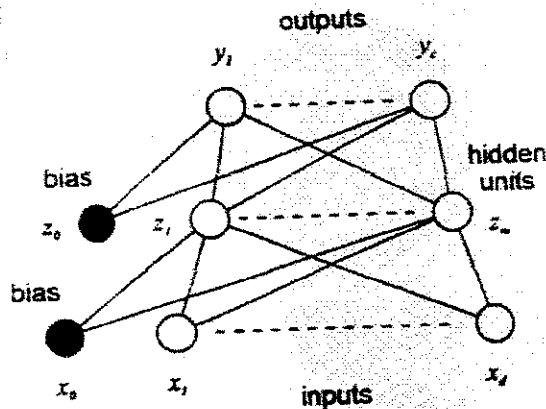


Figura 2.10 - Rede MLP tendo duas camadas de pesos.

2.4.5.2 Neurônios e Funções de Ativação

Para montar o sistema de redes neurais, McCulloch e Pitts simplificaram o que se sabia sobre o neurônio biológico até então. Um neurônio, como já discutido, emite sinais excitatórios ou inibitórios. Tais sinais são emitidos quando é ultrapassado o limiar de excitação (*threshold*).

No sistema de RNAs, o corpo do neurônio é representado pela soma do produto entre os pesos das conexões e a entrada, enquanto a ativação ou sinais emitidos pelo neurônio são obtidos através da aplicação de uma *função de ativação*. Esta função depende do valor da soma ponderada de suas entradas. Portanto, quando o somatório nas Equações 2.15 e 2.16 ultrapassar um determinado limiar, θ , o nodo terá sua saída ativada, conforme ilustra a Figura 2.3.

O modelo proposto inicialmente por McCulloch e Pitts, tanto para a rede quanto para função de ativação dos nodos, foi muito simples e apresentava algumas limitações, entretanto abriu o caminho para especulações. A partir de então foram derivados vários outros modelos que permitem a produção de uma saída qualquer, não necessariamente zero ou um como propunham aqueles pesquisadores, além de diferentes funções de ativação. As quatro principais funções de ativação são mostradas na Figura 2.11.

A função sigmoideal, Figura 2.11-d, também conhecida como *S-shape*, é uma função semilinear, limitada e monotônica. Várias funções sigmoideais podem ser definidas. Uma das formas que esta equação pode tomar é a função logística, segundo a definição abaixo; onde y é a saída, x a entrada e T determina a suavidade da curva.

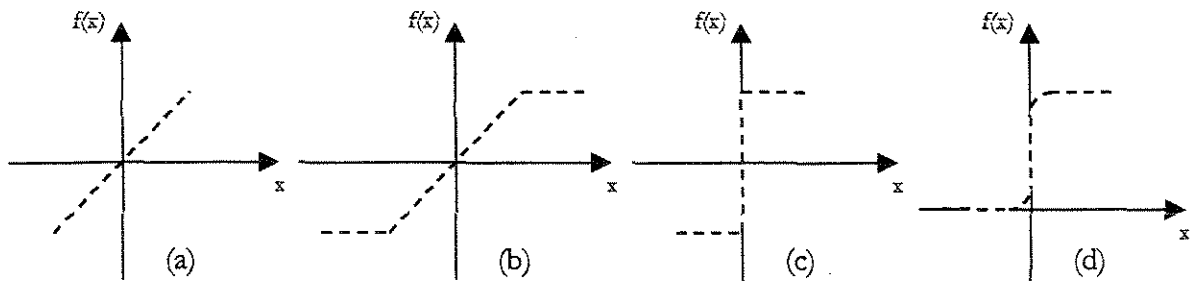


Figura 2.11 - Algumas Funções de Ativação.

$$y = \frac{1}{1 + \exp\left(\frac{-x}{T}\right)} \quad (2.17)$$

As funções linear e rampa são semelhantes, como pode ser visto nas Figuras 2.11 a e b, respectivamente. A primeira é definida pela Equação 2.18, onde α é um número real que determina a saída linear para os valores de entrada. A diferença entre ambas é que a função rampa está restringida para produzir valores constantes em uma faixa $[-\gamma, +\gamma]$, conforme a Equação 2.19.

$$y = \alpha \cdot x \quad (2.18)$$

$$y = \begin{cases} +\gamma, & x \geq +\gamma \\ x, & |x| < +\gamma \\ -\gamma, & x \leq -\gamma \end{cases} \quad (2.19)$$

A função passo ou degrau, Figura 2.11-c, produz uma saída $+\gamma$ para os valores de x maiores que zero e um valor $-\gamma$ para o caso contrário. Esta função é definida pela Equação 2.20.

$$y = \begin{cases} +\gamma, & x > 0 \\ -\gamma, & x \leq 0 \end{cases} \quad (2.20)$$

2.4.6 Minimização de Erros

Embora o algoritmo *backpropagation* leve a bons resultados, algumas características delimitam o seu uso, como por exemplo a velocidade de treinamento para algumas superfícies mais complexas. Uma maneira de minimizar este aspecto, segundo BRAGA et al. (2000), é

considerar efeitos de segunda ordem para o gradiente descendente. Algumas das técnicas que podem tanto acelerar o algoritmo quanto diminuir a incidência de mínimos locais são:

- Utilizar taxa de aprendizado decrescente;
- Adicionar ou retirar nós (neurônios) intermediários;
- Utilizar um termo *momentum*;
- Adicionar ruídos aos dados (CAUDIEL; 1991).

Das técnicas apresentadas, a mais empregada, no entanto, é a adição do termo *momentum* à atualização dos pesos pela rede. Sua representação é dada pela equação a seguir:

$$\psi = \alpha(w_{ji}(t) - w_{ji}(t-1)) \quad (2.21)$$

A fórmula completa de ajuste de pesos utilizando o termo *momentum* fica:

$$w_{ji}(t+1) = w_{ji}(t) + \eta \delta_j(t) x_i(t) + \alpha(w_{ji}(t) - w_{ji}(t-1)) \quad (2.22)$$

A adição deste termo, além de acelerar o treinamento em regiões muito planas da superfície de erro, suprime a oscilação de pesos em vales e ravinas.

Durante o treinamento da rede pode ocorrer outro problema conhecido como *overfitting*, ou seja, isto ocorre quando após um período de treinamento a rede começa a piorar a sua taxa de acertos para os padrões diferentes daqueles utilizados para o ajuste de pesos. Compreende-se então que a rede se acostumou com os padrões de treinamento, gravando sua peculiaridade de ruídos. Conseqüentemente há um detrimento na capacidade de generalização da rede. As alternativas de solução são as seguintes:

- Encerrar o treinamento mais cedo (quando o erro começa a subir);
- Podar (*prunning*) os pesos da rede.

Outro problema que pode ser observado é conhecido como superfície plana, *flat spot*, ou seja, quando a derivada da função sigmoidal de uma unidade se aproxima de zero durante o treinamento. O ajuste de pesos utiliza o valor da derivada, e, portanto, quando a saída da unidade se aproxima de 0.0 ou 1.0, a derivada da função sigmoidal se aproxima de zero. Assim a unidade não pode ter seus pesos ajustados, ou estes são ajustados com um valor muito baixo, o que em consequência retarda o treinamento da rede.

BRAGA et al. (2000), apresentam, no entanto, algumas soluções para este problema:

- Utilizar uma medida de erro que tenda para infinito quando a derivada da sigmoideal tender para zero;
- Adicionar uma constante à derivada, evitando que o erro seja igual a zero;
- Utilizar uma função de erro não-linear.

Uma RNA pode ainda sofrer o problema de *esquecimento catastrófico*, quando ao aprender novas informações, a rede “esquece” aquelas já aprendidas.

Outra sugestão que tenta melhorar o treinamento da rede é a variação da taxa de aprendizado, inicialmente grande, a taxa deve ser reduzida quando a rede parar de diminuir o erro quadrático médio.

Quanto à técnica de *prunning* citada acima, envolve a eliminação de pesos e nodos irrelevantes para a função executada pela rede. Isto é verificado quando após a retirada daquele determinado elemento não ocorre uma grande variação no erro de saída. A desvantagem, como não poderia deixar de existir, é inerente ao fato de que a retirada dos elementos não considera a correlação entre os mesmos. Um exemplo é o caso de dois nodos que anulam as suas saídas em uma determinada camada da rede; como par estes nodos não causam variação à rede, entretanto individualmente podem ter um grande efeito sobre a saída.

2.5 Modelagem Híbrida com Redes Neurais Artificiais

Muito embora as RNAs tenham surgido como uma ferramenta eficaz para a solução de uma grande quantidade de problemas, não é correto afirmar que elas são suficientes para solucionar qualquer problema de inteligência artificial.

Parafraseando o ditado popular que diz: “Duas cabeças pensam melhor do que uma”; a idéia por trás do desenvolvimento de sistemas híbridos é que devido a certas limitações uma técnica pode não ser capaz, por conta própria, de fornecer a solução a um certo problema. Assim, a combinação de duas ou mais técnicas pode levar a uma solução melhor, mais eficiente. Entretanto é importante notar que nem em todos os casos a aplicação de Sistemas Neurais Híbridos (SNH) pode melhorar o desempenho do sistema como um todo sobre aquele obtido pelas técnicas que foram combinadas quando utilizadas isoladamente.

Algumas das formas que a modelagem híbrida neural pode ter, envolvem a combinação dos seguintes sistemas com RNAs.

-
- Estatística;
 - Lógica Fuzzy;
 - Sistemas Baseados em Conhecimento;
 - Algoritmos Genéticos;
 - Raciocínio Baseado em Casos;
 - Lógica Matemática;
 - Linguagens Formais;
 - Agentes Inteligentes;
 - Sistemas Tutores;
 - Linguagem Natural;
 - Equações Determinísticas.

O desenvolvimento e aplicação de um SNH, utilizando equações determinísticas (equações de balanço, por exemplo) oferece algumas vantagens sobre o sistema convencional de redes neurais. Por exemplo, os dados de saída são analisados mais facilmente, de acordo com STUART et al. (1997), o conhecimento de um modelo físico estabelecido a priori, e sua combinação com redes neurais assegura a interpretação física correta dos dados de saída. São necessários também, uma quantidade menor de dados para o treinamento da rede e como consequência disto o tempo de treinamento é geralmente menor. Devido ao fato da incerteza ficar embutida em uma parte da rede somente, os SNHs interpolam e extrapolam melhor os resultados, além de permitirem um melhor uso da rede para a forma adaptativa, já que o esforço para reestimação dos pesos envolvidos na rede é menor.

Desta forma, os sistemas híbridos são melhor empregados quando existe, por exemplo, alguma etapa do processo que possa ser muito obscura. Conforme já comentado, é o caso do comportamento microbiano, pois devido à natureza complexa do seu metabolismo e a natureza altamente não linear da sua cinética, os aspectos do problema cujo comportamento quantitativo é bem entendido são descritos pelas equações determinísticas, enquanto os demais são descritos por redes neurais. A rede neural não apresenta, em geral, dificuldades em estabelecer as relações sobre os aspectos muitas vezes complexos.

PSICHOGIOS & UNGAR (1992) desenvolveram um sistema neural híbrido usando equações determinísticas para modelar um biorreator em batelada alimentada. Segundo os autores, este modelo apresentou melhores resultados do que a rede neural artificial isolada, conhecida como *black-box* (modelo caixa-preta), pois os modelos híbridos desenvolvidos foram capazes de interpolar e extrapolar melhor os dados fornecidos para os testes do modelo, além de serem mais fáceis de analisar e interpretar, requerendo inclusive uma menor quantidade de pares de exemplos para o treinamento. Ainda segundo o artigo, o modelo híbrido é mais confiável, pois é construído também a partir de modelos físicos como, por exemplo, as equações de balanço para o sistema em questão.

TESSIER et al. (1997) desenvolveram um modelo de redes neurais recorrentes para acompanhar o crescimento do fermento em um meio vinho-base. Este modelo possibilitou a estimativa e predição da concentração de fermento em culturas em batelada, baseado em medidas on-line do volume de CO₂ gerado e na concentração inicial do fermento. O erro médio do valor predito ao final da concentração de fermento foi, segundo o artigo, menor do que 5%. Um modelo híbrido combinando este modelo com outro convencional, baseado em correlações lineares refletindo o esquema da reação, também levou a estimativa e predição das concentrações de açúcar e etanol no meio de cultivo com erros médios de 1,6 e 1 g/l, respectivamente. Posteriormente os autores utilizaram estes modelos para complementar uma estratégia de controle do sistema em estudo. No artigo ainda reforçam que a aplicação industrial deste procedimento de modelagem apresenta grandes vantagens, pois pode prever o comportamento da fermentação antecipadamente e indicar como os tanques devem ser alimentados.

SHENE et al. (1999) compararam a modelagem do sistema de fermentação alcoólica com *Zymomonas mobilis* CP4, em batelada, usando redes neurais do tipo “caixa-preta” e modelo híbrido, ambos com estrutura *feedforward*. Os autores empregaram dados experimentais para os treinos e testes com as redes. Os melhores resultados da rede foram com 10 nodos na camada intermediária. Redes treinadas para cada saída deram resultados melhores, além de apresentarem treinamento mais fácil e rápido, pois necessitaram apenas de 5 nodos na camada intermediária. Para o modelo híbrido, desenvolvido com equações de balanço, redes com 30 nodos forneceram bons resultados, mas o erro médio foi maior do que para o primeiro tipo de redes.

2.6 Otimização

Em seu livro "*Optimization of Chemical Processes*", em 1989, EDGAR & HIMMELBLAU, declaram que a ênfase em um futuro próximo seria a melhora da eficiência e aumento da produtividade das plantas de processos existentes, em lugar da expansão destas. Esta ênfase seria impulsionada, entre outros fatores, pelo aumento do custo da energia e exigências ambientais. E a otimização seria uma das mais importantes ferramentas da engenharia para atingir estes objetivos.

Uma forma de otimizar um processo é usando a metodologia das Superfícies de Resposta. As etapas que devem ser seguidas objetivando a otimização de um sistema por esta metodologia podem ser resumidas a seguir (Fonte: Profª. Dra. Maria Isabel Rodrigues; FEA – Unicamp):

1. Analisar profundamente o processo em estudo:
 - a. Determinar quais são as variáveis ou fatores controlados pelo pesquisador;
 - b. Verificar as restrições do processo (se houver);
2. Definir claramente que objetivo se pretende alcançar com os experimentos;
 - a. Função objetivo;
3. Elaborar o planejamento experimental ou fatorial levando em consideração todas as variáveis independentes definidas no item 1;
4. Realizar os experimentos e obter as respostas;
5. Elaborar o modelo das respostas (variáveis dependentes) em função das variáveis independentes;
6. Verificar a validação do modelo (testes estatísticos) → checar as respostas;
7. Empregar algum método de otimização ao modelo elaborado;
8. Definir as faixas ótimas operacionais de cada variável do processo.

Portanto, segundo as etapas acima, um planejamento experimental se faz necessário para gerar dados suficientes para a elaboração do modelo. O planejamento é importante, pois realizar a obtenção dos dados, aleatoriamente, sem critérios, pode gerar números estéréis

incapazes de representar adequadamente o processo. Assim o planejamento experimental consiste em projetar um experimento de forma que ele seja capaz de fornecer exatamente o tipo de informação procurada (BARROS NETO et al., 1995). Além disso, o planejamento pode se mostrar extremamente econômico, uma vez que evita o desperdício de tempo e materiais na elaboração de experimentos que não forneceriam respostas representativas do processo. Entre os vários tipos de planejamentos existentes, o chamado *Fatorial Completo* permite atingir estes objetivos, para a elaboração do modelo, com o mínimo de experimentos.

2.6.1 Planejamento Fatorial Completo, 2^k

O planejamento consiste em realizar experimentos para todas as possíveis combinações dos níveis dos fatores estudados. Os níveis são os valores assumidos pelos fatores, enquanto estes últimos são as variáveis independentes investigadas no processo. Assim, se houver n_1 níveis do fator 1, n_2 níveis do fator 2, ..., e n_k do fator k, o planejamento será um fatorial $n_1 \times n_2 \times \dots \times n_k$. Isso não implica que serão realizados apenas $n_1 \times n_2 \times \dots \times n_k$ experimentos, no entanto, este é um número mínimo para que exista um planejamento fatorial completo.

É possível avaliar o efeito de qualquer variável sobre a resposta com um número menor de ensaios, pois como a resposta é avaliada observando-se a variação dos fatores e seus resultados, e para que isto ocorra é necessário que sejam feitos ensaios em pelo menos dois níveis de cada fator, então é possível realizar um planejamento em que todas as variáveis são estudadas em apenas dois níveis, geralmente máximo e mínimo de cada fator. Este é o planejamento mais simples de todos. Assim havendo k variáveis controladas, fatores, o planejamento de dois níveis irá requerer a realização de $2 \times 2 \times \dots \times 2 = 2^k$ ensaios diferentes, sendo por isso chamado de *planejamento fatorial 2^k* .

Uma vez realizados os ensaios e obtidos os dados, estes podem ser ajustados a um modelo, equação linear ou quadrática, representando o processo. Uma equação quadrática em função de duas variáveis pode ser representada como se segue,

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + b_{11}x_1^2 + b_{22}x_2^2 + b_{12}x_1x_2 \quad (2.23)$$

portanto com seis parâmetros. Para casos onde a quantidade de parâmetros é maior do que o número de "níveis" ou combinações de valores entre os níveis de cada fator, não é possível determinar as estimativas, \hat{y} , logo o planejamento precisa ser ampliado.

Esta ampliação segue o chamado *Planejamento em Estrela*. Para ser feito este planejamento, simplesmente é acrescentado àquele já existente um planejamento idêntico, porém girado de 45 graus em relação à orientação de partida. O resultado é uma distribuição octogonal como mostra a Figura 2.12-b, para o caso de duas variáveis.

Vale ressaltar que no planejamento de dois níveis, quando as variáveis envolvidas apresentam grandezas de ordens diferentes, costuma-se identificar os níveis superior e inferior com (+1) e (-1), respectivamente, e assim codificar todos os níveis presentes. A saída \hat{y} não precisa ser codificada.

Em cada planejamento, geralmente, o ponto localizado no centro do intervalo dos valores estudados é feito em duplicata para que seja possível se ter uma estimativa do erro experimental, e a partir daí avaliar a significância estatística dos efeitos.

Pode ser observado da Figura 2.12-b que os pontos devido à distribuição octogonal do planejamento estrela, estão a uma distância de 1,4142 do ponto central (em variáveis codificadas), para um planejamento com dois fatores. Para um planejamento com três fatores esta distância é de 1,6818.

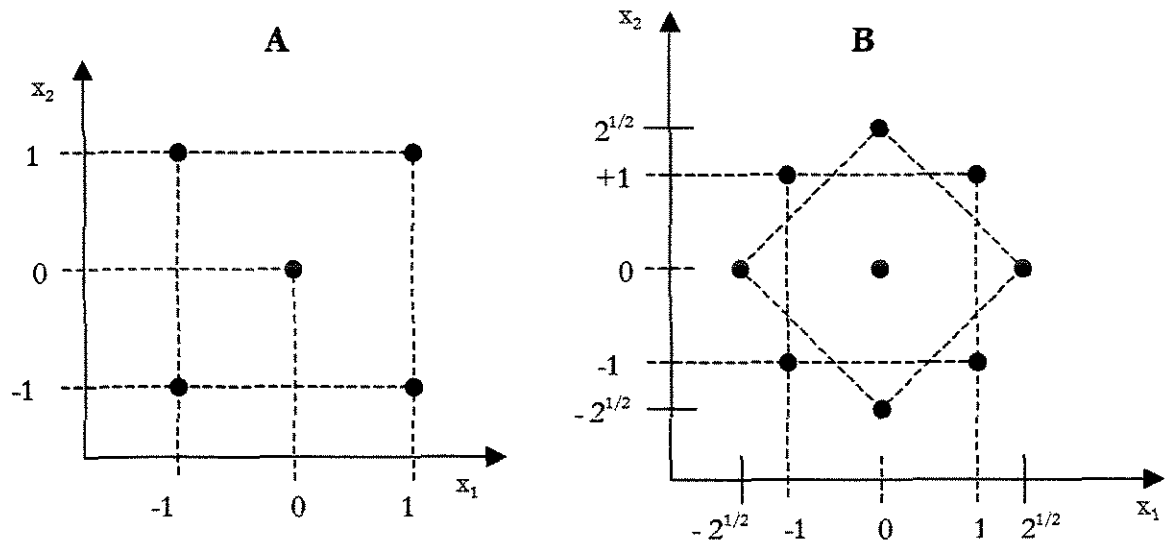


Figura 2.12 - a) Planejamento Fatorial de dois níveis com ponto central
b) Planejamento em Estrela. Cada ponto representa um ensaio.

2.6.2 Método das Superfícies de Resposta (RSM)

De acordo com BARROS NETO et al. (1995), iniciada na década de 1950, este método (*Response Surface Methodology – RSM*) é uma técnica de otimização baseada na utilização

do planejamento fatorial. A técnica é constituída de duas etapas: modelagem e deslocamento. Estas são repetidas o número de vezes necessárias para se atingir uma região ótima de máximo ou mínimo da superfície investigada. A primeira etapa geralmente é feita a partir do ajuste de modelos lineares ou quadráticos aos resultados experimentais obtidos a partir de planejamentos fatoriais. Já a direção do deslocamento é escolhida sempre ao longo do caminho de máxima inclinação de um determinado modelo, sendo esta a trajetória onde a resposta varia de forma mais pronunciada.

Segundo KALIL et al. (2000), em fermentação, especialmente fermentação alcoólica em escala industrial, pouca informação é disponível sobre o uso das técnicas de otimização através do planejamento fatorial e análise das superfícies de resposta, e quando acompanhado de modelagem e simulação para a determinação das respostas ótimas. Os autores, portanto, utilizaram estas técnicas para otimizar um bioprocessos industrial, fermentação alcoólica em múltiplos estágios com reciclo de células. Entre os vários fatores estudados para desenvolver os modelos, os autores avaliaram quais eram aqueles relevantes para o rendimento e produtividade. Os modelos foram desenvolvidos, portanto, com cinco parâmetros estatisticamente significativos para cada tipo de saída desejada. O ponto ótimo obtido para a produtividade representou um aumento de 52% em relação aos dados sem otimização, no entanto sem alterar o alto rendimento previamente apresentado.

ÖZDEMİR & DEVRES (2000) empregaram a metodologia das superfícies de resposta para analisar o processo de tostagem de nozes de avelã, com a finalidade de avaliar o efeito das variáveis do processo sobre o desenvolvimento da cor, e para desenvolver modelos. O método indicou a temperatura como principal fator para o desenvolvimento da cor, enquanto o modelo obtido descreveu satisfatoriamente o problema estudado, em função da temperatura de tostagem e tempo de exposição para as três dimensões da amêndoa, temperaturas da superfície e interna.

COSTA et al. (2001) mostraram com o planejamento e otimização das variáveis do processo, que as condições de operação possuem uma forte influência sobre o desempenho do algoritmo de controle. Para tanto empregaram a metodologia das superfícies de resposta combinada com modelagem e simulação do processo para determinar as condições de operação e maximizar o rendimento e a produtividade de uma fermentação alcoólica extrativa.

IBANOGLU & IBANOGLU (2001) realizaram a fermentação de uma mistura água-farinha do feijão caupi (“cowpea”, *Vigna unguiculata*), fermentado por sua microflora natural em diferentes temperaturas e tempos. A metodologia das superfícies de resposta foi empregada para estabelecer a relação entre as respostas da fermentação (% acidez como ácido lático e o pH) e as variáveis de fermentação (tempo e temperatura). Como resultado o tempo mostrou-se mais significativo do que a temperatura da fermentação, sugerindo que este deveria ser melhor controlado.

2.6.3 Programação Quadrática Sucessiva

A Programação Quadrática (PQ) é o nome dado ao procedimento que minimiza uma função quadrática de n variáveis sujeitas a m restrições lineares de desigualdades, igualdades, ou ambas (EDGAR & HIMMELBLAU, 1989).

De acordo com CONRADIE et al. (1998), a Programação Quadrática Sucessiva (Successive Quadratic Programming – SQP) é um dos algoritmos mais eficientes para a solução de problemas de otimização, pequenos ou médios, não-lineares com restrições. Em um ambiente computacional onde a avaliação de uma função pode despendar um “gasto” equivalente a uma simulação completa, os métodos de SQP são, em geral, favorecidos devido à pequena quantidade de funções e gradientes avaliados.

O método SQP básico resolve uma sequência de problemas de programação quadrática. A solução de cada problema promove um passo em direção ao novo ponto. O modelo do sistema é resolvido simultaneamente com as condições ótimas de Kuhn-Tucker (1ª. e 2ª. ordem) para dar a solução do problema de programação não-linear. Estas condições são satisfeitas na solução otimizada e o código neste ponto encontrará: a função objetivo otimizada; um possível ponto x , as variáveis que fornecem a solução ótima; e os multiplicadores de Lagrange, que são fracamente positivos para os limites inferiores ativos, fracamente negativos para os limites superiores ativos e zero para as restrições inativas. O Hessiano reduzido do Lagrangiano no espaço nulo das restrições ativas (equações e limites ativos) devem ser definidos positivos para uma restrição mínima local (RODRIGUEZ-TORAL et al., 2001).

As equações 2.24 a 2.27, ilustrando uma etapa do método SQP, foram retiradas do artigo de CONRADIE et al. (1998). A cada iteração do método, os valores de (x,y) são fixados e o sub-problema de programação quadrática apresenta a seguinte forma:

$$\text{Minimizar, } \nabla f(x, y)^T s + 0.5s^T Bs \quad (2.24)$$

sujeito às restrições linearizadas

$$\begin{aligned} c_i(x, y) + \nabla c_i(x, y)^T s &= 0, i = 1, \dots, m_{eq} \\ c_i(x, y) + \nabla c_i(x, y)^T s &\geq 0, i = m_{eq} + 1, \dots, m_i \\ h(x, y) + \nabla h(x, y)^T s &= 0 \end{aligned} \quad (2.25)$$

onde s é a direção de procura e B uma aproximação à matriz Hessiana da função de Lagrange para o problema do artigo em questão. A convexidade é garantida calculando B usando o algoritmo BFGS, discutido no item 4.2.2. A cada iteração, as seguintes condições ótimas de Kuhn-Tucker (1ª. ordem) precisam ser atendidas para o sub-problema de programação quadrática:

$$\begin{aligned} \nabla f(x, y) + Bs - \nabla c(x, y)u - \nabla h(x, y)v &= 0 \\ c_i(x, y) + \nabla c_i(x, y)^T s &= 0, i \in A \\ u_i [c_i(x, y) + \nabla c_i(x, y)^T s] &= 0, i = m_{eq} + 1, \dots, m_i \\ u_i &\geq 0, i = m_{eq} + 1, \dots, m_i \\ h(x, y) + \nabla h(x, y)^T s &= 0, \end{aligned} \quad (2.26)$$

onde v e u são os multiplicadores de Lagrange da igualdade e das restrições em geral, respectivamente, e A é o conjunto de restrições gerais, as quais são ativas durante esta iteração (satisfeitas como igualdades). O conjunto ativo consiste de m restrições gerais ativas. A condição de 1ª. ordem necessária para o sub-problema de PQ pode ser apresentado em uma matriz na seguinte forma:

$$\begin{bmatrix} B & -\nabla c_A & -\nabla h \\ -\nabla c_A^T & 0 & 0 \\ -\nabla h^T & 0 & 0 \end{bmatrix} \begin{bmatrix} s \\ u \\ v \end{bmatrix} = \begin{bmatrix} -\nabla f \\ c_A \\ h \end{bmatrix} \quad (2.27)$$

onde c_A representa o vetor geral de restrições ativas a cada iteração, e ∇c_A e ∇h são matrizes $n \times m$ e $n \times r$, respectivamente. B é uma matriz $n \times n$ densa que deve ser atualizada, fatorada e armazenada a cada iteração.

CONRADIE et al. (1998) estudaram técnicas de otimização, SQP e Decomposição da SQP, para obter um melhor custo para o projeto e desempenho de sistemas de arrefecimento com aplicação em plantas de energia. Os autores mostraram que a correta e cuidadosa aplicação dessas estratégias de otimização para modelos complexos, não precisa ser difícil ou demorada. Os resultados ótimos e as tendências que podem ser obtidas, por meios de um procedimento computacional, são discutidas. Os autores estabelecem um novo desafio em práticas de projeto e desempenho para projetistas, fabricantes e operadores, relacionando o nível de otimização desejada em função da sofisticação do programa de projeto, experiência do usuário, procedimento da estimativa de custo além da qualidade dos dados fornecidos.

Dada uma matriz real positiva definida simétrica $n \times n$ (matriz F), quanto pode ser subtraído da diagonal de F e ainda manter uma matriz positiva-semi-definida? AL-HOMIDAN (1998), chama este problema de Problema do Teste Educacional e denota que um problema como este pode surgir em várias situações práticas, como por exemplo, em estatística onde existindo uma matriz F normalmente uma matriz da covariância com os elementos sofrendo variação. Segundo o autor existem aproximações para a solução, entretanto, estes métodos convergem lentamente, mas podem determinar em poucas iterações a correta posição na matriz solução. A partir daí o artigo demonstra como formular o problema como uma suave minimização não-linear, para a qual uma rápida convergência pode ser obtida a partir do método SQP. Então é proposto um método híbrido combinando as melhores características tanto das aproximações previamente conhecidas, como a programação quadrática sucessiva. Os resultados obtidos demonstram que este método desenvolvido consegue melhores resultados para o problema estudado.

CHUAPRASERT et al. (1999) realizaram a reconciliação dos dados de uma planta piloto de um Evaporador com Filme agitado para a concentração de açúcar, usando o sistema *AspenPlus* de simulação. A simulação foi desenvolvida com o rigoroso modelo de trocador de calor *Heatx*, e o modelo de duas fases, *Flash2*. Para a otimização os autores empregaram o algoritmo SQP. A reconciliação dos dados melhorou o ajuste entre os dados experimentais e o modelo simulado entre 30 a 98%, garantindo que as medidas ajustadas do processo satisfizessem os balanços de massa e energia.

3 O PROCESSO FERMENTATIVO E A MODELAGEM MATEMÁTICA

Como dito previamente, este trabalho pretende modelar um processo de fermentação alcoólica. Os dados experimentais utilizados foram obtidos por COSTA (2000) e ATALA (2000). Segundo esse trabalho, foram realizados experimentos para determinar a influência da temperatura na cinética do processo de fermentação alcoólica, uma vez que a maioria dos estudos têm como base um processo isotérmico, o que não corresponde à verdade. Existe grande dificuldade em se manter a temperatura constante nas domas de fermentação, pois na realidade a fermentação é um processo exotérmico e alterações relativamente pequenas na temperatura (2-4°C) podem deslocar o processo das condições ótimas de operação. Para tanto, foram medidas tanto variáveis primárias, como as concentrações de biomassa, substrato e produto, quanto variáveis secundárias: turbidez, pH e Brix. Uma descrição breve da parte experimental segue nas próximas páginas.

3.1 Trabalho Experimental (COSTA, 2000; ATALA, 2000)

Os experimentos e coleta de dados foram realizados nas seguintes temperaturas: 28, 31, 34, 37 e 40°C.

3.1.1 Material

Microorganismo:

Saccharomyces cerevisiae;

Composição do meio de manutenção: 3 g/l de extrato de levedura; 3 g/l de extrato de malte; 5 g/l de peptona; 10 g/l de glicose; 20 g/l de agár.

O meio é esterilizado em autoclave a 121°C por 15 minutos.

Meio de Cultura:

O meio de cultura empregado na fermentação é melaço de cana-de-açúcar diluído e enriquecido com 1g/l de extrato de levedura e 2,4 g/l de sulfato de amônio. A esterilização também é feita a 121°C em autoclave por 30 minutos.

Inóculo:

O microrganismo antes de ser inoculado deve ser ativado, sendo transferido do meio de manutenção para um meio de ativação com a composição semelhante ao primeiro, inclusive sendo esterilizado pelo mesmo tempo e temperatura. O meio de ativação contendo o microrganismo é incubado a 30°C por 24h. O inóculo é então preparado adicionando-se células de levedura do meio de ativação a 200 ml do seguinte meio de cultivo: 50 g/l de glicose, 5 g/l KH_2PO_4 , 1,5 g/l de NH_4Cl , 0,7g/l de $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 1,2 de KCl, 5 g/l de extrato de levedura, pH ajustado em 6 com ácido fosfórico. O inóculo é agitado em shaker a 200 rpm por 24 horas em uma temperatura de 30°C. Ainda segundo COSTA (2000), a transferência do microrganismo de um meio para outro é feita em câmara de fluxo laminar para manter a esterilidade.

Uma descrição completa da unidade experimental não será dada aqui pois ultrapassa o objetivo deste trabalho. Entretanto a descrição detalhada do equipamento pode ser encontrada no trabalho utilizado como referência.

Quanto aos métodos empregados nas análises, para as medidas de biomassa, substrato, etanol e viabilidade, estão colocados em anexo no item 8.1.

3.2 Modelagem Matemática e Parâmetros Cinéticos

A modelagem realizada na parte experimental da fermentação em batelada fez algumas considerações para o desenvolvimento do modelo não-estruturado:

- O reator é de mistura perfeita;
- Reator de volume constante;
- Densidade constante.

A equação empregada para representar a taxa de crescimento é descrita como:

$$r_X = \mu_{\max} \frac{S}{K_s + S} \exp(-K_i S) \left(1 - \frac{X}{X_{\max}}\right)^m \left(1 - \frac{P}{P_{\max}}\right)^n X \quad (3.1)$$

De acordo com SIQUEIRA FILHO (1997), citado por COSTA (2000), a equação acima descreve com sucesso os dados experimentais de processos descontínuos.

A taxa de consumo de substrato tem a seguinte forma:

$$r_s = \frac{r_x}{Y_x} + m_x \cdot X \quad (3.2)$$

E a taxa de produção de etanol é dada por,

$$r_p = Y_{px} \cdot r_x + m_p \cdot X \quad (3.3)$$

As equações acima consideram que não ocorre morte celular, uma vez que, segundo os dados experimentais, durante a batelada a viabilidade estava sempre próxima a 100%.

As equações resultantes dos balanços de biomassa, substrato e produto são:

$$\frac{dX}{dt} = r_x = \mu \cdot X \quad (3.4)$$

$$\frac{dS}{dt} = -r_s = \sigma \cdot X \quad (3.5)$$

$$\frac{dP}{dt} = r_p = \pi \cdot X \quad (3.6)$$

Os parâmetros em função da temperatura, ajustados a partir dos dados experimentais, são mostrados na Tabela 3.1 para a fermentação em batelada.

Tabela 3.1 – Parâmetros em função da Temperatura em °C.

PARÂMETROS	BATELADA
$\mu_{\max} \text{ (h}^{-1}\text{)}$	$1,57 \exp\left(\frac{-41,47}{T}\right) - 1,29 \cdot 10^4 \exp\left(\frac{-431,4}{T}\right)$
$X_{\max} \text{ (g/l)}$	$-0,3279 \cdot T^2 + 18,484 \cdot T - 191,06$
$P_{\max} \text{ (g/l)}$	$-0,4421 \cdot T^2 + 26,41 \cdot T - 279,75$
$Y_x \text{ (g/g)}$	$1,4847 \exp(-0,1005 \cdot T)$
$Y_{px} \text{ (g/g)}$	$-0,0048 \cdot T^2 + 0,8571 \cdot T - 14,292$
$K_s \text{ (g/l)}$	4,1
$K_i \text{ (l/g)}$	0,002
$m_p \text{ (g/(g.h))}$	0,1

m_x (g/(g.h))	0,2
M	1
N	1,5

(Fonte: COSTA, 2000; ATALA, 2000)

4 REDES NEURAIS ARTIFICIAIS, DESENVOLVIMENTO E APLICAÇÃO

Para seguir o objetivo do trabalho, devido à pequena quantidade de dados experimentais disponíveis, estes foram ajustados a curvas e em seguida interpolados, a fim de ser aumentado o conjunto de valores disponíveis para o treinamento.

Quanto ao tipo de redes neurais artificiais implementadas, foram testadas algumas arquiteturas, bem como diferentes algoritmos de treinamento, todos, entretanto, com o método *Backpropagation* para atualização dos pesos das conexões, isto devido à grande disponibilidade de algoritmos que empregam este método, além de ser muito mencionado na literatura.

4.1 Organização e Análise dos Dados Experimentais

Como esperado, a partir dos dados experimentais as variáveis de estado são disponíveis; no entanto, as taxas cinéticas, que foram escolhidas como as saídas das redes, não são mensuráveis, mas podem ser estimadas pela discretização das equações de balanço de biomassa, substrato e produto, equações 3.4, 3.5 e 3.6 (HARADA, 2001).

Entretanto, devido à pequena quantidade de dados, bem como aos intervalos entre os mesmos serem relativamente grandes, a simples discretização não forneceu bons resultados. Portanto, a alternativa encontrada foi o ajuste de uma curva sobre os dados experimentais de forma que foram obtidas expressões de X, S e P em função do tempo. Posteriormente este ajuste permitiu tanto a interpolação das curvas gerando novos dados quanto a discretização dos valores interpolados para se chegar aos valores das taxas. Ainda, a própria derivação da equação ajustada pôde ser usada para calcular as taxas desejadas, ou seja, como mostrado pelas equações $dX/dt=r_x$, $dS/dt=-r_s$ e $dP/dt=r_p$.

Os dados experimentais utilizados, que estão em anexo nas Tabelas 8.1 a 8.5, não consideram a fase *lag* dos microrganismos.

Duas curvas foram ajustadas aos dados experimentais, a fim de fornecerem tanto as taxas quanto uma maior quantidade de dados para o treinamento das redes.

O primeiro ajuste de curvas feito trata-se de um ajuste polinomial. Esta escolha foi feita por se tratar de um ajuste disponível mais facilmente em *softwares* de planilhas eletrônicas, neste caso MS-Excel[®]. Os valores de dados, biomassa (X), substrato (S) e produto (P) foram, portanto, plotados contra o tempo (eixo das abscissas) e polinômios foram ajustados para a

curva de cada temperatura, como mostrado na Tabela 4.1, onde y são os polinômios ajustados e y' as suas derivadas.

Tabela 4.1 - Polinômios ajustados para os dados de cada temperatura e suas derivadas.

TEMPE- RATURA (°C)	VARIAVEIS	POLINOMIO
28	Biomassa	$y = 1,8602 + 0,14614 \cdot x + 0,06095 \cdot x^2 - 0,00207 \cdot x^3$ $y' = 0,14614 + 0,1219 \cdot x - 0,00621 \cdot x^2$
	Substrato	$y = 179,14167 + 2,89392 \cdot x - 1,47996 \cdot x^2 + 0,04599 \cdot x^3$ $y' = 2,89392 - 2,95992 \cdot x + 0,13797 \cdot x^2$
	Produto	$y = 11,56209 - 0,28779 \cdot x + 0,34884 \cdot x^2 - 0,00845 \cdot x^3$ $y' = 0,28779 + 0,69768 \cdot x - 0,02535 \cdot x^2$
31	Biomassa	$y = 1,73497 + 0,32625 \cdot x + 0,06293 \cdot x^2 - 0,00251 \cdot x^3$ $y' = 0,32615 + 0,12586 \cdot x - 0,00753 \cdot x^2$
	Substrato	$y = 216,89154 - 3,48661 \cdot x - 1,10648 \cdot x^2 + 0,03911 \cdot x^3$ $y' = -3,48661 - 2,21296 \cdot x + 0,11733 \cdot x^2$
	Produto	$y = 1,58315 + 2,83944 \cdot x + 0,28853 \cdot x^2 - 0,00939 \cdot x^3$ $y' = 2,83944 + 0,57706 \cdot x - 0,02817 \cdot x^2$
34	Biomassa	$y = 0,77987 + 0,99745 \cdot x - 0,02642 \cdot x^2$ $y' = 0,99745 - 0,05284 \cdot x$
	Substrato	$y = 247,41924 - 9,96924 \cdot x - 0,48236 \cdot x^2 + 0,02173 \cdot x^3$ $y' = -9,96924 - 0,96472 \cdot x + 0,06519 \cdot x^2$
	Produto	$y = -0,86949 + 2,0529 \cdot x + 0,43687 \cdot x^2 - 0,01382 \cdot x^3$ $y' = 2,0529 + 0,87374 \cdot x - 0,04146 \cdot x^2$
37	Biomassa	$y = -0,20897 + 0,59521 \cdot x - 0,01157 \cdot x^2$ $y' = 0,59521 - 0,02314 \cdot x$
	Substrato	$y = 277,15253 - 5,93799 \cdot x - 0,45044 \cdot x^2 + 0,01327 \cdot x^3$ $y' = 5,93799 - 0,90088 \cdot x + 0,03981 \cdot x^2$
	Produto	$y = -1,09501 + 1,65573 \cdot x + 0,21079 \cdot x^2 - 0,00554 \cdot x^3$ $y' = 1,65573 + 0,42158 \cdot x - 0,01662 \cdot x^2$
40	Biomassa	$y = 1,07444 + 0,39054 \cdot x - 0,00928 \cdot x^2$ $y' = 0,39054 - 0,01856 \cdot x$
	Substrato	$y = 250,52578 - 12,73328 \cdot x + 0,21212 \cdot x^2$ $y' = -12,73328 + 0,42424 \cdot x$
	Produto	$y = -3,66565 + 3,85487 \cdot x - 0,04823 \cdot x^2$ $y' = 3,85487 - 0,09646 \cdot x$

O ajuste das curvas experimentais com o modelo polinomial não levou a bons resultados, principalmente quando se derivam as funções polinomiais para a obtenção das taxas cinéticas. Neste caso os polinômios ajustados apresentaram curvas que não são capazes de descrever o comportamento correto da fermentação, segundo os dados experimentais. Um ajuste polinomial inferior ao 3º grau representava os dados muito grosseiramente, pois não acompanhava a curva experimental corretamente, enquanto um ajuste de 3º grau não era capaz de descrever com fidelidade os extremos da função. As Figuras 4.1 a 4.5 mostram os ajustes para os dados de concentração de biomassa, substrato e produto e as Figuras 4.6 a 4.10 mostram as taxas cinéticas calculadas a partir destes ajustes. De uma forma geral, pode ser observado que o ajuste polinomial não foi adequado para interpolar os valores experimentais.

Neste ponto faz-se necessário observar que todos os ajustes feitos sobre os dados experimentais foram comparados com o modelo determinístico apresentado por COSTA (2000) e ATALA (2000), descrito pelas equações 3.1 a 3.6, utilizando os parâmetros da Tabela 3.1. Isto foi feito para que, no caso do cálculo das taxas cinéticas, houvesse um parâmetro de comparação, pois de outra forma não seria possível determinar a melhor curva, já que a discretização dos dados experimentais diretamente leva a resultados ruins devido ao grande intervalo de amostragem.

O segundo ajuste de curvas feito utilizou a função Sigmoidal-Boltzman, que possui a seguinte forma:

$$y = \frac{A_1 - A_2}{1 + \exp\left(\frac{t - x_0}{dx}\right)} + A_2 \quad (4.1)$$

Na equação 4.1 as concentrações (y) eram plotadas em função do tempo (t), sendo os demais parâmetros ajustados para as variáveis de cada temperatura pelo *software* Origin® versão 5.0, apresentando os resultados mostrados na Tabela 4.2.

A derivada desta função pode ser obtida analiticamente:

$$\frac{dy}{dt} = \frac{(A_2 - A_1) \cdot \left[\exp\left(\frac{t - x_0}{dx}\right) \right]}{\left[1 + \exp\left(\frac{t - x_0}{dx}\right) \right]^2} \cdot dx \quad (4.2)$$

Os resultados obtidos para os ajustes por esta função mostraram-se muito melhores com relação àqueles do ajuste polinomial, bem como as taxas cinéticas obtidas a partir da derivada da função Sigmoidal em relação a t (equação 4.2), como pode ser visto nas Figuras 4.1 a 4.10.

Nas Figuras 4.1 a 4.10 pode-se notar que, em alguns casos, o modelo validado por COSTA (2000) e ATALA (2000) apresenta alguma diferença em relação aos resultados das curvas ajustadas a partir dos dados experimentais. Na realidade, este modelo descreve o sistema em toda a faixa de temperatura estudada, apresentando por isso algum desvio dos dados reais. O modelo sigmoidal que foi ajustado representa melhor cada condição, já que os parâmetros foram ajustados independentemente para cada temperatura. Vale ressaltar que o modelo polinomial muito embora também tenha sido ajustado independentemente para cada temperatura, conforme já discutido, não chegou a bons resultados.

Tabela 4.2 – Parâmetros ajustados para as concentrações monitoradas experimentalmente, em cada temperatura, para a equação Sigmoidal-Boltzman.

TEMPERATURA (°C)	VARIÁVEIS	PARÂMETROS			
		A_1	A_2	x_0	Dx
28	Biomassa	1,13688	13,09002	9,56654	3,57281
	Substrato	186,33	6,1736	10,712	3,0846
	Produto	5,0409	97,919	13,710	4,8979
31	Biomassa	0,99818	13,454	8,0636	3,1297
	Substrato	231,18	12,778	9,0608	3,4747
	Produto	-24,350	124,84	10,125	6,4264
34	Biomassa	1,3454	9,9710	6,5084	2,0978
	Substrato	287,91	17,721	7,7526	4,4354
	Produto	-15,244	118,30	10,247	4,7902
37	Biomassa	0,59789	7,2172	9,1525	2,1222
	Substrato	297,22	46,992	11,14837	4,83
	Produto	-10,70852	92,76681	12,20909	5,44022
40	Biomassa	0,42822	5,0664	5,2660	3,1246
	Substrato	314,39	55,73	6,1939	6,0272
	Produto	-10,585	63,461	9,3974	4,8834

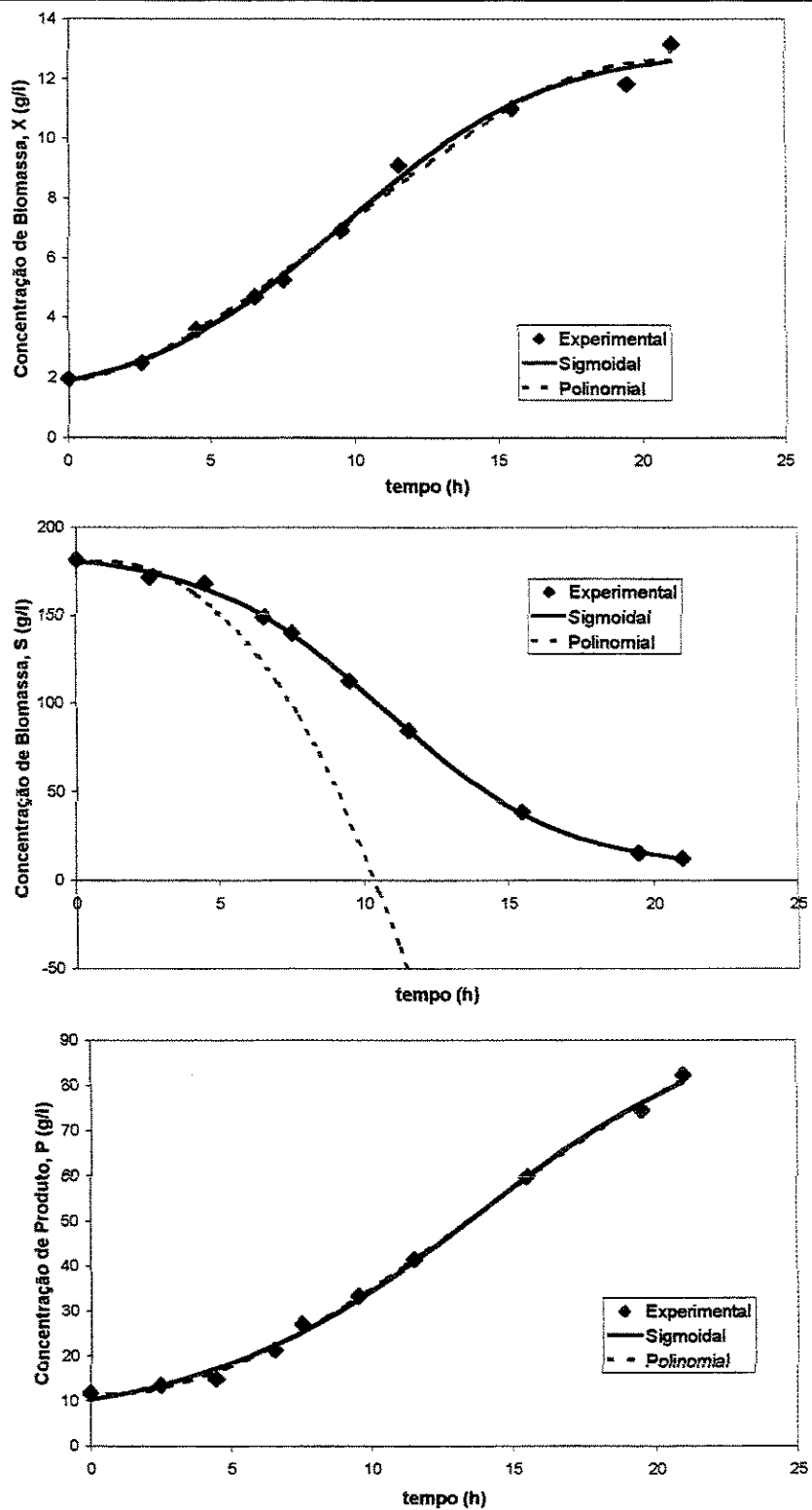


Figura 4.1 - Comparação entre os diversos ajustes, para as concentrações de a) Biomassa; b) Substrato e c) Produto. Temperatura de 28°C.

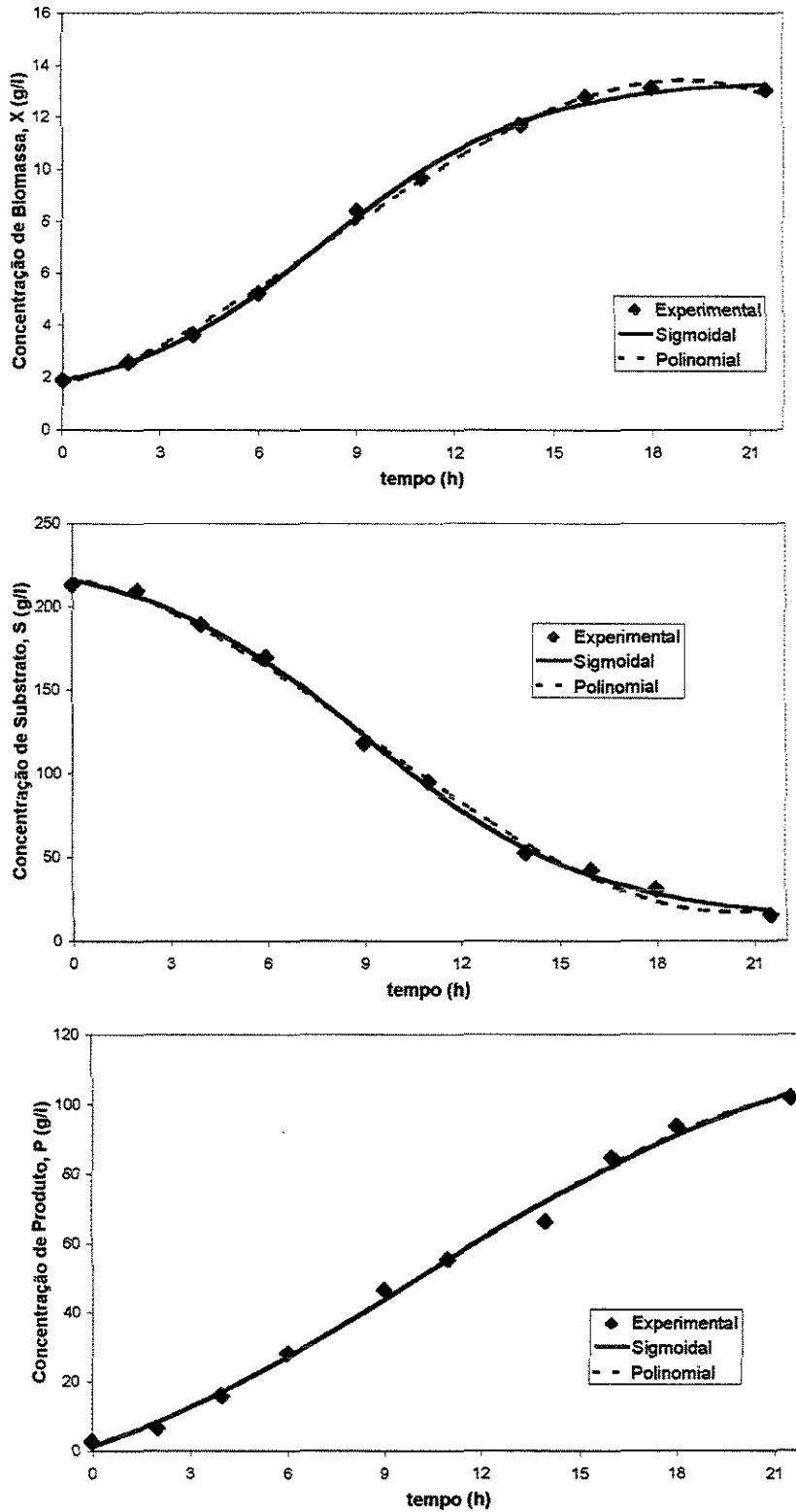


Figura 4.2 - Comparação entre os diversos ajustes, para as concentrações de a) Biomassa; b) Substrato e c) Produto. Temperatura de 31°C.

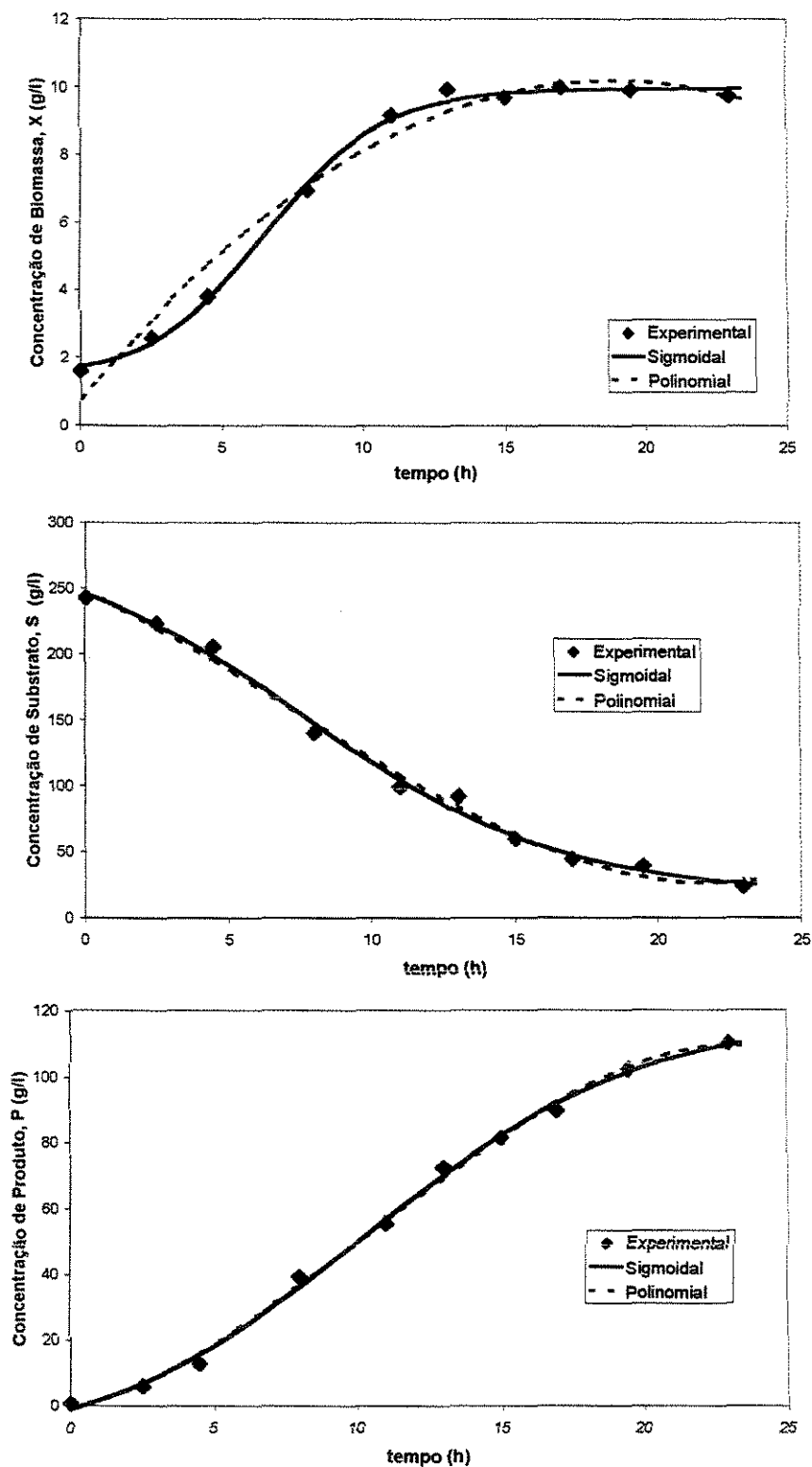


Figura 4.3 - Comparação entre os diversos ajustes, para as concentrações de a) Biomassa; b) Substrato e c) Produto. Temperatura de 34°C.

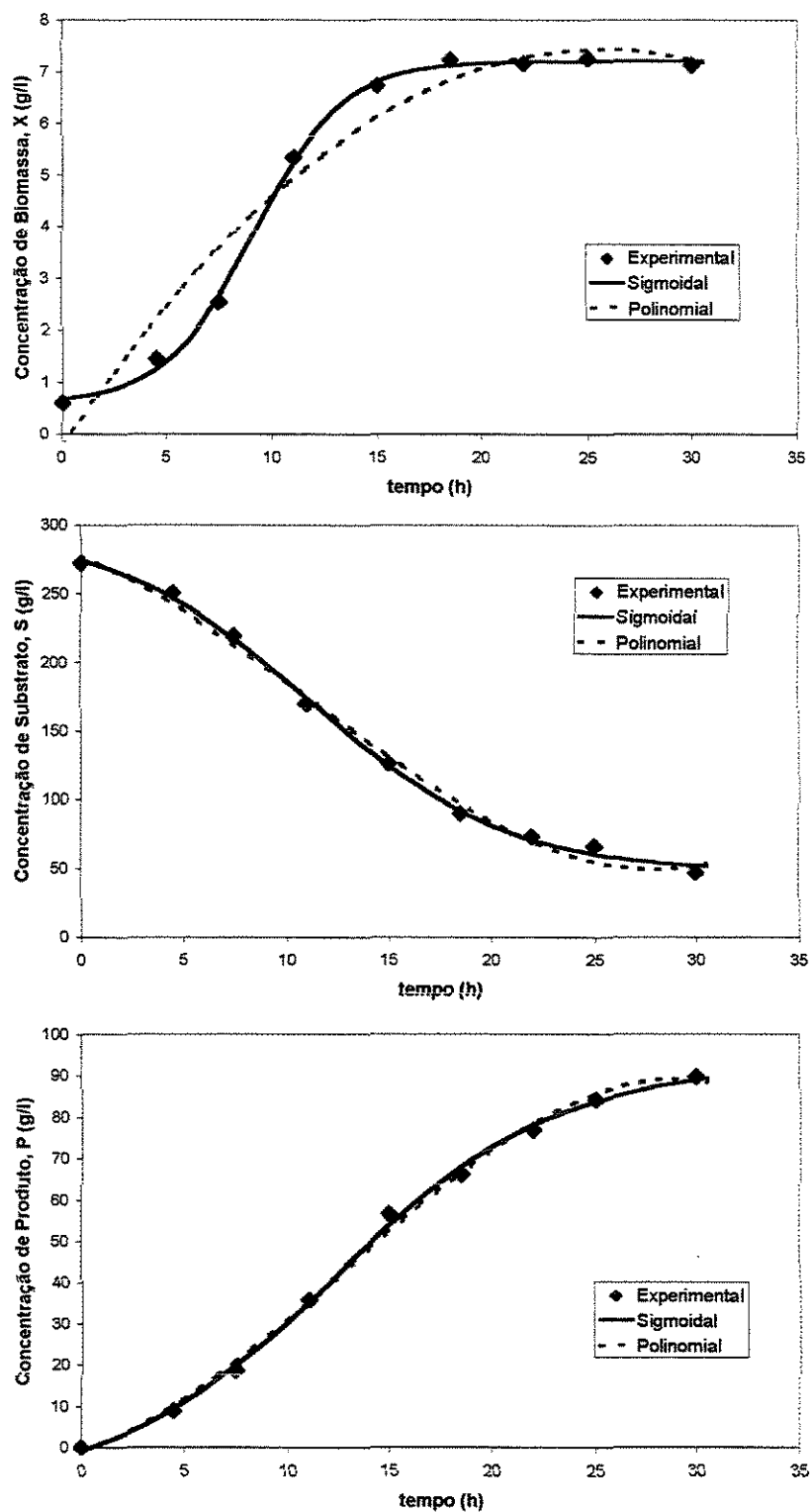


Figura 4.4 - Comparação entre os diversos ajustes, para as concentrações de a) Biomassa; b) Substrato e c) Produto. Temperatura de 37°C.

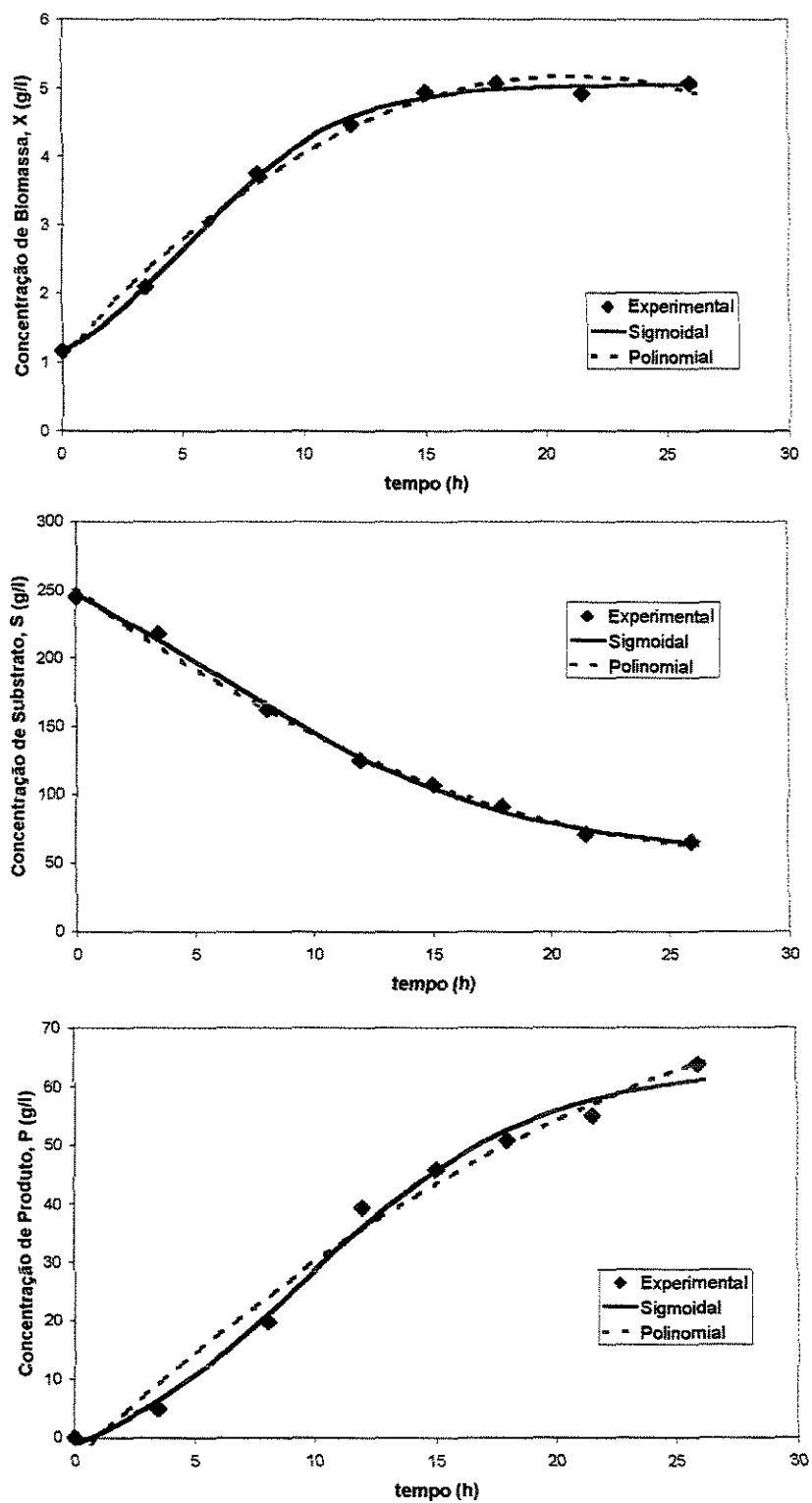


Figura 4.5 - Comparação entre os diversos ajustes, para as concentrações de a) Biomassa; b) Substrato e c) Produto. Temperatura de 40°C.

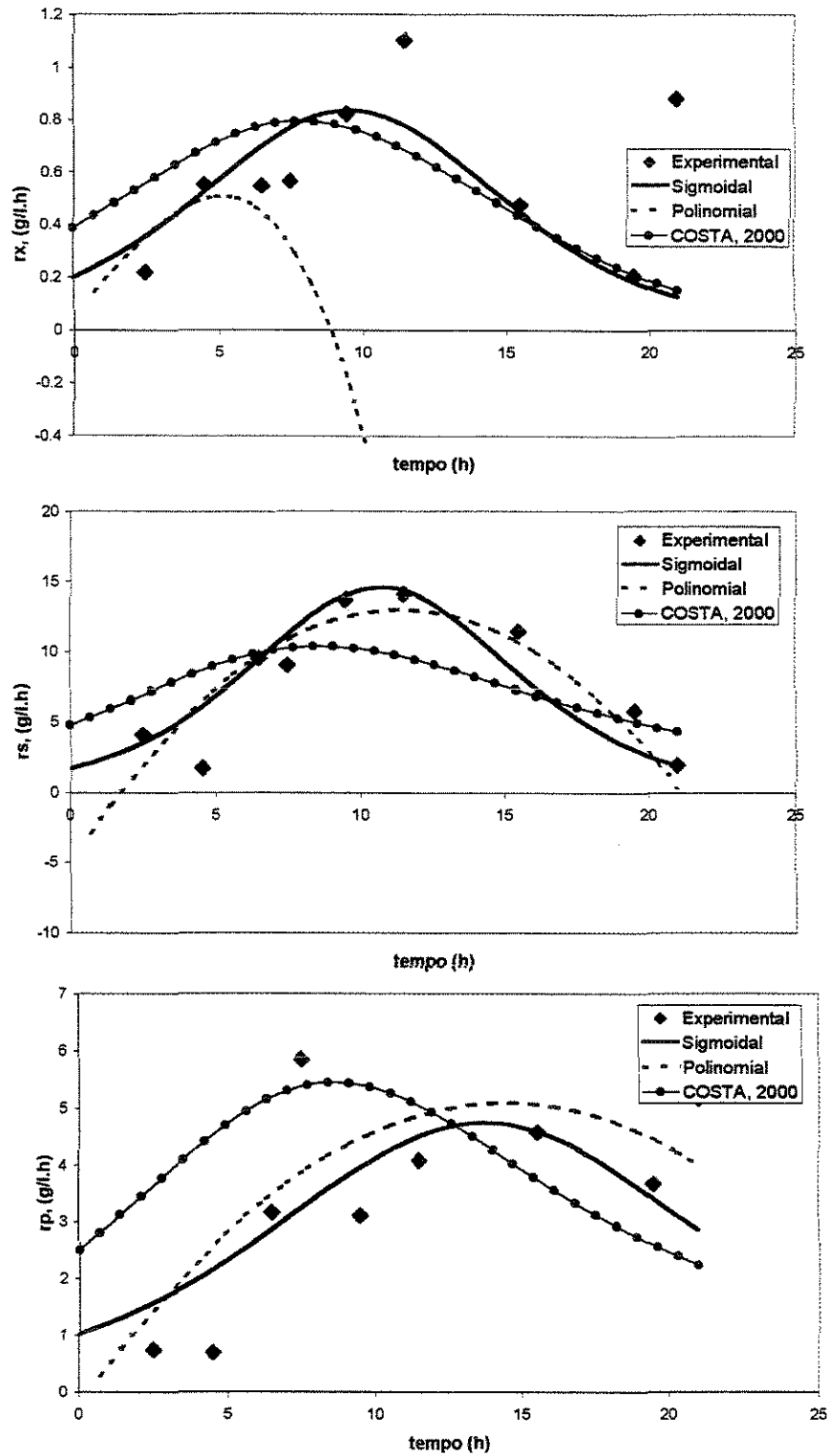


Figura 4.6 - Comparação entre os modelos ajustados para as taxas cinéticas. a) Crescimento de biomassa; b) Consumo de substrato; c) Formação de produto. Temperatura 28° C.

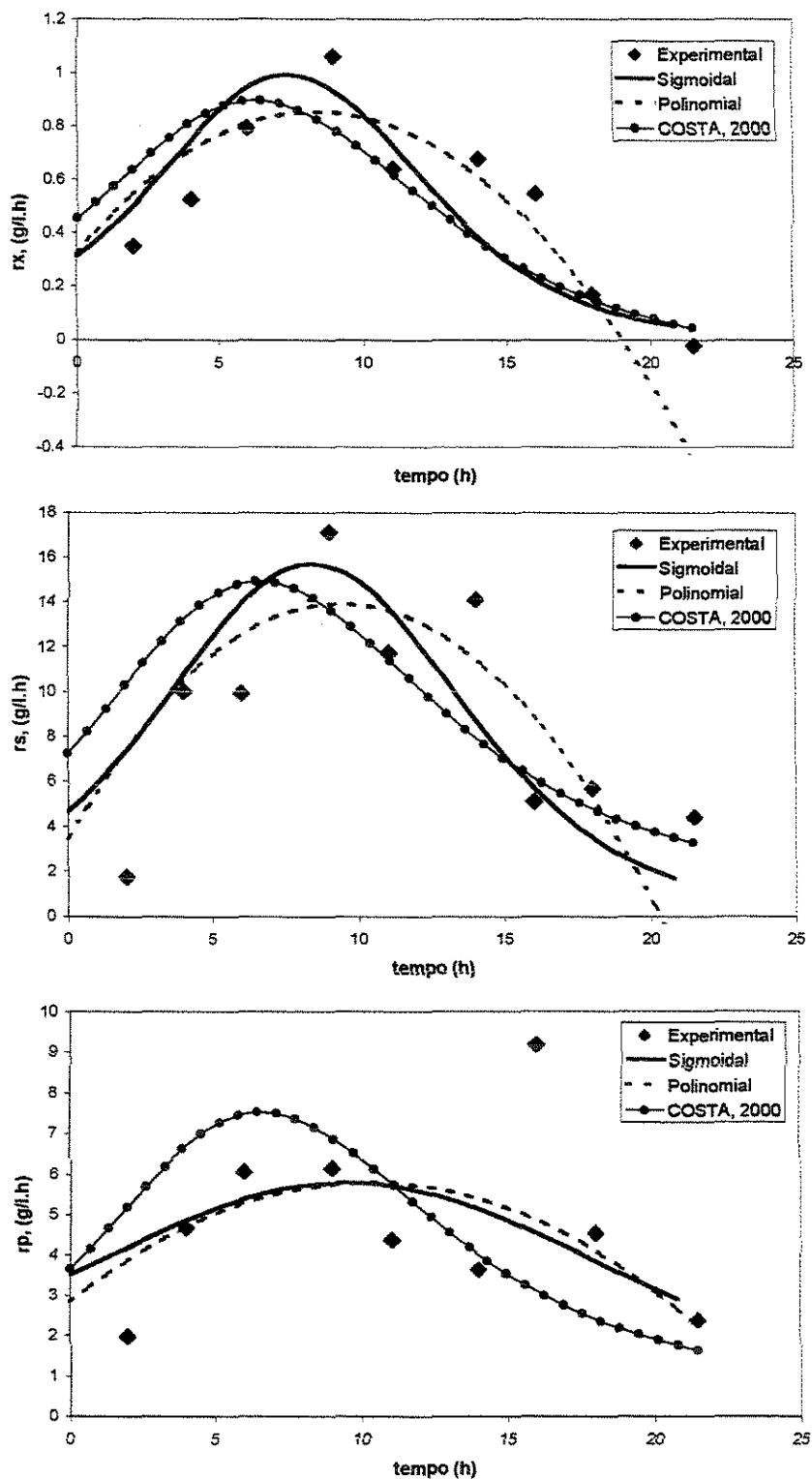


Figura 4.7 - Comparação entre os modelos ajustados para as taxas cinéticas. a) Crescimento de biomassa; b) Consumo de substrato; c) Formação de produto. Temperatura 31° C.

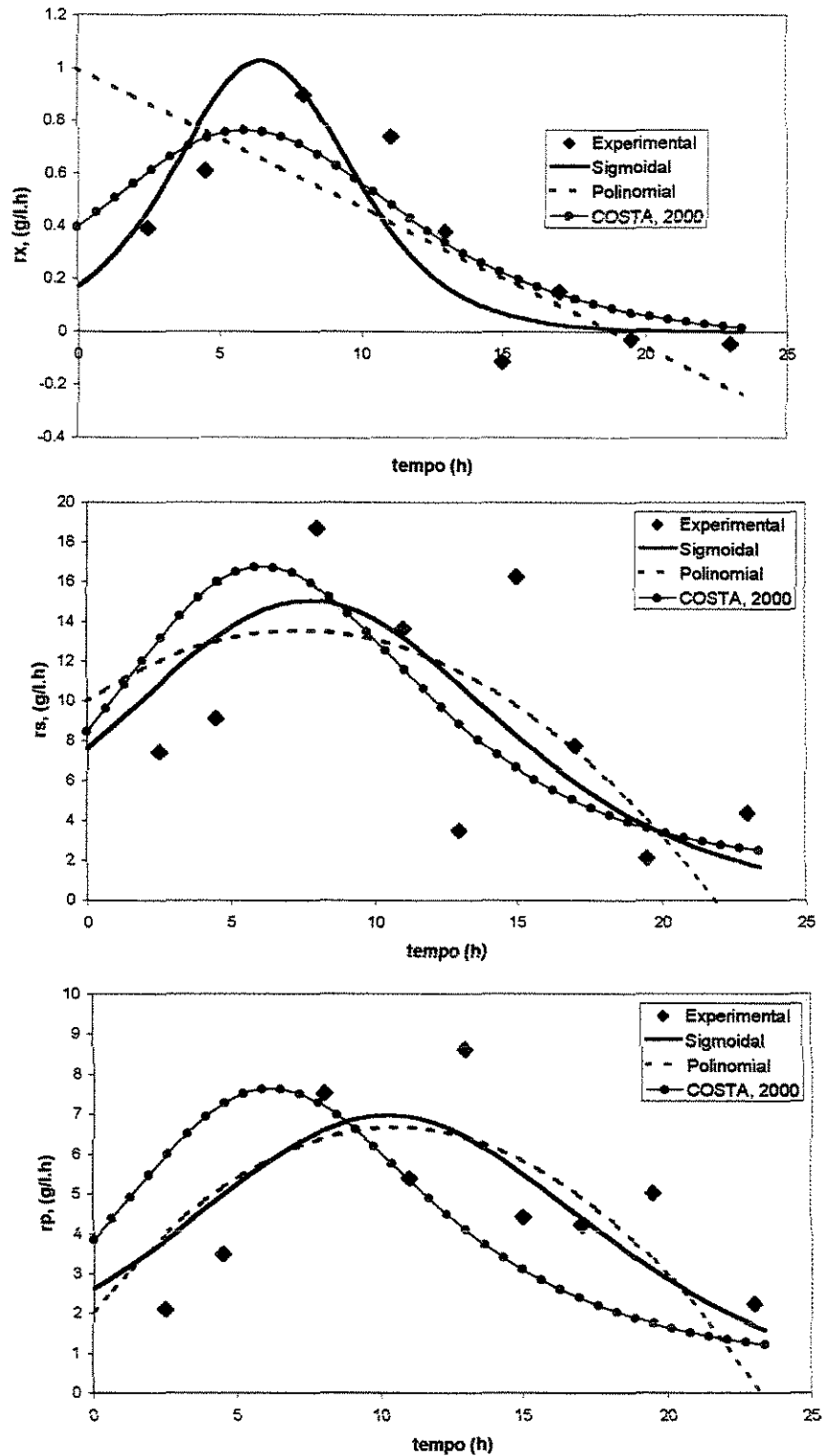


Figura 4.8 - Comparação entre os modelos ajustados para as taxas cinéticas. a) Crescimento de biomassa; b) Consumo de substrato; c) Formação de produto. Temperatura 34° C.

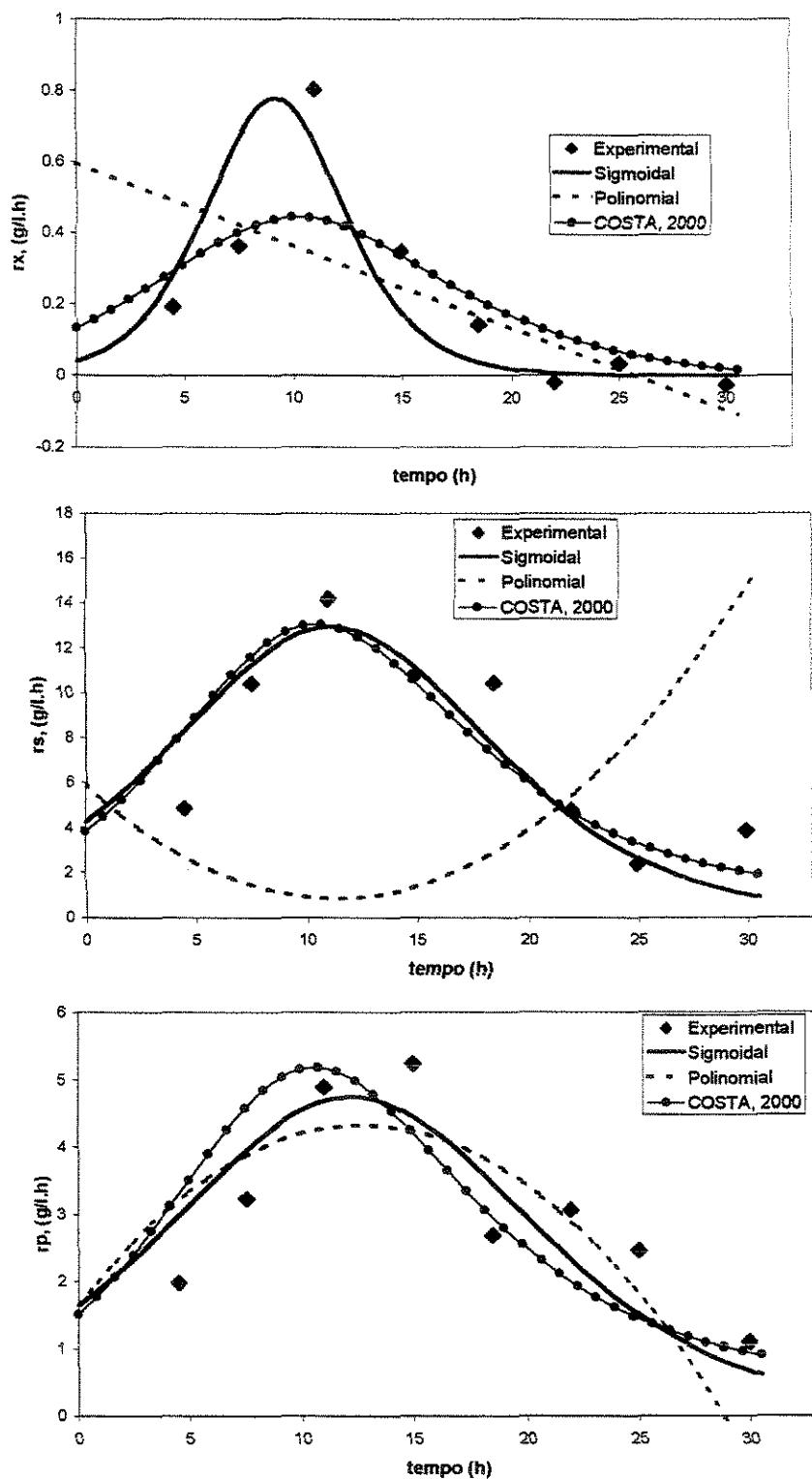


Figura 4.9 - Comparação entre os modelos ajustados para as taxas cinéticas. a) Crescimento de biomassa; b) Consumo de substrato; c) Formação de produto. Temperatura 37° C.

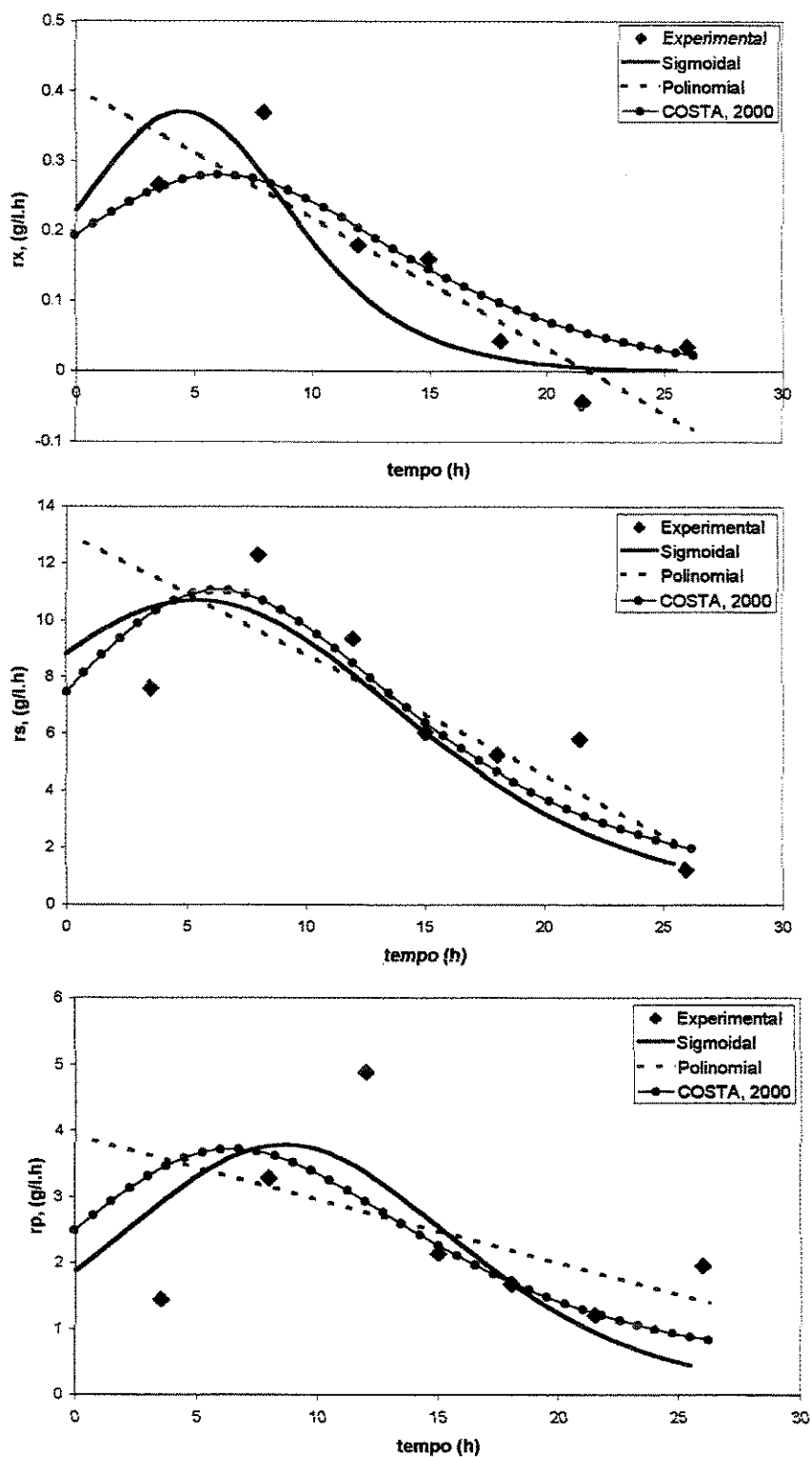


Figura 4.10 - Comparação entre os modelos ajustados para as taxas cinéticas. a) Crescimento de biomassa; b) Consumo de substrato; c) Formação de produto. Temperatura 40° C.

4.2 Modelagem com Redes Neurais Artificiais

Pela observação dos resultados obtidos da análise dos dados experimentais, os dados gerados pelo modelo Sigmoidal foram escolhidos para o treinamento da rede neural artificial.

As concentrações de Biomassa, Substrato e Produto, além da temperatura da fermentação, são as variáveis escolhidas para as entradas da rede. As taxas cinéticas, obtidas pela derivada da função Sigmoidal, são as saídas da rede, formando, portanto, os pares de entrada e saída do conjunto de treinamento.

Durante o treinamento das redes as arquiteturas *feedforward* foram estudadas, assim como as funções de ativação dos neurônios, a quantidade de nodos na camada intermediária, além dos algoritmos de treinamento.

O *software* empregado no desenvolvimento do modelo de rede neural é o Matlab[®] versão 5.3, devido a sua aptidão em trabalhar com matrizes e vetores, além de sua versatilidade. O método de aprendizado escolhido é o supervisionado, pois os conjuntos de entrada e saída são fornecidos por um supervisor externo. O objetivo do treinamento é ajustar os parâmetros da rede, de forma a encontrar uma ligação entre os pares de entrada e saída fornecidos. Os treinamentos são feitos *off-line*, pois os dados do conjunto de treinamento não mudam, e, uma vez obtida uma solução para a rede, esta permanece fixa.

Inicialmente foram treinadas redes com três saídas, ou três neurônios de saída, no entanto, como os resultados destes treinamentos não se mostraram satisfatórios, as saídas foram separadas e redes com apenas uma saída passaram a ser treinadas. Cada saída corresponde a uma taxa cinética específica. Mesmo procedendo desta forma com o aprendizado, as variáveis de entrada eram as mesmas para todas as redes, ou seja, inicialmente quatro entradas e três saídas e depois três redes com quatro entradas e apenas uma saída.

4.2.1 Normalização, Homogeneização e Conjunto de Validação

Para o treinamento da rede com o método *Backpropagation*, três funções de ativação foram pesquisadas. Estas funções estão mostradas na Figura 4.11.

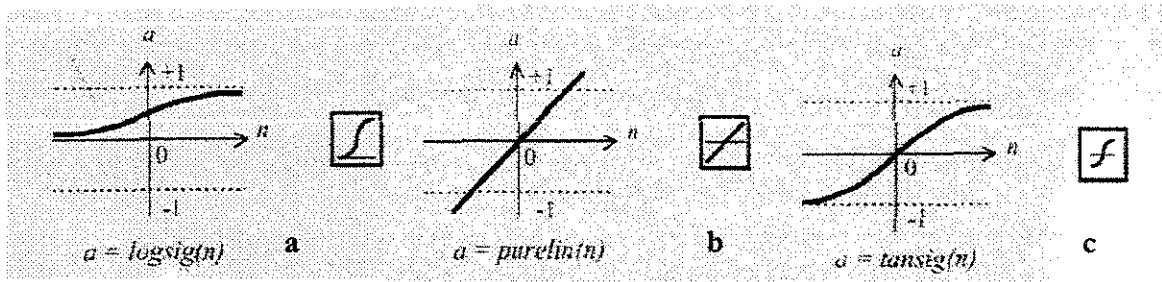


Figura 4.11 - Funções de ativação.

No início, somente a função logsig, também conhecida como sigmoideal, foi empregada na camada intermediária, e portanto a normalização dos dados foi feita dentro do intervalo fechado entre 0,1 e 0,9, conforme recomendado por BASHEER & HAJMEER (2000). A normalização também foi efetuada para que não houvesse discrepância entre os dados de entrada, uma vez que estes continham valores de temperatura e concentrações com magnitudes diferentes.

Na camada de saída da rede a função purelin, ou linear, foi adotada, pois a mesma não restringe os valores de saída, ao contrário das outras funções, como pode ser observado na Figura 4.11b. Desta forma, pode-se evitar que as saídas se tornem saturadas com valores em torno dos limites de máximo ou mínimo das funções não-lineares quando se estiver trabalhando com dados muito próximos dos limites superior ou inferior.

Foram feitos testes para verificar qual a melhor combinação de funções de ativação entre as camadas intermediária e de saída. Entretanto, mesmo testando as três funções, a normalização permaneceu dentro da faixa [0,1; 0,9], apesar da função tansig admitir valores da normalização para uma faixa entre -1 e 1 (Figura 4.11c).

A Figura 4.12 mostra alguns resultados para os testes de comparação feitos com as funções de ativação, onde estão mostrados o número de épocas e o erro quadrático para diferentes números de neurônios intermediários; o número de entradas foi sempre 4 e o de saídas igual a 1. As condições sobre as quais as redes foram treinadas são as mesmas, e conforme pode ser observado não houve, em geral, uma combinação entre as funções que evidenciasse um melhor desempenho de treinamento. As diferenças existentes entre os resultados expostos nessa figura para o número de épocas são devidas ao conjunto de inicialização dos pesos durante o aprendizado, já que a cada novo treino usa-se de um conjunto diferente, gerado aleatoriamente pelo programa de treinamento. Como o treinamento das redes é

uma busca pelo ponto de mínimo sobre uma superfície de erro, a estimativa inicial dos pesos pode estar próxima a este ponto de mínimo ou não. Este fato se traduz em um menor ou maior número de épocas, ou iterações, e diferentes erros de treinamento.

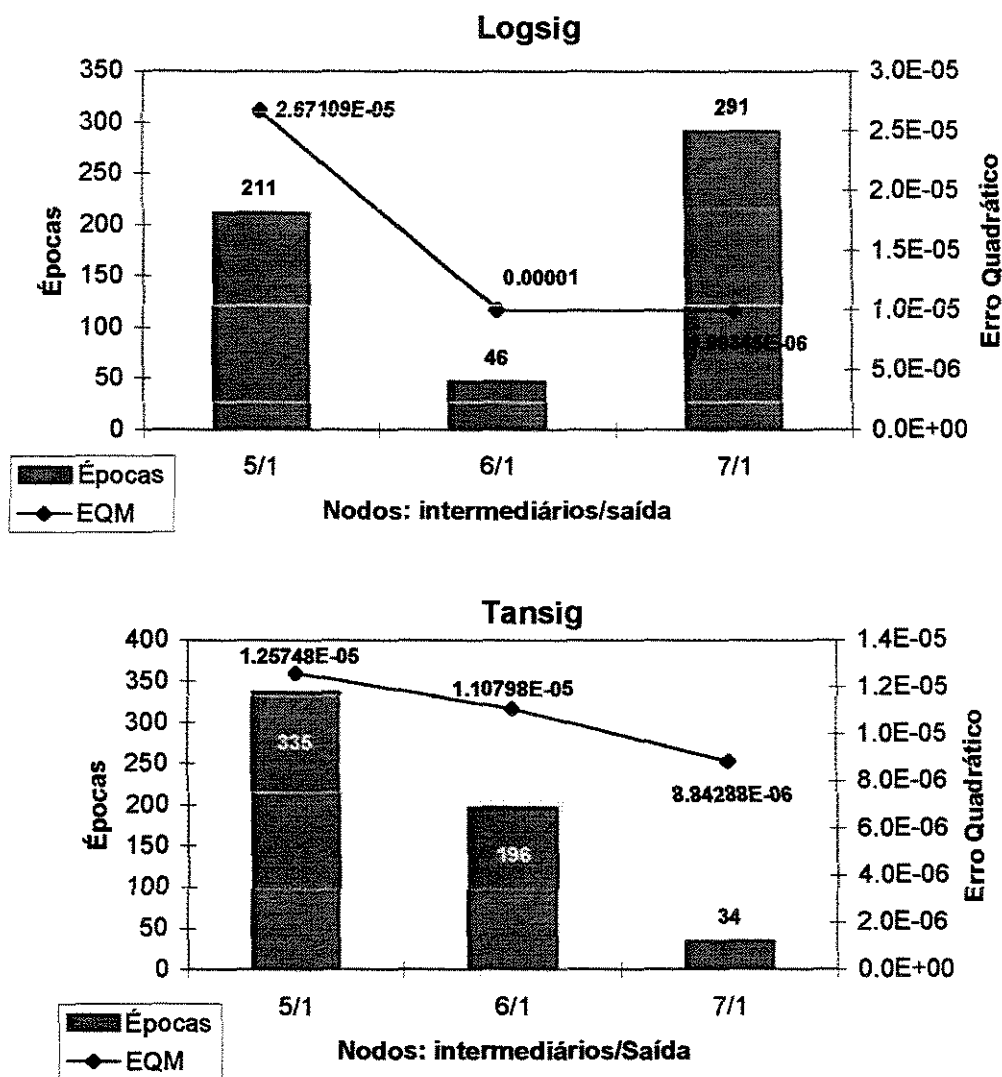


Figura 4.12 – Comparação entre as funções de ativação na camada intermediária. À saída a função purelin foi empregada. Algoritmo de aprendizado: Levenberg-Marquardt. Arquitetura Feedforward.

Embora no treinamento o uso das funções tansig ou logsig pareça levar a resultados semelhantes, a utilização da função tansig na camada intermediária, combinada com a purelin à saída, demonstrou para o caso estudado, um melhor desempenho na validação, quando a rede foi

testada frente a dados não usados no treinamento, como pode ser observado na Figura 4.13. Esta figura apresenta as curvas com os dados de validação de duas redes treinadas com as mesmas condições, com exceção das funções de ativação na camada intermediária. Os dados experimentais são os dados determinados usando o ajuste sigmoidal.

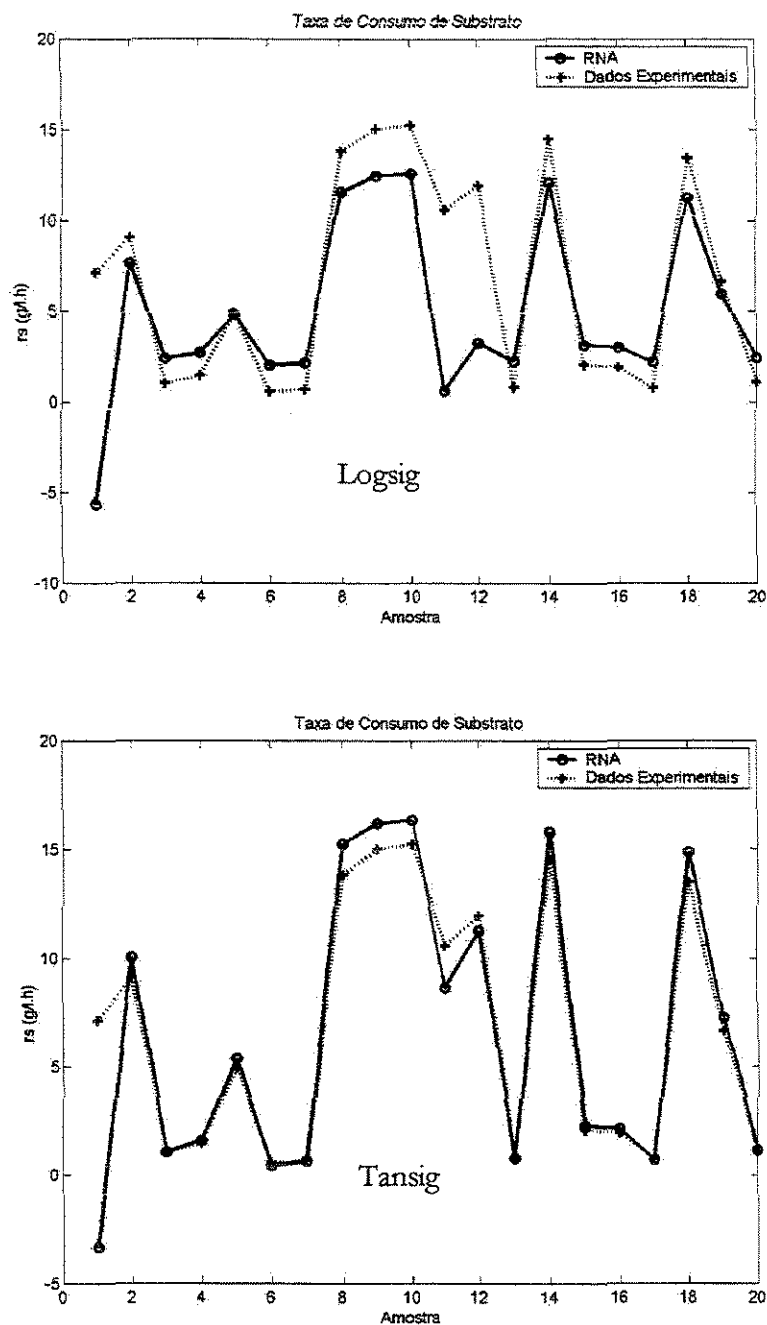


Figura 4.13 – Comparação entre as funções de ativação na camada intermediária. Arquitetura: Feedforward. Algoritmo: Levenberg-Marquardt. 5 nodos na camada intermediária e 1 na de saída.

A partir do conjunto de dados (pares de entrada e saída obtidos pela interpolação, como descrito no item 4.1) foram gerados conjuntos que foram usados no treinamento e validação das redes. Os treinamentos foram realizados empregando dois tipos de conjuntos. No primeiro caso os dados excluídos do conjunto de treinamento e usados para validação consistiam em pares de entrada (T, X, S e P) e saída (rx, rs e rp) aleatoriamente escolhidos entre os pares de exemplos disponíveis. No segundo caso, os dados da temperatura de 34°C foram excluídos do conjunto de treinamento, e foram usados para testar ou validar o modelo.

Na Figura 4.14 o erro quadrático é mostrado para diferentes números de neurônios intermediários e de saída. A quantidade de neurônios na entrada foi sempre igual a 4. As saídas incluíam as três taxas cinéticas. Entretanto, como pode ser observado na figura, quando comparamos as arquiteturas com 3 e 1 neurônios na saída, nota-se que o treinamento com várias saídas é mais difícil e levou a resultados menos satisfatórios. Por isso, em uma segunda fase de treinamento foram criadas redes com apenas uma saída, mas contendo em sua entrada todas as demais variáveis, como temperatura, biomassa, substrato e produto. Ou seja, foram criadas redes específicas para cada taxa cinética.

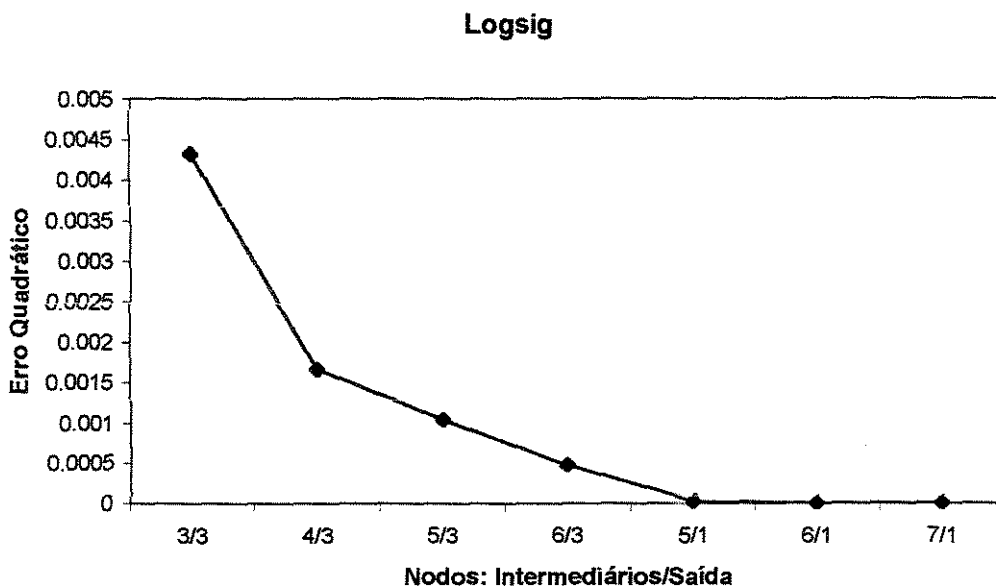


Figura 4.14 – Relação entre o erro quadrático com o número de nós na camada de saída. Arquitetura: Feedforward. Algoritmo: Levenberg-Marquardt.

Os conjuntos de dados gerados para o treinamento foram divididos, conforme já comentado, em dois tipos. O primeiro retirando aleatoriamente pares de exemplos entre as diversas temperaturas disponíveis para montar o conjunto de validação, que não era incluso no treino (Conjunto Tipo I). Estes conjuntos possuíam 976 linhas para o treino e 244 linhas para a validação. Já o segundo tipo de conjunto excluía os dados de uma temperatura (34°C), para serem empregados na validação (Conjunto Tipo II). Assim, o treino com estes dados foram realizados com as quatro temperaturas restantes, tendo 989 linhas de pares de exemplos e 236 linhas da temperatura de 34°C usados para validação.

No artigo de GOH (1995), os dados que formaram o conjunto de validação ou teste foram retirados aleatoriamente do próprio conjunto inicial de dados, apresentando um tamanho de normalmente 1/3 deste conjunto inicial.

CONTANT (1999) sugere que o conjunto de treinamento apresentado à rede tenha os dados homogeneizados, ou seja, misturados aleatoriamente, a fim de evitar que a rede se torne tendenciosa. Após a homogeneização dos dados, a normalização é feita e em seguida o treinamento. Este procedimento foi adotado neste trabalho.

4.2.2 Algoritmos de Treinamento e Arquitetura da Rede

Foram estudados inicialmente três algoritmos de treinamento: Levenberg-Marquardt, Bayesian *framework* e Gradiente Descendente com taxa de aprendizado variável, todos empregando a arquitetura *feedforward* nas conexões entre os pesos, e o método *backpropagation*. Posteriormente um algoritmo específico do Método Quasi-Newton foi igualmente empregado.

A respeito da quantidade de camadas, as redes implementadas foram treinadas com apenas uma camada intermediária e outra de saída, no entanto, com a quantidade de nodos ou neurônios variáveis, principalmente na camada intermediária. Segundo CYBENKO (1989), citado por BRAGA et al. (2000), uma rede com uma camada intermediária é capaz de implementar qualquer função contínua. A Figura 4.15 apresenta um exemplo de rede *feedforward* com duas camadas intermediárias.

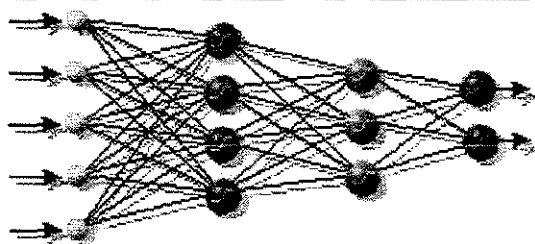


Figura 4.15 – Exemplo de uma rede Feedforward, completamente conectada, com duas camadas intermediárias.

Levenberg-Marquardt: este algoritmo pode fazer uma aproximação com uma velocidade de treinamento de segunda ordem sem precisar computar a matriz Hessiana. Quando a função de desempenho tem a forma da soma dos quadrados, $f(x) = h_{11}x_1^2 + h_{12}x_1x_2 + h_{22}x_2^2$ (típico de redes *feedforward*), a matriz Hessiana é aproximada da seguinte maneira,

$$H(x) \equiv \nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} \end{bmatrix} = \begin{bmatrix} 2h_{11} & h_{12} \\ h_{12} & 2h_{22} \end{bmatrix} \quad (4.3)$$

ou com uma notação mais simples,

$$H = J^T J \quad (4.4)$$

e o gradiente pode ser computado como,

$$g = J^T e \quad (4.5)$$

A matriz jacobiana (J) contém as derivadas de primeira ordem dos erros da rede com relação aos pesos e aos *biases*, e e é o vetor dos erros da rede.

O algoritmo Levenberg-Marquardt usa esta aproximação para a matriz Hessiana da seguinte maneira:

$$x_{k+1} = x_k - [J^T J + \mu I]^{-1} J^T e \quad (4.6)$$

Gradiente Descendente: como o próprio nome já diz, este algoritmo promove a variação dos pesos e *biases* na direção do gradiente negativo da função erro. Para tanto, o treinamento da rede começa com pesos aleatórios e procede com pequenas variações neste vetor até que um mínimo local ou global seja alcançado.

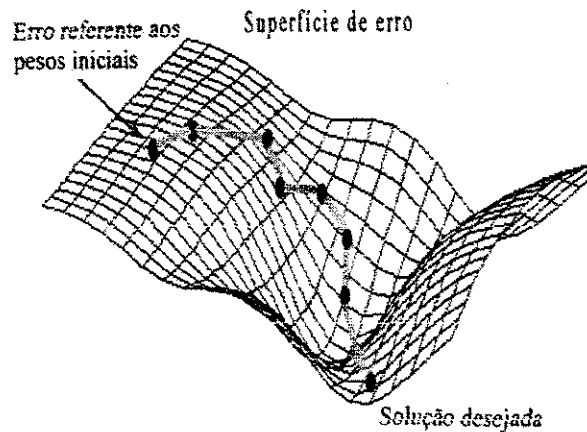


Figura 4.16 - Exemplo de superfície de erro e passos do gradiente descendente. (Fonte: BRAGA et al., 2000).

O desempenho deste algoritmo é melhorado quando a taxa de aprendizado é variada durante o treinamento, pois se a mesma é mantida constante em um valor muito alto o algoritmo pode oscilar e provocar instabilidade. Já se a taxa é muito pequena, o tempo de convergência torna-se muito grande.

Algoritmo BFGS (Broyden, Fletcher, Goldfarb e Shanno): este algoritmo também é conhecido como Método Quasi-Newton (ou secante), pois é baseado no método de Newton, no entanto, sem calcular as derivadas de segunda ordem. O passo básico do método de Newton é:

$$x_{k+1} = x_k - A_k^{-1} g_k \quad (4.5)$$

onde A_k é a matriz Hessiana (derivadas de segunda ordem) do índice de desempenho com os valores atuais dos pesos e *biases* (eq. 4.3). O método de convergência de Newton apresenta uma boa eficiência principalmente com relação à quantidade de iterações necessárias para a otimização. Infelizmente, no entanto, é difícil computar a matriz Hessiana para uma rede neural *feedforward*. Assim uma classe de algoritmos baseada neste método atualiza uma aproximação desta matriz a cada iteração. Esta atualização é computada como uma função do gradiente. Este algoritmo requer um maior esforço computacional e de memória, pois a aproximação da matriz Hessiana precisa ser armazenada. Assim sua indicação é para redes pequenas, para grandes RNAs algum dos algoritmos que empregam o método do gradiente conjugado são recomendados.

Em todas as redes o *Bias* foi usado. Este parâmetro funciona como um peso, com a diferença de que ele possui uma entrada constante com valor 1. Assim a função transferência tem como entrada um escalar, ou seja, a soma entre o produto $\mu \cdot x$ (peso (w), variável de entrada

(\mathbf{x}) e o *Bias*. De acordo com DIETTERICH & KONG (1995), os algoritmos de aprendizado devem adotar esta entrada para permitir uma melhor capacidade de generalização.

4.2.2.1 Melhorando a Capacidade de Generalização

Um dos problemas que ocorrem durante o treinamento de uma rede é o sobreajuste. O erro para o conjunto de treinamento se torna muito pequeno, mas quando novos dados são apresentados à rede, o erro é grande; ou seja, a rede não possui capacidade de generalização.

Uma forma de melhorar a capacidade de generalização da rede é chamada de regularização. Isto envolve a modificação do critério de desempenho usado durante o treinamento da rede de forma que a esta apresente uma resposta mais suave e tenha menos probabilidade de sobreajustar os dados. Uma abordagem que pode ser usada é chamada de *Bayesian Framework* (McKAY, 1992), uma característica deste algoritmo é que ele fornece uma medida de quantos parâmetros (pesos e biases) estão sendo efetivamente usados pela rede.

4.2.2.2 Métodos de Avaliação das Redes

Para a avaliação das redes treinadas e do próprio treinamento podem ser adotados alguns parâmetros ou critérios. Estes são:

- Erro Quadrático;

$$EQ = \frac{1}{2} \sum_m [(d_i - y_i)^2] \quad (4.6)$$

onde m representa o número de variáveis de saída da rede, y é a saída estimada pela rede e d a saída desejada.

- Gráficos comparando a resposta da rede em relação aos dados experimentais interpolados;
- Desvio Padrão Residual (%);

$$DPR = \sqrt{\frac{\sum_{i=1}^n (d_i - y_i)^2}{n}} \quad (4.7)$$

$$DPR(\%) = \frac{DPR}{\bar{y}_i} \cdot 100 \quad (4.8)$$

- Coeficiente de Correlação para o teste de validação, entre as respostas da rede para a validação e os dados experimentais de acordo com HARADA, 2001. Nas equações 4.9-a,b,c $y_e(k)$ é a saída experimental k, $y(k)$ é a saída da rede correspondente, \bar{y}_e é o valor médio das saídas experimentais e N é o número de dados experimentais.

$$cor = \left(1 - \frac{SEE}{S\tau\tau}\right) \cdot 100\% \quad (a)$$

$$SEE = \sum_{k=1}^N (y_e(k) - y(k))^2 \quad (b) \quad (4.9)$$

$$S\tau\tau = \sum_{k=1}^N (y_e(k) - \bar{y}_e)^2 \quad (c)$$

De acordo com TEISSIER (1996), a qualidade de predição de um modelo pode ser caracterizada usando o desvio padrão residual (DPR), representado pela equação 4.7. No entanto, para evitar que o DPR seja comparado com magnitudes diferentes, COSTA (2000) sugere analisar o desvio padrão em forma de porcentagem da média dos valores calculados pelo modelo determinístico (\bar{y}_i), como mostra a equação 4.8. Neste trabalho foram adotados como critérios o erro quadrático, a comparação dos gráficos com resultados e o desvio padrão residual.

4.2.3 Modelo Híbrido com Variáveis Primárias

O modelo híbrido desenvolvido utiliza as equações determinísticas (equações 3.4 a 3.6) do processo em batelada, juntamente com as redes neurais treinadas para cada taxa cinética. O fluxograma deste modelo pode ser observado na Figura 4.17.

Para os testes realizados com o modelo, as redes, uma vez treinadas e validadas, eram usadas para prever as taxas cinéticas respectivas, ou seja, como cada rede foi treinada para apenas uma saída, então durante o procedimento de cálculo do modelo híbrido, cada taxa cinética deveria ser prevista por uma rede treinada especificamente para aquela saída. Assim, no modelo híbrido desenvolvido as redes neurais artificiais foram empregadas como ferramentas para descrever as taxas cinéticas do processo, enquanto o modelo convencional, ou equações de balanço, utilizavam os valores das taxas previstas pelas redes como entradas para calcular as concentrações das variáveis primárias, como Biomassa, Substrato e Produto em um novo tempo $t+1$. Como o fluxograma mostrado na Figura 4.17 ilustra o procedimento. Feita a previsão da

rede a partir dos dados inicialmente alimentados $[T, X_n, S_n \text{ e } P_n]_t$ para as taxas cinéticas, as equações de balanço recebiam tais taxas e calculavam novos valores das concentrações em um novo tempo $[T, X_{n+1}, S_{n+1} \text{ e } P_{n+1}]_{t+1}$. Estes dados juntamente com a temperatura eram novamente alimentados às redes e o processo repetia-se até ser atingido o tempo final da fermentação.

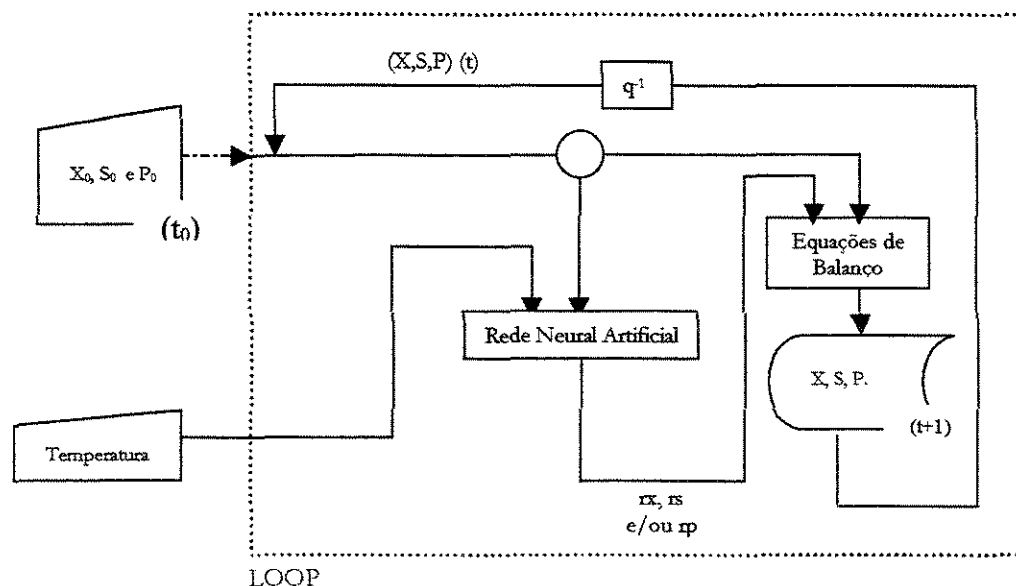


Figura 4.17 - Fluxograma do modelo híbrido.

4.2.3.1 Resultados

As redes com três saídas, muito embora tenham sido estudadas, não apresentaram boas respostas tanto de treino quanto de validação, por isso não têm seus resultados incluídos a seguir.

As Figuras 4.18 a 4.23 apresentam os resultados para a temperatura de 34°C, pois esta temperatura encontra-se no centro do domínio das temperaturas estudadas. Os resultados de treinamento e validação para as demais temperaturas não foram incluídos, pois apresentam o mesmo comportamento, não evidenciando nenhuma outra conclusão diferente a partir dos seus gráficos. Já os resultados obtidos do modelo híbrido para as temperaturas de 28, 31, 37 e 40°C, foram colocados a seguir a partir das Figuras 4.24 a 4.28.

Pode ser observado nestas figuras que na maioria dos casos o algoritmo de treinamento de Levenberg-Marquardt foi empregado para as redes, porque muito embora outros algoritmos também tenham sido estudados, seus resultados não se mostraram tão bons quanto estes.

Entretanto, para aquelas figuras que apresentam outros algoritmos de treinamento, estes também se tratam dos melhores resultados obtidos para aqueles dados.

A quantidade de nodos empregados no treinamento para os resultados de cada taxa cinética foi determinada com o algoritmo Bayesian *framework*. De acordo com o resultado proporcionado por este algoritmo, para o treinamento das redes são necessários em torno de 45 a 55 conexões, para uma rede tipo *feedforward* completamente conectada, incluindo os *biases*. Esta informação pode ser traduzida em 8 a 9 nodos na camada intermediária. A maioria das redes, desta forma, foram treinadas com 9 nodos nesta camada.

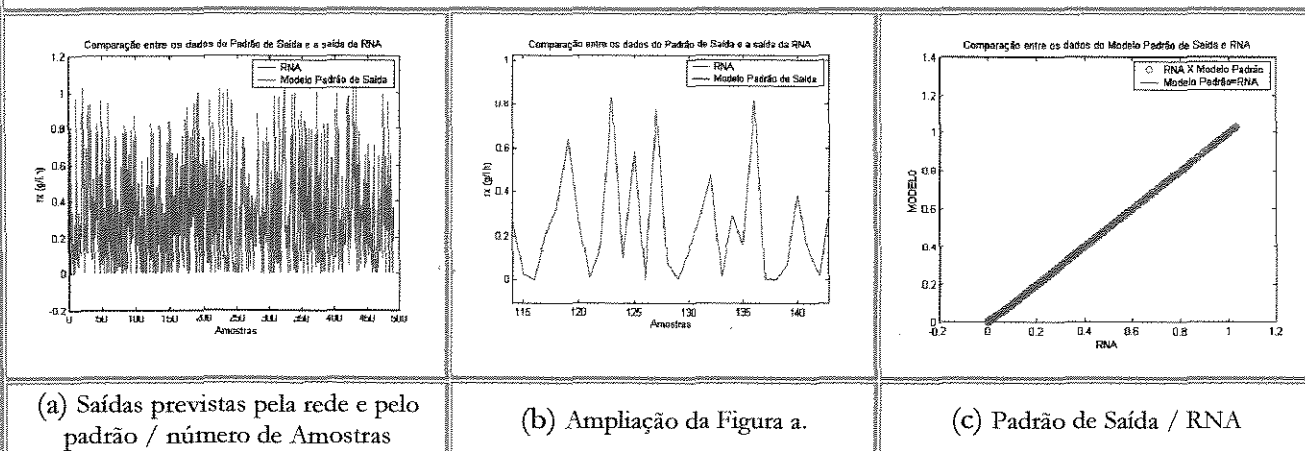
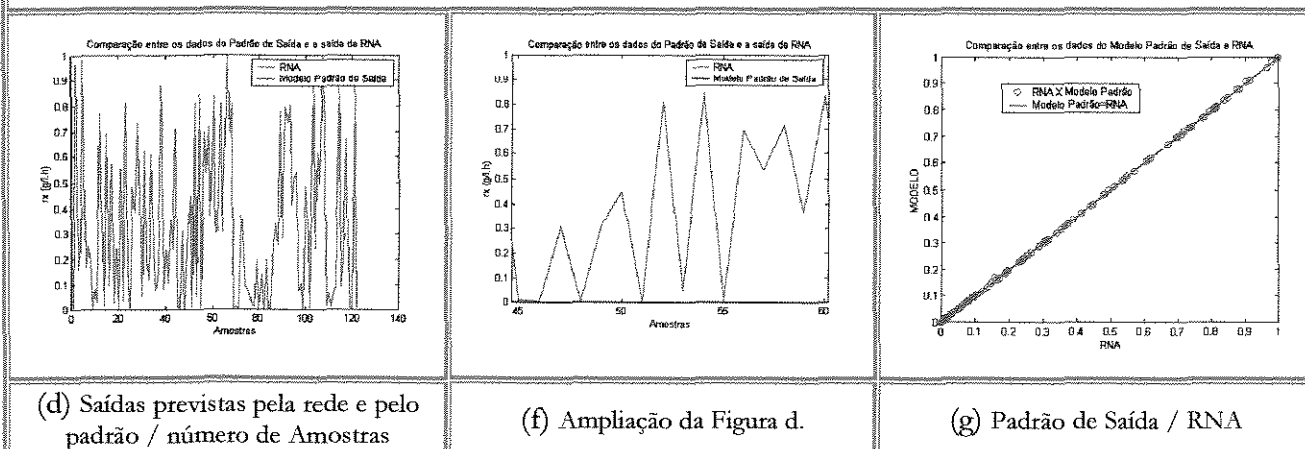
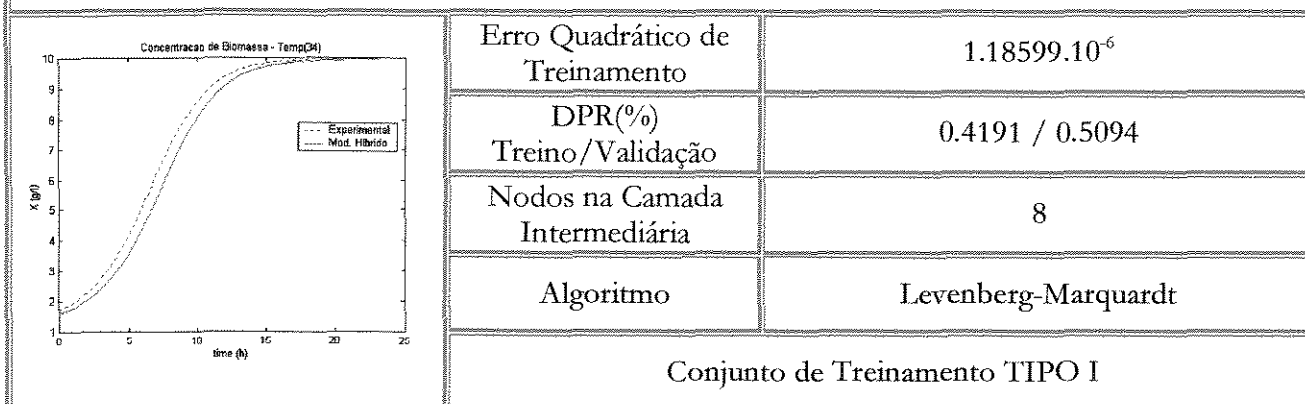
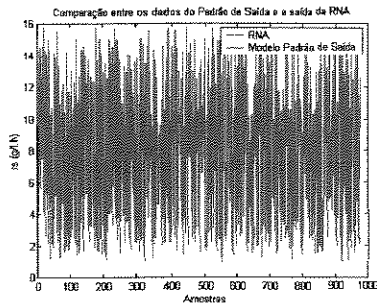
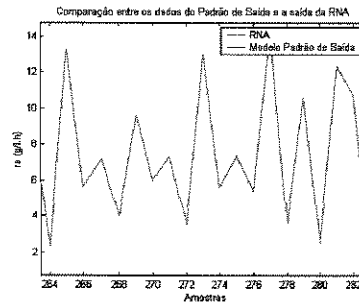
Treino:Validação:Modelo Híbrido:

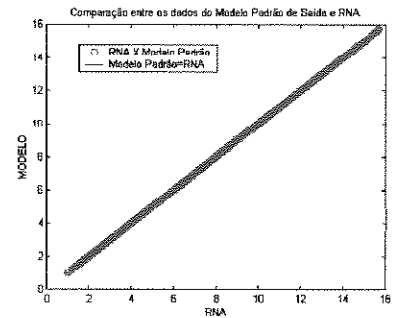
Figura 4.18 – Resultados do treinamento da rede para a Taxa de Crescimento de Biomassa e aplicação do Modelo Híbrido para Concentração de Biomassa. Temperatura de: 34°C. Conjunto de Treinamento TIPO I.

Treino:

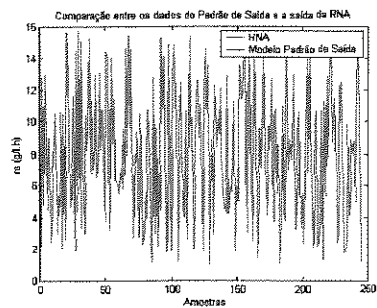
(a) Saídas previstas pela rede e pelo padrão / número de Amostras



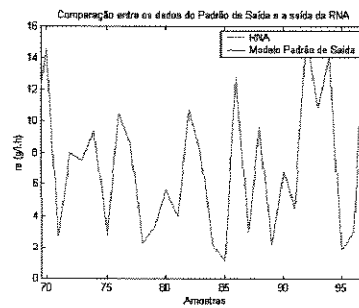
(b) Ampliação da Figura a.



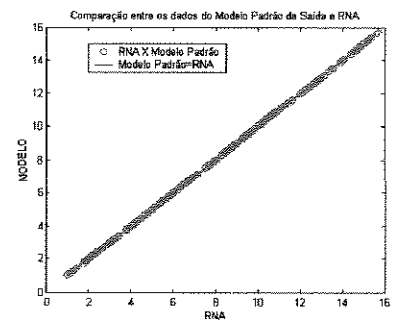
(c) Padrão de Saída / RNA

Validação:

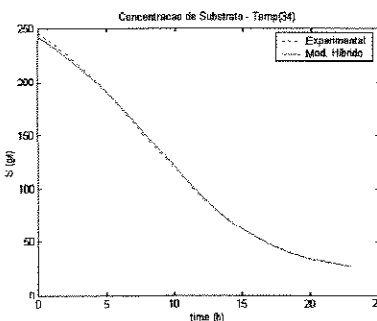
(d) Saídas previstas pela rede e pelo padrão / número de Amostras



(f) Ampliação da Figura d.



(g) Padrão de Saída / RNA

Modelo Híbrido:Erro Quadrático de
Treinamento $9.81101 \cdot 10^{-7}$ DPR(%)
Treino/Validação

0.2239 / 0.2196

Nodos na Camada
Intermediária

9

Algoritmo

Levenberg-Marquardt

Conjunto de Treinamento TIPO I

Figura 4.19 – Resultados do treinamento da rede para a Taxa de Consumo de Substrato e aplicação do Modelo Híbrido para Concentração de Substrato. Temperatura de: 34°C. Conjunto de Treinamento TIPO I.

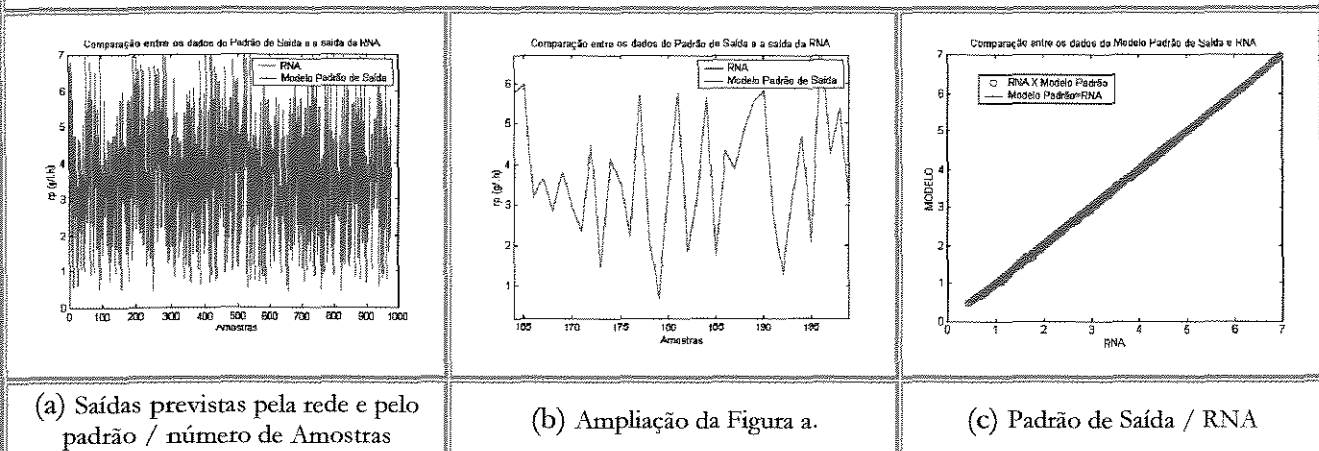
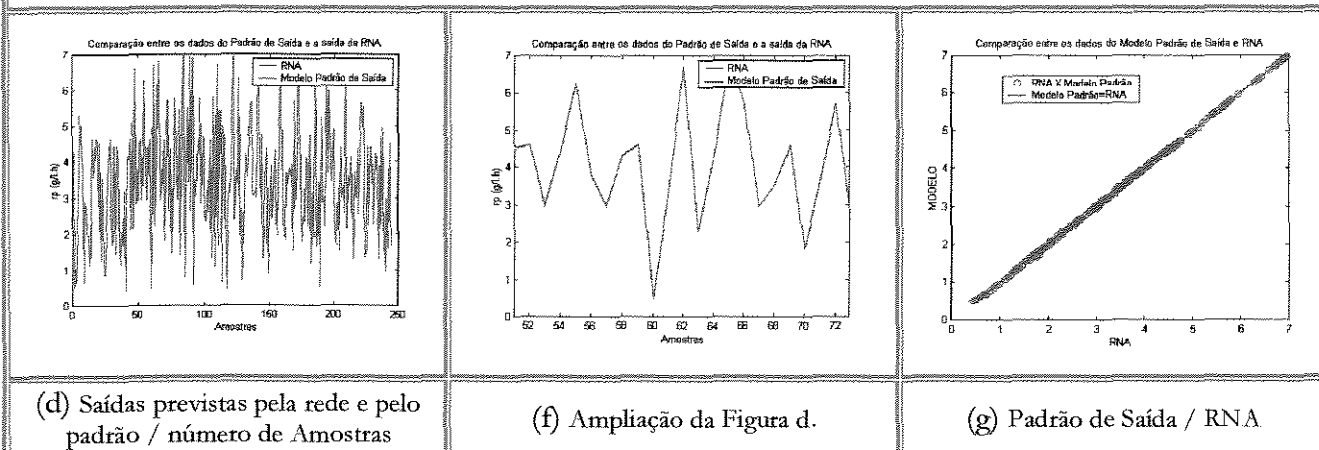
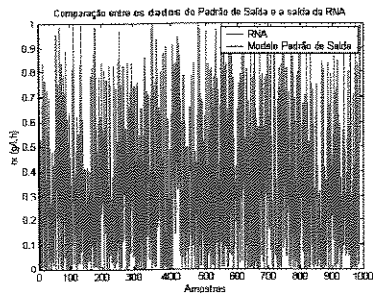
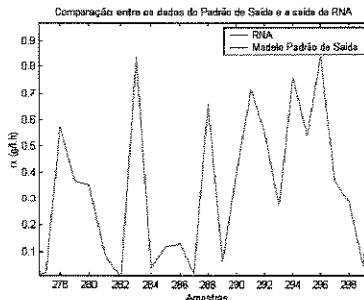
Treino:**Validação:****Modelo Híbrido:**

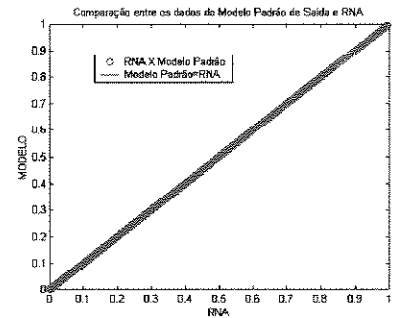
Figura 4.20 - Resultados do treinamento da rede para a Taxa de Formação de Produto e aplicação do Modelo Híbrido para Concentração de Produto. Temperatura de: 34°C. Conjunto de Treinamento TIPO I.

Treino:

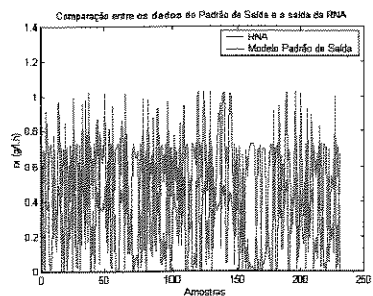
(a) Saídas previstas pela rede e pelo padrão / número de Amostras



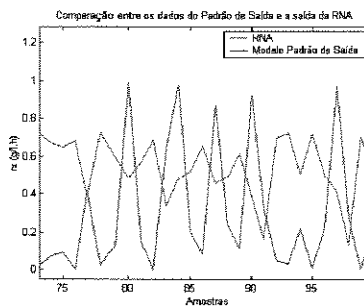
(b) Ampliação da Figura a.



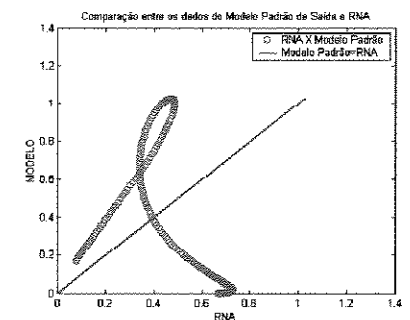
(c) Padrão de Saída / RNA

Validação:

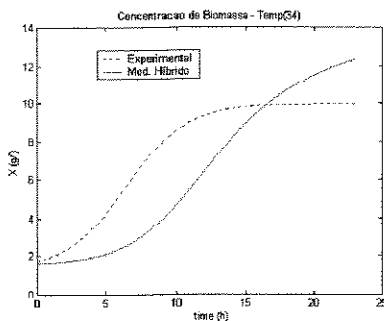
(d) Saídas do padrão e previstas pela rede / número de Amostras



(f) Ampliação da Figura d.



(g) Padrão de Saída / RNA

Modelo Híbrido:Erro Quadrático de
Treinamento $3.17432 \cdot 10^{-7}$ DPR(%)
Treino/Validação

0.0068 / 141.85

Nodos na Camada
Intermediária

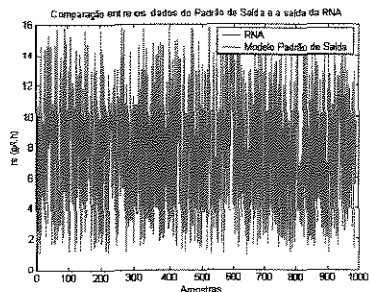
9

Algoritmo

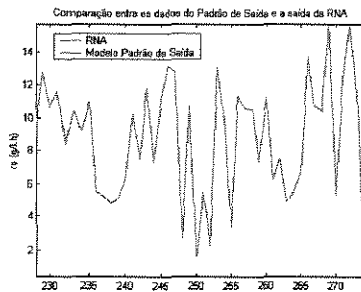
Baesian Framework

Conjunto de Treinamento TIPO II

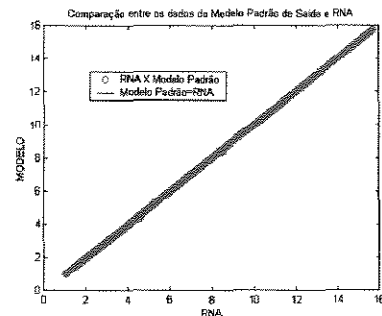
Figura 4.21 - Resultados do treinamento da rede para a Taxa de Crescimento de Biomassa e aplicação do Modelo Híbrido para Concentração de Biomassa. Temperatura de: 34°C. Conjunto de Treinamento TIPO II

Treino:

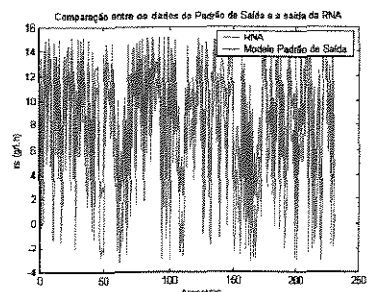
(a) Saídas previstas pela rede e pelo padrão / número de Amostras



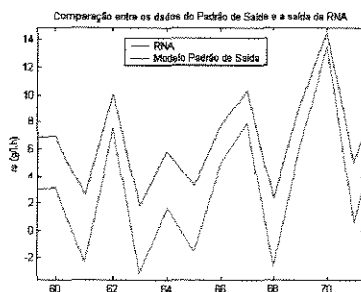
(b) Ampliação da Figura a.



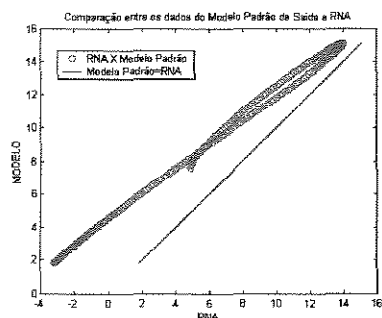
(c) Padrão de Saída / RNA

Validação:

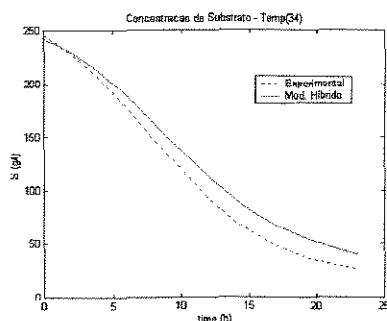
(d) Saídas do padrão e previstas pela rede / número de Amostras



(f) Ampliação da Figura d.



(g) Padrão de Saída / RNA

Modelo Híbrido:Erro Quadrático de
Treinamento $9.80134 \cdot 10^{-6}$ DPR(%)
Treino/Validação

0.2335 / 33.64

Nodos na Camada
Intermediária

9

Algoritmo

Levenberg-Marquardt

Conjunto de Treinamento TIPO II

Figura 4.22 - Resultados do treinamento da rede para a Taxa de Consumo de Substrato e aplicação do Modelo Híbrido para Concentração de Substrato. Temperatura de: 34°C. Conjunto de Treinamento TIPO II

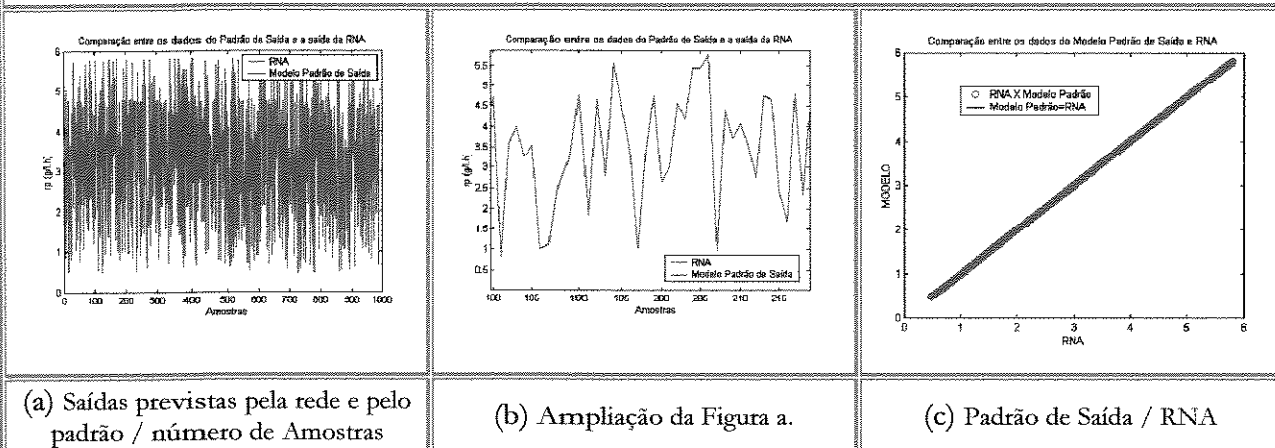
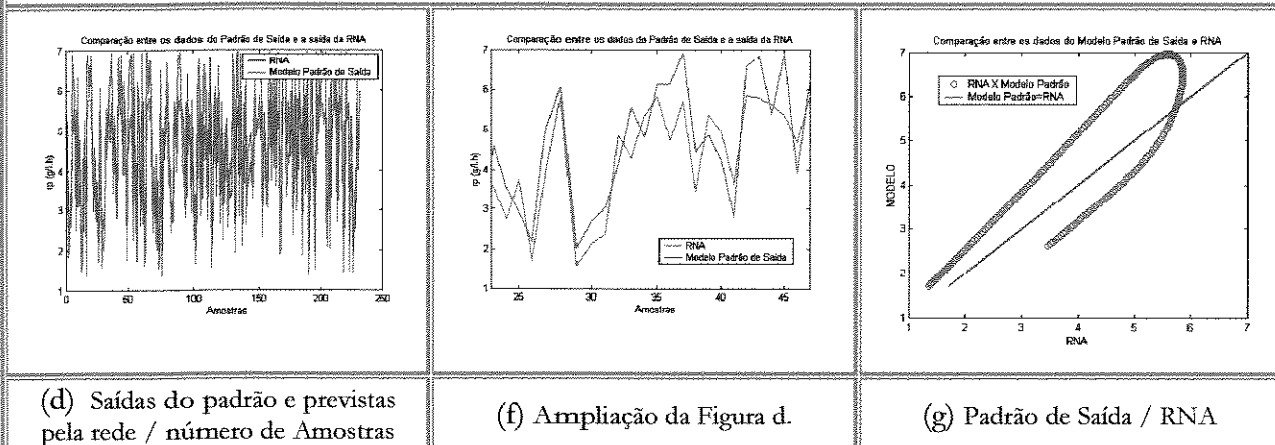
Treino:Validação:Modelo Híbrido:

Figura 4.23 - Resultados do treinamento da rede para a Taxa de Formação de Produto e aplicação do Modelo Híbrido para Concentração de Produto. Temperatura de: 34°C. Conjunto de Treinamento TIPO II.

É conveniente ressaltar que a modelagem de dados experimentais é mais difícil do que o treino de uma rede a partir de dados gerados por simulação.

Os resultados obtidos para o treinamento, validação e para o modelo híbrido usando o conjunto de treinamento onde nenhuma temperatura foi excluída (TIPO I) mostraram-se muito satisfatórios, pois, como pode ser observado das Figuras 4.18 a 4.21, a porcentagem do desvio padrão residual é muito baixa, chegando a valores de até 0.25076%.

Quando os dados de uma temperatura foram omitidos do treinamento (Conjunto TIPO II), como pode ser visto nas Figuras 4.21 a 4.23, com exceção dos resultados de treinamento, os resultados obtidos não se mostram a princípio tão bons quanto os anteriores. Isto pode ser atribuído, conforme já comentado, aos dados da temperatura de validação (34°C) não estarem presentes durante o treino, ou seja, não pertencem ao espaço geométrico reconhecido pela rede, portanto qualquer valor dentro do domínio de temperaturas que não tenha sido adicionado ao treino constitui uma extrapolação para a rede. Contudo, mesmo a rede, não apresentando valores do desvio padrão residual ideais para validação neste caso, os resultados fornecidos para o modelo híbrido foram razoavelmente próximos do desejado, com exceção da concentração de Biomassa (Figura 4.21), o que pode ser mais bem compreendido comparando as Figuras 4.21-f, 4.22-f e 4.23-f. Pode-se reparar que as duas últimas tiveram uma aproximação melhor da saída desejada, contudo não o suficiente para reconhecer perfeitamente os padrões de entrada para o teste realizado. Este comportamento divergente observado para a concentração de biomassa na Figura 4.21, não deve ser atribuído ao algoritmo usado para o aprendizado da rede (Bayesian *framework*), embora para o conjunto de dados TIPO II, esta seja a única saída que usa este algoritmo. Testes anteriores com o algoritmo Levenberg-Marquardt e outros não levaram a resultados melhores.

A seguir as saídas do modelo híbrido são comparadas ao modelo determinístico segundo COSTA (2000) e aos dados experimentais reais, sem interpolação, nas Figuras 4.24 a 4.28. Pode-se notar que para o treinamento feito com o conjunto de dados do TIPO I, os resultados do modelo híbrido foram melhores do que o do modelo determinístico proposto por COSTA (2000).

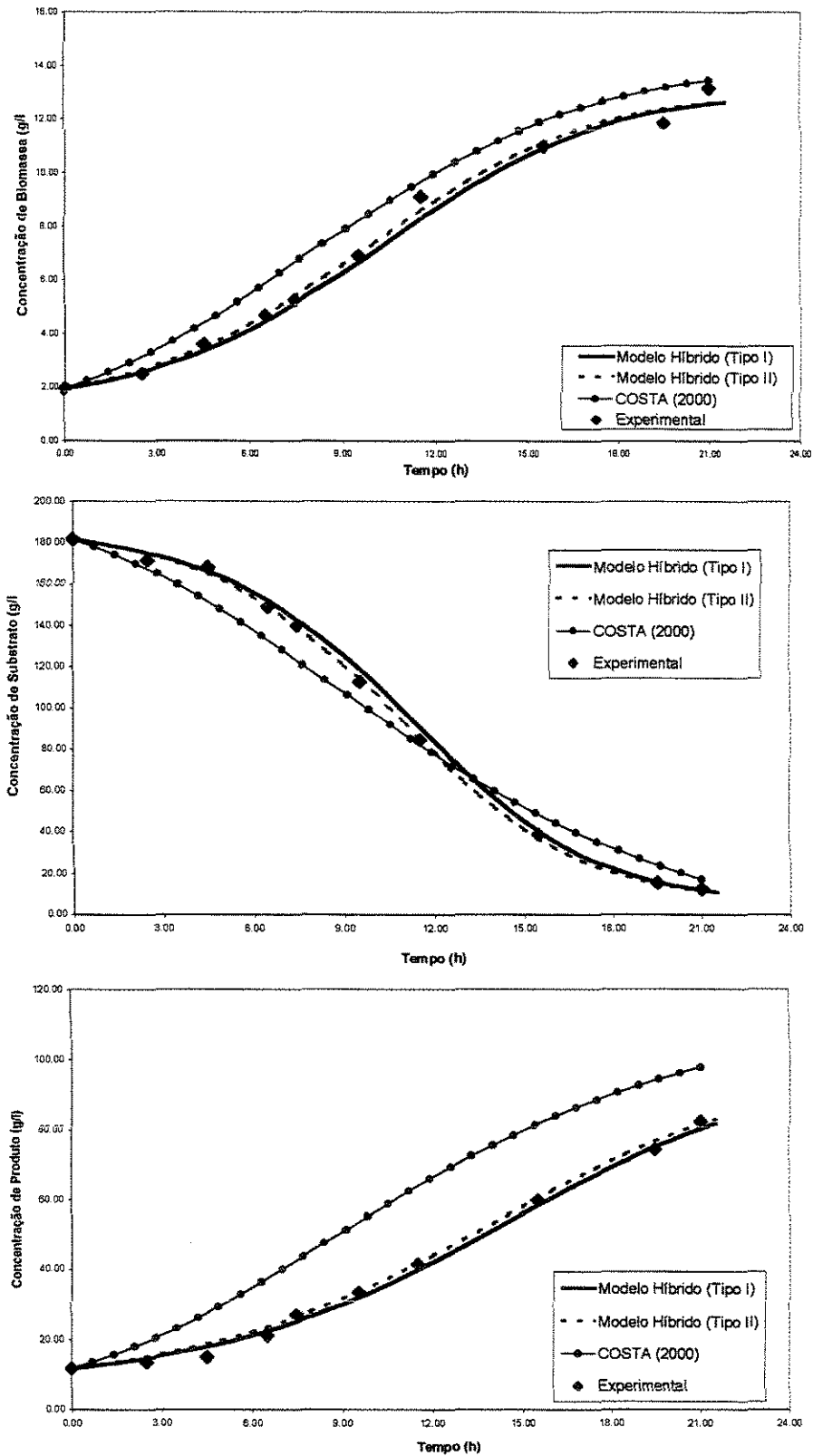


Figura 4.24 – Comparação entre os modelos Híbrido, Determinístico COSTA (2000), e dados Experimentais reais. Temperatura de 28°C.

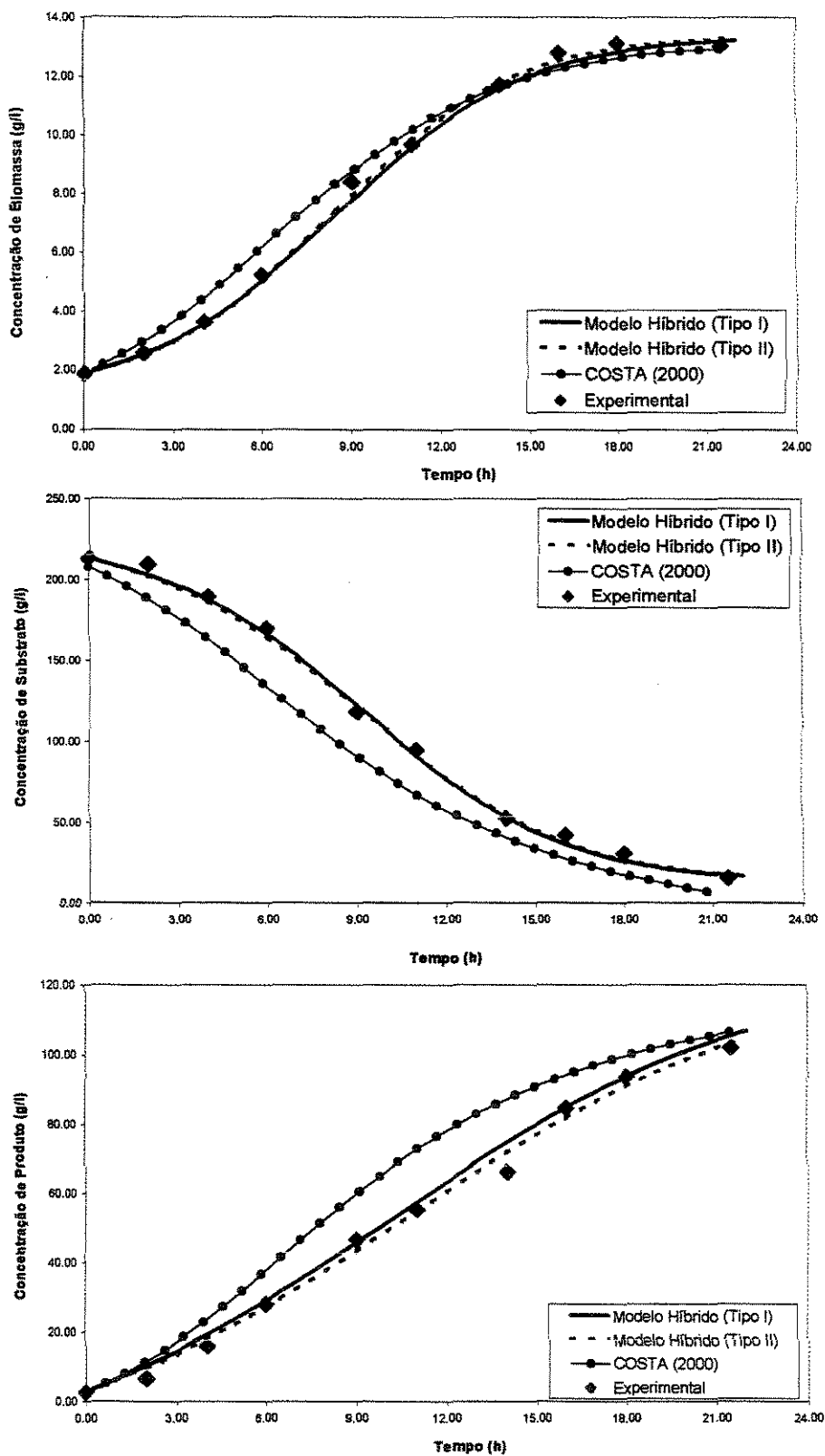


Figura 4.25 – Comparação entre os modelos Híbrido, Determinístico COSTA (2000), e dados Experimentais reais. Temperatura de 31°C.

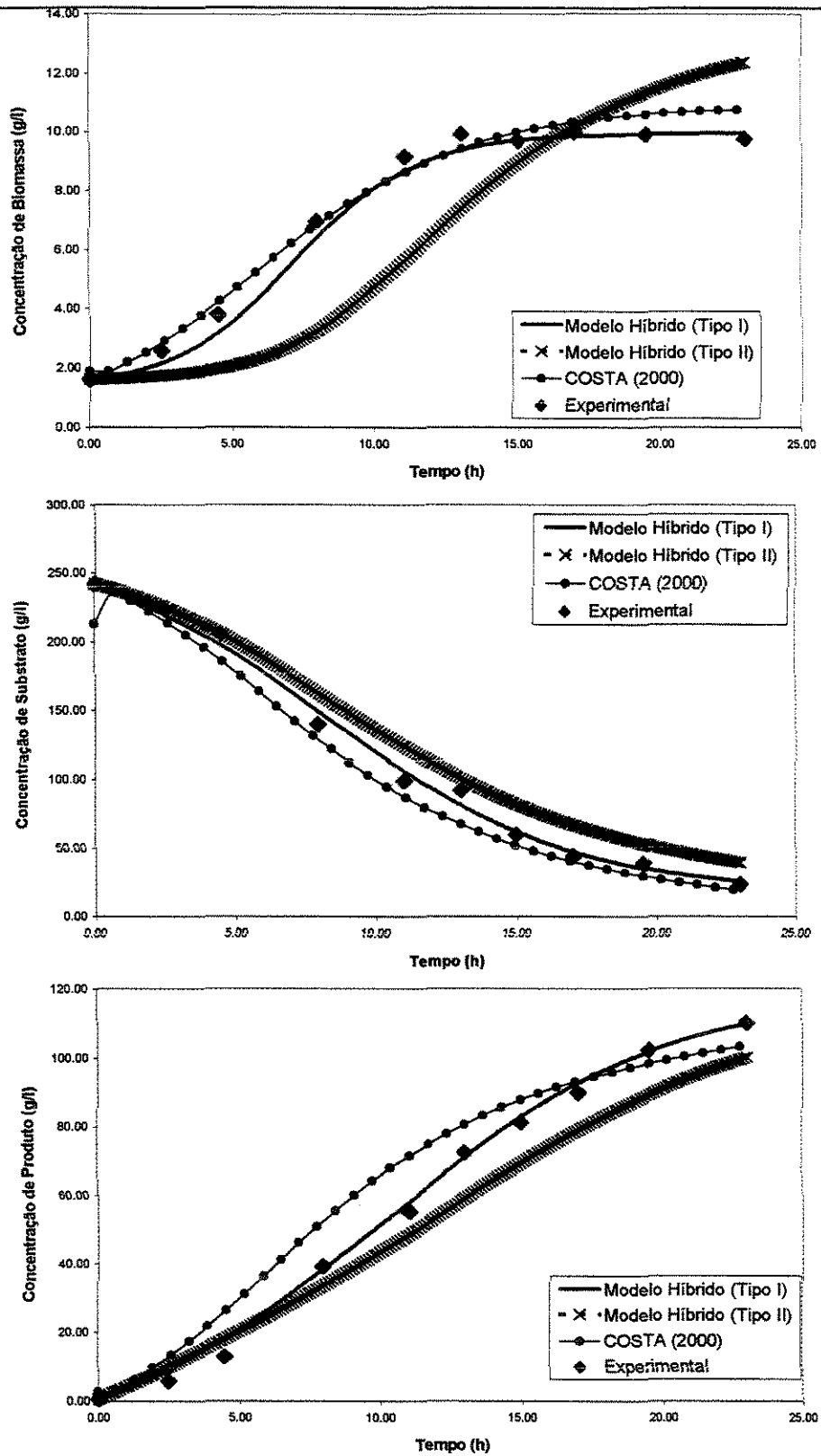


Figura 4.26 – Comparação entre os modelos Híbrido, Determinístico COSTA (2000), e dados Experimentais reais. Temperatura de 34°C.

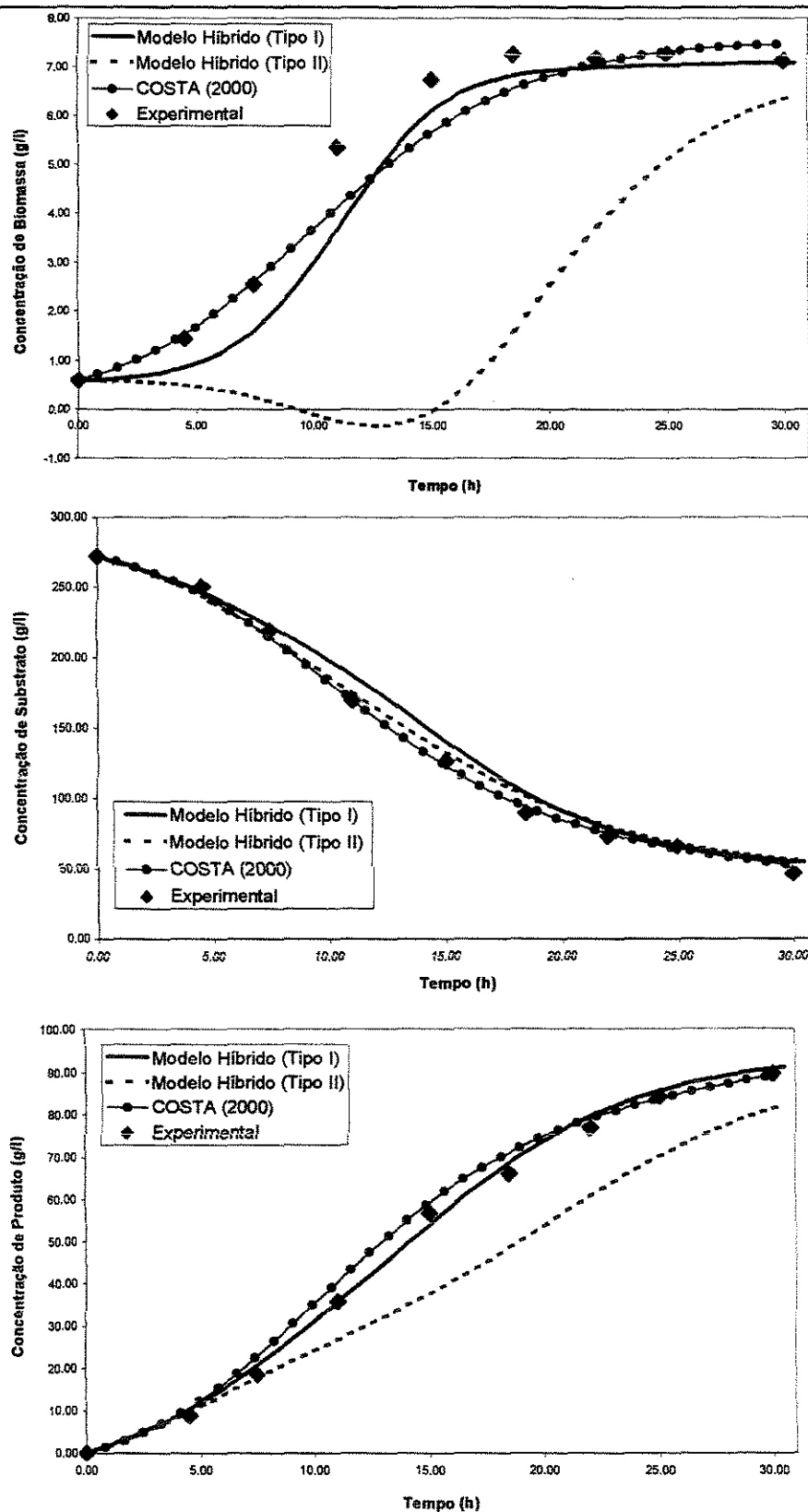


Figura 4.27 – Comparação entre os modelos Híbrido, Determinístico COSTA (2000), e dados Experimentais reais. Temperatura de 37°C.

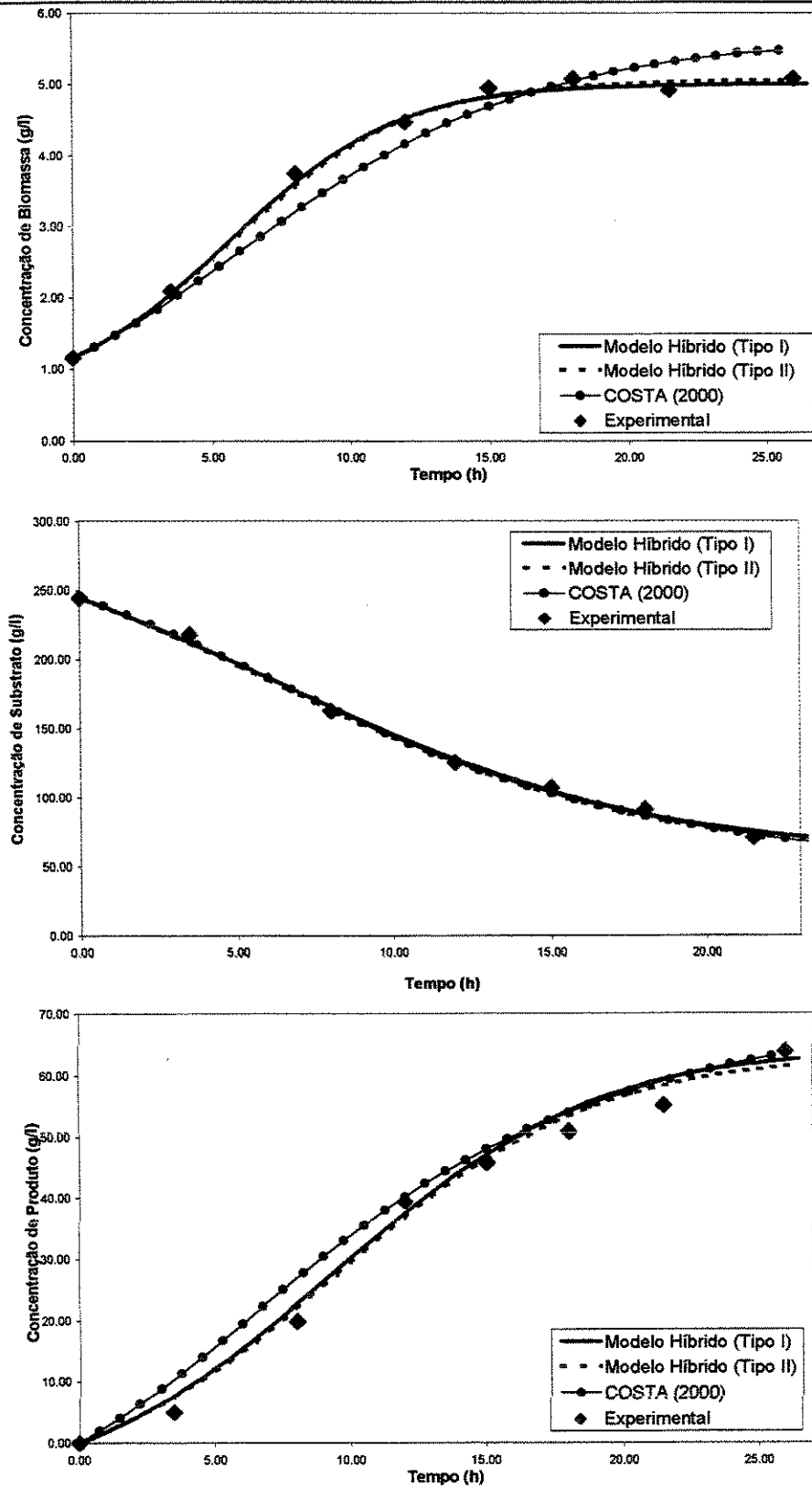


Figura 4.28 – Comparação entre os modelos Híbrido, Determinístico COSTA (2000), e dados Experimentais reais. Temperatura de 40°C.

4.2.4 Modelo Híbrido com Variáveis Secundárias (Software Sensor)

O modelo híbrido com variáveis secundárias também emprega as equações de balanço (equações 3.4 a 3.6) juntamente com as RNAs treinadas para cada taxa cinética. A diferença deste modelo para o anterior (item 4.2.3) está nas variáveis de entrada para a rede neural. Pois como o nome já o diz, este último permite o uso de variáveis secundárias como pH, Turbidez e Brix além da Temperatura. Assim este modelo abre caminho para uma redução de custos com análises laboratoriais além de facilitar o acompanhamento do processo, pois possibilita a estimativa das variáveis de estado a partir de medidas mais simples, como aquelas das variáveis secundárias já citadas. A relativamente fácil disponibilidade comercial de sensores capazes de efetuarem medidas destas variáveis torna ainda mais interessante o desenvolvimento deste modelo, pois permite o acompanhamento do processo em tempo real. O que traz vantagens como economia de matéria prima, e melhora da produção entre outras.

O Fluxograma deste modelo pode ser visto na Figura 4.29.

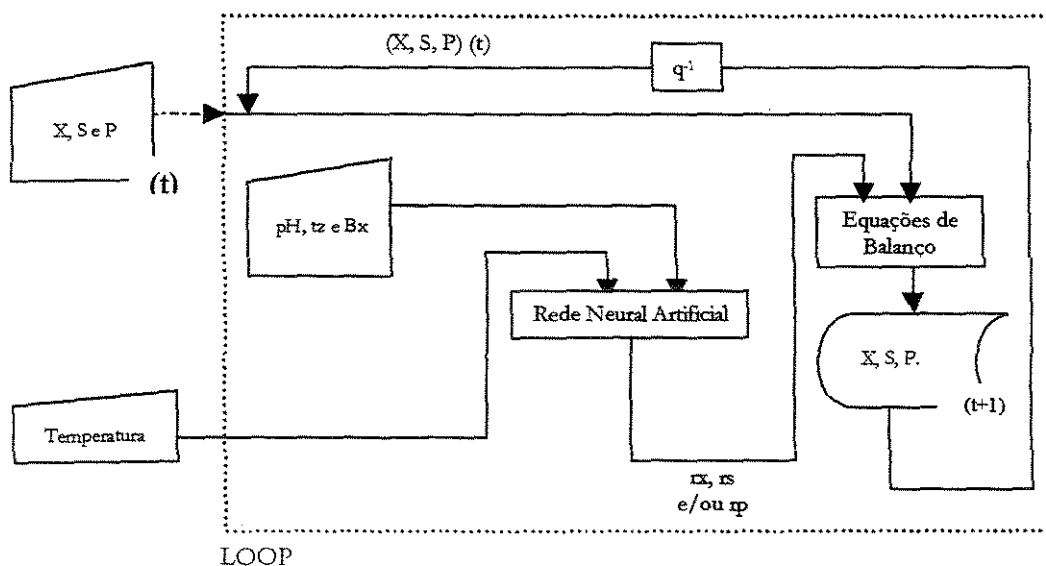


Figura 4.29 - Fluxograma do modelo híbrido com variáveis secundárias na entrada. (tz: turbidez; Bx: Brix; pH: potencial de Hidrogênio).

Para o desenvolvimento do modelo híbrido com variáveis secundárias, novas redes neurais foram treinadas a partir de dados experimentais. Estes dados podem ser observados nas

Tabelas 8.1 a 8.5 em Apêndices e Anexos. O ajuste dos dados experimentais e desenvolvimento do modelo seguiu o mesmo procedimento já relatado para o primeiro modelo híbrido desenvolvido. Portanto os dados de pH, Turbidez e Brix também foram ajustados à equação Sigmoideal-Boltzman (equação 4.1; os parâmetros ajustados estão amostrados na Tabela 4.3) para permitir a interpolação destes e conseqüentemente a expansão do conjunto de dados, além da diminuição dos ruídos próprios de valores experimentais, facilitando o treinamento das redes. Os ajustes feitos às variáveis secundárias estão colocados a seguir nas Figuras 4.30 a 4.34.

Como podem ser observados das Tabelas 8.1 a 8.5, os dados experimentais de pH, Turbidez e Brix estão relacionados aos mesmos instantes de amostragem dos dados de concentrações previamente trabalhados. Assim estas variáveis podem ser igualmente relacionadas às taxas cinéticas já calculadas seguindo o princípio das equações de balanço para a fermentação em batelada, a partir da derivada da equação Sigmoideal-Boltzman (equação 4.2). Desta forma puderam ser estabelecidos os pares de entrada e saída para a elaboração do modelo com redes neurais.

Tabela 4.3 – Parâmetros ajustados para as variáveis secundárias monitoradas experimentalmente, em cada temperatura, para a equação Sigmoideal-Boltzman.

TEMPERATURA (°C)	VARIÁVEIS	PARÂMETROS			
		A_1	A_2	x_0	Dx
28	pH	4,34064	3,81858	11,43347	0,83246
	Turbidez	0,45392	14,52804	13,49574	4,3355
	Brix	24,54016	6,53535	11,86805	6,2346
31	pH	4,73943	3,93437	9,82654	1,86211
	Turbidez	0,63407	15,87943	12,26541	3,78213
	Brix	25,40052	9,79798	10,97877	4,02388
34	pH	4,53993	4,01676	8,64366	1,23799
	Turbidez	-0,15268	11,52025	11,94718	4,67584
	Brix	27,32539	7,60856	13,22505	5,94544
37	pH	4,67789	4,13274	11,37029	1,59428
	Turbidez	-0,70414	8,71047	13,88637	7,21819
	Brix	27,41104	12,86506	12,31955	5,14013
40	pH	4,6293	4,31096	9,93449	2,84118
	Turbidez	-1,67221	6,85237	10,63516	13,28548
	Brix	25,00623	11,849	9,21855	7,361

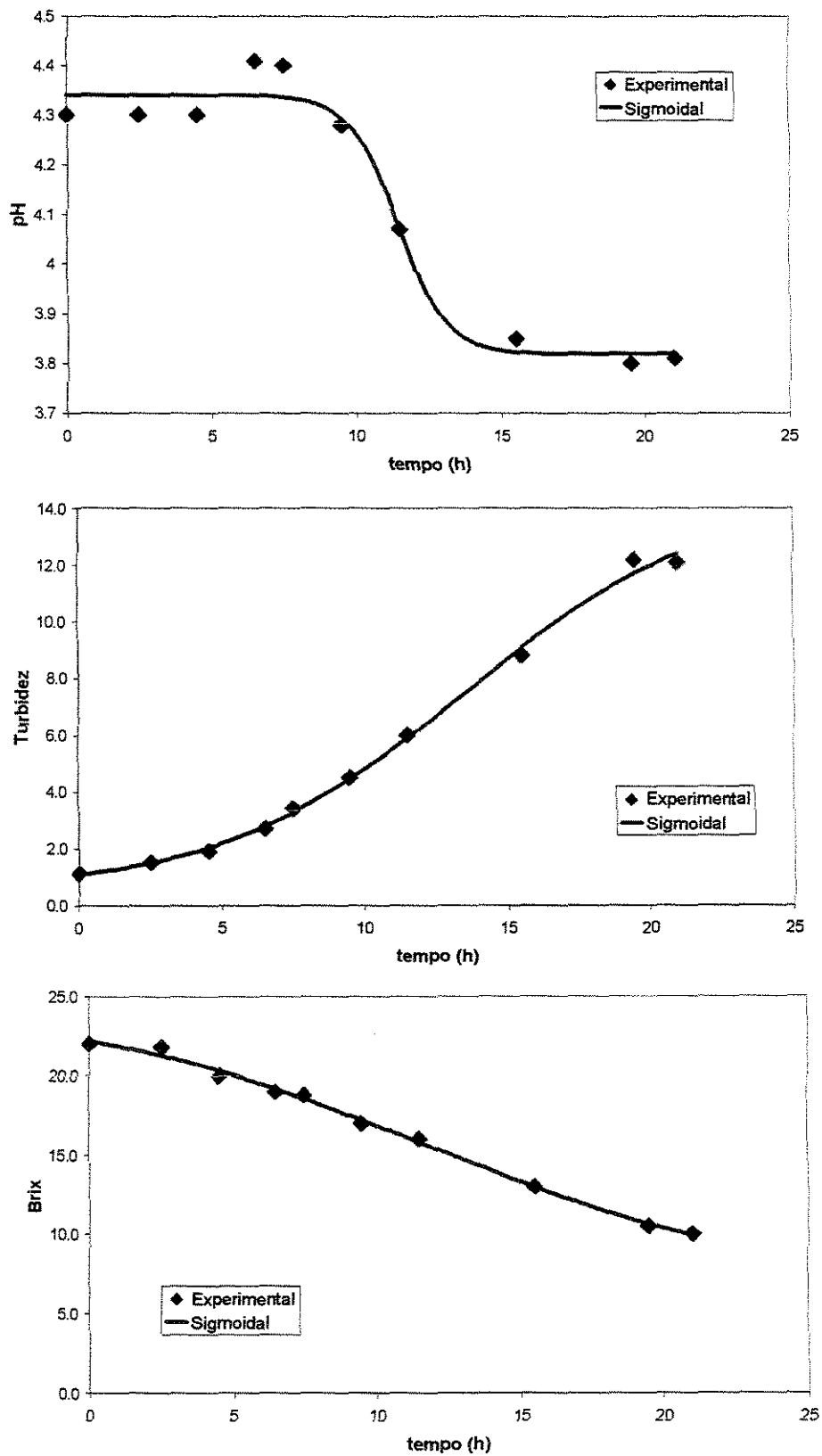


Figura 4.30 - Ajustes para as variáveis secundárias de a) pH; b) Turbidez e c) Brix. Temperatura de 28°C.

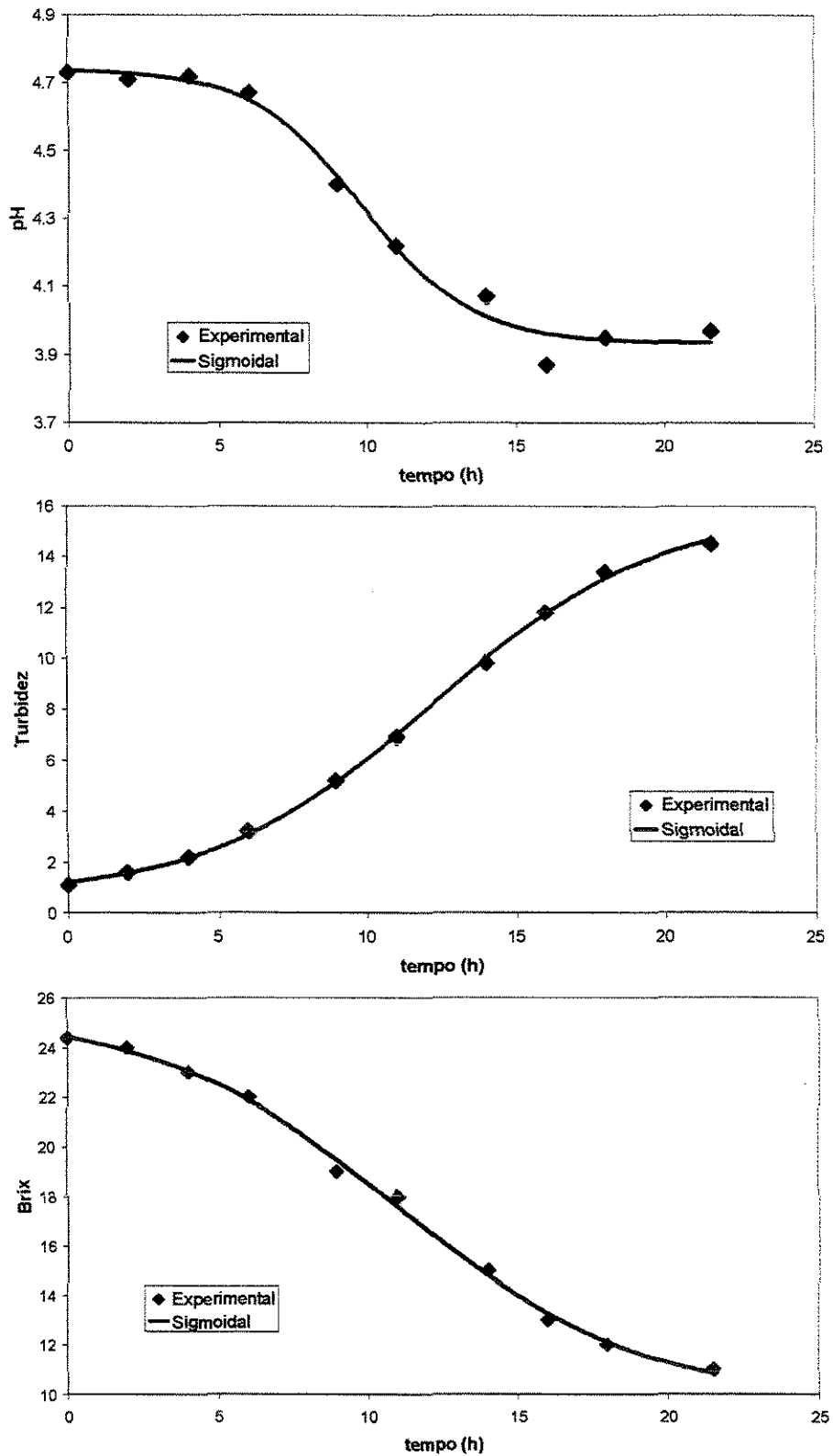


Figura 4.31 - Ajustes para as variáveis secundárias de a) pH; b) Turbidez e c) Brix. Temperatura de 31°C.

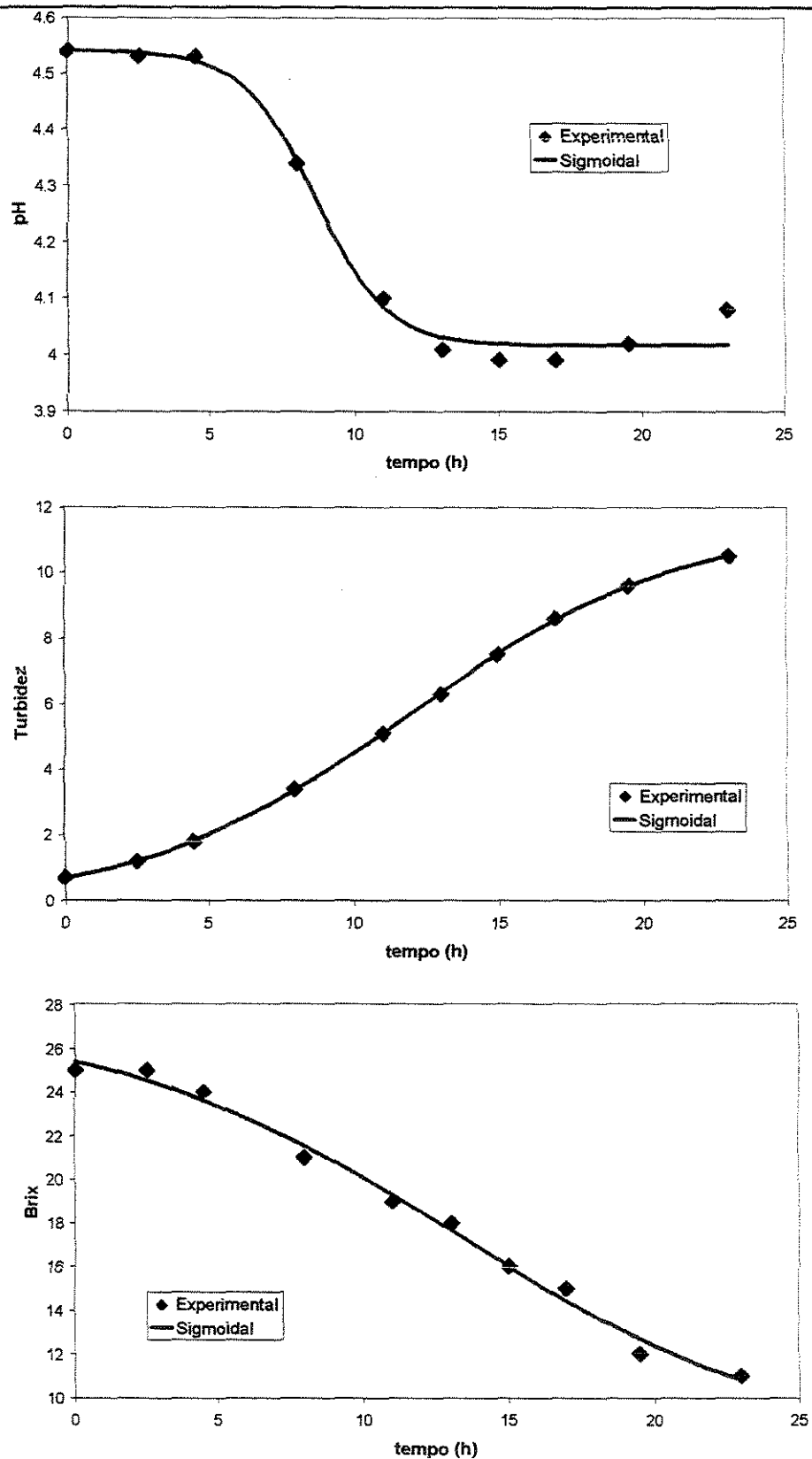


Figura 4.32 - Ajustes para as variáveis secundárias de a) pH; b) Turbidez e c) Brix. Temperatura de 34°C.

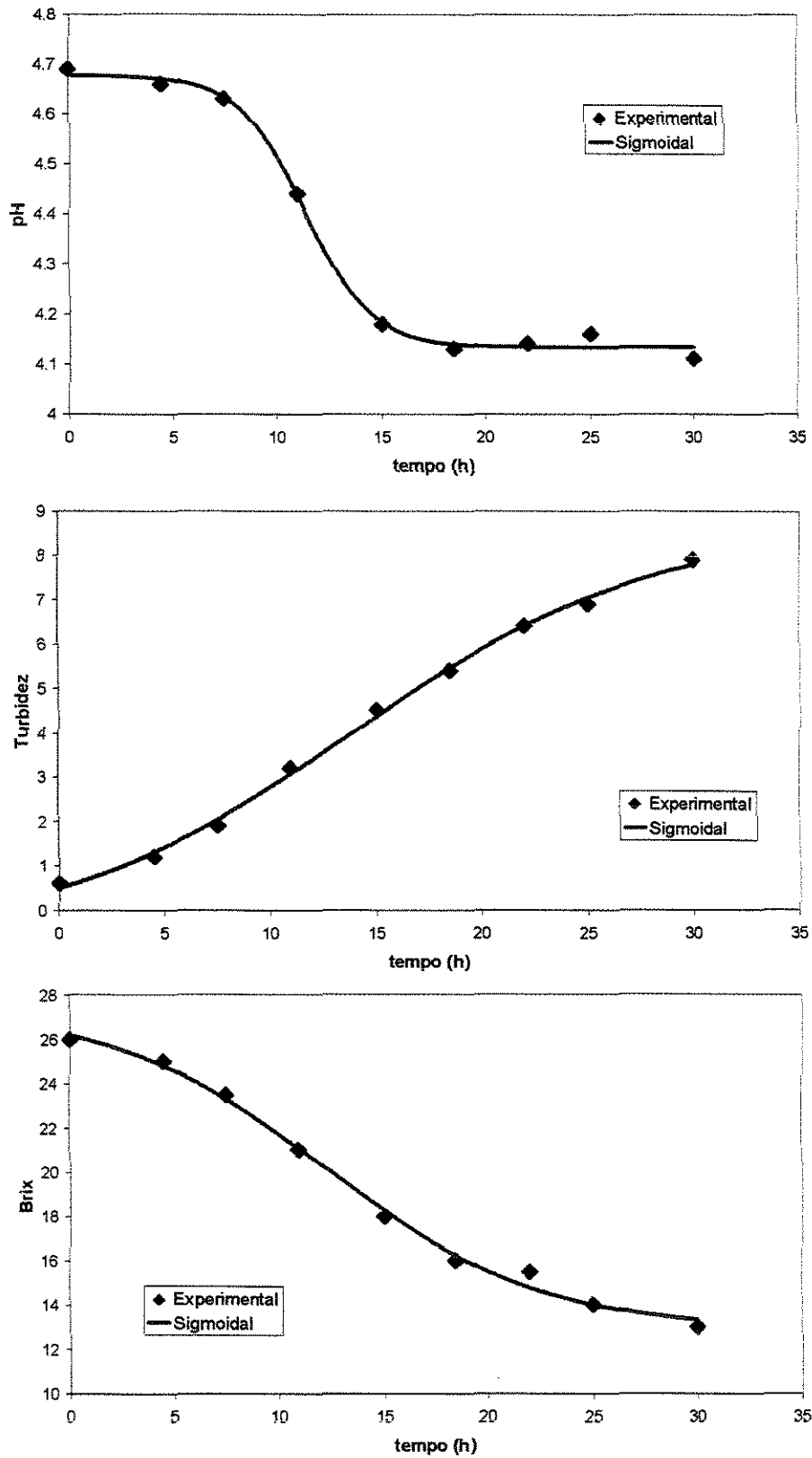


Figura 4.33 - Ajustes, para as variáveis secundárias de a) pH; b) Turbidez e c) Brix. Temperatura de 37°C.

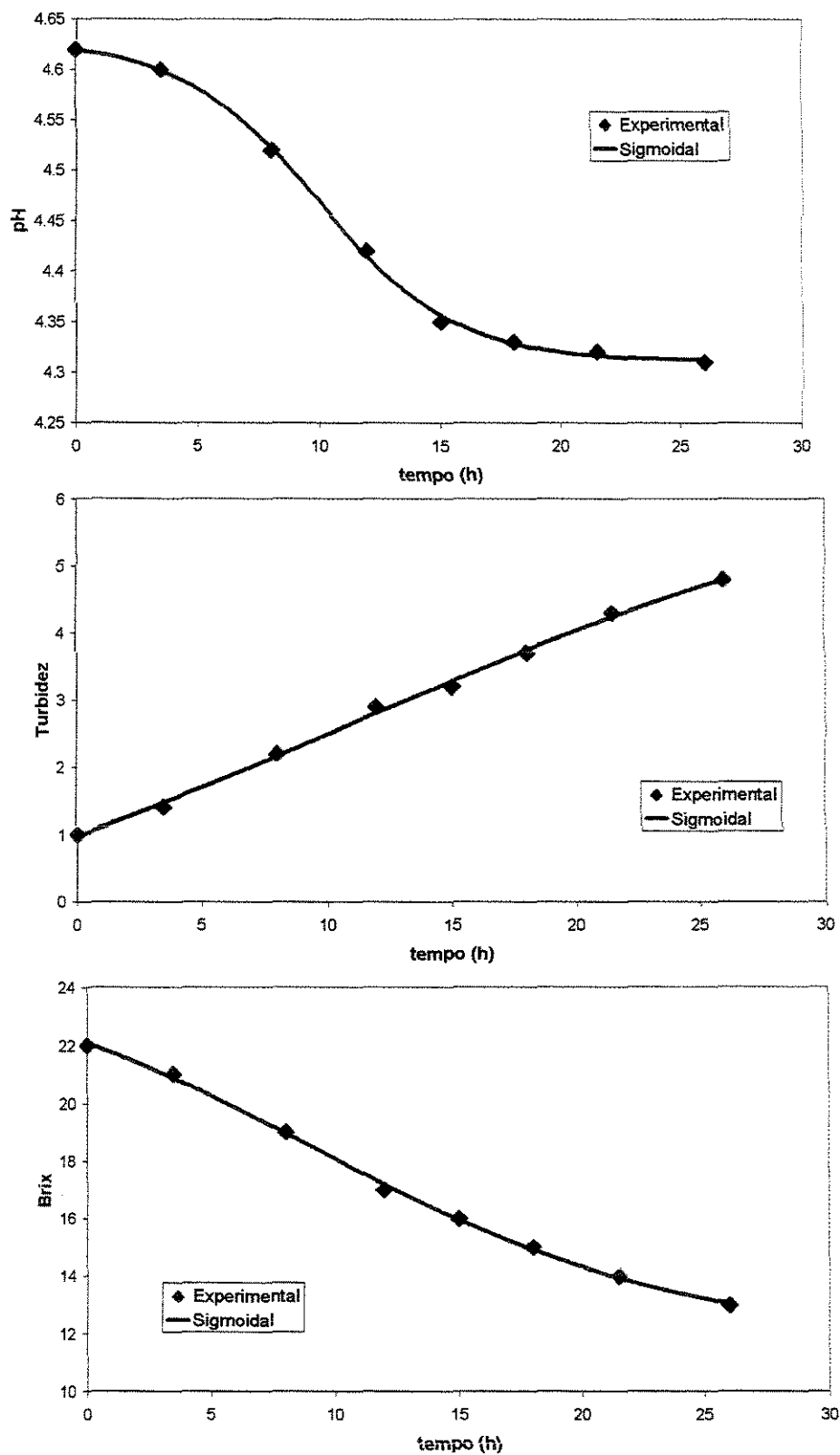
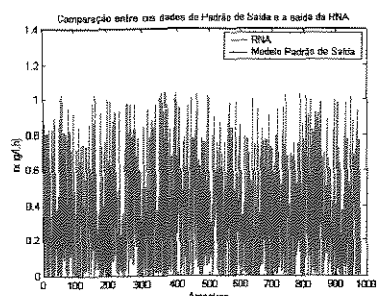
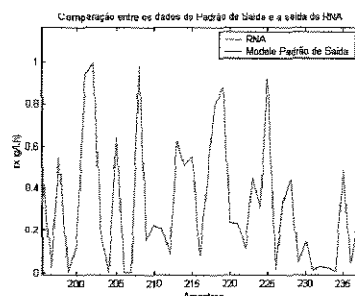


Figura 4.34 - Ajustes, para as variáveis secundárias de a) pH; b) Turbidez e c) Brix. Temperatura de 40°C.

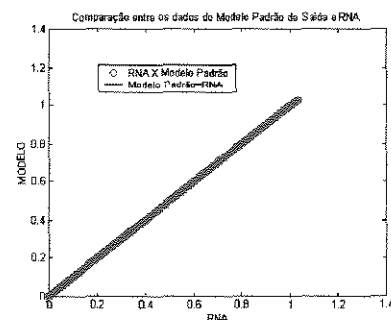
A partir dos dados obtidos pelo ajuste da equação Sigmoidal-Boltzman, as redes neurais artificiais, para as variáveis secundárias, foram treinadas, sendo criadas novamente redes específicas para cada taxa cinética. Devido aos resultados obtidos durante o desenvolvimento do primeiro modelo híbrido (com variáveis primárias), desta vez não foram omitidos os dados referentes a alguma temperatura em específico do conjunto de treinamento, como antes havia sido retirados os dados da temperatura de 34°C. Assim, permitindo a rede mapear por completo o espaço geométrico estudado. Deve ser ressaltado, entretanto que para a validação foram criados conjuntos contendo valores retirados aleatoriamente da matriz de dados inicial, porém não de uma única temperatura. Tais conjuntos não participaram do treinamento, servindo apenas à validação. São apresentadas a seguir, Figuras 4.35 a 4.37, os resultados para a temperatura de 34°C, ilustrando o treinamento e validação das redes além do respectivo modelo híbrido com variáveis secundárias desenvolvido. Os resultados apresentados novamente são referentes somente à temperatura de 34°C, pois como já justificado, esta se encontra no centro do domínio das temperaturas estudadas, além dos gráficos para as demais temperaturas não evidenciarem nenhuma outra conclusão diferente.

Treino:

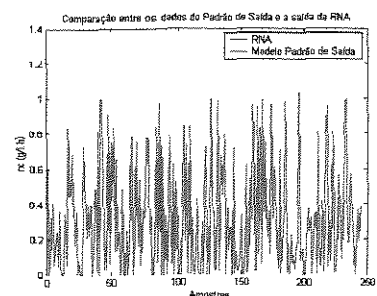
(a) Saídas previstas pela rede e pelo padrão / número de Amostras



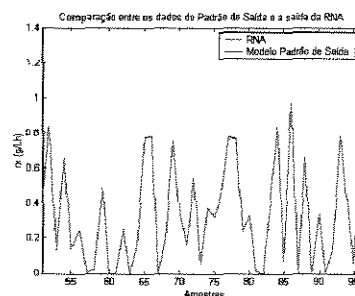
(b) Ampliação da Figura a.



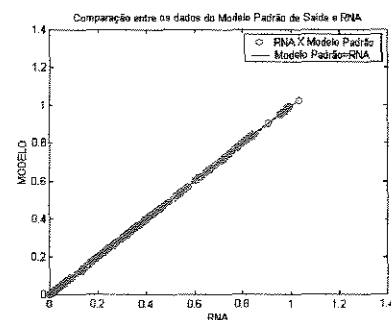
(c) Padrão de Saída / RNA

Validação:

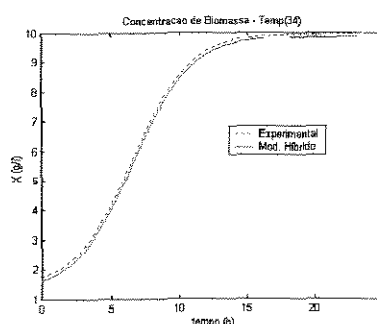
(d) Saídas do padrão e previstas pela rede / número de Amostras



(f) Ampliação da Figura d.



(g) Padrão de Saída / RNA

Modelo Híbrido com Variáveis Secundárias:Erro Quadrático de
Treinamento $9,99189 \cdot 10^{-7}$ DPR(%)
Treino/Validação

0.37715 / 0.37774

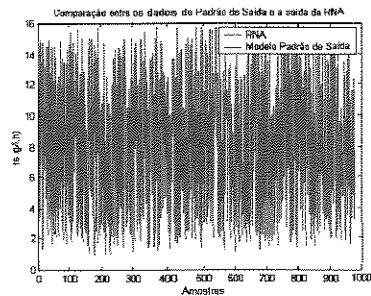
Nodos na Camada
Intermediária

9

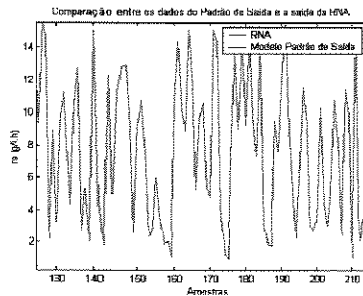
Algoritmo

Levenberg-Marquardt

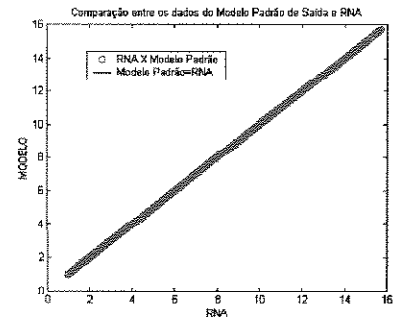
Figura 4.35 - Resultados do treinamento da rede para a Taxa de Crescimento de Biomassa e aplicação do Modelo Híbrido com variáveis secundárias para Concentração de Biomassa. Temperatura de: 34°C.

Treino:

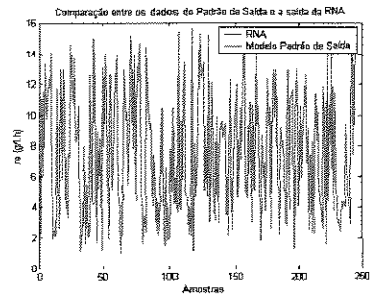
(a) Saídas previstas pela rede e pelo padrão / número de Amostras



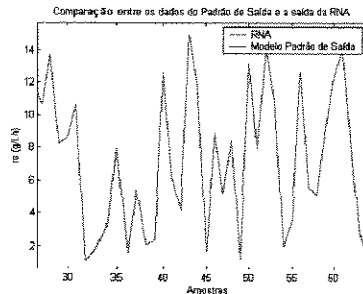
(b) Ampliação da Figura a.



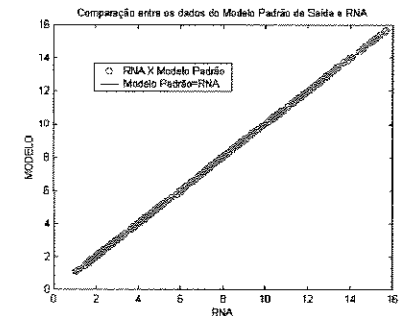
(c) Padrão de Saída / RNA

Validação:

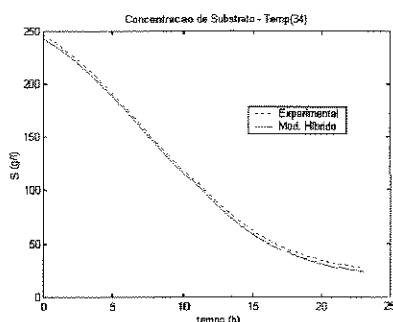
(d) Saídas do padrão e previstas pela rede / número de Amostras



(f) Ampliação da Figura d.



(g) Padrão de Saída / RNA

Modelo Híbrido com Variáveis Secundárias:Erro Quadrático de
Treinamento $9,99145 \cdot 10^{-7}$ DPR(%)
Treino/Validação

0.2231 / 0.2484

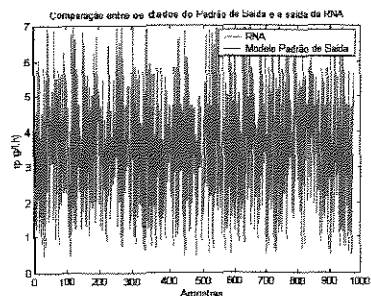
Nodos na Camada
Intermediária

9

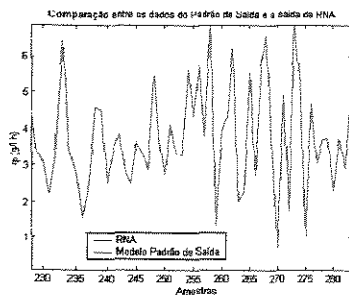
Algoritmo

Levenberg-Marquardt

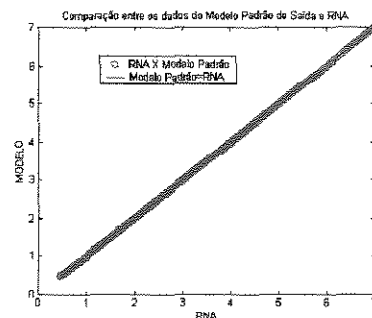
Figura 4.36 - Resultados do treinamento da rede para a Taxa de Consumo de Substrato e aplicação do Modelo Híbrido com variáveis secundárias para Concentração de Substrato. Temperatura de: 34°C.

Treino:

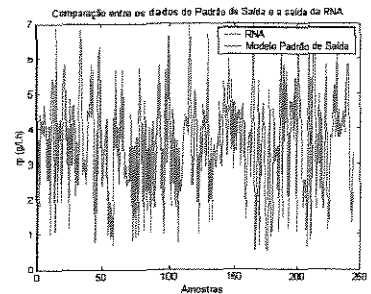
(a) Saídas previstas pela rede e pelo padrão / número de Amostras



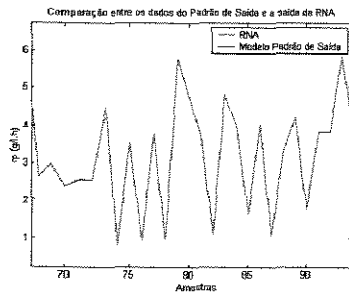
(b) Ampliação da Figura a.



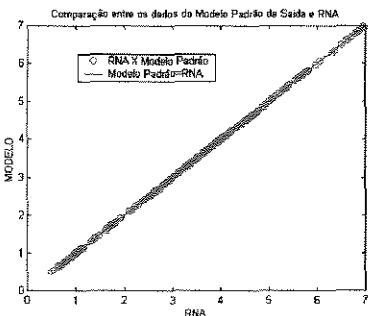
(c) Padrão de Saída / RNA

Validação:

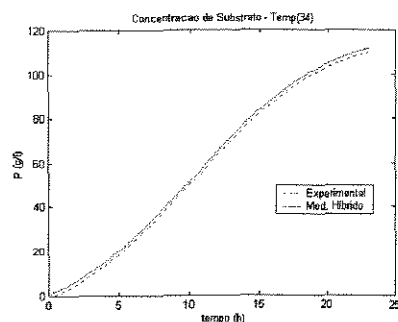
(d) Saídas do padrão e previstas pela rede / número de Amostras



(f) Ampliação da Figura d.



(g) Padrão de Saída / RNA

Modelo Híbrido com Variáveis Secundárias:Erro Quadrático de
Treinamento $9,57335 \cdot 10^{-7}$ DPR(%)
Treino/Validação

0.2230 / 0.2494

Nodos na Camada
Intermediária

9

Algoritmo

Levenberg-Marquardt

Figura 4.37 - Resultados do treinamento da rede para a Taxa de Formação de Produto e aplicação do Modelo Híbrido com variáveis secundárias para Concentração de Produto. Temperatura de: 34°C.

As Figuras 4.38 a 4.42 comparam as saídas do modelo híbrido que usa variáveis secundárias, ao modelo determinístico segundo COSTA(2000) e ATALA (2000), aos dados experimentais reais e aos dados experimentais ajustados a equação Sigmoidal-Boltzman.

O modelo desenvolvido que emprega as variáveis secundárias como pH, Turbidez e Brix, apresenta resultados melhores quando comparados ao modelo híbrido Tipo I com variáveis primárias. Isto pode ser evidenciado comparando-se os resultados obtidos para as temperaturas de 34 e 37°C, Figuras 4.26 e 4.27, aos resultados para as mesmas temperaturas das Figuras 4.40 e 4.41, do modelo híbrido que utiliza as variáveis secundárias; para as demais temperaturas o desempenho foi semelhante. Esta melhora de desempenho do modelo não deve considerar qualquer diferença de função, arquitetura ou algoritmo de treinamento das redes neurais, pois as redes empregadas para o desenvolvimento deste segundo modelo foram exatamente iguais aquelas já relatadas anteriormente (item 4.2.3). Os resultados obtidos mostram que as variáveis secundárias são adequadas para o desenvolvimento do modelo híbrido, levando a resultados semelhantes aos obtidos com as variáveis primárias. Por se tratarem de variáveis secundárias são, portanto, de fácil medida e não embutem um erro muito grande em suas leituras, além de, como pode ser observado na Figura 4.29, serem adicionadas ao modelo em tempo real, permitindo assim um melhor acompanhamento do processo. Nisto se traduz uma das vantagens da utilização de variáveis secundárias para a modelagem do processo.

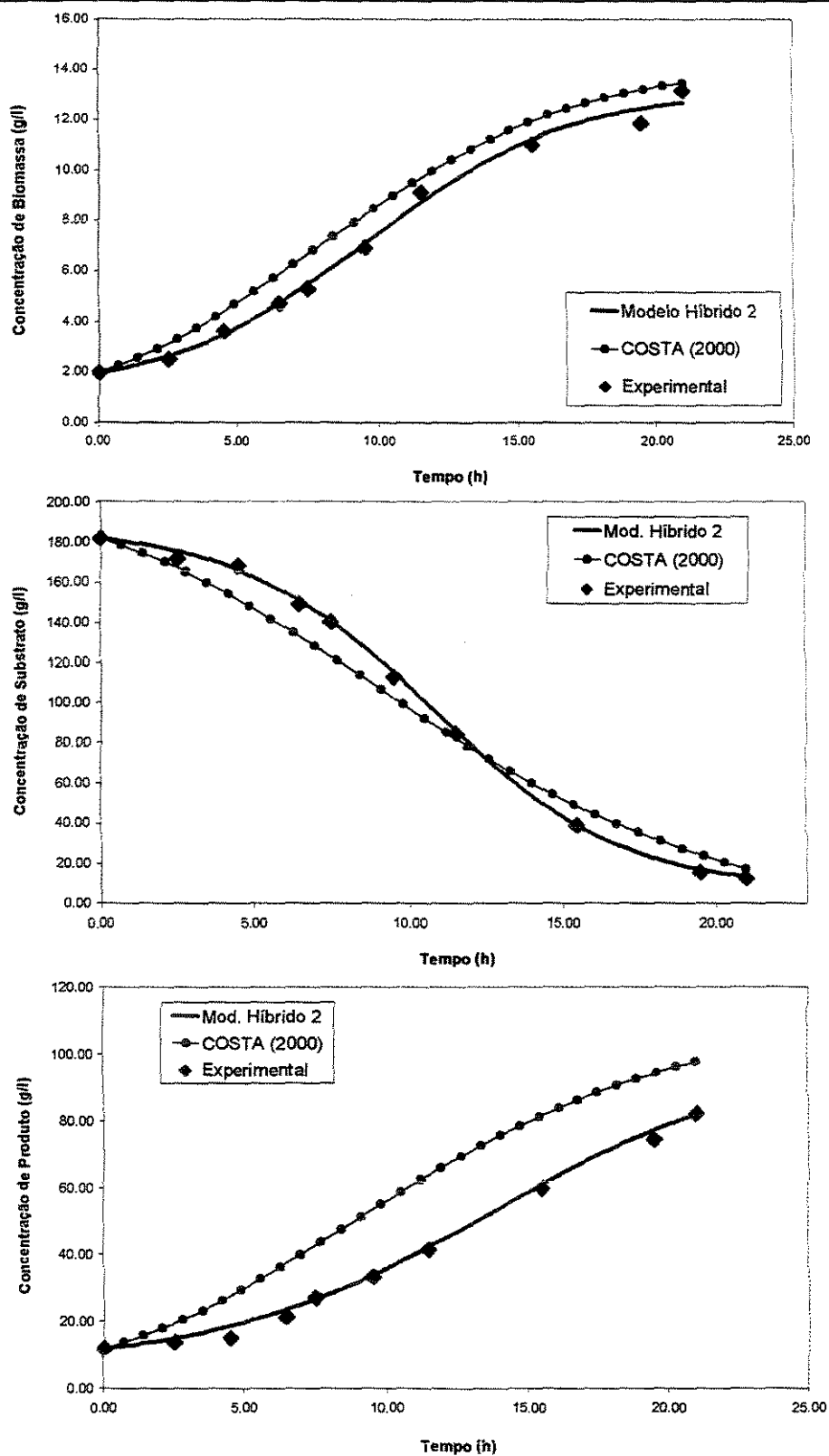


Figura 4.38 – Comparação entre os modelos Híbrido com variáveis secundárias, ajuste aos dados experimentais com a Equação Sigmoidal-Boltzman, Determinístico COSTA (2000) e dados experimentais. Temperatura de 28°C.

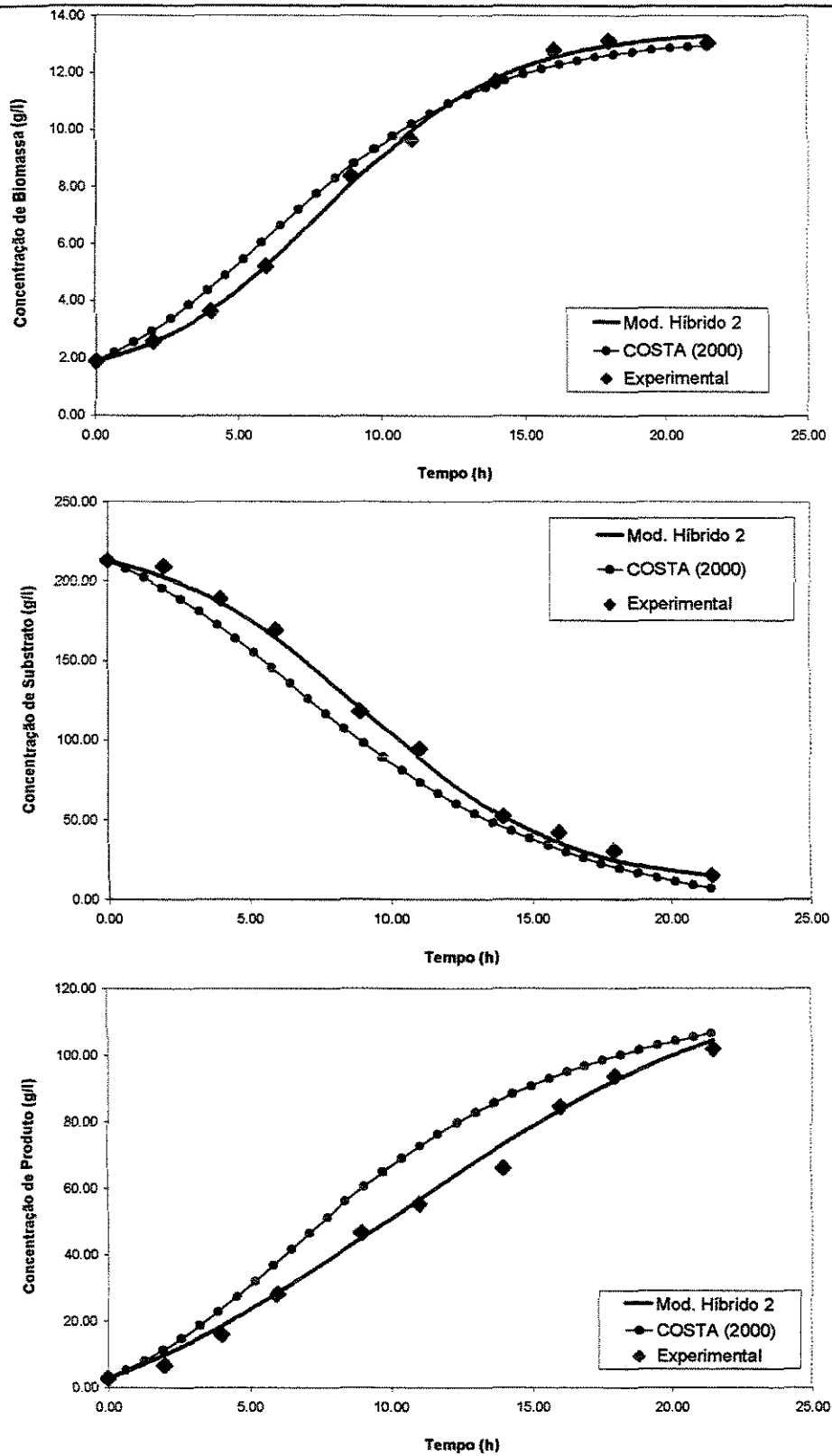


Figura 4.39 – Comparação entre os modelos Híbrido com variáveis secundárias, ajuste aos dados experimentais com a Equação Sigmoidal-Boltzman, Determinístico COSTA (2000) e dados experimentais. Temperatura de 31°C.

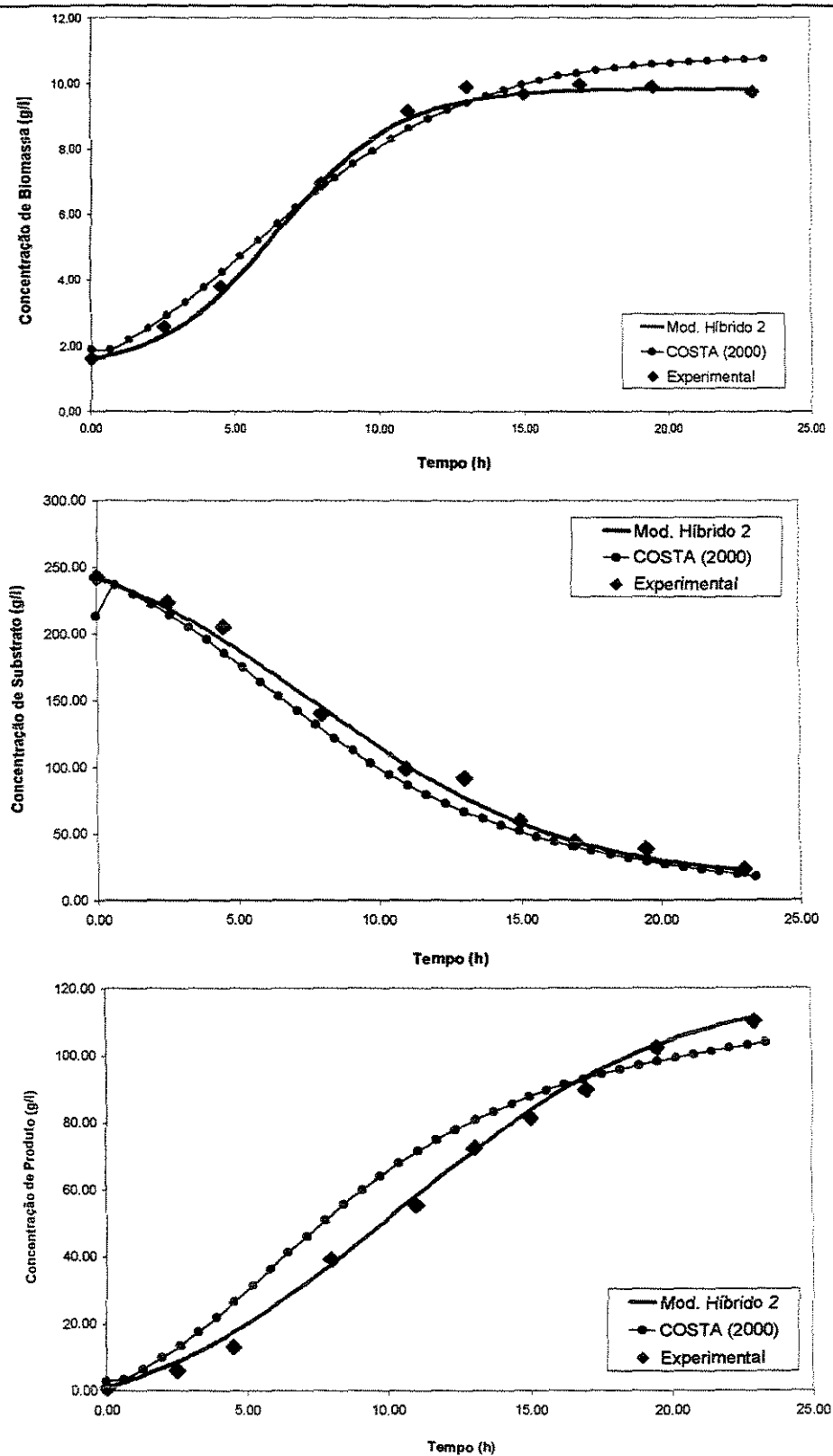


Figura 4.40 – Comparação entre os modelos Híbrido com variáveis secundárias, ajuste aos dados experimentais com a Equação Sigmoidal-Boltzman, Determinístico COSTA (2000) e dados experimentais. Temperatura de 34°C.

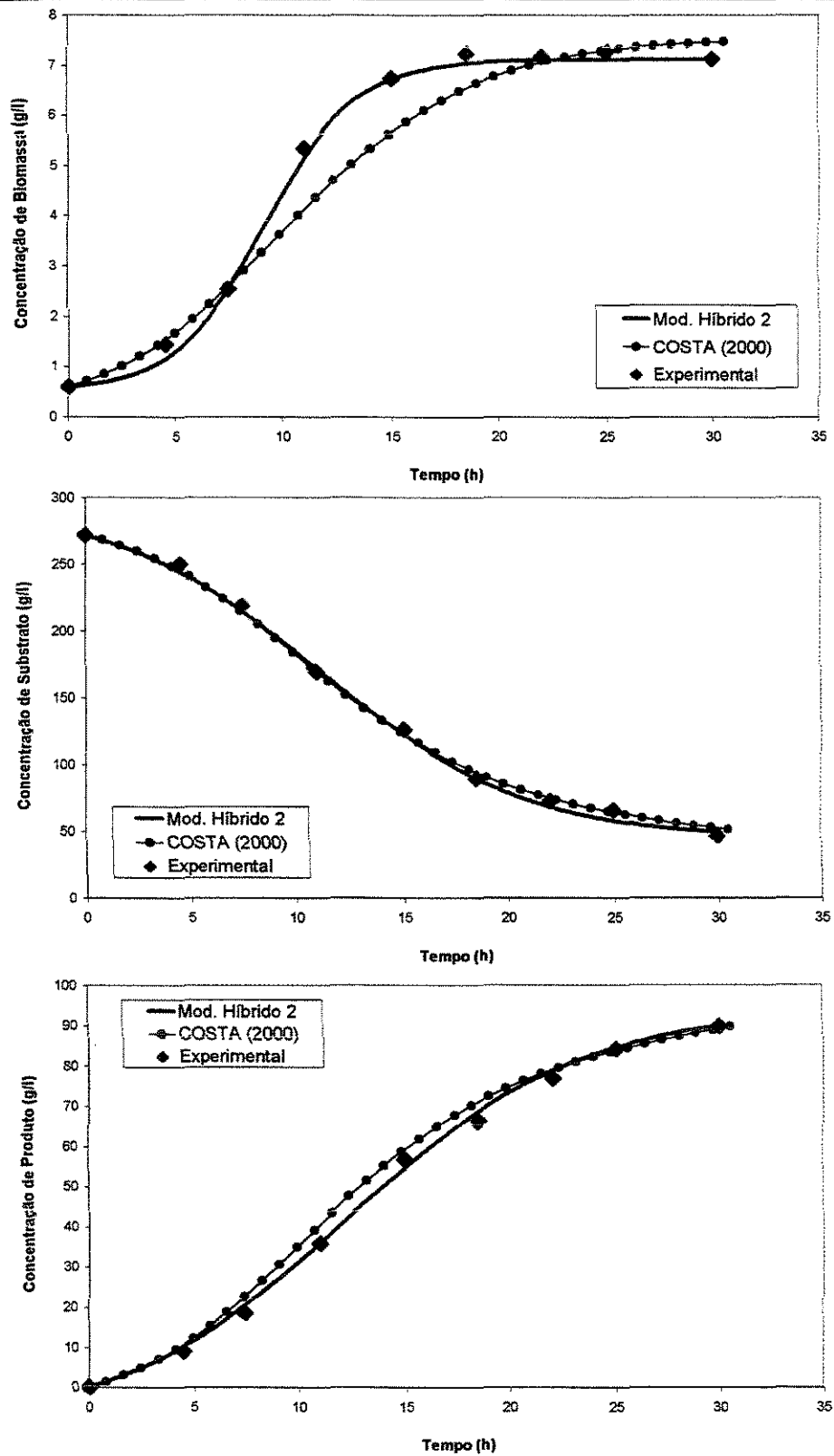


Figura 4.41 – Comparação entre os modelos Híbrido com variáveis secundárias, ajuste aos dados experimentais com a Equação Sigmoidal-Boltzman, Determinístico COSTA (2000) e dados experimentais. Temperatura de 37°C.

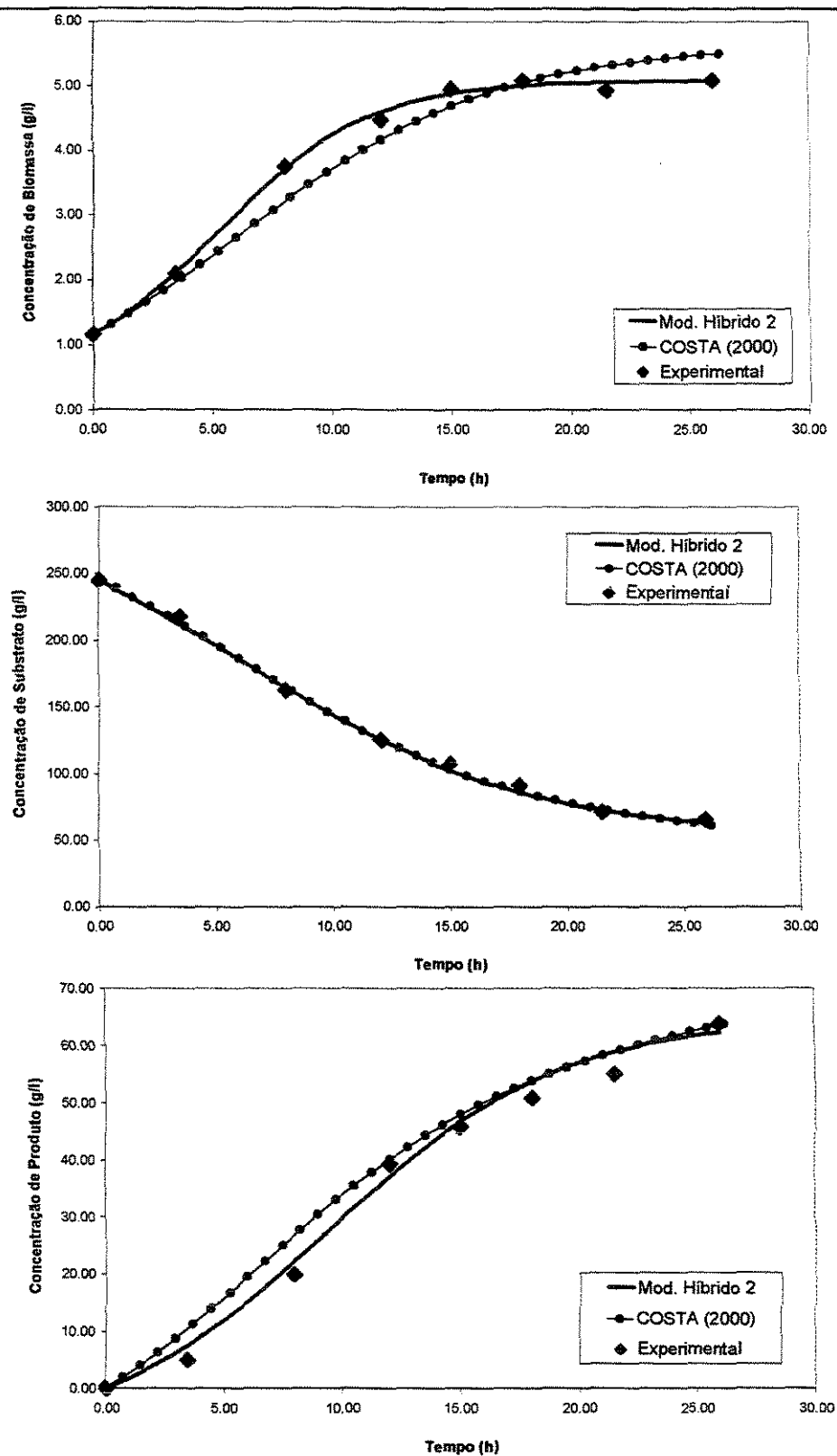


Figura 4.42 – Comparação entre os modelos Híbrido com variáveis secundárias, ajuste aos dados experimentais com a Equação Sigmoidal-Boltzman, Determinístico COSTA (2000) e dados experimentais. Temperatura de 40°C.

5 ESTUDO DE TÉCNICAS DE OTIMIZAÇÃO USANDO MODELOS GERADOS POR PLANEJAMENTO FATORIAL

Um dos objetivos deste trabalho é avaliar o uso de modelos gerados a partir de planejamento fatorial para a otimização do processo usando programação quadrática sucessiva (EDGAR & HIMMELBLAU, 1989). Planeja-se comparar os resultados desta abordagem com o da otimização usando-se a metodologia das superfícies de resposta (BARROS NETO et al., 1995), que normalmente é o método empregado usando-se estes modelos.

Pretende-se, nesta seção, inicialmente, determinar um modelo quadrático da concentração de produto no final da batelada em função das variáveis de entrada de interesse usando-se os conceitos de planejamento fatorial. Para tanto um estudo da influência das principais variáveis do modelo cinético através de um planejamento fatorial 2^{4+0} de resolução completa com um total de 17 ensaios sendo um no ponto central foi realizado. Durante este estudo o objetivo foi verificar o efeito, a importância e a interferência de cada variável sobre a resposta, concentração de produto final, etanol. As respostas de concentração necessárias ao planejamento foram fornecidas pela simulação em computador do modelo matemático de COSTA (2000) e ATALA (2000).

O passo seguinte é estudar formas de otimização, usando a metodologia das superfícies de resposta e programação quadrática sucessiva.

5.1 Estudo para Identificação das Variáveis Relevantes

As variáveis independentes estudadas são:

- T (Temperatura, °C)
- X_0 (Concentração de Biomassa, g/l)
- S_0 (Concentração de Substrato, g/l)
- P_0 (Concentração de Produto, g/l)

A variável dependente é: P_F (Concentração de Produto Final, g/l)

O planejamento experimental permite estudar muitos fatores simultaneamente, quantificar seus efeitos e investigar suas possíveis interações, KALIL et al. (2000) citando BOX et al. (1978).

Os efeitos das variáveis acima foram verificados para a produção de álcool etílico utilizando um planejamento fatorial (2^{4+0}). Cada variável independente é testada em dois níveis, um alto (+) e um baixo (-), como mostra a Tabela 5.1. A faixa dos valores estudados obedece aos dados experimentais. O *software* STATISTICA (Statsoft, v. 5.0) foi empregado para analisar os dados.

Tabela 5.1 – Variáveis e Níveis para o Planejamento Fatorial.

Nível	T (°C)	X_0 (g/l)	S_0 (g/l)	P_0 (g/l)
-1	28	0.59	181.68	0
0	34	1.271	226.9	5.85
+1	40	1.952	272.12	11.7

Deve-se ressaltar que a faixa usada para concentração inicial de biomassa (X_0) não corresponde à realidade industrial, que trabalha com valores de X_0 em torno de 40 g/l. No entanto, o modelo de COSTA (2000) e ATALA (2000), usado neste estudo, não é válido nesta faixa, e trabalhou-se com a faixa de X_0 usada no trabalho experimental realizado para desenvolver este modelo.

A seguir é mostrada a Tabela 5.2 com o primeiro planejamento experimental. Este planejamento pôde definir as variáveis relevantes.

A Tabela 5.2 apresenta os experimentos (simulações) bem como eles devem ser conduzidos para o planejamento dos 17 ensaios; a última coluna contém os resultados obtidos para a concentração de produto final (g/l).

Tabela 5.2 – Primeiro Planejamento Fatorial.

Ensaio	Variáveis independentes				Variável dependente
	T (°C)	X ₀ (g/l)	S ₀ (g/l)	P ₀ (g/l)	P _F (g/l)
1	40	1.952	272.12	11.70	69.11
2	40	1.952	272.12	0	68.99
3	40	1.952	181.68	11.70	69.14
4	40	1.952	181.68	0	63.04
5	40	0.59	272.12	11.70	69.24
6	40	0.59	272.12	0	69.11
7	40	0.59	181.68	11.70	69.24
8	40	0.59	181.68	0	62.79
9	28	1.952	272.12	11.70	113.09
10	28	1.952	272.12	0	112.86
11	28	1.952	181.68	11.70	105.98
12	28	1.952	181.68	0	94.76
13	28	0.59	272.12	11.70	112.92
14	28	0.59	272.12	0	112.68
15	28	0.59	181.68	11.70	106.68
16	28	0.59	181.68	0	94.98
17 (Central)	34	1.271	226.90	5.85	107.04

Os efeitos das variáveis sobre a saída estão apresentados na Tabela 5.3 para um nível de confiança de 95%. Na Figura 5.1 está apresentado um histograma dos efeitos observados.

Tabela 5.3 – Efeitos principais para o Primeiro planejamento Fatorial.

	Efeito	Desvio Padrão	t(12)	p	-95%	+95%
Média	88.2976*	1.636183*	53.9656*	0.000000*	84.7327*	91.8626*
T	-39.0875*	3.373077*	-11.5881	0.000000*	-46.4368*	-31.7382*
X ₀	-0.0100	3.3736077	-0.0030	0.997683	-7.3593	7.3393
S ₀	7.7475*	3.3736077*	2.2969*	0.040426*	0.3982*	15.0968*
P ₀	4.4500	3.3736077	1.3193	0.211698	-2.8993	11.7993

* $p < 0,05$ (estatisticamente significativo a 95% de confiança);

$R^2 = 0.92172$

MS Residual = 45.5106

Como pode ser observado da tabela acima, a concentração de produto final foi afetada pela temperatura, com efeito estatisticamente significativo. A Figura 5.1 mostra, em geral, que um aumento da temperatura leva a um conseqüente decréscimo na concentração final de

produto, sendo muito pronunciado este efeito negativo. Este resultado demonstra que o controle da temperatura é crítico, pois o seu aumento leva a perdas importantes na produção, podendo-se chegar a valores de até aproximadamente 40 g/l.

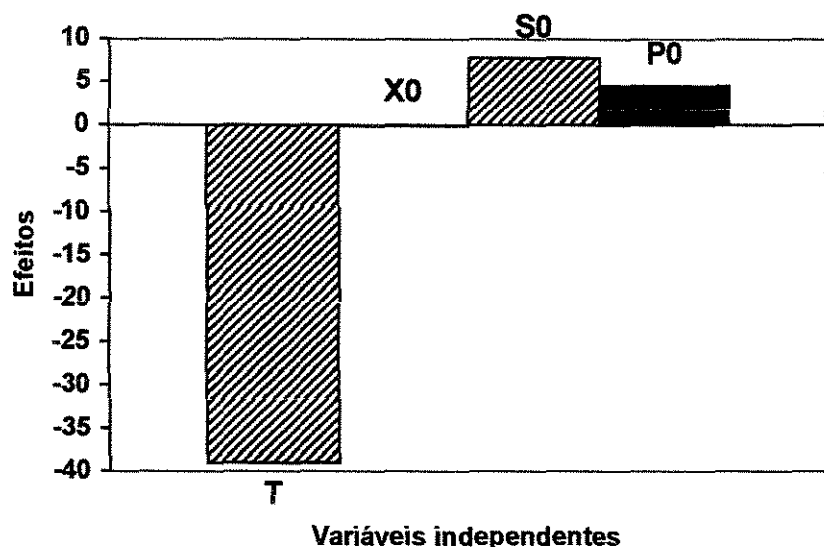


Figura 5.1 - Histograma dos efeitos principais das variáveis sobre a concentração de produto final.

A concentração inicial de substrato mostra-se como a segunda variável com maior efeito sobre a quantidade de produto final, Figura 5.1. Esta variável também apresenta resultados significativos do ponto de vista estatístico, pois como observado seu valor de p é menor do que 0,05. Seu efeito significa que quando a concentração inicial de substrato passa do nível mais baixo (-1) para o mais alto (+1) há um aumento de aproximadamente 8 g/l ou, em termos percentuais, em torno de 9% na concentração de produto final.

Muito embora este não seja o caso aqui, em *screening* de variáveis é provavelmente melhor aceitar valor de $p < 0,1$ do que deixar algum fator importante fora da análise ou modelo, vale ressaltar também que o bom senso para o julgamento prático do pesquisador deve ser o árbitro final para a interpretação das análises estatísticas.

As demais variáveis não se mostraram estatisticamente significativas a um nível de confiançaⁱ de 95%, segundo os valores apresentados para p ⁱⁱ, pois como os limites do intervalo de confiança possuem sinais contrários, e nenhum dos valores neste intervalo é mais provável do

ⁱ Nível de Confiança: ver descrição no item 8.4.

ⁱⁱ Nível- p : ver descrição no item 8.4.

que outro, pode muito bem ser que o verdadeiro valor dos coeficientes devidos a estas variáveis, seja zero. A literatura recomenda e geralmente considera significativos parâmetros ou variáveis com $p < 0,05$, pois pequenos valores para este parâmetro de análise estão associados a grandes valores de t^{iii} , isto implica que o coeficiente é muito maior que o erro padrão. Esta análise pode ser acompanhada na Tabela 5.3.

Os resultados para as variáveis estatisticamente significativas tornam possível definir as variáveis que possuem um efeito sobre a resposta de interesse, assim o planejamento fatorial completo pode ser elaborado para determinar a faixa ótima de operação para o aumento da concentração de etanol ao término da fermentação, combinando as variáveis selecionadas T e S_0 .

A conclusão de que X_0 não tem influência significativa na concentração de produto final, entretanto, somente é válida para a faixa estudada. Sabe-se, pelo conhecimento prévio que se tem do processo, que na faixa de valores usados industrialmente (em torno de 40g/l), esta variável influencia a concentração de produto no tempo final.

5.2 Determinação de um Modelo Empírico Usando Planejamento Fatorial

Da mesma forma como no item anterior, para se chegar ao modelo é necessário desenvolver um planejamento fatorial. Assim, este segundo planejamento foi conduzido para as duas variáveis selecionadas no planejamento anterior tendo como resposta a concentração final de produto, álcool, para a fermentação em batelada. Experimentos (simulações), portanto, foram planejados para fornecer como resposta um modelo quadrático, consistindo de 2^2 ensaios mais uma configuração estrela com ponto central. Novamente o *software* STATISTICA (Statsoft, v 5.0) foi utilizado para analisar os resultados.

A Tabela 5.4 apresenta a faixa das variáveis usadas para otimizar a concentração de produto final.

A Tabela 5.5 apresenta o planejamento experimental e os resultados obtidos para a concentração de produto final. O ponto central é normalmente usado com repetição para a estimativa do erro puro. Neste caso, entretanto, somente um ponto foi usado por não existirem erros desta natureza em simulações computacionais, de acordo com KALIL et al. (2000). Os valores reais para os níveis dos fatores codificados também são apresentados nesta tabela. As

ⁱⁱⁱ t (Student): ver descrição no item 8.4.

demais variáveis controladas como concentração inicial de Biomassa (X_0) e Produto (P_0), foram fixados em 1.44 e 0 g/l respectivamente.

Tabela 5.4 – Variáveis e níveis para o planejamento fatorial completo.

Nível	T (°C)	S_0 (g/l)
-1.41	25.52	38.579
-1	28	80.000
0	34	180.000
+1	40	280.000
+1.41	42.49	321.421

Os dados foram analisados usando a regressão não-linear múltipla do *software* STATISTICA. O modelo empírico gerado a seguir (equação 5.1), representa a concentração de produto final como função das variáveis mais significativas.

$$P_F = -240,445 + 14,263 \cdot T - 0,217 \cdot T^2 + 1,083 \cdot S_0 - 0,001 \cdot S_0^2 - 0,012 \cdot S_0 \cdot T \quad (5.1)$$

A análise de variância (ANOVA), está colocada na Tabela 5.6 para a concentração final de produto.

Tabela 5.5 – Planejamento Experimental, fatorial completo. Segundo planejamento.

Ensaio	Níveis				P _F (g/l)	Tempo (h)*
	Codificado		Decodificado			
	T	S ₀	T (°C)	S ₀ (g/l)		
1	-1	-1	28	80.000	42.030	8.5
2	+1	-1	40	80.000	26.830	6.5
3	-1	+1	28	280.000	113.000	32.5
4	+1	+1	40	280.000	69.250	33.5
5	0	-1.41	34	38.579	17.730	3.5
6	0	+1.41	34	321.421	107.030	29.0
7	-1.41	0	25.52	180.000	91.140	24.5
8	+1.41	0	42.49	180.000	44.560	149.0
9 (Central)	0	0	34	180.000	82.270	13.5

* Tempo de fermentação segundo as condições iniciais de cada ensaio, e de acordo com o modelo de COSTA (2000) simulado.

Tabela 5.6 – Análise da Variância (ANOVA) para o ajuste de um modelo quadrático de duas variáveis, sem repetição no ponto central.

Fonte de variação		Soma Quadrática	Graus de Liberdade	Média Quadrática	F _{calc.} (calculado)	p
R E G R E S S Ã O	T	1947.631	1	1947.631	154.9036	.001118
	T ²	176.734	1	176.734	14.0564	.033135
	S ₀	7180.769	1	7180.769	571.1178	.000161
	S ₀ ²	322.524	1	322.524	25.6518	.014858
	S ₀ · T	203.776	1	203.776	16.2072	.027541
Resíduo		37.720	3	12.573		
Total		9698.269	8			
Coeficiente de Correlação: 0.99611						
F _{tab} (0.95,5,3): 9.01 ∴ F _{calc} / F _{tab} : 781.9368/9.01 = 86.8						

Um modelo estatístico para ser válido deve possuir uma correlação elevada e uma relação entre o F calculado e o tabelado superior a 4 vezes (ATALA, 2000). Na Tabela 5.6 isto pode ser observado, e o modelo estatístico está na equação 5.1. Esta tabela também apresenta os valores do nível p do modelo; pode ser visto que todos são inferiores a 0,05.

5.3. Metodologia das Superfícies de Resposta

A equação 5.1 é usada para determinar a melhor condição de operação do processo em batelada. A Figura 5.2 apresenta a superfície de resposta utilizando esta equação.

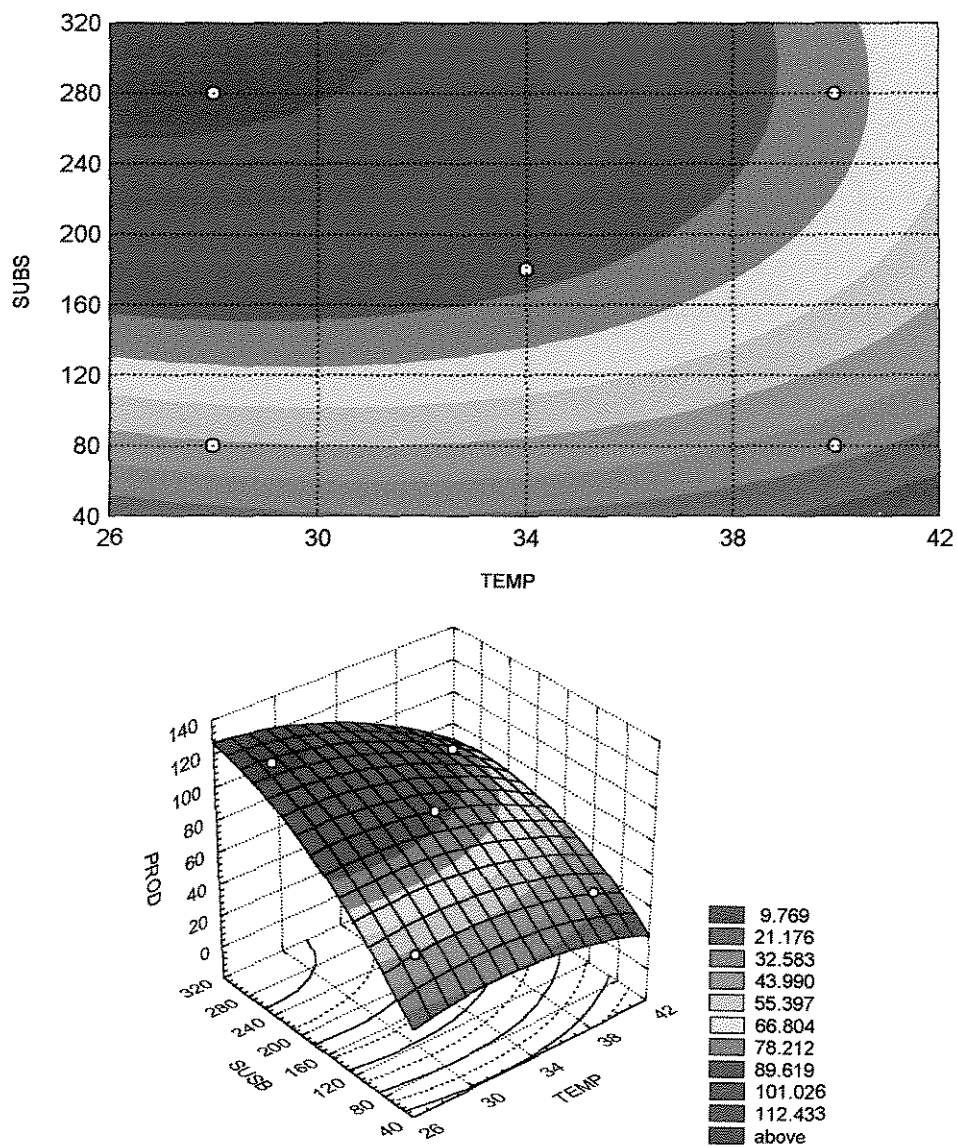


Figura 5.2 – Superfície de resposta e diagrama de contorno para a concentração de produto final (PROD, g/l) em função da concentração inicial de substrato (SUBS, g/l) e temperatura (TEMP, °C).

Podemos observar da Figura 5.2 que um aumento na concentração inicial de substrato pode levar a um aumento na quantidade final de etanol, da mesma forma um aumento da temperatura provoca o efeito contrário, conforme já era esperado pela análise dos efeitos das variáveis envolvidas no primeiro planejamento fatorial. Ainda de acordo com a Figura 5.2 podemos denotar que a região com a maior quantidade de produto final está entre 26 e 32°C e 260 a 320 g/l de S_0 . No entanto, esta não é a melhor condição, pois a quantidade, ou melhor, a concentração de substrato inicialmente necessária para a fermentação é muito alta em vista dos resultados obtidos, ou seja, no final da fermentação sobra uma quantidade de substrato não convertido a etanol. Isto pode ser atribuído a uma possível inibição pelo excesso de substrato presente inicialmente, assim, ainda que a concentração de produto final seja maior, este aumento não é proporcional à quantidade de substrato adicionado ao meio. Desta forma, a melhor condição de operação do sistema, pela Figura 5.2, está entre 26 e 32°C e 210 a 260 g/l de S_0 , uma vez que visualmente pode-se concluir que a partir de uma quantidade menor de substrato inicial produz-se quase a mesma quantidade de etanol. Segundo ALVES (1996), citado por ATALA (2000), a temperatura ótima de crescimento celular para a levedura *Saccharomyces cerevisiae* é ao redor de 32°C.

Outro fator igualmente importante nesta análise é o tempo de fermentação. Por exemplo, sabemos que a cinética microbiana é fortemente influenciada pela temperatura, assim aumentando este fator muito provavelmente o tempo de fermentação para uma determinada concentração de produto final irá se expandir. Com este intuito, uma superfície tendo como resposta o tempo em função da concentração inicial de substrato e temperatura está apresentada na Figura 5.3. A superfície de resposta a seguir não se trata de um modelo validado como a equação 5.1, mas apenas uma boa aproximação para ilustrar o tempo de fermentação médio em cada condição inicial de operação observado. A Figura 5.4 apresenta a superposição das Figuras 5.2 e 5.3, permitindo a avaliação da melhor condição inicial e conseqüente desempenho da fermentação.

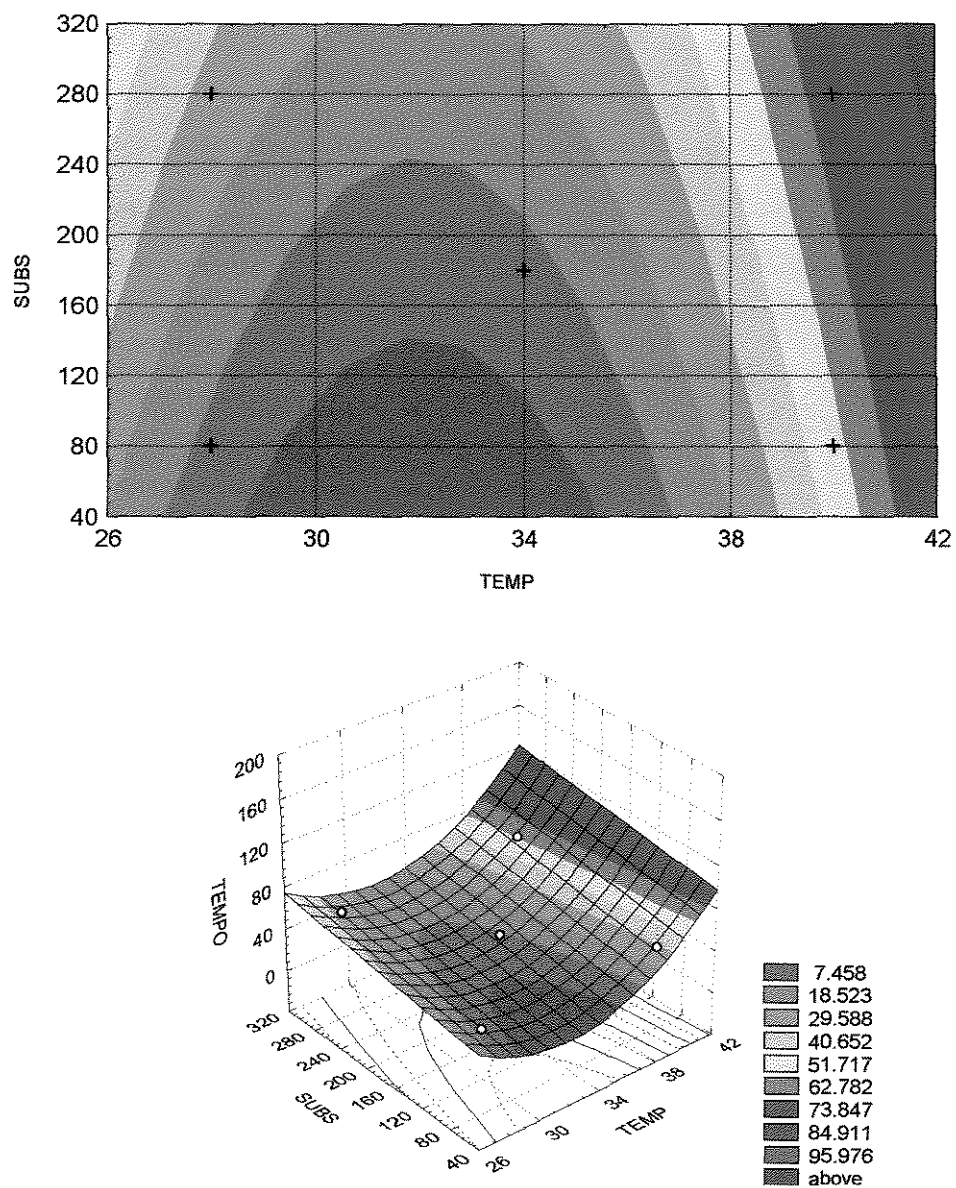


Figura 5.3 – Superfície de resposta e diagrama de contorno para o tempo de fermentação (TEMPO, horas) em função da concentração inicial de substrato (SUBS, g/l) e temperatura (TEMP, °C).

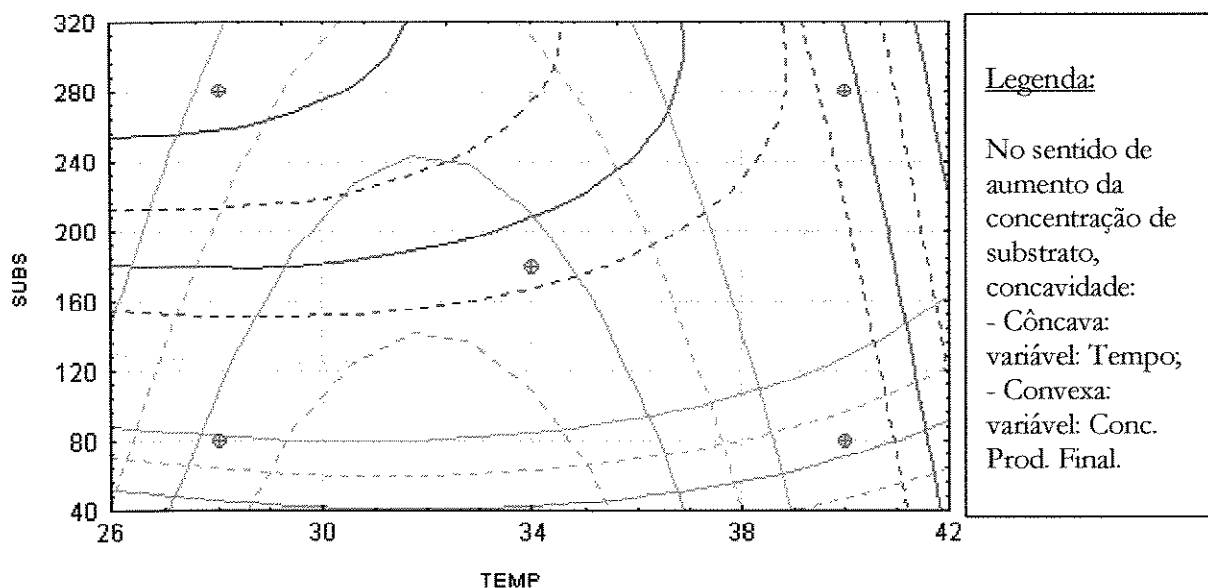


Figura 5.4 – Superposição dos diagramas de contorno das Figuras 5.2 e 5.3.

Com o efeito de comparação permitido na Figura 5.4, aliado a possível inibição pelo excesso de substrato inicial, conforme já discutido, pode-se destacar a expansão do tempo de fermentação para as concentrações de S_0 muito altas, mesmo dentro da faixa de temperatura já recomendadas, 26 a 32°C. Assim, tendo como objetivo a economia de matéria prima, ganho em tempo e maior quantidade de produto final, a região em torno do seguinte ponto oferece a melhor condição:

$$S_0 = 216 \text{ g/l}$$

$$T = 32^\circ\text{C}$$

Obedecendo a estas condições iniciais de operação, os resultados finais esperados estariam dentro da seguinte faixa, lembrando que os extremos abaixo são os valores médios de cada nível nas superfícies de resposta.

$$101,026 \leq P_f \text{ (concentração final de etanol, g/l)} \leq 112,43$$

$$7,46 \leq \text{Tempo da batelada de fermentação (horas)} \leq 18,52$$

Simulando a fermentação com o modelo matemático de COSTA (2000) e ATALA (2000) nas condições acima e mantendo as demais variáveis controladas, concentração de

biomassa e etanol inicial com os mesmos valores do planejamento experimental, 1,44 e 0 g/l respectivamente, foram obtidas as seguintes respostas.

Tempo de fermentação: 22,5 horas.

Concentração de produto final: 105,77 g/l.

A quantidade final de álcool está dentro da faixa esperada, conforme descreve o modelo validado, equação 5.1. O tempo, entretanto não está dentro da faixa esperada, isto pode ser explicado devido ao modelo empregado para descrever a superfície de resposta, da variável tempo, não ser validado, além de apresentar um coeficiente de correlação muito baixo, somente 0.62647. No entanto, mesmo assim fornece uma aproximação razoável, uma vez que, como já dito, a faixa dos valores esperados para o tempo de fermentação tem como limites os valores médios de cada nível da superfície de resposta.

5.4 Programação Quadrática Sucessiva (SQP)

A programação quadrática é um método da programação não-linear que minimiza (ou maximiza) uma função quadrática de n variáveis sujeita a m restrições, que podem ser de igualdade ou desigualdade. Para tanto a SQP usa a programação quadrática recursivamente (EDGAR & HIMMELBLAU, 1989).

A partir do modelo quadrático em função de duas variáveis, equação 5.1, a Programação Quadrática Sucessiva pôde ser empregada para encontrar o ponto de máxima concentração de álcool ao final da batelada.

O *software* Matlab[®] versão 5.3 foi usado, devido a sua versatilidade, para implementação da SQP. As equações a seguir são a função objetivo (equação 5.2) e suas restrições (equações 5.3).

Função Objetivo:

$$\max F(x_1, x_2) = -240,445 + 14,263 \cdot x_1 - 0,217 \cdot x_1^2 + 1,083 \cdot x_2 - 0,001 \cdot x_2^2 - 0,012 \cdot x_1 \cdot x_2 \quad (5.2)$$

Restrições:

$$\begin{aligned} 26 &\leq x_1 \leq 42 \\ 40 &\leq x_2 \leq 320 \end{aligned} \quad (5.3)$$

Onde x_1 refere-se a Temperatura (em °C) e x_2 a Concentração inicial de substrato (em g/l). As restrições consideram o intervalo de temperatura e concentração indicado a partir dos dados do planejamento experimental.

Segundo a sintaxe do Matlab as restrições devem ser transformadas em inequações *menor ou igual* a uma constante, dessa forma:

$$\begin{array}{l} x_1 \leq 42 \\ -x_1 \leq -26 \\ x_2 \leq 320 \\ -x_2 \leq -40 \end{array} \longrightarrow \begin{cases} 1 \cdot x_1 + 0 \cdot x_2 \leq 42 \\ -1 \cdot x_1 + 0 \cdot x_2 \leq -26 \\ 0 \cdot x_1 + 1 \cdot x_2 \leq 320 \\ 0 \cdot x_1 - 1 \cdot x_2 \leq -40 \end{cases} \quad (5.4)$$

Como mostrado acima, as inequações obtidas a partir das restrições podem ser formuladas como um sistema de inequações, e a partir disto, desde que as restrições ou seu sistema de inequações formado sejam lineares, pode ser estabelecida a matriz do sistema de inequações $A \cdot x \leq b$, equação 5.5.

$$A = \begin{bmatrix} 1 & 0 \\ -1 & 0 \\ 0 & 1 \\ 0 & -1 \end{bmatrix} \quad b = \begin{bmatrix} 42 \\ -26 \\ 320 \\ -40 \end{bmatrix} \quad (5.5)$$

A estimativa inicial para ser iniciada a rotina de otimização foi $T=34^\circ\text{C}$ e $S_0=180$ g/l. Estes valores foram escolhidos por estarem no centro dos intervalos estudados. Outros valores iniciais para a rotina de otimização foram igualmente testados, sempre levando ao mesmo resultado.

Fornecendo a matriz **A**, o vetor **b**, os valores para o ponto inicial e a função objetivo para o *software* Matlab, a rotina de otimização SQP chegou ao seguinte resultado:

$$T_{\max} = 26^\circ\text{C};$$

$$S_{0(\max)} = 320 \text{ g/l}.$$

A otimização com a Programação Quadrática Sucessiva obteve como resultados os limites fornecidos nas restrições. Este resultado já era esperado devido a forma tomada pela superfície do modelo quadrático maximizado. Isto pode ser concluído observando o diagrama de

contorno da Figura 5.2, onde o ponto de máxima concentração de produto final encontra-se no canto superior esquerdo, exatamente onde as variáveis independentes assumem os valores acima. Podemos então denotar que os dois métodos chegaram ao mesmo resultado.

A diferença entre o resultado apresentado nesta seção e aquele discutido como o melhor a partir da superfície de resposta (seção 5.2), está no fato de que o RSM permite avaliar o comportamento da função em todo o intervalo estudado e desta forma a melhor região de operação foi escolhida levando em consideração alguns fatores adicionais como bom senso e conhecimento do processo. Por exemplo, na seção anterior o ponto $T=32^{\circ}\text{C}$ e $S_0=216\text{ g/l}$ foi apontado como a melhor condição inicial de operação, porque a superfície permitia avaliar que não haveria um aumento proporcional entre a concentração inicial de Substrato e a quantidade final de etanol, provavelmente devido a inibição por excesso de substrato inicial.

Deve-se ressaltar que neste capítulo não se pretende chegar a um ponto ótimo de operação para o processo de produção de etanol e sim avaliar o uso de modelos obtidos por planejamento fatorial para a otimização usando programação quadrática sucessiva. Estes modelos são simples de desenvolver e tornam o espaço de busca menor do que se o método SQP fosse aplicado usando o modelo rigoroso (determinístico), o que diminui a probabilidade do SQP chegar a um ponto de máximo local. O método das superfícies de resposta, normalmente usado para otimização usando modelos gerados por planejamento fatorial, foi usado para comparação com os resultados do SQP. Na verdade, o ponto ótimo obtido não foi bom, já que a função objetivo escolhida para o estudo de otimização do processo não foi adequada. O modelo quadrático desenvolvido e aplicado à otimização não considera as variáveis tempo de fermentação e quantidade final de substrato não convertido a etanol. Uma forma de levar em conta estes aspectos seria otimizar duas funções objetivos diferentes, conversão e $P_F/\text{tempo de fermentação}$, ao invés de considerar somente a concentração de etanol no tempo final. ATALA (2000), trabalhando com processos fermentativos em batelada alimentada estendida, usou três variáveis para otimizar o processo, rendimento (%), conversão e produtividade (g/l/h). O autor, desta forma, pôde incluir em sua análise a variável tempo; e através do cruzamento das informações das superfícies obtidas para estas variáveis, foi capaz de estimar a melhor região para operação do processo considerado, indicando, inclusive, que altos valores destas variáveis são conseguidos quando se trabalha com temperaturas entre 28 a 34°C .

6 CONCLUSÕES

6.1 Ajuste dos Dados Experimentais e Treinamento das RNAs

Pôde ser observado o bom desempenho das redes neurais artificiais na descrição da cinética do processo estudado. Estas foram capazes de aproximar as taxas cinéticas daquelas obtidas pela derivada da função Sigmoidal-Boltzman, que tinha sido ajustada para modelar os dados experimentais.

Quanto a este ajuste pode-se concluir que:

- Devido à pequena quantidade de dados experimentais estes não puderam ser simplesmente discretizados para fornecerem as taxas cinéticas;
- Polinômios não ajustaram adequadamente os dados experimentais, principalmente quando foram necessárias as suas derivadas;
- A Função Sigmoidal-Boltzman ajustou-se muito bem aos dados experimentais.

As redes neurais mostraram-se de fácil implementação e treinamento e os seguintes pontos puderam ser observados durante este procedimento:

- Nenhuma combinação das funções de ativação estudadas evidenciou um melhor desempenho durante o treinamento;
- Os resultados de validação obtidos com as redes empregando a função Tansig na camada intermediária mostraram-se melhores do que aqueles com a função Logsig;
- Para o treinamento com os dados experimentais, as redes treinadas com apenas um neurônio na camada de saída forneceram tanto os melhores desempenhos de treinamento quanto apresentaram os melhores resultados para os testes de validação;

O seguinte aspecto foi igualmente avaliado durante a fase de treinamento das redes:

- Chegar a um número ótimo de neurônios na camada intermediária;

A quantidade ótima de neurônios empregados na camada intermediária pode ser estabelecida com o algoritmo de treinamento Bayesian *framework*, ficando entre 8 e 9 nodos.

6.2 Modelo Híbrido com Variáveis Primárias

Quanto ao modelo híbrido desenvolvido que utiliza as variáveis primárias como entrada (X , S e P), este se mostrou eficiente aproximando muito bem seus resultados dos dados experimentais. Seus principais aspectos observados foram:

- Aproximação muito boa para o modelo desenvolvido com o conjunto de treinamento onde nenhuma temperatura era excluída (TIPO I);
- Aproximação razoável para o modelo desenvolvido com o conjunto de treinamento onde a temperatura de 34°C era excluída, para posteriormente ser usada na validação e teste do modelo híbrido (TIPO II);
- Excluir uma dada informação do conjunto de treinamento, no caso uma temperatura, constitui uma exigência muito grande para a rede, uma vez que ela não reconhece o padrão daqueles dados;
- A rede neural artificial mostrou-se como uma ferramenta poderosa, pois foi capaz de aproximar razoavelmente dados desconhecidos do seu espaço geométrico.

6.3 Modelo Híbrido com Variáveis Secundárias

O modelo híbrido neural desenvolvido, que emprega as variáveis secundárias como pH, Turbidez e Brix à entrada das redes, apresentou os seguintes resultados:

- Os valores obtidos à saída foram satisfatórios e inclusive ligeiramente melhores quando comparados ao modelo híbrido Tipo I com variáveis primárias;
- As variáveis secundárias por serem de fácil medida podem ser adicionadas ao modelo em tempo real;
- Os modelos desenvolvidos com variáveis de fácil medida permitem um melhor acompanhamento do processo.

6.4 Otimização

- Os modelos gerados por planejamento fatorial são facilmente desenvolvidos e podem ser usados para otimização do processo usando a técnica de Programação Quadrática Sucessiva. No entanto, a função objetivo do problema de otimização deve ser escolhida com cuidado para se obterem resultados representativos;
- O método da Programação Quadrática Sucessiva mostrou-se de fácil implementação usando-se o *software* Matlab;
- O método SQP não permite uma avaliação do comportamento da função objetivo dentro do intervalo estudado;
- Pode-se concluir que o método das Superfícies de Resposta possibilita avaliar o comportamento da função em todo o intervalo estudado e desta forma permite que se leve em consideração o conhecimento prévio que se tem do processo. O método da Programação Quadrática Sucessiva, por outro lado, fornece um resultado pontual.

6.5 Sugestões

Para a realização de trabalhos futuros sugere-se:

- Otimização com duas funções objetivo: Conversão e Pf/tempo;
- Aplicar um controlador usando o modelo desenvolvido;
- Aplicar o modelo híbrido desenvolvido para acompanhar um processo real, principalmente o modelo usando as variáveis secundárias;

7 BIBLIOGRAFIA

- AL-HOMIDAN, S. Hybrid methods for solving the educational testing problem. *Journal of Computational and Applied Mathematics*, 91, p. 34-45, **1998**.
- ATALA, DANIEL IBRAIM PIRES. *Fermentação Alcoólica com Alta Densidade Celular: Modelagem Cinética, convalidação de Parâmetros e Otimização do Processo*. Campinas: Faculdade de Engenharia de Alimento, Universidade Estadual de Campinas, **2000**. 135 p. Dissertação (Mestrado)
- BARROS NETO, B. de, SCARMINIO, I. S., BRUNS, R. E. *Planejamento e Otimização de Experimentos*. Campinas: Ed. Da Unicamp, **1995**.
- BASHEER, I.A., HAJMEER, M. Artificial neural networks: fundamentals, computing, design, and application. *Journal of Microbiological Methods*, 43, p. 3-31, **2000**
- BERG, C. World Ethanol Production and Trade to 2000 and Beyond. <http://www.distill.com/berg/>, january **1999**
- BISHOP, C.M. Neural Networks and Their Applications. *Rev. Sci. Instrum.*; Vol. 65, n. 6, USA, June **1994**.
- BLAESI, J., JENSEN, B. Can Neural Networks Compete With Process Calculations. *Intech/ Applying Technology*, december, **1992**.
- BRAGA, A.P., LUDEMIR, T. B., CARVALHO, A. C. P. de L. F. *Redes Neurais Artificiais – Teoria e Aplicações*. Rio de Janeiro: LTC – Livros Técnicos e Científicos Editora S.A., **2000**, 262p.
- CAUDIEL, M. Neural Network Training Tips and Techniques, *AI Expert*, January **1991**.
- CHUAPRASERT, S., DOUGLAS, P., NGUYEN, M. Data reconciliation of an agitated thin film evaporator using Aspenplus. *Journal of Food Engineering*, 39, p. 261-267, **1999**.
- CLAUSET, C., LÜDERS, G. Seu Bolso pede Álcool. *Revista Quatro Rodas*, p. 82-85, junho **1999**.

- CONRADIE, A.E., BUYS, J.D., KRÖGER, D.G. Performance Optimization of Dry-Cooling Systems for Power Plants Through SQP Methods. *Applied Thermal Engineering*, vol. 18, nos. 1-2, p. 25-45, **1998**.
- CONTANT, S. *Predição e Controle da Formação de Hidratos em Dutos através da Aplicação de Redes Neurais Artificiais*, Prêmio Petrobrás de Tecnologia em Dutos, DPQ – FEQ – UNICAMP, **1999**.
- CONTANT, S., BATISTA, L.M.F.L., FERNANDES, F.A.N. *Aplicação de Redes Neurais Artificiais na Modelagem de Reator de Leito Fluidizado para Produção de Polietileno*, DPQ – FEQ – UNICAMP.
- COSTA, A.C., ATALA, D.I.P., MAUGERI, F., MACIEL FILHO, R. Factorial design and simulation for the optimization and determination of control structures for an extractive alcoholic fermentation. *Process Biochemistry*, 37, p. 125-137, **2001**.
- COSTA, A. C. *Controle Ótimo Adaptativo em Processos Fermentativos*. Rio de Janeiro, Brasil: PEQ/COPPE/UFRJ, **1998**. Tese (Doutorado)
- COSTA, A.C. *Determinação de Estruturas de Controle em um Processo de Fermentação Alcoólica*, Campinas, SP: Faculdade de Engenharia Química, Universidade Estadual de Campinas, **2000**. Projeto (Pós-doutoramento)
- COSTA, A.C., HENRIQUES, A.S.W., ALVES, T.L.M., MACIEL FILHO, R., LIMA, E.L. A Hybrid Neural Model for the Optimization of Fed-Batch Fermentations. *Brazilian Journal of Chemical Engineering*, v.16, p. 53-63, **1999**
- COTTIS, R. A., QING, L., OWEN, G., GARTLAND, S. J., HELLIWELL, I.A., TUREGA, M. Neural network methods for corrosion data reduction. *Materials and Design*, 20, p. 169-178, **1999**.
- CRUZ, A. J. G. da, HOKKA, C. O., GIORDANO, R.C. A Simulation of the Penicillin G Production Bioprocess Applying Neural Networks. *Brazilian Journal of Chemical Engineering*, vol 14, no. 4, December **1997**.

DIEGUEZ, C. Foi apenas um susto. *Revista Veja*, p. 130-131, 23/02/2000

DIETTERICH, T., KONG, E.B. Machine Learning Bias, Statistical Bias, and Statistical Variance of Decision Tree Algorithms. <http://citeseer.nj.nec.com/141236.html>, 1995

ECHEGARAY, O. F., CARVALHO, J. C. M., FERNANDES, A. N. R., SATO, S., AQUARONE, E., VITOLO, M. Fed-batch culture of *Saccharomyces cerevisiae* in sugar – cane blackstrap molasses: invertase activity of intact cells in ethanol fermentation. *Biomass and Bioenergy*, 19, p. 39-50, 2000.

EDGAR, T.F., HIMMELBLAU, D.M. *Optimization of Chemical Process*. New York: McGraw-Hill Book Company, 1989, 652 p.

FURTADO, M. R. Óleo Caro faz o Brasil cair no Álcool. *Química e Derivados*, p. 8-15, setembro 1999.

GOH, A. T. C. Back-propagation neural networks for modeling complex systems. *Artificial Intelligence in Engineering*, 9, p. 143-151, 1995.

GOUVÊIA, VERA LÚCIA REIS de. *Modelagem Dinâmica de Bioprocessos por Modelos Híbridos*. Campinas: Faculdade de Engenharia Química, Universidade Estadual de Campinas, 2000. 91 p . Dissertação (Mestrado)

HARADA, LAYSE HARUMI PEREIRA. *Modelagem híbrido Neuronal Aplicada a Processos Fermentativos*. Campinas: Faculdade de Engenharia Química, Universidade Estadual de Campinas, 2001. 75 p. Dissertação (Mestrado)

HAYKIN, S. *Neural Networks – A Comprehensive Foundation*. 2nd edition. New Jersey: Prentice-Hall, 1999, 842p

I, Y., WU, W., LIU, Y. Neural network modelling for on-line state estimation in fed-batch culture of L-lysine production. *The Chemical Engineering Journal*, 61, p. 35-40, 1996.

IBANOGLU, S., IBANOGLU, E. Modelling of natural fermentation in cowpeas using response surface methodology. *Journal of Food Engineering*, 48, p. 277-281, 2001.

- KALIL, S. J., MAUGERI, F., RODRIGUES, M. I. Response surface analysis and simulation as a tool for bioprocess design and optimization, *Process Biochemistry*, 35, p. 539-550, **2000**.
- KARIM, M. N., RIVERA, S. L. Comparison of Feedforward and Recurrent Neural Networks for Bioprocess State Estimations, *Proceedings of Escape*, S369-S377, **1992**.
- KASPARIAN, V., BATUR, C. Model reference based neural network adaptative controller. *ISA Transactions*, 37, p. 21-39, **1998**.
- KUTNEY, P. Pense Duas Vezes. *Revista Quatro Rodas*, p. 86-87, junho **1999**.
- LANOUELLE, R., THIBAUT, J., VALADE, J. L. Process modeling with neural networks using small experimental datasets. *Computers and Chemical Engineering*, 23, p. 1167-1176, **1999**.
- LAUER, W.; <http://williamlauer.hpg.com.br/neurais.htm>; **2000**;
- LEE, W., HUANG, C. Modeling of ethanol fermentation using *Zymomonas mobilis* ATCC 10988 grown on the media containing glucose and fructose. *Biochemical Engineering Journal*, 4, p. 217-227, **2000**.
- LINKO, P. ZHU, Y. Neural network programming in bioprocess variable estimation and state prediction. *Journal of Biotechnology*, 21, p., 253-270, **1991**.
- LINKO, S., LUOPA, J., ZHU, Y. Neural networks as 'softwares sensors' in enzyme production. *Journal of Biotechnology*, 52, p. 257-266, **1997**.
- MACKAY, D. J. C., Bayesian Interpolation, *Neural Computation*, vol. 4, n° 3, pp. 415-447, **1992**.
- MELEIRO, L. A. C., MACIEL FILHO, R., CAMPELLO, R. J. G. B., AMARAL, W. C. *Hierarchical Neural Fuzzy Models as a Tool for Process Identification: A Bioprocess Application*. In Application of Neural Network and Other Learning Technologies in Process Engineering. Imperial College Press. Mujtaba, I. M. and Hussain, M. A. (Editors), april **2001**.

- MONTAGUE, G., MORRIS, J. Neural –network contributions in biotechnology. *Tibtech*, vol 12, p. 312-324, august **1994**.
- OLIVEIRA, K.P.S., MORI, M., BRUNS, R.E., *Industrial Effluent Treatment System Modeling using PCA and Neural Networks*; DPQ - FEQ – UNICAMP; **1999**.
- OLIVEIRA, M. E. de, *Redes Neurais Artificiais*; www.geocities.com/CapeCanaveral/Runway/4303/; Brasil, **2000**.
- ÖZDEMİR, M., DEVRES, O. Analysis of color development during roasting of hazelnuts using response surface methodology. *Journal of Food Engineering*, 45, pp. 17-24, **2000**.
- PARKINSON, G., FOUHY, K., Advanced Process Control: Ingenious, yet Accessible. *Chemical Engineering*, p. 28-31, may **1995**.
- PHAM, D. T. *Neural Networks for Chemical Engineers*. Editor: A. B. Bulsari, **1995**, Cap. 1: An introduction to artificial neural networks, p. 3-19.
- PSICHOGIOS, C. D., UNGAR, L. H. A Hybrid Neural Network-First Principles Approach to Process Modeling. *AIChE Journal*, Vol. 38, No. 10, p. 1499-1511, October **1992**.
- RIBEIRO, S. K., ROSA, L. P. A. Activities implemented jointly and the use of fuel alcohol in Brazil for abating CO₂ emissions. *Energy Policy*, vol 26, no. 2, p. 103-111, **1998**
- RODRIGUEZ-TORAL, M.A., MORTON, W., MITCHELL, D.R. The use of new SQP methods for the optimization of utility systems. *Computers and Chemical Engineering*, 25, p. 287-300, **2001**.
- SAXÉN, B., SAXÉN, H. A Neural –Network Based Model of Bioreaction Kinetics. *The Canadian Journal of Chemical Engineering*, Vol 74, p. 124-131, February **1996**
- SCHENKER, B., AGARWAL, M. Cross-Validated Structure Selection for Neural Networks. *Computers and Chemical Engineering*, vol 20, no. 2, p. 175-186, **1996**.

- SHENE, C., DIEZ, C., BRAVO, S. Neural networks for the prediction of the state of *Zymomonas mobilis* CP4 batch fermentations. *Computers and Chemical Engineering*, (23), p. 1097-1108, **1999**.
- SHREVE, R. N., BRINK, J. A. *Indústrias de Processos Químicos*. Rio de Janeiro: Editora Guanabara Koogan, 4ª. edição, **1997**, 717 p.
- STUART, G., MACHADO, R., OLIVEIRA, J.V. de, ULLER, A.C., LIMA, E.L. Hybrid Artificial Neural Network Applied to Modeling SCFE of Basil and Rosemary Oils, *Ciência e Tecnologia de Alimentos*, Vol. 17, n.4, Brasil, **1997**.
- SYU, M., CHANG, J. B. Recurrent Backpropagation Neural Network Adaptive Control of Penicillin Acylase Fermentation by *Arthrobacter vicousus*. *Ind. Eng. Chem. Res.*, 36, p. 3756-3761, **1997**.
- SYU, M., TSAO, G. T. Neural network Modeling of Batch Cell Growth Pattern. *Biotechnology and Bioengineering*, vol 42, p. 376-380, **1993**.
- TEISSIER, P., PERRET, B., LATRILE, E., BARILLERE, J. M., CORRIEU, G. A hybrid recurrent neural network model for yeast production monitoring and control in a wine base medium. *Journal of Biotechnology*, (55), p. 157-169, **1997**.
- TEISSIER, P., PERRET, B., LATRILLE, E., BARILLERE, J. M., CORRIEU, G. Yeast concentration estimation and prediction with static and dynamic neural network models in batch cultures. *Bioprocess Engineering*, 14, p. 231-235, **1996**.
- THOLUDUR, A., RAMIREZ, W. F., McMILLAN, J. D. Interpolated parameter functions for neural network models. *Computer and Chemical Engineering*, 24, p. 2545-2553, **2000**.
- TSAPTSINOS, D. *Back-Propagation and its variations*; Neural Networks for Chemical Engineers, Chapter 3, A.B. Bulsari (Editor); Elsevier Science B. V.; **1995**.
- WHEALS, A. E., BASSO, L. C., ALVES, D.M.G., AMORIM, H. V. Fuel ethanol after 25 years. *Tibtech*, vol 17, p. 482-487, December **1999**.

YODA, K., FURUYA, T. *Neural Networks for Chemical Engineers*. Editor: A. B. Bulsari, **1995**, Cap. 20: Modeling Chemical Processes using Multiresolution Representation Neural Networks, p. 515-523.

YOU, Y., NIKOLAOU, M. Dynamic Process Modeling with Recurrent Neural Networks. *AIChE Journal*, vol 39, no. 10, p. 1654-1667, October **1993**.

8 APÊNDICES E ANEXOS

8.1 Métodos Analíticos

Os métodos analíticos utilizados foram compilados do trabalho COSTA (2000), pois não foram utilizados no presente trabalho e, portanto participam aqui somente como complemento necessário ao conhecimento da maneira pela qual os dados experimentais foram coletados.

8.1.1 *Concentração de células totais e viáveis:*

A concentração de células na base seca é medida por gravimetria. Amostras do meio fermentativo são retiradas e centrifugadas. O sobrenadante é separado e o precipitado, constituído de células, é lavado, transferido para um recipiente previamente pesado e, então, seco em estufa por aproximadamente 24 horas (até peso constante). Após este intervalo, a placa é novamente pesada e a razão entre a diferença dos pesos (placa com precipitado menos placa sem precipitado) e o volume da amostra inicial resulta na concentração celular em g célula seca/l.

A massa de células viáveis é dada por,

$$\text{massa viável (base seca)} = \text{massa total} \cdot \text{viabilidade}$$

8.1.2 *Viabilidade:*

A viabilidade celular é medida através do método da coloração das células mortas por corante azul de metileno conforme descrito por LEE et al (1981) citado por COSTA (2000). Este método baseia-se no princípio de que células mortas não conseguem destruir o corante que absorvem, tornando-se azuis. Uma amostra do meio é diluída com solução salina de Ringer e uma gota da solução resultante é colocada em uma câmara de Neubauer para contagem de células e levada ao microscópio. A viabilidade é calculada por,

$$\text{viabilidade(\%)} = \frac{\text{número de células brancas}}{\text{número de células totais (brancas e azuis)}}$$

8.1.3 Concentração de açúcares redutores totais (ART) e etanol:

As concentrações de açúcares redutores e de etanol foram determinadas através de análises em HPLC modelo Varian 9010. Para isso utilizou-se colunas da marca SHODEX modelo KS 801 na temperatura de 70°C e, como eluente, água deionizada e dareada com um fluxo de 0.50 ml/min. Foram usados como padrões soluções coquetel de sacarose, glicose, frutose, etanol e glicerol, nas concentrações de 0,01% a 4%, sendo o software Millennium® v.2.1 usado para integrar e quantificar os dados obtidos.

8.2 Tabelas

Tabela 8.1 - Dados experimentais do processo batelada, variáveis de estado e variáveis secundárias para a temperatura de 28°C.

TEMPO (h)	Variáveis Primárias ou de Estado			Variáveis Secundárias		
	X (g/l)	S (g/l)	P (g/l)	pH	Turbidez	Brix
0	1.952	181.7	11.7	4.3	1.1	22.0
2.5	2.5	171.5	13.55	4.3	1.5	21.8
4.5	3.603	168.1	14.94	4.3	1.9	20.0
6.5	4.695	149.1	21.26	4.41	2.7	19.0
7.5	5.259	140.1	27.108	4.4	3.4	18.8
9.5	6.903	112.7	33.32	4.28	4.5	17.0
11.5	9.11	84.5	41.48	4.07	6.0	16.0
15.5	11.008	38.7	59.78	3.85	8.8	13.0
19.5	11.842	15.4	74.44	3.8	12.2	10.5
21	13.163	12.4	82.184	3.81	12.1	10.0

X: Biomassa; S: Substrato; P: Produto Formado.

Tabela 8.2 - Dados experimentais do processo batelada, variáveis de estado e variáveis secundárias para a temperatura de 31°C.

TEMPO (h)	Variáveis Primárias ou de Estado			Variáveis Secundárias		
	X (g/l)	S (g/l)	P (g/l)	pH	Turbidez	Brix
0	1.88	212.94	2.651	4.73	1.1	24.4
2	2.581765	209.451	6.6	4.71	1.6	24
4	3.632414	189.465	15.942	4.72	2.2	23
6	5.217267	169.64	28.074	4.67	3.2	22
9	8.393585	118.25	46.548	4.4	5.2	19
11	9.674652	94.76	55.29	4.22	6.9	18
14	11.70492	52.45	66.2	4.07	9.8	15
16	12.79636	42.22	84.553	3.87	11.8	13
18	13.13314	30.85	93.621	3.95	13.4	12
21.5	13.05782	15.44	101.9	3.97	14.5	11

X: Biomassa; S: Substrato; P: Produto Formado.

Tabela 8.3 - Dados experimentais do processo batelada, variáveis de estado e variáveis secundárias para a temperatura de 34°C.

TEMPO (h)	Variáveis Primárias ou de Estado			Variáveis Secundárias		
	X (g/l)	S (g/l)	P (g/l)	pH	Turbidez	Brix
0	1.6	242.18	0.58	4.54	0.7	25
2.5	2.58	223.62	5.85	4.53	1.2	25
4.5	3.8	205.41	12.82	4.53	1.8	24
8	6.94	139.97	39.13	4.34	3.4	21
11	9.16	99.04	55.31	4.1	5.1	19
13	9.92	92.05	72.54	4.01	6.3	18
15	9.693642	59.56	81.35	3.99	7.5	16
17	9.986777	44.09	89.77	3.99	8.6	15
19.5	9.923909	38.81	102.30	4.02	9.6	12
23	9.762821	23.55	110.09	4.08	10.5	11

X: Biomassa; S: Substrato; P: Produto Formado.

Tabela 8.4 - Dados experimentais do processo batelada, variáveis de estado e variáveis secundárias para a temperatura de 37°C.

TEMPO (h)	Variáveis Primárias ou de Estado			Variáveis Secundárias		
	X (g/l)	S (g/l)	P (g/l)	pH	Turbidez	Brix
0	0.59	272.12	0	4.69	0.6	26
4.5	1.45	250.28	8.91	4.66	1.2	25
7.5	2.54	219.12	18.612	4.63	1.9	23.5
11	5.35	169.52	35.739	4.44	3.2	21
15	6.74	126.17	56.7	4.18	4.5	18
18.5	7.24	89.69	66.102	4.13	5.4	16
22	7.16965	73.03	76.84	4.14	6.4	15.5
25	7.26843	65.94	84.2	4.16	6.9	14
30	7.1345	46.61	89.68	4.11	7.9	13

X: Biomassa; S: Substrato; P: Produto Formado.

Tabela 8.5 - Dados experimentais do processo batelada, variáveis de estado e variáveis secundárias para a temperatura de 40°C.

TEMPO (h)	Variáveis Primárias ou de Estado			Variáveis Secundárias		
	X (g/l)	S (g/l)	P (g/l)	pH	Turbidez	Brix
0	1.16	244.41	0.000	4.62	1	22
3.5	2.09	217.83	5.038	4.6	1.4	21
8	3.75	162.52	19.816	4.52	2.2	19
12	4.47	125.1	39.350	4.42	2.9	17
15	4.95	107	45.770	4.35	3.2	16
18	5.08	91.266	50.805	4.33	3.7	15
21.5	4.924563	70.97	55.000	4.32	4.3	14
26	5.083839	65.46	63.798	4.31	4.8	13

X: Biomassa; S: Substrato; P: Produto Formado.

8.3 Algoritmo Backpropagation

Segundo TSAPTSINOS (1995), um sumário do algoritmo *Backpropagation* fica da seguinte forma:

- Inicializar todos os pesos;
- Para cada par de entrada-saída do conjunto de dados
- Apresente o vetor de entrada x para os nodos da camada de entrada. Cada entrada x corresponde a um nodo de entrada;
- Calcular a saída de cada nodo de entrada usando a seguinte equação:

$$z_k = x$$

Simplesmente, a saída de um nodo de entrada é a igual a entrada recebida;

- Calcular a saída de cada nodo intermediário, segundo a equação a seguir:

$$z_I = \frac{1}{1 + \exp\left(-\left(\sum_k z_k w_{kI}\right) + \theta_I\right)}$$

A saída de um nodo da camada intermediária é a transformada não-linear da soma de todos os produtos (entradas que chegam ao nodo provenientes das camadas anteriores multiplicados pelos pesos respectivos de cada entrada) mais o termo *bias*.

- Calcular a saída de cada nodo de saída usando a seguinte equação:

$$y_m = \frac{1}{1 + \exp\left(-\left(\sum_I z_I w_{Im}\right) + \theta_m\right)}$$

- Comparar as saídas produzidas com as saídas desejadas;
- Atualizar os pesos dos nodos através da fase *backward*;

8.4 Estatística

Significância Estatística (Nível – p): A significância estatística de um resultado é uma medida estimada do grau do que é “verdadeiro” (do senso da “representatividade da população”). Mais tecnicamente, o valor do nível-p representa um índice decrescente da confiança de um resultado. Quanto mais alto o nível-p, menos podemos acreditar que a relação observada entre as variáveis da amostra é um indicador confiável da relação existente entre as respectivas variáveis na população. Especificamente, o nível-p representa a probabilidade de erro que está envolvida em aceitar os resultados observados como válidos, ou seja, como “representativos da população”. Por exemplo, o nível-p de 0.05 (1/20) indica que existe uma probabilidade de 5% que a relação entre as variáveis dadas em uma amostra sejam fracassadas ou contrárias a relação esperada. Em muitas áreas de pesquisa, o nível-p de 0,05 é costumeiramente tratado como o nível de erro máximo aceitável.

Intervalo de confiança: Os intervalos de confiança para a média fornecem uma faixa de valores em torno da média onde esperamos que a “verdade” (população) média está localizada (com um determinado nível de certeza). Podem ser requisitados intervalos de confiança para qualquer nível-p; por exemplo, se a média de uma amostra é 23, e os limites superior e inferior, de $p=0.05$ do intervalo de confiança, são 19 e 27 respectivamente, então pode ser concluído que existe 95% de probabilidade que a média da população está entre 19 e 27. Se o nível-p for determinado para um valor menor, então o intervalo se tornaria maior aumentando a confiança dos valores estimados, e vice-versa. O tamanho do intervalo de confiança depende do tamanho da amostra e da variação dos valores dos dados. O cálculo dos intervalos de confiança assume que a variável está normalmente distribuída na população. Esta estimativa pode não se tornar válida se esta distribuição não acontece, a menos que o tamanho da amostra seja grande, por exemplo $n \geq 100$.

Distribuição de Student (t): A distribuição de Student é empregada para testar a significância do valor estimado para um determinado coeficiente e mede o “quão grande” é o efeito em relação ao erro padrão. O nível-p, probabilidade de significância, é baseado no t calculado.

$$t = \frac{\text{Valor do Coeficiente de Regressão}}{\text{Erro Padrão}}$$